JOURNAL OF BACTERIOLOGY, Mar. 1988, p. 1268–1274
0021-9193/88/031268-07$02.00/0
Copyright © 1988, American Society for Microbiology
Vol. 170, No. 3

First Committed Step of Lipid A Biosynthesis in *Escherichia coli*: Sequence of the *lpxA* Gene

JACK COLEMAN AND CHRISTIAN R. H. RAETZ*

Department of Biochemistry, College of Agricultural and Life Sciences, University of Wisconsin-Madison, Madison, Wisconsin 53706

Received 1 October 1987/Accepted 1 December 1987

The min 4 region of the *Escherichia coli* genome contains genes (*lpxA* and *lpxB*) that encode proteins involved in lipid A biosynthesis. We have determined the sequence of 1,350 base pairs of DNA upstream of the *lpxB* gene. This fragment of DNA contains the complete coding sequence for the 28.0-kilodalton *lpxA* gene product and an upstream open reading frame capable of encoding a 17-kilodalton protein (ORF_{17}). In addition there appears to be an additional open reading frame ($ORF_?$) immediately upstream of ORF_{17}. The initiation codon for *lpxA* is a GUG codon, and the start codon for ORF_{17} is apparently a UUG codon. The start and stop codons overlap between $ORF_?$ and ORF_{17}, ORF_{17} and *lpxA*, and *lpxA* and *lpxB*. This overlap is suggestive of translational coupling and argues that the genes are cotranscribed. Crowell et al. (D. N. Crowell, W. S. Reznikoff, and C. R. H. Raetz, J. Bacteriol. 169:5727–5734, 1987) and Tomasiewicz and McHenry (H. G. Tomasiewicz and C. S. McHenry, J. Bacteriol. 169:5735–5744, 1987) have demonstrated that there are three similarly overlapping coding regions downstream of *lpxB* including *dnaE*, suggesting the existence of a complex operon of at least seven genes: 5′-$ORF_?$-ORF_{17}-*lpxA*-*lpxB*-ORF_{23}-*dnaE*-ORF_{37}-3′.

Escherichia coli cells are surrounded by two distinct lipid bilayers, the inner membrane and the outer membrane. The two monolayers of the outer membrane differ in their lipid composition (30). The inner monolayer consists primarily of glycerophospholipids (30), as does the inner membrane, but the outer monolayer of the outer membrane is composed primarily of lipopolysaccharide (28). The hydrophobic anchor for lipopolysaccharide is lipid A, a glucosamine disaccharide substituted with phosphate at positions 1 and 4′ (30) and with β-hydroxymyristoyl moieties at the 2, 3, 2′, and 3′ positions (30). Lipid A is responsible for the toxic and immunostimulatory properties of lipopolysaccharide (28, 30). The *lpxA* gene codes for the enzyme responsible for the first committed step in lipid A biosynthesis, UDP-*N*-acetylglucosamine (UDP-GlcNAc) acyltransferase (30). This enzyme catalyzes the transfer of a β-hydroxymyristoyl moiety from *R*-3-hydroxymyristoyl acyl carrier protein to the 3 position of the glucosamine ring of UDP-GlcNAc. The fifth step in the biosynthesis of lipid A is catalyzed by lipid A disaccharide synthase, encoded by *lpxB* (30). These two genes, *lpxA* and *lpxB*, have previously been shown to be adjacent to each other in the min 4 region of the chromosome (7).

The DNA sequence from *lpxB* through *dnaE* indicated that these genes may be part of a complex operon with many overlapping genes (7a, 38). Between *lpxB* and *dnaE* there is an open reading frame coding for a 23-kilodalton (kDa) polypeptide. Similarly, after *dnaE* there is another open reading frame able to code for a 37-kDa protein. All of these genes have overlapping or closely positioned stop and start codons, indicative of cotranslation. Crowell et al. (7a) have suggested that *lpxA* is also translationally coupled in this operon.

The *firA* gene also maps at the min 4 region on the chromosome. Temperature-sensitive mutations in *firA* render RNA synthesis thermosensitive and eliminate high-level resistance to rifampin (13, 18). Lathe and Lecocq (19) have shown that a 7.2-kilobase (kb) *Hin*dIII DNA fragment from the chromosome will correct a *firA* mutation. This same fragment will also complement some *dnaE* mutant alleles in *recA*$^+$ strains. Both *lpxA* and *lpxB* are also on this 7.2-kb *Hin*dIII fragment of DNA immediately upstream (counterclockwise) of *dnaE* (7, 25, 33). Lathe and Lecocq (20) have shown that this 7.2-kb *Hin*dIII DNA fragment, when cloned in either orientation in bacteriophage lambda, will direct synthesis of both a 27-kDa protein and a 17-kDa protein (*firA*) protein (18). The level of expression of both of these proteins is higher when inserted in one orientation as opposed to the other orientation, due to expression from a vector promoter (19, 20). In addition, if the synthesis of the 17-kDa protein is reduced by spontaneous mutation of the λ-*firA* hybrid, synthesis of the 27-kDa protein is also reduced in all cases examined. This indicates that the coding region of the two proteins are in close proximity. Crowell et al. (7) have shown using minicells that the *lpxA* product (UDP-GlcNAc acyltransferase) is a 27-kDa protein, possibly the same as that coexpressed with *firA*.

We have now determined the complete nucleotide sequence of *lpxA* and an upstream open reading frame, and we have identified the initiation codon of *lpxA* as a GUG. Amino-terminal sequence analysis of an *lpxA*–β-galactosidase fusion protein confirmed this assignment. However, amino-terminal sequence analysis of purified FirA protein indicates that the open reading frame that encodes the 17-kDa protein (ORF_{17}) does not code for FirA. Sequence analysis described here and elsewhere (7a, 38) indicates that there are at least seven overlapping genes. Presumably these genes are transcriptionally and/or translationally coupled.

MATERIALS AND METHODS

Strains and media. All strains used are listed in Table 1. Cells were grown in LB or YT medium (24) containing 100 μg of ampicillin per ml, 150 μg of rifampin per ml, 40 μg of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside per ml, 30 μg of tetracycline per ml, or 50 μg of diaminopimelic acid per ml where appropriate.

