JOURNAL OF BACTERIOLOGY, Sept. 1992, p. 5971–5977
0021-9193/92/185971-07$02.00/0

Vol. 174, No. 18

Cloning, Sequence Analysis, and Overexpression of *Escherichia coli folK*, the Gene Coding for 7,8-Dihydro-6-Hydroxymethylpterin-Pyrophosphokinase

TODD L. TALARICO,[1] PAUL H. RAY,[1] INDERJIT K. DEV,[1] BARBARA M. MERRILL,[2] AND WALTER S. DALLAS[1]*

Department of Molecular Genetics and Microbiology[1] and Department of Organic Chemistry,[2] The Wellcome Research Laboratories, Burroughs Wellcome Co., Research Triangle Park, North Carolina 27709

Received 29 May 1992/Accepted 17 July 1992

The gene coding for the *Escherichia coli* enzyme 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase has been cloned and sequenced. This gene, designated *folK*, codes for a protein of 159 amino acids, including an amino-terminal methionine. The protein was overexpressed in *E. coli* MC4100 by cloning the gene behind the *lacUV5* promoter in a high-copy-number plasmid. The enzyme was purified to homogeneity. Amino-terminal analysis of the purified protein showed that the amino-terminal methionine had been removed. The compositional molecular mass (17,945 Da) was identical to the molecular mass determined by mass spectrometry. The enzyme was observed to have a large number of proline residues and migrated anomalously in sodium dodecyl sulfate-polyacrylamide gels, with an apparent molecular mass of 23,000 Da.

All cells require reduced folate cofactors for the syntheses of a variety of essential nutrients, such as thymidylate, methionine, glycine, purines, and pantothenic acid (13). Unlike mammalian cells which take up folate cofactors, most microorganisms must synthesize folates de novo, because they lack the transport pathway. For these reasons, enzymes in this pathway have been and are potential targets for selective antimicrobial chemotherapy. A number of compounds have been developed as antimicrobial agents which target two of the seven enzymes in the folate pathway. Compounds directed toward two of these enzymes have found widespread use as antimicrobial agents. Trimethoprim inhibits dihydrofolate reductase, thereby preventing the formation of tetrahydrofolate (7, 13). The sulfonamides exert their antimicrobial effect on a different folic acid biosynthetic enzyme, dihydropteroate synthase (DHPS) (6, 11, 27, 29). DHPS catalyzes the condensation of *para*-aminobenzoic acid and 7,8-dihydro-6-hydroxymethylpterin-pyrophosphate to form dihydropteroate. Sulfonamides are substrate analogs that compete with *para*-aminobenzoic acid and form sulfapterin adducts (5, 27). The enzyme which catalyzes the synthesis of 7,8-dihydro-6-hydroxymethylpterin-pyrophosphate is 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase (HPPK) (26).

HPPK activity was first detected in cell extracts from a number of organisms (25–27) including *Escherichia coli* (26, 30, 31). Recently, the *Streptococcus pneumoniae* HPPK chromosomal gene, *sulD*, was cloned and sequenced (20). The *sulD* locus was shown to map close to the DHPS gene in this bacterium. *S. pneumoniae* HPPK was overexpressed in *E. coli* and purified, and the molecular mass, as determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), was shown to be the same as the compositional mass (31,000 Da). An apparent folic acid biosynthetic operon has also been cloned from *Bacillus subtilis*. This operon contained an open reading frame which exhibited sequence similarity to the *S. pneumoniae* DHPS-encoding gene and also included an open reading frame that was speculated to encode HPPK (19, 33). The prediction was based upon amino acid similarity of the putative gene product with HPPK from *S. pneumoniae*. However, *B. subtilis* HPPK was significantly smaller than the *S. pneumoniae* enzyme and closer in size to the HPPK protein species reported from *E. coli* (20, 33, 38). At that time, the only molecular size estimation for the *E. coli* protein was based on size-exclusion chromatography of crude extracts in which HPPK was found to elute as a 15,000-Da entity. Reports on HPPK in other organisms suggested the enzyme possessed a much larger molecular mass. Okinaka and Iwai (24) reported that the HPPK activity in pea seedlings was associated with a multifunctional enzyme of 180,000 Da. HPPK from *Pneumocystis carinii* and *Toxoplasma gondii* have also been suggested to be a part of large multifunctional protein (1, 36).

One problem associated with studying the folate pathway is that some of the enzymes are present in minute quantities and are therefore difficult to obtain in pure form. Recently, HPPK from *E. coli* was purified to homogeneity (35). The molecular mass of the enzyme was estimated to be 23,000 Da by SDS-PAGE, and size-exclusion chromatography indicated that the protein was a monomer. The sequence of the first 28 amino acids of this protein was determined. In this communication, we report the cloning, mapping, sequencing, and overexpression of the *E. coli* HPPK gene. The gene has been designated *folK*.

MATERIALS AND METHODS

Bacterial strains, plasmids, bacteriophages, and enzymes. The bacterial strains used were *E. coli* MC4100 (32) and JM110 (39). All strains were grown in Luria broth (32) (GIBCO BRL, Gaithersburg, Md.) supplemented with ampicillin (100 μg/ml) when appropriate. The ampicillin resistance (Amp^r) vector, pGem7Zf(+) (Promega, Madison, Wis.) was used for cloning. The Kohara mini set library of bacteriophage lambda (version 9010) was supplied by the National Institute of Genetics (Shizuoka, Japan) (17). Restriction enzymes, T4 DNA ligase, and T4 DNA polymerase

* Corresponding author.

were purchased from New England Biolabs (Beverly, Mass.). Sequenase was purchased from United States Biochemical Corporation (Cleveland, Ohio). DHPS was purified as previously described (35).

