Nucleotide Sequence and Transcription Start Point of the Phosphoglycerate Transporter Gene of *Salmonella typhimurium*

DIANNE GOLDRICK, GUAN-QIAO YU, † SHU-QIN JIANG, AND JEN-SHIANG HONG*

Department of Cell Physiology, Boston Biomedical Research Institute, 20 Staniford Street, Boston, Massachusetts 02114

Received 7 December 1987/Accepted 27 April 1988

We identified the phosphoglycerate transporter gene of Salmonella typhimurium and its polypeptide product and determined the nucleotide sequence of the gene. The predicted translation product was a protein of 406 amino acid residues and was extremely hydrophobic, a feature that is consistent with its role in membrane transport. Hydropathy analysis suggested that there are eight transmembrane segments of at least 20 amino acid residues for the protein. The transcription start point was mapped to lie at position -44 relative to the putative translational initiation start point. Comparison of PgtP with UhpT and GlpT, the membrane-bound proteins involved in the transport of hexose-6-phosphate and glycerol-3-phosphate, respectively, revealed a very high degree of amino acid sequence similarity among them, reflecting not only similar structures and functions among these polypeptides but also a common evolutionary origin for them.

Phosphoenolpyruvate, 2-phosphoglycerate, and 3-phosphoglycerate (3-PG) are transported into Salmonella typhimurium via the inducible transport system pgt (14). Induction of transport occurs only when inducer is present extracellularly; no induction occurs in the absence of inducer, even though phosphoglycerates are present intracellularly at millimolar concentrations (14).

The pgt system has been cloned previously (11). In this report we describe the identification of the transporter gene pgtP, its polypeptide product and cellular location, the nucleotide sequence of the gene, and the transcription start point. Expression of the pgtP gene requires a functional pgtA gene. The nucleotide sequence of this gene, which encodes an activator protein, has been determined previously (21).

MATERIALS AND METHODS

Bacterial strains and phages. The bacterial strains used in this study were all *Escherichia coli* K-12 derivatives: BK9MDG (F^- *thi hsdR hsdM endB metC*) (13) and JM103 (*thi pro leu endA*). Phages M13mp18 and M13mp19 were used for gene sequencing.

Plasmids. The plasmids used in this study were derivatives of pBR322, pACYC184 (2), and pT7-1 or pT7-2 (17) and were constructed by standard methodologies. Plasmid pGP1-2 was a gift from S. Tabor and C. C. Richardson (Harvard Medical School, Boston, Mass.).

Media. The bacterial strains were grown in nutrient broth, YT, or medium E (18) containing 0.5% succinate or 0.4% 3-PG. When required, amino acids were added to final concentrations of 30 to 50 μ g/ml. The following antibiotics were used at the indicated concentrations: ampicillin, 35 μ g/ml; tetracycline, 15 μ g/ml; chloramphenicol, 30 μ g/ml.

Enzymes and chemicals. Restriction endonucleases and DNA enzymes were obtained from Bethesda Research Laboratories (Gaithersburg, Md.) and New England BioLabs, Inc. (Beverly, Mass.). All chemicals were reagent grade and were obtained from commercial sources.

Manipulations of DNA. Plasmid DNA was prepared from cleared lysates by CsCl-ethidium bromide centrifugation, as described by Davis et al. (4). The methods described by Maniatis et al. (12) were used for DNA manipulations.

Identification of gene products. The phage T7 RNA polymerase-T7 promoter coupled system of Tabor and Richardson (17) was used to identify gene products encoded by pgtP, with the exception that labeling with [³⁵S]methionine was done for 10 min instead of 5 min.

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis. For sodium dodecyl sulfate-polyacrylamide gel electrophoresis, 12% slab gels were used, and samples were boiled for 3 min prior to application. A series of cross-linked cytochrome c's were used as molecular weight standards. Gels were run at 30 mA of constant current for 4 h, stained with Coomassie brilliant blue, treated with En³Hance (New England Nuclear Corp., Boston, Mass.), dried, and exposed to X-ray film for autoradiography.

3-PG transport assays. Strain CSR603 harboring particular plasmids was grown at 37°C in minimal medium (medium E) that contained 0.5% succinate as a carbon source and that was supplemented with thiamine, threonine, leucine, proline, arginine, and the appropriate antibiotics. When growth reached the exponential phase, cells were collected by centrifugation and washed twice with and suspended in medium E to an optical density at 660 nm of 3.0. When induction of the *pgt* transport system was required, 0.2% 3-PG was added to exponentially growing cells, and the cells were harvested 2 h later.

