JOURNAL OF BACTERIOLOGY, Mar. 1994, p. 1630–1638
0021-9193/94/$04.00+0

Vol. 176, No. 6

A Gene at 59 Minutes on the *Escherichia coli* Chromosome Encodes a Lipoprotein with Unusual Amino Acid Repeat Sequences

JEFFREY K. ICHIKAWA, CHUAN LI, JUNE FU, AND STEVEN CLARKE*

Department of Chemistry and Biochemistry and the Molecular Biology Institute, University of California, Los Angeles, California 90024-1569

Received 9 December 1992/Accepted 10 January 1994

We report a 1.432-kb DNA sequence at 59 min on the *Escherichia coli* chromosome that connects the published sequences of the *pcm* gene for the isoaspartyl protein methyltransferase and that of the *katF* or *rpoS* (*katF/rpoS*) gene for a sigma factor involved in stationary-phase gene expression. Analysis of the DNA sequence reveals an open reading frame potentially encoding a polypeptide of 379 amino acids. The polypeptide sequence includes a consensus bacterial lipidation sequence present at residues 23 to 26 (Leu-Ala-Gly-Cys), four octapeptide proline- and glutamine-rich repeats of consensus sequence QQPQIQPV, and four heptapeptide threonine- and serine-rich repeats of consensus sequence PTA(S,T)TTE. The deduced amino acid sequence, especially in the C-terminal region, is similar to that of the *Haemophilus somnus* LppB lipoprotein outer membrane antigen (40% overall sequence identity; 77% identity in last 95 residues). The LppB lipoprotein binds Congo red dye and has been proposed to be a virulence determinant in *H. somnus*. Utilizing a plasmid construct with the *E. coli* gene under the control of a phage T7 promoter, we demonstrate the lipidation of this gene product by the incorporation of [^{3}H]palmitic acid into a 42-kDa polypeptide. We also show that treatment of *E. coli* cells with globomycin, an inhibitor of the lipoprotein signal peptidase, results in the accumulation of a 46-kDa precursor. We thus designate the protein NlpD (new lipoprotein D). *E. coli* cells overexpressing NlpD bind Congo red dye, suggesting a common function with the *H. somnus* LppB protein. Disruption of the chromosomal *E. coli nlpD* gene by insertional mutagenesis results in decreased stationary-phase survival after 7 days.

We are interested in the widely distributed L-isoaspartyl protein methyltransferase that can modify altered aspartyl residues arising from spontaneous chemical degradation of proteins (3, 6). The methylation reaction can initiate steps leading to the restoration of the normal L-configuration of such residues and may participate as such in a protein repair step (12, 22, 32). The *pcm* gene for this enzyme had been cloned, sequenced, and mapped to the 59-min region on the chromosome of *Escherichia coli* (11). It is about 1 kb upstream of the *katF* (or *rpoS*) (*katF/rpoS*) gene whose product shows strong homology to the *E. coli rpoD* σ^{70} protein (34) and appears to be a sigma factor in the central regulation of stationary-phase gene expression (28, 37, 43). It is interesting that *pcm* mutant cells demonstrate two of the phenotypes of *katF/rpoS* mutant cells: reduced stationary-phase survival and lessened heat-shock resistance (28, 29, 36). Genes with similar functions can be adjacent on the *E. coli* chromosome, even if they are not present in one operon. For example, four operons containing genes for chemotaxis and mobility are adjacent at 41.5 min; similarly, three operons containing genes for ribosomal proteins are adjacent at about 72.4 to 73.4 min on the *E. coli* chromosome (1). We were thus interested in exploring whether a polypeptide or polypeptides with a function potentially similar to that of *pcm* or *katF/rpoS* might be encoded in the region between these genes.

In this work, we have determined and analyzed the DNA sequence of this region between the *pcm* and the *katF/rpoS* genes. An open reading frame of 379 amino acid residues with a consensus lipidation sequence was found. Preliminary results have been presented showing that the insertion of a kanamycin resistance cassette at a *Sna*BI site in the middle of this open reading frame does not affect exponential-growth phase but does affect long-term survival (29). These results suggest that this gene may also be involved in stationary-phase survival. Here, we show that the open reading frame does in fact encode a lipoprotein with unusual octapeptide and heptapeptide sequence repeats in its N-terminal region. The deduced sequence of its C-terminal region is very similar to that of a Congo red dye-binding outer membrane lipoprotein from *Haemophilis somnus*.

MATERIALS AND METHODS

Bacterial strains and plasmids. The *E. coli* strains and plasmids used in this study and their sources are listed in Table 1. Plasmid DNAs used as templates for dideoxy sequencing or in coupled in vitro transcription-translation reactions were prepared by column chromatography following the instructions of the manufacturer (Qiagen).

DNA sequencing. DNA sequences were determined by the manual dideoxy methodology for *E. coli* chromosomal DNA inserts in plasmids derived from pMMkatF1 as described previously (11). Specifically, the Sanger reaction was carried out with Sequenase version 2.0 (U.S. Biochemicals) with 5′-α-^{35}S-dATP (~1,000 Ci/mmol; New England Nuclear) using both dGTP and dITP protocols. The oligonucleotide primers were synthesized by using a Pharmacia Gene Assembler Plus and purified as described elsewhere (11). Part of the insert in

* Corresponding author. Mailing address: Department of Chemistry and Biochemistry, UCLA, Los Angeles, CA 90024-1569. Phone: (310) 825-8754. Fax: (310) 206-7286. Electronic mail address: clarke@ewald.mbi.ucla.edu.

TABLE 1. Strains and plasmids used in this study

Strain or plasmid	Description	Source or reference
Strains		
MC1000	F⁻ *araD139* Δ(*araABC-leu*)*7679 galU galK* Δ(*lac*)*X74 rpsL thi*	29
CL3010	MC1000; *orf2*(*nlpD*)::Km^r	29
C600	*thr leu lacY thi supE hsdR tonA*	S. Tabor; 42
BL21 (DE3)	F⁻ *hsdS gal ompT* r_B^- m_B^- with the λ phage DE3 lysogen containing the T7 RNA polymerase gene under the control of the *lacUV5* promoter	41
Plasmids		
pMMkatF1	10.2-kb *Bam*HI *E. coli* chromosomal fragment in pAT153	P. Loewen; 35
pCL1	Religation of a 7.5-kb *Eco*RI-*Sna*BI fragment from pMMkatF1	29
pCL2	4.5-kb *Bam*HI-*Bsp*HI fragment from pMMkatF1 inserted in the *Bam*HI-*Sma*I polylinker site of pGEM-7Zf(+) (Promega) (contains the *pcm* gene and upstream regions and the 5′ portion of the *katF*/*rpoS* gene)	This study
pJF3	1.4-kb *Eco*RV fragment from pMMkatF1 inserted in pUC19	11
pGP1-2	A derivative of pACYC177 containing gene *1* of phage T7 under the control of the inducible λ p_L promoter and the *cI857* gene for the heat-sensitive λ repressor	S. Tabor; 42
pJKI10	1.4-kb blunt-ended *Cla*I-*Dra*I fragment of pMMkatF1 inserted into the *Sma*I site of pBluescript II SK⁺ (Stratagene) so that the insert *nlpD* gene is under the control of the vector T7 promoter	This study

