GTP cyclohydrolase I (GCYH-I) is the first enzyme of the de novo tetrahydrofolate biosynthetic pathway present in bacteria, fungi, and plants, and encoded in *Escherichia coli* by the *folE* gene. It is also the first enzyme of the bioppterin (BH4) pathway in *Homo sapiens*, where it is encoded by a homologous *folE* gene. A homology-based search of GCYH-I orthologs in all sequenced bacteria revealed a group of microbes, including several clinically important pathogens, that encoded all of the enzymes of the tetrahydrofolate biosynthesis pathway but GCYH-I, suggesting that an alternate family was present in these organisms. A prediction based on phylogenetic occurrence and physical clustering identified the COG1469 family as a potential candidate for this missing enzyme family. The GCYH-I activity of COG1469 family proteins from a variety of sources (Thermotoga maritima, Bacillus subtilis, Acinetobacter baylyi, and Neisseria gonorrhoeae) was experimentally verified in vitro and/or in vivo. Although there is no detectable sequence homology with the canonical GCYH-I, protein fold recognition based on sequence profiles, secondary structure, and solvation potential information suggests that, like GCYH-I proteins, COG1469 proteins are members of the tunnel-fold (T-fold) structural superfamily. This new GCYH-I family is found in ~20% of sequenced bacteria and is prevalent in Archaea, but the family is to date absent in Eukarya.

Folic acid, in the form of various tetrahydrofolate (THF) derivatives, serves as a cofactor in one-carbon transfer reactions during the synthesis of purines, thymidylate, pantothe-
Materials and Methods

Bioinformatics—Analysis of the folate subsystem was performed in the SEED database (14) with SEED version cvs.1144925141 (05:45:41 on April 13, 2006) (available on the World Wide Web at anno-3.nmpdr.org/anno/FIG/index.cgi). The phylogenetic pattern search was performed on the following SEED server: www.nmpdr.org/FIG/sigs.cgi?SPROUT—Analysis of the folate subsystem was performed with a three-step PCR thermocycling protocol was utilized: 1) 94 °C for 1 min; 2) 30 cycles of denaturation at 94 °C for 1 min, annealing at 50 °C for 2 min, and extension at 72 °C for 4 min. The PCR product was purified from a 1% agarose gel containing ethidium bromide using the Qiagen Inc. PCR purification kit and cloned into a linearized pET-30 Xa/LIC expression vector (Novagen). The primary structures of the resulting constructs, pSAB-7-189 (T. maritima), pSAB-8-142 (N. gonorrhoeae), and pSAB-9-61 (B. subtilis), were confirmed by sequencing.

Cloning of Thermotoga maritima, Neisseria gonorrhoeae, and Bacillus subtilis COG1469 Genes for Protein Expression—The COG1469 genes from T. maritima (TM0039; GenBankTM accession number gi|15642814), N. gonorrhoeae (nego387; GenBankTM accession number gi|59800831), and B. subtilis (yciA; GenBankTM accession number gi|2632620) were amplified by PCR from genomic DNA of the respective organisms. The following primers, corresponding to the 5’ and 3' regions of the gene, were used for PCR amplification: T. maritima sense primer (5’-TCCCTCCAGAACGCAACATACGCTGCTCCCG-3’), T. maritima antisense primer (5’-AGAGGAGAGTTAGAGCCACCATGACAACG-3’), N. gonorrhoeae sense primer (5’-GATGCTCAACTTCCCGAATGCTGCTAAGGAGTTAGATGTTCAACACATACGCCG-3’), and B. subtilis antisense primer (5’-AGAGGAGAGTTAGAGCCACCATGACAACG-3’). The phylotype patterns were performed on the following SEED server: www.nmpdr.org/FIG/sigs.cgi?SPROUT—Analysis of the folate subsystem was performed with a three-step PCR thermocycling protocol was utilized: 1) 94 °C for 1 min; 2) 30 cycles of denaturation at 94 °C for 1 min, annealing at 50 °C for 2 min, and extension at 72 °C for 4 min. The PCR product was purified from a 1% agarose gel containing ethidium bromide using the Qiagen Inc. PCR purification kit and cloned into a linearized pET-30 Xa/LIC expression vector (Novagen). The primary structures of the resulting constructs, pSAB-7-189 (T. maritima), pSAB-8-142 (N. gonorrhoeae), and pSAB-9-61 (B. subtilis), were confirmed by sequencing.