DNA methods. The degenerate oligonucleotide 5′ CCGC TGGA(G/A)CA(G/A)GT(C/G/T/A)AA(C/T)GC(C/G/T/A) TG(C/G/T) CTGAA(G/A)GC3′ was synthesized using a synthesizer (model 8700; Milligen/Biosearch, Burlington, Mass.). Hybridization was performed in 2× SSC– (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% SDS at 30°C for 16 h, and the filters were washed twice in the same buffer for a total of 60 min. DNA fragments were isolated from 1% SeaPlaque GTG low-melting-point agarose gel using a kit from Qiagen (Studio City, Calif.). Plasmids were isolated by using a magic miniprep kit from Promega. Digestions and ligations were performed essentially as described by O'Farrell et al. (23). Plasmids were introduced into bacteria by electroporation using a Gene Pulser Transfection Apparatus (Bio-Rad, Richmond, Calif.). The nucleotide sequences for both strands were determined with a dideoxynucleotide sequencing kit from United States Biochemical Corporation using double-stranded template. The SP6 and T7 primers (New England Biolabs) were used, as well as primers synthesized in our laboratory: CGTTTTAC CGCACCCCACCGCT, ATTTATGCTGTGGCCGCTGTT TGAAA, GCGAATATAAATATTCTGAAATATCAATC, TAATTTTGAATTAAGTAAATTTAATTTGTTCAA, CGT TCAGCTTTGCGACGCGACC, CGCAAAGAGATGTTTA AACTAAAATGTAAT, TAGGAGAAACAAATTTAGAT ATTAATGAG, and TCCTTTAATAAAATTCCATTAGTC AGACT. The oligonucleotide probe was endlabeled with ^{32}P using T4 polynucleotide kinase by the protocol recommended by the supplier (New England Biolabs) (21).

Purification of HPPK. *E. coli* MC4100(pTT5) (see Fig. 1) was grown overnight at 35°C in Luria broth containing 50 μg of ampicillin per ml. Twenty-four liters of culture (A_{600} of 5.7) were harvested in two 12-liter batches from a 16-liter fermentor (New Brunswick Scientific, New Brunswick, N.J.) and concentrated by diafiltration on 0.6-m^2 UltraSart cellulose triacetate NMWCO 20kDa filters (Sartorius, Long Island, N.Y.) to a volume of 1.2 liters. The cells were pelleted by centrifugation (180-g [wet weight] pellet) and resuspended in 200 ml of 20 mM Tris-HCl, pH 8. Cells were disrupted by passage through a French pressure cell (Aminco, Urbana, Ill.) at 16,000 psig and pelleted by centrifugation at 40,000 × *g*. The crude extract was diluted to 2 liters in 20 mM Tris-HCl, pH 8.0, and applied to a 250-ml column of Q Sepharose (Pharmacia, Piscataway, N.J.). The column was washed with 20 mM Tris-HCl, pH 8.0, until the absorbance returned to baseline, and the protein was eluted with a 2.5-liter solution with a linear gradient (0 to 0.2 M NaCl in 20 mM Tris-HCl, pH 8.0). The fractions containing HPPK activity were pooled, precipitated with 70% ammonium sulfate, and dialyzed twice against 4 liters of 25 mM imadazole acetate, pH 7.6. The dialyzed material was applied to a 75-ml column of PBE 94 chromatofocusing resin (Pharmacia) and eluted with PB 74 polybuffer (pH 4) (Pharmacia) diluted eightfold. The fractions containing HPPK activity were pooled, precipitated with ammonium sulfate, and dialyzed against 20 mM Tris-HCl (pH 8.0)–150 mM NaCl. The pool was applied to a column (2.5 by 120 cm) of G-75 size-exclusion resin equilibrated in the same buffer. Elution of HPPK from the G-75 column yielded a preparation which was estimated to be greater than 98% pure, as judged by SDS-PAGE analysis.

Analysis of cloned HPPK. Purified HPPK was sequenced on an Applied Biosystems 477A sequencer, and the amino acid composition was determined on a Beckman 6300 amino acid analyzer. The mass of the intact protein was determined by ion-spray mass spectrometry on a Sciex API III mass spectrometer (9). SDS-PAGE was performed as described previously (2) on 12% polyacrylamide gels and stained with Coomassie blue G-250. HPPK activity was assayed by a modified procedure of Ferone (10), using excess purified DHPS. The reactions were initiated by the addition of enzyme, and the dihydropteroate formed was separated from substrates by paper chromatography. Radioactivity was measured in a Packard 1500 Scintillation Counter. One unit of activity was defined as the production of 1 μmol of dihydropteroate per min at 37°C. Protein concentrations were determined by the method of Bradford (4). Heat stability studies were performed with 60 μg of HPPK in 1 ml of 10 mM sodium phosphate buffer, pH 7. Samples were placed in a heating block at 96°C for 30 min, the sample was diluted 1 to 100 in 25 mM phosphate buffer (pH 7.6), and HPPK activity at 37°C was determined.

RESULTS

Cloning of the HPPK gene. The amino-terminal amino acid sequence of HPPK, Thr-Val-Ala-Tyr-Ile-Ala-Ile-Gly-Ser-Asn-Leu-Ala-Ser-Pro-Leu-Glu-Gln-Val-Asn-Ala-X-Leu-Lys-Ala-Leu-Gly-Asp-Ile (35) was used to search the GenBank EMBL library using the Genetics Computer Group software package from the University of Wisconsin. The nucleotide sequence immediately following *pcnB*, a gene involved in plasmid copy number control, was predicted to encode a peptide which matched the first 20 amino acids of HPPK before the published sequence ended (18). March et al. (22) had mapped *pcnB* to approximately 3.6 min on the *E. coli* chromosome (3). The region around 3.6 min is contained between Kohara phages from 110 through 120 (28). A degenerate (1,024-fold) 32-base oligonucleotide which began with the first proline codon in the HPPK amino acid sequence was synthesized and used to probe Kohara phages from 110 to 120 (17). Positive hybridization signals were obtained for phages 115 and 116.