pJF3 was sequenced automatically by Robert Pringle (University of California at Los Angeles Core Sequencing Facility) using an Applied Biosystems model 370A and the fluorescence-labeled M13 reverse primer. The 5% Long Ranger (AT Biochem, Malvern, Pa.) sequencing gel was prepared, and electrophoresis was conducted at 48 W in 54 mM Tris–54 mM boric acid–1.2 mM EDTA (pH 8.0) (0.6× TBE) buffer for about 2 h. DNA sequences were analyzed on Macintosh computers utilizing DNAStar programs.

Primer extension analysis of the transcriptional initiation. Total RNA from *E. coli* DH5α cells (BRL Life Sciences) containing plasmid pCL2 (Table 1) was prepared by hot phenol extraction (2). Cells were grown in Luria-Bertani (LB) media with 50 μg of ampicillin per ml to mid-log phase and were then harvested and resuspended in 0.15 M sucrose–0.01 M sodium acetate, pH 5.4 (1 volume of buffer per 12.5 volumes of original culture). An additional volume of the resuspension buffer containing 2% sodium dodecyl sulfate (SDS) at 65°C was added to lyse cells. Extraction with phenol equilibrated with 50 mM sodium acetate (pH 5.2) was performed three times at 65°C. The aqueous phase was then incubated with RNase-free DNase (10 U/ml; Stratagene) for 40 min at 37°C, followed by phenol-chloroform and chloroform extraction. Usually, 20 μg of RNA per milliliter of *E. coli* culture could be obtained. The final ethanol-precipitated RNA was stored dry at −20°C until use.

Oligonucleotide primer 102 (5′-CATAGCGAAACCAGT GACAAAGCCG) complementary to the region 41 to 65 bp downstream of the start codon of the *nlpD* open reading frame was end labeled by T4 polynucleotide kinase (BRL Life Sciences) with [γ-^{32}P]ATP (>3,000 Ci/mmol; ICN). A total of 10^6 cpm of ^{32}P-labeled primer and 100 μg of *E. coli* RNA were annealed in the presence of RNasin RNase inhibitor (final concentration, 4 U/μl; Promega) at 70°C, with a gradual decrease of temperature to 37°C in 90 min. Avian myeloblastosis virus reverse transcriptase (final concentration, 0.625 U/μl; Promega), four deoxyribonucleotide triphosphates (0.6 mM each; Pharmacia), and 5 mM $MgCl_2$ were included for the extension reaction mixture at 43°C for 40 min. The reaction mixture was quenched with 4 volumes of 4 M ammonium acetate (pH 5.9)–20 mM EDTA, and the product was precipitated with ethanol and then resuspended in 10 μl of stop dye (47.5% formamide, 10 mM EDTA, 0.025% bromophenol blue, 0.025% xylene cyanol FF). Primer extension products were analyzed on a sequencing gel with parallel standards of DNA dideoxy-sequencing reactions using the end-labeled primers described above and denatured pCL1 as the template DNA.

In vitro-coupled transcription and translation. Each reaction mixture included 1 to 5 μg of plasmid DNA, 30 to 40 μCi of Tran^{35}S-label (>400 Ci/mmol, a mixture of [^{35}S]methionine and [^{35}S]cysteine; ICN), 0.033 U of *E. coli* RNA polymerase (Boehringer Mannheim) per μl, a mixture of amino acids, nucleotides, and essential small molecules, *E. coli* S100 fraction, and ribosomes kindly supplied by Marilyn Leonard and William Wickner, University of California at Los Angeles (cf. reference 15) in a final volume of 37.5 μl. After incubation at 37°C for 45 min, the reaction was stopped by the addition of an equal volume of 10% ice-cold trichloroacetic acid. The reaction products were pelleted by a centrifugation at 14,000 × *g* for 5 min at 4°C and washed with 90% ethanol. The ^{35}S-labeled products were analyzed by SDS-polyacrylamide gel electrophoresis (26) and then by autoradiography.

Overexpression of the 59-min gene product. A 1.4-kb *Cla*I-*Dra*I fragment from pMMkatF1 including the region between the *pcm* and *katF* genes (11) was blunt ended by treatment with the Klenow fragment of DNA polymerase and ligated into the *Sma*I site of pBluescript II SK⁺ (Stratagene). From a number of white colonies on LB agar plates containing 50 μg of ampicillin and 40 μg of X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) (BRL Life Sciences) per ml, we selected one clone (pJKI10) that was shown by restriction endonuclease analysis to contain the *nlpD* gene in an orientation such that its expression was controlled by the vector T7 polymerase promoter. *E. coli* C600 containing pGP1-2 and BL21 (DE3) were transformed with pJKI10 according the procedure described by Chung et al. (5).

Lipoprotein analysis. Overnight cultures of *E. coli* cells harboring pJKI10 grown in LB broth (Difco) at 37°C were washed twice and diluted 100-fold into 5 ml of fresh M9 medium containing 0.4% glucose, 1% Casamino Acids (Difco), and 1 μg of thiamine per ml (40). These cultures were grown to mid-log phase (optical density at 600 nm of 0.4) and heat shocked at 42°C for 30 min (42). [9,10-^{3}H]palmitic acid (52.4 Ci/mmol, 5 mCi/ml in ethanol; New England Nuclear) was

added to a final radioactivity of 100 μCi/ml to a 2-ml aliquot of the culture. The labeling reaction was allowed to continue for 20 min at 37°C. Cells were harvested by centrifugation and washed in 1.0 M Tris-HCl (pH 8.0) at room temperature and resuspended in 200 μl of Laemmli gel sample buffer (26). Aliquots (60 μl each) were applied to 1.2-cm-wide wells of 1.5-mm-thick polyacrylamide slab gels. Alternatively, cell cultures were radiolabeled with Tran^{35}S-label (see above) according to the procedure described by Tabor and Richardson (42). Cells (3 ml) were grown to mid-log phase (optical density at 600 nm of 0.4) in LB broth at 37°C and then washed once and resuspended in M9 medium containing 0.4% glucose and 0.01% of each of the 20 amino acids except cysteine and methionine. Portions of these cells (1 ml) were adapted in this medium for 30 min at 37°C and then heat shocked at 42°C for 20 min. At this point, rifampin was added to a final concentration of 200 μg/ml, and the mixture was incubated for 10 min at 42°C. When indicated, globomycin was added at this point to a final concentration of 100 μg/ml from a stock solution of 20 mg/ml in methanol. After a 20-min incubation at 37°C, the Tran^{35}S-label was added to a final radioactivity of 10 μCi/ml, and cells were incubated for 10 min at 37°C. Cells were harvested and resuspended in sample buffer as described above.

