JOURNAL OF BACTERIOLOGY, Jan. 1991, p. 435–442
0021-9193/91/020435-08$02.00/0

Vol. 173, No. 2

Primary Sequence of the *Escherichia coli fadL* Gene Encoding an Outer Membrane Protein Required for Long-Chain Fatty Acid Transport

PAUL N. BLACK

Department of Biochemistry, College of Medicine, University of Tennessee, Memphis, Tennessee 38163

Received 30 July 1990/Accepted 26 October 1990

The *fadL* gene of *Escherichia coli* encodes an outer membrane protein (FadL) that plays a central role in the uptake of exogenous long-chain fatty acids. The nucleotide sequence of the *fadL* gene revealed a single open reading frame of 1,344 bp encoding a protein with 448 amino acid residues and a molecular weight of 48,831. The transcriptional start, analyzed by primer extension, was shown to be 95 bp upstream from the translational start. Apparent −10 and −35 regions were found at −12 and −37 bp upstream from the transcriptional start. Three regions with hyphenated dyad symmetry (two between the transcriptional start and the translational start and one upstream from the −10 and −35 regions) were identified that may play a role in the expression of *fadL*. The protein product of the *fadL* gene contained a signal sequence and signal peptidase I cleavage site similar to that defined for other *E. coli* outer membrane proteins. The N-terminal sequence of the mature FadL protein was determined by automated amino acid sequencing of protein purified from the outer membrane of a strain harboring *fadL* under the control of a T7 RNA polymerase-responsive promoter. This amino acid sequence, Ala-Gly-Phe-Gln-Leu-Asn-Glu-Phe-Ser-Ser, verified the signal peptidase I cleavage site on pre-FadL and confirmed the N-terminal amino acid sequence of FadL predicted from the DNA sequence. Mature FadL contained 421 amino acid residues, giving a molecular weight of 45,969. The amino acid composition of FadL deduced from the DNA sequence suggested that this protein contained an abundance of hydrophobic amino acid residues and lacked cysteinyl residues. The hydrophobic amino acids within FadL were predicted to contribute to at least five regions of the protein with an overall hydrophobic character. The amino acid sequence of FadL was used to search GenBank for other proteins with amino acid sequence homology. These data demonstrated that FadL and the heat-modifiable outer membrane protein P1 of *Haemophilus influenzae* type b were 60.5% conserved and 42.0% identical over 438 amino acid residues.

The uptake of exogenous long-chain fatty acids into *Escherichia coli* requires the outer membrane protein FadL (1–4, 18, 25–28). FadL acts to bind long-chain fatty acids with high affinity (2) and subsequently allows these compounds to traverse the outer membrane. The biochemical mechanism that governs the activity of FadL in this transport process remains undefined. We are interested in defining how long-chain fatty acids specifically traverse the cell envelope of *E. coli*. Central to this problem is understanding how FadL functions in the long-chain fatty acid transport process.

The *fadL* gene was identified by Nunn and co-workers and shown to be required for the uptake of exogenous long-chain fatty acids (3, 4, 18, 25–28). We have shown that the product of the *fadL* gene is localized in the outer membrane (4), acts as a high-affinity long-chain fatty acid-binding protein (2, 26), is a receptor for the bacteriophage T2 (1), and is likely to be peptidoglycan associated (1). These data imply that FadL forms a specific channel for the uptake of long-chain fatty acids across the outer membrane. Black et al. (3) cloned the *fadL* gene and identified the *fadL* gene product by using maxicell analysis of plasmid-encoded proteins. FadL, like many outer membrane proteins of gram-negative bacteria, exhibits a heat-modifiable property in the presence of sodium dodecyl sulfate (3, 4). When FadL is heated at 100°C, its M_r is 43,000, whereas when it is heated at 30°C, its M_r is 33,000.

Said et al. reported the sequence of the *fadL* gene and defined the transcriptional initiation site by S1 protection (31). In the course of our studies with in vivo and in vitro mutagenesis of the cloned *fadL* gene to define regions of functional importance on FadL (16), it became apparent that the sequence published by Said et al. was incorrect (31). In particular, the N-terminal amino acid sequence of FadL, the transcriptional start of *fadL*, and a number of base pairs throughout the sequence proposed were inconsistent with our data. Additionally, the data presented by Said et al. (31) showed a FadL protein without a signal sequence normally associated with outer membrane proteins. In this report I present the correct nucleotide sequence of the *fadL* gene, define the transcriptional start of FadL by primer extension, and define the N-terminal amino acid sequence of FadL by using automated amino acid sequencing. These data defined a signal sequence on the unprocessed form of FadL that shares homology with other signal sequences of outer membrane proteins. Furthermore, I present evidence that FadL and the outer membrane protein P1 of *Haemophilus influenzae* type b share considerable amino acid sequence identity, suggesting that these two heat-modifiable proteins may share structural and functional features.

MATERIALS AND METHODS

Bacterial strains, media, and growth conditions. *E. coli* JM103 was used for transformation and propagation of M13 clones containing fragments of the *fadL* gene as described by Messing (21). *E. coli* BL21/plysS (35) harboring the *fadL*$^+$ plasmid pN132 (see below) was used for purification of FadL. Total cellular RNA was isolated from *E. coli* C600. Transformation with M13 clones and plasmids of interest used the standard $CaCl_2$ procedures described by Daghert

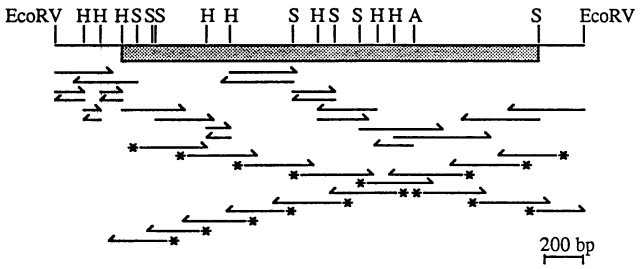


FIG. 1. Sequencing strategy of the *fadL* gene. The shaded area of the 2.8-kb *Eco*RV fragment represents the 2,197 bp presented in Fig. 2. Arrows preceded by an asterisk indicate regions of sequence determined by using *fadL*-specific oligonucleotides and the *fadL*$^+$ plasmid pN103; H, S, and A refer to restriction sites of *Hpa*II, *Sau*3A, and *Acc*I, respectively.

and Ehrlich (5). Cells were routinely grown in Luria broth or tryptone broth (TB) (22) at 37°C in a LabLine gyratory shaker. Bacterial growth was monitored with a Klett-Summerson colorimeter and a blue filter.

