1997

Identification of the ADP-L-Glycero-D-Manno-Heptose-6-Epimerase (rfaD) and Heptosyltransferase II (rfaF) Biosynthesis Genes from Nontypeable Haemophilus Influenzae 2019

W A. Nichols

B W. Gibson

William Melaugh
University of San Francisco, melaugh@usfca.edu

N G. Lee

M Sunshine

See next page for additional authors

Recommended Citation

Nichols WA, Gibson BW, Melaugh W, Lee NG, Sunshine M, Apicella MA. Identification of the ADP-L-glycero-D-manno-heptose-6-epimerase (rfaD) and heptosyltransferase II (rfaF) biosynthesis genes from nontypeable Haemophilus influenzae 2019. Infect Immun. 1997 Apr;65(4):1377-86.

Authors
W A. Nichols, B W. Gibson, William Melaugh, N G. Lee, M Sunshine, and M A. Apicella

INFECTION AND IMMUNITY, Apr. 1997, p. 1377–1386 Vol. 65, No. 4
0019-9567/97/$04.00+0

Identification of the ADP-L-*glycero*-D-*manno*-Heptose-6-Epimerase (*rfaD*) and Heptosyltransferase II (*rfaF*) Biosynthesis Genes from Nontypeable *Haemophilus influenzae* 2019

WADE A. NICHOLS,[1] BRADFORD W. GIBSON,[2] WILLIAM MELAUGH,[3] NA-GYONG LEE,[1] MELVIN SUNSHINE,[1] AND MICHAEL A. APICELLA[1]*

Department of Microbiology, University of Iowa College of Medicine, Iowa City, Iowa 52242[1]; Department of Pharmaceutical Chemistry, School of Pharmacy, University of California at San Francisco, San Francisco, California 94143[2]; and Department of Chemistry, University of San Francisco, San Francisco, California 94117[3]

Received 13 March 1996/Returned for modification 18 April 1996/Accepted 25 January 1997

***Haemophilus influenzae* is an important human pathogen. The lipooligosaccharide (LOS) of *H. influenzae* has been implicated as a virulence determinant. To better understand the assembly of LOS in nontypeable *H. influenzae* (NtHi), we have cloned and characterized the *rfaD* and *rfaF* genes of NtHi 2019, which encode the ADP-L-*glycero*-D-*manno*-heptose-6-epimerase and heptosyltransferase II enzymes, respectively. This cloning was accomplished by the complementation of *Salmonella typhimurium* lipopolysaccharide (LPS) biosynthesis gene mutants. These deep rough mutants are novobiocin susceptible until complemented with the appropriate gene. In this manner, we are able to use novobiocin resistance to select for specific NtHi LOS inner core biosynthesis genes. Such a screening system yielded a plasmid with a 4.8-kb insert. This plasmid was able to complement both *rfaD* and *rfaF* mutants of *S. typhimurium*. The LPS of these complemented strains appeared identical to the wild-type *Salmonella* LPS. The genes encoding the *rfaD* and *rfaF* genes from NtHi 2019 were sequenced and found to be similar to the analogous genes from *S. typhimurium* and *Escherichia coli*. The *rfaD* gene encodes a polypeptide of 35 kDa and the *rfaF* encodes a protein of 39 kDa, as demonstrated by in vitro transcription-translation studies. Isogenic mutants which demonstrated truncated LOS consistent with inner core biosynthesis mutants were constructed in the NtHi strain 2019. Primer extension analysis demonstrated the presence of a strong promoter upstream of *rfaD* but suggested only a very weak promoter upstream of *rfaF*. Complementation studies, however, suggest that the *rfaF* gene does have an independent promoter. Mass spectrometric analysis shows that the LOS molecules expressed by *H. influenzae rfaD* and *rfaF* mutant strains have identical molecular masses. Additional studies verified that in the *rfaD* mutant strain, D-*glycero*-D-*manno*-heptose is added to the LOS molecule in place of the usual L-*glycero*-D-*manno*-heptose. Finally, the genetic organizations of the inner core biosynthesis genes of *S. typhimurium*, *E. coli*, and several strains of *H. influenzae* were examined, and substantial differences were uncovered.**

Nontypeable *Haemophilus influenzae* (NtHi) is the causative agent of a wide range of diseases, including infant meningitis, otitis media, and respiratory tract infections. Lipooligosaccharide (LOS) is a major component of the outer membrane of NtHi, and LOS expressed by the bacteria has been implicated as a virulence factor in these infections (10, 11, 26).

