ACS chemical biology

Letters

pubs.acs.org/acschemicalbiology

Comparative Genomics Guided Discovery of Two Missing Archaeal Enzyme Families Involved in the Biosynthesis of the Pterin Moiety of Tetrahydromethanopterin and Tetrahydrofolate

Valérie de Crécy-Lagard,*[,†] Gabriela Phillips,[†] Laura L. Grochowski,[‡] Basma El Yacoubi,[†] Francis Jenney,[§] Michael W. W. Adams,[∥] Alexey G. Murzin,[⊥] and Robert H. White[‡]

[†]Department of Microbiology and Department of Microbiology and Cell Science, University of Florida, P.O. Box 110700, Gainesville, Florida 32611-0700, United States

[‡]Department of Biochemistry (0308), Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, United States

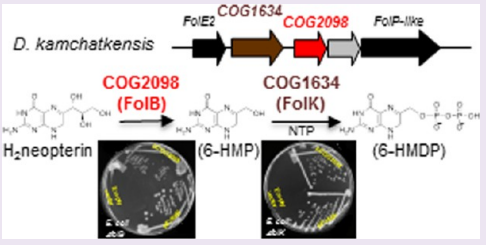
[§]Department of Basic Sciences, Georgia Campus, Philadelphia College of Osteopathic Medicine, Suwanee, Georgia 30024, United States

[∥]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602, United States

[⊥]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, U.K.

Supporting Information

ABSTRACT: C-1 carriers are essential cofactors in all domains of life, and in Archaea, these can be derivatives of tetrahydromethanopterin (H_4-MPT) or tetrahydrofolate (H_4-folate). Their synthesis requires 6-hydroxymethyl-7,8-dihydropterin diphosphate (6-HMDP) as the precursor, but the nature of pathways that lead to its formation were unknown until the recent discovery of the GTP cyclohydrolase IB/MptA family that catalyzes the first step, the conversion of GTP to dihydroneopterin 2′,3′-cyclic phosphate or 7,8-dihydroneopterin triphosphate [El Yacoubi, B.; *et al.* (2006) *J. Biol. Chem.*, *281*, 37586−37593 and Grochowski, L. L.; *et al.* (2007) *Biochemistry 46*, 6658−6667]. Using a combination of comparative genomics analyses, heterologous complementation tests, and *in vitro* assays, we show that the archaeal protein families COG2098 and COG1634 specify two of the missing 6-HMDP synthesis enzymes. Members of the COG2098 family catalyze the formation of 6-hydroxymethyl-7,8-dihydropterin from 7,8-dihydroneopterin, while members of the COG1634 family catalyze the formation of 6-HMDP from 6-hydroxymethyl-7,8-dihydropterin. The discovery of these missing genes solves a long-standing mystery and provides novel examples of convergent evolutions where proteins of dissimilar architectures perform the same biochemical function.



The availability of over 3000 published genome sequences[1] has enabled the use of comparative genomic approaches to drive the biological function discovery process.[2,3] Classically, one used to link a gene with function by genetic or biochemical approaches, a lengthy process that often took years. Phylogenetic distribution profiles, physical clustering, gene fusion, coexpression profiles, structural information and other genomic or postgenomic derived associations can be now used to make very strong functional hypotheses that can then be quickly validated by simple genetic and/or biochemical tests.[4,5] The whole procedure can occur in just weeks, taking advantage of the constantly growing available postgenomic resources such as gene deletion or expression libraries.[5] Here, we illustrate this paradigm shift with the discovery of two archaeal protein families involved in the synthesis of 6-hydroxymethyl-7,8-dihydropterin diphosphate (6-HMDP), the precursor of the pterin containing moiety of the essential C_1-carriers tetrahydrofolate (H_4-folate) and tetrahydromethanopterin (H_4-MPT) (Figure 1). These enzymes had eluded classical genetic and biochemical approaches and had been missing for decades.[6]

Most organisms use H_4-folate (Figure 1) as the essential carrier of C_1 fragments in both anabolic and catabolic reactions. The known exceptions are the methanogenic Archaea that use H_4-MPT (Figure 1)[7] and methylotrophic bacteria that use dephospho-H_4-MPT.[8] The situation in Archaea is quite diverse. Halophilic Archaea such as *Halobacterium* species harbor folates.[9] Hyperthermophiles like *Pyrobaculum* or *Sulfolobus* species use C_1-carriers lacking the C-7 methyl group on the pterin as seen in methanopterin.[10] Methanogenic Archaea such as *Methanobacterium thermoautotrophicum* Δ*H* (now called *Methanobacterium thermoautotrophicus*) use H_4-MPT, whereas *Thermococcus litoralis*[11] and *Pyrococcus furiosus* use only a more exotic derivative of methanopterin containing poly-β-(1→4)-*N*-acetylglucosamine as side chains on their C_1-carrier coenzyme.[12−14] Certain Archaea such as *Methanosarcina barkeri* contain both H_4-MPT and H_4-folate derivatives.[15] *Sulfolobus*