* Corresponding author.

TABLE 1. Strains of *E. coli* K-12

Strain	Genotype	Reference or source
MF6	*thr-1 dapD4 his-4 rpsL136 leuB6 eda thi-1 lacY1 Δgal ara-14 xyl-15 mtl-1 tonA supE*	10
RX800	MF6 *dapD*$^+$ *lpxA*	This study
RX805	MF6 *dapD*$^+$	This study
RX801	RX800 *srlA*::Tn*10 recA56*	This study
MF6R	MF6 *rpoB* (spontaneous high-level rifampin-resistant mutant of MF6)	This study
JCR20	MF6R *dapD*$^+$ *lpxA*	This study
JCR21	JCR20 *srlA*::Tn*10 recA56*	This study
DC1	*thr-1 his-4 rpsL136 leuB6 eda thi-1 lacY1 lpxB1 pgsA srlA*::Tn*10 recA56 Δgal-1 ara-14 xyl-15 mtl-1 tonA supE*	7
JC10240	Hfr PO45 *recA56 srlA*::Tn*10 thr-300 ilv-318 rpsE300*	CGSC[a]
AT982	*dapD4 thi-1 relA* Hfr K46	4
RX8a	AT982 *dapD*$^+$ *lpxA*	This study
JW353	*thr-1 leuB6 zae*::Tn*10 thyA6 met-89 thi-1 deoC1 lacY1 rpsL67 tonA21 supE44*	CGSC
SB4288	*recA1 thi-1 relA mal-24 rpsE supE50 DE5(Δlac proB)*	11
JM103	*Δlac-pro thi rpsL endA sbcB15 hsdR4 supE* F′ (*traD36 proAB lacI*^{q} *ZΔM15*)	39
GM33	*dam-3*	CGSG
RL5	*firA200* λ *c*$^+$	19

[a] *E. coli* Genetic Stock Center, Yale University, New Haven, Conn.

Materials. Restriction enzymes were purchased from Bethesda Research Laboratories, Gaithersburg, Md., or from New England Biolabs, Beverly, Mass. The Klenow fragment of DNA polymerase I was purchased from Boehringer-Mannheim Biochemicals, Indianapolis, Ind. T4 DNA ligase was purchased from International Biotechnologies, Inc., New Haven, Conn. [α^{32}P]dCTP for sequencing was purchased from Amersham Corp., Arlington Heights, Ill. Kodak XAR film was used for autoradiography. Nucleotides used in sequencing were purchased from P-L Biochemicals, Piscataway, N.J.

DNA manipulations and transformation. All DNA manipulations and cell transformations were done as described by Maniatis et al. (22). DNA fragments used for DNA sequencing and cloning were prepared by electroelution into 7.5 M ammonium acetate with an apparatus purchased from International Biotechnologies, Inc. DNA sequencing was performed by the dideoxy method of Sanger et al. (32) with [α-^{32}P]dCTP. Sequencing reactions were conducted at 42°C.

Enzyme assays. UDP-GlcNAc acyltransferase activity was assayed as described previously (2).

Computer programs. Computer programs used in analyzing nucleotide and amino acid sequences were provided by the University of Wisconsin Genetics Computer group (8, 12).

Protein purification and amino-terminal sequencing. The UDP-GlcNAc acyltransferase–β-galactosidase protein hybrid was purified from a culture of SB4288 containing pSR1-β-gal induced for 3 h with 0.5% arabinose as described previously (7) with a monoclonal anti-β-galactosidase immunoaffinity column (a gift from Promega Biotech, Madison, Wis.), followed by purification on a preparative sodium dodecyl sulfate-polyacrylamide (7.5%) gel (17).

Amino-terminal sequence analysis of UDP-GlcNAc acyltransferase was accomplished by using automated Edman degradation at the University of Wisconsin Biotechnology Center with an Applied Biosystems 470A protein sequencer.

FirA was purified as described previously (18) and then separated on a preparative sodium dodecyl sulfate-polyacrylamide (17.5%) gel (1). The gel was briefly stained in 45% methanol–10% acetic acid–0.5% Coomassie brilliant blue G. The FirA band was excised and soaked in 0.15 M ammonium acetate for 4 h, followed by an overnight soak in distilled water at room temperature (21°C). The water and ammonium acetate fractions were pooled and precipitated on ice for 2 h with 20% trichloroacetic acid, followed by centrifugation in a microfuge for 5 min. The pellet was washed with acetone and then suspended in distilled water. Again the protein was precipitated with 20% trichloroacetic acid, followed by five washes with acetone. Amino-terminal sequence analysis of FirA was accomplished manually as described by Tarr (37) at the University of Michigan Protein Sequencing Facility.

Strain construction. P1 transduction and Hfr-mediated gene transfer were performed as described previously (24, 35).

RESULTS

Nucleotide sequence upstream of *lpxB*. It has recently been reported that the *lpxA* gene of *E. coli* is located immediately upstream of *lpxB* (7). The region of the *E. coli* chromosome containing the *lpxA* gene was sequenced (Fig. 1). The previously described *lpxA*$^+$ *lpxB*$^+$ plasmid pDC4 (7) was the source of DNA for all fragments sequenced downstream of the *Pst*I site (Fig. 1). The lambda clone λ*fir24* (a kind gift from M. Nomura [19]) was used as the source of DNA for all fragments sequenced upstream and across the *Pst*I site. Fragments to be sequenced were inserted into M13mp18 or M13mp19 (Bethesda Research Laboratories) or pTZ18U or pTZ19U (United States Biochemical Corp.).