Cloning and sequencing. The strategy used to sequence the *fadL* gene is shown in Fig. 1. *Sau*3A and *Hpa*II fragments of the *fadL* gene were cloned into M13mp18 restricted with *Bam*HI and *Acc*I, respectively, and selected as clear plaques. Single-stranded DNAs were isolated from recombinant phage and sequenced by the dideoxy-chain termination method of Sanger et al. (33). Sequencing reactions were preformed with [α-^{35}S]dATP and Sequenase (United States Biochemicals) and the 17-mer universal primer. A series of *fadL*-specific oligonucleotides (17-mers) was synthesized on a Pharmacia-LKB Gene Assembler Plus and used to sequence the *fadL*$^+$ plasmid pN103 to provide overlap between M13 clones and resolve regions of ambiguity. Before sequencing and annealing to *fadL*-specific primers, 1 to 2 μg of plasmid (pN103) DNA was denatured in 0.2 M NaOH, ethanol precipitated, dried, and suspended in 2 μl of glass-distilled water. Sequencing reactions were loaded on a 0.2-mm by 30-cm by 60-cm standard sequencing gel (19) and electrophoresed at 90 to 95 W (constant power).

The sequence information was analyzed by using programs through the Genetics Computer Group Sequence Analysis Software Package (8).

Mapping the transcriptional initiation site of *fadL*. The initiation site of transcription was identified by primer extension of the 5′ end-labeled *fadL*-specific oligonucleotide 5′-ACCGGACCGGCCCGAATAATCCCC-3′ hybridized to in vivo-synthesized mRNA by reverse transcriptase. Purified oligonucleotide (0.5 to 1.0 μg) was 5′ end labeled with [γ-^{32}P]ATP as described by Maniatis et al. (19) with T4 polynucleotide kinase. Total in vivo-synthesized RNA was isolated from a 100-ml midlog culture of C600 grown in TB supplemented with 5 mM oleate–0.5% Brij 58 by a modification of the procedure of Kassavetis and Geiduschek (14). After cell lysis, sample volumes were adjusted to 3.4 ml with 100 mM Tris hydrochloride (pH 8.0)–2 mM EDTA, and 2-mercaptoethanol was added to 1.0%. Solid CsCl (4.56 g) was added to each sample, dissolved thoroughly, and dispensed into two TLS-50 polyallomer tubes. Samples were centrifuged at 50,000 rpm in a TL-100 table-top ultracentrifuge (TLS-50 rotor; Beckman) for 2.5 h. Supernatants were carefully removed with a siliconized Pasteur pipette, and the RNA pellet was suspended in 20 mM Tris hydrochloride (pH 8.0)–100 mM NaCl–0.1 mM EDTA. RNA was precipitated by the addition of 0.1 volume of 10 M LiCl and 2.5 volumes of absolute ethanol overnight at −80°C. RNA was collected by centrifugation, dried, and suspended in 0.1 of mM EDTA to a final concentration of 10 mg/ml.

Construction of *fadL*$^+$ expression vector. A 2.8-kb *Bam*HI fragment containing the entire *fadL* gene was isolated from the *fadL*$^+$ plasmid pN130 (unpublished data) and cloned into *Bam*HI-restricted and bacterial alkaline phosphatase-treated pCD130. pCD130 is derived from pT7-5 and contains the *fadR* gene (encoding the transcriptional repressor of the fatty acid transport and degradative enzyme genes) cloned in the opposite orientation from that of the T7 RNA polymerase-responsive promoter (9). In the resultant plasmid, pN132, the control of *fadL* was placed under the T7 RNA polymerase-responsive promoter in addition to its own promoter. *fadR* was maintained on the plasmid to provide control of the *fadL* promoter, because otherwise this gene was toxic. pN132 was transformed into strain BL21/plysS and, after induction with isopropyl-β-D-thiogalactopyranoside as described by Studier and Moffat (35), synthesized FadL at high levels.

Purification of FadL and N-terminal amino acid sequence analysis. FadL was isolated from a 0.5-liter midlog culture of strain BL21/plysS (pN132) grown in 2YT containing 100 μg of ampicillin per ml and 15 μg of chloramphenicol per ml after induction with isopropyl-β-D-thiogalactopyranoside for 2 h by a modification of published methods (4). After growth, cells were collected by centrifugation, washed once in minimal salts, and frozen at −80°C for 1 h. Cells were thawed on ice and suspended in 50 ml of cold 0.75 M sucrose–10 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*-2-ethanesulfonic acid) (pH 7.5)–10 mM EDTA–1 mM phenylmethylsulfonyl fluoride (PMSF)–1 μg of DNase I per ml and were subjected to two cycles of sonication (30 s, full power). The total cell envelope was isolated from the clarified cell extract by centrifugation at 22,000 rpm for 90 min (SW28 rotor; Beckman). The total cell envelope was suspended by sonication in 10 mM HEPES (pH 7.5), applied to a discontinuous sucrose gradient (17, 54, and 70%), and centrifuged at 22,000 rpm (SW28 rotor; Beckman) for 18 h, resulting in inner and outer membrane fractions. FadL was partially purified from the outer membrane fraction by using the different detergent extraction protocols previously described (4). With this approach, nearly 70% of the protein in the Tween 20-insoluble, Triton X-100-soluble extract was FadL. A sample containing 500 μg of total extracted protein was separated on a preparative discontinuous 5% sodium dodecyl sulfate–12% polyacrylamide gel (3) and electroblotted to a prewetted polyvinylidene difluoride membrane (Immobilon Transfer; Whatman) in 10 mM 3-(cyclohexylamino)-1-propanesulfonic acid (pH 11.0)–10% methanol as described by Matsudaira (20). FadL was identified by staining the polyvinylidene difluoride membrane with 1% Ponceau S red in 5% acetic acid for 5 min and destaining in 5% acetic acid for 10 min. The strip containing FadL was excised with a razor blade, rinsed with glass-distilled water for 15 min, air dried, and stored at −20°C. The N-terminal amino acid sequence of FadL (200 to 250 pmol) on the polyvinylidene difluoride membrane was determined by automated Edman degradation with an Applied Biosystems protein sequencer with a 120A online PTH-AA analyzer at the Harvard University microchemistry facility.

Materials. Reagents and enzymes used for sequencing and transcriptional mapping were purchased from United States Biochemicals, Bethesda Research Laboratories, or New England BioLabs. Regents used for oligonucleotide synthesis were obtained from Pharmacia-LKB. [α-^{35}S]dATP and

[γ-^{32}P]ATP were purchased from New England Nuclear. Antibiotics and other supplements for bacterial growth were obtained from Difco and Sigma. All other chemicals were obtained from standard suppliers and were of reagent grade.

