Sulfate and Thiosulfate Transport in *Escherichia coli* K-12: Identification of a Gene Encoding a Novel Protein Involved in Thiosulfate Binding

MONIKA HRYNIEWICZ,¹ AGNIESZKA SIRKO,¹ ANDRZEJ PAŁUCHA,¹ AUGUST BÖCK,² AND DANUTA HULANICKA¹*

Institute of Biochemistry and Biophysics, Polish Academy of Sciences, 02-532 Warsaw, Rakowiecka 36, Poland,¹ and Lehrstuhl für Mikrobiologie der Universität München, Maria-Ward-Strasse 1a, D-8000 Munich 19, Federal Republic of Germany²

Received 16 October 1989/Accepted 22 February 1990

The sequence of 1,973 nucleotides encompassing the region at and directly adjacent to the CysB-dependent promoter controlling expression and synthesis of the sulfate-thiosulfate transport system of *Escherichia coli* has been determined. The transcription start site has been mapped by primer extension. One open reading frame representing the first gene of the presumed sulfate transport operon was identified and designated *cysP*. The deduced amino acid sequence of the CysP polypeptide indicates the presence of a signal peptide. Expression of the *cysP* gene in the T7 promoter-polymerase system revealed the location of the gene product in the periplasm. Construction of a *cysP* insertional mutant and assays of binding and uptake of sulfate and thiosulfate by this mutant allowed the identification of the *cysP* gene product as a thiosulfate-binding protein. The TGA termination codon of *cysP* was found to overlap the putative ATG initiation codon of the next open reading frame, inferred as being essential for the sulfate transport system, and it was designated *cysT*. Preliminary sequence data from the corresponding region of the *Salmonella typhimurium* chromosome showed strictly homologous counterparts of the *E. coli cysP* and *cysT* genes.

Escherichia coli and Salmonella typhimurium possess a sulfate transport system which is controlled in parallel with cysteine-biosynthetic enzymes and which is a part of the cysteine regulon (19); gene expression of this region requires sulfur limitation and the positive regulator CysB, the product of the cysB gene. Mutations impairing sulfate transport are localized in the "cysA" region, which maps at min 49 of S. typhimurium (8, 9, 25) and min 52 of the E. coli chromosome (4, 18). In E. coli a second locus, cysZ, situated about 10 kilobases (kb) from "cysA," has been reported to be essen-tial for sulfate transport (34). Early genetic studies on the "cysA" locus in S. typhimurium showed the presence of three complementation groups (25). Furthermore, these studies indicated that integrity of the whole three-cistron region is required for both sulfate and thiosulfate transport (8). Subsequently, the specific sulfate-binding protein (SBP) was isolated from S. typhimurium (32) and studied in detail (16, 29, 33, 35). This analysis raised the possibility that the sulfate transport system might have similarities with other well-characterized binding-protein-dependent transport systems (1), where a single operon contains genes encoding the binding protein and several membrane-bound components. However, a direct relationship between SBP and the genetics of sulfate transport remains to be demonstrated because (i) none of the three "cysA" cistrons of S. typhimurium has been shown to encode this activity, and (ii) transportdefective mutants with a specific lesion in SBP could not be isolated despite a thorough search (29).

We have cloned the region containing the "cysA" and cysM (the latter specifying the O-acetylserine sulfhydrylase B) loci of E. coli (40) and identified the CysB-dependent promoter and two complementation groups within the "cysA" locus (M. Hryniewicz, A. Sirko, and D. Hulanicka,

Acta Biochim. Polon., in press). The sequence analysis of this region presented in the accompanying paper (41) revealed the presence of four open reading frames (ORFs) coding for three membrane components of the sulfate transport system and of an ORF for cysM. Here, we report the identification of an unknown locus, designated cysP, specifying the first gene of the region encoding the sulfate-thiosulfate transport system. It is expressed from a CysB-dependent promoter and encodes a thiosulfate-binding protein.

(The nucleotide sequence data reported here will appear in the EMBL, GenBank, and DDBJ nucleotide sequence data bases under accession no. M32101.)

MATERIALS AND METHODS

Bacterial strains and phages. All strains used in this study were derivatives of E. coli K-12 and S. typhimurium LT2 and are listed in Table 1. P1 vir phage was used for transductions by the method of Miller (24). Phage M13 derivatives mp10, mp11, mp18, and mp19 (48) were used to generate single-stranded templates of subcloned DNA fragments.

Media. Media used for bacterial cultures were described by Miller (24) and consisted of LB for plasmid transformation, $2 \times YT$ for M13 DNA transformation and phage production, and M9 for pulse-labeling of plasmid-encoded proteins. Sulfate-free minimal medium (SF) (14) was supplemented with 0.5% glucose and 1 mM L-djenkolic acid or 0.1 mM L-cystine as sole sulfur sources. These two compounds cause derepression and repression of the cysteine regulon, respectively (19). Amino acids were added to 0.2 mM, ampicillin to 50 µg/ml, and chloramphenicol to 30 µg/ml final concentrations when required. Solid media contained 1.5% agar or 0.6% agar for overlays.

Plasmids. Plasmid pGP1-2 (carrying the gene for the T7 polymerase cloned into pBR322 under the inducible control

^{*} Corresponding author.

Strain	Genotype or relevant characteristics	Source or reference	
E. coli			
JM101	F' traD36 pro A^+B^+ lac $I^{q}Z\Delta M15/\Delta$ (pro-lac) thi supE	48	
JC7623	recB recC sbcB15	20	
K38	HfrC phoA4 pit-10 tonA22 ompF627 relA1 (λ)	22	
EC1119	$\Delta trp E5$ leu-6 thi hsdR hsdM ⁺ hsdM ⁺ cvsA15	Stock collection	
EC1250	F^{-} araD139 $\Delta lacU169$ rpsL thi fla trp-3	17	
EC2256	EC1250 cysT329::lac $imm^{\lambda p1(209)}$	12	
EC2275	EC1250 $cysB$ trp^+	Stock collection	
EC2297	EC1250 cvsP	This work	
EC2299	EC2297(pMHH3222)	This work	
S. typhimurium TL117	$\Delta(proAB47)$	7	

TABLE 1. Bacterial strains

of the bacteriophage λp_L promoter) and PT7 vector, for placing the target gene under control of the T7 promoter, were described before (42) and were gifts of S. Tabor. Plasmids containing DNA fragments from the 52-min region of the E. coli chromosome are shown in Fig. 1a. Plasmid pAMH2 has been described previously (40); it carries an approximately 7.5-kilobase (kb) chromosomal DNA fragment containing the entire "cysA"-cysM region on a mini-Mu-lac replicon. A 2.3-kb EcoRV fragment spanning the region of the CysB-dependent promoter was subcloned into the HincII site of pUC18 to create plasmid pMHH1822. This plasmid was subsequently used as a source of restriction fragments for DNA sequencing (Fig. 1b). As the handling of some strains with pMHH1822 appeared to be difficult, presumably because of overproduction of the cysP gene product, the 2.1-kb HindIII-BamHI fragment (HindIII site from the polylinker of pUC18) was recloned into pBR322 to create pMHH3222. pASB1 was a derivative of pT7-6 containing the same chromosomal DNA fragment as pMHH3222 (the entire cysP region) inserted into the multiple cloning site of the vector in the same orientation as the T7 promoter (not shown). pASB2 was constructed by insertion of the 1.3-kb HaeII fragment of pACYC184 (6), containing the cat gene, into the AsuII site of pMHH1822. Plasmid pAPH1 (Fig. 1c), containing the region spanning "cysA"-cysM of S. typhimurium, according to a previously published restriction map (15), was obtained by cloning an 8-kb chromosomal ClaI fragment into the ClaI site of pBR322. Positive clones were selected on the basis of recovery of a Cys^+ phenotype in transformants of *E. coli* mutant strain EC1119 (*cysA15*).

Recombinant DNA techniques. Plasmid DNA preparations, restriction enzyme digestions, ligations, gel electrophoresis of DNA fragments, and ³²P-labeling of DNA probes (nick translation) were performed by standard methods (23).

Southern hybridization. Transfer of restriction fragments from agarose gels onto nitrocellulose sheets and hybridization with radiolabeled DNA fragments were done as described before (2). Hybridization was performed at 65°C in the following buffer: 1% crystalline bovine serum albumin, 1 mM EDTA, 0.5 M NaHPO₄ (pH 7.2), and 7% sodium dodecyl sulfate (SDS). Filters were washed twice with low-stringency wash buffer and then five to eight times with high-stringency wash buffer at room temperature.

DNA sequencing. Techniques used for cloning and sequence determination in recombinant M13 phages have been described (38, 43, 48). The sequencing procedures followed the protocols recommended in the M13 Sequencing Kit (Amersham) for Klenow reactions and United States Biochemical Corp. for Sequenase reactions. Most sequencing reactions were done with the 17-mer universal primer; in some instances DNA synthesis was initiated at sites internal to a subcloned fragment by using a synthetic primer.

5'-End mapping of cysP mRNA. Transcript mapping was performed with avian myeloblastosis virus reverse transcriptase (13). Synthetic primers 5'-TACCGTTTGGTGAT TTGGAAGTTGAAAAGG-3' and 5'-TGAGTTCTTTTTCA GTAAGTTAACGGCCAT-3' were supplied by the Centre for Molecular and Macromolecular Studies Łódź, Poland). Total RNA was isolated by the diethylpyrocarbonate method (2) from E. coli EC1250 grown on SF medium supplemented with either djenkolic acid or cystine. The primers were labeled with $[\gamma^{-32}P]dATP$ by using polynucleotide kinase (23). Hybridization of the labeled primer (10⁶ cpm) to RNA and extension of the primer with reverse transcriptase were performed by the method of Ausubel et al. (2). The reaction products were analyzed on denaturing polyacrylamide gels containing 7 M urea and compared with a sequence ladder of M13mp18 template DNA.

Analysis of plasmid-encoded proteins. The principle for exclusive labeling of plasmid-encoded proteins with T7 RNA polymerase-T7 promoter system was that described by Tabor and Richardson (42). E. coli K38, containing both pGP1-2 and the pT7-derived recombinant plasmid pASB1, was grown at 30°C in LB medium with 40 µg of ampicillin and kanamycin per ml. At an OD_{590} of 0.5, 0.2-ml portions of the culture were harvested, washed with 5 ml of M9 medium, and suspended in 1 ml of M9 medium supplemented with thiamine (20 µg/ml) and 18 amino acids (0.01% [wt/vol] each) lacking cysteine and methionine. Cells were grown with shaking at 30°C for 60 min, and then the temperature was shifted to 42°C for 15 min, rifampin was added at a final concentration of 200 µg/ml, and incubation was continued for 10 min. The temperature was lowered to 30°C for 20 min, and samples were pulsed with 10 μ Ci of [³⁵S]methionine for 5 min. Subsequently, cells were centrifuged, suspended in 120 µl of disaggregation buffer (60 mM Tris hydrochloride [Tris-HCl, pH 6.8], 1% SDS, 1% 2-mercaptoethanol, 10% glycerol, 0.01% bromophenol blue), heated to 95°C for 3 min, and applied to 12.5% SDS-polyacrylamide gels. Polyacrylamide-SDS gel electrophoresis was performed (21), and the radiolabeled proteins were visualized by autoradiography.