Congo red binding. Overnight cultures of *E. coli* C600 cells containing the pGP1-2-pBluescript II SK$^+$ or pJKI10 dual-plasmid system were diluted 10^6- and 10^7-fold and incubated on LB broth plates containing 50 μg of ampicillin and 100 μg of kanamycin per ml and 0.005% Congo red dye (Sigma Chemicals) at either 30, 37, or 40°C for up to 48 h. Alternatively, overnight cultures of *E. coli* BL21 (DE3) cells containing either pBluescript II SK$^+$ or pJKI10 were diluted 10^6- and 10^7-fold and incubated on LB broth plates containing 50 μg of ampicillin per ml and 0.005% Congo red dye at 37°C in the presence or absence of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG; Sigma Chemicals) for up to 48 h.

Nucleotide sequence accession number. The nucleotide sequence of the 1,432-bp *Cla*I-*Dra*I fragment of the *E. coli* chromosome at 59 min shown in Fig. 1 has been deposited in the GenBank data base and has been given the accession number L07869.

RESULTS

Identification of a gene between *pcm* and *katF* at 59 min on the *E. coli* chromosome. The DNA sequence was determined for a 1.432-kb region between the *Cla*I site in the *pcm* gene and the *Dra*I site in the *katF* gene. Both strands of the appropriate DNA of either plasmid pJF10 or pCL2 were sequenced, corresponding to positions 2882.80 to 2881.35 kb on the *E. coli* chromosome physical map (25, 33). This sequence connects the sequences published for the downstream region of the *pcm* gene (11) and the upstream region of the *katF* gene (31, 34) (Fig. 1). A single open reading frame encoding 379 amino acids is present, separated by 141 bp from the coding region of *pcm* and by 64 bp from the coding region of *katF*. The transcriptional start site was mapped by primer extension analysis. When an oligonucleotide primer complementary to the +41-to-+65 region from the ATG start codon of the open reading frame was used for the extension reaction, the transcription start site was mapped to a G residue 44 bp upstream from the ATG start codon (Fig. 2). A σ^{70} promoter upstream of this transcriptional start site is suggested by the −10 region TAG GGT and a −35 region TGAACA spaced by 17 bp (Fig. 1), corresponding to consensus sequences TATAAT and TT GACA, respectively (16).

To determine whether the open reading frame that we identified from DNA sequence analysis is expressed, plasmid pCL2 containing this region was constructed and used as a template for coupled in vitro transcription-translation reactions. As shown in Fig. 3, a 44-kDa polypeptide is expressed as the product of DNA from pCL2 but not as that of DNA from pGEM, the cloning vector of pCL2. The calculated molecular weight of the unmodified 379-amino-acid polypeptide (40,151) is about 4,000 less than the molecular weight determined by the mobility of the product on gel electrophoresis. This difference may result from a large net positive charge on the protein (+9.39 calculated at pH 7.0) or from possible covalent modifications.

The 59-min gene encodes a lipoprotein. Analysis of the deduced N-terminal amino acid sequence revealed a typical bacterial signal sequence with a Leu-Ala-Gly-Cys sequence at residues 23 to 26 from the initiator Met residue (Fig. 1). This tetrapeptide sequence fits well with the consensus lipidation sequence of bacterial lipoproteins (17, 46). These lipoproteins are modified by glycerol additions to the Cys residue, fatty acylation of the glyceryl Cys residue, and cleavage of the bond between the modified Cys residue and the preceding residue by the globomycin-sensitive lipoprotein signal peptidase (17). This suggests that the mature gene product is a lipoprotein of 354 amino acid residues with a modified N-terminal cysteine residue. We attempted to directly demonstrate the lipidation of the gene product by labeling intact cells with [^{3}H]palmitic acid. However, no differences in the patterns of radioactive bands were detected on fluorographs of SDS gels of the labeled proteins either from MC1000 (wild-type) cells or from MC1000 cells with plasmid pCL2 containing the gene or CL3010 cells in which this gene was disrupted (data not shown). However, when we overexpressed the gene in a plasmid construct under a T7 polymerase promoter, we could clearly see incorporation of [^{3}H]palmitic acid into a 42-kDa polypeptide (Fig. 4). This polypeptide mass can be compared with that observed for the unprocessed polypeptide in the transcription-translation experiment (44 kDa) (Fig. 3). Additionally, we were able to show that globomycin treatment results in the appearance of a larger 46-kDa polypeptide species and the inhibition of the generation of the 42-kDa mature form from the inhibition of the lipoprotein signal peptidase (Fig. 5).

We have thus designated this gene *nlpD* for new lipoprotein D. The NlpD species described here is now the fourth *E. coli* lipoprotein (or putative lipoprotein) identified by its consensus lipidation sequence. This group includes the lipoproteins NlpA (21, 48) and NlpB (4) as well as the putative lipoprotein NlpC assigned as ORF-17 adjacent to the *btu* operon (GenBank accession number M14031 [10]). It appears that our failure to detect NlpD lipidation except when the T7 polymerase-driven overexpression vector was used probably results from a combination of the poor expression of the *nlpD* gene and its comigration with a more abundant lipidated species with a polypeptide that has a similar molecular mass (data not shown). This latter species may represent the lipoprotein NLP2 described previously as a 39-kDa outer-membrane species (19). Although we have not determined the subcellular localization of the NlpD lipoprotein, the absence of an aspartate residue following the cysteine residue in the consensus lipidation sequence is consistent with an outer membrane localization (13, 14). However, the presence of an aspartate residue in the second position after the modified cysteine residue can lead to a weak inner membrane sorting signal in a hybrid lipo-β-lactamase protein, and context-specific effects have also been previously noted (14).