Nucleotide sequence accession number. The nucleotide sequence shown in Fig. 2 has been submitted to GenBank under accession number M37714.

RESULTS

Sequence of the *fadL* gene. Based on our previous complementation analyses, the *fadL* gene was localized to a 2.8-kb *Eco*RV fragment (3). In the present work, we sequenced this DNA fragment and identified the *fadL* coding region. The *fadL* gene is located within the 2,197-bp fragment presented in Fig. 2 and encodes a single polypeptide of 448 amino acid residues. The calculated molecular weight of this protein (48,831) was somewhat higher than that estimated from sodium dodecyl sulfate-polyacrylamide gel electrophoresis of the purified protein (3, 4). The protein sequence deduced from the DNA sequence demonstrated that FadL had no cysteinyl residues and contained a large number of hydrophobic residues. Both findings were in agreement with total amino acid compositional analysis (4). Sequence analysis identified a second, incomplete reading frame at the end of the 2.8-kb *Eco*RV fragment that, based on analysis with the GCG program TFASTA, was fused to lambda DNA. Since the original *fadL* clone was identified from a lambda library and subsequently subcloned into pACYC177 (3), the lambda DNA was likely to be a cloning artifact. Although the incomplete reading frame identified was closely linked to *fadL*, it was not required for complementation of *fadL* mutants and thus is presented in the paper.

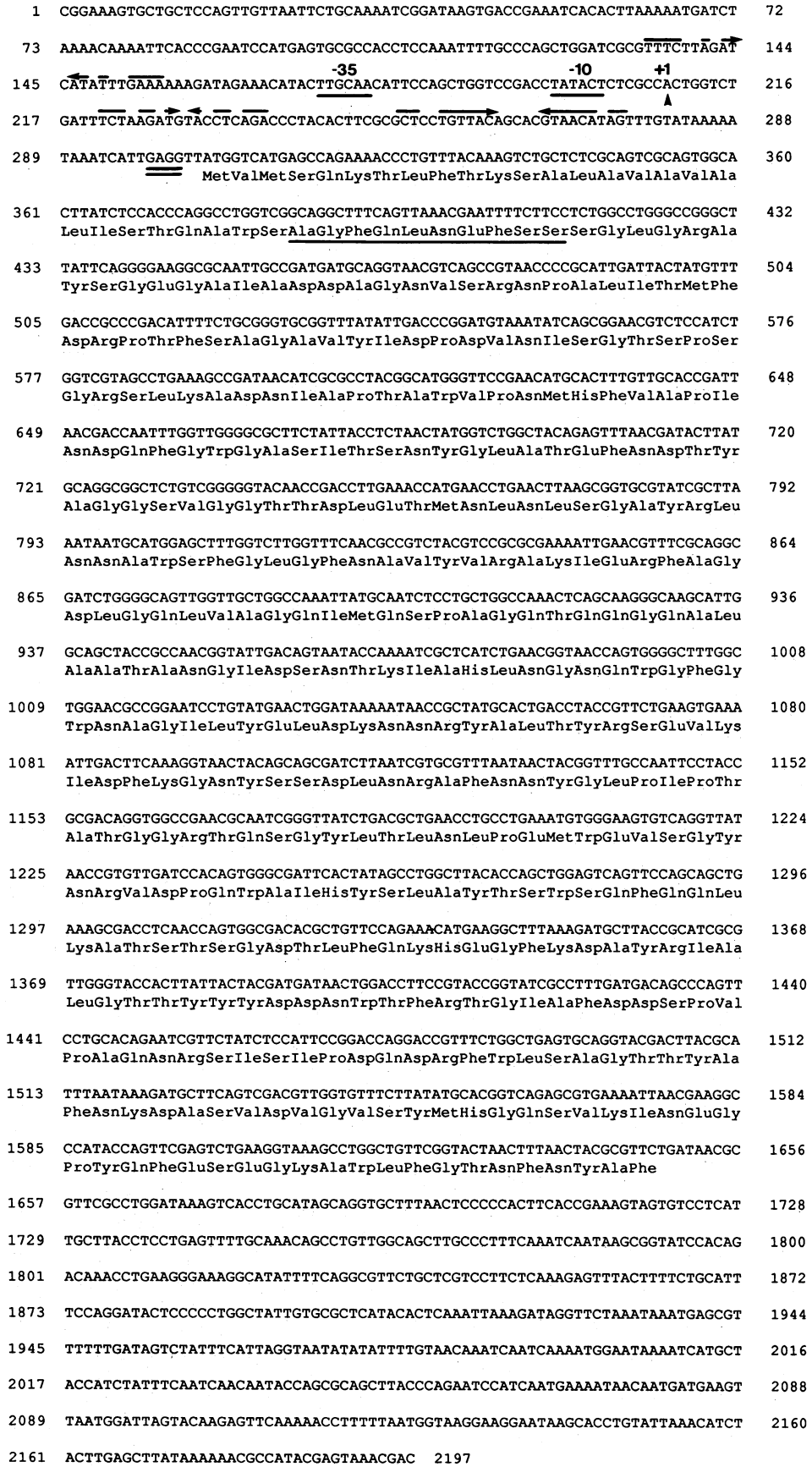
Features of the *fadL* sequence. The open reading frame identified contained two possible translational start sites (nucleotides 304 and 310) separated by GTC. We cannot say with certainty which ATG represents the initiation methionine. Six base pairs upstream from the first ATG is a purine-rich segment (5′-GAGG-3′) that shares homology with ribosomal binding sites (34). The transcriptional start, defined by primer extension (Fig. 3), was positioned 95 bp upstream from the first ATG. The adenine residue at this position has been designated +1 (Fig. 2). The −10 and −35 sequences were defined by consensus at positions −12 (5′-TATACT-3′) and −37 (5′-TTGCAA-3′), respectively, from the transcriptional start. The spacing between these two regions is consistent with that documented for other *E. coli* promoters. A region of hyphenated dyad symmetry at −74 (5′-TTTCTTAGATCATATTTGAAA-3′) was identified that shares some homology with sequences involved in the binding of cyclic AMP and catabolite gene activator protein (6). Two regions with hyphenated dyad symmetry found at positions +12 (5′-TCTAAGATGTACCTCAGA-3′) and +44 (5′-CTCCTGTTACAGCACGTAACATAG-3′) may play a role in the regulation of the *fadL* gene. Between the proposed transcriptional start and the translational start of *fadL*, a reading frame of 14 amino acid residues was identified (between positions +16 and +58). This reading frame was not preceded by sequences that resemble a ribosome binding site and thus more likely represents untranslated sequences. The sequence downstream from the translational stop was characterized by two inverted repeats.