Sequence analysis. We sequenced 1,350 nucleotides up-

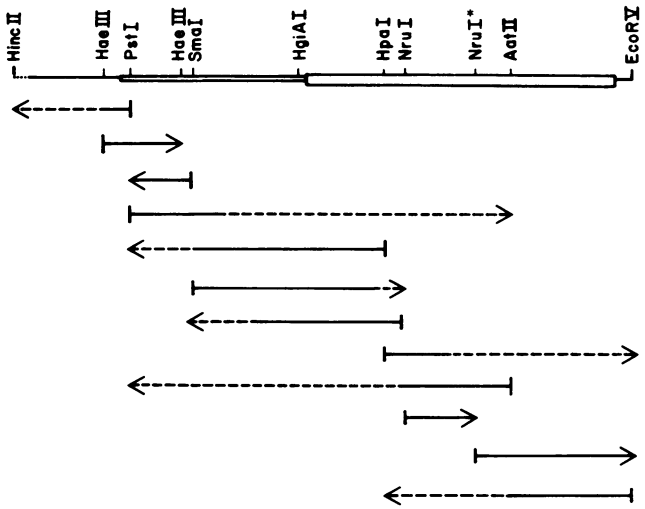


FIG. 1. Sequencing strategy. Lines with arrows represent sequenced fragments. Arrows show direction of sequencing. Dashed lines indicate regions of sequenced fragments yielding ambiguous sequence. Relevant restriction enzyme recognition sites are shown at the top. An asterisk indicates a methylated restriction site not cleaved on DNA isolated from *dam*$^+$ *E. coli*. The open reading frames corresponding to *lpxA* and ORF_{17} are indicated by boxes of the appropriate size.

```
   1 GGCCGTGTTATTATTGTCGTTTCTTATATTTTGACAGGAAGAGTATCTTGACTACTAACACTCATACTCTGCAGATTGAAGAGATTTTAGAACTTC  96
     GlyArgValIleIleValValSerTyrIleLeuThrGlyArgValSer *
                                          ORF17- MetThrThrAsnThrHisThrLeuGlnIleGluGluIleLeuGluLeuLeu

  97 TGCCGCACCGTTTCCCGTTCTTACTGGTGGATCGCGTGCTGGATTTTGAAGAAGGTCGTTTTCTGCGCGCAGTAAAAAATGTCTCTGTCAATGAGC 192
       ProHisArgPheProPheLeuLeuValAspArgValLeuAspPheGluGluGlyArgPheLeuArgAlaValLysAsnValSerValAsnGluPro

 193 CATTCTTCCAGGGCCATTTCCCTGGAAAACCGATTTTCCCGGGTGTGCTGATTCTGGAAGCAATGGCACAGGCAACAGGTATTCTGGCGTTTAAAA 288
       PhePheGlnGlyHisPheProGlyLysProIlePheProGlyValLeuIleLeuGluAlaMetAlaGlnAlaThrGlyIleLeuAlaPheLysSer

 289 GCGTAGGAAAACTGGAACCGGGTGAGCTGTACTACTTCGCTGGTATTGACGAAGCGCGCTTCAAGCGCCCGGTCGTGCCTGGCGATCAAATGATCA 384
       ValGlyLysLeuGluProGlyGluLeuTyrTyrPheAlaGlyIleAspGluAlaArgPheLysArgProValValProGlyAspGlnMetIleMet

 385 TGGAAGTCACTTTCGAAAAAACGCGCCGCGGCCTGACCCGTTTTAAAGGGGTTGCTCTGGTCGATGGTAAAGTAGTTTGCGACAGAACGATGATGT 480
       GluValThrPheGluLysThrArgArgGlyLeuThrArgPheLysGlyValAlaLeuValAspGlyLysValValCysAspArgThrMetMetCys

 481 GTGCTCGTAGCCGGAGGCGTGATACGTGATTGATAAATCCGCCTTTGTGCATCCAACCGCCATTGTGGAAGAGGGCGCGTCAATTGGCGCGAACGC 576
       AlaArgSerArgArgArgAspThr *
                             lpxA- MetIleAspLysSerAlaPheValHisProThrAlaIleValGluGluGlyAlaSerIleGlyAlaAsnAla

 577 ACACATTGGTCCTTTTTGTATCGTTGGACCCCATGTCGAAATTGGTGAGGGTACCGTACTGAAATCTCACGTTGTCGTGAATGGTCATACTAAAAT 672
       HisIleGlyProPheCysIleValGlyProHisValGluIleGlyGluGlyThrValLeuLysSerHisValValValAsnGlyHisThrLysIle

 673 TGGCCGCGATAATGAGATTTATTCAGTTGCCTCCATCGGCGAAGTTAACCAGGATCTGAAATATGCTGGCGAACCGACCCGTGTGGAAATCGGCGA 768
       GlyArgAspAsnGluIleTyrSerValAlaSerIleGlyGluValAsnGlnAspLeuLysTyrAlaGlyGluProThrArgValGluIleGlyAsp

 769 TCGTAACCGCATTCGCGAAAGCGTCACCATTCATCGTGGCACAGTCCAGGGCGGTGGATTGACGAAGGTGGGCAGCGACAACTTACTGATGATCAA 864
       ArgAsnArgIleArgGluSerValThrIleHisArgGlyThrValGlnGlyGlyGlyLeuThrLysValGlySerAspAsnLeuLeuMetIleAsn

 865 CGCGCACATTGCCGACGATTGTACGGTAGGTAACCGCTGTATTCTCGCCAACAACGCAACGCTGGCGGGTCACGTATCGGTTGACGACTTCGCGAT 960
       AlaHisIleAlaAspAspCysThrValGlyAsnArgCysIleLeuAlaAsnAsnAlaThrLeuAlaGlyHisValSerValAspAspPheAlaIle

 961 CATCGGCGGCATGACCGCAGTCCATCAGTTCTGCATCATTGGTGCGCACGTGATGGTTGGCGGCTGCTCCGGTGTGGCGCAGGACGTCCCTCCTTA 1056
       IleGlyGlyMetThrAlaValHisGlnPheCysIleIleGlyAlaHisValMetValGlyGlyCysSerGlyValAlaGlnAspValProProTyr

1057 TGTCATTGCGCAGGGTAACCACGCAACGCCGTTCGGTGTCAATATCGAAGGGCTGAAGCGCCGCGGATTCAGCCGTGAGGCGATTACCGCTATCCG 1152
       ValIleAlaGlnGlyAsnHisAlaThrProPheGlyValAsnIleGluGlyLeuLysArgArgGlyPheSerArgGluAlaIleThrAlaIleArg

1153 CAATGCGTATAAGCTGATTTATCGTAGCGGTAAAACGCTCGATGAAGTGAAACCGGAAATTGCTGAACTGGCGGAAACATATCCGGAAGTGAAAGC 1248
       AsnAlaTyrLysLeuIleTyrArgSerGlyLysThrLeuAspGluValLysProGluIleAlaGluLeuAlaGluThrTyrProGluValLysAla

1249 CTTTACCGATTTCTTTGCACGCTCAACGCGCGGTCTGATTCGTTAATGACTGAACAGCGTCCATTAACGATTGCCCTGGTCGCCGGAGAAACCTCC 1344
       PheThrAspPhePheAlaArgSerThrArgGlyLeuIleArg *
                                          lpxB- MetThrGluGlnArgProLeuThrIleAlaLeuValAlaGlyGluThrSer

1345 GGCGAT 1350
     GlyAsp...
```

