Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 9151–9155, October 1992
Biochemistry

Expression cloning of a human cDNA encoding folylpoly(γ-glutamate) synthetase and determination of its primary structure

TIMOTHY A. GARROW*, ARIE ADMON†, AND BARRY SHANE*‡

*Department of Nutritional Sciences and †Howard Hughes Medical Institute, Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720

Communicated by Jesse C. Rabinowitz, June 25, 1992

ABSTRACT **A human cDNA for folylpoly(γ-glutamate) synthetase [FPGS; tetrahydrofolate:L-glutamate γ-ligase (ADP forming), EC 6.3.2.17] has been cloned by functional complementation of an *Escherichia coli folC* mutant. The cDNA encodes a 545-residue protein of M_r 60,128. The deduced sequence has regions that are highly homologous to peptide sequences obtained from purified pig liver FPGS and shows limited homology to the *E. coli* and *Lactobacillus casei* FPGSs. Expression of the cDNA in *E. coli* results in elevated expression of an enzyme with characteristics of mammalian FPGS. Expression of the cDNA in AUXB1, a mammalian cell lacking FPGS activity, overcomes the cell's requirement for thymidine and purines but does not overcome the cell's glycine auxotrophy, consistent with expression of the protein in the cytosol but not the mitochondria.**

Cellular folates exist primarily as poly(γ-glutamate) derivatives with typical peptide chains ranging from five to nine residues in mammalian tissues (1). Metabolism of pteroylmonoglutamates to polyglutamates, catalyzed by the enzyme folylpoly(γ-glutamate) synthetase [FPGS; tetrahydrofolate:L-glutamate γ-ligase (ADP forming), EC 6.3.2.17], allows tissues to concentrate folate at higher levels than in plasma. In addition, folylpolyglutamates are the active coenzymatic forms of the vitamin and display increased affinities or lowered K_m values for most of the enzymes of one carbon metabolism (reviewed in refs. 1 and 2). The essential role of folylpolyglutamate synthesis was first demonstrated in a Chinese hamster ovary (CHO) cell mutant (AUXB1) that lacks FPGS activity (3). This cell has an impaired ability to accumulate folate and is auxotrophic for products of one carbon metabolism such as purines, thymidine, and glycine (3–5).

Metabolism of antifolate drugs to polyglutamate forms also plays a role in their cytotoxic efficacy due to their increased affinity for target enzymes and the increased retention of drug within the cell, whereas a decrease in FPGS activity can lead to resistance to antifolates (refs. 6–9; J.-S. Kim and B.S., unpublished data). FPGS is a potential target for antifolate drugs and an understanding of the catalytic mechanism and the specificity of the substrate binding sites of FPGS should aid in the design of antifolate agents.

FPGS has been purified to homogeneity from *Corynebacterium* sp. (10), *Lactobacillus casei* (11), and *Escherichia coli* (12), and the *E. coli* and *L. casei* genes have been cloned and sequenced and the proteins have been overexpressed (12–14). However, the bacterial enzymes are poor models for mammalian FPGS as they can only metabolize folates to short polyglutamate derivatives and they display a folate substrate specificity quite distinct from the mammalian enzyme. Some characterizations of crude or partially purified rat (15), mouse (16), beef (17), and human (18) liver FPGS have been reported and pig liver FPGS has been purified to homogeneity (19) but, in each case, only small amounts of protein have been obtained, which has limited characterization to kinetic analyses (15–22). The low abundance and instability of mammalian FPGS have complicated its purification in sufficient quantity to carry out mechanistic studies.

CHO–human (23) and CHO–mouse (24) hybrids have been used to localize the human and murine FPGS genes to chromosomes 9 and 2, respectively. To aid in the further characterization of the mammalian protein and to study its regulation, we initially attempted to clone the human FPGS gene by purifying human sequences capable of complementing CHO AUXB1 cells to the wild-type phenotype using multiple rounds of DNA transfection (25). Although CHO cells expressing human FPGS were obtained, we were unable to isolate unique size fragments when restricted DNA from independent transfectants was probed with human *Alu* sequences or bulk human DNA, and we were unable to complement AUXB1 cells with various human cosmid libraries. As an alternative approach, we have attempted to purify and sequence peptides derived from pig liver FPGS to develop oligonucleotide probes for the isolation of a porcine cDNA. During the course of these studies, λ-YES, an efficient system for the expression cloning of cDNAs in bacteria and yeast, was described by Elledge *et al.* (26).

In this report we describe the cloning of a human FPGS cDNA[§] by its complementation of an *E. coli* FPGS mutant and the ability of the cDNA to functionally complement the CHO FPGS mutant AUXB1, and we compare its deduced amino acid sequence with that of pig liver FPGS peptides and bacterial FPGS proteins.

MATERIALS AND METHODS

Materials. L-[U-^{14}C]Glutamic acid (270 mCi/mmol; 1 Ci = 37 GBq) and [α-^{32}P]dATP (7000 Ci/mmol) were obtained from Amersham. *Staphlococcus* V8 protease was obtained from Promega. Pig livers were generously supplied by the Department of Animal Science at the University of California, Davis.

Bacterial, Bacteriophage, and Plasmid Strains and Media. The *E. coli* mutant SF4 (F^-, *folC, strA, recA, tn10:srlC*) and its parent strain W1485 (F^-) have been described (12, 13). SF4 is defective in FPGS and dihydrofolate synthetase (DHFS) activities, both of which are encoded by the *folC* gene (13), and requires methionine (50 μg/ml) and glycine (50 μg/ml) for growth when cultured in minimal medium (12).

