JOURNAL OF BACTERIOLOGY, Dec. 1987, p. 5727–5734
0021-9193/87/125727-08$02.00/0

Vol. 169, No. 12

Nucleotide Sequence of the *Escherichia coli* Gene for Lipid A Disaccharide Synthase

DRING N. CROWELL, WILLIAM S. REZNIKOFF, AND CHRISTIAN R. H. RAETZ*

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

Received 5 February 1987/Accepted 15 September 1987

The *lpxB* gene of *Escherichia coli*, believed to be the structural gene for lipid A disaccharide synthase, is located in the min 4 region of the chromosome. It is adjacent to and clockwise of the *lpxA* gene, which is thought to encode UDP-*N*-acetylglucosamine acyltransferase. Preliminary evidence suggests that *lpxA* and *lpxB* are cotranscribed in the clockwise direction and thus constitute part of a previously unknown operon (D. N. Crowell, M. S. Anderson, and C. R. H. Raetz, J. Bacteriol. 168:152–159, 1986). We now report the complete nucleotide sequence of a 1,522-base-pair *Pvu*II-*Hin*cII fragment known to carry the *lpxB* gene. This sequence contained an open reading frame of 1,149 base pairs, in agreement with the predicted size, location, and orientation of *lpxB*. There was a second open reading frame 5′ to, and in the same orientation as, *lpxB* that corresponded to *lpxA*. The ochre codon terminating *lpxA* was shown to overlap the methionine codon identified as the initiation codon for *lpxB*, suggesting that these genes are cotranscribed and translationally coupled. A third open reading frame was also shown to begin at the 3′ end of *lpxB* with analogous overlap between the opal codon terminating *lpxB* and the methionine codon that putatively initiates translation downstream of *lpxB* in the clockwise direction. These results argue that at least three genes constitute a translationally coupled operon in the min 4 region of the *E. coli* chromosome. The accompanying paper by Tomasiewicz and McHenry (J. Bacteriol. 169:5735–5744, 1987) presents 4.35 kilobases of DNA sequence, beginning at the 3′ end of *lpxB*, and argues that *dnaE* and several other open reading frames may be members of this operon.

The outer membrane of *Escherichia coli* is composed of two lipid monolayers that are chemically distinct (25). The inner monolayer of the outer membrane consists primarily of glycerophospholipids and is thus similar in lipid composition to the two monolayers of the inner membrane. In contrast, lipopolysaccharide (LPS) is the major component of the outer monolayer of the outer membrane (19, 24). LPS is a complex molecule that has three structural domains: an O-antigen domain that extends into the growth medium, a core oligosaccharide that is conserved among gram-negative bacteria, and a lipid A moiety (27). Lipid A, which is a phosphorylated glycolipid, anchors the LPS molecule to the outer membrane and causes LPS to have endotoxic and immunostimulatory properties (9, 18, 24, 25, 27).

Although the pathway leading to lipid A biosynthesis has not been completely elucidated, several enzymes believed to be involved in lipid A biosynthesis have been detected in crude extracts of *E. coli* (1, 5, 26,). The first putative step in lipid A biosynthesis is catalyzed by UDP-*N*-acetylglucosamine acyltransferase (1) (Fig. 1). This enzyme transfers a β-hydroxymyristoyl moiety from β-hydroxymyristoyl acyl carrier protein to the 3 position of the glucosamine ring of UDP-*N*-acetylglucosamine. The UDP-3-*O*-acyl-*N*-acetylglucosamine product thus formed then undergoes substitution at the 2 position of the glucosamine ring (Fig. 1.) in the presence of *E. coli* crude extracts and β-hydroxymyristoyl acyl carrier protein to form UDP-2,3-diacylglucosamine (1, 1a, 2). Extracts of *E. coli* also catalyze the hydrolysis of UDP-2,3-diacylglucosamine (Fig. 1) to 2,3-diacylglucosamine-1-phosphate (1). The first putative disaccharide precursor of lipid A is generated by lipid A disaccharide synthase (Fig. 1), which catalyzes the reaction UDP-2,3-diacylglucosamine + 2,3-diacylglucosamine-1-phosphate → 2′,3′-diacylglucosamine (β1 → 6) 2,3-diacylglucosamine-1-phosphate + UDP (26). This tetraacyl-disaccharide-1-phosphate compound is then converted by a series of reactions (Fig. 1) to mature lipid A (25, 26a).

The elucidation of the lipid A biosynthetic pathway began with the isolation of a mutation in the *lpxB* gene (21, 25). This mutation, called *lpxB1*, causes *E. coli* cells to accumulate UDP-2,3-diacylglucosamine and 2,3-diacylglucosamine-1-phosphate (22, 30) and, in the presence of mutations in the gene encoding phosphatidylglycerophosphate synthase (*pgsA*), also causes temperature-sensitive growth (20). The interaction between mutations in *lpxB* and *pgsA* is unclear, but cells harboring the *lpxB1* lesion lack lipid A disaccharide synthase activity (26, 30). Overproduction of lipid A disaccharide synthase by increased gene dosage (i.e., plasmid-borne copies of *lpxB*) has been demonstrated, arguing that *lpxB* is the structural gene for this enzyme (5).

The *lpxB* gene is located 631 base pairs (bp) counterclockwise of *dnaE* on the *E. coli* chromosome (5, 20, 28, 31). Furthermore, a recently discovered gene called *lpxA*, which directs the synthesis of UDP-*N*-acetylglucosamine acyltransferase activity, is located immediately counterclockwise of *lpxB* (1, 5). The *lpxA* and *lpxB* genes are both transcribed in the clockwise direction, toward *dnaE*, and preliminary evidence suggests that they may be cotranscribed (5).

We have now determined the complete nucleotide sequence of the *lpxB* gene and identified the codon that initiates translation of *lpxB*. Sequences flanking *lpxB* in both directions suggest that there are several overlapping genes, including *lpxA*, *lpxB*, and *dnaE* (31), in the min 4 region of the *E. coli* chromosome. Presumably, these genes are cotranscribed and constitute part of a previously unknown operon.

* Corresponding author.