Cloning of Thermotoga maritima, Neisseria gonorrhoeae, and Bacillus subtilis COG1469 Genes for Protein Expression—The COG1469 genes from T. maritima (TM0039; GenBankTM accession number gi|15642814), N. gonorrhoeae (nego387; GenBankTM accession number gi|59800831), and B. subtilis (yciA; GenBankTM accession number gi|2632620) were amplified by PCR from genomic DNA of the respective organisms. The following primers, corresponding to the 5’ and 3’ regions of the gene, were used for PCR amplification: T. maritima sense primer (5’-TCCCTCCAGAACGCAACATACGCTGCTCCCG-3’), T. maritima antisense primer (5’-AGAGGAGAGTTAGAGCCACCATGACAACG-3’), N. gonorrhoeae sense primer (5’-GATGCTCAACTTCCCGAATGCTGCTAAGGAGTTAGATGTTCAACACATACGCCG-3’), and B. subtilis antisense primer (5’-AGAGGAGAGTTAGAGCCACCATGACAACG-3’). The phylotype patterns were performed on the following SEED server: www.nmpdr.org/FIG/sigs.cgi?SPROUT—Analysis of the folate subsystem was performed with a three-step PCR thermocycling protocol was utilized: 1) 94 °C for 1 min; 2) 30 cycles of denaturation at 94 °C for 1 min, annealing at 50 °C for 2 min, and extension at 72 °C for 4 min. The PCR product was purified from a 1% agarose gel containing ethidium bromide using the Qiagen Inc. PCR purification kit and cloned into a linearized pET-30 Xa/LIC expression vector (Novagen). The primary structures of the resulting constructs, pSAB-7-189 (T. maritima), pSAB-8-142 (N. gonorrhoeae), and pSAB-9-61 (B. subtilis), were confirmed by sequencing.
PCR using the following oligonucleotides located upstream and downstream from the folE gene (ChkDfole-ol1 \(5'\text{-CTCTTT-GTTGTGTTTGGCA}3'\)) and ChkDfole-ol2 \(5'\text{-GGGG-CACAAATTTCGAGG}3'\)) or located upstream and downstream of the polylinker in the pBAD derivatives (pBA-Drev2 \(5'\text{-TCTCTATTATCCTATCGCC}3'\)) and pBADols5 \(5'\text{-AGTATACGCGATCTACTCC}3'\)).

**Purification and Overexpression of Recombinant T. maritima, N. gonorrhoeae, and B. subtilis COG1469 Proteins**—The plasmids pSAB-7-189, pSAB-8-142, and pSAB-9-61 were transformed into *E. coli* BL21 (DE3) for expression of His6 tag fusion proteins. Cultures of the transformed cells were grown at 37 °C with shaking (250 rpm) until an \(A_{600}\) of 0.9 was attained. Isopropyl-\(\beta\)-D-thiogalactopyranoside was added to a final concentration of 0.1 mM, and the cultures were incubated for an additional 4 h at 37 °C with shaking (250 rpm). The cells were harvested by centrifugation at 5000 × g for 10 min at 4 °C. The cell paste was flash frozen in liquid nitrogen and stored at −80 °C until needed.

Frozen cell paste was thawed and suspended in lysis buffer (50 mM Tris acetate (pH 8.0), 50 mM KCl, and 1 mM \(\beta\)-mercaptoethanol) at a concentration of 250 mg/ml. The cells were lysed by the addition of lysozyme and DNase to a final concentration of 0.25 mg/ml and 10 \(\mu\)g/ml, respectively. The lysate was centrifuged at 15,000 × g for 30 min at 4 °C, and the resulting supernatant was filtered (low protein binding, 0.45 \(\mu\)m). The cell-free extract was loaded onto an Ni\(^{2+}\)-nitrilotriacetic acid-agarose column (Qiagen) that had been equilibrated with Buffer A (100 mM Tris-acetate (pH 8.0), 300 mM KCl, 2 mM \(\beta\)-mercaptoethanol, 1% Triton X-100, 1 mM phenylmethysulfonyl fluoride, and 10% glycerol). The column was washed with 5 column volumes of Buffer A, 5 column volumes of Buffer B (100 mM Tris acetate (pH 8.0), 300 mM KCl, 2 mM \(\beta\)-mercaptoethanol, 1% Triton X-100, 1 mM phenylmethysulfonyl fluoride, 10% glycerol, and 20 mM imidazole), and finally 5 column volumes of Buffer C (100 mM Tris acetate (pH 8.0), 300 mM KCl, 2 mM \(\beta\)-mercaptoethanol, 10% glycerol, and 20 mM imidazole). The protein was eluted from the column with 1 column volume of Buffer C containing 250 mM imidazole. The protein was concentrated in a Centricon YM-10 ultracentrifugation device and dialyzed at 4 °C against 50 mM Tris acetate (pH 8.0), 50 mM KCl, and 4 mM dithiothreitol.