Received: April 18, 2012
Accepted: August 29, 2012
Published: August 29, 2012

 © 2012 American Chemical Society

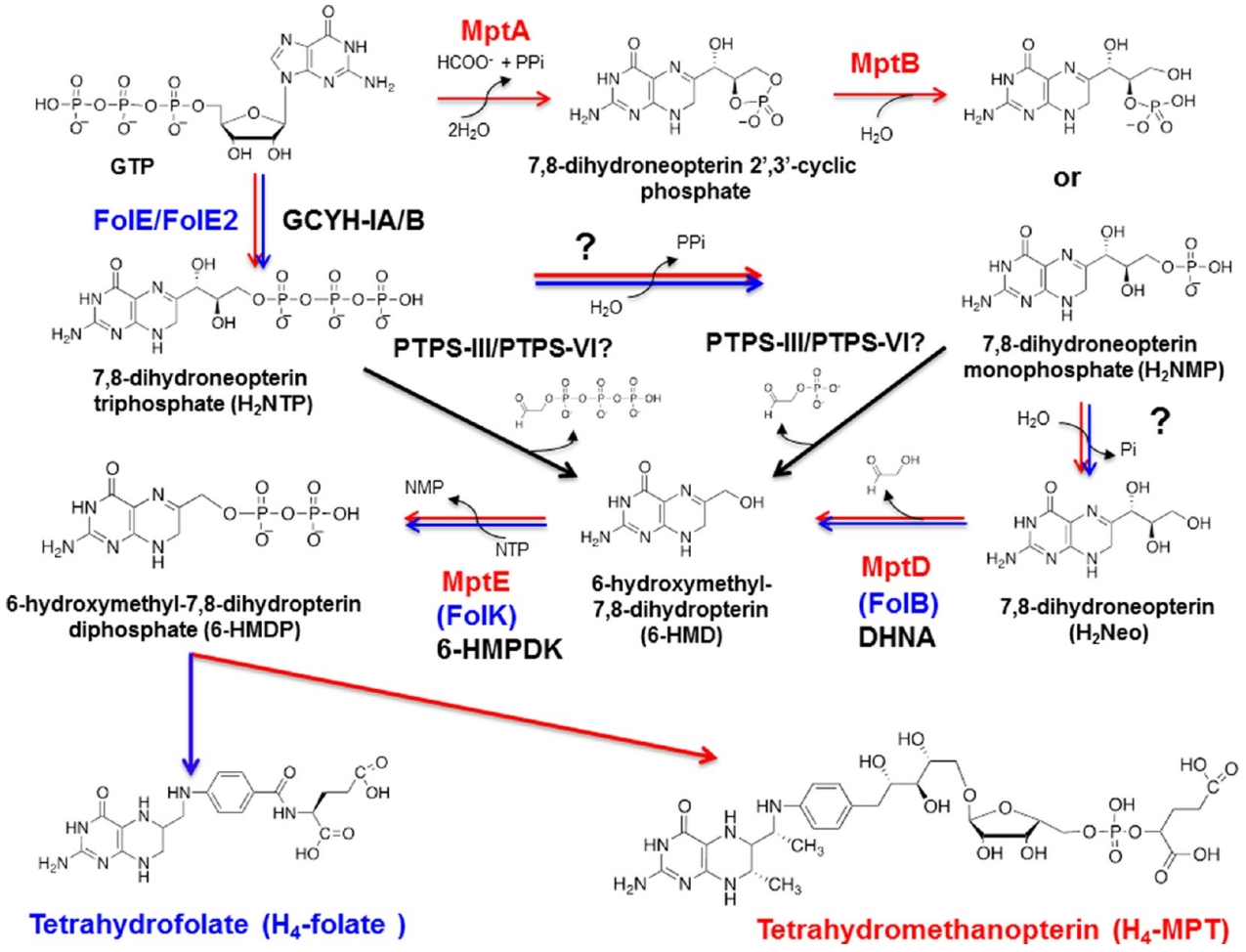


Figure 1. Early steps of tetrahydrofolate and tetrahydromethanopterin pathways in Bacteria and Archaea. Most bacteria use the FolE (or FolE2)/FolB/FolK route (in blue) to 6-HMDP even if some use the bacterial PTPS-III shunt (in green). Several routes to the common 6-HMDP intermediate in tetrahydrofolate and tetrahydromethanopterin are found in Archaea. A common pathway is the FolE2/MptD/MptE route (in red) such as in *H. volcanii* paralleling the bacterial pathway. However, some methanogens such as *M. jannaschii* use the MptA/MptB/MptD/MptE route, whereas *P. furiosus* uses the archaeal PTPS-III shunt. Phosphatases still to be identified are noted by a question mark (?). FolE/FolE2, GTP cyclohydrolase IA/IB (GCYH-IA/B); FolB, 7,8-dihydroneopterin aldolase (DHNA); FolK, 7,8-dihydro-6-hydroxymethylpterin diphosphokinase (6-HMDPK); MptA, archaeal GTP cyclohydrolase I (Fe(II)-dependent enzyme); MptB, Fe(II) dependent-cyclic phosphodiesterase; MptD, archaeal specific DHNA; MptE, archaeal specific 6-HMDPK; PTPS-III/PTPS-V/PTPS-VI, pyruvoyltetrahydropterin synthase paralogs involved in 6-HMDP synthesis.

solfataricus contains a hybrid coenzyme C_1-carrier coenzyme harboring a nonmethylated pterin and the same arylamine moiety found in methanopterin.[16] Although numerous variations in the C_1-carrier structures exist among the various archaeal lineages, the early steps in the syntheses of H_4-folate and of H_4-MPT and its derivatives, leading to the formation of the 6-HMDP intermediate, have been predicted to be similar [17] (Figure 1). The 6-HMDP pathway is well characterized in bacteria, plants, and fungi. GTP cyclohydrolase IA (GCYH-IA or FolE) or GTP cyclohydrolase IB (GCYH-IB or FolE2) catalyze the first step of the pathway producing 7,8-dihydroneopterin triphosphate (H_2NTP) from GTP.[18−20] H_2NTP produces 7,8-dihydroneopterin (H_2Neo) after the lost of a diphosphate and a phosphate. Then, 7,8-dihydroneopterin aldolase (DHNA) encoded in *Escherichia coli* by *folB*[21] catalyzes the formation of 6-hydroxymethyl-7,8-dihydropterin (6-HMD) from H_2Neo. A derivation from the classical bacterial 6-HMDP synthesis pathway occurs in *Plasmodium falciparum* and various bacteria. The DHNA step is bypassed by PTPS-III that cleaves the side chain of H_2NTP to form 6-HMD[22−24] (Figure 1). In all cases, 6-HMD is then diphosphorylated with ATP by a 7,8-dihydro-6-hydroxymethylpterin diphosphokinase (6-HMDPK), encoded in *E. coli* by *folK*[25] to form 6-HMDP.

Methanocaldococcus jannaschii was the first Archaea with a sequenced genome. It was immediately apparent that this organism lacked homologues of FolE, FolB, and FolK and used nonorthologous enzymes to catalyze the same reactions.[26] This prediction was confirmed as more archaeal genomes became available (Figure 2). As shown in Figure 2, a minority of Archaea (16 out of 58 analyzed) contained homologues of the canonical FolE and expression of the corresponding gene from *Sulfolobus solfataricus* P2 (*sso0364*) complemented the deoxythymidine (dT) auxotrophy of an *E. coli* Δ*folE* mutant.[27] Most Archaea (40/58 analyzed) contained homologues of the more recently discovered FolE2 (Figure 2) that were experimentally validated in a few species. The *folE2* mutant of *Haloferax volcanii* (Δ*HVO*_2348) is a dT and hypoxanthine auxotroph,[28] and the *M. jannaschii* FolE2 homologue MptA (MJ0775) is a unique Fe(II)-dependent GTP cyclohydrolase IB that forms

dx.doi.org/10.1021/cb300342u | *ACS Chem. Biol.* 2012, 7, 1807−1816

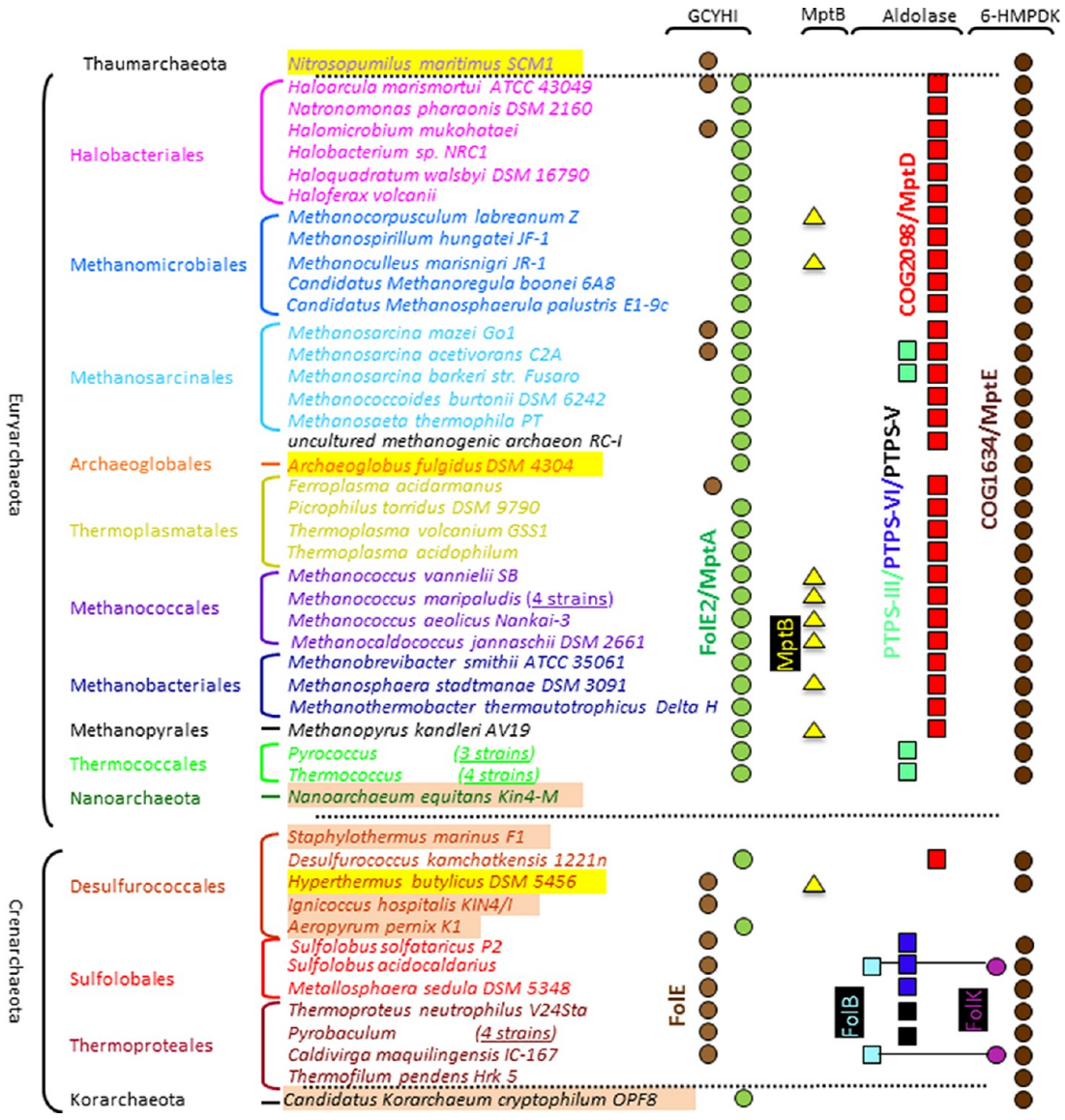


Figure 2. Phylogenetic distribution of predicted 6-HMDP synthesis genes in a subset of archaeal genomes. The presence of a symbol denotes the presence of a member of the protein family represented in that specific column in the genome covered in the corresponding line. Symbols and corresponding protein family are in the same color. Abbreviations have been defined in the Figure 1 legend. 6-HMDP synthesis genes might still be unidentified in organisms highlighted in yellow. Organisms highlighted in beige are most certainly pterin auxotrophs. Symbols linked by a line represent fused proteins. The full analysis is available in the Public SEED database in the Subsystem: "Pterin Biosynthesis Archaea".