Isolation of periplasmic proteins. Periplasmic proteins were isolated from cultures grown on SF medium with djenkolic acid as a sulfur source and were released by osmotic shock (46). Protein concentration was estimated by the method of Whitaker and Granum (45).

Assays of sulfate and thiosulfate uptake and binding. The



FIG. 1. Restriction maps of the "cysA" region of the (a) *E. coli* (40) and (c) *S. typhimurium* (15) chromosomes. Plasmids used in this study containing cloned fragments of chromosomal DNA are shown below the restriction maps. pMHH1822 is derivative of pUC18, pASB1 is a derivative of pT7-6, and pMHH3222 and pAPH1 are derivatives of pBR322. They are aligned with respect to the corresponding restriction sites on the chromosomal maps. Plasmid pASB2 contains the 1.3-kb insert of pACYC184 spanning the *cat* gene cloned in the *Asu*II site of pAHH1822. Positions of previously localized mutations *cysT329::lac* and *cysA15* are given above the *E. coli* map. The position of the CysB-dependent promoter is shown by P, and the direction of transcription is shown by a horizontal arrow. The dashed line shows the 0.72-kb fragment used as a probe for the hybridization experiments. The sequencing strategy of the *E. coli* (b) *S. typhimurium* (c) fragments is shown; segments marked by 0-1, 0-2, 0-3, and 0-4 were sequenced by using synthetic oligonucleotides as primers specific for that particular sequence. Abbreviations: *A. Asu*II; B, *Bam*HI; C, *Cla*I; E1, *Eco*RI; E5, *Eco*RV; H, *Hind*III; Hc, *Hinc*II; P, *Pst*I; S, *Sma*I; Sa, *Sal*I; the *Eco*RI site homologous in *E. coli* and *S. typhimurium* is circled.

indirect assay of sulfate and thiosulfate uptake by whole cells and the assay of binding activity of osmotic shock fluids were performed by the method of Pardee et al. (33) with minor modifications. For the uptake assay, *E. coli* cells were grown in SF medium with djenkolic acid. At an OD₆₅₀ of 0.4 to 0.6, cells were harvested and resuspended in SF medium with glucose (2 mg/ml) and chloramphenicol (30 μ g/ml). Cells (1.7 to 3.4 mg of protein per ml) were incubated in a final volume of 0.2 ml with 0.01 mM sodium [³⁵S]sulfate (0.14 μ Ci) or 0.01 mM sodium [³⁵S]thiosulfate (0.014 μ Ci) for 3 min at 37°C. The incubation was terminated by a 30-s centrifugation in an Eppendorf centrifuge; samples of the supernatant were transferred to GF/C fiberglass filters,

placed in scintillation vials, dried, and counted in a scintillation counter. The protein content of the cell suspension was estimated by assuming that 10⁹ cells contained 150 µg of protein (24). Binding activity of cell-free shock fluids was measured by incubation of 0.2-ml samples of shock fluid (concentrated about 50 times) with 25 µl of 4 µM sodium [³⁵S]sulfate (0.067 µCi) or sodium [³⁵S]thiosulfate (0.007 µCi) for 3 min at 37°C. Incubation was terminated by addition of 0.5 ml of an aqueous suspension of AG1-X8 resin (0.3 g/ml). The resin was allowed to settle for 2 min, and samples from the supernatant were analyzed for radioactivity bound by proteins present in the shock fluid.

Chemicals. Antibiotics were from Sigma Chemical Co.

L-[³⁵S]methionine (>800 Ci/mmol), sodium [outer³⁵S]thiosulfate (10.4 mCi/mmol), and $[\alpha^{-35}S]dATP$ (400 Ci/mmol) were purchased from Amersham Corp. Carrier-free sodium [³⁵S]sulfate was from the Centre for Nuclear Studies (Swierk, Poland), and $[\gamma^{-32}P]dATP$ was from the Institute of Biochemistry and Biophysics (Warsaw, Poland). Anionexchange resin AG1-X8, C1⁻ form, 100–200 mesh, was from Bio-Rad Laboratories. Restriction endonucleases and other enzymes for DNA manipulations were from Bethesda Research Laboratories and U.S. Biochemical or Amersham (England). Other chemicals were of analytical-grade purity.

RESULTS

DNA sequence of the 52-min region of *E. coli* chromosome spanning the CysB-dependent promoter. The restriction map of the chromosomal region containing genes essential for sulfate transport along with the previously localized CysBdependent promoter is shown in Fig. 1a. This DNA fragment contains the genes necessary for complementation of the *cys-329::lac* and *cysA15* mutations (40; Hryniewicz et al., in press). The nucleotide sequence was determined by the chain termination method (38) for both strands of the 2-kb (*Hinc*II-*Eco*RV) DNA insert present in plasmid pMH1822 (Fig. 1a). The sequencing strategy used is outlined in Fig. 1b and the DNA sequence is given in Fig. 2.

Deduced protein products. Two large ORFs were identified on the same DNA strand of the sequenced region (Fig. 2). The first (ORF1), extending from positions 559 to 1572, codes for a 338-amino-acid polypeptide with a deduced M_r of 37,614. The predicted initiation codon for ORF1 was ideally separated from a well-defined ribosome-binding site (39). ORF1 was located directly downstream from the region containing the CysB-dependent promoter, mapped within the 0.56-kb HincII fragment (Fig. 1a) that is between nucleotide positions 1 and 565 (Hryniewicz et al., in press). In agreement with our previous data concerning the organization of the "cysA" locus in E. coli, ORF1 was expected to specify the first gene in the sulfate transport operon and was designated cysP. The NH₂-terminal portion of the CysP polypeptide contained a 25-amino-acid sequence with features typical of signal sequences of E. coli, including a positively charged N-terminus followed by a hydrophobic region and an alanine residue at a potential site of proteolytic cleavage (44). The sequence distal to the ORF1 termination codon contained a region of dyad symmetry, but it did not include the stretch of T's typical of rho-independent termination signals (37). The presumed termination codon of cysP, TGA at position 1573, overlapped the putative ATG codon of the next ORF (ORF2), which extended to the end of the DNA strand sequenced and lacked any in-frame translation termination codons. The restriction mapping of the cys-329::lac mutation (EC2256), isolated by in vivo mutagenesis with Mu d1(Ap lac) and characterized as giving rise to a sulfate permease-deficient phenotype (12), localized the fusion joint to just downstream of the EcoRV site (Fig. 1a), that is, in the region specified by ORF2. As the cys-329::lac mutant was a cysteine auxotroph, ORF2 was presumed to specify the essential component of the sulfate transport system, which has been proposed to be different from that specified by the cysA15 mutation (40). ORF2 was designated cysT (and consequently, the cys-329 mutation was named cysT329), and its location in the sulfate transport operon was confirmed by further sequencing of the whole region (41).