Bacteriophage λYES-R, a human cDNA library in λYES-R, and *E. coli* BNN132 containing λKC were gifts from

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: FPGS, folylpoly(γ-glutamate) synthetase; DHFS, dihydrofolate synthetase; $H_4PteGlu_n$, tetrahydropteroylpoly(γ-glutamate), *n* indicating the number of glutamate moieties; H_2Pte, dihydropteroate; CHO, Chinese hamster ovary.

‡To whom reprint requests should be addressed.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M98045).

S. Elledge (Baylor College of Medicine, Waco, TX). The construction of λKC and λYES-R has been described (26). The *cre* gene on λKC allows automatic subcloning of plasmid pSE936, contained between *lox* sites on λYES-R, when *E. coli* is infected with λYES-R (26). The human cDNA library, containing *Eco*RI–*Xho* I–*Bgl* I linkers, was made from mRNA derived from Epstein–Barr virus-transformed B lymphocytes and was cloned into a unique *Eco*RI site located downstream from the *lac* promoter in the pSE936 region of λYES-R (26). λKC was rescued from BNN132 by mitomycin C induction and used to infect SF4 (26). Kanamycin-resistant colonies were tested for the correct phenotype on VB plates containing kanamycin with or without methionine and glycine supplementation. Single cells of SF4(λKC) were isolated.

DNA Sequencing and Mutagenesis. *Eco*RI inserts of pSE936 that complemented SF4(λKC) were subcloned into pTZ18U and transformed into *E. coli* MV1190 (Bio-Rad). Single-stranded DNA, produced using helper phage M13KO7, was sequenced by the method of Sanger *et al.* (27) using Sequenase 2 (United States Biochemical). Primers were synthesized by the Micro-Chemical Facility (University of California, Berkeley).

An *Nco* I site was introduced at an ATG in the cDNA by oligonucleotide-directed mutagenesis using the method of Eckstein (28). *Nco* I–*Sal* I fragments were cloned into similarly treated pTrc99A (Pharmacia), a vector that contains the *trc* promoter and *lacZ* ribosome binding site for high expression of nonfusion proteins in *E. coli*.

FPGS and DHFS Assays. FPGS and DHFS activities were measured by incorporation of [^{14}C]glutamate (250 μM) into folate products using assay conditions described for pig liver FPGS (19) using 40 μM tetrahydropteroylmonoglutamate [(6*RS*)-H_4PteGlu (FPGS)] or 25 μM dihydropteroate [H_2Pte (DHFS)] as the substrate. The glutamate concentration was not saturating (12, 20). Reaction mixtures were incubated at 37°C for 1 hr. One unit of activity catalyzes the formation of 1 nmol of product per hr.

Purification of Pig Liver FPGS. FPGS was purified from pig liver (1.3 kg) as described (19) with the following modifications. Extraction and column buffers up to the chromatofocusing step contained Triton X-100 (0.02%) to stabilize enzyme activity. Enzyme eluted from phenyl-agarose was concentrated by reapplication to the phenyl-agarose column. Ammonium sulfate was added to 10% saturation to peak activity fractions and the extract was reapplied to a smaller phenyl-agarose column (2.5 cm × 1 cm) equilibrated with 100 mM Tris·HCl buffer (pH 8.4) containing 10% saturated ammonium sulfate and 50 mM 2-mercaptoethanol. The column was washed with equilibration buffer (3 ml) and enzyme was eluted with 80 mM Tris·HCl buffer (pH 8.4) containing 10% ethylene glycol (vol/vol), 0.02% Triton X-100, and 40 mM 2-mercaptoethanol.

Protein was determined by a modified Lowry procedure (29). Protein purity was determined by gradient (8–20%) SDS/PAGE with a 4% stacking gel. The discontinuous buffer system of Laemmli (30) was used. Protein bands were visualized by silver staining.

Sequencing of Pig Liver FPGS Peptides. One molar Tris buffer pH 8 (1.5 μl), 5% SDS (1.5 μl), and 100 mM EDTA (1.5 μl) were added to pig liver FPGS (2 μg) in 30 mM potassium phosphate buffer, pH 7.5/50 mM 2-mercaptoethanol (150 μl), and the mixture was heated at 95°C for 2 min and allowed to cool to room temperature. Twenty percent 4-vinylpyridine in isopropyl alcohol (3 μl) was added and the mixture was incubated at 37°C for 5 hr. V8 protease (0.5 μg) was added and FPGS was digested overnight at room temperature. Peptides were separated on a C_4 reverse-phase column using a gradient from 10% acetonitrile/0.1% trifluoroacetic acid (TFA) to 60% acetonitrile/30% *n*-propyl alcohol/0.1% TFA. Peptides were also separated by electrophoresis. In these cases, the 4-vinylpyridine step was omitted and, after proteolysis, the sample was lyophilized and peptides were separated by SDS/PAGE using a 4% stacking gel and a linear gradient separating gel (15–20%) using the buffer system of Schagger and von Jagow (31). Peptides were blotted onto a Pro-Blott membrane (Applied Biosystems) using 25 mM Tris·HCl/25 mM Tricine/15% methanol transfer buffer. Peptides were visualized by staining with 0.05% Coomassie blue/50% methanol followed by destaining in 50% methanol. The membrane was washed for 2 min in 10 mM NaCl followed by water, the membrane was then air dried, and the stained peptide bands were cut out and stored at 4°C until sequenced. Automated amino acid sequencing was performed using an Applied Biosystems 477A protein sequencer.