3-PG transport was measured as follows. A portion $(25 \ \mu)$ of the cell suspension prepared as described above was incubated at 37°C for 2 min, when 1 μ l of 250 mM glucose was added. Fifteen seconds later, 1 μ l of 3-phospho[¹⁴C]glycerate (2.3 mM; specific activity, 55 mCi/mM) was added, and incubation was continued for the desired time intervals. To terminate transport, the mixture was diluted with 2 ml of medium E. Cells were collected on cellulose acetate membranes (pore size, 0.45 μ m; Schleicher & Schuell, Inc., Keene, N.H.) and washed once with 2 ml of medium E. Membranes were dried and counted in toluene-based Omnifluor (New England Nuclear) in a liquid scintillation counter.

Determination of cellular locations of pgt proteins. Cells (10

^{*} Corresponding author.

[†] Present address: Department of Molecular Genetics, Shanghai Institute of Plant Physiology, Academia Sinica, Shanghai, People's Republic of China.

ml) with [35S]methionine-labeled, plasmid pJH587-encoded proteins synthesized in the T7 RNA polymerase-T7 promoter coupled system (17) were pelleted by centrifugation, suspended in 0.6 ml of 0.1 M Tris hydrochloride (pH 7.6), and subjected to sonication 4 times for 15 s each time, with a 1-min cooling interval between sonications. After lowspeed centrifugation (4,000 rpm in a rotor [RC-5B; Ivan Sorvall, Inc., Norwalk, Conn.]) for 10 min to remove unbroken cells, the supernatant was layered on top of a 0.2-mlthick cushion in a tube containing 0.25 M sucrose, 10 mM Tris hydrochloride, (pH 7.6), 50 mM KCl, 10 mM MgCl₂, and 0.01% L-methionine and subjected to high-speed centrifugation (37,000 rpm in an SW50.1 rotor) for 2 h. The membrane fraction was suspended in 0.2 ml of 0.1 M Tris hydrochloride (pH 7.6) containing 0.4 M NaCl and then subjected to centrifugation as described above. Washing of the membrane with the buffer containing 0.4 M NaCl was repeated once.

Transcription start point determination. The site of transcription initiation was determined by the primer extension method of Hu and Davidson (10). RNA was isolated from plasmid pJH6-harboring strain BK9MDG grown on minimal medium containing 3-PG as the sole carbon and energy source by the method of Chen et al. (3), with the modification that phenol replaced *m*-cresol. The RNA was hybridized

to single-stranded M13mp18 carrying the 3.0-kilobase-pair (kbp) *Hin*dIII-*Pst*I fragment, and the RNA-DNA complex was used as a template for the extension by T4 DNA polymerase of the ³²P end-labeled hexadecameric primer 5'-TTCACCACACCCTTCA-3' (corresponding to positions -157 to -142 in Fig. 2) that were annealed to it. The reaction mixture was subjected to polyacrylamide gel electrophoresis with a control from which RNA was omitted and in parallel with corresponding sequence ladders, as described previously (10).

RESULTS AND DISCUSSION

Subcloning. We previously described (11) the cloning of the phosphoglycerate transport system of *S. typhimurium* LT-2 into pBR322. The plasmid, pBR322-pgt2, which contained a 14.4-kbp insert in pBR322 at the *Bam*HI site, was found to be capable of conferring on *E. coli* K-12 the ability to transport 3-PG and to utilize 3-PG as the sole carbon and energy source, suggesting that the genes for the entire *pgt* system are contained in the 14.4-kbp insert (11). This plasmid was renamed pJH5.

Subcloning of the 14.4-kbp insert was undertaken to localize genes of the pgt system. A series of subclones was constructed from plasmid pJH5, and their ability to confer