The His6 tag was cleaved from the *T. maritima*, *N. gonorrhoeae*, and *B. subtilis* COG1469 proteins in reactions that contained fusion protein (20 mg), Factor Xa protease (20 \(\mu\)g), 50 mM Tris acetate (pH 8.0), 100 mM KCl, 2 mM CaCl\(_2\) in a final volume of 1 ml. After incubating for 20 h at room temperature, the reactions were loaded onto a column containing 2 ml of Ni\(^{2+}\)-nitrilotriacetic acid-agarose equilibrated in Buffer A. Wild-type protein was eluted from the column with 10 column volumes of Buffer A. The protein was concentrated and dialyzed against 50 mM Tris acetate (pH 8.0), 50 mM KCl, and 10% glycerol.

**Radiochemical Analysis**—The radiochemical release of \([^{14}C]\)formic acid from \([8-^{14}C]\)GTP was conducted according to the methods of Ferre et al. (20) and Burg et al. (21). Assays were performed at 37 °C (*E. coli*, *N. gonorrhoeae*, *B. subtilis*) or 75 °C (*T. maritima*) in 100 mM Tris-HCl (pH 8.0), 100 mM KCl, 2.0 mM MgCl\(_2\), 1.0 mM \([8-^{14}C]\)GTP, and 5.0 \(\mu\)M enzyme in a total volume of 250 \(\mu\)l. Aliquots of 50 \(\mu\)l were removed at specific time points and terminated by the addition of 62.5 \(\mu\)l of 0.5 M formic acid. After chilling on ice for 5 min, the samples were loaded onto a column containing 250 \(\mu\)l of charcoal that had been equilibrated in 50 mM Tris-HCl (pH 7.5). Formate was washed from the column with 50 mM Tris-HCl (pH 7.5), and the eluent was collected and quantified by liquid scintillation counting.

**UV-visible Analysis**—Assays were performed in 100 mM Tris-HCl (pH 8.0), 100 mM KCl, 2.0 mM MgCl\(_2\), 0.1 mM GTP, and 20 \(\mu\)M *E. coli* FoE, or 40 \(\mu\)M COG1469 enzyme from *B. subtilis* or *N. gonorrhoeae* in a volume of 0.5 ml. The reactions were allowed to proceed for 30 min at 37 °C while monitoring the absorbance from 200 to 400 nm (22).

**Fluorescence**—Assays were performed in 100 mM Tris-HCl (pH 8.0), 100 mM KCl, 2.0 mM MgCl\(_2\), and either 10 \(\mu\)M *E. coli* FoE or 40 \(\mu\)M COG1469 enzyme in a total volume of 100 \(\mu\)l. Reactions were initiated with the addition of GTP to a final concentration of 1.0 mM and incubated for 60 min at 37 °C (*E. coli*, *N. gonorrhoeae*, *B. subtilis*) or 75 °C (*T. maritima*). Alkaline phosphatase (5 units) was added, and the reaction mixtures were further incubated for 30 min at 37 °C. To the reactions was then added 12 \(\mu\)l of a 1% I\(_2\) 2% KI solution in 1.0 M HCl, and the samples were incubated at room temperature in the dark for 15 min. Excess iodine was reduced by the addition of 6 \(\mu\)l of 2% ascorbic acid. The samples were then analyzed by fluorescence spectroscopy (excitation at 365 nm and emission at 446 nm).

**HPLC Analysis**—Assays were performed in 100 mM Tris-HCl (pH 8.0), 100 mM KCl, 2.0 mM MgCl\(_2\), 1.0 mM GTP, and 60 \(\mu\)M *E. coli* FoE or 100 \(\mu\)M COG1469 protein in 0.5 ml. Reactions were incubated at 37 °C for 60 min followed by the addition of alkaline phosphatase, and the reactions were incubated an additional 60 min at 37 °C. The reaction products were analyzed by reversed phase HPLC on a Gemini C18 (Phenomex; 250 × 3.90 mm, 5 \(\mu\)m) column equilibrated in 25 mM ammonium acetate (pH 6.0). The column was developed at 1 ml/min with the following solvent gradient: 0–10 min, 0% acetonitrile; 30 min, 4% acetonitrile; 35 min, 50% acetonitrile.