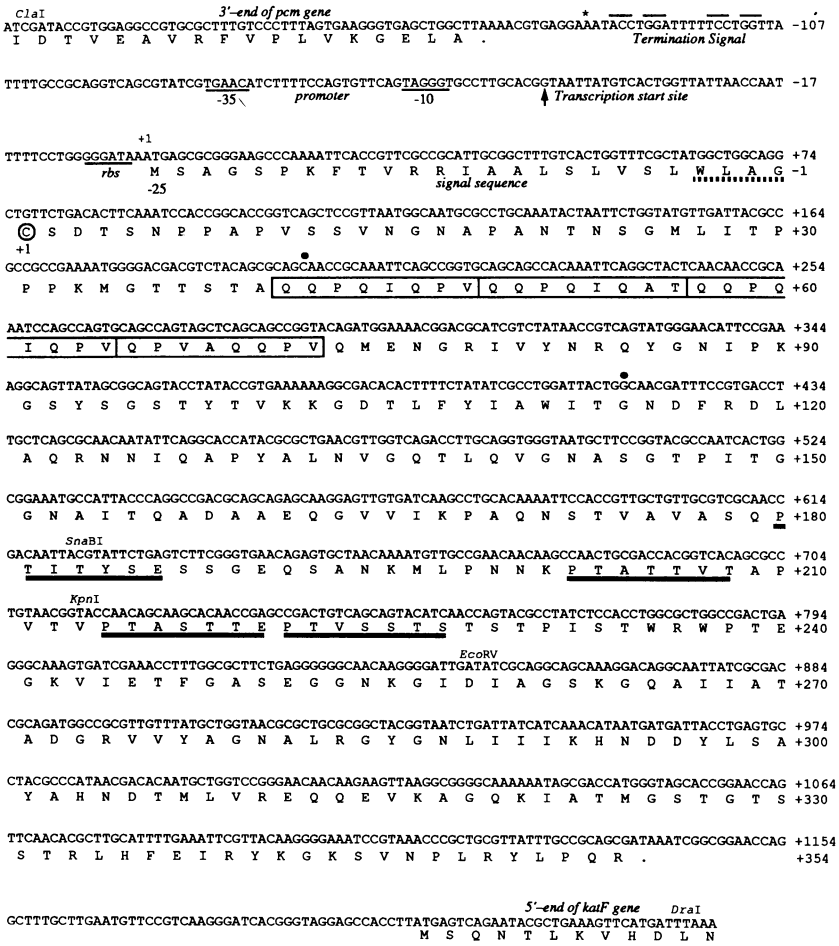

FIG. 1. DNA sequence analysis in the region of the *E. coli* chromosome at 59 min between the *Cla*I site of the *pcm* gene and the *Dra*I site of the *katF/rpoS* gene. A single open reading frame defining a new gene coding for a potential lipoprotein, designated NlpD, is found between these two genes. Features of this gene include a transcriptional start site mapped by primer extension at base −44 (arrow), a potential sigma 70 promoter at bases −78 to −55 (with the −35 and −10 elements underlined), a potential ribosomal binding site at bases −8 to −3 (with bases corresponding to 16S RNA underlined), and a coding region from bases +1 to +1137. A possible signal sequence from amino acids −25 to −1 and a bacterial lipidation consensus sequence from −4 to +1 (dashed underline) are indicated. The cysteine residue at position +1 (circled) is the putative site of lipidation. Boxed amino acid residues represent the glutamine- and proline-rich octapeptide repeats, and the heavy underlined amino acid residues represent the serine and threonine heptapeptide repeats. This sequence overlaps with the previously published sequences for the *pcm* gene (11) from nucleotides −196 to +32 (corrected by addition of a single A at the position indicated by an asterisk on the top line) and for the *katF/rpoS* gene from nucleotides +878 to +1236 (34). A number of differences in sequence from that given in reference 34 were found here between nucleotides +880 and +1180. However, this sequence matches that of the corrected *katF/rpoS* sequence in the GenBank data base (accession number X16400 [28a]) in this region. The entire sequence is also identical to that recently submitted to the EMBL database (accession number D17549 [42a]), with the exception of the two positions marked here by dots; the C nucleotide at position +199 is not present, and the G at position +416 is a C. We have rechecked our data at these two positions to confirm the sequence shown above.

Homology of the *E. coli* NlpD lipoprotein with the LppB lipoprotein of *H. somnus*. When the nucleotide and the deduced amino acid sequences of NlpD were compared with those in the nonredundant Swiss Prot/PIR/SPUpdate/GenPept/GPUpdate data bases in the BLAST network search service (National Center for Biotechnology Information), a strong similarity was found to the translated amino acid sequence of the LppB 40-kDa antigenic outer membrane lipoprotein of *H. somnus*, an intracellular pathogen of cattle (44). Additionally, nucleic acid similarity between the *nlpD* and *lppB* genes that extended beyond the assigned 3′ end of the *lppB*-coding region (cf. reference 44) was found. Thus, we examined alternate reading frames of the *lppB* gene and found that modification of its nucleotide sequence by two base deletions and one base insertion leads to an open reading frame with a striking resemblance to the NlpD-deduced protein sequence (Fig. 6). When the encoded protein sequences of the LppB and NlpD lipoproteins are aligned as shown in Fig. 6, we find identity at 154 positions of the total of 384 amino acid residues compared (40%) and identity at 73 of the C-terminal 95 residues (77%). Further sequencing of the 3′ end of the *lppB* gene may reveal additional similarity with the terminal 5′ coding end of the *nlpD* gene.

The binding of the aromatic dye Congo red by the LppB protein has been suggested as evidence for the interaction of this lipoprotein with hydrophobic molecules such as hemin (44). We thus asked whether the NlpD protein would similarly interact with Congo red dye. We found that colonies of *E. coli* cells containing the plasmid pJKI10 overexpressing the *nlpD* gene under the control of the T7 polymerase turned red on

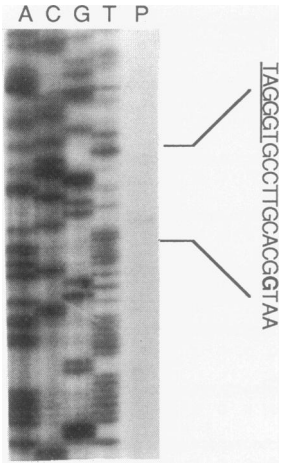

FIG. 2. Primer extension analysis of the putative lipoprotein mRNA. A 20-bp segment of the DNA sequence of the sense strand starting from the −10 region of the putative promoters is illustrated. DNA dideoxy-sequencing reactions with denatured pCL1 as the template and with the same end-labeled primer for primer extension were run in parallel with the primer extension products on a sequencing gel as standards. The ACGT label on each lane indicates the terminating dideoxyribonucleotide included in the sequencing reaction. The product of the primer extension reaction was electrophoresed in lane P.

