## Revised Nucleotide Sequence of the *gltP* Gene, Which Encodes the Proton-Glutamate-Aspartate Transport Protein of *Escherichia coli* K-12

B. TOLNER, B. POOLMAN, B. WALLACE, † AND W. N. KONINGS\*

Department of Microbiology, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands

Received 16 October 1991/Accepted 22 January 1992

The gene encoding the proton-glutamate carrier (GltP) of *Escherichia coli* K-12 was sequenced, and the primary structure of the protein was analyzed. The nucleotide sequence was found to differ in several aspects from the previously published sequence (B. Wallace, Y. Yang, J. Hong, and D. Lum, J. Bacteriol. 172:3214–3220, 1990). The corrected open reading frame encodes a protein of 437 (instead of 395) amino acids. Hydropathy analysis predicts 12 membrane-spanning  $\alpha$ -helical regions. The complementary strand does contain an open reading frame possibly encoding a highly hydrophilic polypeptide of 272 amino acids.

In *Escherichia coli*, three L-glutamate transport systems have been identified: (i) a binding-protein-dependent, sodium-independent glutamate-aspartate system (inhibited by cysteate); (ii) a binding-protein-independent, sodium-independent glutamate-aspartate system (inhibited by  $\beta$ -hydroxyaspartate and cysteate), designated GltP; and (iii) a bindingprotein-independent, sodium-dependent glutamate-specific system (inhibited by  $\alpha$ -methylglutamate), designated GltS (6, 9, 11). Genes encoding the last two transport proteins have been cloned, and their nucleotide sequences have been reported (2, 3, 8, 15).

So far, two types of L-glutamate transport mechanisms have been identified in thermophilic bacteria. In Bacillus stearothermophilus, L-glutamate (or L-aspartate) transport proceeds via a sodium-proton symport mechanism with a 1:1:1 stoichiometry (4, 7). In Clostridium fervidus, an electrogenic sodium-glutamate symport mechanism with a stoichiometry of 2 has been identified (12). To further characterize the molecular properties of the sodium-proton-L-glutamate symport system of B. stearothermophilus, the gene encoding the transport protein was cloned and sequenced (14). Comparison of the primary structure of the sodium-proton-L-glutamate symport system (GltT) of B. stearothermophilus with the published sequence of the proton-glutamate carrier (GltP) of E. coli K-12 (15) revealed regions of homology, but other regions differed completely. By translating the nucleotide sequence of gltP in different reading frames and by comparing the translated se-quences with that of GltT, it became apparent that the sequence divergence between GltP and GltT was likely to be caused by sequencing errors, i.e., base substitutions, deletions, and insertions, in the L-glutamate transport gene (gltP) of E. coli. To test this explanation, the MluI-NsiI fragment of pBW8, containing the gene encoding the protonglutamate carrier protein of E. coli K-12 as described by Wallace et al. (15), was resequenced. The MluI-NsiI fragment of pBW8 was removed as an EcoRI-PstI fragment and ligated into the corresponding sites of pUC18. The nucleo-

† Present address: School of Microbiology, The University of New South Wales, Kensington, New South Wales, 2033 Australia.

tide sequences of both strands of the *Eco*RI-*Pst*I fragment or subclones of this fragment in pUC18 or M13mp18/19 (*RsaI*, *Sau3A*, *HpaI*, and *Hinc*II fragments) were determined (Fig. 1) by using the dideoxy-chain termination method (10). A T7 sequencing kit (Pharmacia) was used in sequencing either single- or double-stranded DNA, using T7 forward and pUC/M13 reverse synthetic oligonucleotide primers. The programs Micro Genie (release 5.0) and PC-Gene (release 6.26) were used for computer-assisted sequence analysis. The sequence of the *MluI-NsiI* fragment is shown in Fig. 2.