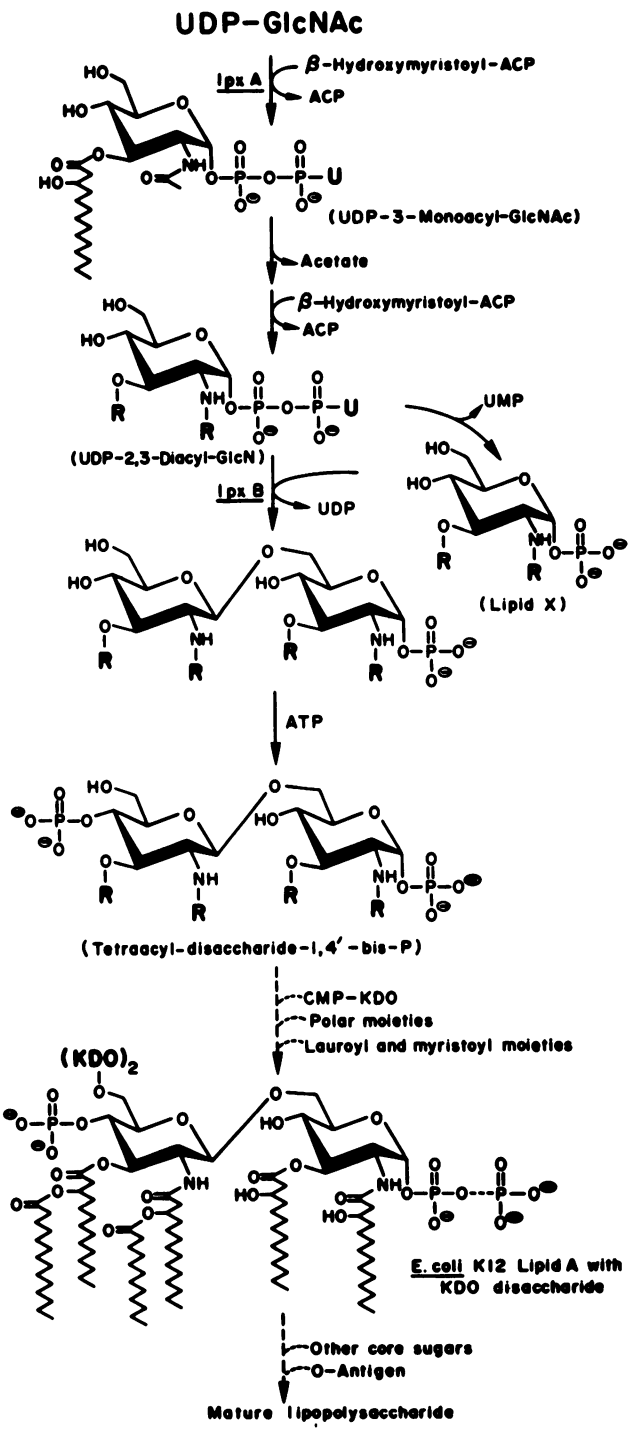

FIG. 1. Biosynthesis of lipid A disaccharides from monosaccharide precursors in extracts of *E. coli*. Evidence for this scheme has been presented previously (1, 1a, 2, 25, 26, 26a, 30). Abbreviations: ACP, acyl carrier protein; R, β-hydroxymyristoyl moiety; U, uridine; KDO, 2-keto-3-deoxyoctulosonic acid; GlcNAc, *N*-acetylglucosamine; GlcN, glucosamine; P, phosphate.

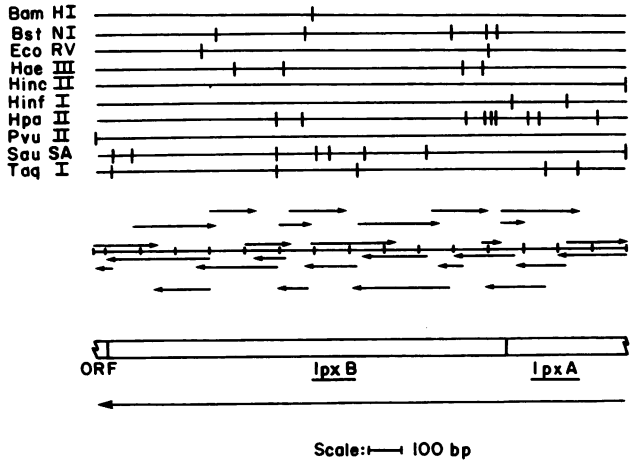

FIG. 2. Strategy for sequencing the *lpxB* gene. Small arrows represent sequenced fragments. These arrows point in the 5′ to 3′ direction. Relevant restriction enzyme recognition sites are shown at the top. The open reading frames corresponding to *lpxA*, *lpxB*, and the unidentified gene downstream of *lpxB* are indicated by boxes of the appropriate size. The large arrow at the bottom indicates the clockwise direction of transcription of these genes. Abbreviation: ORF, Open reading frame.

MATERIALS AND METHODS

Materials. Restriction enzymes were from Bethesda Research Laboratories, Gaithersburg, Md., or from New England Biolabs, Beverly, Mass. The Klenow fragment of DNA polymerase I, calf intestine alkaline phosphatase, and T4 polynucleotide kinase were from Boehringer-Mannheim Biochemicals, Indianapolis, Ind. T4 DNA ligase was purchased from New England Nuclear Corp., Boston, Mass. [γ-^{32}P]ATP was from Amersham Corp., Arlington Heights, Ill. Chemical reagents for Maxam-Gilbert sequencing reactions (e.g., dimethyl sulfate, formic acid, hydrazine, piperidine, etc.) were obtained from Eastman Kodak, Rochester, N.Y., or from Fisher Scientific Co., Pittsburgh, Pa. Agarose and reagents for polyacrylamide gel electrophoresis of DNA fragments and proteins were from Bethesda Research Laboratories. Kodak XAR-5 film was used for autoradiography. Tryptone, yeast extract, and agar were from Difco Laboratories, Detroit, Mich.

Bacterial strains. All *lpxB*$^+$ plasmids were stored in strain DC1 (*pqsA444 lpxB1 recA56 rpsL136 srl-300*::Tn*10*) as described previously (5). The plasmid pMC1403-22 was grown in strain RZ211 [Δ(*lac-pro*) *recA56* Str^r Srl$^-$] (12). All strains were grown in LB (17) medium at 37°C, unless otherwise indicated.