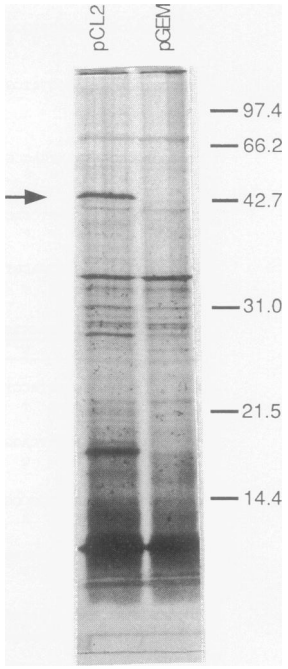

FIG. 3. Products of in vitro transcription-translation reactions of DNA from the 59-min region of the *E. coli* chromosome. ^{35}S-labeled polypeptides were analyzed by Laemmli gel electrophoresis (12.5% acrylamide–0.43% *N,N*-methylenebisacrylamide) and then by autoradiography. The template plasmid included in the reaction is indicated on top of the lane. The position at which the precursor of the putative lipoprotein electrophoreses is indicated by an arrow. The 17-kDa polypeptide in the pCL2 lane is due to the expression of a gene upstream of *pcm*. Protein molecular mass standards were electrophoresed in parallel lanes and include rabbit muscle phosphorylase *b* (97.4 kDa), bovine serum albumin (66.2 kDa), hen egg white ovalbumin (42.7 kDa), bovine carbonic anhydrase (31.0 kDa), soybean trypsin inhibitor (21.5 kDa), and hen egg white lysozyme (14.4 kDa).

agar plates containing 0.005% Congo red dye when the polymerase was induced either by heat shock induction in *E. coli* C600 cells with pGP1-2 or by IPTG induction in the BL21 (DE3) strain (data not shown). However, colonies containing cells in which the T7 polymerase was not induced did not turn red, nor did colonies with cells containing the parent pBluescript II SK$^+$ in place of pJKI10 containing the *nlpD* gene (data not shown). These results suggest that homologous regions of the LppB and NlpD lipoproteins can bind this aromatic dye and might reflect endogenous affinity of these proteins for hemin or other hydrophobic molecules (see Discussion).

Unusual repeats in the deduced polypeptide sequence. The N-terminal part of the polypeptide is extraordinarily rich in glutamines and prolines; 17 Gln residues and 15 Pro residues are present in the first 90 amino acid residues starting from the +1 cysteine. A Gln- and Pro-rich octamer with a consensus sequence QQPQIQPV is directly repeated four times between residues 41 and 72 (Fig. 1). From positions 180 to 227, four Ser- and Thr-rich heptamer repeats with the consensus sequence PT(I,A,V) (T,S) (T,S) (T,S) (E,T,S) are found (Fig. 1). We found several proteins with sequence similarities in the PQ repeat region, including the circumsporozoite proteins of the malarial parasites *Plasmodium yoelii* (27) and *Plasmodium berghei* (47), as well as the outer membrane protein pertactin from *Bordetella parapertussis*, a causative agent of whooping cough (30) (Fig. 7). Each of these proteins is thought to interact directly with the eucaryotic cell surface and to be important in the invasiveness of these organisms. Significantly, the regions containing the octamer and heptamer repeats in the NlpD protein are not highly conserved with the LppB lipoprotein of *H. somnus*, suggesting that these proteins can have different physiological functions.

Stationary-phase survival of *E. coli* cells with a disrupted *nlpD* gene. It is interesting that the genes both upstream and downstream of the *nlpD* gene described here are involved in stationary-phase survival. The upstream *pcm* gene can catalyze the removal of altered aspartyl residues that can accumulate in age-damaged proteins and peptides (29). The downstream *katF/rpoS* gene encodes a sigma factor that plays a central role in starvation and stationary-phase survival (18, 28, 34, 37, 43). In a study of mutations of *pcm* and adjacent genes, evidence was presented that the insertion of a kanamycin cassette into the lipoprotein gene (described initially as *orf2*) results in no change in exponential growth in rich or minimal media but does affect its stationary-phase survival to a limited extent (29). In Table 2, we present the results of more extensive studies that indicate significant differences in the viability of the *nlpD*$^+$ (MC1000) and *nlpD* (CL3010) mutant cells after 7 or 9 days in stationary phase. Since stationary-phase CL3010 cells survive well after H_2O_2 challenge while the *katF* mutant cells do not (29), the reduced viability of CL3010 cells in stationary phase does not appear to be due to defective expression of the downstream *katF/rpoS* gene, although polar effects of the *nlpD* disruption resulting in reduced *katF/rpoS* expression cannot be ruled out at present. The stationary-phase defect is not due to reduced levels of the upstream *pcm* gene product, because its methyltransferase activity is not affected by the *nlpD* disruption in CL3010 cells (29). Although the stationary-phase defect in *nlpD* mutant cells is not as pronounced as that seen in mutations of the *pcm* or *katF/rpoS* genes, there is a fourfold

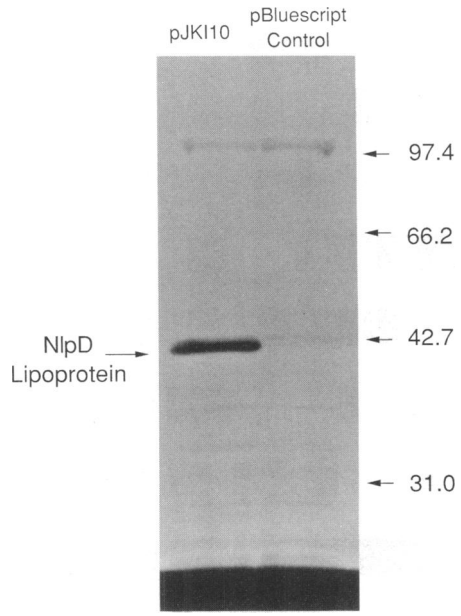

FIG. 4. Incorporation of [^{3}H]palmitic acid into polypeptides of *E. coli* cells overexpressing the *nlpD* gene. Labeling reactions and SDS-gel electrophoresis (8% acrylamide–0.28% *N,N*-methylenebisacrylamide) were performed as described in Materials and Methods with *E. coli* C600 containing pGP1-2 and either the parent control vector (pBluescript II SK$^+$) or its derivative containing the *nlpD* gene under the control of the vector T7 promoter (pJKI10) that overexpresses the NlpD protein. After electrophoresis, the gels were treated with En^{3}Hance (Dupont/NEN, Boston, Mass.), dried, and fluorographed for 7 days at −80°C. The position of the NlpD lipoprotein is marked by an arrow. The molecular masses of marker proteins electrophoresed in parallel lanes are given in kilodaltons as in Fig. 3. The large radioactive band at the bottom of the gel probably corresponds to [^{3}H]palmitic acid incorporated into phospholipid.