The restriction map of *pcnB* and flanking sequences (18) was aligned with the restriction map of phage 115, and we concluded that the HPPK gene would be on a 7-kb *Kpn*I fragment. Phage 115 was digested with *Kpn*I, the 7-kb fragment was isolated on gels, and the fragment was ligated with *Kpn*I-digested pGEM7Zf(+). After electroporation of the ligation mixture and subsequent plating, white transformants on X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) plates were screened by plasmid minipreparations for the appropriate plasmid. A plasmid was identified with a 7-kb *Kpn*I insert, and it was designated pTT10 (Fig. 1). The presence of the correct fragment and the orientation of the insert were confirmed by DNA sequence determination with the SP6 and T7 primers. By restriction analysis of pTT10 and estimation of the size of an open reading frame necessary to encode HPPK (35, 38), we determined that a deletion of a 5-kb *Eco*RI fragment from pTT10 would form a smaller plasmid which would contain the complete HPPK gene. pTT10 was digested with *Eco*RI, diluted, ligated, and electroporated into *E. coli*. A deletion plasmid was isolated and designated pTT5 (Fig. 1).

Sequence of the HPPK gene. The nucleotide sequence of the insert was determined by sequentially reading a nucleotide sequence and making a new primer based on the 3′ end of the newly sequenced portion. A 477-bp open reading

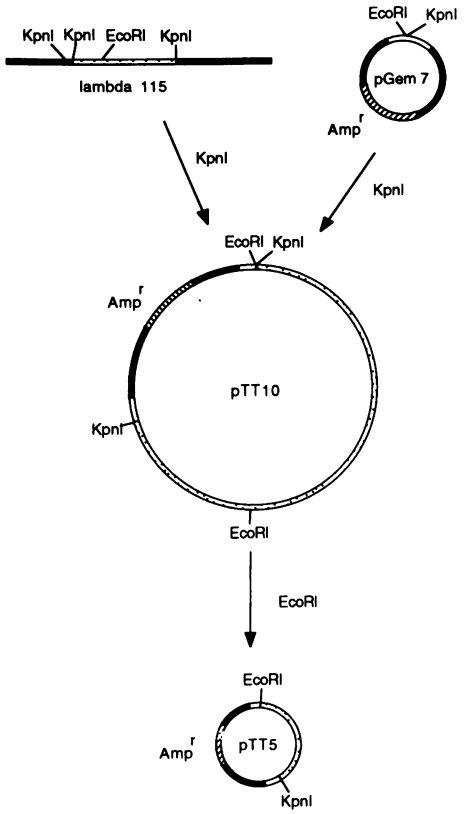

FIG. 1. Cloning of *folK*. A 7-kb *Kpn*I fragment from lambda 115 was subcloned into the *Kpn*I site of pGem7f(+). Deletion of an *Eco*RI fragment from pTT10 yielded a 5-kb plasmid, pTT5. The β-lactamase coding region is hatched and labeled Amp^r. The *lacZ* region is shown as the solid black portion. The multiple cloning site lies near the center of the *lacZ* region. Only the *Kpn*I and *Eco*RI restriction sites are shown in this figure.

frame was identified beginning 8 nucleotides 3′ to the termination codon of *pcnB* (18) (Fig. 2 [the termination codon of *pcnB* is from bp 122 to 124]). When the open reading frame was translated, the first 28 amino acids, excluding the amino-terminal methionine, were identical to the amino acids in the amino-terminal sequence of purified HPPK (35). The HPPK gene was designated *folK*. The region 3′ to *folK* was examined for the presence of other open reading frames, and the largest found was 86 nucleotides. This region was also searched for tracts which could form inverted repeats characteristic of transcription terminators. Three tracts were found, one of which started 2 bp from the *folK* termination codon (bp 614 to 637 in Fig. 2).

Enhanced expression of *folK*. The orientation of *folK* in pTT5 resulted in transcription from the *lacUV5* promoter. Plasmid pTT5 was introduced into *E. coli* MC4100, a strain with *lacI* deleted, and HPPK levels were determined. Expression of the enzyme was found to be 1,000-fold greater than for the wild-type strain on the basis of specific activity in overnight cultures.

Purification of HPPK. The purification of HPPK from wild-type *E. coli* MC4100 has been described (35). HPPK had been shown to compose less than 0.01% of the total soluble cell proteins, and the overall recovery was only 2%. By using a strain which overexpressed HPPK, enzyme yields increased substantially and the purification protocol was simplified. The total number of purification steps was reduced from seven to three, as shown in Table 1. The initial step, chromatography on Q Sepharose, yielded a large increase in the relative amount of HPPK (Fig. 3), but this was mitigated by a 50% loss of the total activity. Further purification of the protein was achieved by chromatofocusing followed by size-exclusion chromatography. Application of the new protocol resulted in the isolation of a single protein species, as shown in Fig. 3 (arrow). The purity of the preparation was estimated to be greater than 98% by SDS-PAGE analysis, and the overall recovery of HPPK activity was 22%. HPPK purified from MC4100(pTT5) was observed to migrate in SDS-polyacrylamide gels with the same molecular mass as HPPK from MC4100, 23,000 Da (35). However, this mass was inconsistent with the compositional mass of HPPK predicted from the nucleotide sequence, 17,945 Da (excluding the initial methionyl residue).