FIG. 1. Construction of *pgt* subclones from plasmid pJH5. Plasmids pJH12, pJH13, and pJH14 were derived from pJH5 by deleting *Eco*R1, *Hind*III, and *Sal*I fragments, respectively. Plasmid pJH6 was constructed by inserting the 7.6-kbp *Pst*I fragment into pBR322 at the *Pst*I site; plasmid pJH71 was constructed by inserting the 5.1-kbp *Bam*H1-*Hind*III fragment at the *Bam*H1 and *Hind*III sites of pBR322; and pJH72 was constructed by inserting the 7.6-kbp *Eco*R1-*Sal*I fragment at the *Eco*RI and *Sal*I sites of pBR322. Plasmid pJH501 was constructed by cloning the 4.5-kbp *Hind*III-*Sal*I-*Pst*I fragment from pJH6 into pACYC184. A 2.6-kbp *Sal*I fragment was then removed from pJH501, to yield pJH502, which then contained only a 2.7-kbp *Sal*I-*Pst*I fragment of the original 7.6-kbp insert of pJH6. Construction of pJH7 and pJH8 are described in the text. Restriction sites are abbreviated as follows: B, *Bam*H1; E, *Eco*R1; H, *Hind*III; P, *Pst*I; Hp, *Hpa*I; S, *Sal*I; S3, *Sau*3A. The broken bars indicate the locations of the *pgtP* and *pgtA* genes. Transport phenotypes are indicated as inducible (1), constitutive (C), or nonexpression (-).