Plasmids. The plasmid pDC4 (5) was the source of all DNA fragments used in the sequencing of *lpxB*. The 2.9-kilobase-pair (kb) *Bam*HI insert of pMC1403-22 was also obtained from pDC4. The vector used in the construction of pMC1403-22 was pMC1403 (4).

DNA sequencing techniques. DNA fragments used in the sequencing of *lpxB* were dephosphorylated by digestion with calf intestine alkaline phosphatase and 5′-end labeled with $^{32}P_i$ by T4 polynucleotide kinase reaction in the presence of [γ-^{32}P]ATP. Labeled DNAs were then treated with an appropriate restriction enzyme and separated by polyacrylamide (8%) gel electrophoresis, or the strands were separated on a 6% strand-separating gel (15). DNA sequencing

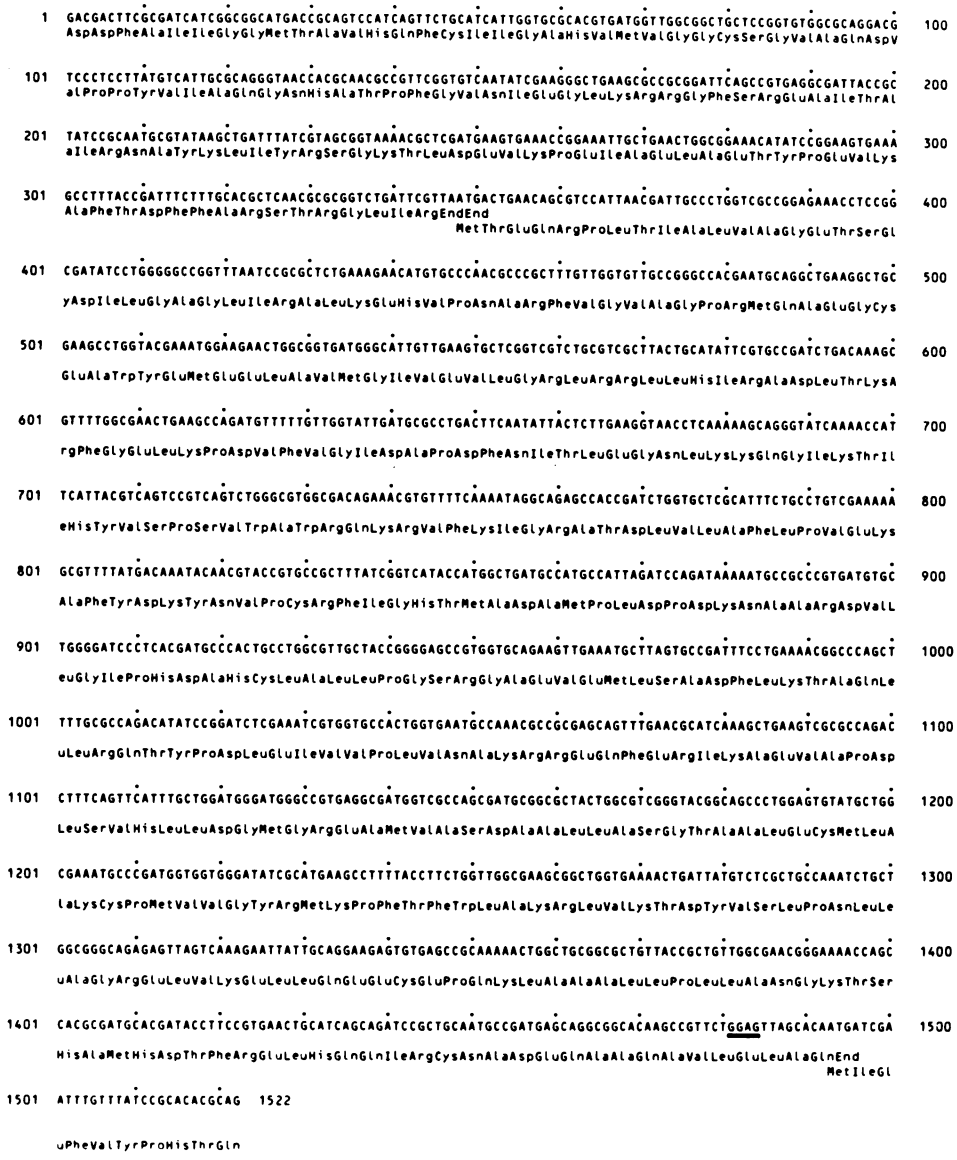

FIG. 3. Nucleotide sequence of the *lpxB* gene. The complete 1,522-bp nucleotide sequence described in the text is shown. The predicted amino acid sequences of the three open reading frames found in this nucleotide sequence are staggered to illustrate the overlap between them. The three open reading frames correspond to the 3′ end of *lpxA*, bases 1 to 351; *lpxB*, bases 348 to 1496; and the unidentified open reading frame downstream of *lpxB*, bases 1493 to 1522. A putative ribosome-binding site (29), found immediately upstream of this unidentified open reading frame, is underlined. Transcription of these genes proceeds in the clockwise direction with respect to the *E. coli* chromosome (i.e., from top to bottom of the figure).

was performed as described by Maxam and Gilbert (16) and Maniatis et al. (15).

Computer programs. Computer programs used in analyzing nucleotide and amino acid sequences were provided by the University of Wisconsin Genetics Computer Group (6). The hydropathy plot was generated by the program of Kyte and Doolittle (6, 13), as modified by Michael Gribskov, McArdle Laboratory, University of Wisconsin-Madison.

Recombinant DNA techniques. Plasmid preparation and cloning techniques were as described by Maniatis et al. (15). DNA fragments used for sequencing or cloning were prepared by the method of Dretzen et al. (7) or by electroelution into 7.5 M ammonium acetate with an apparatus purchased from International Biotechnologies, Inc. Transformation of *E. coli* cells was performed as described (15). Transformants were spread onto LB plates containing ampicillin (50 μg/ml) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (50 μg/ml) and incubated at 37°C.

Miscellaneous techniques. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis was done by the method of Laemmli (14). β-Galactosidase assays were as described by Miller (17).