Transfection of CHO Cells. The 2.2-kilobase (kb) *Eco*RI insert from pSE936-25 (see below) was cloned into the *Eco*RI site of pSVK3 (Pharmacia) to give pSVK-hFPGS, and the orientation of the insert was checked by *Sma* I digestion. The *Eco*RI site in pSVK3 is downstream from the simian virus 40 early promoter and upstream from a small tumor antigen splice site and a poly(A) signal allowing expression of inserts in mammalian cells. CHO AUXB1 cells were cultured in α-minimum essential medium (α-MEM) medium containing 10% dialyzed fetal calf serum, PteGlu (2 μM), and glycine, thymidine, and hypoxanthine (5, 25). The cells were transfected with pSVK-hFPGS (10 μg) as described (25), and the ability of transfectants to grow in the absence of glycine, thymidine, and purines was assessed.

RESULTS

Cloning of Human FPGS by Complementation of SF4(λKC). SF4(λKC) cells (≈10^{10}) were infected with the human cDNA library in λYES-R (4 × 10^7 phage) as described by Elledge *et al.* (26) and cultured for 2 hr at 30°C in nonselective medium (plus methionine/glycine) containing 1 mM isopropyl β-D-thiogalactoside. Washed cells were then plated on selective agar plates (minimal medium plus 50 μg of ampicillin per ml). After 6 days at 30°C, 50 colonies were restreaked onto selective plates. Ten of the colonies continued to grow without methionine and glycine supplementation. Plasmid was isolated from the 10 transformants and used to transform SF4. Four plasmids retained the ability to complement the SF4 phenotype (pSE936-3, -8, -10, and -25).

Three plasmids contained *Eco*RI inserts of ≈2.2 kb and one (pSE936-10) contained an insert of ≈2.3 kb. All four plasmids had similar restriction maps. Table 1 shows FPGS activity in crude extracts of the SF4 transformants and in SF4 and its parent strain (W1845) when transformed with pSE936 lacking a cDNA insert. W1845 showed typical substrate specificities expected for the *E. coli* protein (12, 13). Activity was observed with H_2Pte as a substrate, whereas H_4PteGlu was a poor substrate (10-formyl-H_4PteGlu is the preferred monoglutamate substrate for *E. coli* FPGS), and activities were increased at higher KCl concentrations. SF4 displayed neg-

Table 1. FPGS and DHFS activities in *E. coli* transformants

	Specific activity, nmol/hr per mg			
	20 mM KCl		200 mM KCl	
Transformant	H_2Pte	H_4PteGlu	H_2Pte	H_4PteGlu
W1845/pSE936*	0.50	0.08	0.74	0.11
SF4/pSE936*	0	0	0	0
SF4/pSE936-3	0	24	0	15
SF4/pSE936-8	0	18	0	11
SF4/pSE936-10	0	26	0	15
SF4/pSE936-25	0	27	0	16

Bacteria were cultured overnight in LB medium containing ampicillin and cell extracts were prepared by sonication.
*No cDNA insert.

ligible DHFS and FPGS activity under these assay conditions. SF4 transformed with pSE936-3, -8, -10, and -25 displayed elevated FPGS activities but lacked DHFS activity and activities were reduced at higher KCl concentrations. These properties are typical of mammalian FPGS (1, 2, 19). Sequence analysis (below) indicated that the inserts in pSE936-3, -10, and -25 were in frame with an ATG translation start site in the pSE936 vector (26) and consequently the transformants may have expressed fusion proteins. The cDNA insert in pSE936-8 was out of frame with this ATG and translation must have started at an internal ATG in the cDNA, which may explain why FPGS activity in pSE936-8 transformants was lower than in other transformants.

Nucleotide Sequence of Human FPGS cDNA. The *Eco*RI inserts of the pSE936 vectors were cloned into an *Eco*RI site in pTZ18U for generating single-stranded DNA for sequencing. One insert (from pSE936-25) was completely sequenced in both orientations. Approximately 200 base pairs (bp) of the 5′ and 3′ ends of the other inserts were sequenced to verify that they contained essentially identical inserts except for a few base pairs at the 5′ and 3′ ends. The inserts in pTZ18U complemented SF4 if cloned in an orientation that allowed expression from the *lac* promoter.