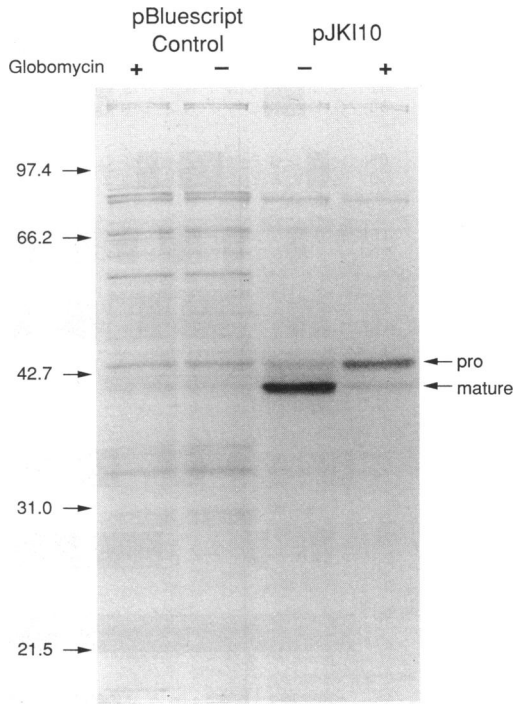

FIG. 5. Effect of globomycin on the maturation of the NlpD lipoprotein. *E. coli* C600 cells containing pGP1-2 and either the control vector pBluescript II SK$^+$ or this vector with an insert containing the *nlpD* gene under the control of the vector T7 promoter (pJKI10) were labeled with [^{35}S]cysteine and [^{35}S]methionine as described in Materials and Methods in the presence or absence of 100 μg of globomycin per ml. Proteins were analyzed by SDS-polyacrylamide gel electrophoresis as shown in Fig. 4. After electrophoresis, gels were treated with En^{3}Hance, dried, and fluorographed for 2 h at −80°C. The positions of the radiolabeled mature and precursor (pro) forms of the NlpD lipoprotein are indicated by arrows. The molecular masses of marker proteins electrophoresed in parallel lanes are given in kilodaltons as in Fig. 3.

difference in the survival of *nlpD*$^+$ and *nlpD* mutant cells after 9 days.

DISCUSSION

We describe here a new lipoprotein gene (*nlpD*) at 59 min on the *E. coli* chromosome. This gene encodes a 379-residue polypeptide with a typical lipidation consensus sequence at its amino terminus. On the basis of the conserved maturation reactions of bacterial lipoproteins (17, 45), we propose that amino acid residues −25 to −1 represent a signal sequence that directs the modification of the cysteine residue in position +1 to give a fatty acyl glyceride as well as the cleavage of the bond between residues −1 and +1 by the globomycin-sensitive lipoprotein signal peptidase. We have now experimentally demonstrated the incorporation of [^{3}H]palmitic acid into the NlpD polypeptide chain and the inhibition of the conversion of the lipidated precursor to the mature form by globomycin. Interestingly, this proposed signal sequence of 25 residues is longer than any of the 14 bacterial lipoproteins described in reference 46.

The physiological function of the NlpD lipoprotein is unclear at present. We find a small, but significant, defect in the ability of *E. coli* cells with a disrupted *nlpD* gene to form colonies after extensive incubation under stationary-phase conditions. However, we find no differences in the growth of the mutant stain under exponential-phase conditions (29).

Significantly, the *E. coli* NlpD lipoprotein shares identities at 154 of 384 residues (40%) with the *H. somnus* LppB lipoprotein (44). This latter protein is an outer membrane component of this pathogen that may be an immunodominant surface antigen for its bovine host (7, 44). Its expression has been identified with a Congo red dye-binding phenotype that has been associated with virulence in a number of pathogenic bacteria (44). We have demonstrated that *E. coli* cells overexpressing its NlpD lipoprotein also bind Congo red dye. It thus appears that the Congo red dye-binding elements are present in the conserved sequence domains of these two proteins. On the other hand, the unusual octamer and heptamer repeats of the *E. coli* NlpD lipoprotein are not conserved with the *H. somnus* LppB lipoprotein.

The physiological relevance of the Congo red dye-binding phenotype is unclear at present, although its association with virulence is seen in a variety of cells, including *Shigella flexneri* (8), *Yersinia enterocolitica* (39), and *Aeromonas salmonicida* (24), as well as *E. coli* (38). In *A. salmonicida* (24) and *S. flexneri* (8), hemin and protoporphyrin can compete with Congo red dye binding, suggesting that some aspect of porphyrin or iron assimilation may be mediated by these binding proteins. However, it should be stressed that there is no evidence to date that the Congo red dye-binding proteins from these organisms are structurally related to either the NlpD or

FIG. 6. Comparison of the deduced amino acid sequences of the *E. coli* NlpD lipoprotein described here and the *H. somnus* LppB lipoprotein (44). To maximize the similarity of these sequences, three alterations were made to the *lppB* nucleotide sequence (44); a C at position 788 and an A from position 796 were deleted, and a G was inserted at position 983. Identical residues are indicated by the dark shading.

the LppB lipoproteins. Thus, it would be useful to directly determine whether the NlpD and LppB species can function as hemin- or porphyrin-binding and transport proteins. If these proteins do serve in an alternative iron assimilation pathway, it is possible that their loss in stationary-phase cells limits iron intake and cell viability. Interestingly, *Salmonella* strains containing *katF* mutations have reduced virulence in mice (9).

Congo red dye-binding is also characteristic of β-amyloid proteins, such as the amyloid component in Alzheimer β-protein or the prion protein (20). In fact, it has recently been shown that a peptide corresponding to residues 28 to 44 of the *E. coli* OsmB lipoprotein with a sequence similar to that present in the mammalian Alzheimer's amyloid and prion proteins can bind Congo red dye (20). This observation suggests that the Congo red dye-binding phenotype need not be always associated with hemin binding. The OsmB lipoprotein has no apparent sequence similarity to the NlpD and LppB lipoproteins (23), although all three proteins have regions rich in glycine and alanine residues that have been associated with Congo red dye binding (20). The OsmB protein provides resistance to osmotic stress and may thus be important itself in stationary-phase survival (23).