RESULTS

Nucleotide sequence of the *lpxB* gene. The *lpxB* gene of *E. coli* is located 631 bp counterclockwise of *dnaE* in the min 4 region of the chromosome on a 1.7-kb *Pvu*II-*Nru*I fragment

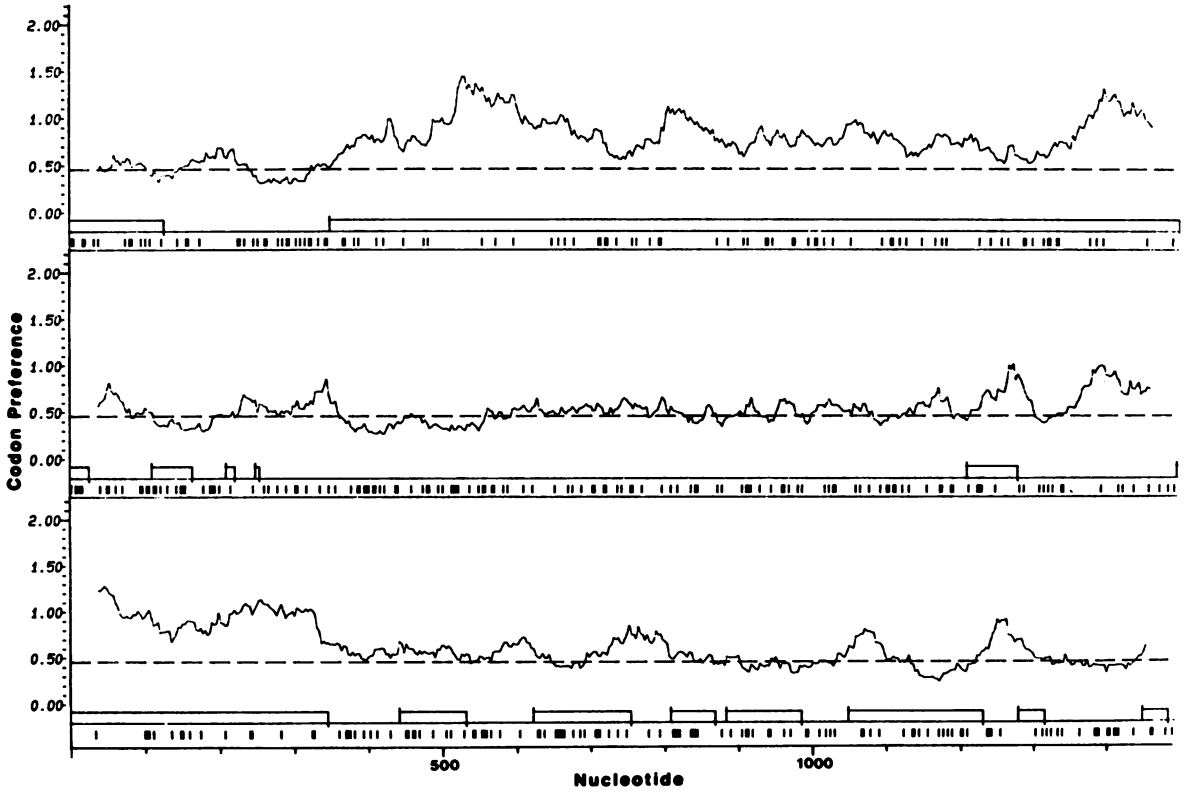

FIG. 4. Frequency of rare codons in the *lpxA* and *lpxB* open reading frames. This figure is a plot of codon preference (10) in each of the three forward reading frames on the *y* axis versus nucleotide number on the *x* axis (see Fig. 3). Codon preference values reflect a computer-generated comparison of the codons found in the three forward reading frames, averaged over 25 consecutive codons, with codons found in *E. coli* genes that are highly expressed. Increasing positive values represent an increasing correlation between the codons in a reading frame and these common, or "preferred," codons. Codons not commonly found in highly expressed *E. coli* genes are called rare, or "poor," codons. Rare codons in the three forward reading frames are indicated underneath each plot by small, vertical line segments. All possible open reading frames beginning upstream of the sequence, or beginning in the sequence with ATG, are indicated by boxes of the appropriate size. The open reading frames identified as *lpxA* and *lpxB* appear in the lower plot (nucleotide 1 to nucleotide 351) and in the upper plot (nucleotide 348 to nucleotide 1496), respectively. No significant open reading frames appear in the middle plot.

(5, 31). An overlapping 2.5-kb *Hin*cII fragment also carries the *lpxB* gene (5). The 1.5-kb overlap between these fragments was sequenced by the method of Maxam and Gilbert (16) to identify the *lpxB* coding region. The *lpxB*$^+$ plasmid pDC4, which has been described (5), was the source of all fragments used in the sequencing of *lpxB*. The sequencing strategy and the locations of relevant restriction enzyme recognition sites are shown in Fig. 2.

The sequence of the 1,522-bp *Pvu*II-*Hin*cII overlap described above is shown in Fig. 3. This sequence revealed an open reading frame that was 1,149 bp in length and oriented (5′ to 3′) in the clockwise direction. Previous studies have shown that the protein product of the *lpxB* gene has a molecular weight of 42,000, suggesting that *lpxB* is 1.1 to 1.2 kb in length. In addition, *lpxB* has been shown to have a clockwise direction of transcription (5). Hence, this open reading frame agrees with the predicted size, location, and orientation of *lpxB*. The frequency of rare codons (10) throughout the *lpxB* open reading frame was low (Fig. 4), arguing that no frame shift errors exist in the sequence.

The sequence shown in Fig. 3 revealed another open reading frame oriented in the clockwise direction that was counterclockwise of, or 5′ to, the *lpxB* coding sequence. The ochre codon terminating this open reading frame overlapped

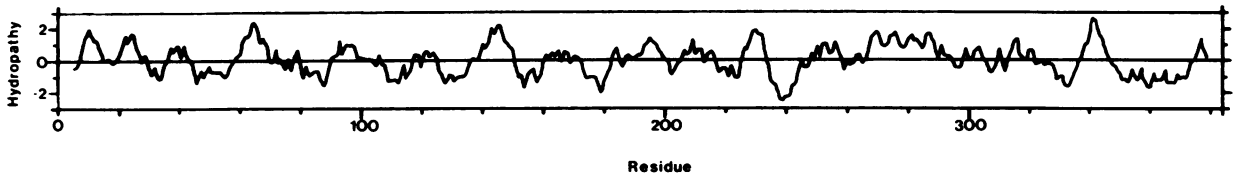

FIG. 5. Hydrophilic nature of the *lpxB* gene product. The protein product of the *lpxB* gene, deduced from the nucleotide sequence of *lpxB*, was analyzed for hydropathy by the computer program of Kyte and Doolittle (6, 13), as modified by Michael Gribskov, McArdle Laboratory, University of Wisconsin-Madison. Hydropathy is plotted on the *y* axis versus amino acid residue number on the *x* axis. Residue 1 corresponds to the amino terminus of the *lpxB* gene product. Hydropathy values are averaged over nine amino acid residues. Positive values indicate hydrophobic regions, and negative values indicate hydrophilic regions.