The cDNA sequence of human FPGS and the deduced protein sequence are shown in Fig. 1. All inserts contained a potential poly(A) signal and one insert (from pSE936-10) contained a stretch of about 80 As that accounted for the slightly larger size of this insert. The open reading frame codes for a protein of 545 amino acid residues with a predicted M_r of 60,128. This is similar to the molecular weight of pig liver FPGS (M_r 60,000) and larger than that of the bacterial FPGS proteins (M_r ≈45,000).

```
   GCGCGGCATA ACGACCCAGG TCGCGGCGCG GCGGGGCTTG AGCGCGTGGC CGGTGCCGCA GGAGCCGAGC ATG GAG TAC CAG GAT    85
   *       **      *                                                          MET Glu Tyr Gln Asp     5

GCC GTG CGC ATG CTC AAT ACC CTG CAG ACC AAT GCC GGC TAC CTG GAG CAG GTG AAG CGC CAG CGG GGT GAC CCT   160
Ala Val Arg Met Leu Asn Thr Leu Gln Thr Asn Ala Gly Tyr Leu Glu Gln Val Lys Arg Gln Arg Gly Asp Pro    30

CAG ACA CAG TTG GAA GCC ATG GAA CTG TAC CTG GCA CGG AGT GGG CTG CAG GTG GAG GAC TTG GAC CGG CTG AAC   235
Gln Thr Gln Leu Glu Ala Met Glu Leu Tyr Leu Ala Arg Ser Gly Leu Gln Val Glu Asp Leu Asp Arg Leu Asn    55

ATC ATC CAC GTC ACT GGG ACG AAG GGG AAG GGC TCC ACC TGT GCC TTC ACG GAA TGT ATC CTC CGA AGC TAT GGC   310
Ile Ile His Val Thr Gly Thr Lys Gly Lys Gly Ser Thr Cys Ala Phe Thr Glu Cys Ile Leu Arg Ser Tyr Gly    80

CTG AAG ACG GGA TTC TTT AGC TCT CCC CAC CTG GTG CAG GTT CGG GAG CGG ATC CGC ATC AAT GGG CAG CCC ATC   385
Leu Lys Thr Gly Phe Phe Ser Ser Pro His Leu Val Gln Val Arg Glu Arg Ile Arg Ile Asn Gly Gln Pro Ile   105

AGT CCT GAG CTC TTC ACC AAG TAC TTC TGG CGC CTC TAC CAC CGG CTG GAG GAG ACC AAG GAT GGC AGC TGT GTC   460
Ser Pro Glu Leu Phe Thr Lys Tyr Phe Trp Arg Leu Tyr His Arg Leu Glu Glu Thr Lys Asp Gly Ser Cys Val   130

TCC ATG CCC CCC TAC TTC CGC TTC CTG ACA CTC ATG GCC TTC CAC GTC TTC CTC CAA GAG AAG GTG GAC CTG GCA   535
Ser Met Pro Pro Tyr Phe Arg Phe Leu Thr Leu Met Ala Phe His Val Phe Leu Gln Glu Lys Val Asp Leu Ala   155

GTG GTG GAG GTG GGC ATT GGC GGG GCT TAT GAC TGC ACC AAC ATC ATC AGG AAG CCT GTG GTG TGC GGA GTC TCC   610
Val Val Glu Val Gly Ile Gly Gly Ala Tyr Asp Cys Thr Asn Ile Ile Arg Lys Pro Val Val Cys Gly Val Ser   180

TCT CTT GGC ATC GAC CAC ACC AGC CTC CTG GGG GAT ACG GTG GAG AAG ATC GCA TGG CAG AAA GGG GGC ATC TTT   685
Ser Leu Gly Ile Asp His Thr Ser Leu Leu Gly Asp Thr Val Glu Lys Ile Ala Trp Gln Lys Gly Gly Ile Phe   205

AAG CAA GGT GTC CCT GCC TTC ACT GTG CTC CAA CCT GAA GGT CCC CTG GCA GTG CTG AGG GAC CGA GCC CAG CAG   760
Lys Gln Gly Val Pro Ala Phe Thr Val Leu Gln Pro Glu Gly Pro Leu Ala Val Leu Arg Asp Arg Ala Gln Gln   230

ATC TCA TGT CCT CTA TAC CTG TGT CCG ATG CTG GAG GCC CTC GAG GAA GGG GGG CCG CCG CTG ACC CTG GGC CTG   835
Ile Ser Cys Pro Leu Tyr Leu Cys Pro Met Leu Glu Ala Leu Glu Glu Gly Gly Pro Pro Leu Thr Leu Gly Leu   255

GAG GGG GAG CAC CAG CGG TCC AAC GCC GCC TTG GCC TTG CAG CTG GCC CAC TGC TGG CTG CAG CGG CAG GAC CGC   910
Glu Gly Glu His Gln Arg Ser Asn Ala Ala Leu Ala Leu Gln Leu Ala His Cys Trp Leu Gln Arg Gln Asp Arg   280