FIG. 2. Nucleotide sequence upstream of *lpxB*. The 1,350-nucleotide sequence upstream of *lpxB* is shown. The predicted amino acid sequence of the open reading frames discussed in the text are shown and labeled. The putative Shine-Dalgarno sequences (34) are underlined. Transcription of these genes proceeds in the clockwise direction with respect to the *E. coli* chromosome.

stream of *lpxB* (Fig. 2). Immediately upstream of *lpxB* is *lpxA*. The *lpxA* gene extends from nucleotides 506 to 1294 and encodes a 28.0-kDa protein. This is in agreement with the reported molecular size of 27 kDa previously determined by electrophoretic mobility on a sodium dodecyl sulfate-polyacrylamide gel (7). The initiation codon for *lpxA* is a GUG codon preceded 8 bases upstream by a typical Shine-Dalgarno (34) sequence GGAGG. Immediately upstream of *lpxA* is another open reading frame. This open reading frame codes for a highly charged protein with a net positive charge. The presumed initiation codon for this 17.4-kDa protein is a UUG codon (as found in several other *E. coli* proteins [16]) 8 bases downstream of a typical Shine-Dalgarno sequence (AGGA).

Figure 3 is a plot of the codon preference for all three forward reading frames. Regions where the codon preference is consistently above 0.5 indicates high use of codons found in highly expressed *E. coli* proteins (12). Figure 3 shows good codon usage in the third reading frame up until nucleotide 48, where the first reading frame appears to encode a 17.4-kDa protein. At the termination codon of the 17.4-kDa open reading frame, the second reading frame shows good codon usage for the expression of *lpxA*. Although the first reading frame contains no stop codons from the first nucleotide, codon usage indicates the initiation codon to be near nucleotide 48.

Identification of the initiation codon of *lpxA*. To confirm the location of the *lpxA* initiation codon, a hybrid *lpxA-lacZ* gene was constructed (Fig. 4). This *lpxA-lacZ* gene produced a protein consisting of an amino-terminal domain of 71 amino acid residues starting from the presumed initiation codon of *lpxA* attached to the enzymatically active carboxy terminus of β-galactosidase. The hybrid protein was purified as described in Materials and Methods, and the amino terminus was sequenced. The analysis demonstrates that the amino-terminal sequence of the hybrid protein is Met-Ile-Asp-Lys-