the methionine codon identified as the initiation codon for *lpxB* (see below). Previous work has shown that *lpxA* is transcribed in the clockwise direction and lies between *lpxB* and the genomic *Sma*I site 1.0 kb counterclockwise of *lpxB* (5). In addition, the protein product of the *lpxA* gene has been shown to have a molecular weight of 28,000, arguing that *lpxA* is approximately 800 bp in length (5), and this open reading frame has recently been shown to be 792 bp in length (J. Coleman and C. R. H. Raetz, submitted for publication). Hence, this open reading frame corresponds to the *lpxA* gene, since it agrees with the predicted size, location, and orientation of *lpxA*. The frequency of rare codons (10) in the *lpxA* open reading frame was also low (Fig. 4), arguing against frameshift errors in the sequence.

The overlap between the *lpxA* and *lpxB* open reading frames suggested that these genes were cotranscribed, because transcription termination downstream of *lpxA* would prevent *lpxB* expression. This hypothesis was supported by the observation that no obvious ribosome-binding sites (29) or promoter sequences (11) were found upstream of *lpxB* in the *lpxA* coding region, suggesting that perhaps these genes are translationally as well as transcriptionally coupled (23).

A methionine codon that presumably initiates translation downstream of *lpxB* in the clockwise direction overlapped the opal codon that terminates the *lpxB* coding region (Fig. 3), arguing that at least three genes are cotranscribed and translationally coupled (31). This methionine codon was preceded by a consensus ribosome-binding site (Fig. 3) (29). In addition, the overlap between *lpxB* and the open reading frame that begins with this codon was strikingly analogous to the overlap between *lpxA* and *lpxB*. The recently determined nucleotide sequence of *dnaE* and its flanking DNA (31) overlaps and agrees with the nucleotide sequence reported here by 191 bp. These two sequences predict that this downstream gene is 642 bp in length and support our hypothesis that a gene clockwise of and overlapping with *lpxB* is translated in the clockwise direction toward *dnaE*.

Properties of the *lpxB* gene product. The nucleotide sequence of the *lpxB* gene allowed certain predictions to be made about the physical nature of the *lpxB* gene product. The amino acid sequence predicted by the nucleotide sequence of *lpxB* was analyzed by the computer programs of the University of Wisconsin Genetics Computer Group and the computer programs of Kyte and Doolittle (6, 13), as modified by Michael Gribskov, University of Wisconsin-Madison. These analyses predicted that the *lpxB* gene product is a predominantly hydrophilic protein with a molecular weight of 42,339 (Fig. 5). This prediction agrees with the observation that 70% of the lipid A disaccharide synthase activity in *E. coli* crude extracts remains in the supernatant after centrifugation at 100,000 × *g* for 2 h (26). The Kyte and Doolittle analysis also predicted that the *lpxB* gene product has regions of hydrophobicity. This is consistent with the knowledge that lipid A disaccharide synthase converts membrane-associated substrates (22) into an extremely hydrophobic product, which suggests that the enzyme interacts, at least transiently, with membranes.

Construction of a hybrid *lpxB-lacZ* gene. To confirm the location of the *lpxB* initiation codon, a hybrid *lpxB-lacZ* gene was constructed (Fig. 6). The protein product of this gene consisted of an amino-terminal domain corresponding to the amino terminus of the *lpxB* gene product and a carboxy-terminal domain corresponding to β-galactosidase. This fusion protein was purified as described below, and its amino-terminal sequence was determined.