Features of FadL. The open reading frame defined a protein with a molecular weight of 48,831, exceeding the M_r of 43,000 predicted from sodium dodecyl sulfate-polyacrylamide gels (3, 4). The predicted sequence of FadL suggested the presence of a signal sequence consisting of 27 amino acid residues. This signal sequence shared features common to other outer membrane proteins in that it had a charged segment followed by a hydrophobic segment and the signal peptidase I recognition site Ala-Trp-Ser (29). In an effort to define the N-terminal amino acid sequence of the mature FadL protein, *fadL* was cloned into a T7 RNA polymerase expression system (see Materials and Methods). Upon induction of the *fadL* gene, FadL is produced in substantial quantities. Using this expression system, FadL was partially purified from isolated outer membranes and electroblotted to polyvinylidene difluoride membranes for N-terminal amino acid sequence analysis. Data from these experiments demonstrated that the N-terminal amino acid sequence of FadL was Ala-Gly-Phe-Gln-Leu-Asn-Glu-Phe-Ser-Ser and confirmed the predicted cleavage site of the signal peptide. The N-terminal alanyl residue of the mature FadL protein has been designated +1. These data predict that mature FadL contains 421 amino acid residues, giving a molecular weight of 45,969. Amino acid compositional analysis of purified FadL (4) and that defined from the DNA sequence were in close agreement, with the exception that the DNA sequence predicted 8 mol% of Glx and 13 mol% of Asx rather than 12.5 and 7 mol%, respectively.

Previous data demonstrated that FadL consisted of a large number of hydrophobic residues (4), and therefore it was of interest to determine whether there was any pattern of hydrophobic residues along the linear sequence of the protein. The predicted amino acid sequence of FadL was subjected to a hydrophilicity analysis with the algorithms of Hopp and Woods and an averaging length of 6 amino acid residues (12). Data from this analysis indicated that FadL contained a number of regions with an overall hydrophobic character (Fig. 4). The role(s) of these hydrophobic regions in the transport and binding of exogenous long-chain fatty acids is unknown. We have, however, recently identified an *fadL* mutant by using linker mutagenesis of the cloned *fadL* gene that has a lesion in the last hydrophobic region (16). This mutant binds long-chain fatty acids at wild-type levels but is defective for long-chain fatty acid transport and encodes a protein that is not heat modifiable (16). These data imply that transport of exogenous long-chain fatty acids and the heat-modifiable property of FadL may be related.

FadL, in addition to the outer membrane protein OmpF, serves as a receptor for the bacteriophage T2 (1). In this respect, we were interested in determining whether there were any regions of amino acid homology between these two proteins that might suggest common regions for T2 binding. Comparisons of the amino acid sequences of OmpF and FadL by using the GCG programs TFASTA and BESTFIT did not reveal any regions of significant homology. These data suggest that regions of FadL and OmpF that are involved in the binding of T2 differ in primary amino acid sequence. This does rule out the possibility that FadL and OmpF share some secondary or tertiary structure that is required for the binding of T2.