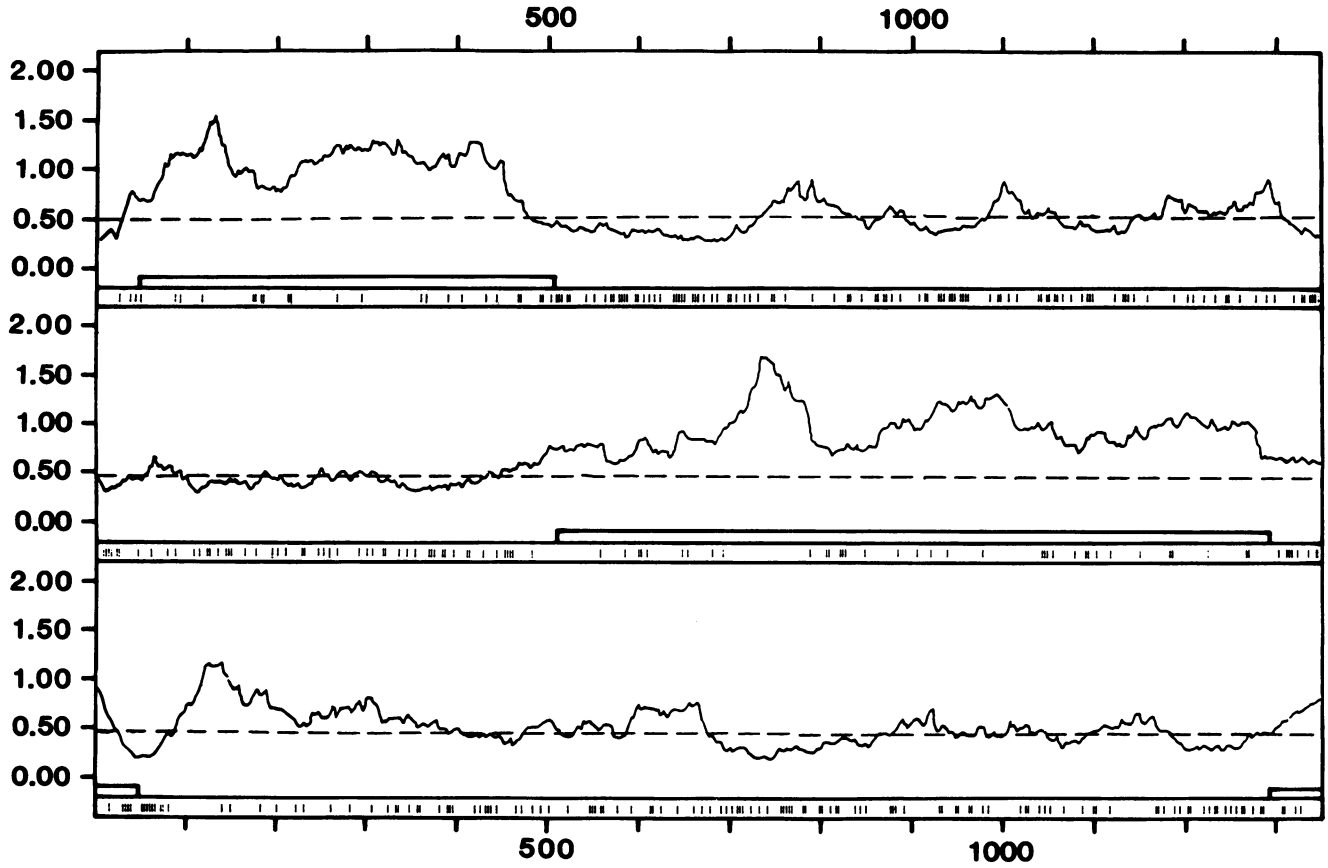


FIG. 3. Codon preference plot. This figure is a plot of codon preference (12) in each of the three forward reading frames on the vertical axis versus the nucleotide number on the horizontal axis (numbering is as in Fig. 2). Codon preference values reflect a computer-generated comparison of the codons found in each of the three forward reading frames, averaged over 25 consecutive codons, to codons found in *E. coli* genes that are highly expressed. Increasing positive values represent an increasing correlation between codons in a reading frame and these *E. coli* preferred codons. Codons not commonly found in highly expressed *E. coli* genes are termed rare. Rare codons are indicated underneath each plot by a small vertical line segment. Open reading frames described in the text are indicated.

X-Ala-Phe-Val. This sequence is in agreement with the amino-terminal sequence predicted by the nucleotide sequence of *lpxA* shown in Fig. 2. It should be noted that the GUG start codon encodes a methionine residue as expected (16) and that the residue is not removed from the completed hybrid protein.

Demonstration that the 28-kDa open reading frame is *lpxA*. Crowell et al. (7) demonstrated that UDP-GlcNAc acyltransferase is encoded on the fragment of DNA between *Sma*I and *Eco*RV (Fig. 1). To further demonstrate that the 28-kDa open reading frame is indeed *lpxA*, an *lpxA*(Ts) mutant strain was created. The *lpxA*(Ts) allele was isolated by localized mutagenesis (35). Strain JW353 was mutagenized with nitrosoguanidine, followed by P1 transduction into AT982, correcting the nearby *dapD* phenotype. Transductants were tested for the temperature-sensitive phenotype, and all temperature-sensitive strains were assayed for UDP-GlcNAc acyltransferase activity. One such temperature-sensitive strain, RX8a, showed less than 10% of the wild-type UDP-GlcNAc acyltransferase activity. The *lpxA*(Ts) allele was 60% cotransducible with *dapD* from this strain into MF6 as expected. One such transductant, RX800, was made *recA* by a 10-min Hfr cross with JC10240, selecting for the nearby *srlA*::Tn*10* insertion with tetracycline. A *recA* (UV-sensitive) recombinant, RX801, was used for complementation analysis. Strain RX801 exhibited less than 10% UDP-GlcNAc acyltransferase activity when compared with the nearly isogenic strain RX805. UDP-GlcNAc acyltransferase activity was as follows (nanomoles per minute, per milligram of protein): for RX805, 1.0; for RX801(pINGI) with or without arabinose, <0.1; for RX801(pLPXA) without arabinose, <0.1; for RX801(pLPXA) with arabinose, 7.0.

Previously described plasmids pDC4 (7) and pLC26-43 (6) were used to transform RX801. Both of these plasmids complement the RX801 *lpxA* mutation by restoring the UDP-GlcNAc acyltransferase activity and concomitantly relieving the temperature-sensitive phenotype (Fig. 5). To further clarify the region of DNA containing *lpxA*, we removed all the DNA to within seven nucleotides of the putitive Shine-Dalgarno sequence of the 28-kDa protein. This *Hgi*AI-*Eco*RV fragment was inserted in the arabinose-inducible expression vector pINGI (15) (Fig. 6). The resulting plasmid, pLPXA, was able to complement the temperature-sensitive phenotype of RX801 only in the presence of the inducer arabinose. The UDP-GlcNAc acyltransferase activity of RX801 containing pLPXA was sevenfold higher than wild-type level only when the cells were grown in the presence of arabinose (see above), demonstrating that the *lpxA* product is encoded by the 28-kDa open reading frame.

The 17.4-kDa open reading frame is not *firA*. Lathe and Lecocq (19) have demonstrated that *firA* is located on a 7.2-kb *Hin*dIII fragment containing a portion of *dnaE*, as are

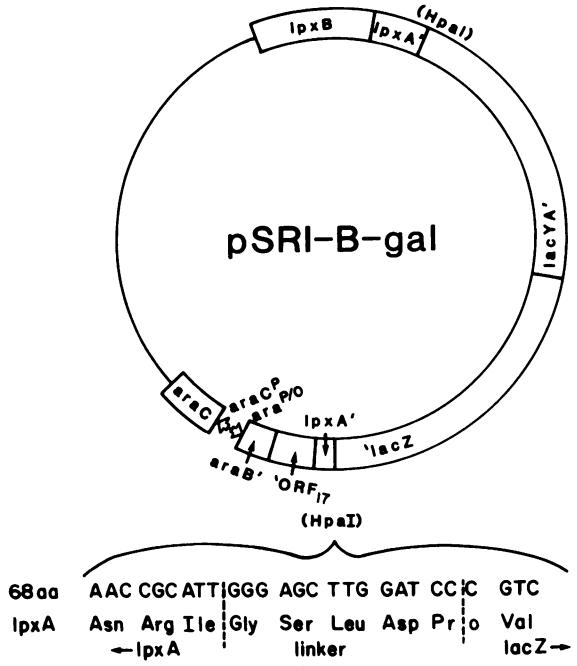


FIG. 4. Structure of pSR1-β-Gal. The *lacZ* gene from plasmid pICIII (23) was removed with restriction endonucleases *Sma*I and *Sal*I. The *Sal*I end was made blunt with a DNA polymerase I Klenow fragment in the presence of the four deoxyribonucleotide triphosphates. This *lacZ* DNA fragment was ligated into the unique *Hpa*I site in *lpxA* in pSRI (7). The DNA and protein sequence at the *lpxA-lacZ* junction is shown. The linker DNA is from pICIII. Relevant protein coding regions are boxed. Promoters are indicated by arrows. Abbreviations: *araC*^{P}, *araC* promoter; *ara*$^{P/O}$, arabinose promoter-operator.