Aat II -600 GGGGGATGAGGTGAGATCAGATGGACATTTTCCGCATTCGG	-550 CCTATCAAGTAATTGCTCTAAAGAACCCCCTCGTACCGTTGAGCGTACGAATCATGACGGAACCC		
-500 GGCTCOGTCTGCCAGOGCTGGATAATCCAGGOGGTGGGCCTGGG	-450 DTGAAAAGGT0GTTGCCATCACCAGCT0G00GCT0TA0GCCT0GCAAGAAC0GAACATTGC00GC		
-400 CAGCCATAGOGTAAAAGOGCCCCACTCTGGGGAAAOGACTACAG	-350 CIB I		
-300 -250 Acaaaacttacccititatttctgttggcatccagaacaacgitatttgaataataacggggtggcggaacaatgagggggggg			
-200 AMATTTGTTAATATÜÜGEACATGCTAACGATAAAACGGCCATGCCAATAATTAAAGTTTCACCACACCCCTTCAGGATGGGTCAAAAATAACCTTTAAATTCACCCAA			
-100 <u>-50</u> ISP -1 <u>TTGACTCAATTICTTGACACATTICTTGACTCAATTICTTATCTCACTCCTTCACCTCCTCATCTTCACTCAC</u>			
ATG TTA ACA ATA TTA AAA ACA GGG CAA TOG GGG Met Leu Thr Ile Leu Lys Thr Gly Gln Ser Als	CAT AAA GTC CCA COG GAA AAA GTC CAG GCC ACA TAT GGT GCA TAT GOT His Lys Vel Pro Pro Glu Lys Vel Gln Ale Thr Tyr Gly Arg Tyr Arg		
100 ATA CAG GCA TTG TTA AGT GTA TTT CTG GGA TAT Ile Gin Ale Leu Leu Ser Val Phe Leu Giv Tyr	CTC GGT TAC TAT ATA GTC AGA AAC AAC TTC ACA TTA TOG AOG GGT TAT Lau Ala TYT II VI II VAL ATA AAC AAC TAC TTC ACA TTA TOG AOG GGT TAT		
TTA AAA GAG CAA TTG GAT CTC AGC GCT AGG CAA	(40) 200 ATC GEC CTG CTG AGT AGC TGT ATG CTT ATT GGT TAC OGA ATC AGT AAA		
Leu Lys Glu Gln Phe Asp Leu Ser Als Thr Gln (60) 250	Ile Gly Leu Leu Ser Ser Cys Het Leu Ile Als Tyr Gly Ile Ser Lys (80) 300		
GGC GTA ÅTG AGC AGT CTG GGC GAT AAA CCC AGC Gly Val Met Ser Ber Leu Ala Asp Lys Ala Ser	CCA AAA GTC TTC ATG GGG TGC GGT GTG GTA GTC TGC GGG ATT GTT AAC Pro Lys Val Phe Net Als Cys Cly Leu Cys Als Ile Val Ass (100)		
GTT GOG CTG GGA TTC AGT AGC GCA TTC TGG ATA Val Gly Leu Gly Phe Ser Ser Als Phe Trp Ile	400 TTT GCC GCT CTG GTG GTC TTC AAT GGC CTT TTT CAG GGC ATG GOT GOG Phe Ale Ale Leu Vel Vel Phe Asn Gly Leu Phe Gin Gly Het Arg Arg		
CCC CTC GTT TAT TAC TAT TGC AAA CTG GTT CCT Pro Lew Val Tyr Tyr Tyr Cys Lys Lew Val Pro	450 657 COG GAG COC GTA GGC GCC TTC TOG AAT ATC TOG CAT AAC Are Are Giu Are Giu Are Tai Giu Ate Pas Tra Ame Tia Mer Ha Ame		
Val Gly Gly Gly Ile Val Ale Pro Ile Val Gly	Ale Ale Phe Ale Ile Leu Gly Ser Gin His Trp Gin Ser Ale Ser Tyr (180)		
ATC OTT COG GCC TGC GTC GCC GTC ATT TTT GCC Ile Vel Pro Ala Cys Vel Ala Vel Ile Phe Ala (200)	TTA ATT GTT CTG GTC TTG GGA AAA GGT TOG GGG GGC AAA GAA GGT GTT Leu lle Val Leu Val Phe Gly Lys Gly Ser Pro Arg Lys Amp Gly Les		
650 CCC TCC CTG GAA CAG ATG ATG CCG GAA GAA AAA Pro Ser Leu Glu Gln Met Met Pro Glu Glu Lys	TOO TO GTA CTG AAA AOG AAA AOG GCA AAA COG COG GAA AAT ATG AOC Vel Vel Leu Lys Thr Lys Asm Thr Ale Lys Ale Pro Clu Asm Met Ser		
(220) 750 GCA TGG CAA ATC TTC TGT ACT TAT GTC GTG CGC Ala Two Cia Lia Pha Cra Thr Twr Wal Wal Ara	ANT ANA ANT GCC TOG TAT ATT TOG CTG GTG GAT GTC TTC GTC TAT ATG Ant Ana Ant GCC TOG TAT ATT TOG CTG GTG GAT GTC TTC GTC TAT ATG Ant Ing Ann Ale Tro Try Tie Far Len Vel Tat Her Vel Try Her		
	(260)		
GTG CGG TTT GGC ATG ATT AGC TGG TTG CCT ATC Val Arg Phe Glu Met Ile Ser Trp Leu Pro Ile (280)	TAT CTG ${\rm \bar{T}TG}$ ACA GTA AAA CAC TTT TCA AAA GAA CAG ATG AGC GTC GOG Tyr Leu Leu Thr Vel Lys His Phe Ser Lys Glu Gln Met Ser Vel Ale		
TTT CTC TTT TTC GAG TGG GOG GCG ATT CCC TCC	AGE CTA CTE GCA CCC TCE CTE TCA GAT ANA CTE TTT ANG GGC CGC AGA		
Phe Leu Phe Phe Glu Trp Als Als Ile Pro Ser (300) 1000	Thr Leu Leu Ale Gly Trp Leu Ser Asp Lys Leu roe Lys Gly Arg Arg (320)		
ATG COG TTA GCC ATG ATT TGC ATG GCG CTG ATT Het Pro Leu Als Met Ile Cys Met Als Leu Ile	TTT GTC TGT GTG ATC GGA TAC TAG AAA AFT GAA TGT TTG GTG ATG GTA Phe Vel Cys Leu Ile GJY TYT Trp Lys Ser Glu Ser Leu Leu Met Vel (340)		
ACC ATT TTC GCC GCC ATT GTA GGT TGT CTG ATT Thr Ile Phe Als Als Ile Val Gly Cys Leu Ile (360)	TAC GTC COG CAG TTC CTC GCG TCC GTA CAG ACA ATG GAA ATA GTC CCC Tyr Vel Pro Gin Phe Leu Ale Ser Vel Gin Thr Net Giu Ile Vel Pro		
1150 AGC TTT GCC GTA GGT TCC GCC GTC GGT TTA GGT Ser Phe Ale Vel Gly Ser Ale Vel Gly Leu Arg	1200 GGG TTC ATC AGC TAT ATT TTC GGC GCC TOG TTC GGC ACC AGC CTG TTT Gly Phe Het Ser Tyr Ile Phe Gly Als Ser Leu Gly Thr Ser Leu Phe		
(380) 1250			
TOT TAATGGTGGTAAAACTTGGCTGGTACGGCGGATTTTATCTTUTUATGGCGGCATGJTCTGCTGCATTUTATTCTTTTACTCTUCCATGGCGGCACGGCA			
ACTEGAACECCAGGCCAGGATGCGTTACATAATCAGGACTCACTECAC			