FadL and other heat-modifiable outer membrane proteins from gram-negative bacteria. The heat-modifiable characteristic of FadL is similar to that seen for several other outer membrane proteins identified in gram-negative bacteria. Using the GCG program BESTFIT, we compared the amino acid sequence of FadL to the *E. coli* outer membrane proteins OmpA and phospholipase (7, 24) and the *H. influenzae* type b outer membrane protein P1 (23). This analysis revealed no significant homology between FadL, OmpA, and phospholipase. This analysis did, however, reveal strik-




```
1    CGGAAAGTGCTGCTCCAGTTGTTAATTCTGCAAAATCGGATAAGTGACCGAAATCACACTTAAAAATGATCT   72

73   AAAACAAAATTCACCCGAATCCATGAGTGCGCCACCTCCAAATTTTGCCCAGCTGGATCGCGTTTCTTAGAT  144

145  CATATTTGAAAAAAGATAGAAACATACTTGCAACATTCCAGCTGGTCCGACCTATACTCTCGCCACTGGTCT  216

217  GATTTCTAAGATGTACCTCAGACCCTACACTTCGCGCTCCTGTTACAGCACGTAACATAGTTTGTATAAAAA  288

289  TAAATCATTGAGGTTATGGTCATGAGCCAGAAAACCCTGTTTACAAAGTCTGCTCTCGCAGTCGCAGTGGCA  360
                MetValMetSerGlnLysThrLeuPheThrLysSerAlaLeuAlaValAlaValAla

361  CTTATCTCCACCCAGGCCTGGTCGGCAGGCTTTCAGTTAAACGAATTTTCTTCCTCTGGCCTGGGCCGGGCT  432
     LeuIleSerThrGlnAlaTrpSerAlaGlyPheGlnLeuAsnGluPheSerSerSerGlyLeuGlyArgAla

433  TATTCAGGGGAAGGCGCAATTGCCGATGATGCAGGTAACGTCAGCCGTAACCCCGCATTGATTACTATGTTT  504
     TyrSerGlyGluGlyAlaIleAlaAspAspAlaGlyAsnValSerArgAsnProAlaLeuIleThrMetPhe

505  GACCGCCCGACATTTTCTGCGGGTGCGGTTTATATTGACCCGGATGTAAATATCAGCGGAACGTCTCCATCT  576
     AspArgProThrPheSerAlaGlyAlaValTyrIleAspProAspValAsnIleSerGlyThrSerProSer

577  GGTCGTAGCCTGAAAGCCGATAACATCGCGCCTACGGCATGGGTTCCGAACATGCACTTTGTTGCACCGATT  648
     GlyArgSerLeuLysAlaAspAsnIleAlaProThrAlaTrpValProAsnMetHisPheValAlaProIle

649  AACGACCAATTTGGTTGGGGCGCTTCTATTACCTCTAACTATGGTCTGGCTACAGAGTTTAACGATACTTAT  720
     AsnAspGlnPheGlyTrpGlyAlaSerIleThrSerAsnTyrGlyLeuAlaThrGluPheAsnAspThrTyr

721  GCAGGCGGCTCTGTCGGGGGTACAACCGACCTTGAAACCATGAACCTGAACTTAAGCGGTGCGTATCGCTTA  792
     AlaGlyGlySerValGlyGlyThrThrAspLeuGluThrMetAsnLeuAsnLeuSerGlyAlaTyrArgLeu

793  AATAATGCATGGAGCTTTGGTCTTGGTTTCAACGCCGTCTACGTCCGCGCGAAAATTGAACGTTTCGCAGGC  864
     AsnAsnAlaTrpSerPheGlyLeuGlyPheAsnAlaValTyrValArgAlaLysIleGluArgPheAlaGly

865  GATCTGGGGCAGTTGGTTGCTGGCCAAATTATGCAATCTCCTGCTGGCCAAACTCAGCAAGGGCAAGCATTG  936
     AspLeuGlyGlnLeuValAlaGlyGlnIleMetGlnSerProAlaGlyGlnThrGlnGlnGlyGlnAlaLeu

937  GCAGCTACCGCCAACGGTATTGACAGTAATACCAAAATCGCTCATCTGAACGGTAACCAGTGGGGCTTTGGC  1008
     AlaAlaThrAlaAsnGlyIleAspSerAsnThrLysIleAlaHisLeuAsnGlyAsnGlnTrpGlyPheGly

1009 TGGAACGCCGGAATCCTGTATGAACTGGATAAAAATAACCGCTATGCACTGACCTACCGTTCTGAAGTGAAA  1080
     TrpAsnAlaGlyIleLeuTyrGluLeuAspLysAsnAsnArgTyrAlaLeuThrTyrArgSerGluValLys

1081 ATTGACTTCAAAGGTAACTACAGCAGCGATCTTAATCGTGCGTTTAATAACTACGGTTTGCCAATTCCTACC  1152
     IleAspPheLysGlyAsnTyrSerSerAspLeuAsnArgAlaPheAsnAsnTyrGlyLeuProIleProThr

1153 GCGACAGGTGGCCGAACGCAATCGGGTTATCTGACGCTGAACCTGCCTGAAATGTGGGAAGTGTCAGGTTAT  1224
     AlaThrGlyGlyArgThrGlnSerGlyTyrLeuThrLeuAsnLeuProGluMetTrpGluValSerGlyTyr

1225 AACCGTGTTGATCCACAGTGGGCGATTCACTATAGCCTGGCTTACACCAGCTGGAGTCAGTTCCAGCAGCTG  1296
     AsnArgValAspProGlnTrpAlaIleHisTyrSerLeuAlaTyrThrSerTrpSerGlnPheGlnGlnLeu

1297 AAAGCGACCTCAACCAGTGGCGACACGCTGTTCCAGAAACATGAAGGCTTTAAAGATGCTTACCGCATCGCG  1368
     LysAlaThrSerThrSerGlyAspThrLeuPheGlnLysHisGluGlyPheLysAspAlaTyrArgIleAla

1369 TTGGGTACCACTTATTACTACGATGATAACTGGACCTTCCGTACCGGTATCGCCTTTGATGACAGCCCAGTT  1440
     LeuGlyThrThrTyrTyrTyrAspAspAsnTrpThrPheArgThrGlyIleAlaPheAspAspSerProVal

1441 CCTGCACAGAATCGTTCTATCTCCATTCCGGACCAGGACCGTTTCTGGCTGAGTGCAGGTACGACTTACGCA  1512
     ProAlaGlnAsnArgSerIleSerIleProAspGlnAspArgPheTrpLeuSerAlaGlyThrThrTyrAla

1513 TTTAATAAAGATGCTTCAGTCGACGTTGGTGTTTCTTATATGCACGGTCAGAGCGTGAAAATTAACGAAGGC  1584
     PheAsnLysAspAlaSerValAspValGlyValSerTyrMetHisGlyGlnSerValLysIleAsnGluGly

1585 CCATACCAGTTCGAGTCTGAAGGTAAAGCCTGGCTGTTCGGTACTAACTTTAACTACGCGTTCTGATAACGC  1656
     ProTyrGlnPheGluSerGluGlyLysAlaTrpLeuPheGlyThrAsnPheAsnTyrAlaPhe

1657 GTTCGCCTGGATAAAGTCACCTGCATAGCAGGTGCTTTAACTCCCCCACTTCACCGAAAGTAGTGTCCTCAT  1728

1729 TGCTTACCTCCTGAGTTTTGCAAACAGCCTGTTGGCAGCTTGCCCTTTCAAATCAATAAGCGGTATCCACAG  1800

1801 ACAAACCTGAAGGGAAAGGCATATTTTCAGGCGTTCTGCTCGTCCTTCTCAAAGAGTTTACTTTTCTGCATT  1872

1873 TCCAGGATACTCCCCCTGGCTATTGTGCGCTCATACACTCAAATTAAAGATAGGTTCTAAATAAATGAGCGT  1944

1945 TTTTTGATAGTCTATTTCATTAGGTAATATATATTTTGTAACAAATCAATCAAAATGGAATAAAATCATGCT  2016

2017 ACCATCTATTTCAATCAACAATACCAGCGCAGCTTACCCAGAATCCATCAATGAAAATAACAATGATGAAGT  2088

2089 TAATGGATTAGTACAAGAGTTCAAAAACCTTTTTAATGGTAAGGAAGGAATAAGCACCTGTATTAAACATCT  2160

2161 ACTTGAGCTTATAAAAAACGCCATACGAGTAAACGAC  2197
```

FIG. 2. Nucleotide sequence of the *fadL* gene; +1 refers to the transcriptional start defined by primer extension, and −10 and −35 regions are noted as are regions with hyphenated dyad symmetry (convergent arrows) and the proposed ribosome binding site (double underline). The amino acid sequence of pre-FadL is given below the DNA sequence. The underlined amino acid sequence represents the N-terminal sequence of FadL determined by using automated protein sequencing.

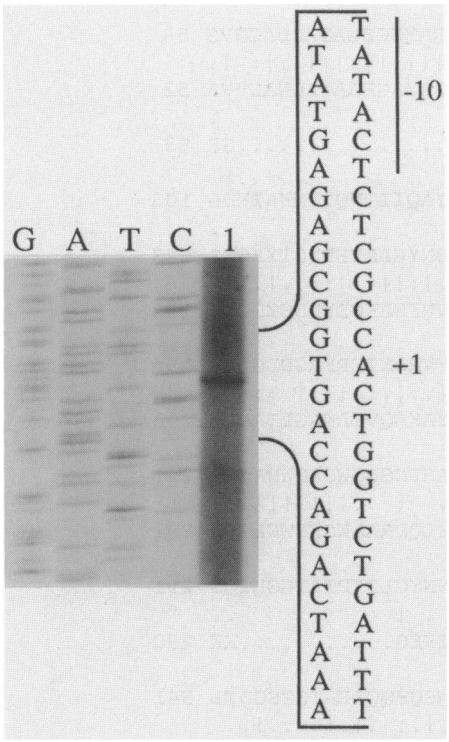


FIG. 3. Primer extension of the 5′ end-labeled *fadL*-specific oligonucleotide 5′-ACCGGACCGGCCCGAATAATCCCC-3′. Lanes G, A, T, and C represent the *fadL* sequence with the indicated oligonucleotide and the *fadL*$^+$ plasmid pN103; lane 1 represents the primer extended fragment. The sequence to the right indicates the transcriptional start (+1) and the −10 (−10) sequences.