The details of the cloning procedure were as follows. The

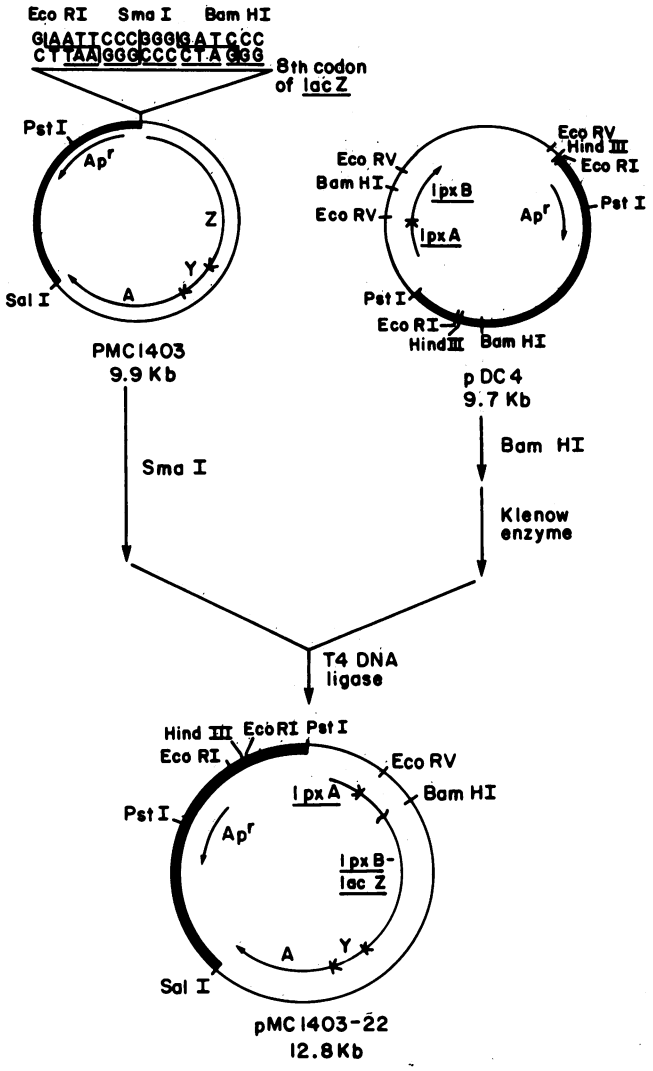

FIG. 6. Construction of plasmid pMC1403-22. All restriction and modification enzymes used in the construction of pMC1403-22 are indicated. The sequence of pMC1403 (4) at the point of insertion of the 2.9-kb *Bam*HI fragment of pDC4 (5) is also shown. A description of these steps is given in the text. Plasmid sizes, ampicillin resistance (Ap^r) genes, and relevant restriction enzyme recognition sites are shown. Arrows represent relevant cistrons. The wavy line drawn inside the *lpxB-lacZ* cistron of pMC1403-22 indicates the point of fusion between *lpxB* and *lacZ*. Fine lines represent *E. coli* chromosomal DNA, and heavy lines represent vector DNA. The chromosomal DNA of pDC4 is shown with correct clockwise orientation. Abbreviations: Z, *lacZ*; Y, *lacY*; A, *lacA*.

2.9-kb *Bam*HI fragment of pDC4 was prepared as described in Materials and Methods. This DNA fragment was treated with the Klenow fragment of DNA polymerase I in the presence of the four deoxyribonucleoside triphosphates. This treatment was expected to generate a blunt end within the *lpxB* coding region that would result in a translational fusion between *lpxB* and *lacZ* when ligated to pMC1403 (4) at the *Sma*I site. The fusion vector pMC1403 was digested with *Sma*I and then mixed with the blunt 2.9-kb *Bam*HI fragment of pDC4. This mixture was treated with T4 DNA ligase and then used to transform strain RZ211 to Amp^r lac^+. As expected, only the correct orientation of the insert DNA (pMC1403-22) resulted in a Lac^+ phenotype. In addition,

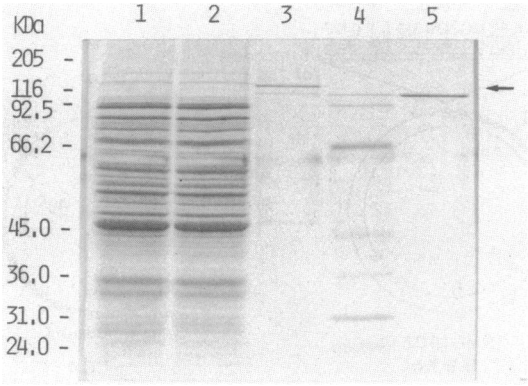

FIG. 7. SDS–10% polyacrylamide gel electrophoresis of column purified β-galactosidase fusion protein. Lanes: 1, 5 μl of sample (i.e., the crude extract that was loaded onto the column); 2, 5 μl of runthrough fraction (i.e., the fraction that did not bind to the column); 3, 1 μl of concentrated eluant; 4, molecular weight standards (1 μg each); 5, 2.5 μg of purified β-galactosidase. A description of the column conditions is given in the text. The β-galactosidase fusion protein was present in the concentrated eluant at approximately 2 μg/μl. This value was determined by comparing the intensity of Coomassie blue-stained fusion protein in lane 3 with the intensity of Coomassie blue-stained β-galactosidase in lanes 4 and 5. Hence, the concentrated eluant, which had a volume of 200 μl, contained approximately 400 μg of fusion protein. Two major contaminants with molecular weights of 116,000 and 48,000 can be seen in lane 3. The other apparent contaminants are in the sample loading buffer (compare lane 3 with lanes 4 and 5). The location of the 136,000-molecular-weight fusion protein is indicated by an arrow. Molecular size standards are: myosin, 205 kilodaltons (kDa); β-galactosidase, 116 kDa; phosphorylase B, 92.5 kDa; bovine serum albumin, 66.2 kDa; ovalbumin, 45 kDa; glyceraldehyde-3-phosphate dehydrogenase, 36 kDa; carbonic anhydrase, 31 kDa; trypsinogen, 24 kDa.

polyacrylamide gel electrophoresis of proteins from RZ211 cells carrying either the hybrid plasmid pMC1403-22 or the vector plasmid pMC1403 revealed a 136,000-molecular-weight protein present only in cells carrying pMC1403-22 (data not shown).

Purification of the *lpxB-lacZ* gene product. *E. coli* RZ211 cells carrying pMC1403-22 were grown in LB medium to late log phase (A_{550}, 1.0), sedimented, and suspended (10 ml/liter of culture) in 10 mM Tris hydrochloride (pH 8.0) containing the protease inhibitors α_2-macroglobulin (30 μg/ml) and phenylmethylsulfonyl fluoride (50 μg/ml). The cell suspension was then passed through a French pressure cell at 18,000 lb/in^2, and unbroken cells were sedimented by centrifugation at 2,000 × *g* for 10 min. Triton X-100 was added to the crude extract to a final concentration of 2%, and the membrane fraction was removed by centrifugation at 100,000 × *g* for 2 h. The sample was then divided into 5-ml portions and stored at −70°C. This crude extract contained approximately 18 nmol [(A_{420} × 380)/min] of β-galactosidase activity (17) per min per μl, suggesting that each liter culture of RZ211(pMC1403-22) cells produced 0.6 mg of fusion protein. This calculation, which is consistent with the result that 0.4 mg of fusion protein was purified from a 1-liter culture of RZ211(pMC1403-22) cells (see below), assumes that the specific activity of the fusion protein is the same as the specific activity of pure β-galactosidase (3.0×10^5 nmol/min per mg). Given that a liter culture of *E. coli* cells produces 100 mg of soluble protein, this fusion protein constituted 0.6% of the soluble protein in RZ211(pMC1403-22) cells.