The structure of LOS from NtHi has been elucidated (5) (Fig. 1). The principal differences between the LOS/lipopolysaccharide (LPS) core structure of *H. influenzae* and other gram negative bacteria is the single 3-deoxy-D-manno-octulosonic acid (KDO) and the branching of its triheptose region (5). Studies have shown that oligosaccharide chain extension can occur from any one of the three heptoses. Thus, the heptose structure is crucial to the ultimate configuration of the LOS. An important step in the assembly of LOS is the production of the L-glycerol form of heptose, which is accomplished by the ADP-L-*glycero*-D-*manno*-heptose-6-epimerase. Without the activity of this enzyme, L-*glycero*-D-*manno*-heptose (LD-heptose) molecules are not readily incorporated into the LOS molecule and subsequent additions cannot take place. This results in a highly truncated LOS structure which has hydrophobic characteristics. The gene that encodes this enzyme has been identified in *Salmonella typhimurium* (23) and *Escherichia coli* (17) and called *rfaD*.

The addition of heptose molecules is achieved by the action of enzymes called heptosyltransferases, and three are involved in LOS biosynthesis of NtHi. The addition of the proximal heptose to the KDO is achieved by heptosyltransferase I, the addition of the second heptose to the proximal heptose by heptosyltransferase II, and the addition of the third heptose to the second by heptosyltransferase III. The gene encoding the heptosyltransferase II has also been cloned and characterized in *S. typhimurium* and *E. coli* and called *rfaF*.

While little is currently understood about the genetics of LOS biosynthesis in NtHi, the genetics of LPS biosynthesis has been well characterized in *E. coli* and *Salmonella* species. Structural analyses of the LOS of *H. influenzae* and LPS of *E. coli* and *Salmonella* have determined that the inner core structures of the molecules are very similar, and this similarity provides an opportunity to use some of the many *S. typhimurium* LPS mutants to study genes involved in inner core assembly of NtHi LOS. The method uses the hydrophobic nature of deep rough mutants of *S. typhimurium*. Such mutants are highly susceptible to hydrophobic antibiotics such as novobiocin; however, they become resistant to novobiocin if trans-

* Corresponding author. Mailing address: Department of Microbiology, BSB 3-403, University of Iowa College of Medicine, 51 Newton Road, Iowa City, Iowa 522442. Phone: (319) 335-7807. Fax: (319) 335-9006. E-mail: Michael-Apicella@uiowa.edu.

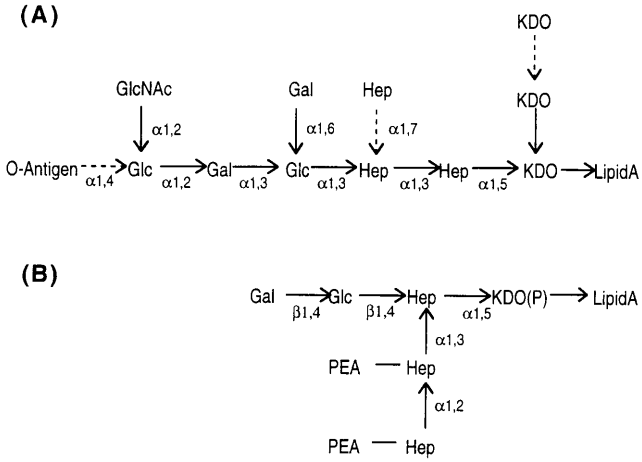

FIG. 1. Schematic illustration of the structures of *S. typhimurium* LPS (A) and NtHi 2019 LOS (B). Abbreviations not given in the text: Gal, galactose; Glc, glucose; GlcNAc, *N*-acetylglucosamine; Hep, LD-heptose; P, phosphate. Possible partial substitutions are indicated with dashed arrows. The higher-molecular-weight forms of NtHi 2019 LOS have not been fully characterized.

formed with a plasmid which bears the appropriate complementing gene. In this manner, we are able to use novobiocin resistance to select inner core biosynthesis genes from an NtHi genomic DNA library. We have previously used this technique with success in isolation and identification of LOS biosynthesis genes from *H. influenzae* (13).

In this work, we describe the cloning and identification of the *rfaD* and *rfaF* genes, which encode the ADP-L-*glycero*-D-*manno*-heptose-6-epimerase and the heptosyltransferase II of NtHi 2019.