Finally, the local sequence similarity of proteins in the QP-rich tandem repeat region shown in Fig. 7 suggests possible roles for NlpD protein in adhesion to eucaryotic cells. The

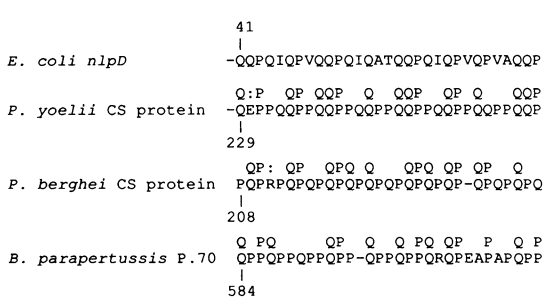

FIG. 7. Similarity between the deduced amino acid sequence of the QP octapeptide repeat region of the putative *E. coli* NlpD lipoprotein described here, the circumsporozoite (CS) protein from *P. yoelii* (27) and *P. berghei* (47), and the outer membrane protein pertactin from *B. parapertussis* (30).

TABLE 2. Long-term viability of MC1000 and CL3010 cells[a]

Time in stationary phase (days)	% Maximum viability[b]		
	MC1000 (*nlpD*$^+$)	CL3010 (*nlpD*)	*P*[c]
2	92.15 ± 15.14	83.12 ± 21.85	NS
3	74.77 ± 8.69	50.63 ± 14.20	NS
5	53.24 ± 6.95	40.48 ± 14.93	NS
7	39.24 ± 4.25	20.36 ± 4.66	<0.05
9	23.63 ± 5.08	6.35 ± 2.17	<0.05

[a] Cells were grown at 37°C in 0.2% glucose in M9 minimal medium supplemented with 40 μg of L-leucine and 0.2 μg of thiamine per ml.

[b] Percent maximum viability is the number of CFU of the culture formed on LB plates at each time point divided by the number of CFU of the same culture after 30 h in stationary phase, corresponding to 50 h after the initial inoculation of the culture. Values are the means ± standard deviations from three independent experiments.

[c] *P* values indicate the statistical probability that the difference between the percent maximum viability of MC1000 cells and that of CL3010 cells at each indicated time point would occur, given that no difference actually exists. NS, not significant.

other proteins described here with the QP repeat motif are membrane proteins of animal pathogens or parasites that are involved with microbial attachment to the host cell membrane.

ACKNOWLEDGMENTS

We thank Marilyn Leonard and William Wickner (Dartmouth University) for their help with coupled in vitro transcription-translation experiments and their gift of globomycin, Donald Guiney (University of California, San Diego) and Virginia Miller (University of California, Los Angeles) for valuable discussions, and Kenneth E. Rudd (National Institutes of Health) for suggesting the *nlpD* designation for the lipoprotein gene.

This work was supported by grant GM26020 from the National Institute of General Medical Sciences.

REFERENCES

1. **Bachmann, B. J.** 1987. Linkage map of *Escherichia coli* K-12, edition 7, p. 807–876. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
2. **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.
3. **Barten, D. M., and R. F. O'Dea.** 1990. The function of protein carboxyl methyltransferase in eucaryotic cells. Life Sci. **47:**181–194.
4. **Bouvier, J., A. P. Pugsley, and P. Stragier.** 1991. A gene for a new lipoprotein in the *dapA-purC* interval of the *Escherichia coli* chromosome. J. Bacteriol. **173:**5523–5531.
5. **Chung, C. T., S. L. Niemela, and R. H. Miller.** 1989. One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. Proc. Natl. Acad. Sci. USA **86:**2172–2175.
6. **Clarke, S.** 1985. Protein carboxyl methyltransferases: two distinct classes of enzymes. Annu. Rev. Biochem. **54:**479–506.
7. **Corbeil, L. B., S. A. Kania, and R. P. Gogolewski.** 1991. Characterization of immunodominant surface antigens of *Haemophilus somnus*. Infect. Immun. **59:**4295–4301.
8. **Daskaleros, P. A., and S. M. Payne.** 1987. Congo red binding phenotype is associated with hemin binding and increased infectivity of *Shigella flexneri* in the HeLa cell model. Infect. Immun. **55:**1393–1398.
9. **Fang, F. C., S. J. Libby, N. A. Buchmeier, P. C. Loewen, J. Switala, J. Harwood, and D. G. Guiney.** 1992. The alternative sigma factor KatF (RpoS) regulates *Salmonella* virulence. Proc. Natl. Acad. Sci. USA **89:**11978–11982.
10. **Friedrich, M. J., L. C. DeVeaux, and R. J. Kadner.** 1986. Nucleotide sequence of the *btuCED* genes involved in vitamin B_{12} transport in *Escherichia coli* and homology with components of periplasmic-binding-protein-dependent transport systems. J. Bacteriol. **167:**928–934.
11. **Fu, J., L. Ding, and S. Clarke.** 1991. Purification, gene cloning, and sequence analysis of an L-isoaspartyl protein carboxyl methyltransferase from *Escherichia coli*. J. Biol. Chem. **266:**14562–14572.
12. **Galletti, P., A. Ciardiello, D. Ingrosso, A. Di Donato, and G. D'Alessio.** 1988. Repair of isopeptide bonds by protein carboxyl O-methyltransferase: seminal ribonuclease as a model system. Biochemistry **27:**1752–1757.
13. **Gennity, J. M., and M. Inouye.** 1991. The protein sequence responsible for lipoprotein membrane localization in *Escherichia coli* exhibits remarkable specificity. J. Biol. Chem. **266:**16458–16464.
14. **Gennity, J. M., H. Kim, and M. Inouye.** 1992. Structural determinants in addition to the amino-terminal sorting sequence influence membrane localization of *Escherichia coli* lipoproteins. J. Bacteriol. **174:**2095–2101.
15. **Gold, L. M., and M. Schweiger.** 1971. Synthesis of bacteriophage-specific enzymes directed by DNA in vitro. Methods Enzymol. **20:**537–542.
16. **Harley, C. B., and R. P. Reynolds.** 1987. Analysis of *E. coli* promoter sequences. Nucleic Acids Res. **15:**2343–2361.
17. **Hayashi, S., and H. Wu.** 1990. Lipoproteins in bacteria. J. Bioenerg. Biomembr. **22:**451–471.
18. **Hengge-Aronis, R.** 1993. Survival of hunger and stress: the role of *rpoS* in early stationary phase gene regulation in *E. coli*. Cell **72:**165–168.
19. **Ichihara, S., M. Hussain, and S. Mizushima.** 1981. Characterization of new membrane lipoproteins and their precursors of *Escherichia coli*. J. Biol. Chem. **256:**3125–3129.
20. **Jarrett, J. T., and P. T. Lansbury, Jr.** 1992. Amyloid fibril formation requires a chemically discriminating nucleoation event: studies of an amyloidogenic sequence from the bacterial protein OsmB. Biochemistry **31:**12345–12352.
21. **Jin, T., K. E. Rudd, and M. Inouye.** 1992. The *nlpA* lipoprotein gene is located near the *selC* tRNA gene on the *Escherichia coli* chromosome. J. Bacteriol. **174:**3822–3823.
22. **Johnson, B. A., E. D. Murray, Jr., S. Clarke, D. B. Glass, and D. W. Aswad.** 1987. Protein carboxyl methyltransferase facilitates conversion of atypical L-isoaspartyl peptides to normal L-aspartyl peptides. J. Biol. Chem. **262:**5622–5629.
23. **Jung, J. U., C. Gutierrez, and M. R. Villarejo.** 1989. Sequence of an osmetically inducible lipoprotein gene. J. Bacteriol. **171:**511–520.
24. **Kay, W. W., B. M. Phipps, E. E. Ishiguro, and T. J. Trust.** 1985. Porphyrin binding by the surface array virulence protein of *Aeromonas salmonicida*. J. Bacteriol. **164:**1332–1336.
25. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50:**495–508.
26. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
27. **Lal, A. A., V. F. de la Cruz, J. A. Welsh, Y. Charoenvit, W. L. Malay, and T. F. McCutchan.** 1987. Structure of the gene encoding the circumsporozoite protein of *Plasmodium yoelii*. J. Biol. Chem. **262:**2937–2940.
28. **Lange, R., and R. Hengge-Aronis.** 1991. Identification of a central regulator of stationary-phase gene expression in *Escherichia coli*. Mol. Microbiol. **5:**49–59.