Characterization of HPPK. In order to resolve the inconsistency in the compositional mass of HPPK, the protein was analyzed by a number of other methods. Purified HPPK was subjected to N-terminal amino acid sequence analysis and the first 20 residues were found to be identical to those reported for native HPPK, with the first residue being threonine (35). The first residue predicted from the nucleotide sequence was methionine, but the second codon specifies threonine. The methionine must be removed posttranslationally. The amino acid composition was determined and was found to match the composition predicted from the nucleotide sequence (Table 2). The comparison indicated that the proteins were identical and in addition showed HPPK to contain an unusually high number of proline residues. The molecular mass of HPPK was determined by ion-spray mass spectrometry and found to be 17,945 ± 4 Da.

Weisman and Brown (38) reported that HPPK partially purified by DEAE-cellulose chromatography was heat stable, retaining 70% of its activity after a 30-min incubation at 100°C. We also found the purified recombinant enzyme to be heat stable, retaining 75% of its activity after a 30-min incubation at 96°C.

Homology of HPPK with other proteins. The amino acid sequences predicted from the open reading frames believed to encode HPPK in *E. coli*, *S. pneumoniae*, *B. subtilis*, and *P. carinii* were examined for regions of similarity by alignment of the protein sequences using the Microgenie software package (Beckman Instruments, Palo Alto, Calif.) (Fig. 4). The *E. coli* HPPK shared 43% identity with the protein predicted from translation of open reading frame 2 (ORF2) in the folate operon of *B. subtilis* (33). It shared 36% identity with the carboxyl-terminal portion of *S. pneumoniae* HPPK (20). The percentage of similar amino acid residues (identical matches plus mismatches which showed group identity such as a substitution of leucine with isoleucine) was approximately 50% when *E. coli* HPPK was compared with either *B. subtilis* or *S. pneumoniae* HPPK. The FasC domain from *P. carinii* (amino acids 292 to 462) also shared amino acid homology with HPPK from *E. coli* (36). The HPPK amino acid sequence from the four organisms are completely conserved for 32 residues, and 71 residues showed at least 75% conservation. One 8-residue stretch is conserved near the amino terminus in all of the organisms as well as one 6- and two 4-residue stretches near the middle of the protein (Fig. 4). The homology based upon alignment of regions of HPPK from these four organisms is shown in Fig. 5.

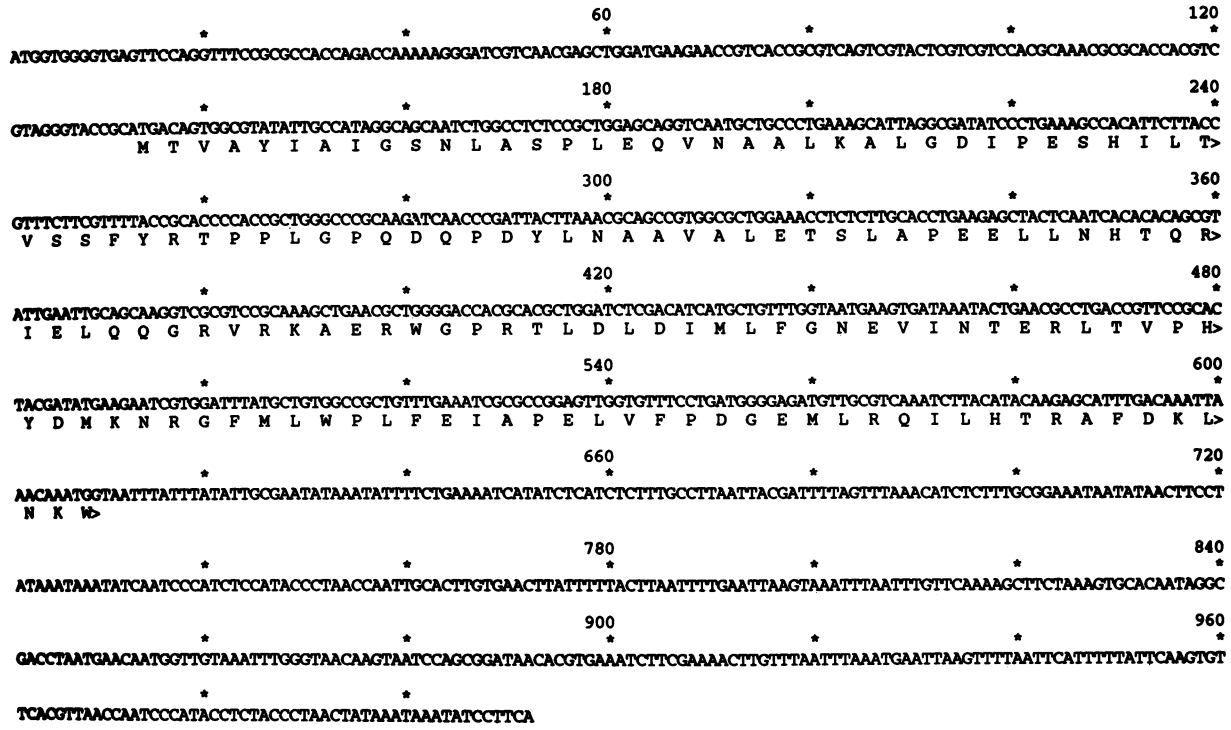

FIG. 2. Nucleotide sequence of the *folK* region. The sequence 5′ to *folK* codes for *pcnB*: the termination codon starts at nucleotide 122. *pcnB* sequence data for bases 1 to 125 was reported by Liu and Parkinson (18) and was combined with sequence data obtained for *folK* to show the proximity of the two genes. The open reading frame which codes for HPPK begins at nucleotide 133. The predicted amino acid sequence for the open reading frame is given in the one-letter code below the DNA sequence.