7,8-dihydroneopterin 2′,3′-cyclic phosphate.[29] In *M. jannaschii*, MptB (MJ0837), a cyclic phosphodiesterase, is required to cleave the cyclic phosphate to form a mix of 7,8-dihydroneopterin 2′-monophosphate and 7,8-dihydroneopterin 3′-monophosphate.[30] This pathway, involving a 7,8-dihydroneopterin 2′,3′-cyclic phosphate intermediate, could be specific to a subset of methanogens because homologues of MptB are mostly found in Methanococcales (Figure 2). Even if the first archaeal 6-HMDP biosynthesis enzymes have been characterized, the remaining steps encoded in bacteria by FolB and FolK remain to be discovered in most Archaea. The identification and characterization of these missing gene families is the focus of this study.

Only two sequenced Archaea (*Sulfolobus acidocaldarius* DSM 639 and *Caldivirga maquilingensis* IC-167) contain homologues of bacterial FolB proteins fused with homologues of bacterial FolK proteins (Saci_1101 and Cmaq_0517, respectively) (Figure 2) that certainly derive from a lateral gene transfer event (the closest homologue to these two proteins is the fused FolKB from *Pneumocystis carinii* f. sp. macacae (AAN38834.1) with a Blastp E-value of 9e-40). A few Archaea such as *P. furiosus* or *Methanosarcina barkeri* str. fusaro harbor proteins of the PTPS-III family (Figure 2) that in bacteria function in a DHNA bypass where H_2NTP is converted directly to 6-HMDP[22−24] (Figure 1). Surprisingly, *in vitro*, the PTPS-III homologue from *P. furiosus*, PF1278, catalyzed the cleavage of 7,8-dihydroneopterin monophosphate (H_2NMP) to 6-HMD, but H_2NTP was not a substrate (Supporting Information and Figure 1). Finally, we recently showed that close homologues of PTPS-III with a slightly different active site motif named PTPS-VI[31] were found in a few *Sulfolobus* species (Figure 2). Expressing the PTPS-VI gene from *S. solfataricus* (*sso*2412) partially complemented the dT auxotrophy of a Δ*folB E. coli* mutant[31] suggesting a role of PTPS-VI proteins in 6-HMDP

dx.doi.org/10.1021/cb300342u | *ACS Chem. Biol.* 2012, 7, 1807−1816

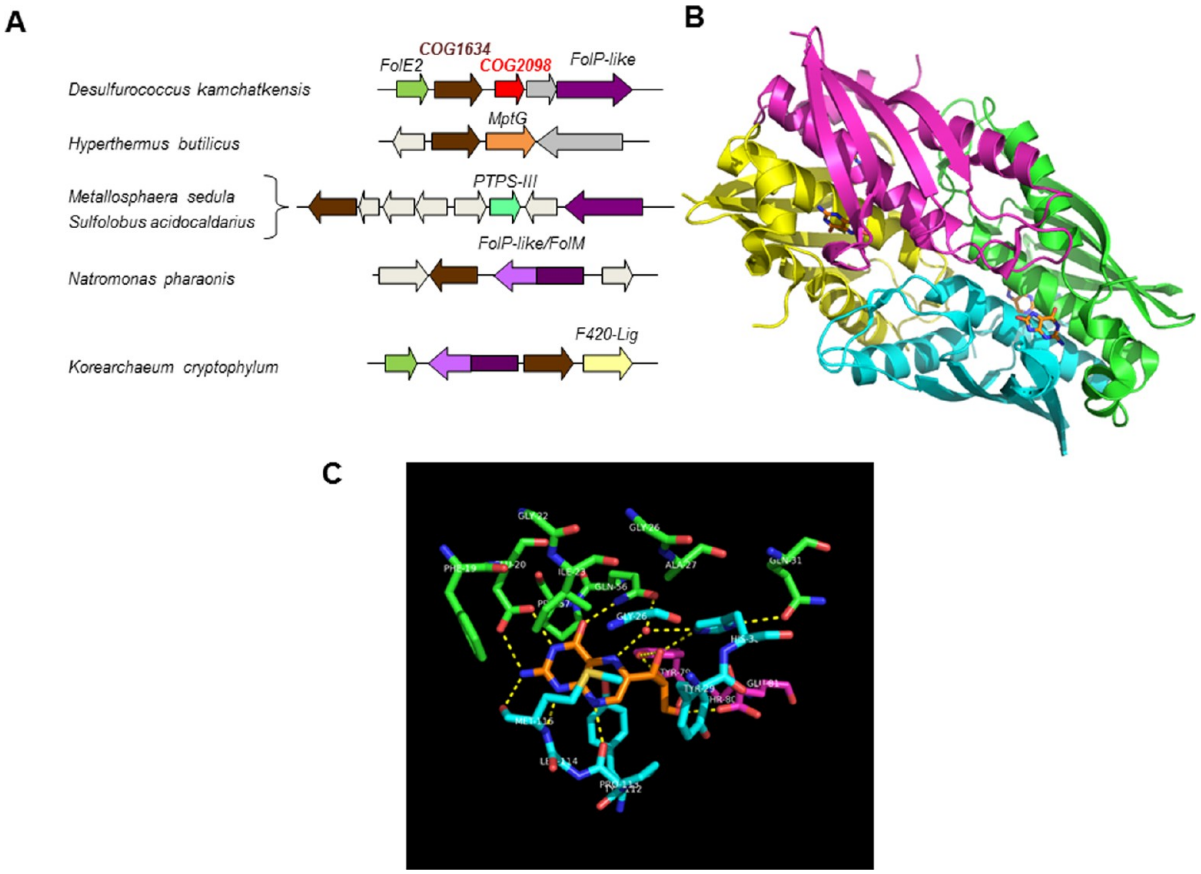


Figure 3. Comparative genomic evidence. (A) Clustering of COG1634 and COG2098 genes with pterin and cofactor biosynthetic related genes. Abbreviation not found in the text: FolP-like, dihydropteroate synthase-like enzyme homologous to the bacterial folate enzyme FolP but of unknown function;[59] MptG, β-ribofuranosylaminobenzene 5′-phosphate synthase;[60] FolM, alternative dihydrofolate reductase;[61] F420-lig, coenzyme F420-0: L-glutamate ligase.[62] (B) The archaeal DHNA (MptD) tetramer with bound pterin ring mimic (PDB 2OGF). The individual subunits of MJ0408 are shown with differently colored cartoons, the bound ligand 8-oxoguanine with orange carbons. (C) Putative active site of the archaeal DHNA with manually docked neopterin. The MptD structure is from PDB 2IEC, and the neopterin ligand (orange carbons) is from PDB 2O90 (in alternative conformation B). The active site residues contributed by three different subunits are shown with green, cyan, and magenta carbons, respectively (as in panel B and Supplemental Figure 3).

synthesis even if the substrate specificity of this family is yet to be experimentally determined (Figure 1). In summary, 56 out of 58 of the archaeal genomes analyzed lacked a FolK homologue, and 47 out of 58 lacked a FolB, PTPS-III, or PTPS-VI homologue. Hence, we set out to identify these missing archaeal 6-HMDP synthesis enzymes using a combination of comparative genomic approaches.