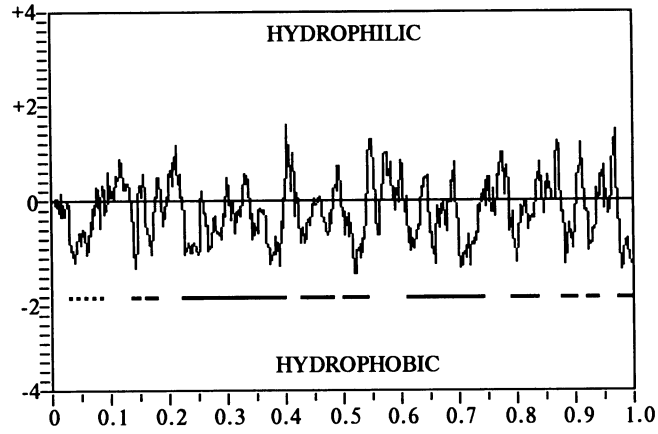


FIG. 4. Hexapeptide profile of pre-FadL. Hydrophilicity was determined using the algorithms of Hopp and Woods (12) and an averaging length of 6 amino acid residues. The area defined by the dashed line represents the hydrophobic region of the signal peptide, and the solid line represents regions of FadL with an overall hydrophobic character.

ing homology between FadL and P1 (Fig. 5). FadL and P1 had 61.5% similarity with 42.0% perfect identity over 438 amino acid residues.

DISCUSSION

The FadL protein of *E. coli* plays a central role in the uptake of exogenous long-chain fatty acids. As an outer membrane protein, FadL acts to bind long-chain fatty acids with a relatively high affinity (2) and by some unknown mechanism allows these hydrophobic molecules to traverse this membrane layer. After transport across the outer membrane, long-chain fatty acids traverse the periplasmic space and inner membrane by an undefined process and are delivered to fatty acyl coenzyme A synthetase or acyl-acyl carrier protein synthetase (10, 13, 15, 30). The majority of exogenous long-chain fatty acids are derivatized to coenzyme A thioesters and are broken down by cyclic β-oxidation (10, 15). A small proportion of exogenous long-chain fatty acids enter the phospholipid pool directly by acyl-acyl carrier protein synthetase (30). In either case, FadL is required. We envision FadL acting as a common component for the uptake of exogenous long-chain fatty acids into these two metabolic systems.

In the present study, we determined the nucleotide sequence of the *fadL* gene. Our sequence analysis demonstrated that *fadL* is contained within a 2,197-bp fragment and encodes a protein with a molecular weight of 48,831. Based on the DNA sequence of *fadL*, two possible translational starts for FadL, separated by the sequence GTC, were identified. We cannot say with certainty which ATG represents the translational start. Total amino acid compositional analysis of FadL deduced from the DNA sequence of the *fadL* gene demonstrated that this protein was rich in hydrophobic residues and lacked cysteinyl residues. Both findings were in agreement with our earlier work describing the amino acid composition of purified FadL (4). The predicted amino acid sequence of FadL suggested the presence of a signal peptide of 27 amino acids in length. In an effort to confirm the presence of a signal peptide and the FadL reading frame, the N-terminal amino acid sequence of FadL was determined after partial purification from a strain harboring a T7-*fadL* expression system. These data demonstrated that a pre-FadL was processed giving rise to a mature protein with a molecular weight of 45,969. The processing site of pre-FadL was in agreement with the consensus cleavage site define for signal peptidase I (29).

The transcriptional start site of *fadL* was defined at 95 bp upstream from the translational start of FadL. The −10 and −35 regions were identified by comparison with consensus sequences for RNA polymerase and sigma 70. In addition, three regions of hyphenated dyad symmetry (one upstream from the −10 and −35 sequences and two downstream from the transcriptional start) were identified that may play a role in the regulation of this gene. Some similarities exist between these dyads and those defined for the *fadBA* genes (encoding the β-oxidation multienzyme complex) (9) and thus may represent potential binding sites for the FadR repressor.

The sequence of the *fadL* gene presented in this paper is different from that published by Said et al. (31). In the course of our studies defining a series of mutants within *fadL* at the level of DNA sequence (16), we found that Said et al. had (i) incorrectly positioned the promoter of *fadL*, (ii) incorrectly defined the transcriptional start of *fadL*, (iii) incorrectly defined the translational start of FadL, (iv) omitted a number of base pairs (five) in the coding sequence for the amino terminus of FadL, and (v) included several extra base pairs in the noncoding and coding regions of *fadL*. The deficien-

```
          .         .         .         .         .
  9 FTKSALAVAVALISTQAWSAGFQLNEFSSSGLGRAYSGEGAIADDAGNVS 58
    |..| ||.|: | ...| .|:|||.|.|.|||||||.||:||||:|: |.
  4 FNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVA 53
          .         .         .         .         .
 59 RNPALITMFDRPTFSAGAVYIDPDVNISGTSPS.............GR 93
     ||||:.:|. : ||.|:||||. :|:.|. .|              :
 54 TNPALMSLFKTAQFSTGGVYIDSRINMNGDVTSYAQIITNQIGMKAIKDG 103
          .         .         .         .         .
 94 SLKADNIAPTAWVPNMHFVAPINDQFGWGASITSNYGLATEFNDTYAGGS 143
    | .. |:.|.|:|||:.||||:||.|::||::. |:|| .|::|.|.:|
104 SASQRNVVPGAFVPNLYFVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGV 153
          .         .         .         .         .
144 VGGTTDLETMNLNLSGAYRLNNAWSFGLGFNAVYVRAKIERFAGDLGQLV 193
    .||.|||..:|||||||||:.:::|:|||.||||.:|.:|| || ::: |
154 FGGKTDLSAINLNLSGAYRVTEGLSLGLGVNAVYAKAQVERNAGLIADSV 203
          .         .         .         .         .
194 AGQIMQSPAGQTQQGQALAATANGIDSNTK..IAHLNGNQWGFGWNAGIL 241
     :.  | ..: . |.::: .  . :.|..|  :.  : ..|||||||::
204 KDN..QITSALSTQQEPFRDLKKYLPSKDKSVVSLQDRAAWGFGWNAGVM 251
          .         .         .         .         .
242 YELDKNNRYALTYRSEVKIDFKGNYSSDLNRAFNNYGLPIPTATGGRTQS 291
    |:::..||.:|.|:|.|.||| :. ...|: .. . |           ..
252 YQFNEANRIGLAYHSKVDIDFADRTATSLEANVIKEG...........KK 290
          .         .         .         .         .
292 GYLTLNLPEMWEVSGYNRVDPQWAIHYSLAYTSWSQFQQLKATSTSGDTL 341
    | ||:.||: :|:||::.:...:|:|||. || ||.: .|.|. ..|..
291 GNLTFTLPDYLELSGFHQLTDKLAVHYSYKYTHWSRLTKLHASFEDGKKA 340
          .         .         .         .         .
342 FQKHEGFKDAYRIALGTTYYYDDNWTFRTGIAFDDSPVPAQNRSISIPDQ 391
    |:|. .:.:. |:|||..| . :.:|:|.|||:|:.: . ::|| .|||
341 FDKELQYSNNSRVALGASYNLYEKLTLRAGIAYDQAASR.HHRSAAIPDT 389
          .         .         .         .         .
392 DRFWLSAGTTYAFNKDASVDVGVSYMHGQSVKINE.............GP 428
    || |.| |.|| |..: |||:| .|:.|..|.:.|           ..
390 DRTWYSLGATYKFTPNLSVDLGYAYLKGKKVHFKEVKTIGDKRTLTLNTT 439
          .         .
429 YQFESEGKAWLFGTNFNYAF 448
     .:.|::.| |:| |:||.|
440 ANYTSQAHANLYGLNLNYSF 459
```