A comparison of the sequence (Fig. 2) with the reported sequence of gltP (15) revealed 10 differences (indicated by roman numerals in Fig. 2): (i) at position 447, a G should be



FIG. 1. Sequencing strategy of the *MluI-NsiI* fragment of pBW8 containing *gltP*. Part of the vector sequence (open box), the cloned fragment (black box), the position and direction of transcription of *gltP* (large arrow above the sequence), and the positions and directions of transcription of putative ORFX and ORFY genes (small arrows above the sequence) are shown. The starts and extent of the data derived from individual sequencing reactions are shown by arrows below the sequence. S. *Sau3A*; R, *RsaI*; H, *HindIII*.

<sup>\*</sup> Corresponding author.

<b>A</b> <sub>1</sub>	10 CAAGGG <u>TTGCGC</u> -35	20 AACATACC	30 GCGCAAAT <u>GATACT</u> -10	40 GATCAT	50 AGCGTTAAA	60 AAAATCTACA	<b>B</b> <sub>1</sub> {	GC GTC	10 SCA TCA	GCC GT	20 T GAT	GTC GGA	30 TGC GGC	ATA AAC	40 GCC T	TA TCC	50 GAC C	TA CAA L Q	AAT C
61	AACCAACGCA	ACACAATTCA	TGCCCTGGCA G	STATGTCACG 1	TCTCGCGTT	TCTGAACGGG	61	AC GCT	K S	GGC TA	т тес	GCT GTC	TGG CGT W R	AGG CCT	GAT A	AG ACG	CGG C. R	AA GCG	TCG C
121	GAACGGCGCT	CCATTG <u>ACGAAC</u> RBS	S TCATTCAT	TG AAA AAT ATA M K N I	AAA TTC AGC	CTG GCC TGG L A W	121	AT CAG	A TTO	ACA CA	т тес	CGG GTG	GAT ATC	CCC GCC	AAT C	TT CAA	TTA T	TG ATC	CGC A
181	CAG ATT CTG	TTT GCT ATG O	STG CTG GCC ATT C	TC CTG GGA AGC	тас стс сас	TẠC CẠT AGC	181	GT TTT . S F	ATC AAA	TTT GC	C CAG	CAC TTO	ACG CTC	ATA AGC	CAG C	GC TTT	CTT A	CG GTC	KAAA T
241	GAC AGC CGC	GAC TGG CTG (	STC GTC AAT TIG C	TC TCT CCG GCG	GGT GAT ATC	TTC ATC CAT	241	TT GTG	TTC CCA	CTT GG	C AAT		CAC CGO	CAG CGC	ATT A		CAC G	TT CAG	CGC A
301	CTG ATT AAA	ATG ATT GTT C	TC CCC ATT CTC A	TC TCC ACG CTG	сте сте сет	ATC GCG GGT	301	GT ACG	CGC CAT	GTC GA	G GAT	ACG GTC	AAC ACC	AGC AAT	AAA C	C CAG	ACC T	TC CAG	CGG A
361	GTT GGT GAT	GCC AAA CAG (	CTC GGG CGT ATT G	GC GCG AAA ACC	ATT ATC TẠC	TTC GAG GTG	361	AT ACC	TẠC GCT	ACC CA	c cet	TGC CAG	CAA CAC	CÁC AÃA	CGA C	AC GCC	AGG C	AC GCC	AGC A
421	AŢC AÇC AÇC	GIC GCC AIC A	АТТ ТТС ССС АТС А	rt cie cce výc tř	GTC TTC CAG (	ссс ест есс	421	AT CCC	TTT CGA	GGT CA	C CAT	CÁG CGI	che che	CAG ANT	GAT T	TC CTG	CCA G	AT GGA	CAGG
481	GGG GTG GAT	ATG TCG CAG 1	TE CC AC CTC C	AT ATC TCG AAA	TẠT CẠG AỆC A	ACT ACG GAA	481	TC AAT	GCC ATA	CAA CT	e cec	GĂT GẦ	GAT AG	GC ANT	ACT T	ŢG AŢA	CAG C	GT CGA	A DOA V
541	GCG GTA CAA	AGC AGT TCC (	CAC GCC ATT ATC G	ec ace at tre	TCG CTG GTG	cçe açe aac	541	TC AAG	GTT AM	AGA GI	A ACC	eet cee	CĂC CĂC	GAA ACT	GGT G	AT CGA	CặC C	çe tçc	TCCG
601	ATT GTG GCG	TCG ATG GCG	AAA GGC GAA ATG C	TE CES ATC ATC	TTT TTC TCG	ЭТС СТС ТТТ L	601	I K TA GGC	V K TTC CAT	K V CTT CI	T CAAT	AAT TO	CGC CĂC	CAC GCT	TTC A	GA GCT	GČC V	IGT GGA	GTA C
661	egt cte ege	CTT TCT TCC C	TE CEC GEG AGE C	AT CET GAA CEG	CTG GTG ACC	ste tic cec		V G	<b>г</b> н	L 1	. N	N 5	к Q		-35	к л		5 6	•