FIG. 2. The nucleotide sequences of the *pgtP* gene and its 5'-flanking region and the deduced amino acid sequence of PgtP polypeptide. The antisense (mRNA-like) strand is shown. The 3.0-kbp *PstI-Hind*III and the 3.0-kbp *PstI-Bgl*II fragments from pJH6 were cloned into M13mp19 at the *PstI* and *Hind*III sites and the *PstI* and *Bam*HI sites, respectively. A series of deletions from each clone was generated by the method of Henikoff (9). Nucleotide sequences were determined by the M13-dideoxynucleotide chain-termination method (15). A complementary universal 15-bp oligodeoxynucleotide was used as primer, and $[^{35}S]$ dATP was used to label the products. Fractionation of the single-stranded DNA products of the primer elongation reaction was performed on 8% polyacrylamide gels. The numbers in parentheses indicate amino acid residues beginning from the N terminus. TSP indicates the transcription start point (see Fig. 3). The potential catabolite activator protein binding site is underlined with lines, with arrows, and the potential -10 and -35 promoter regions are indicated with thin lines.

3-PG transport was examined (Fig. 1). Deletion of the 3.3-kbp *Eco*RI fragment from the left arm of pJH5 yielding pJH12 had no effect on the inducible expression of 3-PG transport, indicating that this fragment contains none of the information needed for 3-PG transport or its regulation. However, deletion of the 9.0-kbp *Hind*III fragment from the left arm of pJH5, yielding pJH13, abolished 3-PG transport, indicating that the right half of the deleted fragment, namely, the 4.9-kbp *Eco*RI-*Hind*III fragment, contains information that is required for 3-PG transport, regulation, or both.

Deletion of the 3.9-kbp SalI fragment from the right arm of pJH5, yielding pJH14, also abolished 3-PG transport. When the 7.6-kbp PstI fragment was subcloned into pBR322 at the PstI site, the resulting clone conferred inducible 3-PG transport. However, subclones with inserts containing less than the full complement in pJH6, such as pJH71, pJH72, pJH501, or pJH502, conferred no 3-PG transport. Thus, it is evident that the genetic information necessary for inducible expression of 3-PG transport is contained in a 7.6-kbp PstI fragment. We have previously identified (21) a regulatory

gene that is needed for the expression of 3-PG transport in this region; this gene, pgtA, encodes an activator protein and is located on the right arm of the 7.6-kbp PstI fragment. Constitutive transport conferred by pJH7 and pJH8 (see below) suggests that a regulatory sequence is localized in the region of Sau3A-HindIII-SalI-PstI.

Localization of the transporter gene pgtP. To localize the structural gene(s) for the 3-PG transporter pgtP, subclones that were capable of conferring constitutive expression of 3-PG transport were sought. For this purpose, plasmid pJH5 was partially digested with Sau3AI. After electrophoresis fragments of 2 to 3 kbp in length were purified and ligated to pBR322 at the BamHI site. The ligation mixture was then used to transform strain BK9MDG, and transformants that were able to utilize 3-PG as a source of carbon and energy were selected. Several of the 3-PG⁺ clones were grown on minimal succinate medium in the absence of inducer (3-PG) and assayed for their ability to transport 3-PG. The two smallest plasmids, pJH7 and pJH8, which contained 2.2- and 3.3-kbp inserts, respectively, conferred a constitutive 3-PG transport ability (Fig. 1). Restriction analysis indicated that the inserts in these two plasmids correspond to the 1.6-kbp left arm of the 7.6-kbp insert in pJH6, as shown in Fig. 1, plus a short segment contiguous to its left end. For pJH7 this segment was 0.6 kbp, and for pJH8 it was 1.7 kbp. Thus, the structural gene(s) for the 3-PG transporter pgtP is contained within the 1.6-kbp PstI-HpaI-HpaI-Sau3AI sequence. The sequences to the left of the 1.6-kbp arm are not present in pJH6 and are therefore not required for pgt expression. The location of the pgtA gene reported previously (21) is also indicated in Fig. 1. Located between the pgtP and pgtA genes are two pgt genes that are involved in the induction process of the *pgtP* gene expression (unpublished data). Sequencing of these genes is in progress.