**MS Analysis of GCYH-I Reactions**—The preparation of products from GCYH-I reaction assays for liquid chromatography-mass spectrometry analysis was carried out in reaction mixtures containing 100 mM Tris-HCl (pH 8.0), 100 mM KCl, 2.0 mM MgCl\(_2\), 1.0 mM GTP, and either 20 \(\mu\)M *E. coli* FoE or 40 \(\mu\)M COG1469 from *N. gonorrhoeae* in a volume of 500 \(\mu\)l. The reaction mixtures were incubated at 37 °C for 3 h in the dark and treated with activated charcoal in a modification of the method of Yim and Brown (10). After incubating the charcoal-treated reactions at 4 °C for 30 min in the dark, the mixtures were filtered through a Millipore type HA filter (0.45 \(\mu\)m). The filtrate was washed sequentially with 10 ml of water, 10 ml of 5% ethanol, 10 ml of 50% ethanol containing 3.1% ammonium hydroxide (pH 8.0), and 5 ml of 50% ethanol containing 3.1% ammonium hydroxide (pH 12). The filtrate from the final wash was immediately neutralized with acetic acid. The filtrates were frozen in liquid nitrogen and lyophilized to dryness and then dissolved in 20 mM ammonium acetate (pH 8.0) and 50% methanol.
and filtered through Millipore Amicon ultrafree-MC spin filters. The filtrates were analyzed by MS with an LCQ Advantage ion-trap mass spectrometer (Thermo Electron, San Jose, CA) equipped with an electrospray ionization source. The ion interface was operated in the negative mode using the following settings: needle voltage of 4.5 kV; sheath and auxiliary gas flow rates of 25 and 3.0 p.s.i., respectively; tube lens voltage of 50 V; capillary voltage of 3.0 V; and capillary temperature of 275 °C. An instrument method was created to scan the range m/z 50–1100. An isocratic LC mobile phase system consisted of methanol and water (pH 9.0) (1:1 by volume) delivered at a flow rate of 0.4 ml/min. The injection volume was 20–50 μl in aqueous 20 mM ammonium acetate (pH 8.0).

RESULTS

Comparative Genomic Analysis of folE—The signature genes of the de novo folate pathway are folP and folK, which encode dihydropterate synthase and 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, respectively. All organisms that possess these two genes should have a homolog of the folE gene, since none of the metabolic intermediates, from 7,8-dihydrofolate to 7,8-dihydrohydroxymethylpterin pyrophosphokinase, are transported in bacteria (23). Analysis of the distribution of the folE gene among all sequenced genomes that possessed folKP homologs revealed a large class of organisms (Table 1 and supplemental Table 1) that lacked folE homologs, suggesting that folE was “locally missing” (24) in these organisms. Using a SEED tool that allows identification of protein families that follow a defined phylogenetic distribution profile, we searched the available genomes for protein families that clustered physically with folate metabolism genes in several organisms. The combination of phylogenetic distribution and clustering suggested that the COG1469 family might encode the missing GCYH-I enzyme.

COG1469 Genes Complement an E. coli ΔfolE Mutant—Because folate is not transported in most bacteria (23), it cannot be supplied in the medium to enable growth of a folate auxotroph. However, on rich medium, all of the folate-derived metabolites are present in sufficient quantities except for dT, which allows a ΔfolE mutant to be maintained on LB/dT (19, 25). Nevertheless, the E. coli ΔfolE::KanR strain has a slow growth phenotype on LB/dT (colonies take 2 days instead of one to form at 37 °C), presumably due to the absence of formylation of the initiator tRNA. The ΔfolE::KanR strain was transformed with pBAD derivatives expressing the COG1469 homolog from T. maritima (TM0039). Although complementation of both the dT auxotrophy (data not shown) and the slow growth phenotype was observed (Fig. 3C), it was not robust and depended on high arabinose levels. This is not surprising, since T. maritima is a thermophile, and many of the enzymes from thermophiles exhibit low activity at 37 °C. To achieve better complementation, the COG1469 orthologs from the mesophiles B. subtilis and A. baylyi (formally known as Acinetobacter sp. ADP1) were cloned and transformed into the E. coli ΔfolE::KanR strain. Robust complementation of dT auxotrophy (Fig. 3A) and poor growth (Fig. 3B) was observed with these constructs and is consistent with COG1469 family proteins catalyzing GCYH-I activity.