721	TCC ATC TCT	GAA ACC ATG 1	TT ANA GTG ACT C	AC ATG GTG ATG	CGT TAT GCA	ссе сте сет	661	GC <u>CAG</u> R Q -1	N Q N Q		F	Q N	T N	Q D	P		100		CAUAGO
781	GTG TTT GCG	CTG ATT GCG G	TG ACG GTG GCT A	AC TIT GGT TIC	TCG TCT CTG	IGG CCA CTG	721	CGCGCCA	CAA	TTCCCAG	CAC	TACCA	CCCC	AAGAACAG	<u>аа</u> т 5	GGCGAA	ATCO	AC CAG	CAG C
841	GCG AAA CTG	сте сте сте с	TG CAT TIC GCC A	т сте трс трс	GCG CTG GTA	STG CTG GGA	781	AC CAG	TTT CGO	CAG TO	G CCA	CAG AGA	CGA GA	ACC AAA	GTT A	GC CAC	CGT C R	H R	N N N
901	ATT GTG GCG	CGC CTG TGC G	GG TTA AGC GTC T	GG ATC CTG ATT	CGT ATT CTG	AAA GAT GAG K D E	841	AG CGC	AAA CAG	ACC CA	C CGG	TGC AT	ACG CA	CAC CAT	GTG A	GT CAC	T <u>T</u> T #	AA CAT	GGT T
961	CTG ATT CTG	GCG TAC TCC A	ACT GCC AGC TCT G	AA AGC GTG CTG	CCG CGA ATT	TT GAG AAG	901	Q R TC AGA	K H	T E	I R	C I GGT CA	T H LC CAG C	H H G TTC AC	V G ATG	S H CGT CG	F C GGG	K H CAG G(	G A AGA A
	LIL	A Y S	TASS	ESVL	PRI	I E K vi	961	F R		5 A 5 AAA CA	E H G CAC	G I		L F T	M CAG C	R R AT TTC	G GCC T	QG NDCTT	; R CAT C
1021	ATG GAA GCC	TẠC GGA GCA (	CCG GTG TCG ATC A	CC AGT TTC GTG	GTG CCG ACC (	GT THE TET	1021	K P	Q T	K C	H CGG	R E	K D	D R	Q RCC C	H P	A GCC C	F R	H
1081	VII TTT AAC CTT	GAT GGT TCG	ACG CTG TAT CAA A	GT ATT GCC GCT			1021	R R	H N	V I	R	H Q	R Q	N R	A	H N	A (	V G	T
	P N L	D G S	τ· τ· τ· τ· τ	S I A A vi	-10 -10 -10 - 11	A Q L	1081	A L	Y R	FI	S	A L	I F	R D	I	D G	R	Q L	R
1141	TAT GGC ATT	GAC CTG TCC A	TC TGG CAG GAA A	TC ATT CTC CTC	CTG ACG CTG	TG GTG ACC	1141	AT ATC	CAC CCC	GGC AC	C GGG	L E	D V	RQ	AGT G	D P	Q	N D	GGC G
1201	TCG AAA GGG	ATT GCT GCC G	те срт сес ете т	сс ттт стс стс	TTG CTG GCA	ACG CTG GGT	1201	AC GGT D G	GGT GA		C GAA	GTA GA	N G	FR	GCC A	AT ACG N T	P	E L	F
1261	AGC GTA GGT	ATT CCG CTG G	AA GGT CTG GCG T	TT ATT GCT GGT	GTT GAC CGT	TC CTC GAC	1261	GC ATC G I	ACC AAC T N		C GAT	ACC CA	CAC CA	R G	D GAT C	AC AAT H N	CGG C R	AC AAC	; AAT C N
			E G L A	F I A G	V D R	Te che ane	1321	AT TTT H F	AAT CAG	ATG GA	T GAA D E	GAT AT	CACC CG	CGG AGA	GAGC	AA ATT	GAC G	AC CAG	CCAG
1321	ATG GCG CGT	ACT CC CTC A	AC GTG GTG GGT A	AT GCG CTG GCG	ere cre ere A	TT GCC AAG	1381	TC GCG V A	GCT GT	GCT AT	IG GTA	GTG CA	GTA GC	TCC CAG	GAG A	AT GCC	CAG C Q	H H	S AGC A
1381	TGG GAA CAC	AAA TTT GAC C	COT ANG ANN GCG C	TG GCT TAT GAG	CGT GAA GTG (	TG GGC AAA	1441	AA CAG K Q	AAT CTO	G CCA GO	C CAG	GCT GA	TTT TA	ATT TTT	CAT A	TG AAT M N	GAC T D	F L	S AAT G N
1441	TTT GAT AAA F D K	ACT GCG GAT C	Q TAA TTG A	AG ATTGCCGGGGAT	ATCCACCCGGCA	GTGT	1501	GA GCG G A	CCG TT	C CCC G	T CAG	AAA CG	GAG AA	GTG ACA	TAC T	GC CAG	GGC A	TG AAT M N	C TGT G
1501 1561 1621	GAATGCCTGA TGATTTAGCG TGATGCGACG	TGCGACGCTT TGATTTTGTA C	GCCGCGTCTT A GGTCGGATAA G	TCAGGCCTA C GCGTTTATG C	GCCAGACAG CGCATCCGA	CGCAATAGCC CATCAACGCC	1561 1621	TT GCG V A CGCAACO	ТТС СТ L V	T TGT AG	A TTT	111 <u>14</u>	<u>а</u> сест	TATGAT	CAGT	ATCAT	TTGCG	CGG1	ATGTTG