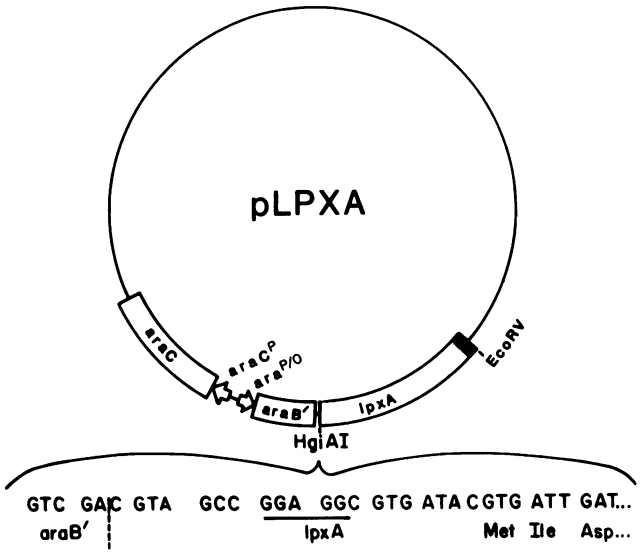
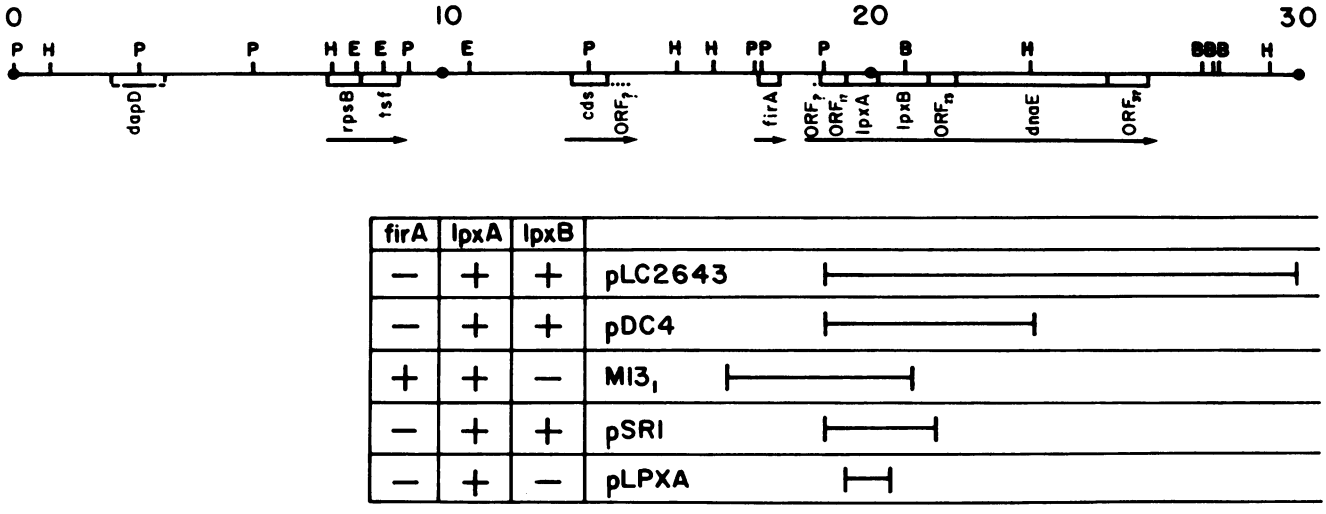


FIG. 6. Structure of pLPXA. The *lpxA* gene was excised from pDC4 (7) by cleavage with restriction endonucleases *Hgi*AI and *Eco*RV. The ends of the fragment were made blunt by digestion with a DNA polymerase I Klenow fragment. This fragment of DNA was inserted into the unique *Sal*I site, also made blunt, in the pINGI expression vector (15). The DNA and protein sequence at the *araB-lpxA* junction are shown. Relevant protein coding regions are boxed. Promoters are indicated by arrows. Abbreviations are as in Fig. 4.

lpxA and *lpxB*. Lathe et al. (18) have shown the *firA* gene product to be a 17-kDa highly charged protein with a net positive charge. The 17.4-kDa open reading frame immediately upstream of *lpxA* codes for a highly charged protein with a net positive charge. This led us to suspect that this open reading frame is *firA*.

To determine if ORF$_{17}$ was *firA*, we constructed the *rpoB firA recA* strain JCR21 (Table 1). This strain is normally



firA	lpxA	lpxB	
−	+	+	pLC2643
−	+	+	pDC4
+	+	−	M13$_1$
−	+	+	pSRI
−	+	−	pLPXA

FIG. 5. Genetic organization of the min 4 region of the *E. coli* chromosome. Genes with precisely defined locations are boxed with solid lines. Genes with locations that are not precisely defined are boxed with dashed lines. Direction of transcription is indicated by the arrows. The scale is in kilobases. The sources of information upstream of *firA* are references 3 and 14. Information downstream of *lpxA* is from reference 7a and 33. Below the genetic map are shown the regions of DNA used in this study for complementation analysis, with an indication of which genes are complemented. All complementation studies were done in *recA* mutant strains. Abbreviations: P, *Pst*I; H, *Hin*dIII; E, *Eco*RI.

temperature sensitive and susceptible to rifampin. However, when a wild-type *firA* allele is added to the strain extragenically it becomes temperature resistant and rifampin resistant. An M13 phage recombinant was constructed to test for complementation of *firA*. $M13_1$ contains a *Hin*dIII-*Bam*HI DNA fragment from the λ *firA24* hybrid phage inserted into the M13 vector mp18 (Fig. 5). To determine whether the M13 hybrid phage $M13_1$ complements the *firA* mutation, naked double-stranded DNA was used to transfect JCR21. Transformants able to complement *firA* were selected at high temperature and subsequently screened for rifampin resistance. $M13_1$ DNA produced temperature resistant colonies significantly above the background frequency for the reversion of the *firA* allele (approximately 10,000 fold more colonies were formed when $M13_1$ was used to transform JCR21 than when the vector mp18 was used) (Fig. 5); this demonstrates that *firA* is encoded upstream of *lpxB*.