COG1469 Proteins Have GCYH-I Activity in Vitro—In parallel with the in vivo experiments, COG1469 genes were cloned into protein expression vectors to allow unambiguous assignment of catalytic function through the direct investigation of putative GTP cyclohydrolase I activity with in vitro enzymatic assays of purified proteins. Thus, the genes encoding COG1469 proteins from T. maritima, N. gonorrhoeae, and B. subtilis were cloned from genomic DNA into the pET30 system, and the recombinant His6 fusion proteins were overproduced and purified. All

### TABLE 1

| Organism | folE | COG1469 | folK | folP |
|----------|------|---------|------|------|
| E. coli K12 | +    | +       | +    | +    |
| B. subtilis subsp. subtilis strain 168 | +    | +       | +    | +    |
| A. baylyi | +    | +       | +    | +    |
| Bordetella bronchiseptica RB50 | +    | +       | +    | +    |
| Neisseria meningitidis MC58 | +    | +       | +    | +    |
| Nitrosomonas europaea ATCC 19718 | +    | +       | +    | +    |
| Oceanobacillus iheyensis HTE831 | +    | +       | +    | +    |
| S. aureus subsp. aureus MW2 | +    | +       | +    | +    |
| T. maritima MS8B | +    | +       | +    | +    |
| Desulfotalea psychrophila L5v54 | +    | +       | +    | +    |
| Desulfuromonas acetoxidans | +    | +       | +    | +    |
| Esigooebacter sp. 255-15 | +    | +       | +    | +    |
| Geobacter sulfurreducens PCA | +    | +       | +    | +    |
| Magnetococcus sp. MC-1 | +    | +       | +    | +    |
| Methylobacillus flagellatus KT | +    | +       | +    | +    |
| Neisseria lactamica ST-640 | +    | +       | +    | +    |
| Neisseria meningitidis FAM18 | +    | +       | +    | +    |
| Silicibacter sp. TM1040 (B) | +    | +       | +    | +    |
| Rhodobacter sphaeroides 2.4.1 | +    | +       | +    | +    |
| N. gonorrhoeae FA 1090 | +    | +       | +    | +    |
| Geobacter metallireducens GS-15 | +    | +       | +    | +    |
| Idiomarina lobifera l2TR | +    | +       | +    | +    |
| Oceanicola battersi HTCC2597 | +    | +       | +    | +    |
| Rhodobacterales bacterium HTCC2664 | +    | +       | +    | +    |
| Staphylococcus haemolyticus JCSCI435 | +    | +       | +    | +    |
| Thiomicrospira crunogena XCL-2 | +    | +       | +    | +    |
| Sulfitobacter sp. EE-36 | +    | +       | +    | +    |
| Roseovarius nubinhibens ISM (B) | +    | +       | +    | +    |
| Staphylococcus saprophyticus LK70 | +    | +       | +    | +    |
| Lkalanella vestfoldensis SKA53 | +    | +       | +    | +    |
| Roseobacter sp. MED193 | +    | +       | +    | +    |
| Desulfuvibrio vulgaris | +    | +       | +    | +    |

5 A. Hanson and V. de Crécy-Lagard, unpublished results.

---

FIGURE 2. Physical clustering of COG1469 encoding genes with folate biosynthetic genes, such as folK, folP, and folM (an alternative dihydrofolate reductase gene (53)).
three of the recombinant proteins were obtained as soluble, active enzymes both as the His6 fusion and the cleaved wild type.

Radiochemical assays using [8-14C]GTP (10) of each of the COG1469 proteins, along with E. coli FolE as a positive control, demonstrated that [14C]formate was released in each assay and that its production was both time- and enzyme-dependent (data not shown), consistent with enzyme-catalyzed hydrolytic reactions. From these data, specific activities of 2.3–5.3 nmol min⁻¹ ng⁻¹ were calculated for the COG1469 proteins, roughly an order of magnitude lower than that reported for FolE (28–30) and our FolE control. To confirm that the product of the COG1469 catalyzed reactions was in fact 7,8-dihydropterin triphosphate, we analyzed the enzyme assays with UV-visible (22) and fluorescence (26) spectroscopy. Shown in Fig. 4A are UV-visible spectra for enzyme assays of E. coli FolE and COG1469 proteins under standard GTP cyclohydrolase I assay conditions. The spectra are essentially identical, with the characteristic absorption spectrum of GTP replaced by that of H2NTP (22). When enzyme assays were subjected to postreaction dephosphorylation and oxidation to convert the putative enzymatically produced H2NTP to the fluorescent neopterin, the fluorescent spectra from the COG1469 assays were identical to the spectrum of the E. coli FolE assay (Fig. 4B) and to that of authentic neopterin (26). Furthermore, HPLC analysis of the enzyme assays after dephosphorylation showed that the product from each enzyme-catalyzed reaction had the same retention time as the putative authentic neopterin (under the analysis conditions dihydroneopterin is oxidized to neopterin) (Fig. 4C). Finally, mass spectrometry analysis of the E. coli FolE and N. gonorrhoeae COG1469 reactions revealed identical constituents, with ions corresponding to GTP ([M + H]⁺ at 522 m/z) and neopterin ([M - 2H]⁻ at 544 m/z) for both enzymes.