DISCUSSION

We have cloned and sequenced the HPPK gene which codes for the enzyme that catalyzes the conversion of 7,8-dihydro-6-hydroxymethylpterin to 7,8-dihydro-6-hydroxymethylpterin-PP_i and named it *folK*. *folK* was shown to map 3′ to *pcnB*, a plasmid copy number control gene located at 3.6 min on the *E. coli* chromosome (18, 22). March et al. had predicted the presence of a gene 3′ to *pcnB* after identifying an open reading frame with a potential ribosome binding site followed by an appropriately spaced methionine codon (22). The *folK* initiation codon was 8 bp from the *pcnB* stop codon. *pcnB* appears to be the first gene transcribed in a polycistronic mRNA, since it is preceded by −35 and −10 regions which closely resemble the consensus sequence for promoters (18). We sequenced 300 bp 3′ to the *folK* termination codon, and the longest open reading frame was 86 bp in length. However, this region contained three tracts which could form hairpin structures, one of which was a 10-bp inverted repeat with a 3-bp loop and started with the second base after the *folK* termination codon. The inverted repeat was followed by four consecutive thymidine nucleotides but did not contain the G-C-rich stretch of nucleotides in the stem region normally associated with transcription terminators (40). On the basis of this analysis, we predict that a polycistronic message would encompass only *pcnB* and *folK*.

We found that HPPK was overexpressed in MC4100 with either pTT10 or pTT5. The expression was dependent upon transcription from the *lacUV5* promoter, since inversion of

TABLE 1. Purification of HPPK

Fraction no. and description	Total protein (g)	Total U (μmol/min)	HPPK sp act (U/mg)	Recovery (%) of HPPK
1. Total crude extract	14.4	12,484	0.9	100
2. Q Sepharose	0.63	6,562	10.5	53
3. Chromatofocusing	0.16	5,322	32.7	43
4. G-75	0.08	2,738	34.2	22

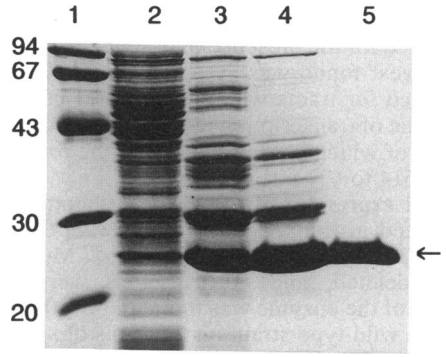

FIG. 3. Protein profile by SDS-PAGE of a fraction from each purification step in the isolation of HPPK. Lane 1, molecular mass markers (numbers to the left indicate the mass in kilodaltons); lane 2, crude extract; lane 3, Q-Sepharose pool; lane 4, chromatofocusing pool; lane 5, G-75 pool. The gels were stained with Coomassie blue G-250, and lanes 3, 4, and 5 were purposely overloaded with respect to HPPK to show contaminating proteins.

TABLE 2. Amino acid composition of HPPK

Amino acid	No. of amino acids	
	Predicted from composition analysis	Predicted from nucleotide sequence
Ala	13.4	13
Arg	10.5	10
Asn		8
Asp		8
Asx	16.2	16
Cys	ND[a]	0
Gln		7
Glu		12
Glx	19.9	19
Gly	8.9	9
His	4.1	4
Ile	8.1	9
Leu	21.7	22
Lys	5.5	5
Met[b]	2.3	5
Phe	6.2	6
Pro	12.1	12
Ser	6.0	6
Thr	8.3	9
Trp	ND	3
Tyr	4.1	4
Val	7.6	8

[a] ND, not determined.
[b] Amino-terminal methionine removed posttranslationally.

the insert (pTT10R) and introduction of this plasmid into MC4100 yielded only wild-type levels of HPPK. We chose MC4100 for these experiments, because *lacI* had been deleted from the strain and no induction was necessary.

Enhanced expression of HPPK enabled us to simplify a previously reported purification scheme (35). The purification procedure was reduced to three steps, and the overall recovery of enzyme was 22% compared with 2% with the original protocol (35). Additionally, the specific activity of the purified material from the recombinant source was 10-fold higher than previously reported (35). This may be a consequence of the increased speed and the reduced number of steps associated with the new protocol.

Eighty milligrams of protein was purified from 24 liters of cells. Mass spectrometric analysis showed the protein had a mass of 17,945 Da. Sequencing of the amino terminus confirmed that the first residue was threonine and that the next 19 residues were identical to those from HPPK purified from the wild-type strain. The nucleotide sequence of *folK* predicted an N-terminal methionine that is absent from the purified protein. The presence of threonine in the second position is expected to direct removal of the *N*-formylmethionine from cytoplasmic proteins by methionyl-aminopeptidase (14). Since HPPK purified from wild-type and recombinant sources contained the same amino terminus, migrated similarly on SDS-polyacrylamide gels, and had HPPK enzymatic activity, we concluded that the two sources yielded equivalent proteins, except that a portion of the protein from the former source was inactive.