To determine whether the 17.4-kDa open reading frame is *firA*, we purified FirA as described in Materials and Methods, and the amino terminus was sequenced. The amino-terminal sequence was Ala-Asp-Lys-Ile-Ala-Ile-Val-Asn-Met-Gly-Ser-Leu-Phe-Gln-Gln. This sequence is not found in the 17.4-kDa open reading frame, demonstrating that *firA* must be coded upstream of this open reading frame. Restriction mapping of the independently isolated *firA* complementing clones pDB2 (3) and λ *fir24* (19) indicates two closely spaced *Pst*I restriction sites upstream of ORF_{17} (Fig. 5). These two closely spaced *Pst*I sites are not shown on the Bendiak and Friesen (3) map of this portion of the chromosome. Preliminary sequence data of the small *Pst*I fragment indicate that the amino terminus of the *firA* protein is encoded within this fragment of DNA.

DISCUSSION

The DNA sequence upstream of *lpxB* reveals at least two overlapping genes in addition to *lpxB*: *lpxA* and ORF_{17}. The stop codon of *lpxA* UAA overlaps with the start codon AUG of *lpxB* (underlines indicate overlapping bases). Upstream of *lpxA*, the termination codon of ORF_{17}, UGA overlaps the start codon of *lpxA* GUG. The presumed start codon of ORF_{17} UUG overlaps the stop codon UGA of a possible open reading frame upstream of ORF_{17}. Work is currently in progress to determine whether all the overlapping genes are indeed translationally coupled. The hypothesis that ORF_{17} and *lpxA* are translationally coupled is supported by the fact that expression of UDP-GlcNAc acyltransferase is induced 140-fold over wild-type levels in pSR1 (7), in which translation occurs in the ORF_{17} reading frame. However, under the same conditions and with the same promoter, pLPXA increases UDP-GlcNAc acyltransferase production only sevenfold over wild-type levels. In the pLPXA case there should be no translation in the ORF_{17} reading frame.

Both ORF_{17} and *lpxA* start with codons other than AUG. ORF_{17} apparently starts with UUG, and *lpxA* starts with the more common GUG. Both ORF_{17} and *lpxA* have appropriately spaced Shine-Dalgarno sequences of four and five contiguous nucleotides, respectively, complementary with the 3′ end of the 16S rRNA.

Anderson and Raetz (2) have shown that UDP-GlcNAc acyltransferase of *E. coli* is predominantly cytoplasmic. The *lpxA* product is a 28.0-kDa protein with no long hydrophobic stretches.

Lathe et al. (18) have indicated the *firA* product to be a 17-kDa, highly charged, basic soluble protein. Sequence analysis shows the ORF_{17} product to be a 17.4-kDa protein consisting of 153 amino acid residues. It contains 41 charged residues, 22 of which are basic amino acids. The charged residues are fairly evenly distributed throughout the protein, with a higher proportion of basic amino acid residues at the carboxy terminus and a higher proportion of acidic residues at the amino terminus. Although ORF_{17} is similar to FirA in that it is highly charged with a net basic charge, amino-terminal sequence analysis shows FirA to be distinct from ORF_{17}. Preliminary sequence data show *firA* to be coded approximately 1 kb upstream of ORF_{17} (Fig. 5). The function of ORF_{17} is as yet unknown.

The DNA sequence shown in Fig. 2 contains no sequences matching the consensus sequence for *boxA* (*nusA* binding) (27), cyclic AMP-CAP (31), *lexA* (21), *dnaA* (9), or REP (36).

Similar overlapping genes have been previously reported in many operons (26). In most of these cases the genes are all cotranscribed from one promoter. However, in other cases such as the *thrS-infC* operon (29) it is known that there are promoters within the first gene transcribing only the downstream gene. It is unknown whether expression of the downstream gene *infC* is affected by the expression of the upstream gene *thrS*. Tomasiewicz and McHenry (38) have found a functional promoter just upstream of *dnaE*, in ORF_{23} (Fig. 5). It has yet to be determined whether this promoter is active in vivo or functions in concert with other promoters. It has been shown previously that *E. coli* will coordinately express growth-rate-dependent functions by clustering certain genes into operons (5). ORF_{17}, *lpxA*, *lpxB*, and *dnaE* may be part of such an operon, with possible internal promoters to conditionally decouple expression of certain genes.

Work is currently being done in our laboratory to determine whether the overlapping genes $ORF_?$-ORF_{17}-*lpxA*-*lpxB*-ORF_{23}-*dnaE*-ORF_{37} are coupled transcriptionally or translationally. We are also searching for the 5′ end of this string of overlapping genes. In addition we are trying to determine the function of the unidentified open reading frames.

ACKNOWLEDGMENTS

This research was supported by Public Health Service grant DK-19551 to C.R.H.R. from the National Institutes of Health. J.C. is an American Cancer Society Postdoctoral Fellow (PF-2876).

We thank Sue Schlaffman for her help in the construction of strain RX8a.

LITERATURE CITED

1. **Anderson, C. W., P. R. Baum, and R. F. Gesteland.** 1973. Processing of adenovirus 2-induced proteins. J. Virol. **12**:241–252.
2. **Anderson, M. S., and C. R. H. Raetz.** 1987. Biosynthesis of lipid A precursors in *Escherichia coli* A cytoplasmic acyl transferase that converts UDP-*N*-acetylglucosamine to UDP-3-*O*-(*R*-3-hydroxymyristoyl)-*N*-acetylglucosamine. J. Biol. Chem. **262**: 5159–5169.
3. **Bendiak, D. S., and J. D. Friesen.** 1981. Organization of the genes in the four minute region of the *Escherichia coli* chromosome: evidence that *rpoB* and *tsf* are cotranscribed. Mol. Gen. Genet. **181**:356–362.
4. **Bukhari, A. I., and A. L. Taylor.** 1971. Genetic analysis of diaminopimelic acid and lysine requiring mutants of *Escherichia coli*. J. Bacteriol. **105**:844–854.
5. **Burton, Z. F., C. A. Gross, K. K. Watanabe, and R. R. Burgess.** 1983. The operon that encodes the sigma subunit of RNA polymerase also encodes ribosomal protein S2 and DNA primase in *Escherichia coli* K12. Cell **32**:335–349.
6. **Clarke, L., and J. Carbon.** 1976. A colony bank containing synthetic ColEI hybrid plasmids representative of the entire

Escherichia coli genome. Cell **9**:91–99.