Mapping of the transcription start. Several putative se-

1 GTCAACGCCATTTGCCCGGGATACGTGCCCAACCAACGAGCGGAAAGC 61 TCGAACCCGGAAGATCCAGAGTCGGTGACTGAATAGGCGAAAGC 121 CGCCTCGCCGATCCGCTGGAAGTCGGCGAACTGGCGGCCTTCCTCGC 181 AGCTATTTAACCGGTACCACAGAATGTGATTGATGGCGGCAGCACTT 241 AGCGTCGGTACTACCATGTTCCTCCCTGCATTTGGGGG 301 GAACTAAGTTCACCAGGCTATTTATTGTCATTTGGCCCGGCGAGTGTCGA 361 TCACGTACTATGGTGACGCCGGTTTCTTCACCTCCGGCG 421 AATTACGCTCTTTGACCCAGGTTTTATTGCCCTTTCAACTTCCAACT 423 -35 -10 TATAAAACCGTTACTCCTTTCACGCCGGTTATAAATATGATGGCGTAT	CATTGCCCGCCAG AATCCCGATGCGG ATCGGATGAATCC GCCGGAGACGGT AGGATTTCGTCTT GTGCTCGAAATCC AACTGACTGCAACC TCACCAAACCGGT
61 TEGAACCEGGAAGATCCAGAGTCGGTGCTGACTGAAATGGCGAAACGC 121 CGCCTGCCGATCGGCTGGAAGTGGGAACTGGCGACTGCGGCCTCCTCGC; 181 AGCTATTTAACCGGTACACAGAATGTGATTGATGGGGCAGCACACT 241 AGCGTCGGTATCTGATTCACCTCGTGTTCCCTCGTGACTACTG 301 GAACTAAGTTCACCAGGTATTTTATTGTCATTTTGGCGCGGGGA 361 TCACGTACTATGTGACCCCGGTTTCTCACGGCGTGTCGTGTCCAA 421 AATTACGCCTGTTGACACCAAGTTCTTATTCCACTTCCACTTCCAACT -35 -10 TATAAAACCGTTACTCCTTTCACGCCGGTTATAAATATGGTGGCTAT	AATCCCCGATGCG F ATCGGATGAATCC GCCGGAGACGGTT AGGATTTCGTCTT GTGCTCGAAATCC AACTGACTGCAAC
121 CGCCTCGCCGATCCGCTGGAAGTCGGCGAACTGGCGGCCCTCCTCCGC, 181 AGCTATTTAACCGGTACCAGAATGGATGGAGGCAGCAGCACCC 241 AGCGTCGGTATCTGATGCAGCGTGGCGGCAGCCAGTTGGGGG 301 GAACTAAGTTCACCTCCGGTATTATTGTCACTTTGGCCCGGCGCAGTCGTCGTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	ATCGGATGAATCC GCCGGAGACGGTT AGGATTTCGTCTT GTGCTCGAAATCC AACTGACTGCAAC
181 AGCTATTTAACCGGTACCACAATGTGATTGATGGCGCAGCACATT 241 AGCGTCGGTATCTGATCACCTCGTGTTCCTCCCCGCGCTATTGGGGGG 301 GAACTAAGTTCACCAGGCTATTTATTGTCATTTGGCCCGGGCAC 361 TCACGTACTATGGTACCGCCCGGGTTTCCCCGGGCTGTCGGGCGG 361 TCACGTACTATGTGTACGCCCCGGGTTTCCCCGGGCTGTCGTGTCCAC 421 AATTACGCCTGTTGTACCCCGTTTCAATTCCACTTCCACATCCAAGTTCCACAGCTGTCCAA 361 TATAAAACCGGTACCCAGGTCTCTATGCCGTTATAATATGATGGCTAT	GCCGGAGACGGTT AGGATTTCGTCTT GTGCTCGAAATCC AACTGACTGCAAC
101 AGCGTCGGTATCTGATCACCTCTGTTTCCTCCCTGCATTTGTGGGGG 211 AGCGTCGGTATCTGATCACCTCGTTTCTCTCCCGCCGGGA 301 GAACTAAGTTCACCAGGCTATTTATTGTCATTTGGCCCCGGGAA 361 TCACGTACTATGTGTGGCCCGGTTTTTTTTCTCCGGCCTGTTGGTGCGTCCA 421 AATTACGCCTGTTGAACCAAGTTCTTATTCCCCTTTCAACTTCCAAC -35 -10 481 TATAAAACCGTTACCCTTTTCACGCCGTTATAATGGCCAAT	AGGATTTCGTCTT GTGCTCGAAATCC AACTGACTGCAAC
241 AGCENGUTACTOCACAGGCTATTTATTGCATTTGGCCCGGGCA 301 GAACTAAGTTCACCAGGCTATTTATTGCATTTGGCCCGGGCA 361 TCACGTACTATGGTACGCCCCGGGTTTCCCGGGCTGTCGAC 421 AATTACGCCGTTGCACCAGGTCTTTATCATTCCACTTCCAC -35 -10 181 TATAAAACCGTTACTCCTTTCACGCCGGTTATAAATATGATGGCCTAT	GTGCTCGAAATCC
301 GAACTAAGTCACCAGGCTATTATTGTTGTCCGCGCTGCTCCGGCG 361 TCACGTACTTGGTGACGCTCCGGTTTCTCCGCGCGTGTCGTGCAA 421 AATTACGCCTGTTGAACCAAGTTCTTATTCCCTTTTCAACTTCCAAC -35 -10 -11 TATAAAACCGTTACTCCTTTCACGTCCGTTATAAATATGATGGCTAT	AACTGACTGCAAC
361 ΤΕΛΕGTACTATGTGTACGCTCCGGGTT1CTCCGGGGTGTCCAC 421 ΑΑΤΤΑCGCCTGTGACACCAGGTTCTTATTCCCTTTCACTCCACT -35 -10 181 ΤΑΤΑΔΙΑΛCCGTTACTCCTTTCACGCCGTTATAAATTATGATGGCTAT	AACTGACTGCAAC
421 AATTACGCCTGTTGAACCAAGTTCTTATTCCCTTTTCAACTTCCAA -35 -10 -11 -12 -13 -10 -10 -10 -10 -10 -10 -10 -10	TCACCABACCCTA
- 35 - 10 JR1 TATAAAACCGTTACTCCTTTCACGTCCGTTATAAATATGATGGCTAT	increcentacoo in
181 TATAAAACCGTTACTCCTTTCACGTCCGTTATAAATATGATGGCTAT	↓ -1
	ТАĞАААGTCАТТА
	=
***** ORF1 (cysP) Hc	
541 AATTTATAAGGGTGCGCAATGGCCGTTAACTTACTGAAAAAGAACTCA	ACTCGCGCTGGTC
MAVNLLKKNS	I L A L V
601 GCTTCTCTGCTGCTGGCGGGCCATGTACAGGCAACGGAACTGCTGAA	CAGITCTTATGAC
A S L L L A G H V Q A T E L L N	ISSYD
	ATCCCCANANCAT
BBI GILICUGUGAGLIGITIGUCUCULUAAICOCOCOTTIGAGCAACA	
V S R E L F A A L N P P F E Q Q	J W A K D
721 ANGGGGGGGGACANACTGACGATANAACANTCTCATGCGGGGTCATC	ANNAC AGGCGCTG
ROOPKEIIKQSIIKOSS	
781 GCGATTTTACAGGGCTTAAAAGCCGACGTTGTCACTTATAACCAGGT	IGACCGACGTACAA
	/ T D V O
841 ATCCTGCACGATAAAGGCAAGCTGATCCCGGCCGACTGGCAGTCGCG	GCCTGCCGAATAAT
I L H D K G K L I P A D W Q S R	R L P N N
901 AGCTCGCCGTTCTACTCCACCATGGGCTTCCTGGTGCGTAAGGGTAA	ACCCGAAGAATATC
S S P F Y S T M G F L V R K G N	N P K N I
961 CACGATIGGAACGACCIGGIGCGCICCGACGIGAAGCIGATTITCCC	GAALLEGAAAALG
HDWNDLVRSDVKLIFP	PNPKT
1021 TCCCCTANCCCCCCTATCTCCCCCCCATCCCCCCCCCCC	TAAAGCTGACGGT
5 G N A K I I I L A A W O A A D	
1081 GGTGACAAAGGCAAAACCGAACAGTTTATGACCCAGTTCCTGAAAAA	CGTTGAAGTGTTC
G D K G K T F O F M T O F I K N	VFVF
1141 GATACTGGCGGTCGTGGCGCGACCACCACTTTTGCCGAGCGCGGCCT	
	FGGGCGATGTGCTG
D T G G R G A T T T F A E R G L	TGGGCGATGTGCTG L G D V L
D T G G R G A T T F A E R G L A	IGGGCGATGTGCTG L G D V L
D T G G R G A T T T F A E R G L 1201 ATTACTOGRAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGC	IGGGCGATGTGCTG L G D V L
D T G R G A T T T F A E R G L 1201 ATTAGCTTCGAATGGGAAGTGAACAACAATCCGGTAAACAGTATGGAAGC	IGGGCGATGTGCTG L G D V L CGCAGGGCTTTGAA
D T G G R G A T T T F A E R G L A 1201 ATTACCTITCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGC I S F E S E V N N I R K Q Y E A	ГGGGCGATGTGCTG L G D V L CGCAGGGCTTTGAA A Q G F E
D T G A R G A T T T F A E R G L 1201 ATTAGÉTTEGAATEGGAAGTGAACAACATECGTAAACAGTATGAAGE I S F E S E V N I R K Q Y E A 1000 F E S E V N I I R K Q Y E A	IGGGCGATGTGCTG L G D V L CGCAGGGCTTTGAA A Q G F E
D T G G R G A T T T F A E R G L A 1201 ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGC I S F E S E V N N I R K Q Y E A E1 1261 GTGGTGATTCCGAAAACCAACATTCTGGCGGGAATTCCCGGTGGCGTG	FGGGCGATGTGCTG L G D V L CGCAGGGCTTTGAA A Q G F E GGGTTGATAAAAAC
D T G A R G A T T T F A E R G L 1201 ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGC I S F E S E V N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATCTCGGCGGAATTCCCGGCGGGTG V V I P K T N I L A E F T V A W	IGGGCGATGTGCTG L G D V L CGCAGGGCTTTGAA A Q G F E GGGTTGATAAAAAC W V D K N
D T G G R G A T T T F A E R G L A 1201 ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGC I S F E S E V N N I R K Q Y E A E1 1261 GTGGTGATTCCGAAAACCAACATTCTGGCGGAATTCCCGGTGGCGGG V V I P K T N I L A E F T V A W	IGGGCGATGTGCTG L G D V L CGCAGGGCTTTGAA A Q G F E GGGTTGATAAAAAA W V D K N
D T G A R G A T T T F A E R G L 1201 ATTAGÉTTEGAATEGGAAGTGAACAACATCEGTAAACAGTATGAAGE I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTECGAAAACCAACATTETGGGGGAATTECEGGGGACTGGCGGG V V I P K T N I L A E F T V A W 1321 GTGEAGGECAACGGTACGGAAAAAGECGECAAAGECTATETGAACTEG	IGGGCGATGTGCTG L G D V L CGCAGGGCTTTGAA A Q G F E GGGTTGATAAAAAC W V D K N GGCTCTATAGCCCG
D T G A R G A T T T F A E R G L A ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGC I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATTCTGGCGGAATTCCCGGTGGCGTG V V I P K T N L A E F T V A W 1321 GTGCAGGCCAACGGTACGGAAAAGCCGCAAAGCCTATCTGAACGG V Q A N G T E K A A K A Y L N W	TGGGCGATGTGCTG L G D V L CGCAGGGCTTTGAA A Q G F E GGGTTGATAAAAA W V D K N GGCTCTATAGCCCG W L Y S P
D T G A R G A T T T F A E R G L 1201 ATTAOCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGG I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATCTCGGGGAATTCCCGGTGGGCGGG V V I P K T N I L A E F T V A W 1321 GTGCAGGCCAACGGTACGGAAAAAGCCGCCAAAGCCTATCTGAACTG V Q A N G T E K A A K A Y L N W 1381 CAGCGGCAAACCATATCGGATATATGCTGACTGTIAAACC	IGGGCGATGTGCTG L G D V L CGCAGGGCTTTGAA A Q G F E GGGTTGATAAAAAC W V D K N GGCTCTATAGCCCG W L Y S P TGGAGGTGATGGAG
D T G G R G A T T T F A E R G L A TTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGC I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATTCTGGCGGAATTCCCGGTGGGCTG V V I P K T N I L A E F T V A W 1321 GTGCAGGCCAACGGTACGGAAAAGCCGCCAAAGCCTATCTGAACTG V Q A N G T E K A A K A Y L N W 1381 CAGGGGCAAACCATCATCACGGACTATTACTACCGGGTGAATAACCC	IGGGCGATGTGCTG L G D V L CGCAGGGCTTTGAA A Q G F E SGGTTGATAAAAAA W V D K N SGCTCTATAGCCCC W L Y S P CGGAGGTGATGGAG
$ \begin{array}{ccccccc} D & T & G & A & R & G & A & T & T & T & F & A & E & R & G & L \\ 1201 & ATTAGÉTTEGAATCGGAAGTGAACAACAACCATCCGTAAACAGTATGAAGC \\ I & S & F & E & S & V & N & I & R & K & Q & Y & E & A \\ 1261 & GTGGTGATTCCGAAAACCAACATTCTGGGCGGAATTCCCGGTGGCGGG \\ V & V & I & P & K & T & N & I & L & A & E & F & T & V & A & W \\ 1321 & GTGCAGGCCAACGGTACGGAAAAAGCCGCCAAAGCCTATTCTGAACTG \\ V & Q & A & N & G & T & E & K & A & K & A & Y & L & N & W \\ 1381 & CAGGCGCAAACCATCATCATCAGCGTAATTACCACCGGTGAATAACCC \\ Q & A & Q & T & I & I & D & Y & Y & R & V & N & P \end{array} $	TOGGCGATGTGCTU L G D V L CGCAGGGCTTTGAA A Q G F E SGGTTGATAAAAAC W V D K N SGCTCTATAGCCCC W L Y S P CGGAGGTGATGGAC P E V M D