FIG. 5. Amino acid sequence homology between FadL and *H. influenzae* type b P1. Solid lines indicate complete amino acid identity. The top sequence represents pre-FadL, and the bottom sequence represents *H. influenzae* type b P1. The numbers to the left and right of the sequences refer to amino acid residue numbers. The GenBank accession number for the *H. influenzae* type b P1 gene is JO3381 (23).

cies found in the work of Said et al. (31) reflect errors in DNA sequencing and S1 protection analysis and not in the identification of the *fadL* clone. In the present work (like that of Said et al. [31]), we sequenced the 2.8-kb *Eco*RV fragment of genomic DNA previously shown to encode the entire *fadL* gene (3). We defined the start of transcription for the *fadL* gene by using an *fadL*-specific oligonucleotide and primer extension with reverse transcriptase. This transcriptional start is nearly 280 bp upstream from that defined by Said et al. (31). The S1 nuclease protection data presented by Said et al. are difficult to interpret due to a smearing in the lane containing the S1 nuclease-protected fragment and an inability to decipher the Maxam and Gilbert sequence. Our data defining the transcriptional start were clear and well defined. The size of the primer extended fragment was determined by using both 5′ end-labeled ϕX174 *Hae*II-restricted DNA as size markers and sequencing reactions with the same oligonucleotide and the *fadL*$^+$ plasmid pN103. Based on the data presented in this paper, the transcriptional start site proposed by Said et al. (31) codes for the leucine and isoleucine residues at amino acid positions −7 and −8 in the signal peptide. Based on two criteria, the translational start of FadL proposed by Said et al. also appears to be incorrect. First, the translational start defined by our data would place the translational start proposed by Said et al. 39 amino acid residues downstream from the sequence we defined as the signal peptidase I processing site. Second, Said et al. claim that the amino acid sequence they presented predicts a signal sequence. Although their proposed signal sequence contains a basic region followed by a small hydrophobic region, the presence of two aspartic acid residues and one asparagine residue within this hydrophobic region results in a considerable deviation from consensus. The DNA sequence data of *fadL*, the primer extension data using a *fadL*-specific oligonucleotide, and the N-terminal sequence data of FadL determined in our laboratory support the conclusion that the transcriptional start is 95 bp upstream from the translational start. The translational start site is preceded by a polypurine-

rich segment that has homology with known ribosome binding sites (34). Based on our data, we feel confident that the transcriptional and translational starts proposed in this paper for the *fadL* gene are correct.

FadL, like a number of proteins found in the outer membrane of gram-negative bacteria, is heat modifiable in the presence of sodium dodecyl sulfate. Using the amino acid sequence comparison program BESTFIT supplied by GCG, we found that FadL and the heat-modifiable outer membrane protein P1 of *H. influenzae* type b had remarkable similarity (61.5% similarity with 42.0% perfect identity) over 438 amino acid residues. P1 is not required for pathogenicity of *H. influenzae* type b (11), although antibodies raised against this protein are protective against bacteremia (17). Despite the work done on this *H. influenzae* type b protein, the precise physiological role of P1 remains largely undefined. The similarity between FadL and *H. influenzae* type b P1 suggests that these proteins may have a common function, perhaps in long-chain fatty acid transport. The expression of *fadL* is under the control of the FadR repressor and is induced in the presence of long-chain fatty acids (32). By analogy, P1 may be subject to similar control, and its expression may be induced by long-chain fatty acids (like FadL) or some other metabolically useful compound. The similarity between FadL, OmpA, and phospholipase were also compared, because these two latter outer membrane proteins of *E. coli* are also heat-modifiable, no significant homology was found. In addition, the amino acid sequences of FadL and OmpF were compared, since both proteins serve as receptors for the bacteriophage T2. No regions of significant homology between these proteins were identified, suggesting that the domains that give rise to T2 receptor activity are different of the level of amino acid sequence.

FadL is a central component of the long-chain fatty acid transport machinery in *E. coli*. This protein binds exogenous long-chain fatty acids with a relatively high affinity (2) and by some unknown mechanism allows these compounds to traverse the outer membrane. The proposed amino acid sequence of FadL suggests that this protein contains several stretches of hydrophobic residues. Although the roles of these regions are unknown, hexameric linker mutagenesis of the *fadL* gene suggests that one of these regions is involved in the heat-modifiable characteristic of FadL, whereas another is involved in the actual binding of long-chain fatty acids (16). With the sequence of *fadL* defined and amino acid sequence of FadL determined, we are now using random and directed mutagenesis to investigate the functional and phenotypic properties of this protein further: long-chain fatty acid binding, long-chain fatty acid transport, bacteriophage T2 binding, and heat modifiability.