We first searched for genes that physically clustered with pterin related genes using the clustering tool of the SEED platform[2] and identified the COG1634/DUF115 gene family as a candidate (Figure 3A). Members of COG1634 are uncharacterized proteins found in most Archaea (Figure 2) and are part of the thiamin pyrophosphokinase (TPK, thiamin diphosphokinase) catalytic domain superfamily.[32] TPK is a thiamin salvage enzyme that transfers the diphosphate group of ATP to thiamin to form thiamin diphosphate, the active form of the cofactor.[33] TPK consists of two domains: the N-terminal catalytic domain that binds ATP and the C-terminal substrate-binding domain that binds thiamin.[34−37] The COG1634 family members show sequence similarity to the TPK catalytic domain but not to the C-terminal domain (Supplemental Figure 1). Moreover, fold recognition servers, *e.g.*, FFAS,[38] predict the TPK catalytic domain being a good template for the COG1634 subunit fold; they also suggest even higher-scoring hits to the structures of bacterial GST-II-like sialyltransferases[39] (Supplemental Figure 1). These bacterial enzymes use CMP-NeuAc as a sugar donor and have a distinct fold that is also found in mammalian sialyltransferases.[40] Comparison of the sialyltransferase and TPK structures revealed a common structural core and similar binding modes of their respective products, CMP and AMP, suggesting a distant evolutionary relationship of these protein families (Supplemental Figure 2). The COG1634 members are predicted to share the NMP-binding site (Supplemental Figure 1). In addition, COG1634 and TPK family members share the metal (Mg(II)) ion-binding site, involved in binding and transfer of the diphosphate group (Supplemental Figures 1 and 2). On the basis of the physical clustering evidence and fold homology, we predicted that COG1634 was the missing archaeal 6-HMDPK family. The homology with the sialyltransferase family opens the possibility that members of the COG1634 family may utilize other nucleoside triphosphates, *e.g.*, CTP. There are documented cases in archaeal biosynthetic pathways where CTP substitutes for ATP; for example, the Archaeon-specific riboflavin kinase uses CTP as its phosphoryl donor,[41] and the archaeal FAD synthetase (RibL) catalyzes the cytidylation of FMN with CTP.[42]

dx.doi.org/10.1021/cb300342u | *ACS Chem. Biol.* 2012, 7, 1807−1816

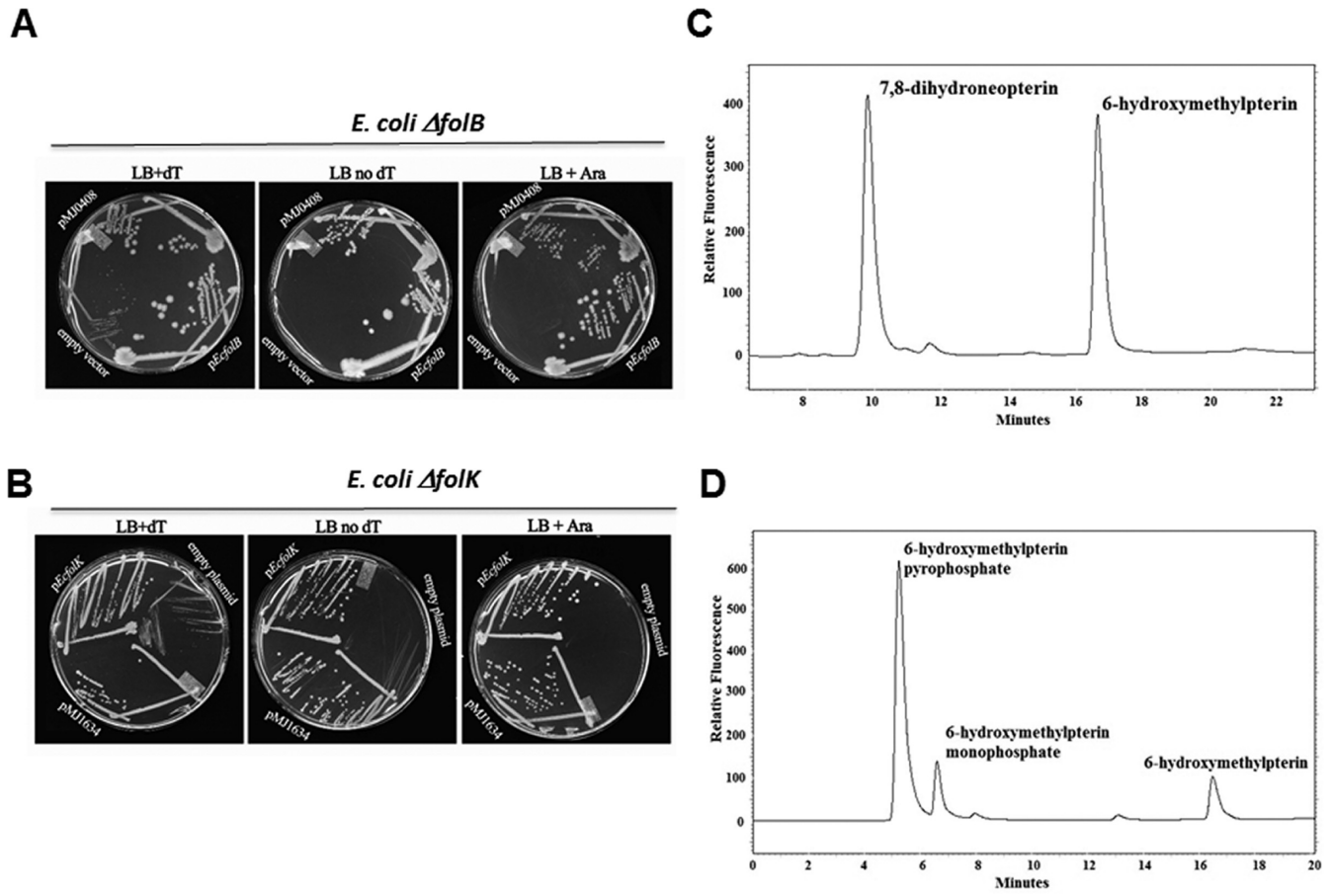


Figure 4. Experimental validations. (A) The MG1655 Δ*folB*::Kan^R strain (VDC3276) was transformed with the empty vector pBAD24 (top), p$folB_{Ec}$ (middle), or pMJ0408 (bottom). The resulting strains were plated on LB (with appropriate antibiotics), LB supplemented with arabinose (0.2%), or LB supplemented with dT and grown for 48 h. (B) The C600 Δ*folK*::*tetB* [45] strain was transformed with empty plasmid pBAD24 (top), p$folK_{Ec}$ (middle), or pMJ1634 (bottom). The resulting strains were plated on LB (with appropriate antibiotics) with or without dT supplementation and grown for 48 h. In both cases, complementation of the dT auxotrophy phenotypes by the archaeal clones were observed even in the absence of the inducer, arabinose. (C) Purified MJ0408 derived protein was incubated with H_2Neo as described in the methods section. After incubation, the sample was oxidized with iodine to convert the dihydropterins to the fluorescent pterins and assayed by HPLC with fluorescence detection. The first peak to elute at ∼10 min was neopterin, and the second at ∼16.5 min was 6-hydroxymethylpterin. The figure shows an assay where about half of the substrate was converted into product. No product was observed at zero time or in an assay run without added enzyme. The MonoQ fraction of the purified enzyme produced from an *E. coli* extract not expressing the MJ0408 derived protein likewise did not show any activity. (D) Purified PF0930 derived protein was incubated with 6-HMD and ATP as described in the methods section. After incubation, the sample was oxidized with iodine to convert the dihydropterins to the fluorescent pterins and assayed by HPLC with fluorescence detection. The first peak to elute at ∼5.2 min was 6-hydroxymethylpterin-PP, the second at ∼6.6 min was 6-hydroxymethylpterin-P, and the third at 16.5 min was 6-hydroxymethylpterin. The figure shows an assay where about 90% of the substrate was converted into product. The origin of the 6-hydroxymethylpterin-P is not clear but could arise from the hydrolysis of the 6-hydroxymethylpterin-PP during sample preparation. No product was observed at zero time or in an assay run without added enzyme. The MonoQ fraction of the purified enzyme produced from an *E. coli* extract not expressing the PF0930 derived protein likewise did not show any activity. Similar results were obtained with the MJ1634 protein.