D T G A R G A T T T F A E R G L ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGC I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATTCTGGCGGAATTCCCGGTGGGTG	TOGGCGATGTGCTG L G D V L COCAGGGCTTTGAA A Q G F E SGGTTGATAAAAAA W V D K N SGCTCTATAGCCCG W L Y S P CGGAGGGTGATGGGA CG A CGTTGGCCC CACAATTGGCTCC
D T G A R G A T T T F A E R G L 1201 ATTAOCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGG I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATCTCGGGGAATTCCGGGGGGGG	TOGGCGATGTGCTG L G D V L CGCAGGGCTTTGAA A Q G F E SGGTTGATAAAAAC W V D K N SGCTCTATAGCCCC W L Y S P CGGAGGTGATGGAC P E V M D ACAAATTTGGCTCC D K F G S
D T G A R G A T T T F A E R G L 1201 ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGC I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATTCTGGCGGAATTCCCGGTGGGTG	TOGGCGATGTGCTG L G D V L COCAGGGCTTTGAA A Q G F E GGGTGATAAAAAA W V D K N GGCTCTATAGCCCG W L Y S P Cogaggtgatggatg P E V M D AccastTtggctcc D K F G S
D T G A R G A T T T F A E R G L 1201 ATTAOCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGG I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATCTCGGGGAATTCCGGCGGGGTG V V I P K T N I L A E F T V A W 1321 GTGCAGGCCAACGGTACGGAAAAAGCCGCCAAAGCCTATCTGAACTG V Q A N G T E K A A K A Y L N M 1381 CAGGCGCAAACCATCTACCGACGTATTACTACCGCGGAATAACCC Q A Q T I I T D Y Y R V N P 1441 AAACTGAAAGCCAAATTCCGCGAGCGGCGGCGGAGTAGAA K L K D K F P Q T E L F R V E D 1501 TGGCGGGAAGTCATAAAACCCACTTACCACCGGCGGCGGAGTAGAA	TOGGCGATGTGCTG L G D V L CGCAGGGCTTGATAAAAC A Q G F E SGGTTGATAAAAC W V D K N SGCTCTATAGCCCC W L Y S P CGCAGGTGATGGAC P E V M D ACAAATTGGCTCC D K F G S ACAAGCGTTAGGCA
D T G CAGGCAACGAACGAACAACAACCCGTAAACAGTATGAAGG 1201 ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGG I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATTCTGGCGGAATTCCCGGTGGGTG	TOGGCGATGTGCTG L G D V L COCAGGGCTTTGAA A Q G F E SGGTTGATAAAAAA W V D K N SGCTCTATAGCCCG W L Y S P CGGAGGTGATGGAC P E V M D ACCAAATTTGGCTCC D K F G S ACCAGGCTGTTAGCG N L L A
D T G A R G A T T T F A E R G L 1201 ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGG I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATCTCGGGGAATTCCGGCGGGCG	TOGGEGATGTGCTG L G D V L CGCAGGGCTTTGAA A Q G F E GGGTTGATAAAAAC W V D K N GGCTCTATAGCCCC W L Y S P CGGAGGTGATGGAC P E V M D ACAAATTTGGCTCC D K F G S ACAAGCTGTTAGGC D K L L A
D T G C GGGGGGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGG 1201 ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGG 1 S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATTCTGGCGGAATTCCCGGTGGGTG	TOGGCGATGTGCTG L G D V L CGCAGGGCTTTGAAA A Q G F E SGGTTGATAAAAAC W V D K N SGCTCTATAGCCCG W L Y S P CGGAGGGGATGGACG P E V N D ACAAATTTGGCTCC D K F G S ACAAGCGTTAGCGC D K L A
D T G G R G A T T T F A E R G L 1201 ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGG I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAATCTGGGGGAATTCCGGCGGGATCGGGTGGG V V I P K T N I L A E F T V A W 1321 GTGCAGGCCAACGGTACGGAAAAAGCCGCCAAAGCCTATCTGAACTG V Q A N G T E K A A K A Y L N M 1381 CAGGGCAAACAATCACCGACTATTACTACCGGGGAAAAACCC Q A Q T I I T D Y Y Y R V N P 1441 AAACTGAAAGCAAAATTCCGCGAGACGCGGCGGGGGGGGG	TOGGCGATGTGCTG L G D V L CGCAGGGCTTTGAAA A Q G F E CGGTTGATAAAAAC W V D K N CGCTCTATAGCCCC W L Y S P CGGAGGTGATGGAC P E V M D ACAAATTTGGCTCC D K F G S ACAAGCTGTTAGCC D K L L A CGGCTTTACCTTAA
D T G CAGGCCAACCAACCACCACCACCCGAACCCAACCC	TOGGCGATGTGCTG L G D V L CGCAGGGCTTTGAA A Q G F E GGGTGATAAAAAA W D K N GGCTCTATAGCCCG W L Y S P CGGAGGTGATAGGAC P E V M D ACAAATTTGGCTCC D K F G S ACAAGCTGTTAGCG D K L L A
D T G G R G A T T T F A E R G L 1201 ATTAOCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGC I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAATCTGGGGGAATTCCGGCGGGATGCCGGTG V V I P K T N I L A E F T V A W 1321 GTGCAGGCCAACGGTACGGAAAAAGCCGCCAAAGCCTATCTGAACTG V Q A N G T E K A A K A Y L N W 1381 CAGGCGCAAACATCTACCGACTATTACTACCGCGTGAATAACCC Q A Q T I I T D Y Y R V N N P 1441 AAACTGAAAGAAATTCCGCGAGACGAGCTGTTCCGGCGGGGAGGAATTAACCCAGGCTGTCCGGCGGCGGGGGGGG	TOGGECGATGTGCTG L G D V L CGCAGGGCTTTGAAA A Q G F E CGGTTGATAAAAAC W V D K N SGCTCTATAGCCCC W L Y S P CGGAGGTGATGGAC P E V M D ACAAATTTGGCTCC D K F G S ACAAGCTGTTAGCG D K L L A G F T L
D T G CAGGCCACAGCCACCTCCCCCAACCCGCGCGCGCGCGCGC	TOGGCGATGTGCTG L G D V L CGCAGGGCTTTGAA A Q G F E SGGTTGATAAAAAC W V D K N SGCTCTATAGCCCC W L Y S P CGGAGGTGATGGAC P E V M D ACAAATTTGGCTCC D K F G S ACAAGCGTTAGCG D K L L A CGGCTTTACCTTAA G F T L SCTCTCCCCGCGCTGG
D T G G R G A T T T F A E R G L A 1201 ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGC I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATTCTGGCGGAATTCCCGGTGGCGTG V V I P K T N I L A E F T V A W 1321 GTGCAGGCCAACGGTACGGAAAAGCCGCCAAGCCTATCTGAACTG V Q A N G T E K A A K A Y L N W 1381 CAGGGGCAAACATCATCACGACGATTATCAACGGGGAGAATAACCC Q A Q T I I T D Y Y R V N N P 1441 AAACTGAAAGCAAATTCCCGCGAACGACGAGCTGTCCGCGTGGAGGA K L K D K F P Q T E L F R V E D 1501 TGGCCGGAAGTATTCCAAACCCACTTCACCAGCGGGCGGG	TOGGEGATGTGCTG L G D V L CGCAGGGCTTTGAAA A Q G F E CGGTTGATAAAAAC W V D K N SGCTCTATAGCCCC W L Y S P CGGAGGTGATGGAC P E V M D K F G S NCCAAGCTGTTAGCC D K F G S NCCAAGCTGTTAGCC D K L L A G F T L SGCCTTTACCTTAA G F T L SCTCTCCCCCGCGGGG L S A L
D T G G R G A T T T T F A E R G L 1201 ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGC I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATTCTGGCGGAATTCCCGGTGGCGGG V V I P K T N I L A E F T V A M 1321 GTGCAGGCCAACGGTACGGAAAAAGCCGCCAAAGCCTATCTGAACTG V Q A N G T E K A A K A Y L N M 1381 CAGGCGCAACGATCATCACCGCAGATTACTACTGCGGTGAATAACCC Q A Q T I I T D Y Y R V N N P 1441 AAACTGAAAGACAAATTCCGCGAAGCCGATCTGCGGGGGAGAA K L K D K F P Q T E L F R V E D 1501 TGGCCGGAAGTAAAAACCCATCTCACCAGGCGGGGGGGGG	TOGGCGATGTGCTG L G D V L CGCAGGGCTTTGAA A Q G F E SGGTTGATAAAAAC W V D K N SGCTCTATAGGCCG W L Y S P E V M D ACAAATTTGGCTCC D K F G S ACAAGCTGTTAGGC D K L L A CGGCTTTACCTTAA G F T L SGCTCTCCCGCGCTGG L S A L
D T G G R G A T T T F A E R G L A TTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGG I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATTCTGGCGGAATTCCCGGTGGGTG	TOGGEGATGTGCTG L G D V L CGCAGGGCTTTGAAA A Q G F E CGGTTGATAAAAAC W V D K N SGCTCTATAGCCCC W L Y S P CGGAGGTGATGGAC P E V M D ACAAATTGGCTGTCG D K F G S ACAAGCTGTTACCG D K L L A G F T L SGCCTTTACCTTAA G F T L SGCCCCCCCCGCGGG L S A L CACCAACCCGCAGG
D T G G A CARACTACCACCACCACCGACGCGCGCGCGCGCGCGCGCGCCCCGCCG	The second secon
D T G G R G A T T T F A E R G L ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGG I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATTCTGGCGGAATTCCCGGTGGCTGG V V I P K T N I L A E F T V A W 1321 GTGCAGGCCAACGGTACGGAAAAGCCGACCAAGCCTATCTGAACTG V Q A N G T E K A A K A Y L N W 1381 CAGGGGCAACATCATCACCGCAGCTATTACTACCGCGTGAATAACCC Q A Q T I I T D Y Y R V N N P 1441 AAACTGAAAGACAATTCCCGCAGACGACGACGTACTCGGAAGA K L K D K F P Q T E L F R V E D 1501 TGGCGGGAGTGATGAAAACCCACTTCCCCAGCGCGGCGGCGGTGAGTTAGA W P E V M K P H F T S G G E L D 0RF2 (cysT) 1561 GCGGGGCCACGACGTGCTGCTGCTGCCGCG A G R N M F A V S S R R V L P 1621 GCCTGGCACCAGTCTGCTGCTGCTGCCGCGG S L G T S L L F V C L I L L L P 1681 TGATGCAACTGGCCCAGTGGCGGCGCGGTACTGGGGGGTGATC	The second secon
D T G G G G G G G C A C T T T T F A E R G L 1201 ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGG I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATTCTGGCGGAATTCCCGGTGGCGTG V V I P K T N I L A E F T V A W 1321 GTGCAGGCCAACGGTACGGAAAAAGCCGCCAAAGCCTATCTGAACTG V Q A N G T E K A A K A Y L N W 1381 CAGGCGCAACCATCATCACCGGCATATTACTACCGCGTGAATAACCC Q A Q T I I T D Y Y R V N N P 1441 AAACTGAAAGACAAATTCCCGCGAGACCGATCTGGCGGAAGA K L K D K F P Q T E L F R V E D 1501 TGGCGGGAGTGATCGAAAACCCATCTCCACCAGGCGGGGGGGG	TOGGCGATGTGCTG L G D V L CGCAGGGCTTTGANA A Q G F E SGGTTGATANANAC W V D K N SGCTCTATAGCCCC W L Y S P CGGAGGTGATGGAC P E V M D ACAAATTTGGCTCC D K F G S ACAAGCTGTTAGCC D K F G S ACAAGCTGTTACCTTAA G F T L SGCTCTCCCCGCGCGGC L S A L CACCAACCCGCAGG T N P Q SATTTTTAACGCCC