7. **Crowell, D. N., M. A. Anderson, and C. R. H. Raetz.** 1986. Molecular cloning of the genes for lipid A disaccharide synthase and UDP-*N*-acetylglucosamine acyltransferase in *Escherichia coli*. J. Bacteriol. **168**:152–159.

7a. **Crowell, D. N., W. S. Reznikoff, and C. R. H. Raetz.** 1987. Nucleotide sequence of the *Escherichia coli* gene for lipid A disaccharide synthase. J. Bacteriol. **169**:5727–5734.

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. **Fuller, R. S., B. E. Funell, and A. Kornberg.** 1984. The *dnaA* protein complex with the *Escherichia coli* chromosomal replication origin (*oriC*) and other DNA sites. Cell **38**:889–900.

10. **Ganong, B. R., and C. R. H. Raetz.** 1982. Massive accumulation of phosphatidic acid in conditionally lethal CDP-diglyceride synthetase mutants. J. Biol. Chem. **257**:389–394.

11. **Green, P. J., and M. Inouye.** 1984. Roles of the 5′ leader region of the *ompA* mRNA. J. Mol. Biol. **176**:431–442.

12. **Gribskov, M., J. Devereux, and R. R. Burgess.** 1984. The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. Nucleic Acids Res. **12**:539–549.

13. **Hayward, R., and J. Scaife.** 1976. Systematic nomenclature for the RNA polymerase gene of prokaryotes. Nature (London) **260**:646–647.

14. **Icho, T., C. P. Sparrow, and C. R. H. Raetz.** 1985. Molecular cloning and sequencing of the gene for CDP-diglyceride synthetase of *Escherichia coli* and cytodine auxotrophs of *E. coli*. J. Biol. Chem. **260**:12078–12083.

15. **Johnston, S., J.-H. Lee, and D. S. Ray.** 1985. High level expression of M13 gene II protein from an inducible polycistronic messenger RNA. Gene **34**:137–145.

16. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47**:1–45.

17. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.

18. **Lathe, R., H. Buc, J. P. Lecocq, and E. K. F. Bautz.** 1980. Prokaryotic histone-like protein interacting with RNA polymerase. Proc. Natl. Acad. Sci. USA **77**:3548–3552.

19. **Lathe, R., and J. P. Lecocq.** 1977. The *firA* gene a locus involved in the expression of rifampicin resistance in *Escherichia coli*. I. Characterisation of λ*firA* transducing phages constructed *in vitro*. Mol. Gen. Genet. **154**:43–51.

20. **Lathe, R., and J. P. Lecocq.** 1977. The *firA* gene a locus involved in the expression of rifampicin resistance in *Escherichia coli*. II. Characterisation of bacteriol proteins coded by λ*firA* transducing phages. Mol. Gen. Genet. **154**:53–60.

21. **Little, J. W., and D. W. Mount.** 1982. The SOS regulatory system of *Escherichia coli*. Cell **29**:11–22.

22. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

23. **Masui, Y., J. Coleman, and M. Inouye.** 1983. Multipurpose expression cloning vehicles in *Escherichia coli*, p. 15–32. *In* M. Inouye (ed.), Experimental manipulation of gene expression. Academic Press, Inc., New York.

24. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

25. **Nishijima, M., C. E. Bulawa, and C. R. H. Raetz.** 1981. Two interacting mutations causing temperature-sensitive phosphatidylglycerol synthesis in *Escherichia coli* membranes. J. Bacteriol. **145**:113–121.

26. **Normark, S., S. Bergstrom, T. Edlund, T. Grundstrom, B. Jaurin, F. P. Lindberg, and O. Olsson.** 1983. Overlapping genes. Annu. Rev. Genet. **17**:499–525.

27. **Olson, E. R., E. L. Flamm, and D. I. Friedman.** 1982. Analysis of *nutR*: a region of phage lambda required for antitermination of transcription. Cell **31**:61–70.

28. **Osborn, M. J.** 1979. Biosynthesis and assembly of the lipopolysaccharide of the outer membrane, p. 15–34. *In* M. Inouye (ed.), Bacterial outer membranes. John Wiley & Sons, Inc., New York.

29. **Pramanik, A., S. J. Wertheimer, J. J. Schwartz, and I. Schwartz.** 1986. Expression of *Escherichia coli infC*: identification of a promoter in an upstream *thrS* coding sequence. J. Bacteriol. **168**:746–751.

30. **Raetz, C. R. H.** 1986. Molecular genetics of membrane phospholipid synthesis. Annu. Rev. Genet. **20**:253–295.

31. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.

32. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

33. **Shepard, D., R. W. Oberfelder, M. M. Welch, and C. S. McHenry.** 1984. Determination of the precise location and orientation of the *Escherichia coli dnaE* gene. J. Bacteriol. **158**:455–459.

34. **Shine, J., and L. Dalgarno.** 1974. The 3′ terminal sequence of *Escherichia coli* 16s ribosome RNA: complementary to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.

35. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

36. **Stern, M. J., G. F.-L. Ames, N. H. Smith, E. C. Robinson, and C. F. Higgins.** 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. Cell **37**:1015–1026.

37. **Tarr, G. E.** 1986. Manual Edman sequencing system, p. 155–194. *In* J. E. Shively (ed.), Microcharacterization of polypeptides: a practical manual. Humana Press, Inc., Clifton, N.J.

38. **Tomasiewicz, H. G., and C. S. McHenry.** 1987. Sequence analysis of the *dnaE* gene of *Escherichia coli*. J. Bacteriol. **169**:5735–5744.

39. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13 mp18 and pUC19 vectors. Gene **33**:103–119.