ACKNOWLEDGMENTS

I thank William Lane of the Harvard University Microchemistry facility for the N-terminal sequence analysis of FadL and Concetta DiRusso for synthesizing the *fadL*-specific oligonucleotides and for providing plasmid pCD130.

This work was supported by grant DCB-8811714 from the National Science Foundation.

REFERENCES

1. **Black, P. N.** 1988. The *fadL* gene product of *Escherichia coli* is an outer membrane protein required for uptake of long-chain fatty acids and involved in sensitivity to bacteriophage T2. J. Bacteriol. **170**:2850–2854.
2. **Black, P. N.** 1990. Characterization of FadL-specific fatty acid binding in *Escherichia coli*. Biochim. Biophys. Acta **1046**:97–105.
3. **Black, P. N., S. K. Kianian, C. C. DiRusso, and W. D. Nunn.** 1985. Long-chain fatty acid transport in *Escherichia coli*: cloning, mapping and expression of the *fadL* gene. J. Biol. Chem. **260**:1780–1789.
4. **Black, P. N., B. Said, C. R. Ghosn, J. V. Beach, and W. D. Nunn.** 1987. Purification and characterization of an outer membrane-bound protein involved in long-chain fatty acid transport in *Escherichia coli*. J. Biol. Chem. **262**:1412–1419.
5. **Daghert, M., and S. D. Ehrlich.** 1979. Prolonged incubation with calcium chloride improves the competence of *Escherichia coli* cells. Gene **6**:23–28.
6. **deCrombrugghe, B., S. Busby, and H. Buc.** 1984. Cyclic AMP receptor protein's role in transcription activation. Science **224**: 831–838.
7. **deGeus, P., I. van Die, H. Bergmans, J. Tommassen, and G. deHaas.** 1983. Molecular cloning of *pldA*, the structural gene for outer membrane phospholipase in *E. coli* K-12. Mol. Gen. Genet. **190**:150–155.
8. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the vax. Nucleic Acids Res. **12**:387–395.
9. **DiRusso, C. C.** 1990. Primary sequence of the *Escherichia coli fadBA* operon, encoding the fatty acid-oxidizing multienzyme complex, indicates a high degree of homology to eucaryotic enzymes. J. Bacteriol. **172**:6459–6468.
10. **Frerman, F. E., and W. Bennett.** 1973. Studies on the uptake of fatty acids by *Escherichia coli*. Arch. Biochem. Biophys. **159**: 434–443.
11. **Hanson, M. S., L. D. Cope, and E. J. Hansen.** 1989. Expression of the heat-modifiable major outer membrane protein of *Haemophilus influenzae* type b is unrelated to virulence. Infect. Immun. **57**:1639–1646.
12. **Hopp, T. P., and K. R. Woods.** 1981. Predictions of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78**:3824–3828.
13. **Kameda, K., and W. D. Nunn.** 1981. Purification and characterization of acyl CoA synthetase from *Escherichia coli*. J. Biol. Chem. **256**:5702–5707.
14. **Kassavetis, G. A., and E. P. Geiduschek.** 1982. Bacteriophage T4 late promoters: mapping 5′ ends of T4 gene 23 mRNAs. EMBO J. **1**:107–114.
15. **Klein, K., R. Steinberg, B. Feithan, and P. Overath.** 1971. Fatty acid degradation in *Escherichia coli*: an inducible system for the uptake of fatty acids and further characterization of *old* mutants. Eur. J. Biochem. **19**:442–450.
16. **Kumar, G. B., and P. N. Black.** J. Biol. Chem., in press.
17. **Loeb, M. R.** 1987. Protection of infant rats from *Haemophilus influenzae* type b infection by antiserum to purified outer membrane protein a. Infect. Immun. **55**:2612–2618.
18. **Maloy, S. R., C. L. Ginsburgh, R. W. Simons, and W. D. Nunn.** 1981. Transport of long and medium chain fatty acids by *Escherichia coli* K12. J. Biol. Chem. **256**:3735–3742.
19. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1981. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Matsudaira, P.** 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J. Biol. Chem. **262**:10035–10038.
21. **Messing, J. H.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
22. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Munson, R., Jr., and S. Grass.** 1988. Purification, cloning and sequence of outer membrane protein P1 of *Haemophilus influenzae* type b. Infect. Immun. **56**:2235–2242.
24. **Nakamura, K., and S. Mizushima.** 1976. Effects of heating in dodecyl sulfate solution on the conformation and electrophoretic mobility of isolated outer membrane proteins from *Escherichia coli* K-12. J. Biochem. **80**:1411–1422.
25. **Nunn, W. D.** 1986. A molecular view of fatty acid catabolism in *Escherichia coli*. Microbiol. Rev. **50**:179–192.
26. **Nunn, W. D., R. W. Colburn, and P. N. Black.** 1986. Transport of long-chain fatty acids in *Escherichia coli*: evidence for role of

fadL gene product as long-chain fatty acid receptor. J. Biol. Chem. **261**:167–171.

27. **Nunn, W. D., and R. W. Simons.** 1978. Transport of long-chain fatty acids by *Escherichia coli*: mapping and characterization of mutants in the *fadL* gene. Proc. Natl. Acad. Sci. USA **75**:3377–3381.
28. **Nunn, W. D., R. W. Simons, P. A. Egan, and S. R. Maloy.** 1979. Kinetics of the utilization of medium and long-chain fatty acids by a mutant of *Escherichia coli* defective in the *fadL* gene. J. Biol. Chem. **254**:9130–9134.
29. **Oliver, D.** 1985. Protein secretion in *Escherichia coli*. Annu. Rev. Microbiol. **39**:615–648.
30. **Rock, C. O., and S. Jackowski.** 1985. Pathways of the incorporation of exogenous fatty acids into phosphatidylethanolamine in *Escherichia coli*. J. Biol. Chem. **260**:12720–12724.
31. **Said, B., C. R. Ghosn, L. Vu, and W. D. Nunn.** 1988. Nucleotide sequencing and expression of the *fadL* gene involved in long-chain fatty acid transport in *Escherichia coli*. Mol. Microbiol. **2**:363–370.
32. **Sallus, L., R. J. Haselbeck, and W. D. Nunn.** 1985. Regulation of fatty acid transport in *Escherichia coli*: analysis by operon fusion. J. Bacteriol. **155**:1450–1454.
33. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. USA **74**:5463–5467.
34. **Shine, J., and L. Dalgarno.** 1975. Determinants of cistron specificity in bacterial ribosomes. Nature (London) **254**:34–38.
35. **Studier, W. F., and B. A. Moffat.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. **189**:113–130.