FIGURE 3. Complementation experiments. Complementation of the dT auxotrophy on LB (A) or the slow growth phenotype on LB dT (B) of the E. coli ΔfolE::KanH by pBAD derivatives expressing the COG1469 genes from B. subtilis (BsfolE2) and A. baylyi (AcfolE2). C, complementation of the slow growth on LB in the presence of dT of the E. coli ΔfolE::KanH by a plasmid expressing TM0039 (TmfolE2). The control is ΔfolE::KanH transformed by pBAD24. The plates were incubated at 37 °C for 48 h. D, PCR amplification to check for the presence of the ΔfolE::KanH allele (lanes 2, 4, 6, and 8) and to check for the presence of an insert of the expected size in the pBAD derivatives (lanes 3, 5, 7, and 9). Each PCR reaction was performed as described under "Materials and Methods" on the ΔfolE::KanH strain transformed with plasmids expressing AcfolE2 (lanes 2 and 3), BsfolE2 (lanes 4 and 5), and TmfolE2 (lanes 6 and 7) and with the pBAD24 control (lanes 8 and 9). The expected size of the PCR product detecting ΔfolE::KanH is about 3.5 kb, whereas the same primers amplify a 2.5-kb product in the wild type strain. The sizes of the PCR products resulting from having AcfolE2, BsfolE2, or TmfolE2 in pBAD24 are 1072, 1105, and 960 bp, respectively.

FIGURE 4. GCYH-I activity assays. A, UV-visible spectra of cyclohydrolase assays. a, E. coli FolE; b, COG1469 protein from N. gonorrhoeae; c, COG1469 protein from B. subtilis; d, no enzyme. B, fluorescence spectra of authentic neopterin and cyclohydrolase assays following postreaction dephosphorylation and oxidation. Fluorescence was measured with excitation at 365 nm. a, T. maritima COG1469; b, B. subtilis COG1469; c, E. coli FolE; d, authentic neopterin; e, N. gonorrhoeae COG1469. C, HPLC chromatograms of authentic neopterin (a) and cyclohydrolase assays with E. coli FolE (b). B. subtilis COG1469 added to authentic neopterin (c), and B. subtilis COG1469 alone (d). Enzyme assays were subjected to postreaction dephosphorylation and oxidation.
Na⁺; m/z 566, M – 3H⁻ + 2Na⁺), neopterin triphosphate (m/z 492, M – H⁻; under the conditions of the analysis dihydronicotin is oxidized to neopterin), and neopterin cyclic monophosphate (m/z 314, M – H⁻; it has been previously documented that under the alkaline conditions of the work-up neopterin triphosphate is converted to the cyclic monophosphate (31–33)).

Taken together, the data clearly demonstrate that the COG1469 proteins catalyze GTP cyclohydrolase I activity, and thus they represent a new structural class of GTP cyclohydrolase enzymes, distinct from the canonical GCYH-I enzyme exemplified by human and E. coli FolE. To differentiate these two cyclohydrolase families, we propose that the canonical type I cyclohydrolase be renamed GCYH-IA, that the COG1469 members belong to the bimodular subfamily of the T-fold superfamily in general (43). Both the size of COG1469 enzymes (∼250–300 amino acids) and the fact that two T-fold domains can be detected in their sequences suggest that COG1469 members belong to the bimodular subfamily of the T-fold superfamily, which includes urate oxidase (46), the plant GCYH-IA enzyme (31), and the novel nitrile oxidoreductase (class 2; e.g. YqCD from E. coli) recently reported (27). Preliminary sedimentation velocity and crystallographic analyses of N. gonorrhoeae GCYH-IB suggest either a trimeric or a tetrameric quaternary structure (data not shown).