The anomalous migration of HPPK on SDS-polyacrylamide gels (23,000 Da) was shown to be an inherent characteristic of the protein. Inspection of the amino acid composition of the enzyme showed that 36% of the residues were hydrophobic amino acids and that there were 12 proline residues. The composition may confer a unique structure to the protein, resulting in the anomalous migration characteristics during electrophoresis (35). Furthermore, the large number of proline residues (7.5 mol%) in HPPK probably influences other physical properties, such as heat stability. Suzuki et al. (34) have proposed the proline hypothesis

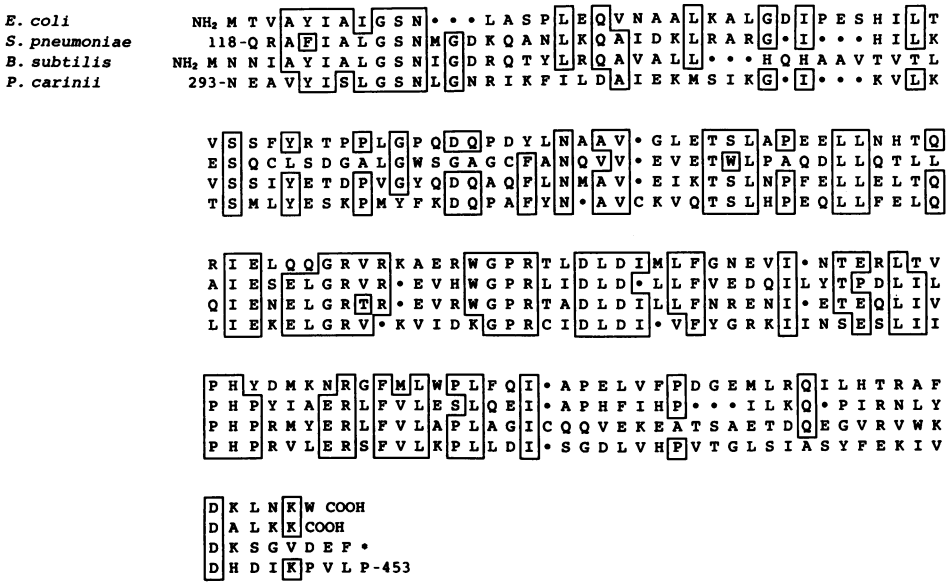

FIG. 4. Comparison of the predicted HPPK amino acid sequences of *E. coli*, *S. pneumoniae*, *B. subtilis*, and *P. carinii*. The boxed areas indicate regions of homology. Dots indicate spaces inserted into the sequence to provide optimal homology among the sequences. Numbers preceding and following a sequence indicate the first and last amino acid residues of the region used for comparison. NH_2 and COOH indicate the respective terminal amino acids of the protein. The asterisk following the *B. subtilis* HPPK sequence indicates that the complete sequence has not been determined. Protein sequence data for *S. pneumoniae*, *B. subtilis*, and *P. carinii* were reported in references 20, 33, and 36, respectively.

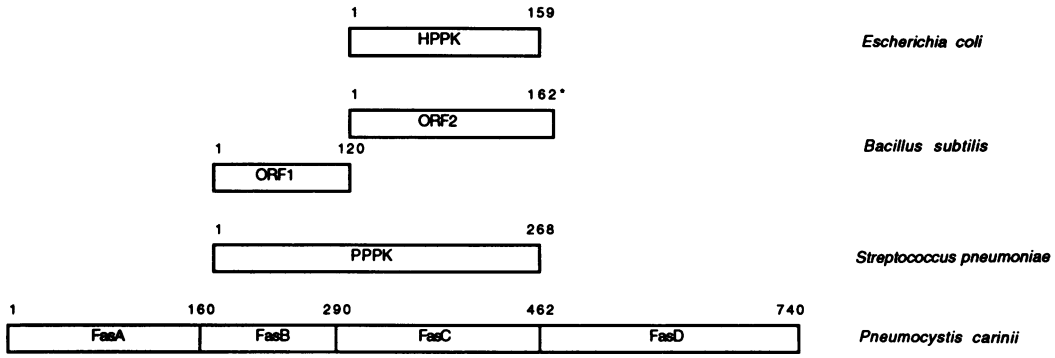

FIG. 5. Homology of regions of amino acids of monomeric subunits of HPPK from *E. coli*, *S. pneumoniae*, *B. subtilis*, and *P. carinii*. Proteins are aligned to indicate the regions of predicted amino acid sequence which show significant homology. Both ORF1 and ORF2 are from the folic acid operon in *B. subtilis*. The numbers above each region indicate the amino acid residue which ends the protein or region. The asterisk after the number for ORF2 indicates that the size of this protein has not been determined, because the cloned DNA fragment from *B. subtilis* did not include all of ORF2 (no termination codon for ORF2 was found).

which states that heat stability is proportional to the number of proline residues in a protein. This has been demonstrated by isolating the same enzyme from a number of different microorganisms, determining the proline content, and showing that thermal stability increased as proline content increased (15, 16, 37). Weisman and Brown first reported thermal stability for *E. coli* HPPK (38), and we have confirmed their initial observation with purified enzyme. Unfortunately, no heat stability data is available for other HPPKs, and we cannot test the proline hypothesis for the enzyme. The putative *B. subtilis* HPPK appears to be similar in size to *E. coli* HPPK but has only half the number of proline residues. If the proline hypothesis were applicable to HPPK, then we would predict the *E. coli* enzyme to be more heat stable.