D T G G R G A T T T F A E R G L 1201 ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGG I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCACATTCTGGCGGAATTCCCGGTGGGTG	TOGGEGATGTGCTG G G D V L CGCAGGGCTTTGAA A Q G F E CGGTTGATAAAAAC W V D K N SGCTCTATAGCCCG W L Y S P CGGAGGTGATGGAC P E V M D ACAAATTTGGCTGTG D K F G S ACAAGCTGTTACGCC D K F G S ACAAGCTGTTACGTAA G F T L SGCCTTTACCTTAA G F T L SGCCTTCCCGCGCTGGG L S A L CACCAACCCGCAGG T N P Q CGATTTTAACGGCG I F N G
D T G A R G A T T T F A E R G L 1201 ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGC I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATTCTGGCGGAATTCCCGGTGGCGGCG V V I P K T N I L A E F T V A W 1321 GTGCAGGCCAACGGTACGGAAAAAGCCGCCAAAGCCTATCTGAACTG V Q A N G T E K A A K A Y L N W 1381 CAGGCGCAAACCATCATCACCGGCTATTACTACCGCGTGAATAACCC Q A Q T I I T D Y Y R V N N P 1441 AAACTGAAAGACAAATTCCCGCGAGACCGAGCTGTTCCGGCGGGAAGA K L K D K F P Q T E L F R V E D 0 RF2 (cysT) 1561 TGGCGGGAACGAGCTGCTGCTCCTCCACGGCGGGTATGCCG A G R N M F A V S S R R V L P 1681 TGATGCAACGGCCCAGTCTGCCGCAGTCGCGGGGGGGTATGCCG S L G T S L L F V C L I L L L P 1681 TGGTCGCGGCCTACAAAGTAGCCCCACTGCCGGCGGAGTTGTCGCGG S L G T S L L F V C L I L L P 16741 TGGTCGCGGCCTACAAAGTAGCCACGTGCTGCCGGGAGTTGCGGGGGCGTACCGGCCTACAAGGCCGAGTCGCGGGGGGTATCG V N Q L A Q M S W A Q Y W E V I 1741 TGGTCGCGGCCTACAAAGTAACGCTGCTGCCGGCGGTTTGGGCGATCGG V V A A Y K V T L L S A F V A S B	TOGGCGATGTGCTG L G D V L CGCAGGGCTTGATAAAAAC A Q G F E SGGTTGATAAAAC W V D K N GGCTCTATAGCCCC W L Y S P CGGAGGTGATGGAC P E V M D ACAAATTGGCTCC D K F G S ACAAGCTGTTAGCC D K L L A SGGCTTTACCTTAA G F T L SGCCTCCCGCGCGGG L S A L CACCAACCGCAGG C N P Q GATTTTTAACGGCG I F N G
D T G C A R G A T T T F A E R G L 1201 ATTAGÉTTEGAATEGGAAGTGAACAACATECEGTAAACAGTATGAAGE 1 S F E S E V N N I R K Q Y E A 1261 GTGGTGATTECGAAAACCACATETGGCGGAATTECEGGTGGGTG V V I P K T N I L A E F T V A W 1321 GTGCAGGECAACGTAEGGAAAAGECGECAAAGECTATEGAACTG V Q A N G T E K A A K A Y L N W 1381 CAGGEGCAACATEATECACGGACTATTAETACEGEGTGAATAACEC Q A Q T I I T D Y Y R V N N P 1441 AAACTGAAAGACAAATTECEGCAGACGAGETGTTECEGGTGGAATAACEC 0 A Q T I I T D Y Y R V N N P 1441 AAACTGAAAGACAAATTECEGCAGACCGACGTGEGGTGGAAGA K L K D K F P Q T E L F R V E D 1501 TGGECGGGAGTGAAAACCCACTTECACAGGEGGEGGGGGGGAGTAAG N P E V M K P H F T S G E L D 0 R 2 (cysT) 1561 GEGGGGCAACCAGTEGTGTGTGTGCCTGCCTGCCTGCCG A G R N M F A V S S R R V L P 1621 GECCTGGGCACCAGTEGTGTGTGTGCCTGATTTTGCTGCTGCCGCG S L G T S L L F V C L I L L V P 1681 TGATGCAACTGGCCCCAGATGAGACGGGCGGCGGGGAGTACTGGC V W Q L A Q M S W A Q Y W E V I 1741 TGGTCGGCGCTAACTGAAAGTAACCCCGCTGTCGCGGCGTTGTGGGCATGGGACTGAACTGGCCTGACTGGCGCGCGC	Togge Coart of Getting L G D V L Cock of Goge CTTTGATA A Q G P L Cock of Goge CTTGATA A Q G P E Cock of Goge CTTGATA A Q G P E Cock of CTTA A Q G P E Soc CTTA A V D K N Soc CTTA A V D K N Soc CTTA A V D K N D Soc CTA A V D K N D CAAAGCTGTTA A C G T L D A C G F T L S C Soc CTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
D T G A R G A T T T F A E R G L 1201 ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGC I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATTCTGGCGGAATTCCCGGTGGCGGCG V V I P K T N I L A E F T V A W 1321 GTGCAGGCCAACGGTACGGAAAAAGCCGCCAAAGCCTATCTGAACTG V Q A N G T E K A A K A Y L N W 1381 CAGGCGCAAACCATCATCACCGGCTATTACTACCGCGTGAATAACCC Q A Q T I I T D Y Y R V N N P 1441 AAACTGAAAGACAAATTCCCGCGAGACCGATCCGGCGGGAAGAA K L K D K F P Q T E L F R V E D 1501 TGGCGGAAGTAAAACCCACTCACCACGGCGGGGGGGGAGAA W P E V M K P H F T S G G E L D 0 RF2 (cysT) 1621 GCCTCGGCAACAGTCTGCTCTCTCCCCAGACGCGGGTATGCCG A G R N M F A V S S R R V L P 1631 TGGCGGGACCAGTCTGCCGCAGACCGGGCGGATCGCGG S L G T S L L F V C L I L L L P 1641 TGATGCAACGGCCCAGTCTGAGCTGGCCGGAGTATGCCG Y M Q L A Q M S W A Q Y W E V I 1741 TGGTCGCGGCCTACAAAGTACGCCTCTCCCGCGGCGTTTGTGGCATCG V V A A Y K V T L L S A F V A S 1801 TTTTCGGTCTGCTGATGGCGGGAGTCTCACCGCCTACTGCCGCTCCCCG Y F G L L M A M I T A V S B 1801 TTTTCGGTCTGCTGATGGCGGGAGTCTCACCGCCTACCGGCCTGCCCGCCTACCGGCCTACCGGCCTACCGGCCTACCGGCCTACCGGCCTACCGCCTCCCCGCCTACCGCCTCCCCGCCTCCCCGCCTCCCCGCCCCCCCC	TOGGEGATOTOCTO L G D V L COCAGGGCTTTGATA A Q G F E SGGTTGATATAAAAC W V D K N GGTCTATAGCCCC W L Y S P CGGAGGTGATGGAC P E V M D ACAAATTTGGCTCC D K F G S ACAAGCTGTTAGCC D K F G S ACAAGCTGTTAGCC D K L L A G F T L SGCCTCTCCGCGCTGGC L S A L CACCAACCCGCAGG T N P Q SATTTTTAACGCCG I F N G AGGCCGCCCGCGCGCGCGCG G G T L
D T G C G C C G C C A C T T T T F A E R G L 1201 ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGC I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATTCTGGCGGAATTCCCGGTGGGTG	TOGGCCCATCTGCTC L G D V L CGCAGGGCTTTGAN A Q G F E SGGTTGATAAAAAC N V D K N SGCTCTATAGCCCC N L Y S P CGGAGGTGATGGAC P E V N D ACAAATTTGGCTCC D K F G S ACAAGCTGTTAGCC D K F G S ACAAGCTGTTAGCTC D K F G S ACAAGCTGTTACCTTAA G F T L SGCTCTCCGCGCTGGC L S A L CACCAACCCGCAGC T N P Q SATTTTAACGGCCC I F N G AGGCCCCACCGCCGCC G R T L
D T G A R G A T T T F A E R G L 1201 ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGC I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATCTCGGCGGAATTCCCGGTGGCGGCG V V I P K T N I L A E F T V A W 1321 GTGCAGGCCAACGGTACGGAAAAAGCCGCCAAAGCCTATCTGAACTG V Q A N G T E K A A K A Y L N W 1381 CAGGCGCAAACCATCATCACCGGCTATTACTACCGCGTGAATAACCC Q A Q T I I T D Y Y R V N N P 1441 AAACTGAAAGACAAATTCCCGCAGACCGAGCTGTTCCGCGTGGAAGA K L K D K F P Q T E L F R V E D 1501 TGGCCGGAAGTAAAACCCACTACTCACCAGCGGCGGGTATGACGG W P E V M K P H F T S G G E L D 0 RF2 (cysT) 1621 GCCTCGGCACCAGTCTGCTGTTGTGTGCCGCAAGCGGCTGTTCCGCGGGGGTATGCCG A G R N ^ M F A V S S R R V L P 1631 TGGCCGGAACCAGTCTGCCGTGTTGTGTGCCGCAGTATGCCG G G G S L G T S L L F V C L I L L L P 1641 TGATGCACTGGCCCCAAAGTAGCCCGAGTCGGCGGGTATGCCG S L G T S L L F V C L I L L P 1641 TGGTCGCGGCCCAACAGTCTGACGTGGCGGGGTATGCGC V V A A Y K V T L L S A F V A S 1801 TTTTCGGTCTGCTGATTTACCACCGCCTACTGCCGCTCCCCG V V A A Y K V T L L S A F V A S 1801 TTTCGGTCGCGAGTGGCGGGAGTTTGCCGCAACGGCCGATCGCGCTGCCGCGACCAAGTCTGCCGGATCGCCGCATGCCGCCTGCTGCCGCGCTGCCGCGCCTGCCGCGCCTGCTG	The second secon
D T G C G G C A C A T T T T F A E R G L 1201 ATTAGCTTCGAATGGAAGTGAACAACATCCGTAAACAGTATGAAGG I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATTCTGGCGGAATTCCCGGTGGGTG	THE CONTRACT OF CO
D T G A R G A T T T F A E R G L 1201 ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGC I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATTCTGGGGGAATTCCCGGTGGGGGGGG	THE CONTRACT OF CONTRACT OF CONTRACT OF CONTRACTOR OF CONTRACT OF CONTACT OF CONTRACT OF CONTRACT OF CONTRACT OF CONTRACT OF C
D T G G R G A T T T F A E R G L 1201 ATTAGÉTTEGAATEGGAAGTGAACAACATCCGTAAACAGTATGAAGG I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCACATTCTGGCGGAATTCCCGGTGGGTG	THE CONTRACT OF CONTRACT ON CO
D T G G R G A T T T T F A E R G L 1201 ATTAGCTTCGAATCGGAAGTGAACAACATCCGTAAACAGTATGAAGC I S F E S E V N N I R K Q Y E A 1261 GTGGTGATTCCGAAAACCAACATCTCGGCGGAATTCCCGGTGGCGGGGT V V I P K T N I L A E F T V A W 1321 GTGCAGGCCAACGGTACGGAAAAAGCCGCCAAAGCCTATCTGAACTG V Q A N G T E K A A K A Y L N W 1381 CAGGCGCAAACCATCATCACCGCGATTACTACCGCGTGAATAACCC Q A Q T I I T D Y Y R V N N P 1441 AAACTGAAAAGCAAATTCCGCGAGACCGACCGGTGTCCGCGGGAAGA K L K D K F P Q T E L F R V E D 1501 TGGCGGAGAGTATTACCACCACCACCGAGCCGATCCTGCCGGGAGAGA W P E V M K P H F T S G G E L D 0RF2 (cysT) 1561 GCCGGGACCAGTCTGCCGCATTTGCTGCCGGAGCGGGTAGTCGC A G R N M F A V S S R R V L P 1621 GCCTGGGCACCAGTCTGCCGTGTTGTGTGCCGCAGTCTGCCG S L G T S L L F V C L I L L L P 1681 TGATGCAACTGGCCTACAGAGCTGGCTGGCGGGAGTCTGC V M Q L A Q H S M A Q Y W E V I 1741 TGGTCGCGGCCTACAAAGTAACGCTCTTGCGCGCGGCTTGTGGGCATCG V V A A Y K V T L L S A F V A S 1801 TTTCGGTCGCCGAGTGATGCCGTGTTGCGCGCAGTCTGCCG V V A A Y K V T L L S A F V A S 1801 TTTCGGTCGCCGAGTGGAGCGTGTTGCGCGCGAGTGCTGCCG V F G L L M A W I L T R Y R F P 1861 TGATGCCGTGATGGGCTGGTTTGCGCCGCAGCTGCCGCCGCTGCCGCCGATGCCGCCTGATGGCGCTGATGGCCTGCCGCCAACGGCTGCTGCCGCCTGCCGCCTGCCGCCTGCCGCCTGCCGCC	TOGGCGATGTGCTG G G D V L CGCAGGGCTTGATAAAAAC W Q G F E CGGTTGATAAAAAC W D K N CGCTCTATAGCCCC W L Y S P CGGAGGTGATGGAC P E V M D ACAAATTTGGCTCC D K F G S ACAAGCTGTTAGCC D K F G S ACAAGCTGTTAGCC D K L L A G F T L CGCTTTACCCCCCACGC G R T L CGCTTTAACGCCG G L T L (E5) STTTGAT F D