Based on a superposition of the predicted tertiary structure of GCYH-IB with the crystal structure of E. coli GCYH-IA (47) and that of DHNA (45), a sequence alignment of the C-terminal half of GCYH-IB with GCYH-IA could be generated (Fig. 5B). The alignment reveals ∼16% sequence similarity between the two GCYH-I families and shows that the GCYH-IB family contains the conserved Glu characteristic of the substrate-binding pocket of GCYH-IA enzymes (Glu¹⁵², E. coli numbering) and T-fold enzymes in general (43). In GCYH-IA, two conserved motifs, CEHH and HXXC, contain the zinc-coordinating and catalytic residues and are separated by ∼70 residues. These motifs are missing from GCYH-IB sequences and are replaced by CP-C/H/S-A/S and ESXHH, which are separated by ∼100 residues (e.g. Cys¹⁴⁶, Pro¹⁴⁷, Cys¹⁴⁸, Ser¹⁴⁹, Xaa⁵⁰, Glu²⁴², Ser²⁴³, Ile²⁴⁴, His²⁴⁵, Asn²⁴⁶, His²⁴⁷ in N. gonorrhoeae residue numbers). Although in some organisms the two motifs in GCYH-IB collectively lack only a single cysteine or histidine in the putative active site, in other organisms the combined motifs

**DISCUSSION**

**Phylogenetic Distribution of folE and folE2 Genes**—The role of folate as an essential cofactor, coupled with the historical importance of the pathway in the development of antibacterial, antiparasitic, and anticancer therapies (34), has led to folate metabolism being an especially well characterized area of biology. Thus, the discovery of a novel, widely distributed folate biosynthetic enzyme is a particularly compelling illustration of the power of comparative genomic approaches to link genes and function.

We analyzed the distribution of the folE/folE2 genes among all sequenced organisms in the SEED data base (26 archaeal, 363 bacterial, and 29 eukaryal more or less complete genomes). No FolE2 homolog is present to this date in any of the eukaryotic genomes, and as shown in supplemental Table 1, there is significant variation in the distribution of the folE/folE2 genes among bacteria. The first and largest group, which includes E. coli, has only a folE homolog. A second group, which includes Staphylococcus aureus and N. gonorrhoeae, has only a folE2 homolog. A third group, including B. subtilis and A. baylyi, has a homolog of each gene, whereas a fourth group can possess multiple copies of the two genes (e.g. Pseudomonas aeruginosa has two folE genes and one folE2 gene). The need for several genes encoding type I cyclohydrolase enzymes in many organisms is still not clear, but it may be due to differential expression under specific environmental conditions or their involvement in pathways other than folate biosynthesis; for example, a GTP cyclohydrolase has been implicated in the biosynthesis of 7-deazaguanosine derivatives, such as the modified tRNA nucleoside archaeosine (35) and the secondary metabolites toyocamycin and tubercidin (36, 37). In B. subtilis, it has been shown that the yciA gene is not essential (38), as expected because a folE gene (mtrA) (25) is also present in this organism. No folE2 deletions are available in bacteria that do not have another identified folE gene; construction of the corresponding folE2 mutant is currently under way.

Most archaeal genomes possess either a folE or a folE2 homolog (see supplemental Table 1). Several GTP-derived metabolites are synthesized in Archaea, including folate in the halophiles and Sulfolobii (39), tetrahydromethanopterin in the methanogens (40), and the 7-deazaguanosine tRNA-modified nucleoside archaeosine (41), which is found in the majority of archaeal tRNA. The archaeal folE/folE2 genes may be involved in one or more of these biosynthetic pathways.

**Structural Prediction of COG1469 Proteins**—The primary structure of COG1469 proteins presents no homology to any other known protein family. Direct alignment of COG1469 and GCYH-IA sequences yields no detectable similarity. However, protein fold recognition analysis using one- and three-dimensional sequence profiles, coupled with secondary structure and solvation potential information (using the 3D-PSSM server available on the World Wide Web at www.sbg.bio.ic.ac.uk/3dpssm/index2.html (42)), indicates potential three-dimensional structural homology with two tunnel-fold (T-fold) enzymes, a structural superfamily of enzymes that includes GCYH-IA (43). T-fold enzymes bind planar purine and pterin-like substrates but catalyze disparate reactions (43), and although they characteristically exhibit low sequence homology, their tertiary structural homology is very high. Using the N. gonorrhoeae sequence as a bait, the N-terminal half of COG1469 is most similar in predicted tertiary structure to 7,8-dihydronicotin triphosphate epimerase (Protein Data Bank code 1B9L (44), PSSM E value 0.39), whereas the C-terminal half is similar to 7,8-dihydroneopterin aldolase (DHNA; Protein Data Bank code 1NBU (45), PSSM E value 0.3) (Fig. 5A). These were the only PSSM hits with a qualifying E value (i.e. lower than the detection threshold E value of 1.00). Both hits are folate biosynthetic enzymes with homo-octameric structures. When aligned with the predicted fold of COG1469, the two enzymes exhibit comparable low overall sequence identities (10%) and similarities (23%) with the COG1469 family, consistent with the low sequence homology observed within the T-fold superfamily in general (43). Both the size of COG1469 proteins (∼250–300 amino acids) and the fact that two T-fold domains can be detected in their sequences suggest that COG1469 members belong to the bimodular subfamily of the T-fold superfamily, which includes urate oxidase (46), the plant GCYH-IA enzyme (31), and the novel nitrile oxidoreductase (class 2; e.g. YqCD from E. coli) recently reported (27). Preliminary sedimentation velocity and crystallographic analyses of N. gonorrhoeae GCYH-IB suggest either a trimeric or a tetrameric quaternary structure (data not shown).