FIG. 2. Nucleotide sequence and flanking regions of the *gltP* gene encoding the proton-glutamate transport system of *E. coli* K-12 (A) and of the reverse strand on which ORFX and ORFY are located (B). The putative promotor regions (-35 and -10), the possible ribosome-binding site (RBS), and the terminator sequences (<-- and -->) are underlined. The translated amino acid sequences of GltP, ORFX, and ORFY are shown below the DNA sequence. The positions of corrected sequence elements are indicated by asterisks and roman numerals above the sequence.

inserted; (ii) at position 462, a C should be deleted; (iii) the CG at positions 999 to 1000 should be GC; (iv) the CA at positions 1009 to 1010 should be AT; (v) at position 1014, a T should be inserted; (vi) the AACC at positions 1076 to 1077 should be AC; (vii) at position 1085, an A should be inserted; (viii) the CG at positions 1182 to 1183 should be GC; (ix) at position 1272, a T should be inserted, and (x) the CG at positions 1335 to 1336 should be GC. The amino acid sequence deduced from the revised nucleotide sequence differs greatly from the published sequence because of the frameshifts (15). On the basis of the ribosome-binding site at positions 137 to 142 and the amino acid sequence homology between GltP and GltT (14), the translation initiation codon was assigned to the ATG at positions 151 to 153. Putative -35 and -10 promotor regions can be found upstream of gltP. The first stop codon (TAA at positions 1462 to 1464) is followed immediately by a putative rho-independent transcription terminator sequence with a  $\Delta G^{\circ}$  of -27.8 kcal (1 cal = 4.184 J/mol according to the calculations of Tinoco et al. (13). The revised primary structure of GltP contains 437

(instead of 395 [15]) amino acids with a molecular mass of 47.2 kDa.