We then observed that, in *Desulfurococcus kamchatkensis*, a gene in the COG2098 family was in a predicted operon with both the *folE2* and COG1634 genes (Figure 3A). Physical clustering in only one organism is not very strong evidence; nonetheless, further structural analysis suggested that COG2098 was the missing archaeal DHNA family. The COG2098 family previously was targeted by Structural Genomics Initiatives resulting in the determination of three representative structures: one from *Picrophilus torridus* (PTO0218; PDB: 2I52), one from *M. jannaschii* (MJ0408; PDB: 2OGF), and one from *Methanopyrus kandleri* (MK0786; PDB: 2IEC). The subunit fold comprising two α-helices and a four-stranded β-sheet somewhat resembles the fold of bacterial DHNA in architecture but differs from it in topology, as its secondary structure elements are connected in a very different order. Moreover, the COG2098 subunits assemble in a compact homotetramer, unlike the tunnel-like architectures of the canonical DHNA octomer. The tetramer is an apparent biological unit of the COG2098 family. The most conserved residues are scattered across the subunit surface but come together in the subunit interfaces. There are four equivalent putative active sites in the tetramer, each formed by the residues from three different subunits (Figure 3B). Fortuitously, in one of the determined structures, 2OHG, there is a ligand bound to each of the four sites that was tentatively identified as 8-oxoguanine (8-oxoG). Using the bound ligand as a guide, we manually docked the predicted substrate molecule in the COG2098 active site (Figure 3C). The dihydroneopterin molecule is in essentially the same conformation as in the structures of canonical DHNA complexes and fits almost

dx.doi.org/10.1021/cb300342u | *ACS Chem. Biol.* 2012, 7, 1807−1816

perfectly in the active site pocket when its pterin ring is aligned with the 8-oxoG mimic. The environment of bound substrate is also similar to that of the canonical FolB, suggesting a similar enzymatic mechanism for the predicted archaeal DHNA.[43]

A genetic approach was first used to validate these predictions. The COG2098 and the COG1634 encoding genes from *M. jannaschii* and represented, respectively, by *mj*0408 and *mj*1634, were cloned into pBAD24 under the P_{BAD} promoter[44] and tested for complementation of the dT auxotrophy phenotypes of the Δ*folB*::Kan^R *E. coli* strain (VDC3267[31]) and of the Δ*folK*::Tet^R *E. coli* strain (C600 Δ*folK*::*tetB*[45]), respectively. *E. coli* strains deleted in H_4-folate biosynthesis genes can grow on rich medium if dT is added to the medium, albeit poorly because of the absence of formylation of the initiator tRNA.[46] As shown on Figure 4A, expression of *mj*0408 complemented the dT auxotrophy phenotype of the *folB* deletion as did expression of the *E. coli folB* positive control. Overexpression of *mj*0408 seemed to be toxic with cells showing better growth with no arabinose (Figure 4A). Similarly, overexpression of *mj*1634 complemented the dT auxotrophy phenotype of the *folK* deletion as did expression of the *E. coli folK* positive control (Figure 4B).

A biochemical validation strategy was then used to confirm the genetic results. *mj*0408, *mj*1634, and its homologue from *P. furiosis* (*pf*0930) were all expressed in *E. coli*. The respective gene products, MJ0408, MJ1634, and PF0930 were purified by heating the extract to 80 °C followed by anion exchange chromatography. The resulting proteins were greater than 95% pure as judged by polyacrylamide gel electrophoresis with coomassie staining (Supplemental Figure 4). The identity of the purified proteins was confirmed by MALDI MS of the tryptic-digested protein band. The protein product from *mj*0408 was confirmed to be a DHNA as it was found to catalyze the formation of 6-HMD from H_2Neo (Figures 1 and 4C). The retention time of the 6-hydroxymethylpterin was identical to that of the 6-hydroxymethylpterin standard under two separate chromatographic systems utilizing either a Varian PursuitXRs C18 column or a Varian Pursuit polyfluorophenyl (PFP) column. The formation of 6-hydroxymethylpterin was linear with respect enzyme concentration, and no product was observed in control samples that were incubated in the absence of enzyme. In order to confirm that the observed activity was due to the *mj*0408 gene product and not a result of the *E. coli* DHNA activity, the activity of a cell extract and identical purified fractions from *E. coli* expressing the *mj*0408 gene to those of *E. coli* expressing a different gene (*mj*0929) were compared. The DHNA activity was greater than 4-fold higher in cell extracts of *E. coli* expressing the *mj*0408 gene. The identical MonoQ-purified fraction of *mj*0929 exhibited no activity relative to the purified *mj*0408 gene product.

The protein product from *pf*0930 was confirmed to be a 6-HMDPK as it was found to catalyze the formation of 6-HMDP from 6-HMD and ATP (Figures 1 and 4D). Of the four nucleotide phosphates tested, the maximum activity was observed with ATP. Relative to ATP, CTP, UTP, and GTP had, respectively, 41%, 40%, and 12% of the ATP activity. The fact that the oxidized product peak was 6-hydroxymethylpterin diphosphate (6-hydroxymethylpterin-PP) was confirmed by the following observations. The product peak had the same UV/visible absorbance and fluorescence spectra as a known sample of 6-hydroxymethylpterin-PP. Attempts to confirm the identity of the product peak by LC-ESI-MS analysis of crude incubation mixtures was not successful most likely due to ion suppression from the many salts in the sample. Thus, a reaction mixture was applied to a DEAE-Sephadex HCO_3^- column (2 × 5 mm), and the column was washed with 0.5 mL of 0.1 M NH_4HCO_3 and the 6-hydroxymethylpterin-PP eluted with 0.5 mL of 0.4 M NH_4HCO_3. After evaporation of the NH_4HCO_3, LC-ESI-MS analysis of this sample showed the expected MH^+ ion at 354.1 m/z and $(M - H)^-$ ion at 352.1 m/z for 6-hydroxymethylpterin-PP. The identity of the compound was further confirmed by the measurement of the MRM 354/125 and 354/176 fragments, the same as observed for the known sample of 6-hydroxymethylpterin-PP.

The combination of genetic and biochemical data presented here strongly validates our comparative genomic derived predictions solving the long-standing mystery of 6-HMDP biosynthesis in most Archaea. We therefore renamed the two families MptD for the archaeal specific DHNA and MptE for the archaeal specific 6-HMDPK. The discovery of the missing archaeal 6-HMDP synthesis genes completes the picture for the initial steps in the pterin pathways in the third kingdom of life (Figures 1 and 2). A great diversity in the metabolic and enzymatic solutions used to produce this molecule is observed, and the picture may be even more complex as some genes are still missing in specific lineages. MtpE homologues are found in almost all archaeal genomes (Figure 2). The handful of organisms that lack this gene like *Nanoarchaeum equitans*, or *Staphylothermus marinus* have lost all other 6-HMDP biosynthesis genes[47] and must certainly salvage pterin cofactors, even if one cannot rule out missing genes without further studies. For MptD, the situation is more complex. As discussed above, if the majority of archaeal genomes analyzed encode an MptD homologue, a minor subset does not. It seems most Thermococcales use a PTPS-III dependent bypass (Figures 1 and 2 and Supplemental Information), whereas *Sulfolobus solfataricus* uses a PTPS-VI dependent one.[31] A few archaeal species known to synthesize pterin containing C1-carrier coenzymes such as *Archaeoglobus fulgidus*[48] or *Pyrobaculum* species[10] lack homologues of MptD, FolB, PTPS-III, or PTPS-VI. A few of these (mainly Thermoproteales) encode members of the PTPS-V family (Figure 2), but initial validation tests with the *Pyrobaculum calidifontis* PTPS-V encoding gene, *Pcal*_1063, were negative.[31] Some archaeal genomes such as *M. barkeri* encode both MptD and PTPS-III homologues and others such as *S. solfataricus* encode both FolB and PTPS-VI homologues (Figure 2). Of note, these organisms synthesize both H_4-folate and H_4-MPT derivatives or hybrid molecules.[13,15,49] There are differences between the uses H_4-folate and H_4-MPT as C1 donors[7] since H_4-MPT derivatives do not form N^{10}-formyl derivatives as a result of thermodynamic differences in the chemical properties of the arylamine nitrogen N^{10} in H_4-folate and H_4-MPT. In organisms that use both cofactors, it might be necessary to use two different gene families in order to independently control the production of these molecules.

The discovery of MptD and MptE provides new examples of both divergent and convergent enzyme evolution.[50] Further characterization of these enzymes therefore will be of interest for structural biologists and biochemists. Finally, folate biosynthesis genes are traditional antibacterial targets and recently methanopterin biosynthesis has been proposed as a target to eliminate the dominant archaeon in the human gut *Methanobrevibacter smithii*.[51] Both MptD and MptE represent new targets found neither in human nor other members of the bacterial flora. As new roles of the gut Euryarchaeota emerge,[52]

inhibiting methanopterin pathway enzymes might be a viable solution to selectively eliminate Archaea from the flora.