Identification of the gene product. To identify the pgtP gene product that is encoded in the insert of plasmid pJH7 and the transcriptional direction of the gene, the 1.8-kbp PstI-HindIII fragment of pJH7 was placed behind the phage T7 promoter of plasmids pT7-1 and pT7-2 at the PstI and HindIII sites, generating pJH586 and pJH587, respectively, and the plasmid-encoded products were identified by the T7 RNA polymerase-T7 promoter coupled system of Tabor and Richardson (17). Plasmid pJH587, which carried the insert with the T7 promoter proximal to the HindIII site, encoded a rather diffused, 37-kilodalton product, but no product was observed with pJH586, which carried the insert in the opposite orientation (data not shown). Thus, it is evident that the 37-kilodalton product is the 3-PG transporter encoded by *pgtP* and that the direction of transcription of the pgtP gene is from right to left in Fig. 1.

The product of the *pgtP* gene is membrane bound. To determine the cellular location of the *pgtP* gene product, cells carrying plasmid pJH587 were labeled with [³⁵S]methionine in the T7 RNA polymerase-T7 promoter coupled system as described above, sonicated, and centrifuged to separate the membrane fraction from the soluble fraction. The transporter expressed by pJH587 was found to be associated with the membrane fraction, whereas the mature periplasmic β -lactamase was found in the soluble fraction, as expected (data not shown). Repeated washing of the membranes with buffer containing 0.4 M NaCl did not dissociate the proteins from the membranes (data not shown).

Nucleotide sequence of pgtP gene. The entire nucleotide sequence of the 3.0-kbp PstI-HindIII fragment containing the pgtP gene was determined by using the M13-dideoxynucleotide chain-termination method (15). Both strands were sequenced. The sequences of the pgtP gene and its flanks are presented in Fig. 2. The pgtP gene is encoded in the sequence from positions 1 to 1218 with 406 amino acid residues. A putative Shine-Dalgarno sequence AGGTG at -10 to -6 precedes the coding frame. Sequences at positions -56 to -51 (5'-CACTCT) and -78 to -72 (5'-TT GAATT) are potential -10 and -35 promoter regions. At 10 bp upstream from these sequences (positions -106 to -89) is the 18-bp sequence 5'-TGAGTCAATTTTGACACA-3', which is a potential catabolite activator protein-binding site.

Transcription start point. The transcription start point of the pgtP gene was determined by the examination of RNA transcripts by hybridization mapping of mRNA isolated from the pJH6-carrying strain grown on 3-PG. We used T4 DNA polymerase to extend a radiolabeled primer annealed to a single-stranded DNA template in the presence of mRNA (10). Because of the inability of T4 DNA polymerase to displace a RNA hybridized to DNA, primer extension should stop at the 5' terminus of the hybridized mRNA, with the 3' end of the growing DNA chain thereby marking its position. The site of the first termination of primer extension in the presence of mRNA was at position -44, whereas no termination was observed in that region in the absence of added mRNA (Fig. 3). Thus, the transcription start point of



FIG. 3. Mapping of the transcription start point of the pgtP gene. The transcription start point was identified by primer extension analysis with T4 DNA polymerase, as described in the text. Lane 1, Primer extension in the absence of mRNA; lane 2, primer extension in the presence of mRNA; lanes 3 to 6, sequence ladders made by the dideoxynucleotide sequencing method with the same primer (unlabeled). Part of the nucleotide sequence deduced from the sequencing lanes is shown on the right, and the shortest extended primer segment is indicated with an asterisk.



FIG. 4. Analysis of the 3-PG transporter amino acid sequence for potential transmembrane regions. This was determined by the method of Engelman et al. (6) by using a window of 20 amino acids. The potential transmembrane hydrophobic regions are above the horizontal zero line, whereas regions with a relatively hydrophilic nature are below the zero line. The amino acid numbers are the same as those in Fig. 2.

G1pT PgtP UhpT pgtP probably lies at position -44 relative to the translational initiation start point. Significant primer extension occurred beyond position -44, probably because of the existence of partially degraded mRNA molecules and possibly because of limited mRNA displacement by T4 DNA polymerase.