We have chosen to designate the *E. coli* HPPK gene *folK* (K for kinase) in keeping with the convention established for genes coding for other enzymes in the folate pathway (*folA* is the gene which encodes dihydrofolate reductase, and *folC* is the gene which encodes folylpolyglutamate synthetase). The HPPK genes from four different organisms have now been sequenced, and there appears to be an interesting progressive relationship between this gene and other genes or sequences (Fig. 5). The *E. coli* HPPK gene is the simplest case. It codes for one of the smallest enzymes with HPPK activity, and the gene is not intimately linked to another open reading frame. The protein predicted from ORF2 in the *B. subtilis* folate operon is closest in size to *E. coli* HPPK. ORF2 is preceded by ORF1, which overlaps the beginning of ORF2. The protein predicted from ORF1 shows no similarity with the gene product from *E. coli pcnB* (which immediately precedes *folK*) but does have striking sequence similarity with the amino-terminal end of the *S. pneumoniae* HPPK gene. In fact, the *S. pneumoniae* HPPK gene resembles an in-frame fusion of the *B. subtilis* ORF1 and ORF2. To date, the most complex HPPK gene reported is from *P. carinii*. In this organism there are apparently four distinct domains in a multifunctional protein. The first domain (FasA) has no known equivalent in either *S. pneumoniae*, *B. subtilis*, or *E. coli*. The second domain (FasB) shares similarity with the amino-terminal end of *S. pneumoniae* HPPK and *B. subtilis* ORF1. The third domain (FasC) is similar to the three bacterial HPPKs. The final domain (FasD) has similarity with *S. pneumoniae* and *E. coli* DHPS (8a). One product of HPPK catalysis is a substrate for DHPS. There is precedent for multifunctional enzymes in the microbial world and more specifically for enzymes that use folate substrates. Several parasites make multifunctional proteins with both thymidylate synthase (N^5N^{10} methylenetetrahydrofolate is a substrate) and dihydrofolate reductase activities (8, 12).

Enzymes of the folate biosynthetic pathway are attractive targets for chemotherapeutic agents because of the requirement for de novo-synthesized reduced folate cofactors by many pathogenic microbes. Mammalian cells, on the other hand, must acquire folates from their surroundings and have efficient uptake systems. These transport systems are absent in most microorganisms. Now that the HPPK gene from *E. coli* has been cloned and the enzyme has been overexpressed, an adequate supply of the enzyme is available for use in drug susceptibility experiments and in experiments designed to explore the mechanism of catalysis. Pterin analogs which directly inhibit HPPK may now be screened more readily. A DNA fragment with cloned *folK* may be used to isolate a strain in which *folK* had been deleted, which may be useful in complementation cloning of the HPPK gene from other microorganisms. The opportunity for success in discovering new antimicrobial and antifungal agents should increase as more information about the enzymes in the folate pathway becomes available.

ACKNOWLEDGMENTS

We thank Lester Taylor and Lisa St. John for mass spectrometric analysis and Bill Chestnut for the amino acid sequence and compositional analysis. Thanks to Sara Egan and Ray Grimaila for helpful technical discussions.

REFERENCES

1. **Allegra, C. J., J. A. Kovacs, P. Morrison, J. Beaver, B. A. Chabner, and H. Masur.** 1989. Interaction of sulfonamide and sulfone compounds with *Toxoplasma gondii* dihydropteroate synthase. J. Clin. Invest. **85**:371–379.
2. **Ames, G. F.** 1974. Resolution of bacterial proteins by polyacrylamide gel electrophoresis on slabs. J. Biol. Chem. **249**:634–644.
3. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54**:130–197.
4. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248–254.
5. **Brown, G. M.** 1962. The biosynthesis of folic acid. II. Inhibition by sulfonamides. J. Biol. Chem. **237**:536–540.
6. **Brown, G. M., R. A. Weisman, and D. A. Molnar.** 1961. The