MATERIALS AND METHODS

Strains, plasmids, and culture conditions. The bacterial strains and plasmids used in this study are described in Table 1. Plasmid pGB19 was kindly provided by Gerald J. Barcak. Plasmid pGB19 is a shuttle vector for allowing the plasmid to be transformed in *E. coli* and *H. influenzae* strains. The plasmid bears a chloramphenicol resistance determinant. *E. coli* and *S. typhimurium* strains were grown in Luria broth (LB) medium containing appropriate antibiotics at 37°C. *S. typhimurium* strains were grown on heart infusion (HI) medium during competence induction and transformation procedures. NtHi 2019 was cultured in brain heart infusion broth supplemented with 2% Fildes reagent (Difco Laboratories) (sBHI) at 37°C with agitation or on sBHI agar at 37°C in humidified 5% CO_2 atmosphere. Novobiocin (100 μg/ml) was used for selection of *S. typhimurium* expressing a complete LPS structure, and kanamycin (20 μg/ml) was used for selection of NtHi isolates containing the kanamycin resistance (Kan^r) cassette. Ampicillin (100 μg/ml) was added for selection and maintenance of recombinant plasmids bearing the beta-lactamase gene (Amp^r). Chloramphenicol was used at 40 μg/ml for *E. coli* strains bearing the shuttle vector pGB19, and chloramphenicol was used at 1.5 μg/ml for *H. influenzae* strains bearing the same plasmid.

Recombinant DNA techniques. Restriction enzymes and DNA-modifying enzymes were purchased from New England Biolabs and were used according to the manufacturer's specifications. Transformation of *E. coli* strains with plasmid DNA was routinely done by the $CaCl_2$ method (7), *Salmonella* strains were transformed by electroporation (1), and *Haemophilus* strains were transformed by the MIV starvation medium procedure (9). Cloning of DNA fragments generated by PCR into a vector plasmid was performed by using the TA cloning system from Invitrogen.

Construction of an NtHi 2019 genomic library. Chromosomal DNA from NtHi 2019 was partially digested with the restriction endonuclease *Apo*I and resolved on a 1.0% agarose gel. DNA fragments with sizes ranging from 4 to 8 kb were excised, purified by passage through DEAE-cellulose membranes (Schleicher & Schuell), and ligated into the λSH*lox* cloning vector (Novagen, Inc.). The ligation mix was packaged by using a Gigapack II Gold packaging kit (Stratagene). The plasmid library was made by in vivo excision as recommended in the manufacturer's protocol. Colonies bearing the excised plasmids were pooled, and plasmid DNA was isolated as described below.

Isolation and purification of DNA. Bacterial genomic DNA was prepared according to the cetyltrimethylammonium bromide method described by Ausubel et al. (1). Plasmid DNA was prepared by the alkaline lysis method (1) and purified on a CsCl gradient. Final DNA pellets were resuspended in sterile, deionized water and stored at 4°C.

Complementation of *S. typhimurium rfa* mutants. Overnight cultures of the *S. typhimurium rfa* mutants were inoculated into 100 ml of fresh HI medium and grown at 37°C with vigorous shaking to an optical density at 600 nm of 0.5. The cells were chilled briefly on ice and centrifuged. The pellet was washed twice with ice-cold water and resuspended in an equal volume of sterile 10% (vol/vol) glycerol-water.

Twenty-five microliters of the cells was electroporated (330 μF, 4 KΩ, rapid discharge) with 10 ng of DNA of the plasmid library by using a Cell Porater (Gibco BRL), incubated in 1 ml of HI medium at 37°C for 4 h with shaking, and then spread on HI agar plates containing ampicillin (50 μg/ml) and novobiocin (100 μg/ml). Plasmid DNA was purified from each transformant and retrans-

TABLE 1. Bacterial strains and plasmids used in this study

Strain or plasmid	Genotype or characteristic(s)[a]	Source or reference
H. influenzae		
2019	Wild type	M. A. Apicella
DK1	*rfaD1* Kan^r	This study
FK1	*rfaF1* Kan^r	This study
3031	Wild type	M. A. Apicella
3198	Wild type	M. A. Apicella
5987	Wild type	M. A. Apicella
7052	Wild type	M. A. Apicella
Eagan	Wild type	Robert Munson
A2	Wild-type type b	M. A. Apicella
A4	Wild-type type b	M. A. Apicella
A9	Wild-type type b	M. A. Apicella
KW20	Wild-type Rd	25
S. typhimurium LT2		
SL3770	rfa^+ wild type	SGSC[b]
SA3789	*rfaF*	SGSC (23)
SA1377	*rfaC*	SGSC
SA3600	*rfaD*	SGSC (14)
E. coli K-12		
DH5α	*hsdR17* (r_k^- m_k^+) *supE44 thi-1 recA1 gyrA* (Nal^r) *relA1* Δ(*lacZYA-argF*) *U169 deoR* [ϕ80dlacΔ(lacA)M15]	Gibco BRL
HAK117	Increased mRNA half-life	16
Plasmids		
pUC19	Cloning vector, Amp^r	Gibco BRL
pSH*lox*-1	Cloning vector, Amp^r	Novagen
pBluescript II SK−	Cloning vector, Amp^r	Stratagene
pGB19	Cloning vector, Cm^r	Gerald Barcak
pD41	4-kb *Apo*I fragment from 2019 cloned into pSH*lox*-1	This study
pDEP	1.6-kb *Eco*RI-*Pst*I fragment from pD41 cloned into pUC19	This study
pDPP	1.4-kb *Pst*I fragment from pD41 cloned into pUC19	This study
pDHP	1.5-kb *Pst*I-*Hin*dIII fragment from pD41 cloned into pUC19	This study
pDMN1	2.4-kb *Mlu*I-*Nsi*I fragment from pD41 into pBluescript SK II−	This study
pDDK1	pDMNI with a Kan^r cassette cloned into the *Pst*I site	This study
pDFK1	pDHP with a Kan^r cassette cloned into the *Mlu*I site	This study
pDEXO3	ExoIII deletion of pD41	This study
pDEXO5	ExoIII deletion of pD41	This study
pDEXO6	ExoIII deletion of pD41	This study
pDEXO16	ExoIII deletion of pD41	This study