**Novel GTP Cyclohydrolase I Family**
lack both a cysteine and a histidine. Furthermore, the locations of specific residues are different, both in primary sequence and deduced three-dimensional structure, and an additional conserved sequence, H\textsuperscript{X}Q-R/K (His158-Asn159-Gln160-Arg161 in \textit{N. gonorrhoeae} residue numbers), is found in GCYH-IB but not in GCYH-IA. The corresponding region in the structures of GCYH-IA and DHNA crystal structures are shown above and below the sequences, respectively. The red boxes indicate conserved residues in all three enzyme families. The invariant Glu is highlighted in red. Zinc binding and catalytic residues in GCYH-IB are highlighted in yellow and cyan, respectively. Black boxes indicate conserved residues in GCYH-IB predicted to be involved in catalysis and/or metal binding based on their predicted spatial homology to active site residues of GCYH-IA and DHNA. Ec, \textit{E. coli}; Bs, \textit{B. subtilis}; Sa, \textit{S. aureus}; Pa, \textit{Pyrococcus abyssi}; Ap, \textit{Aeropyrum pernix}; Tm, \textit{T. maritima}; Ng, \textit{N. gonorrhoeae}; Mt, \textit{M. tuberculosis}.

Mechanistic Implications of the Structural Divergence of GCYH-IA and -IB Enzymes—Of the enzymes involved in folate and bipterin biosynthesis, GCYH-IA has attracted particular attention (47–51) due to the mechanistic complexity inherent in the conversion of GTP to H\textsubscript{2}NTP. GCYH-IA activity is dependent on a catalytic Zn\textsuperscript{2+} atom (52), which functions as a Lewis acid in activating a water molecule for nucleophilic attack at C-8 of GTP in the initial hydrolytic step of the reaction. The Zn\textsuperscript{2+} further serves to facilitate nucleophilic attack of the second water molecule by polarizing the resulting amide carbonyl. The zinc-binding site in GCYH-IA is made up of Cys\textsuperscript{110}, His\textsuperscript{113}, and Cys\textsuperscript{181} (\textit{E. coli} numbering), with water occupying the fourth coordination site. As noted above, the zinc site is disrupted in GCYH-IB members, with His\textsuperscript{113} and Cys\textsuperscript{181} replaced by Ser/Ala and His, respectively, suggesting that metal binding is substantially different or abolished in the GCYH-IB enzymes. His\textsuperscript{115}, which has been identified as a key residue in opening and rearrangement of the ribose ring in GCYH-IA (30, 48), can be His/Cys/Ser in GCYH-IB, indicating that the latter steps of the reaction may also be catalyzed differently by the GCYH-IB enzymes.

The sequence of the GCYH-IB enzymes and resulting structural predictions are consistent with an active site architecture that is, minimally, much different from that in GCYH-IA, potentially involving a change in metal ion binding. We are actively investigating these issues using both biochemical and structural approaches and hope to resolve these questions in the near future. It further remains to be seen whether the apparent differences in the active site architecture of the type 1A and 1B enzymes can be exploited for the design of selective inhibitors of the type 1B enzymes; realization of such a goal would add yet another chapter to the therapeutic importance of the folate pathway.

Acknowledgments—We thank Scott Leslie (Joint Center for Structural Genomics) for the pTM0039 plasmid, Andrei Osterman (Burnham Institute) for insightful discussions, Ross Overbeek (The Fellowship for Integration of Genomes) for help with phylogenetic profile searches, and Andrew Hanson (University of Florida) for the \textit{E. coli} \textDelta\textit{folE} strain.

REFERENCES
1. Nichol, C. A., Smith, G. K., and Duch, D. S. (1985) \textit{Annu. Rev. Biochem.} \textbf{54}, 729–764
2. Clark, B. F., and Marcker, K. A. (1966) \textit{J. Mol. Biol.} \textbf{17}, 394–406