CAT GGT GCT GGG GAG CCA AAG GCA TCC AGG CCA GGG CTC CTG TGG CAG CTG CCC CTG GCA CCT GTG TTC CAG CCC   985
His Gly Ala Gly Glu Pro Lys Ala Ser Arg Pro Gly Leu Leu Trp Gln Leu Pro Leu Ala Pro Val Phe Gln Pro   305

ACA TCC CAC ATG CGG CTC GGG CTT CGG AAC ACG GAG TGG CCG GGC CGG ACG CAG GTG CTG CGG CGC GGG CCC CTC  1060
Thr Ser His Met Arg Leu Gly Leu Arg Asn Thr Glu Trp Pro Gly Arg Thr Gln Val Leu Arg Arg Gly Pro Leu   330

ACC TGG TAC CTG GAC GGT GCG CAC ACC GCC AGC AGC GCG CAG GCC TGC GTG CGC TGG TTC CGC CAG GCG CTG CAG  1135
Thr Trp Tyr Leu Asp Gly Ala His Thr Ala Ser Ser Ala Gln Ala Cys Val Arg Trp Phe Arg Gln Ala Leu Gln   355

GGC CGC GAG AGG CCG AGC GGT GGC CCC GAG GTT CGA GTC TTG CTC TTC AAT GCT ACC GGG GAC CGG GAC CCG GCG  1210
Gly Arg Glu Arg Pro Ser Gly Gly Pro Glu Val Arg Val Leu Leu Phe Asn Ala Thr Gly Asp Arg Asp Pro Ala   380

GCC CTG CTG AAG CTG CTG CAG CCC TGC CAG TTT GAC TAT GCC GTC TTC TGC CCT AAC CTG ACA GAG GTG TCA TCC  1285
Ala Leu Leu Lys Leu Leu Gln Pro Cys Gln Phe Asp Tyr Ala Val Phe Cys Pro Asn Leu Thr Glu Val Ser Ser   405

ACA GGC AAC GCA GAC CAA CAG AAC TTC ACA GTG ACA CTG GAC CAG GTC CTG CTC CGC TGC CTG GAA CAC CAG CAG  1360
Thr Gly Asn Ala Asp Gln Gln Asn Phe Thr Val Thr Leu Asp Gln Val Leu Leu Arg Cys Leu Glu His Gln Gln   430

CAC TGG AAC CAC CTG GAC GAA GAG CAG GCC AGC CCG GAC CTC TGG AGT GCC CCC AGC CCA GAG CCC GGT GGG TCC  1435
His Trp Asn His Leu Asp Glu Glu Gln Ala Ser Pro Asp Leu Trp Ser Ala Pro Ser Pro Glu Pro Gly Gly Ser   455

GCA TCC CTG CTT CTG GCG CCC CAC CCA CCC CAC ACC TGC AGT GCC AGC TCC CTC GTC TTC AGC TGC ATT TCA CAT  1510
Ala Ser Leu Leu Leu Ala Pro His Pro Pro His Thr Cys Ser Ala Ser Ser Leu Val Phe Ser Cys Ile Ser His   480

GCC TTG CAA TGG ATC AGC CAA GGC CGA GAC CCC ATC TTC CAG CCA CCT AGT CCC CCA AAG GGC CTC CTC ACC CAC  1585
Ala Leu Gln Trp Ile Ser Gln Gly Arg Asp Pro Ile Phe Gln Pro Pro Ser Pro Pro Lys Gly Leu Leu Thr His   505

CCT GTG GCT CAC AGT GGG GCC AGC ATA CTC CGT GAG GCT GCT GCC ATC CAT GTG CTA GTC ACT GGC AGC CTG CAC  1660
Pro Val Ala His Ser Gly Ala Ser Ile Leu Arg Glu Ala Ala Ala Ile His Val Leu Val Thr Gly Ser Leu His   530

CTG GTG GGT GGT GTC CTG AAG CTG CTG GAG CCC GCA CTG TCC CAG TAG CCAAGGCCCG GGGTTGGAGG TGGGAGCTTC     1738
Leu Val Gly Gly Val Leu Lys Leu Leu Glu Pro Ala Leu Ser Gln •••                                       545

CCACACCTGC CTGCGTTCTC CCCATGAACT TACATACTAG GTGCCTTTTG TTTTTGGCTT TCCTGGTTCT GTCTAGACTG GCCTAGGGGC  1828

CAGGGCTTTG GGATGGGAGG CCGGGAGAGG ATGTCTTTTT TAAGGCTCTG TGCCTTGGTC TCTCCTTCCT CTTGGCTGAG ATAGCAGAGG  1918

GGCTCCCCGG GTCTCTCACT GTTGCAGTGG CCTGGCCGTT CAGCCTGTCT CCCCCAACAC CCCGCCTGCC TCCTGGCTCA GGCCCAGCTT  2008

ATTGTGTGCG CTGCCTGGCC AGGCCCTGGG TCTTGCCATG TGCTGGGTGG TAGATTTCCT CCTCCCAGTG CCTTCTGGGA AGGGAGAGGG  2098

CCTCTGCCTG GGACACTGCG GCACAGAGGG TGGCTGGAGT GAATTAAAGC CTTTGTTTTT AAAAA..... .....AAAAA             2250
                                            **   *                          *
```