Two 5-ml portions of this crude extract were thawed and combined. Sodium chloride was added to a final concentration of 150 mM, and the sample was loaded (5 ml/h) onto a 1-ml monoclonal anti-β-galactosidase immunoaffinity column (a gift from Promega Biotec, Madison, Wisc.) that was preequilibrated at 4°C in 10 mM Tris hydrochloride (pH 8.0)–2% Triton X-100–150 mM NaCl. After the sample was loaded, the column was washed with 5 ml of 10 mM Tris hydrochloride (pH 8.0)–150 mM NaCl. The fusion protein was then eluted from the column with 3 ml of 100 mM $NaHCO_3$–Na_2CO_3 (pH 10.8). The 3-ml eluant was collected and concentrated on a centricon 30 microconcentrator. The retentate (0.2 ml) was then diluted with 2 ml of 10 mM Tris hydrochloride (pH 8.0)–0.1% sodium dodecyl sulfate (SDS) and re-concentrated. This sample, which contained approximately 400 μg of fusion protein as judged by SDS–10% polyacrylamide gel electrophoresis (Fig. 7), was loaded onto a preparative SDS–7.5% polyacrylamide gel. Following electrophoresis, the 136,000-molecular-weight protein was visualized by staining with Coomassie blue R-250, excised from the gel, and harvested by electroelution into 10 mM ammonium bicarbonate. Fifty micrograms (368 pmol) of purified protein was obtained.

Amino-terminal sequence analysis of the *lpxB-lacZ* gene product. Amino-terminal sequence analysis of the purified fusion protein (200 pmol) was done by automated Edman degradation (8) at the University of Wisconsin Biotechnology Center with an Applied Biosystems 470A protein sequencer. The phenylthiohydantoin-amino acid (PTH-amino acid) product from each cycle of Edman degradation was analyzed by high-pressure liquid chromatography on an IBM Instruments LC/9533 with an IBM C_{18} column.

The yields of PTH-amino acids were somewhat low, decreasing from 57 pmol in the first cycle of Edman degradation to 16 pmol in the fourth cycle. PTH-threonine (40 pmol) was the predominant PTH-amino acid detected in cycle 1, but 17 pmol of PTH-methionine was also present. In cycle 2, 23 pmol of PTH-glutamate was detected. Both PTH-glutamate and PTH-glutamine were detected in cycle 3. In cycle 4, 16 pmol of PTH-arginine was present. The background increased by cycle 5, making further identification impossible. This analysis suggests that most of the sequenced protein had the amino-terminal sequence Thr-Glu-Gln-Arg. These results also suggest that some of the sequenced protein had the amino-terminal sequence Met-Thr-Glu-Gln-Arg. This sequence agrees with the amino-terminal sequence predicted by the nucleotide sequence of *lpxB*. We therefore conclude that we have identified the correct initiation codon for *lpxB*.

DISCUSSION

Previous data have established the sizes, locations, and direction of transcription of *lpxA* and *lpxB* and have argued for their cotranscription (5). We report here the complete nucleotide sequence of *lpxB* and show that it is flanked on both sides by open reading frames, corresponding to *lpxA* in the counterclockwise direction and an unidentified open reading frame in the clockwise direction. Furthermore, we show that the termination codon for *lpxA* overlaps the initiation codon for *lpxB* and that the termination codon for *lpxB* overlaps the initiation codon for the unidentified open reading frame downstream of *lpxB*. These results support our hypothesis that *lpxA* and *lpxB* are cotranscribed and thus

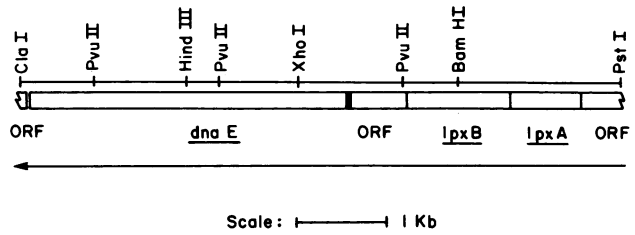

FIG. 8. Genetic organization of *lpxA*, *lpxB*, and *dnaE*. A 6.6-kb *ClaI-PstI* fragment carrying *lpxA*, *lpxB*, and *dnaE* is shown. Relevant restriction enzyme recognition sites are also shown at the top. *lpxA*, *lpxB*, *dnaE*, and the three unidentified open reading frames (ORFs) on this fragment are indicated by boxes of the appropriate size. The heavy line between *dnaE* and the open reading frame immediately downstream of *lpxB* denotes the prediction (31) that these two genes overlap by several codons. The broken line between *dnaE* and the open reading frame immediately downstream of *dnaE* denotes the prediction that these two genes do not overlap. The information in this figure is from the following sources: data to the left of the rightmost *PvuII* site (28, 31); data from the right most *PvuII* site to the middle of *lpxA*, this report; data to the right of the middle of *lpxA* (Coleman and Raetz, submitted). The arrow at the bottom of the figure indicates the clockwise direction of transcription of *lpxA*, *lpxB*, *dnaE*, and all three unidentified open reading frames.

constitute part of an operon consisting of three or more genes. We are currently studying the structure of *lpxA* and *lpxB* transcripts and also studying polarity in this region of the chromosome to determine whether these genes are, in fact, part of an operon. The results described above also suggest that *lpxA* and *lpxB* are translationally coupled (i.e., translation of *lpxB* requires translation of *lpxA*) (23). This hypothesis is supported by the observations that no consensus ribosome-binding site (29) can be found within 20 bp of the *lpxB* initiation codon and that certain plasmids (pDC25 and pDC27) carrying *lpxB* and the 3′ end of *lpxA* do not express *lpxB* in spite of a vector promoter on these plasmids that presumably directs transcription of *lpxB* (5). This promoter efficiently expresses *lpxB* from plasmids (pCR9 and pDC29) carrying the entire *lpxA* gene (D. N. Crowell, C. R. H. Raetz, and W. S. Reznikoff, manuscript in preparation).

Many operons consisting of genes with a common function (e.g., utilization of sugars or biosynthesis of amino acids) have been reported. Some of the genes in these operons overlap in precisely the same way *lpxA* and *lpxB* overlap (23). However, *lpxA* and *lpxB* may be members of a large operon that consists of genes with various functions. The recently determined nucleotide sequence of *dnaE* (31), which encodes the α subunit of DNA polymerase III (32), demonstrated possible overlap between the 3′ end of the unidentified open reading frame downstream of *lpxB* and the 5′ end of *dnaE*. The work of Tomasiewicz and McHenry (31) also revealed an open reading frame immediately downstream of *dnaE*, arguing that five genes may be cotranscribed. In our laboratory, a sixth open reading frame has been found upstream of *lpxA*. The termination codon for this open reading frame overlaps the codon identified as the initiation codon for *lpxA* (Coleman and Raetz, submitted). Hence, *lpxA* and *lpxB* are members of a string of overlapping genes. These observations argue that *lpxA* and *lpxB* constitute part of a 7.0- to 8.0-kb operon that includes *dnaE* and perhaps three other genes (Fig. 8).