[a] Abbreviations for antibiotics: Amp, ampicillin; Cm, chloramphenicol; Kan, kanamycin.

[b] Salmonella Genetic Stock Center, University of Calgary, Calgary, Alberta, Canada.

formed to the mutant strain to confirm the complementation. The transformants were also tested for phage sensitivities and LPS phenotype.

DNA sequencing and analysis. DNA sequence was determined by the dideoxy-chain termination method (21), using a Sequenase version 2.0 kit (U.S. Biochemical). Primers were commercially available or synthesized with a model 381A DNA synthesizer (Applied Biosystems, Inc.). DNA sequence was analyzed with the Genetics Computer Group package (University of Wisconsin) and GeneWorks software (IntelliGenetics, Inc.). Protein sequence alignment was done using the Bestfit program of the Genetics Computer Group sequence analysis software.

Genomic Southern hybridization of NtHi 2019 and isogenic *rfa* mutants. NtHi 2019 genomic DNA was digested with restriction enzymes, resolved on a 1% agarose gel, transferred to a Hybond-N membrane (Amersham) by capillary blotting overnight, and cross-linked to the membrane by using a Stratalinker (Stratagene). After prehybridization, the membrane was hybridized with a digoxigenin-dUTP-labeled DNA probe and washed. The hybridized probe was detected by using a DIG Luminescent Detection kit (Boehringer Mannheim Biochemicals), and the membrane was exposed to a Kodak X-Omat AR film at room temperature.

LPS gel analysis. LPS of *S. typhimurium* and LOS of *H. influenzae* were prepared by extraction with phenol and precipitation in ethanol as described previously (3), separated on a sodium dodecyl sulfate (SDS)–14% polyacrylamide gel (12), and visualized by silver staining as described previously (24).

Primer extension analysis. RNA was extracted from NtHi 2019 or from *E. coli* HAK117 and HAK117 bearing recombinant plasmids. *E. coli* HAK117 (16) has a mutation which increases the chemical half-life of mRNA. The NtHi cells were grown in sBHI broth, and the *E. coli* cells were grown in LB to an optical density at 600 nm of 0.6. The cells were disrupted by sonication, digested with proteinase K, and extracted with phenol-chloroform followed by ethanol precipitation as described previously (1). The purified RNA was quantitated spectrophotometrically, and the quality of RNA was confirmed on a formaldehyde-agarose gel stained with ethidium bromide. To ensure that extension was due to RNA and not DNA contamination, samples of RNA were incubated with heat-treated RNase (0.5 μg/μl) at 37°C for 1 h, and this treated RNA was used as a control for primer extension analysis.

Primer extension analysis was carried out by using a Promega Primer Extension kit as instructed by the manufacturer except for annealing, which was done by heating the reaction mixture to 70°C, incubating it at 60°C for 20 min, and then slowly cooling it to room temperature. Two micrograms of RNA was used for each reaction, and the reaction products were precipitated in ethanol, dissolved in loading dye, and loaded on a 6% sequencing gel. The dideoxy sequencing ladder with the same template and primer was used as a marker to confirm the position of the primer-extended products.

In vitro transcription-translation analysis. Purified plasmid DNA was used for in vitro transcription-translation (Promega). The translation products were labeled by using translation-grade [^{35}S]methionine (Amersham). The reaction mixtures were separated on an SDS–11% polyacrylamide gel (12), which was dried and exposed to an X-ray film. ^{14}C-labeled protein molecular weight Rainbow standards (Amersham) were used to generate a linear regression curve between 14.3 and 200 kDa for determination of relative molecular weights.