METHODS

Bioinformatics. Analysis of the phylogenetic distribution and physical clustering was performed in the SEED database [2] on the 58 genomes available at the time of the analysis (Dec/2011). Results are available in the "Pterin Biosynthesis Archaea" subsystem on the public SEED server (http://pubseed.theseed.org/SubsysEditor.cgi). A subset of the analysis is summarized Figure 2. We also used the BLAST tools and resources at NCBI.[53] Multiple sequence alignments were built using the ClustalW tool.[54] Structure based alignments were performed using the Espript platform (http://espript.ibcp.fr/ESPript/ESPript/).[55] Visualization and comparison of protein structures and manual docking of ligand molecules were performed using PyMol (The PyMOL Molecular Graphics System, Version 1.4.1, Schrödinger, LLC).

Chemicals. 7,8-Dihydroneopterin, 6-hydroxymethylpterin-monophosphate (6-hydroxymethylpterin-P), 6-hydroxymethylpterin-PP, 6-hydroxymethyl-7,8-dihydropterin, 6-hydroxymethyl-7,8-dihydropterin-P, 6-hydroxymethyl-7,8-dihydropterin-PP, and D-neopterin were supplied by Schircks Laboratories, Jona, Switzerland. ATP, GTP, CTP, UTP, and 6-hydroxymethylpterin was supplied by Sigma. Recombinant NgFolE2 was supplied by Dirk Iwata-Reuyl, Department of Chemistry, Portland State University, Portland, Oregon.

***E. coli* Complementation Tests.** The *mj*0408 (NP_247382.1) and *mj*1634 (NP_248644.1) corresponding genes were amplified from *M. jannaschii* genomic DNA by PCR and cloned into the *Nco*I and *Sph*I sites of pBAD24 (Amp^R, ColE1)[44] after digestion with the corresponding enzymes to give plasmids pGPP528 (or pMJ0408) and pGPP541 (or pMJ1634), respectively. The resulting plasmids were verified by Sanger sequencing at the University of Florida core facility. Primers used were MjfolB2_Fwd (5′-CGTGACCATGGGAGTAGAAGAAACAGAAG-3′) and MjfolB2_Rev (5′-GGTCGGCATGCTTATTCCTCAAACTTTTTGACATAC-3′) for the *mj*0408 cloning and MJ1634pBAD24_Fwdol1 (5′-ATCGGCCATGGACATGAAGGAGTGGGA-3′) and MJ1634pBAD24_Revol2 (5′-GGTCGGCATGCTTATTTTAAAAATTCAATCTCTATTT-3′) for the *mj*1634 cloning. Positive controls were used as plasmids expressing the WT *E. coli folB* (pBAD24::$folB_{Ec}$, Amp^R, ColE1)[31] and the WT *E. coli folK* gene (p$folK_{Ec}$, ASKA clone JW0138[56]). *E. coli* derivatives were routinely grown at 37 °C in LB (BD Diagnostic System). Growth media were solidified with 15 g/L agar (BD Diagnostic System) for the preparation of plates. Transformations of *E. coli* were performed following standard procedures.[57] Ampicillin (Amp, 100 μg/mL), thymidine (dT, 80 μg/mL), chloramphenicol (20 μg/mL), kanamycin (Kan, 50 μg/mL), and L-arabinose (0.02% to 0.2%) were used as needed.

Cloning and expression of the *M. jannaschii mj0408* and *mj1634* and expression of their gene products. The *M. jannaschii* genes *mj*0408 and *mj*1634 was amplified from *M. jannaschii* genomic DNA by PCR. The primers used for *mj*0408 were *mj*0408Fwd (5′-GGTCATATGAGAGTAGAAGAAACAGAAG-3′) and MJ0408Rev (5′-GCTGGATCCTTATTCCTCAAACTTTTTGAC-3′). The primers used for *mj*1634 were *mj*1634Fwd (5′-GGTCATATGGACA-TGAAGGAGTG-3′) and *mj*1634Rev (5′-GCTGGATCCTTATTTTAAAAATTCAATCTC-3′). PCR amplification was performed using a 55 °C annealing temperature for *mj*0408 and 50 °C for *mj*1634. The PCR product was purified, digested with *Nde*I and *Bam*HI restriction enzymes, and then ligated into compatible sites in plasmid pT7−7 to make the recombinant plasmid pMJ0408 and pMJ1634. The sequences were verified by sequencing at the University of Iowa DNA core facility. The resulting plasmids were transformed into *E. coli* strain BL21-Codon Plus (DE3)-RIL (Stratagene). Transformed cells were grown in LB-medium (200 mL) supplemented with 100 μg/mL ampicillin at 37 °C with shaking until they reached an OD_{600} of 1.0. Recombinant protein production was induced by the addition of lactose to a final concentration of 28 mM. After an additional 4 h of culture at 37 °C, the cells were harvested by centrifugation (4000 × *g*, 5 min) and frozen at −20 °C. SDS-polyacrylamide gel electrophoresis (SDS-PAGE) of total cellular proteins confirmed induction of the desired protein at approximately 14 kDa for MJ0408 and 27.6 kDa for MJ1634.

Purification of Recombinant MJ0408 and MJ1634 Gene Products. Frozen *E. coli* cell pellets (~0.4 g wet weight from 200 mL of medium) were suspended in 3 mL of 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. Both protein products were found to remain soluble after heating the resulting cell extracts for 10 min at 70 °C followed by centrifugation (16 000 × *g* for 10 min). This process allowed for the purification of the desired enzymes from the majority of *E. coli* proteins, which denature and precipitate under these conditions. The next step of purification was performed by anion-exchange chromatography of the 70 °C soluble fractions on a MonoQ HR column (1 × 8 cm; Amersham Bioscience) using a linear gradient from 0 to 1 M NaCl in 25 mM Tris buffer (pH 7.5), over 55 min at a flow rate of 1 mL/min. Fractions of 1 mL were collected. The different purification steps are shown in Supplemental Figure 4 for MJ1634 as prototypical of the purification of all the proteins described herein. Protein concentrations were determined by Bradford analysis.

Cloning of the *P. furiosus pf*0930 and *pf*1278 Genes and Expression of Their Gene Products. The recombinant plasmids pPF0930 and pPF1278 were constructed as described in Sugar et al.[58] Briefly, the primers used for *pf*0930 were PF0930For (5′-CACCGGATCCAAGTGGGAGGAGTGGAAGCCATTC-3′) and PF0930Rev (5′-AAGCTCGAGCGGCCGCGATTTACGACTTGAGATAATAAAAAC-3′). The primers used for *pf*1278 were PF1278For (5′-CACCGGATCCAAGGCTAGGATTATCTATAGAGCT-3′) and PF1278Rev (5′-AAGCTCGAGCGGCCGCAGTGGTAAGGTCAAGGTGAGGTTTGA-3′). These primers were used to amplify the genes by PCR and cloned into a modified pET24d vector using *Bam*HI and *Not*I restriction enzymes. Proteins encoded by these plasmids were expressed and purified as described [58]

Enzymatic Assay of DHNA Activity. The standard assay used for the measurement of DHNA enzymatic activity was conducted in 200 μL reaction volume and included 7 ng of *M. jannaschii* MJ0408, 40 mM TES/KCl buffer pH 7.4, 8 mM $MgCl_2$, 16 mM DTT, and 110 μM H_2Neo. Samples were sealed under argon and incubated for 10 min at 70 °C. Following incubation, the reactions were quenched by the addition of 20 μL of 1 M HCl. H_2Neo and 6-HMD in the incubation mixture were oxidized to fluorescent neopterin and 6-hydroxymethylpterin by the addition of 8 μL of a saturated solution of iodine in methanol and incubated at RT for 30 min. Following oxidation, the samples were neutralized by the addition of 20 μL of 1 M NaOH and excess iodine removed by reduction with 8 μL of 1 M $NaHSO_3$. Following centrifugation, the samples were combined with water for a final volume of 1 mL and analyzed by HPLC as described below.

Enzymatic Assay of 6-HMDPK Activity. The standard assay used for the measurement of 6-HMDPK enzymatic activity was conducted in 200 μL reaction volume and included 3.5 ng of the PF0930 (or MJ1634) enzyme, 40 mM TES/KCl buffer pH 7.4, 8 mM $MgCl_2$, 16 mM DTT, 100 μM 6-HMD, and 1 mM ATP. Samples were sealed under argon and incubated for 10 min at 70 °C. Following incubation, the reactions were quenched by the addition of 20 μL of 1 M HCl. 6-HMD and 6-HMDP in the incubation mixture were oxidized to the fluorescent pterins by the addition of 8 μL of a saturated solution of iodine in methanol and incubated at RT for 30 min. Following oxidation, the samples were neutralized by the addition of 20 μL 1 M NaOH and excess iodine removed by reduction with 8 μL of 1 M $NaHSO_3$. Following centrifugation, the samples were combined with water for a final volume of 1 mL and analyzed by HPLC as described below.