FIG. 2. Nucleotide sequence of the DNA fragment spanning the CysB-dependent promoter and the gene which lies adjacent. The sequence starts at the *HincII* site (base 3) and ends just after the *Eco*RV site (base 1980). Relevant restriction sites for *HincII* (Hc), *SmaI* (S), *AsuII* (A), *Eco*RI (E1), *Bam*HI (B), and *Eco*RV (EV) are shown above the sequence. Double dashed lines below the sequence show the -35 and -10 elements of the promoter; the vertical arrow indicates the position of the transcription start; and **** indicates the Shine-Dalgarno (39) sequence. An inverted repeat in the 3'-flanking region is indicated by arrows. The deduced amino acid sequence is shown below the open reading frames.

quences showing similarity to the canonical -35 and -10 promoter elements (11) were found in the 5'-flanking region of the *cysP* gene. To determine which of the putative promoters is actually functional in vivo, the mRNA start site was determined by primer extension. Two different 5'-³²P-labeled oligodeoxynucleotides were used as primers: the first extended from and was complementary to nucleotide positions 451 to 480, and the second extended from nucleotide positions 559 to 588 (Fig. 2). Total RNA was isolated



FIG. 3. Primer extension analysis of in vivo transcription start for the cysP gene. RNA was isolated from strain EC1250 grown in SF minimal medium with (lane 1) djenkolic acid or (lane 2) cystine as the sulfur source. Identical amounts of the primer and RNA were used for each reaction. Product sizes were estimated by comparison with M13mp18 sequence initiated from a 17-mer sequencing primer (-40). The relevant portion of the *E. coli* sequence is presented, and the nucleotide whose size corresponds to that of the extension product (indicated by the large arrow) is circled. It is not known whether the minor products indicated by the small arrows represent additional transcription start sites or whether they are artifacts of the primer extension reaction.

from the wild-type strain of E. *coli* grown on either nonrepressive (djenkolic acid) or repressive (cystine) sulfur sources.

Extension of the first of the primers (nucleotides 451 to 480) used did not give any specific product (not shown). Extension of the second primer gave a product of 59 nucleotides (Fig. 3) with the RNA isolated from the culture grown on djenkolic acid but not with that from the culture grown on cystine as a sulfur source. This result suggested the G at position 530 as a possible candidate for the transcription start site of the *cysP* gene.

Identification of the cysP gene product in the T7 promoterpolymerase system. The nature of the cysP gene product was investigated in the system developed by Tabor and Richardson (42) for exclusive expression of target genes placed on a plasmid under the control of the T7 promoter. A labeled product with an approximate M_r of 32,000 was observed (Fig. 4) in the extract of whole cells harboring plasmid pASB1 as well as in the fluid released by osmotic shock; there was only a small residual portion of this product in the extract of shocked cells. The periplasmic location of the CysP polypeptide is consistent with the presence of a signal peptide typical of proteins transported through cytoplasmic membranes (44). The molecular weight of the CysP polypeptide estimated by SDS-gel electrophoresis (M_r of 32,000) is lower than that predicted from the amino acid composition of the mature form of CysP (M_r of 35,057). Such anomalous migration in SDS gels has also been observed for the histidine-binding protein (27).

Analysis of the cysP gene from the S. typhimurium chromosome. Comparison of the deduced amino acid sequence of E. A B C D E F - - 32000

FIG. 4. Autoradiograph of [35S]methionine-labeled proteins from E. coli K38 transformed with pASB1 and pGP1-2. Lane A, Induction of the T7 polymerase (42°C) without rifampin; lane B, without induction of the T7 polymerase (30°C), without rifampin; lane C, induction of the T7 polymerase (42°C), with rifampin; lane D, shocked fluid; lane E, crude extract of osmotically shocked cells; lane F, crude extracts of unshocked cells. Experiments whose results are presented in lanes D, E, and F were performed with induction of the T7 polymerase and with rifampin (see also Materials and Methods). After the labeling of proteins, the cells were divided into two portions. Bacteria from one portion were osmotically shocked; the shocked cells and the unshocked cells from the second portion were sonicated. Samples in lanes D, E, and F correspond to the same amounts of bacterial cells. Molecular weight standards were: bovine serum albumin (68,000), ovalbumin (45,000), aldolase (40,000), trypsin inhibitor (21,500) and lysozyme (14,300).

coli CysP protein with that of SBP of S. typhimurium (16, 35) showed 45% similarity (Fig. 5). This value seemed to be rather low for genes from such closely related organisms as E. coli and S. typhimurium (see also Discussion). Thus, in order to check whether the S. typhimurium genome contains a counterpart to the E. coli cysP gene, Southern blots of EcoRV-digested chromosomal DNAs from both E. coli and S. typhimurium were hybridized with a 32 P-labeled probe from the coding region of cysP (0.72-kb HincII-EcoRI fragment; Fig. 1a). The same 2.3-kb fragment (containing the internal EcoRI site in E. coli) was detected in E. coli and S. typhimurium (Fig. 6). The above result suggested that these regions in E. coli and S. typhimurium are homologous. The 8-kb ClaI fragment (Fig. 1c) presumed to contain the entire cysP region of S. typhimurium was chosen for cloning in pBR322 to create plasmid pAPH1 (Fig. 1c). The strategy for sequencing this portion of the chromosomal insert present in this plasmid is outlined in Fig. 1c. The sequence estimated for one DNA strand of this region (Fig. 7) showed 86% similarity with ORF1 (cysP) and also with the sequenced portion of ORF2 (cysT) of E. coli; the similarity on the amino acid level was even higher (data not shown). These results indicated that the cysP gene of E. coli has a homologous counterpart in S. typhimurium and codes for a gene product distinct from the characterized sulfate-binding protein (16, 35)

Construction of a cysP mutant. A means of determining the precise function of the CysP protein was to construct a defined cysP mutant. For this purpose, plasmid pASB2 was constructed by insertion of the cat gene into the coding region of cysP (Fig. 1a). Crossing of this cysP insertional mutation back onto the chromosome was performed by the technique of Winans et al. (47). pASB2 DNA was linearized by restriction cleavage with PstI and introduced to a recBC sbcB strain (JC7623) by transformation. Chloramphenicol-resistant, ampicillin-sensitive colonies were selected, and phage P1 grown on one of the clones obtained was used to transduce the EC1250 strain to chloramphenicol resistance. One of the transductants (EC2297) was verified as a cysP insertional mutant by Southern analysis of EcoRV-cleaved

M A V N L L K K N S L A L V A S L L L A G H V Q A T E K D I Q L L N S S Y D V S R E L F A A L N P P F E Q Q M A K A D N G G D K L T I K Q S H A G S S K Q A L A I L Q G L K E T G D N V V I D D S H G G S G L Q A T S V I N G I E A D V V T L A L A Y D V N A I A G R G R - U D K N H I S R L P N N S S P Y T S T I V E L V R K G N P K N I H K R L P D N S A P Y T S T I V E L V R K G N P K Q I H D M N D L V R S D V K L I F P N P K T S G N A R Y T Y D M N D L V R S D V K L I F P N P L S S G Q A R W N Y L A A W G A A D G G G N K G K T E Q F M T Q F L K N V E V F D T G G R G A T T T F A E R G L G D V L I S N V E V F D T G G R G A T T T F A E R G L G D V L I S N V E V L D S G A R G S T N T F V E R G I G D V L I S F E S E V N N I R K Q Y E A Q G F E V V I F S S S I I N V E V L D S G A R G S T N T F V E R G I G D V L I S M Y Z A A W G Y A L H H N N N D Q A K A E D E V K A L F K N V E V L D S G A R G S T N T F N E R G L G D V L I S A A W G A A D K A D G G N K G K F E I V T P S E S I L A A W G A A L A T N E L G K G K F E I V T P S E S I L A A W G A C A L A T N E L G K G K F E I V M I F S E S I L A E F P V A W V D K N V Q E A N G T E K A A A V A E A Y L N W L Y S P E G Q E I A A K N F Y R P R D A D Y A K K Y D D K F P Q T E L F R V E D K F G S W A K A Q K D H F A S G G E L D K L L A A G R N S G G E L D K L K L E T I D E V F G G W A K A Q K D H F A S G G E L D K L K L E F T I D E V F G G W A K A Q K D H F A

FIG. 5. Comparison of the amino acid sequences of the CysP polypeptide from *E. coli* (upper sequence) and the SBP from *S. typhimurium* (lower sequence). The sequences are aligned for maximum identity. Identical amino acid positions between two sequences are boxed. An arrow at amino acid position 26 of CysP shows the potential site for cleavage of the signal peptide (44). Amino acids essential for sulfate binding by SBP (35) are marked by vertical arrows below the SBP sequence. The SBP sequence is essentially consistent with that published by Isihara and Hogg (16) with the amendments given by Pflugrath and Quiocho (35).

chromosomal DNA from EC2297 and by comparison with the pattern generated by cleavage of chromosomal DNA of the parental strain (EC1250). The *HincII-EcoRI* 0.72-kb fragment (Fig. 1a) from the coding portion of *cysP* was used as a hybridization probe. This probe detected the predicted 2.3-kb band in the case of strain EC1250 and a 3.6-kb band for strain EC2297 (Fig. 6). These data confirm the insertion site of the *cat* gene fragment and are in agreement with the predicted length of the *cysP-cat* gene hybrid.

The insertional mutation in cysP did not result in the cysteine auxotrophy observed for other known sulfate transport mutants (e.g., cysA15 and cysT329::lac). However, the rate of growth of the cysP strain was about twofold lower in liquid medium with either sulfate (0.04 mM) or thiosulfate (0.02 mM) as the sulfur source compared with the wild-type strain (data not shown).

Uptake and binding of sulfate and thiosulfate by the cysP mutant. The reduced growth rate of the cysP mutant on minimal medium with sulfate or thiosulfate as the sole sulfur source suggested that the CysP protein might be involved in binding of either one or both of these anions. To check this possibility, assays of sulfate and thiosulfate binding and uptake were performed in the wild-type strain, sulfate transport-deficient mutants, and a cysP mutant. As shown in Table 2, the binding of thiosulfate by crude osmotic shock fluid of the cysP strain (EC2297) was reduced more than 10-fold compared with the wild-type strain, whereas the

binding of sulfate was practically unchanged. The presence of the $cysP^+$ gene on a multicopy plasmid in the cysP mutant (EC2299) resulted in a significant increase in thiosulfatebinding activity (more than 100 times), but sulfate-binding activity was unaffected. The above findings strongly suggest that the major function of the CysP polypeptide is the binding of thiosulfate.