FIG. 1. Nucleotide sequence of human FPGS cDNA and derived amino acid sequence. The amino acid sequence is numbered from the first ATG codon. Asterisks (*) indicate the initial nucleotide in the insert in pSE936-3, -8, -25, and -10, respectively, and the final nucleotide in the insert in pSE936-3, -25, -8, and -10, respectively. The presumed poly(A) signal is doubly underlined. A four-base sequence that could potentially act as a Shine–Dalgarno sequence for translation in *E. coli* is singly underlined.

Table 2. Purification of pig liver FPGS

Fraction	Volume, ml	Activity, units*/ml	Protein, mg/ml	Specific activity, units/mg	Purification, fold	Yield, %
Crude extract	2950	19	69	0.28	1.0	100
Streptomycin sulfate	2925	18	42	0.43	1.5	97
20–40% $(NH_4)_2SO_4$	1200	33	27	1.2	4.3	71
Phosphocellulose peak	65	359	19	19	68	42
Chromatofocusing peak	190	56	0.11	510	1,820	19
Phenyl-agarose peak	15	592	0.043	13,770	49,200	16
Hydroxylapatite peak	20	374	0.0073	51,230	183,000	13

*nmol/hr.

A polypurine tract 7 bp prior to the presumptive start ATG has homology with bacterial Shine–Dalgarno sequences (Fig. 1) and may explain the expression of FPGS activity in bacteria transformed with pSE936-8. G-69 in the cDNA sequence (Fig. 1) was changed to a C to generate an *Nco* I site and a 2.1-kb *Nco* I–*Sal* I fragment was cloned into pTrc99A. SF4 transformed with this construct expressed active nonfusion FPGS. Met-37 in the deduced sequence is also preceded by a polypurine tract. However, expression of the cDNA from Met-37 in pTrc99A did not result in active enzyme.

Homology with Pig Liver FPGS Peptides. The purification of pig liver FPGS is summarized in Table 2. The protein was purified about 180,000-fold and displayed a single band of M_r 60,000 on a SDS/PAGE gel. The protein had a blocked N terminus, and multiple attempts at the generation of peptides using various proteases were unsuccessful due to the formation of insoluble peptides or protein and/or very poor recovery of peptides from reverse-phase columns. This mirrored some of our past experience with the *E. coli* protein. Although FPGS is a soluble protein, it behaves like a hydrophobic protein and it is difficult to resolubilize precipitated protein or peptides. Four peptides were isolated and sequenced after V8 protease treatment in the presence of SDS. Under the conditions used (phosphate buffer), V8 protease can cleave on the C-terminal side of Glu and Asp residues. The sequences obtained, AVRILNTLQ*T*NA, GGPPLTLGLEGEHQRTNAA, *X*IRINGQPI*G*PE, and AT*SRP*SL*L*GQLP, were highly homologous to regions of the deduced human sequence (Fig. 2) with one peptide sequence starting at residue six of the deduced sequence. Italicized residues reflect some ambiguity in the assignment.

Homology with Bacterial FPGS. A comparison of the deduced human FPGS protein sequence with the GenBank and EMBL nucleic acid data bases using TFASTN and TBLASTN failed to pick up any significant homology with any sequences except for *E. coli* and *L. casei* FPGS. A comparison of the three protein sequences is shown in Fig. 2. The C-terminal region of the proteins are omitted as there is no significant homology between any of these proteins in this region. In the region shown there is about 25% residue identity between the human and each bacterial sequence. There are only limited areas of homology between the three proteins, and the two major conserved regions, around human FPGS residues 59–67 [VXGTXGKG(S/T)] and 153–171 [DX$_4$EVG(I/L)GGXXDXTN(I/V)(I/V)], are the presumptive A and B nucleotide binding sites (13, 14).

Expression in CHO Cells. Expression of the cDNA (pSVK-hFPGS) in AUXB1 cells restored their ability to grow in the absence of purines and thymidine but the transfectants remained auxotrophic for glycine.

DISCUSSION

We have screened an Epstein–Barr virus-transformed human lymphocyte cDNA library using functional complementation of an *E. coli* FPGS mutant to isolate a cDNA for human FPGS. The deduced protein sequence is highly homologous to pig liver FPGS peptides and is of similar size to the pig liver protein and considerably larger than bacterial FPGSs.

```
Pig          AVRILNTLQTNA                                                                                   IRI
             |||•|||||||||                                                                                  |||
Human   MEYQDAVRMLNTLQTNAGYLEQVKR QRGDPQTQLEAMELYLARSGLQVED LDRLNIIHVTGTKGKGSTCAFTECI LRSYGLKTGFFSSPHLVQVRERIRI  100
        | • • •  • •• •  ••••••    •      |• • |  ••  • •     •  • •|•|| |||•••   • • |   |•• | ••||•••   ||•••
Coli    MIIKRTPQAASPLASWLSYLENL-H SK-TIDLGLERVSLVAAR--LGVLK PAPF-VFTVAGTNGKGTTCRTLESI LMAAGYKVGVYSSPHLVRYTERVRV   95
Casei   MNYTETVAYIHSF-PRLAKTGD--H RR-ILTL-LH--ALGNPQ------Q QGRY-I-HVTGTNGKGSAANAIAHV LEASGLTVGLYTSPFIMRFNERIMI   85

Pig     NGQPIGPE
        ||||| ||
Human   NGQPISPELFTKYFWRLYHRLEETK DGSCVSMPPYFRFLTLMAFHVFLQE KVDLAVVEVGIGGAYDCTNIIRKPV VCGVSSLGIDHTSLLGDTVEKIAWQ  200
        •••••           •   •|     •       •| ••|•••• •| |   •|••••|||•||  | ||••  •• |  ••••••||• •||  ••• |•
Coli    QGQELPESA--HTASF--AEIES-- -ARGDISLTYFEYGTLSALWLFKQA QLDVVILEVGLGGRLDATNIV-DAD VAVVTSIALDHTDWLGPDRESIGRE  187
Casei   DHEPIPDAALVNAVAFVRAALERLQ QQQADFNVTEFEFITALGYWYFRQR QVDVAVIEVGIGGDTDSTNVI-TPV VSVLTEVALDHQKLLGHTITAIAKH  184

Pig                                                         GGPPLTLGLEGEHQRTNAA                       ATSRPSLLGQ
                                                            |||||||||||||||•|||                          ||| || |
Human   KGGIFKQGVPAFTV-LQPEGPLAVL RDRAQQISCP-LYLCPMLEALEE-- GGPPLTLGLEGEHQRSNAALALQLA HCWLQRQDRHGAGEPKASRPGLLWQ  296
        |••|•• ••|• •   •|••• •••  ••••      •               •       ••   ••   | • |      |••   • •  •••     •
Coli    KAGIFRSEKPAIVGE--PEMP-STI ADVAQEKGALLQRRGVEWNYSVTD- HDWA--FS-DAHG--TLENLPLPLV ---PQPNAATALAALRASGLE----  271
Casei   KAGIIKRGIPVVTGNLVPDAA-AVV AAKVATTGSQWLRFDRDFSVPKAKL HGWGQRFTYEDQDGRI-SDLEVPLV GDYQQRNMAIAIQTAKVYAKQTEW-  281

Pig     LP
        ||
Human   LPLAPVFQPTSHMRLGLRNTEWPGR TQVLRRGPLTWYLDGAHTASSAQAC VRWFRQALQGRERPSGGPEVRVLLF NATGDRDPAALLKLLQPCQFDYAVF  396
         •• •       •| |•  • •|•|  •••    |•  • |•||     •••  •  ••              • •••     •|•| |•••  |     •  •
Coli    --VSE-----NAIRDGIASAILPGR FQIVSESPRVIF-DVAHNPHAAEYL TGRMKALPKNGR-------VLAVI- GMLHDKDIAGTLAWLKSVVDDWYCA  355
Casei   -PLTP-----QNIRQGLAASHWPAR LEKISDTPLIVI-DGAHNPDGINGL ITALKQLFSQP--------ITVIA- GILADKDYAAMADRLTAAFSTVYLV  365
```