HPLC Analysis of Pterins. Chromatographic separation of pterins was performed on a Shimadzu HPLC 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, pterins were eluted in the following order (min): 6-hydroxymethylpterin-PPP (4.982), 6-hydroxymethylpterin-PP (5.243), 6-hydroxymethylpterin-P (6.605), D-neopterin (10.10), monapterin (12.012), and 6-hydroxymethylpterin (16.490). Alternately, pterins were separated on a Varian Pursuit polyfluorophenyl (PFP) column (250 × 4.6 mm, 5 μm partical size). The elution profile was isocratic with 95% formic acid in water (0.1%) and 5% MeOH. Pterins were detected by fluorescence using an excitation wavelength of 356 nm and an emission wavelength of 450 nm.

■ ASSOCIATED CONTENT

Supporting Information

Ezymatic and HPLC assays. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: (352) 392 9416. Fax: (352) 392 5922. E-mail: vcrecy@ufl.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank G. Swedberg for the *E. coli* Δ*folK* strain. We thank M. Rasche and A. Hanson for frequent discussions and input and critical reading of the manuscript. We thank C. Brochier-Armanet for providing an archaeal genetic tree template. We thank D. Miller with helping to purify the *M. jannaschii* enzymes and D. Iwata-Reuyl for providing purified FolE2 from *Neisseria gonorrhoeae*. This research was supported in part by the U.K. Medical Research Council U1051922716 to A.G.M., the U.S. National Science Foundation MCB 0722787 to R.H.W., and the National Institutes of Health GM070641 to V.d.C.-L.

■ REFERENCES

(1) Liolios, K., Chen, I.-M. A., Mavromatis, K., Tavernarakis, N., Hugenholtz, P., Markowitz, V. M., and Kyrpides, N. C. (2010) The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res. 38*, D346−D354.

(2) Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A. C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G. D., Rodionov, D. A., Ruckert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., and Vonstein, V. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res. 33*, 5691−5702.

(3) Henry, C. S., Overbeek, R., Xia, F., Best, A. A., Glass, E., Gilbert, J., Larsen, P., Edwards, R., Disz, T., Meyer, F., Vonstein, V., DeJongh, M., Bartels, D., Desai, N., D'Souza, M., Devoid, S., Keegan, K. P., Olson, R., Wilke, A., Wilkening, J., and Stevens, R. L. (2011) Connecting genotype to phenotype in the era of high-throughput sequencing. *Biochim. Biophys. Acta 1810*, 967−977.

(4) Hanson, A. D., Pribat, A., Waller, J. C., and de Crécy-Lagard, V. (2009) "Unknown" proteins and "orphan" enzymes: the missing half of the engineering parts list and how to find it. *Biochem. J. 425*, 1−11.

(5) Blaby-Haas, C. E., and de Crécy-Lagard, V. (2011) Mining high-throughput experimental data to link gene and function. *Trends Biotechnol. 29*, 174−182.

(6) Grochowski, L. L.; White, R. H. (2010) Biosynthesis of the Methanogenic coenzymes. In *Comprehensive Natural Products II: Chemistry and Biology* (Begley, T., Ed.), pp 711−748, Elsevier, New York.

(7) Maden, B. E. (2000) Tetrahydrofolate and tetrahydromethanopterin compared: functionally distinct carriers in C1 metabolism. *Biochem. J. 350*, 609−629.

(8) Chistoserdova, L., Vorholt, J. A., Thauer, R. K., and Lidstrom, M. E. (1998) C1 transfer enzymes and coenzymes linking methylotrophic bacteria and methanogenic Archaea. *Science 281*, 99−102.

(9) Ortenberg, R., Rozenblatt-Rosen, O., and Mevarech, M. (2000) The extremely halophilic archaeon *Haloferax volcanii* has two very different dihydrofolate reductases. *Mol. Microbiol. 35*, 1493−1505.

(10) White, R. H. (1991) Distribution of folates and modified folates in extremely thermophilic bacteria. *J. Bacteriol. 173*, 1987−1991.

(11) White, R. H. (1993) Structures of the modified folates in the extremely thermophilic archaebacterium *Thermococcus litoralis*. *J. Bacteriol. 175*, 3661−3663.

(12) White, R. H. (1993) Structures of the modified folates in the thermophilic archaebacteria *Pyrococcus furiosus*. *Biochemistry 32*, 745−753.

(13) White, R. H. (1997) Purine biosynthesis in the domain Archaea without folates or modified folates. *J. Bacteriol. 179*, 3374−3377.

(14) Thauer, R. (1997) Biodiversity and unity in biochemistry. *Antonie van Leeuwenhoek 71*, 21−32.

(15) Buchenau, B., and Thauer, R. K. (2004) Tetrahydrofolate-specific enzymes in *Methanosarcina barkeri*; and growth dependence of this methanogenic archaeon on folic acid or *p*-aminobenzoic acid. *Arch. Microbiol. 182*, 313−325.

(16) Zhou, D., and White, R. H. (1992) 5-(*p*-Aminophenyl)-1,2,3,4-tetrahydroxypentane, a structural component of the modified folate in *Sulfolobus solfataricus*. *J. Bacteriol. 174*, 4576−4582.

(17) Graham, D. E., and White, R. H. (2002) Elucidation of methanogenic coenzyme biosyntheses: from spectroscopy to genomics. *Nat. Prod. Rep. 19*, 133−147.

(18) Yim, J. J., and Brown, G. M. (1976) Characteristics of guanosine triphosphate cyclohydrolase I purified from *Escherichia coli*. *J. Biol. Chem. 251*, 5087−5094.

(19) Nar, H., Huber, R., Auerbach, G., Fischer, M., Hosl, C., Ritz, H., Bracher, A., Meining, W., Eberhardt, S., and Bacher, A. (1995) Active site topology and reaction mechanism of GTP cyclohydrolase I. *Proc. Natl. Acad. Sci. U.S.A. 92*, 12120−12125.

(20) El Yacoubi, B., Bonnett, S., Anderson, J. N., Swairjo, M. A., Iwata-Reuyl, D., and de Crécy-Lagard, V. (2006) Discovery of a new prokaryotic type I GTP cyclohydrolase family. *J. Biol. Chem. 281*, 37586−37593.

(21) Garcon, A., Levy, C., and Derrick, J. P. (2006) Crystal structure of the bifunctional dihydroneopterin aldolase/6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from *Streptococcus pneumoniae*. *J. Mol. Biol. 360*, 644−653.

(22) Dittrich, S., Mitchell, S. L., Blagborough, A. M., Wang, Q., Wang, P., Sims, P. F. G., and Hyde, J. E. (2008) An atypical orthologue of 6-pyruvoyltetrahydropterin synthase can provide the missing link in the folate biosynthesis pathway of malaria parasites. *Mol. Microbiol. 67*, 609−618.

(23) Hyde, J. E., Dittrich, S., Wang, P., Sims, P. F., de Crécy-Lagard, V., and Hanson, A. D. (2008) Plasmodium falciparum: a paradigm for alternative folate biosynthesis in diverse microorganisms? *Trends Parisitol. 24*, 502−508.

(24) Pribat, A., Jeanguenin, L., Lara-Nunez, A., Ziemak, M. J., Hyde, J. E., de Crécy-Lagard, V., and Hanson, A. D. (2009) 6-Pyruvoyltetrahydropterin synthase paralogs replace the folate synthesis enzyme dihydroneopterin aldolase in diverse bacteria. *J. Bacteriol. 191*, 4158−4165.

(25) Talarico, T. L., Ray, P. H., Dev, I. K., Merrill, B. M., and Dallas, W. S. (1992) Cloning, sequence analysis, and overexpression of *Escherichia coli folK*, the gene coding for 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase. *J. Bacteriol. 174*, 5971−5977.

(26) Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., Weidman, J. F., Fuhrmann, J. L., and Venter, J. C. (1996) Complete genome sequence of the methanogenic archaeon *Methanococcus jannaschii*. *Science 273*, 1058−1073.

(27) Phillips, G., El Yacoubi, B., Lyons, B., Alvarez, S., Iwata-Reuyl, D., and de Crécy-Lagard, V. (2008) Biosynthesis of 7-deazaguanosine-modified tRNA nucleosides: a new role for GTP Cyclohydrolase I. *J. Bacteriol. 190*, 7876−7884.