Indirect assay of uptake of both anions by whole cells confirmed the function of the cysP gene product as a thiosulfate-binding protein. Uptake of thiosulfate by the cysP mutant was decreased about seven times, whereas sulfate uptake was diminished only twofold compared with wild-type activity. Furthermore, the EC2299 strain [cysP $(cysP^+)$] showed the wild-type level of thiosulfate uptake; however, the cysP plasmid had no effect on sulfate uptake, consistent with the findings in the binding experiments. A possible explanation for the decrease in sulfate uptake by the cysP mutant might be an effect of the 1.3-kb fragment insertion on the expression of the genes encoding other components of the transport system common to both anions. The very low levels of binding and uptake of thiosulfate in the cysB mutant (EC2275) indicated that cysP gene expression requires a functional CysB protein.

DISCUSSION

The genetic region of E. coli heretofore designated "cysA" is in fact composed of several genes specifying distinct components of the sulfate-thiosulfate transport system (see accompanying paper [41]). Earlier studies allowed us to approximately localize the CysB-dependent promoter governing the expression of the downstream genes within the restriction map of this region (Hryniewicz et al., in press). In this study we have presented the sequence analysis of the region encompassing this promoter and the gene lying directly adjacent to it and we have identified its gene product.

The DNA sequence of the region directly adjacent to the promoter revealed the presence of a 1,014-bp open reading frame, designated cysP, encoding a protein of 338 amino acid residues (Fig. 2) with a predicted molecular weight of 37,614. As the NH₂-terminal portion of the deduced amino acid sequence showed the features typical of a signal peptide (44), the mature CysP protein may in fact consist of 313 amino





(E.coli) 450 460 470 (E.coli) TCTTATTCCTTTTCAACTTCCAAATCACCAAACGGTA	1020 1030 1040 1050 1060 1070 GTCGGGTAACCGCGGTAAACCTATCTGGCGGCATGGGGCGCAGGGATAAACCTGACGG
(S.typhimurium) TCTTATTCCCTTTTCAACTTTCAAACCATCAAACGGTA 10 20 30	CTCCGGCAACGCCGTTACACGTATCTGGCGGCATGGGGCGCGGGGATAACGCGGACGG 590 600 610 620 630
490 500 510 520 530 TATAAAACCGTTACTCCTTTCACGTCCGTTATAAAATATGATGGCTATT-AGAAAGTCATT 	1080 1090 1100 1110 1120 1130 TGGTGACANAGGCANAACCGAACAGTTTATGACCCAGTTCCTGAAAAACGTTGAAGTGTT CGGCGATAAAGCCAAAAACCGAACAGTTTATGACCCAGTTCCTGAAAAACCGTCGAAGTGTT
50 60 70 80 90 540 550> 570 580 590 ANATTTATAAGGGTGCGCAATGGCCGTTAACTTACTGAANAAGAACTCACTCGCGCTGGT IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	650 660 670 680 590 1140 1150 1160 1170 1180 1190 CGATACTGGCGGCGTGGGGGGGGGGGGGGGGGGGGGGGG
600 610 620 630 640 650 CGCTTCTCTGCTGCGGGGGGCCATGTACAGGCAACGGAACTGCTGAACAGTTCTTATGA IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	1200 1210 1230 1240 1250 1237 1230 1240 1250 1250 1250 <th< td=""></th<>
170 180 190 200 210 660 670 680 690 700 710 CGTCTCCCGCGGCGTGTTTGCCCCCGCATTGGGCAACAATGGGCAAAAGA 	1260 1270 1280 1290 1310 1310 AGTGGTGATTCCGAAAACCAACATTCTGGCGGAATTCCCGGTGGGGTGGGGTGGGT
720 730 740 750 760 770 TAACGGCGGCGACAAACTGACGATAAAACAATCTCATGCCGGGTCATCAAAACAGGCGCT	1320 1330 1340 1350 1360 1370 CGTGCAGGCCAACGGTACGGAAAAAGCCGCCAAAGCCTATCTGAACTGGCTCTATAGCCC
780 790 800 810 820 830 GGCGATTTACAGGGCTTAAAAGCCGACGTTOTCACTATAACCAGGTGACCGACGTACA GGCGATTTTGCAGGGACTGAAAGGCAGCGGACGGACGGAC	1380 1390 1400 1410 1430 GCA8GEGCAAACCATCATCACCACTATTACTACCGGGGAATAACCCGGGAGGGGTGATGAG 111111111111111111111111111111111111
840 850 860 870 880 890 ANTCCTGCACGATAAAGCTAACCTGATCCCGGCCGACTGGCGACTACGCGAATAA IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	1410 1450 1460 1470 1480 CAAACTGAAAGACAAATTCCCGCAGACCGAGCTGTTCCGCGTGGAGAGACAAATTTGGCCC 1
900 910 920 930 940 950 TAGCTCGCCGTTCTACTCCACCATGGGCTTCCTGGTGCGTAAGGGTAACCCGAAGAATAT 1 1111 1111 1111 1111 1111 1111 11	1500 1510 1520 1530 1540 1550 CTGGCCGGAAGTGATGAAAACCCACTCACCAGCGGCGGCGGACTTAGACAACCTGTTAGC
950 970 980 990 1000 1010 CCACGATTGGAACGACCTGGTGCGCTCCGACGTGAAGCTGATTTTCCCGAACCCGAAAAC IIIIIIIIIIIIIIIIIIIIIIII	1560 1570> 1580 1590 1600 1610 GGCGGGGCGTAACTGATGTTGCTGTCTCCTCCCAGACGCGGCGGGCG

FIG. 7. Comparison of the *E. coli cysP* DNA sequence with the homologous region of the *S. typhimurium* chromosome. Identical nucleotides are indicated by a colon between the two sequences. Nucleotide positions of *E. coli* are numbered as in Fig. 2. The initiation codons of cysP and cysT are marked by arrows, and the stop codon of cysP is marked by asterisks above the *E. coli* sequence.

acids with a predicted M_r of 35,057. The periplasmic location of the mature CysP polypeptide, inferred by the presence of a signal peptide, was confirmed experimentally by specific expression of the *cysP* gene in the T7 promoter-polymerase system (Fig. 4). The suggested cleavage site between Ala and Thr (Fig. 5) conforms to the rules for procaryotic signal peptidase sites (44).

In most well-characterized periplasmic binding-proteindependent transport systems, the single operon contains genes for all transport components, and the gene for the binder is usually the first in the cluster (for a review, see

 TABLE 2. Sulfate and thiosulfate uptake and binding activities in strains with different mutations in the "cysA" region^a

Otras in	Relevant	Cys pheno- type	Sulfate		Thiosulfate	
Strain	genotype		Uptake	Binding	Uptake	Binding
EC1250	Wild type	+	1,900	119	1,900	300
EC2275	cvsB	-	80	38	110	46
EC1119	cysA15	_	100	ND^{b}	140	ND
EC2256	cvsT329	_	170	ND	290	ND
EC2297	cysP	+	740	132	260	5
EC2299	$cysP(cysP^+)$	+	650	128	1,500	2,200

^a Uptake and binding activities are expressed as picomoles of [³⁵S]sulfate or [³⁵S]thiosulfate removed from the incubation mixture by 1 mg of whole-cell protein (uptake) or bound by 1 mg of protein of crude osmotic shock fluid at 37°C (binding).

^b ND, Not determined.

reference 1). Thus, our initial assumption was that the cysP gene product was identical to the SBP which was isolated from S. typhimurium and characterized as a sulfate binder and shown to be regulated in parallel with sulfate permease (29). However, the deduced amino acid sequence of mature CysP protein showed only 45% sequence similarity with that of S. typhimurium SBP (16, 35), a value relatively low compared with other pairs of predicted homologous gene products from these organisms, like cysB and cysK, which share, respectively, 95% (30) and 96% (5) sequence similarity. Therefore, the corresponding region of the S. typhimurium chromosome whose restriction map was established previously (15) was cloned and sequenced. The sequence analysis of this region revealed the presence in S. typhimurium of a strictly homologous counterpart of the E. coli cysP gene (over 86% sequence homology on the DNA level). The deduced amino acid sequence of CysP from both E. coli and S. typhimurium is distinct from that of the previously characterized SBP (16, 35).

Assays of sulfate and thiosulfate binding and uptake by E. coli strains with different lesions in the sulfate transport system, including a cysP insertional mutant (Table 2), allowed us to establish that the major activity of CysP protein is the binding of thiosulfate. Therefore, this protein has been designated a thiosulfate-binding protein. It is worth noting that some amino acid residues essential for sulfate binding by SBP of S. typhimurium as estimated by Pflugrath and Quiocho (35) are conserved in the thiosulfate-binding protein

			. 35
S.typh.	cysJIH	AAC AGG'	TTAGTTCATTTGGTTATTTGTTATT ICCAACC-CTICTTTAAT
E.coli	cysJIH	AACAGG	TTAGTCGATTTGGTTATTAGTTATC GCTATCC-CGTCTTTAAT
S.typh.	cysK	ACCATT	ATTTCCCATCAGCATATA-GATATGCGAAATCC-TTACTTCCCC
E.coli	cysK	GTCATT	ATTTCCCTTCTGTATATA-GATATGCTAAATCC-TTAC TTCCGC
S.typh.	cysP	TTCAAC	ТТТ-САААТС-АТСАААС-GGTATA-ТААААССGTTAC TCCTTT
E.coli	cysP	TTCAAC	ТТС-САЛАТС-АССАЛАС-GGTATA-ТАЛААССGTTACTCCTTT
(conserv	ed)	CA	.TTAG.TATA.CCCT
			-10
S.typh.	cysjiH	(cont.)	TGTTATTCCTCTCACCGTTAACCTTATCCTC
S.typh. E.coli	CysJIH CysJIH	(cont.) (cont.)	TGTTATTCCTCTCACCGTTAACCTTATCCTC CCACACCGTTTGCCCCGTTAACCTTACCTT
S.typh. E.coli S.typh.	cysJIH cysJIH cysK	(cont.) (cont.) (cont.)	TGTTATTCCTCTCACCGTTAACCTTATCCTC CCACACCGTTTGCCCCGTTAACCTTACCTT
S.typh. E.coli S.typh. E.coli	сузJIН сузJIН сузК сузК	(cont.) (cont.) (cont.) (cont.)	TGTTATTCCTCTCACCGTTAACCTTATCCTC CCACACCGTTTGCCCCGGT TAACCT TACCTT ATATCTGGCTGGAAGG TATGCT GGGAAG ATATTCTCTGAGCGGG TATGCT ACCTGTTG
S.typh. E.coli S.typh. E.coli S.typh.	cysJIH cysJIH cysK cysK cysP	<pre>(cont.) (cont.) (cont.) (cont.) (cont.)</pre>	ТGTTATTCCTCTCACCGTTAACCTTATCCTC CCACACCGTTTGCCCCGGT TAACCT TACCTT ATATCTGGCTGGAAGG TATGCT GGGAAG ATATTCTCTGAGCGGGG TATGCT ACCTGTTG AGCGCTGGTTATAAA TATGAT GACCAATAA
S.typh. E.coli S.typh. E.coli S.typh. E.coli	CysJIH CysJIH CysK CysK CysP CysP	(cont.) (cont.) (cont.) (cont.) (cont.) (cont.)	ТGTTATTCCTCTCACCGTTAACCTTATCCTC CCACACCGTTTGCCCCGGTTAACCTTACCTT

FIG. 8. Comparison of the cys promoter regions of *E. coli* and *S. typhimurium* (*S. typh*). Sequences are aligned with respect to -35 and -10 regions and have been adjusted by inclusion of gaps. The cysK and cysJIH data are from reference 31.

sequence (Fig. 5). Thus, it is an attractive possibility to consider the sulfate-thiosulfate transport system as a porter interacting with two binding proteins whose functions may partially overlap. Precedents for such a situation exist; e.g., the *liv* system, where two binding proteins, one being specific for leucine and the other for leucine, isoleucine, and valine, interact with the same membrane-bound components (1). Such functional overlap would explain both the failure to isolate transport-defective mutants with specific defects in SBP activity (29) and the cysteine bradytrophy of our *cysP* mutant. Genetic analysis of the sulfate-binding protein will reveal whether this assumption holds true.