FIG. 2. Comparison of the amino acid sequence of human FPGS with the *E. coli* (422 residues) and *L. casei* (428 residues) proteins and pig liver FPGS peptides. The C-terminal residues of the three proteins are omitted. The original alignment was made using a PAM 256 matrix and minor modifications were made manually. Lines (|) indicate amino acid identity between the three proteins. Dots (•) indicate amino acid identity between the human sequence and one of the bacterial sequences or amino acid similarity between the human sequence and one or both of the bacterial sequences. Similarity is defined as (E/D), (N/Q), (K/R), (A/G), (S/T), (F/Y), and (I/L/M/V).

FPGS activity was elevated in *E. coli* transformed with plasmids containing the human cDNA although the specific activities in crude bacterial extracts were <0.1% of that of homogeneous pig liver FPGS. However, these specific activities were still >500-fold higher than we have previously found in crude extracts of human liver (unpublished data). We must attempt to improve its expression further as a prelude to enzyme purification. The substrate specificities noted for the expressed cDNA were those expected for mammalian FPGS and different from bacterial enzyme.

The 2.3-kb cDNA appears to be of near full length. The 3′ poly(A) tail was retained in one clone, and Northern analysis of mRNA from human cells or AUXB1 cells transfected with human genomic DNA and expressing human FPGS activity (8, 25) demonstrated a single band of ≈2.3 kb; the intensity of the signal was proportional to the level of expressed FPGS activity and was consistent with a very low abundance mRNA. No hybridization signal was detected with mRNA from wild-type CHO cells, AUXB1 cells, or pig liver unless the stringency was greatly reduced (J.-S. Kim, L. Chen, and B.S., unpublished data). FPGS activity is located in mitochondria and cytosol of mammalian cells and expression of the enzyme in mitochondria is required for glycine synthesis (refs. 8; B. F. Lin and B.S., unpublished data). Transfection of the cDNA in a mammalian expression vector in AUXB1 cells restored cytosolic folate metabolism but not mitochondrial metabolism. Although this suggests that the isolated cDNA codes for the cytosolic form of FPGS, a mitochondrial isoform cannot be eliminated as the cDNA was obtained by expression cloning, and expression of a full-length mitochondrial mRNA may have been selected against as it may not have produced an active protein in bacteria. The nucleotides 5′ to the presumed ATG start site would code for the sequence RGITTQVAARRGLSAWPVPQEPS, which shares features similar to the motif reported for mitochondrial leader sequences (32), and it is possible that the cDNA codes for the mitochondrial isozyme but lacks the start ATG. Reversion frequencies of the AUXB1 mutant are consistent with a single mutation causing the multiple auxotrophy (3, 4, 25), and in multiple transfection studies with human genomic DNA we have always observed restoration of mitochondrial and cytosolic FPGS activities in transfectants (unpublished data). These data are consistent with a single gene coding for both isozymes. As only one mRNA species was detected by Northern analysis, it is possible that a single mRNA may be responsible for mitochondrial and cytoplasmic forms of the protein (33).

The peptides isolated from pig liver FPGS showed a high degree of homology with the human protein and the alignment of one of the peptides, starting at residue six, supports the deduced start site of the protein. Human FPGS shows only limited homology with the *E. coli* and *L. casei* proteins and the two bacterial proteins are only homologous in limited regions. This may reflect the different substrate specificities of the proteins. The bacterial proteins have different preferred pteroylmonoglutamate substrates, only effectively use 5,10-methylene-H_4PteGlu$_n$ as their polyglutamate substrates, metabolize folates to only the tri- or tetraglutamate, and have a >100-fold difference in their affinities for MgATP (10–12). In addition, the *E. coli* protein also possesses DHFS activity, whereas the *L. casei* and mammalian proteins lack this activity or any affinity for H_2Pte. We have recently cloned two yeast genes that complement the *E. coli* SF4 mutant (A. Brenner and B.S., unpublished data). One of these codes for a DHFS and most closely resembles the bacterial proteins, whereas the other codes for a FPGS and most closely resembles the human protein.