Mutagenesis. The Kan^r cassettes were cloned into restriction sites of the *rfaD* and *rfaF* genes. The Kan^r cassette used was isolated from plasmid pUC4K (Gibco BRL) and possesses its own promoter. The *rfaD* mutation was constructed by inserting the Kan^r marker into the *Pst*I site of pDMN1. The *rfaF* mutation was constructed by inserting the Kan^r marker into the *Mlu*I site of pDHP. NtHi genomic mutations were then constructed by the excision of the insert containing the Kan^r cassette from the plasmid, and this was then transformed into NtHi 2019 by using the MIV competence induction method described by Herriot et al. (9). Transformants were selected on an sBHI agar plate containing 15 μg of kanamycin per ml at 30°C. The locations of cassette insertion sites were confirmed by genomic Southern hybridization.

MS studies. To determine the structures of the oligosaccharide regions of LOS from strains DK1 and FKI, LOS were O-deacylated with hydrazine to form water-soluble forms that are amenable to mass spectrometry (MS) analysis. To prepare O-deacylated LOS, ca. 1-mg samples of LOS from strains DK1 and FK1 were treated with hydrazine at 37°C for 2 h and precipitated with acetone as previously described (5). To determine the molecular masses of the LOS glycoforms contained in the two strains, the O-deacylated LOS samples were analyzed by electrospray ionization MS (ESI-MS). For ESI-MS analysis, the O-deacylated LOS from each strain was resuspended up in 100 μl of water, and a 1% aliquot (1 μl) was mixed with 4 μl of acetonitrile-water containing 1% acetic acid and delivered via a Rheodyne injector to the ESI source. Spectra were acquired in the negative-ion mode on a Quattro quadrupole mass spectrometer (Fison, Manchester, England) running at a flow rate of 5 μl/min over the mass range *m/z* 450 to 1900. Spectra were mass calibrated with an external reference. Details of these procedures have been described elsewhere (5).

Chromatographic analysis of LOS from strains DK1 and FK1. To determine the configuration of the single heptose residue in the DK1 and FK1 LOS, an aliquot of each O-deacylated LOS sample (~100 μg) was HF treated (48% aqueous HF, 12 h at 3°C) and then hydrolyzed into its constituent monosaccharides by using standard condition: 4 h at 100°C in 4 N HCl. An aliquot of each hydrolysate was first analyzed by high-pH anion-exchange chromatography using a Dionex high-pressure liquid chromatography system equipped with a PA-1 column and pulsed amperometric detector operating under conditions previously described (19). In addition, a second aliquot from each of the LOS hydrolysates (~100 μg) was reduced, peracetylated, and analyzed by gas chromatography (GC)-MS using a procedure previously employed to determine the relative amounts of D-*glycero*-D-*manno*-heptose (DD-heptose) and (LD-heptose) in the LOS from *Haemophilus ducreyi* 35000 (15). An alditol acetate standard prepared from an LPS hydrolysate of *Aeromonas liquefaciens* SJ19 (kindly provided by G. Aspinal, York University, North York, Ontario, Canada) containing both LD- and DD-heptose was used to determine the precise retention times of these sugars as well as that of D-glucosamine.

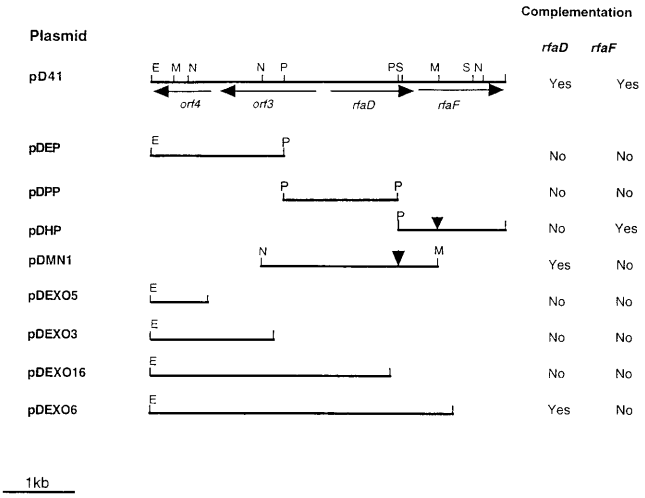

FIG. 2. Physical maps of parental plasmid pD41 and derivative subclones. The insert of pD41 is a 4.8-kb *Apo*I fragment. Also shown are the results of complementation studies of *S. typhimurium rfaD* and *rfaF* mutants. Triangles indicate the relative positions of the Kan^r cassette insertions to construct mutants in the *rfaD* and *rfaF* genes. Abbreviations: E, *Eco*RI; M, *Mlu*I; N, *Nsi*I; P, *Pst*I; S, *Sty*I.