The presence in S. typhimurium of a gene homologous to cysP of E. coli, as well as unchanged sulfate-binding activity of the E. coli cysP mutant (Table 2), indicates the presence of a separate gene encoding the SBP. Our attempts to find the open reading frame encoding SBP of E. coli in the region extending about 1,500 bp upstream of cysP (where it would be expected, by analogy to the *liv* system) were, however, unsuccessful.

The primer extension analysis defined the -35 and -10elements of the promoter as TTACTC and TATGAT, respectively (Fig. 2). The deviation of the -35 region from the consensus sequence (10, 11) was also observed for two other CysB-dependent promoters governing the expression of the cysK and cysJIH genes (31) in both E. coli and S. typhimurium. Comparison of the -35 upstream region with that published by Ostrowski and Kredich (31) for the cysK and cysJIH promoters did not reveal any clear homology; however, some nucleotide sequence arrangements seem to be similar (Fig. 8). The second ORF (cysT) found in the region directly downstream of cysP and inferred to be essential for sulfate transport by identification of a cysT329::lac mutation (12; Hryniewicz et al., in press) also has a strictly homologous counterpart in S. typhimurium. Overlapping of the TGA and ATG at the cysP-cysT junction (28) and the apparent requirement of the cysP promoter for maximum expression of the cysT329::lac fusion (Hryniewicz et al., in press) suggests that cysP and cysT form an operon. A sequence of dyad symmetry was found in the 3' nontranslated region of the cysP gene; however, it does not include the usual run of T's typical of rho-independent terminator signals (36, 37). It was reported that alterations reducing the number of terminal uridines in RNA transcript weaken the termination response (36). The ability of dyad symmetry regions to function as a barrier to degradation of polycistronic transcripts was reported for the region of trp t-trp t' in the tryptophan operon (26) and the intercistronic region between papA and papH loci (3) of E. coli; thus, this potential stem-loop structure may have such a function.

The presence of two separate complementation groups specified by two sulfate transport-defective mutants (Hryniewicz et al., in press) suggests that the region downstream of the cysP gene encodes other components of the sulfate transport system. The sequence analysis of this region and the identification of gene products are presented in the accompanying paper (41).

ACKNOWLEDGMENTS

We thank G. Sawers for his help in correcting the manuscript and R. Mertz for kindly supplying oligonucleotides.

Part of this research was supported by the Polish Academy of Sciences within project 3.13.

LITERATURE CITED

- 1. Ames, G. F.-L. 1986. Bacterial periplasmic transport systems: structure, mechanism and evolution. Annu. Rev. Biochem. 55:397-425.
- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl. 1987. Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
- 3. Baga, M., M. Goransson, S. Normark, and B. E. Uhlin. 1988. Processed mRNA with differential stability in the regulation of *E. coli* pilin gene expression. Cell 52:197-206.
- Britton, P., A. Boronat, D. A. Hartley, M. C. Jones-Mortimer, H. L. Kornberg, and F. Parra. 1983. Phosphotransferase-mediated regulation of carbohydrate utilization in *Escherichia coli* K-12: location of the gsr (tgs) and iex (crr) genes by specialized transduction. J. Gen. Microbiol. 129:349–356.
- Byrne, C. R., R. S. Monroe, K. A. Ward, and N. M. Kredich. 1988. DNA sequences of cysK regions of Salmonella typhimurium and Escherichia coli and linkage of the cysK regions to ptsH. J. Bacteriol. 170:3150-3157.
- Chang, A. C. Y., and S. N. Cohen. 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from P15A cryptic miniplasmid. J. Bacteriol. 134: 1141-1156.
- 7. Csonka, L. N. 1981. Proline over-production results in enhanced

osmotolerance in Salmonella typhimurium. Mol. Gen. Genet. 182:82-86.

- Dreyfuss, J. 1964. Characterization of sulfate and thiosulfatetransporting system from *Salmonella typhimurium*. J. Biol. Chem. 239:2292-2297.
- 9. Dreyfuss, J., and K. J. Monty. 1963. The biochemical characterization of cysteine-requiring mutants of *Salmonella typhimurium*. J. Biol. Chem. 238:1019–1024.
- 10. Harley, C. B., and R. P. Reynolds. 1987. Analysis of *Escherichia* coli promoter sequences. Nucleic Acids Res. 15:2343-2361.
- Horwitz, M. S. Z., and L. A. Loeb. 1988. DNA sequences of random origin as probes of *Escherichia coli* promoter architecture. J. Biol. Chem. 263:14724–14731.
- Hryniewicz, M., A. Palucha, and M. D. Hulanicka. 1988. Construction of cys::lac gene fusions in *Escherichia coli* and their use in the isolation of constitutive cysB^c mutants. J. Gen. Microbiol. 134:763-769.
- 13. 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.
- Hulanicka, M. D., T. KJopotowski, and D. A. Smith. 1972. The effect of triazole on cysteine biosynthesis in Salmonella typhimurium. J. Gen. Microbiol. 72:291–301.
- 15. Hulanicka, M. D., C. Garrett, G. Jagura-Burdzy, and N. M. Kredich. 1986. Cloning and characterization of the cysAMK region of Salmonella typhimurium. J. Bacteriol. 168:322-327.
- Isihara, H., and R. W. Hogg. 1980. Amino acid sequence of sulfate binding protein from Salmonella typhimurium LT2. J. Biol. Chem. 255:4614–4618.
- 17. Jagura-Burdzy, G., and D. Hulanicka. 1981. Use of gene fusions to study expression of *cysB*, the regulatory gene of the cysteine regulon. J. Bacteriol. 147:744-751.
- Karbonowska, H., A. Wiater, and M. D. Hulanicka. 1977. Sulfate permease of *Escherichia coli* K12. Acta Biochim. Polon. 24:329–334.
- Kredich, N. M. 1987. Biosynthesis of cysteine, p. 419-428 In F. C. Neidhardt (ed.), Escherichia coli and Salmonella typhimurium: cellular and molecular biology, vol. 1. American Society for Microbiology, Washington, D.C.
- Kushner, S. R., H. Nagaishi, A. Templin, and A. J. Clark. 1971. Genetic recombination in *Escherichia coli*: the role of exonuclease I. Proc. Natl. Acad. Sci. USA 68:824–827.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227:680-685.
- Lyons, L. B., and N. D. Zinder. 1972. The genetic map of the filamentous bacteriophage f1. Virology 49:45-60.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- 24. Miller, J. H. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Mizobuchi, K., M. Demerec, and D. H. Gillespie. 1962. Cysteine mutants of Salmonella typhimurium. Genetics 47:1617–1627.
- Mott, E. J., J. L. Galloway, and T. Platt. 1985. Maturation of Escherichia coli tryptophan operon mRNA: evidence for 3' exonucleolytic processing after rho-dependent termination. EMBO J. 4:1887-1891.
- 27. Noel, N., K. Nikaido, and G. F.-L. Ames. 1979. A single amino acid substitution in a histidine-transport protein drastically alters its mobility in sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Biochemistry 18:4159–4165.
- Normark, S., S. Bergstrom, T. Edlung, T. Grundstrom, B. Jaurin, F. P. Lindberg, and O. Olsson. 1983. Overlapping genes. Annu. Rev. Genet. 17:499-525.
- 29. Ohta, N., P. R. Galsworthy, and A. B. Pardee. 1971. Genetics of

sulfate transport by Salmonella typhimurium. J. Bacteriol. 105:1053-1062.

- Ostrowski, J., G. Jagura-Burdzy, and N. M. Kredich. 1987. DNA sequences of the cysB regions of Salmonella typhimurium and Escherichia coli. J. Biol. Chem. 262:5999–6005.
- Ostrowski, J., and N. M. Kredich. 1989. Molecular characterization of the cysJIH promoters of Salmonella typhimurium and Escherichia coli: regulation by cysB protein and N-acetyl-L-serine. J. Bacteriol. 171:130–140.
- Pardee, A. B. 1966. Purification and properties of a sulfatebinding protein from Salmonella typhimurium. J. Biol. Chem. 241:5886-5892.
- Pardee, A. B., L. S. Prestidge, M. B. Whipple, and J. Dreyfuss. 1966. A binding site for sulfate and its relation to sulfate transport into Salmonella typhimurium. J. Biol. Chem. 241: 3962-3969.
- 34. Parra, F., P. Britton, C. Castle, M. C. Jones-Mortimer, and H. L. Kornberg. 1983. Two separate genes involved in sulfate transport in *Escherichia coli* K12. J. Gen. Microbiol. 129: 353-358.
- Pflugrath, J. W., and F. A. Quiocho. 1988. The 2Å resolution structure of the sulfate-binding protein involved in active transport in *Salmonella typhimurium*. J. Mol. Biol. 200:163–180.
- 36. Platt, T. 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. 55:339–372.
- 37. Rosenberg, M., and D. Court. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. 13:319–353.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-termination inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
- Shine, J., and L. Dalgarno. 1974. The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome-binding sites. Proc. Natl. Acad. Sci. USA 71:1342-1346.
- Sirko, A. E., M. Zatyka, and M. D. Hulanicka. 1987. Identification of the *Escherichia coli cysM* gene encoding *O*-acetylserine sulfhydrylase B by cloning with mini-Mu-lac containing a plasmid replicon. J. Gen. Microbiol. 133:2719-2725.
- 41. Sirko, A., M. Hryniewicz, D. Hulanicka, and A. Böck. 1990. Sulfate and thiosulfate transport in *Escherichia coli* K-12: nucleotide sequence and expression of the *cysTWAM* gene cluster. J. Bacteriol. 172:3351-3357.
- 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.
- Tabor, S., and C. C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA 84:4767–4771.
- 44. van Heijne, G. 1986. A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. 14:4683-4691.
- 45. Whitaker, J. R., and P. E. Granum. 1980. An absolute method for protein determination based on difference in absorbance at 235 and 280 nm. Anal. Biochem. 109:156–159.
- Willis, R. C., R. G. Morris, C. Cirakoglu, G. D. Schellenberg, N. H. Gerber, and C. E. Furlong. 1974. Preparation of periplasmic binding proteins from *Salmonella typhimurium* and *Escherichia coli*. Arch. Biochem. Biophys. 161:64–75.
- Winans, S. C., S. J. Elledge, J. H. Krueger, and G. C. Walker. 1985. Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli*. J. Bacteriol. 161:1219– 1221.
- Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene 33:103–119.