The availability of a human cDNA for FPGS will allow studies on the regulation of expression of the protein and on the mechanism of subcellular localization of the protein. Studies on the overexpression of the protein are necessary to allow purification and more detailed studies on the mechanism of the FPGS reaction.

We thank Stephen Elledge (Baylor College of Medicine) for his generous gift of the human cDNA library in λ-YES, Yun-Jung Choi for assistance in transfecting CHO cells, Jong-Sang Kim and Linda Chen for the Northern analyses, and Al Brenner for his advice. This work was supported by National Cancer Institute Grant CA41991 to B.S.

1. Shane, B. (1989) *Vitam. Horm. (NY)* **45,** 263–335.
2. McGuire, J. J. & Coward, J. K. (1984) in *Folates and Pterins*, eds. Blakley, R. L. & Benkovic, S. J. (Wiley Interscience, New York), Vol. 1, pp. 135–190.
3. McBurney, M. W. & Whitmore, G. F. (1974) *Cell* **2,** 173–182.
4. Taylor, R. T. & Hanna, M. L. (1977) *Arch. Biochem. Biophys.* **181,** 331–344.
5. Foo, S. K. & Shane, B. (1982) *J. Biol. Chem.* **257,** 13587–13592.
6. Chabner, B. A., Allegra, C. J., Curt, G. A., Clendeninn, N. J., Baram, J., Koizumi, S., Drake, J. C. & Jolivet, J. (1985) *J. Clin. Invest.* **76,** 907–912.
7. Hagan, R. L., Duch, D. S., Smith, G. K., Hanlon, M. H., Shane, B., Freisheim, J. H. & Hynes, J. B. (1991) *Biochem. Pharmacol.* **41,** 781–787.
8. Shane, B., Lowe, K., Osborne, C., Kim, J. & Lin, B. (1990) in *Chemistry and Biology of Pteridines 1989*, eds. Curtius, H. C., Ghisla, S. & Blau, N. (de Gruyter, Berlin), pp. 891–895.
9. McCloskey, D. E., McGuire, J. J., Russell, C. A., Rowan, B. G., Bertino, J. R., Pizzorno, G. & Mini, E. (1991) *J. Biol. Chem.* **266,** 6181–6187.
10. Shane, B. (1980) *J. Biol. Chem.* **255,** 5655–5662.
11. Bognar, A. L. & Shane, B. (1983) *J. Biol. Chem.* **258,** 12574–12581.
12. Bognar, A. L., Osborne, C., Shane, B., Singer, S. C. & Ferone, R. (1985) *J. Biol. Chem.* **260,** 5625–5630.
13. Bognar, A. L., Osborne, C. & Shane, B. (1987) *J. Biol. Chem.* **262,** 12337–12343.
14. Toy, J. & Bognar, A. L. (1990) *J. Biol. Chem.* **265,** 2492–2499.
15. McGuire, J. J., Hsieh, P., Coward, J. K. & Bertino, J. R. (1980) *J. Biol. Chem.* **255,** 5776–5788.
16. Moran, R. G. & Colman, P. D. (1984) *Biochemistry* **23,** 4580–4589.
17. Pristupa, Z. B., Vickers, P. J., Sephton, G. B. & Scrimgeour, K. G. (1983) *Can. J. Biochem. Cell Biol.* **62,** 495–506.
18. Clarke, L. & Waxman, D. J. (1987) *Arch. Biochem. Biophys.* **256,** 585–596.
19. Cichowicz, D. J. & Shane, B. (1987) *Biochemistry* **27,** 504–512.
20. Cichowicz, D. J. & Shane, B. (1987) *Biochemistry* **27,** 513–521.
21. George, S., Cichowicz, D. J. & Shane, B. (1987) *Biochemistry* **27,** 522–529.
22. Moran, R. G., Colman, P. D., Rosowski, A., Forsch, R. A. & Chan, K. K. (1985) *Mol. Pharmacol.* **27,** 156–166.
23. Jones, C., Kao, F.-T. & Taylor, R. T. (1980) *Cytogenet. Cell Genet.* **28,** 181–194.
24. Fournier, R. E. K. & Moran, R. G. (1983) *Somatic Cell Genet.* **9,** 69–84.
25. Sussman, D. J., Milman, G. & Shane, B. (1986) *Somatic Cell Mol. Genet.* **12,** 531–540.
26. Elledge, S. J., Mulligan, J. T., Ramer, S. W., Spottswood, M. & Davis, R. W. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 1731–1735.
27. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
28. Nakamaye, K. & Eckstein, F. (1986) *Nucleic Acids Res.* **14,** 9679–9698.
29. Bensadoun, A. & Weinstein, D. (1970) *Anal. Biochem.* **70,** 241–250.
30. Laemmli, U. K. (1970) *Nature (London)* **227,** 680–685.
31. Schagger, H. & von Jagow, G. (1987) *Anal. Biochem.* **166,** 368–379.
32. Hendrick, J. P., Hodges, P. E. & Rosenberg, L. E. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 4056–4060.
33. Slusher, L. B., Gillman, E. C., Martin, N. C. & Hopper, A. K. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 9789–9793.