Amino acid sequence and protein structure. Examination of the deduced amino acid sequence of the PgtP polypeptide indicates that the polypeptide is extremely hydrophobic, because 67% of its total amino acid residues contain nonpolar side chains. The PgtP protein appears to be highly positively charged, with a total of 35 arginine plus lysine residues but only 17 aspartic plus glutamic acid residues. The calculated M_r of 44,800 for PgtP is larger than the apparent M_r of 37,000 observed on sodium dodecyl sulfatepolyacrylamide gel electrophoresis. Such anomalous behavior can be attributed to the high degree of hydrophobicity of the protein and is apparently characteristic of hydrophobic membrane transport proteins, such as the lactose permease (1) and glucose-6-phosphate carrier of E. coli (7, 19). Examination of the PgtP protein for potential transmembrane regions by the method of Engelman et al. (6) indicated the presence of eight such regions (Fig. 4), suggesting that the polypeptide traverses the membrane 8 times. The amino

HLSIFKPAPHKARLPAAEIDPTYRRLRWQI PLCIPPCTAATTA WAN PALAMPYL VEQOFSROD	0.5
11 1 1 2 2 11 1 1 1 1 1 1 1 1 1 1 1 1 1	65
d : i i i i i i i i i i i i i i i i i i	66
	129
	125
10 LISCHILLIST I I I I I I I I I I I I I I I I I I	131
- COMPO-GNGWPPCGRTNVNWSQKURGCIVSVNNCAHNVCGCIPPLLFLLGNAWFNDWHAALYM	192
PHERE OF GHRRPLY-TYTCKLYPREMECHVGATHING MINNGCGUYAPIVGAAPAILGSQNQQSAS	188
LSOFT OSTGGSCST-STITKWTPARING TFLGTUNG SUID-GAGAAGVALFGAWTLFDGWT	191
PAPCAILVAL FAPANHRDTPQSCGLPPIBETKNDYPDDYN HAEQELTANQI PHOT	248
TITTALITLAPGKGSPRKDGLPSLEQMMPBEKAALKIKHTAKAPEMMSAUQIPCIT	251
I IIIIII I I IIIII -IGNPIFPSIIALIYGFIGLR-YGSDSPESYGLGKAEELFGEEISEEDGETESTDMTKUQIFYEI	254
	313
	316
	319
HEVEREN HEATENEN TO TIATI THE HEAGENET TO MICHITIGHT TO PROVIDE GLEAL ELARK	378
THELPREMEMPLANICHALT PVCLIGTHESESLLNVTIPAAIVGGLIT VOOPLASVQTMEIVIS	379
H-LANGINGCLVACIALALIIATLGVIQHASNETITLASLFALGVLVFGDQLLIGVAAVGFVHK	381
KAADIMADE TOLFOTTLOGSVAASAIVGYTVDFFGWDGGFNVHIGGSILAVILLIVVHIGEKRRHE	443
i i i i i i i i i i i i i i i i i i i	406
II II II II II II II II Maigaadogi kortrati igo sfaklgighiadgtpaygitgwagtfaaldiaaigcicimai⊽a⊽h	446
QLLQEKNGG	452
	463

FIG. 5. Alignment of amino acid sequences of PgtP, UhpT, and GlpT polypeptides. Identical positions between a pair of polypeptides are indicated by double dots, and those among the three polypeptides are boxed. Gaps were introduced to optimize the alignment, which was generated as described by Wilbur and Lipman (20).

terminus, which is relatively hydrophilic in composition (Fig. 4), is assumed to lie in the cytoplasm.

Codon usage. Of the 61 codons, 2 (AGG and GAC) were not used in the pgtP gene. From the analysis of Grosjean and Fiers (8), the pgtP gene appears to preferentially use the degenerate codons found in the weakly expressed genes and has a codon preference statistic of 0.38, which was calculated as described by Sharp and Li (16). This suggests that the pgtP gene belongs to a group of genes with a low codon bias.

Amino acid sequence homology with components of hexose-6-phosphate and glycerol-3-phosphate transport systems. The PgtP polypeptide (406 amino acid residues) has a high degree of amino acid sequence similarity with the UhpT polypeptide (463 amino acid residues), the membrane-bound transporter for the hexose-6-phosphate transport system, which, like the *pgt* system, is expressed only in the presence of exogenous inducers (7, 19); with the UhpC polypeptide (219 amino acid residues), a membrane-bound regulatory protein that is presumably involved in inducer recognition and binding in the regulation of uhpT expression and which has a high degree of sequence homology with UhpT (7); and with the GlpT polypeptide (452 amino acid residues), a membranebound glycerol-3-phosphate transporter (5). Allowing for a few small gaps and a misalignment by one amino acid residue at the N terminus, the sequences of PgtP and UhpT aligned well, with identical amino acid residues occupying 31% of the positions (Fig. 5). Approximately the same degree of sequence similarity was observed between PgtP and UhpC polypeptides (comparison not shown). The sequences of PgtP and GlpT also aligned well, with identical amino acid residues occupying 37% of the positions. As expected from the pairwise similarity observed here and that between GlpT and UhpT reported previously (5, 7), a high degree of similarity was observed among GlpT, PgtP, and UhpT polypeptides; 17% of the 406 positions were occupied by identical amino acids (Fig. 5). In addition to the sequence similarities among GlpT, PgtP, and UhpT, the hydropathy profiles of these polypeptides were also extremely similar (data not shown). Eiglmeier et al. (5) have noted previously that the hydropathy profiles between GlpT and UhpT are similar. These observations reflect not only similar structures and functions among these polypeptides but also a common evolutionary origin for them.