The carrier protein contains 66.3% nonpolar and 32.7% polar residues, an amino acid composition typical for membrane proteins (1). Of the 437 residues, 30 (6.9%) are basic and 28 (6.4%) are acidic, resulting in a basic protein with an excess of two positive charges at neutral pH. Hydropathy analyses by the method of Eisenberg et al. (5) predicts 12 putative membrane-spanning  $\alpha$ -helical regions (Fig. 3).

Interestingly, the complementary strand contains two open reading frames, ORFX and ORFY. The start codon of ORFX is located upstream of the sequenced region, and the stop codon is located at position 708. ORFY is located between positions 768 and 1584 and could encode a highly hydrophilic polypeptide of 272 amino acids. ORFX and ORFY were not homologous to any of the proteins present in the SWISS-PROT protein sequence data bank (version 1.40).



**Residue number** 

FIG. 3. Hydropathy profile of the amino acid sequence of the proton-glutamate of *E. coli* K-12. The hydropathy profile was calculated according to the method of Eisenberg et al. (5), with a window of 21 amino acids. Vertical bars and numbers indicate the positions of the 12 putative membrane-spanning segments.

Nucleotide sequence accession number. The GenBank accession number for the sequence of the *gltP* gene is M84805.

## REFERENCES

- 1. Büchel, D. E., B. Gronenborn, and B. Müller-Hill. 1980. Sequence of the lactose permease gene. Nature (London) 283:541-545.
- Deguchi, Y., I. Yamato, and Y. Anraku. 1989. Molecular cloning of *gltS* and *gltP*, which encode glutamate carriers of *Escherichia coli* B. J. Bacteriol. 171:1314–1319.
- 3. Deguchi, Y., I. Yamato, and Y. Anraku. 1990. Nucleotide sequence of *gltS*, the Na<sup>+</sup>/glutamate symport carrier gene of *Escherichia coli* B. J. Biol. Chem. 265:21704–21708.

- de Vrij, W., R. A. Bulthuis, P. R. van Iwaarden, and W. N. Konings. 1989. Mechanism of L-glutamate transport in membrane vesicles from *Bacillus stearothermophilus*. J. Bacteriol. 171:1118-1125.
- 5. Eisenberg, D., E. Schwarz, M. Komarony, and R. Wall. 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. 179:125–142.
- Halpern, Y. S., H. Barash, S. Dover, and K. Druck. 1973. Sodium and potassium requirements for active transport of glutamate by *Escherichia coli* K-12. J. Bacteriol. 114:53–58.
- Heyne, R. I. R., W. de Vrij, W. Crielaard, and W. N. Konings. 1991. Sodium ion-dependent amino acid transport in membrane vesicles of *Bacillus stearothermophilus*. J. Bacteriol. 173:791– 800.
- Kalman, M., D. R. Gentry, and M. Cashel. 1991. Characterization of the *Escherichia coli* K12 gltS glutamate permease gene. Mol. Gen. Genet. 225:379–386.
- Miner, K. M., and L. Frank. 1974. Sodium-stimulated glutamate transport in osmotically shocked cells and membrane vesicles of *Escherichia coli*. J. Bacteriol. 117:1093–1098.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
- Schellenberg, G. D., and C. E. Furlong. 1977. Resolution of the multiplicity of the glutamate and aspartate transport systems of *Escherichia coli*. J. Biol. Chem. 252:9055–9064.
- 12. Speelmans et al. Unpublished data.
- Tinoco, I., P. N. Borer, B. Dengler, M. K. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla. 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) 246:40–41.
- 14. Tolner, B., B. Poolman, and W. N. Konings. Unpublished data.
- Wallace, B., Y. Yang, J. Hong, and D. Lum. 1990. Cloning and sequencing of a gene encoding a glutamate and aspartate carrier of *Escherichia coli* K-12. J. Bacteriol. 172:3214–3220.