biosynthesis of folic acid. I. Substrate and cofactor requirements for enzymatic synthesis by cell-free extracts of *Escherichia coli*. J. Biol. Chem. **236**:2534–2543.
7. **Burchall, J. J., and G. M. Hitchings.** 1965. Inhibitor binding analysis of dihydrofolate reductases from infectious species. Mol. Pharmacol. **1**:126–136.
8. **Bzik, D. J., W. Li, T. Horii, and J. Inselburg.** 1987. Molecular cloning and sequence analysis of the *Plasmodium falciparum* dihydrofolate reductase-thymidylate synthase gene. Proc. Natl. Acad. Sci. USA **84**:8360–8364.
8a. **Dallas, W.** Unpublished data.
9. **Fenn, J. B., M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse.** 1989. Electrospray ionization for mass spectrometry of large biomolecules. Science **246**:64–71.
10. **Ferone, R.** 1973. The enzymatic synthesis of dihydropteroate and dihydrofolate by *Plasmodium berghei*. J. Protozool. **20**: 459–464.
11. **Griffin, M. J., and G. M. Brown.** 1964. The biosynthesis of folic acid. III. Enzymatic formation of dihydrofolic acid from dihydropteroic acid and tetrahydropteroylpolyglutamic acid compounds from tetrahydrofolic acid. J. Biol. Chem. **239**:310–316.
12. **Grumont, R., W. L. Washtien, D. Caput, and D. V. Santi.** 1986. Bifunctional thymidylate synthase-dihydrofolate reductase from *Leishmania tropica*: sequence homology with the corresponding monofunctional protein. Proc. Natl. Acad. Sci. USA **83**: 5387–5391.
13. **Harvey, R. J.** 1973. Growth and initiation of protein synthesis in *Escherichia coli* in the presence of trimethoprim. J. Bacteriol. **114**:309–322.
14. **Hirel, P.-H., J. M. Schmitter, P. Dessen, G. Fayat, and S. Blanquet.** 1989. Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. Proc. Natl. Acad. Sci. USA **86**: 8247–8251.
15. **Ito, Y., H. Yamada, S. Nakamura, and T. Imoto.** 1990. Purification, amino acid sequence, and some properties of rabbit kidney lysozyme. J. Biochem. **107**:236–241.
16. **Karst, U., H. Schutte, H. Baydoun, and H. Tsai.** 1989. Purification and characterization of leucine dehydrogenase from the thermophile *Bacillus caldolyticus*. J. Gen. Microbiol. **135**:1305–1313.
17. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50**:495–508.
18. **Liu, J., and J. S. Parkinson.** 1989. Genetics and sequence analysis of the *pcnB* locus, an *Escherichia coli* gene involved in plasmid copy number control. J. Bacteriol. **171**:1254–1261.
19. **Lopez, P., M. Espinosa, B. Greenberg, and S. A. Lacks.** 1987. Sulfonamide resistance in *Streptococcus pneumoniae*: DNA sequence of the gene encoding dihydropteroate synthase and characterization of the enzyme. J. Bacteriol. **169**:4320–4326.
20. **Lopez, P., B. Greenberg, and S. A. Lacks.** 1990. DNA sequence of folate biosynthesis gene *sulD*, encoding hydroxymethyldihydropterin pyrophosphokinase in *Streptococcus pneumoniae*, and characterization of the enzyme. J. Bacteriol. **172**: 4766–4774.
21. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **March, J. B., M. D. Colloms, D. Hart-Davis, I. R. Oliver, and M. Masters.** 1989. Cloning and characterization of an *Escherichia coli* gene, *pcnB*, affecting plasmid copy number. Mol. Microbiol. **3**:903–910.
23. **O'Farrell, P. H., E. Kutter, and I. Nakanishi.** 1980. A restriction map of the bacteriophage T4 genome. Mol. Gen. Genet. **179**: 421–435.
24. **Okinaka, O., and K. Iwai.** 1970. The biosynthesis of folic acid compounds in plants. IV. Purification and properties of the dihydropteroate-synthesizing enzyme from pea seedlings. J. Vitaminol. **16**:201–209.
25. **Ortiz, P. J., and R. D. Hotchkiss.** 1966. The enzymatic synthesis of dihydrofolate and dihydropteroate in cell-free preparations from wild-type and sulfonamide-resistant pneumonococcus. Biochemistry **5**:67–73.
26. **Richey, D. P., and G. M. Brown.** 1969. The biosynthesis of folic acid. IX. Purification and properties of the enzymes required for the formation of dihydropteroic acid. J. Biol. Chem. **244**:1582–1592.
27. **Roland, S., R. Ferone, R. Harvey, V. L. Styles, and R. W. Morrison.** 1979. The characteristics and significance of sulfonamides as substrates for *E. coli* dihydropteroate synthase. J. Biol. Chem. **254**:10337–10345.
28. **Rudd, K. E.** Alignment of *E. coli* DNA sequences to a revised integrated genomic restriction map. *In* J. H. Miller (ed.), A short course in bacterial genetics, in press. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **Shiota, T.** 1984. Biosynthesis of folate from pterin precursors, p. 121–134. *In* R. L. Blakley and S. J. Benkovic (ed.), Folates and pterins. John Wiley and Sons, New York.
30. **Shiota, T., and M. N. Disraely.** 1961. The enzymatic synthesis of dihydrofolate from 2-amino-4-hydroxy-6-hydroxymethyl-dihydropteridine and *p*-aminobenzoylglutamate by extracts of *Lactobacillus plantarum*. Biochim. Biophys. Acta **52**:467–473.
31. **Shiota, T., M. N. Disraely, and M. P. McCann.** 1964. The enzymatic synthesis of folate-like compounds from hydroxymethyldihydropteridine pyrophosphate. J. Biol. Chem. **239**: 2259–2266.
32. **Silhavy, T. J., M. J. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
33. **Slock, J., D. P. Stahly, C.-Y. Han, E. W. Six, and I. P. Crawford.** 1990. An apparent *Bacillus subtilis* folic acid biosynthetic operon containing *pab*, an amphibolic *trpG* gene, a third gene required for synthesis of *para*-aminobenzoic acid, and the dihydropteroate synthase gene. J. Bacteriol. **172**:7211–7226.
34. **Suzuki, Y., K. Oishi, H. Nakano, and T. Nagayama.** 1987. A strong correlation between the increase in number of proline residues and the rise in thermostability of five *Bacillus* oligo-1,6-glucosidases. Appl. Microbiol. Biotechnol. **26**:546–551.
35. **Talarico, T. L., I. K. Dev, W. S. Dallas, R. Ferone, and P. H. Ray.** 1991. Purification and partial characterization of 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase and 7,8-dihydropteroate synthase from *Escherichia coli* MC4100. J. Bacteriol. **173**:7029–7032.
36. **Volpe, F., M. Dyer, J. G. Scaife, G. Darby, D. K. Stammers, and C. J. Delves.** 1992. A multifunctional folate biosynthesis gene fas in *Pneumocystis carinii* encodes dihydropteroate synthase and hydroxymethyldihydropterin pyrophosphokinase. Gene **112**: 213–218.
37. **Watanabe, K., K. Chishiro, K. Kitamura, and Y. Suzuki.** 1991. Proline residues responsible for thermostability occur with high frequency in the loop regions of an extremely thermostable oligo-1,6-glucosidase from *Bacillus thermoglucosidasius* KP1006. J. Biol. Chem. **266**:24287–24292.
38. **Weisman, R. A., and G. M. Brown.** 1964. The biosynthesis of folic acid. V. Characterization of the enzyme system that catalyzes the synthesis of dihydropteroic acid. J. Biol. Chem. **239**:326–331.
39. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
40. **Yager, T. D., and P. H. von Hippel.** 1987. Transcript elongation and termination in *Escherichia coli*, p. 1241–1275. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*. American Society for Microbiology, Washington, D.C.