Determination of distance between *rfaD* and *rfaF* genes of *H. influenzae* strains. The distance between genes was initially characterized by PCR amplification. Primers were designed from the region encoding the C terminus of the *rfaD* gene and the N terminus of the *rfaF* gene and called D17LE and D18LE, respectively. The primer sequences are GGTGGTGCAGGTTTTATTGGCAG and CTAATTCAAAGGCCCCGTGTCC, respectively. PCR was performed by using an Expand PCR kit (Boehringer Mannheim Biochemicals) with the genomic DNA from nine *H. influenzae* strains. The region between the *rfaD* and *rfaF* genes was cloned from five strains and sequenced.

Nucleotide sequence accession number. The DNA sequences of the *rfaD* and *rfaF* genes and the adjacent regions have been submitted to GenBank and assigned accession number L76100.

RESULTS

Identification of a plasmid bearing the *rfaD* and *rfaF* genes of NtHi 2019. A plasmid carrying the *rfaD* gene was isolated by complementation of the *S. typhimurium rfaD* mutant strain SA3600. The NtHi 2019 plasmid library DNA was electroporated into strain SA3600. This strain lacks the ADP-L-*glycero*-D-*manno*-heptose-6-epimerase activity and synthesizes a truncated LPS molecule. Deep rough mutants like SA3600 present a more hydrophobic surface and are thus more susceptible to hydrophobic antibiotics such as novobiocin. Transformants carrying recombinant plasmids which bear the gene required for complementation were selected on plates supplemented with novobiocin, and the plasmids were isolated and purified. The purified plasmids were again electroporated into *S. typhimurium* SA3600 to confirm the ability of this plasmid to complement the strain. The plasmid was again purified and subjected to restriction endonuclease characterization. Restriction mapping demonstrated that the plasmid which complemented

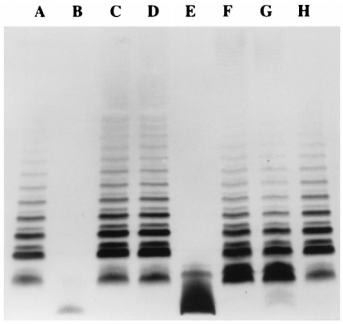

FIG. 3. Silver-stained LPS gel demonstrating complementation of *S. typhimurium rfaD* and *rfaF* mutants by plasmids bearing the *H. influenzae rfaD* and *rfaF* genes. Lane A, wild-type *S. typhimurium* LPS; lane B, *rfaD* mutant SA3600; lane C, SA3600 bearing pD41; lane D, SA3600 bearing pDMN1; lane E, *rfaF* mutant SA3789; lane F, SA3789 bearing pD41; lane G, SA3789 bearing pDHP; lane H, wild-type *S. typhimurium* LPS.

the *rfaD* mutant SA3600 contained an insert of approximately 4.8 kb, and this plasmid was named pD41 (Fig. 2).

It was known that in *S. typhimurium* and *E. coli*, the *rfaD*, *rfaF*, and *rfaC* genes were contiguous on the bacterial chromosome (22); therefore, plasmid pD41 was used in a manner similar to that described above in an attempt to complement the *rfaF* and *rfaC* mutant strains of *S. typhimurium*. Strain SA3789 bears a mutation in the *rfaF* determinant and lacks heptosyltransferase II activity, which attaches the second heptose molecule to the proximal heptose (Fig. 1). Plasmid pD41, when transformed into SA3789, was able to complement the *rfaF* phenotype back to the wild type. This finding suggested that the *rfaF* gene of *H. influenzae* was also present on the insert of pD41. However, pD41 was unable to complement the *rfaC* strain *S. typhimurium* SA1377, and this finding suggested that an intact copy of the *rfaC* gene was not present on pD41.

Subcloning of plasmid pD41. To localize the *rfaD* and *rfaF* genes to smaller regions on pD41, we constructed a series of subclones (Fig. 2). Plasmids pDEP and pDPP were constructed by cloning the 1.6-kb *Eco*RI-*Pst*I and 1.4-kb *Pst*I fragments, respectively, from pD41 into pUC19. Plasmid pDHP was constructed by cloning the 1.5-kb *Hin*dIII-*Pst*I fragment into *Pst*I-*Hin*dIII sites of pUC19. Plasmid pDMN1 was constructed using by cloning the 2.4-kb *Mlu*I-*Nsi*I fragment from pD41 into pBluescript SK II$^-$. The *Mlu*I site used for this clone is located in the vector sequence, and therefore the resultant fragment contains a small amount of vector DNA. In addition, a series of exonuclease III (ExoIII) deletions were constructed from plasmid pD41, and these constructs are also presented in Fig. 2.