28a. **Loewen, P. C., I. van Ossowski, J. Switala, and M. R. Mulvey.** 1993. KatF (σ^S) synthesis in *Escherichia coli* is subject to posttranscriptional regulation. J. Bacteriol. **175:**2150–2153.

29. **Li, C., and S. Clarke.** 1992. A protein methyltransferase specific for altered aspartyl residues is important in *Escherichia coli* stationary-phase survival and heat-shock resistance. Proc. Natl. Acad. Sci. USA **89:**9885–9889.
30. **Li, L. J., G. Dougan, P. Novotny, and I. G. Charles.** 1991. P. 70 pertactin, an outer-membrane protein from *Bordetella parapertussis*: cloning, nucleotide sequence and surface expression in *Escherichia coli*. Mol. Microbiol. **5:**409–417.
31. **McCann, M. P., C. D. Fraley, and A. Matin.** 1993. The putative σ factor KatF is regulated posttranscriptionally during carbon starvation. J. Bacteriol. **175:**2143–2149.
32. **McFadden, P. N., and S. Clarke.** 1987. Conversion of isoaspartyl peptides to normal peptides: implications or the cellular repair of damaged proteins. Proc. Natl. Acad. Sci. USA **84:**2595–2599.
33. **Medigue, C., J. P. Bouche, A. Henaut, and A. Danchin.** 1990. Mapping of sequenced genes (700 kbp) in the restriction map of the *Escherichia coli* chromosome. Mol. Microbiol. **4:**169–187.
34. **Mulvey, M. R., and P. C. Loewen.** 1989. Nucleotide sequence of *Escherichia coli* suggests katF protein is a novel transcription factor. Nucleic Acids Res. **17:**9979–9991.
35. **Mulvey, M. R., P. A. Sorby, B. L. Triggs-Raine, and P. C. Loewen.** 1988. Cloning and physical characterization of *katE* and *katF* required for catalase HPII expression in *Escherichia coli*. Gene **73:**337–345.
36. **Mulvey, M. R., J. Switala, A. Borys, and P. C. Loewen.** 1990. Regulation of transcription of *katE* and *katF* in *Escherichia coli*. J. Bacteriol. **172:**6713–6720.
37. **Nguyen, L. H., D. B. Jensen, N. E. Thompson, D. R. Gentry, and R. R. Burgess.** 1993. In vitro functional characterization of over-

produced *Escherichia coli katF/rpoS* gene product. Biochemistry **32:**11112–11117.

38. **Payne, S. M., and R. A. Finkelstein.** 1977. Detection and differentiation of iron-responsive avirulent mutants on Congo Red agar. Infect. Immun. **18:**94–98.
39. **Prpic, J. K., R. M. Robins-Browne, and R. B. Davey.** 1983. Differentation between virulent and avirulent *Yersinia enterocolitica* isolates by using Congo red agar. J. Clin. Microbiol. **18:**486–490.
40. **Rupp, W. D., C. E. Wilde III, and D. L. Reno.** 1971. Exchanges between DNA strands in ultraviolet-irradiated *Escherichia coli*. J. Mol. Biol. **61:**25–44.
41. **Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff.** 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. **185:**60–89.
42. **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.

42a. **Takahashi, H.** Unpublished EMBL data base submission D17549.

43. **Tanaka, K., Y. Takayanagi, N. Fujita, A. Ishihama, and H. Takahashi.** 1993. Heterogeneity of the principal sigma factor in *Escherichia coli*: the *rpoS* gene product, sigma^{38}, is a second principal sigma factor of RNA polymerase in stationary-phase *Escherichia coli*. Proc. Natl. Acad. Sci. USA **90:**3511–3515.
44. **Theisen, M., C. R. Rioux, and A. A. Potter.** 1993. Molecular cloning, nucleotide sequence, and characterization of *lppB*, encoding an antigenic 40-kilodalton lipoprotein of *Haemophilus somnus*. Infect. Immun. **61:**1793–1798.
45. **Tokunaga, M., H. Tokunaga, and H. C. Wu.** 1982. Post-translational modification and processing of *Escherichia coli* prolipoprotein *in vitro*. Proc. Natl. Acad. Sci. USA **79:**2255–2259.
46. **von Heijne, G.** 1989. The structure of signal peptides from bacterial lipoproteins. Protein Eng. **2:**531–534.
47. **Weber, J. L., J. E. Egan, J. A. Lyon, R. A. Wirtz, Y. Charoenvit, W. L. Maloy, and W. T. Hockmeyer.** 1987. *Plasmodium berghei*: cloning of the circumsporozoite protein gene. Exp. Parasitol. **63:**295–300.
48. **Yu, F., S. Inouye, and M. Inouye.** 1986. Lipoprotein-28, a cytoplasmic membrane lipoprotein from *Escherichia coli*. J. Biol. Chem. **261:**2284–2288.