(28) El Yacoubi, B., Phillips, G., Blaby, I. K., Haas, C. E., Cruz, Y., Greenberg, J., and de Crécy-Lagard, V. (2009) A gateway platform for functional genomics in *Haloferax volcanii*: deletion of three tRNA modification genes. *Archaea 2*, 211−219.

(29) Grochowski, L. L., Xu, H., Leung, K., and White, R. H. (2007) Characterization of an Fe^{2+}-dependent Archaeal-specific GTP Cyclohydrolase, MptA, from *Methanocaldococcus jannaschii*. *Biochemistry 46*, 6658−6667.

(30) Mashhadi, Z., Xu, H., and White, R. H. (2009) An Fe^{2+}-dependent cyclic phosphodiesterase catalyzes the hydrolysis of 7,8-dihydro-D-neopterin 2′,3′-cyclic phosphate in methanopterin biosynthesis. *Biochemistry 48*, 9384−9392.

(31) Phillips, G., Grochowski, L. L., Bonnett, S., Xu, H., Bailly, M., Blaby-Haas, C., El Yacoubi, B., Iwata-Reuyl, D., White, R. H., and de Crécy-Lagard, V. (2012) Functional promiscuity of the COG0720 family. *ACS Chem. Biol. 7*, 197−209.

(32) Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Lu, F., Marchler, G. H., Mullokandov, M., Omelchenko, M. V., Robertson, C. L., Song, J. S., Thanki, N., Yamashita, R. A., Zhang, D., Zhang, N., Zheng, C., and Bryant, S. H. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res. 39*, D225−D229.

(33) Nosaka, K., Kaneko, Y., Nishimura, H., and Iwashima, A. (1993) Isolation and characterization of a thiamin pyrophosphokinase gene, THI80, from *Saccharomyces cerevisiae*. *J. Biol. Chem. 268*, 17440−17447.

(34) Baker, L. J., Dorocke, J. A., Harris, R. A., and Timm, D. E. (2001) The crystal structure of yeast thiamin pyrophosphokinase. *Structure 9*, 539−546.

(35) Timm, D. E., Liu, J., Baker, L. J., and Harris, R. A. (2001) Crystal structure of thiamin pyrophosphokinase. *J. Mol. Biol. 310*, 195−204.

(36) Santini, S., Monchois, V., Mouz, N., Sigoillot, C., Rousselle, T., Claverie, J.-M., and Abergel, C. (2008) Structural characterization of CA1462, the *Candida albicans* thiamine pyrophosphokinase. *BMC Struct. Biol. 8*, 33.

(37) Liu, J. Y., and Hurley, T. D. (2011) A new crystal form of mouse thiamin pyrophosphokinase. *Int. J. Biochem. Mol. Biol. 2*, 111−118.

(38) Jaroszewski, L., Li, Z., Cai, X.-H., Weber, C., and Godzik, A. (2011) FFAS server: novel features and applications. *Nucleic Acids Res. 39*, W38−W44.

(39) Chiu, C. P., Watts, A. G., Lairson, L. L., Gilbert, M., Lim, D., Wakarchuk, W. W., Withers, S. G., and Strynadka, N. C. (2004) Structural analysis of the sialyltransferase CstII from *Campylobacter jejuni* in complex with a substrate analog. *Nat. Struct. Mol. Biol. 11*, 163−170.

(40) Rao, F. V., Rich, J. R., Rakic, B., Buddai, S., Schwartz, M. F., Johnson, K., Bowe, C., Wakarchuk, W. W., Defrees, S., Withers, S. G., and Strynadka, N. C. (2009) Structural insight into mammalian sialyltransferases. *Nat. Struct. Mol. Biol. 16*, 1186−1188.

(41) Mashhadi, Z., Zhang, H., Xu, H., and White, R. H. (2008) Identification and characterization of an Archaeon-specific riboflavin kinase. *J. Bacteriol. 190*, 2615−2618.

(42) Mashhadi, Z., Xu, H., Grochowski, L. L., and White, R. H. (2010) Archaeal RibL: a new FAD synthetase that is air sensitive. *Biochemistry 49*, 8748−8755.

(43) Wang, Y., Li, Y., and Yan, H. (2006) Mechanism of dihydroneopterin aldolase: functional roles of the conserved active site glutamate and lysine residues. *Biochemistry 45*, 15232−15239.

(44) Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J. Bacteriol. 177*, 4121−4130.

(45) Jönsson, M., and Swedberg, G. (2005) Hydroxymethyldihydropterin pyrophosphokinase from *Plasmodium falciparum* complements a *folK*-knockout mutant in *E. coli* when expressed as a separate polypeptide detached from dihydropteroate synthase. *Mol. Biochem. Parasitol. 140*, 123−125.

(46) Klaus, S. M., Kunji, E. R., Bozzo, G. G., Noiriel, A., de la Garza, R. D., Basset, G. J., Ravanel, S., Rebeille, F., Gregory, J. F., III, and Hanson, A. D. (2005) Higher plant plastids and cyanobacteria have folate carriers related to those of trypanosomatids. *J. Biol. Chem. 280*, 38457−38463.

(47) Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., Beeson, K. Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X., Mathur, E., Ni, J., Podar, M., Richardson, T., Sutton, G. G., Simon, M., SÃll, D., Stetter, K. O., Short, J. M., and Noordewier, M. (2003) The genome of *Nanoarchaeum equitans*: Insights into early archaeal evolution and derived parasitism. *Proc. Natl. Acad. Sci. U.S.A. 100*, 12984−12988.

(48) Gorris, L. G., Voet, A. C., and van der Drift, C. (1991) Structural characteristics of methanogenic cofactors in the nonmethanogenic archaebacterium *Archaeoglobus fulgidus*. *Biofactors 3*, 29−35.

(49) White, R. H. (1988) Analysis and characterization of the folates in the nonmethanogenic archaebacteria. *J. Bacteriol. 170*, 4608−4612.

(50) Galperin, M. Y., and Koonin, E. V. (2012) Divergence and convergence in enzyme evolution. *J. Biol. Chem. 287*, 21−28.

(51) Samuel, B. S., Hansen, E. E., Manchester, J. K., Coutinho, P. M., Henrissat, B., Fulton, R., Latreille, P., Kim, K., Wilson, R. K., and Gordon, J. I. (2007) Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. *Proc. Natl. Acad. Sci. U.S.A. 104*, 10643−10648.

(52) Horz, H.-P., and Conrads, G. (2010) The discussion goes on: what is the role of *Euryarchaeota* in humans? *Archaea 2010*, 967271.

(53) Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol. 215*, 403−410.

(54) Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., and Thompson, J. D. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res. 31*, 3497−3500.

(55) Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics 15*, 305−308.

(56) Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., and Mori, H. (2005) Complete set of ORF clones of *Escherichia coli* ASKA library (A complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res. 12*, 291−299.

(57) Sambrook, J., Fritsch, E. F., Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

(58) Sugar, F., Jenney, F., Poole, F., Brereton, P., Izumi, M., Shah, C., and Adams, M. (2005) Comparison of small- and large-scale expression of selected *Pyrococcus furiosus* genes as an aid to high-throughput protein production. *J. Struct. Funct. Genomics 6*, 149−158.

(59) Xu, H., Aurora, R., Rose, G. D., and White, R. H. (1999) Identifying two ancient enzymes in Archaea using predicted secondary structure alignment. *Nat. Struct. Mol. Biol. 6*, 750−754.

(60) Scott, J. W., and Rasche, M. E. (2002) Purification, overproduction, and partial characterization of β-RFAP synthase, a

key enzyme in the methanopterin biosynthesis pathway. *J. Bacteriol. 184*, 4442−4448.
(61) Levin, I., Giladi, M., Altman-Price, N., Ortenberg, R., and Mevarech, M. (2004) An alternative pathway for reduced folate biosynthesis in bacteria and halophilic archaea. *Mol. Microbiol. 54*, 1307−1318.
(62) Li, H., Graupner, M., Xu, H., and White, R. H. (2003) CofE catalyzes the addition of two glutamates to F420-0 in F420 coenzyme biosynthesis in *Methanococcus jannaschii*. *Biochemistry 42*, 9771−9778.

NOTE ADDED AFTER ASAP PUBLICATION

This paper was published ASAP on September 7, 2012. Figure 4 legend has been updated. The revised version was posted on September 12, 2012.