The ability of these recombinant plasmids to complement the *rfaD* and *rfaF* mutants of *S. typhimurium* was examined, and this information is included in Fig. 2. Plasmids pD41, pDMN1, and pDEXO6 were able to complement the *rfaD* mutant strain SA3600, which suggested that the *rfaD* gene of *H. influenzae* is located on the region located on all of these plasmids and therefore was present on the 2.4-kb insert of plasmid pDMN1, which is the smallest of the three plasmids. Only plasmids pD41 and pDHP were able to complement the *rfaF* mutant strain SA3789. This evidence suggested that an intact copy of *rfaF* is present on pDHP.

Characterization of the LPS of *S. typhimurium rfa* mutants complemented with pD41 and derivative plasmids. The LPS was purified from *S. typhimurium rfa* mutant strains and from the same strains complemented with plasmids bearing the *rfaD* and *rfaF* genes of *H. influenzae*. The LPS was analyzed by SDS-polyacrylamide gel electrophoresis (PAGE) followed by silver staining (Fig. 3). The rfaD mutant strain *S. typhimurium* SA3600 expresses a highly truncated LPS molecule, and this molecule corresponds to the major band present on the gel. The LPS from SA3600 transformed with either pD41 or pDMN1 shows an upward shift on the gel corresponding to an increase in molecular mass. In addition, the major band from the mutant, which corresponds to the highly truncated LPS molecule, is not present in significant amounts in the lanes with LPS from the complemented strain. This finding is consistent with the complementation studies which demonstrated that both pD41 and pDMN1 could complement SA3600 to novobiocin resistance.

Similar studies were performed with the *rfaF* mutant strain SA3789, which also expresses a truncated form of LPS that is present as the major band on an SDS-polyacrylamide gel. This strain, when transformed with pD41, expresses an LPS of higher molecular weight which includes the addition of O-antigen groups. The LPS from SA3789/pD41 also had very little of the low-molecular-weight band which corresponded to the truncated mutant form. This finding demonstrates that there is not complete complementation of the mutant phenotype. SA3789 transformed with plasmid pDHP, however, demonstrated an appreciable amount of the truncated LPS still present, which suggests that plasmid pDHP is somehow less effective than pD41 in its ability to complement SA3789.

DNA sequence analysis of the *rfaD* and *rfaF* genes of NtHi. Nucleotide sequence examination of the 4.8-kb insert of pD41 demonstrated the presence of four open reading frames (ORFs). The relative locations of these ORFs are shown in Fig. 2. Two of the ORFs corresponded well with the predicted locations of genes which encode the ADP-L-*glycero*-D-*manno*-heptose-6-epimerase and heptosyltransferase II, based on the ability of subcloned plasmids to complement strains SA3600 and SA3789, and these ORFs were named *rfaD* and *rfaF*, respectively. It should be noted that the entire *rfaD* ORF is present on pDMN1 and that the entire *rfaF* ORF is present on pDHP. Any complementation facilitated by these plasmids is due to the presence of a complete gene on the plasmid rather than truncated determinants.

The ORF of *rfaD* encodes a polypeptide of 308 amino acids which has a predicted molecular mass of 34.8 kDa. The predicted amino acid sequence of the *H. influenzae* RfaD protein was compared to that of the RfaD proteins from *S. typhimurium* and *E. coli* and was found to be 74.5% identical to these proteins. An acceptable Shine-Dalgarno sequence is present 6 bases upstream of the codon of translation initiation.

The ORF of *rfaF* hypothetically encodes a protein of 346 amino acids with a predicted molecular mass of 38.9 kDa. The predicted amino acid sequence of the RfaF protein was compared to that of the RfaF protein of *S. typhimurium*, and the two were found to be 64% identical. There is also an acceptable Shine-Dalgarno sequence present 6 bases from the initial ATG codon of *rfaF*.

The nucleotide sequences of the *rfaD* and *rfaF* genes were 96 and 95% identical to the HI1114 and HI1105 ORFs of the TIGR *H. influenzae* Rd genome database (2). Not surprisingly, these genes from Rd demonstrate the strongest identity to the *rfaD* and *rfaF* genes, respectively, from *E. coli*. The strong identity between these sequences supports the complementation data which suggested that the *rfaD* and *rfaF* genes of *H. influenzae* encoded the ADP-L-*glycero*-D-*manno*-heptose-6-epimerase and heptosyltransferase II enzymes, respectively, involved in LOS biosynthesis.