The *dnaE* gene of *E. coli* encodes the polymerase (α) subunit of DNA polymerase III (32), the major enzyme responsible for chromosomal DNA replication. DNA replication is a process that is dependent on the rate of cell division. Since conditionally lethal mutations that affect lipid A biosynthesis have been reported (20, 21, 25), it seems likely that lipid A is essential for growth and division in *E. coli*. It is thus possible that lipid A biosynthesis, like DNA replication, is dependent on the rate of cell division. It has been proposed that *E. coli* cells coordinate essential, growth-rate-dependent functions such as biosynthesis of macromolecules by clustering certain genes into operons (3). We believe that *lpxA* and *lpxB* may be components of such an operon, consisting of genes involved in lipid A biosynthesis and DNA replication. Hence, a thorough study of this putative operon may explain how *E. coli* cells coordinate membrane biosynthesis and DNA replication.

ACKNOWLEDGMENTS

We thank the University of Wisconsin Genetics Computer Group and the University of Wisconsin Biotechnology Center for computer programs and protein sequencing, respectively.

D.N.C. was supported by Public Health Service Training Grant for Cellular and Molecular Biology 2-T32-GM07215 from the National Institutes of Health. This work was supported in part by Public Health Service grant DK-19551 from the National Institutes of Health to C.R.H.R. and by Public Health Service grant GM-19760 from the National Institutes of Health to W.S.R.

LITERATURE CITED

1. **Anderson, M. S., C. E. Bulawa, and C. R. H. Raetz.** 1985. The biosynthesis of gram-negative endotoxin: formation of lipid A precursors from UDP-GlcNAc in extracts of *Escherichia coli*. J. Biol. Chem. **260**:15536–15541.

1a. **Anderson, M. S., and C. R. H. Raetz.** 1987. Biosynthesis of lipid A precursors in *Escherichia coli*: a cytoplasmic acyltransferase that converts UDP-N-acetylglucosamine to UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamine. J. Biol. Chem. **262**: 5159–5169.

2. **Bulawa, C. E., and C. R. H. Raetz.** 1984. The biosynthesis of gram-negative endotoxin: identification and function of UDP-2,3-diacylglucosamine in *Escherichia coli*. J. Biol. Chem. **259**: 4846–4851.

3. **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 S21 and DNA primase in *E. coli* K12. Cell **32**:335–349.

4. **Casadaban, M. J., J. Chou, and S. N. Cohen.** 1980. *In vitro* gene fusions that join an enzymatically active β-galacatosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signals. J. Bacteriol. **143**:971–980.

5. **Crowell, D. N., M. S. 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.

6. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.

7. **Dretzen, G., M. Bellord, P. Sassone-Corsi, and P. Chambon.** 1981. A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. Anal. Biochem. **112**:295–298.

8. **Edman, P., and G. Begg.** 1967. A protein sequenator. Eur. J. Biochem. **1**:80–91.

9. **Galanos, C., M. A. Freudenberg, O. Luderitz, E. T. Rietschel, and O. Westphal.** 1979. Chemical, physiochemical and biological properties of bacterial lipopolysaccharides, p. 321–332. *In* E. Cohen (ed.), Biomedical application of the horseshoe crab (*Limulidae*). Alan R. Liss, Inc., New York.

10. **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.
11. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.
12. **Johnson, R. C., J. C. P. Yin, and W. S. Reznikoff.** 1982. Control of Tn5 transposition in *Escherichia coli* is mediated by protein from the right repeat. Cell **30**:873–882.
13. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydrophathic character of a protein. J. Mol. Biol. **157**:105–132.
14. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
15. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
17. **Miller, J. H.** 1972.Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **Morrison, D. C., and J. L. Ryan.** 1979. Bacterial endotoxins and host immune responses. Adv. Immunol. **28**:293–450.
19. **Nikaido, H.** 1973. Biosynthesis and assembly of lipopolysaccharide and the outer membrane layer of gram-negative cell wall, p. 131–208. *In* L. Leive (ed.), Bacterial membranes and walls. Marcel Dekker, Inc., New York.
20. **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.
21. **Nishijima, M., and C. R. H. Raetz.** 1979. Membrane lipid biogenesis in *Escherichia coli*: identification of genetic loci for phosphatidylglycerophosphate synthetase and construction of mutants lacking phosphatidylglycerol. J. Biol. Chem. **254**:7837–7844.
22. **Nishijima, M., and C. R. H. Raetz.** 1981. Characterization of two membrane-associated glycolipids from an *Escherichia coli* mutant deficient in phosphatidylglycerol. J. Biol. Chem. **256**: 10690–10696.
23. **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.
24. **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.
25. **Raetz, C. R. H.** 1986. Molecular genetics of membrane phospholipid synthesis. Annu. Rev. Genet. **20**:253–295.
26. **Ray, B. L., G. Painter, and C. R. H. Raetz.** 1984. The biosynthesis of gram-negative endotoxin: formation of lipid A disaccharides from monosaccharide precursors in extracts of *Escherichia coli*. J. Biol. Chem. **259**:4852–4859.
26a. **Ray, B. L., and C. R. H. Raetz.** 1987. The biosynthesis of gram-negative endotoxin: a novel kinase in *Escherichia coli* membranes that incorporates the 4′-phosphate of lipid A. J. Biol. Chem. **262**:1122–1128.
27. **Rietschel, E. T. (ed.).** 1984. Handbook of endotoxin, vol. 1. Elsevier Science Publishers B. V., Amsterdam.
28. **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.
29. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
30. **Takayama, K., N. Qureshi, P. Mascagni, M. A. Nashed, L. Anderson, and C. R. H. Raetz.** 1983. Fatty acyl derivatives of glucosamine 1-phosphate in *Escherichia coli* and their relation to lipid A. J. Biol. Chem. **258**:7379–7385.
31. **Tomasiewicz, H. G., and C. S. McHenry.** 1987. Sequence analysis of the *Escherichia coli dnaE* gene. J. Bacteriol. **169**:5735–5744.
32. **Welch, M. M., and C. S. McHenry.** 1982. Cloning and identification of the product of the *dnaE* gene of *Escherichia coli*. J. Bacteriol. **152**:351–356.