ACKNOWLEDGMENTS

We thank Brian Wallace for helpful discussions and comments on the manuscript.

This study was supported by Public Health Service grant GM31836 from the National Institute of General Medical Sciences. Computer resources used to carry out our studies were provided by the National Institutes of Health-sponsored BIONET National Computer Resource for Molecular Biology.

LITERATURE CITED

- 1. Buchel, D. E., B. Gronenborn, and B. Muller-Hill. 1980. Sequence of the lactose permease gene. Nature (London) 283:541– 545.
- 2. Chang, A. C. Y., and S. N. Cohen. 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles

derived from the P15A cryptic miniplasmid. J. Bacteriol. 134: 1141-1148.

- Chen, L., D. Rhoad, and P. C. Tai. 1985. Alkaline phosphatase and OmpA protein can be translocated posttranslationally into membrane vesicles of *Escherichia coli*. J. Bacteriol. 161:973– 980.
- 4. Davis, R. W., D. Botstein, and J. Roth. 1980. A manual for genetic engineering. Advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- 5. Eiglmeier, K., W. Boos, and S. T. Cole. 1987. Nucleotide sequence and transcriptional startpoint of the glpT gene of *Escherichia coli*: extensive sequence homology of the glycerol-3-phosphate transport protein with components of the hexose-6-phosphate transport system. Mol. Microbiol. 1:251–258.
- Engelman, D. M., T. A. Steitz, and A. Goldman. 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu. Rev. Biophys. Chem. 15:321–353.
- Friedrich, M. J., and R. J. Kadner. 1987. Nucleotide sequence of the *uhp* region of *Escherichia coli*. J. Bacteriol. 169:3556– 3563.
- Grosjean, H., and W. Fiers. 1982. Preferential codon usage in procaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene 18:199–209.
- Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28:351– 359.
- 10. Hu, M. C.-T., and N. Davidson. 1986. Mapping transcription start point on cloned genomic DNA with T4 DNA polymerase: a precise and convenient technique. Gene 42:21–29.
- 11. Hugenholz, J., J.-S. Hong, and H. R. Kaback. 1981. ATP-driven active transport in right-out membrane vesicles. Proc. Natl. Acad. Sci. USA 78:3446–3449.
- 12. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Nohno, T., T. Saito, and J.-S. Hong. 1986. Cloning and complete nucleotide sequence of the *Escherichia coli* glutamine permease operon (glnHPQ). Mol. Gen. Genet. 205:260–269.
- Saier, M. H., Jr., D. L. Wentzel, B. U. Feucht, and J. J. Judice. 1975. A transport system for phosphenolpyruvate, 2-phosphoglycerate, 3-phosphoglycerate in Salmonella typhimurium. J. Biol. Chem. 250:5089-5096.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.
- Sharp, P. M., and W.-H. Li. 1986. Codon usage in regulatory genes in *Escherichia coli* does not reflect seleltion for 'rare' codons. Nucleic Acids Res. 14:7737-7749.
- Tabor, S., and C. C. Richardson. 1985. A bacteriophage T7 RNA polymerase-promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA 82: 1074–1078.
- Vogel, H. J., and D. M. Bonner. 1965. Acetylornithinase of Escherichia coli: partial purification and some properties. J. Biol. Chem. 218:97-102.
- 19. Weston, L. A., and R. J. Kadner. 1987. Identification of Uhp polypeptides and evidence for their role in exogenous induction of the sugar phosphate transport system of *Escherichia coli*. J. Bacteriol. 169:3546–3555.
- Wilbur, W. J., and D. J. Lipman. 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA 80:726-730.
- 21. Yu, G.-Q., and J.-S. Hong. 1986. Identification and nucleotide sequence of the activator gene of the externally induced phosphoglycerate transport system of *Salmonella typhimurium*. Gene **45**:51–57.