Construction and characterization of *H. influenzae rfaD* and *rfaF* mutant strains. To better verify that the *rfaD* and *rfaF* genes of *H. influenzae* encode the ADP-L-*glycero*-D-*manno*-heptose-6-epimerase and heptosyltransferase II proteins, respectively, it was desirable to construct *rfaD* and *rfaF* mutants

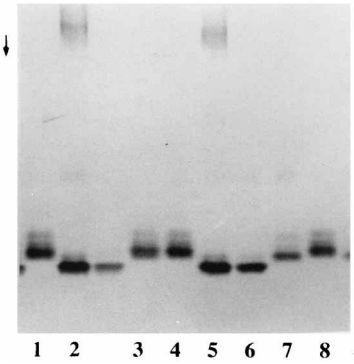

FIG. 4. Silver-stained SDS-polyacrylamide gel of LOS from NtHi 2019 isogenic *rfaD* mutant DK1 and *rfaF* mutant FK1. Lane 1, 2019 (wild-type) LOS; lane 2, *rfaD* mutant DK1; lane 3, DK1 transformed with pDMN1; lane 4, 2019 LOS; lane 5, *rfaF* mutant FK1; lane 6, FK1 bearing pDMN1; lane 7, FK1 bearing pDHP; lane 8, 2019 LOS. The arrow on the left shows the direction of band migration through the gel.

in NtHi 2019. This was accomplished by the incorporation of an intact Kan^r gene into the *rfaD* and *rfaF* genes. The locations of the Kan^r gene insertions into the *rfaD* and *rfaF* genes were confirmed by DNA sequencing and are presented in Fig. 2. The plasmid bearing an insertion into the *rfaD* gene was called pDDK1, and the plasmid with the insertion in the *rfaF* gene was called pDFK1. To verify that the insertion of the Kan^r cassette effectively eliminated the ability of the genes to express a functional product, complementation studies were performed with plasmids pDDK1 and pDFK1. Plasmid pDDK1 was unable to complement SA3600 to novobiocin resistance; likewise, pDFK1 was unable to complement the *rfaF* mutant SA3789. This result demonstrated that the insertion of the Kan^r cassette was eliminating gene expression of the desired genes.

Genomic insertion mutants were next constructed by transforming NtHi 2019 with the excised inserts from pDDK1 and pDFK1 and screening for Kan^r ampicillin-susceptible colonies. The strain which incorporated the Kan^r gene in the *rfaD* gene was called DK1, and the strain which incorporated the Kan^r marker in the *rfaF* gene was called FK1. The location of the insertion into the appropriate genes was verified by genomic Southern hybridization (data not shown). It had to be demonstrated that the plasmids bearing the *rfaD* and *rfaF* genes could complement the genomic mutations in *trans.*

H. influenzae DK1 and FK1 were transformed with pDMN1 and pDHP, respectively, and the LOS from these strains was characterized by SDS-PAGE followed by silver staining. The results of this experiment are shown in Fig. 4. The LOS of wild-type strain 2019 has a simple profile compared to that of *S. typhimurium* (Fig. 3) due to the lack of repeating O-antigen groups on the *H. influenzae* LOS. The *rfaD* mutant DK1 expresses an LOS which is truncated and thus migrates further down the gel than the wild type. Strain DK1 bearing plasmid pDMN1 expresses an LOS which has migration characteristics identical to those of the wild type on this gel. This finding verifies that the *rfaD* gene from NtHi 2019 is able to complement a genomic mutation in *trans.*

Strain FK1 also expresses a truncated form of LOS as expected, which demonstrates that the *rfaF* gene is involved in LOS biosynthesis. FK1 transformed with plasmids pDHP and pDMN1 expresses an LOS which has migration characteristics identical to those of the wild-type LOS expressed by 2019. In addition, no bands corresponding to the mutant LOS are present in FK1 complemented with pDHP. This finding suggested that complementation by pDHP is complete in NtHi when similar experiments in *S. typhimurium* yielded partial complementation.

Identification of the *rfaD* and *rfaF* gene products. In vitro transcription-translation studies were performed on pD41 and derivative plasmids to better characterize the products of the *rfaD* and *rfaF* genes (Fig. 5). Plasmid pDMN1 encoded a polypeptide with a molecular mass of 35 kDa (lane 3), which corresponds well to the predicted mass of 34.8 kDa for the *rfaD* gene product. Plasmid pDDK1 (lane 4) expressed a protein with the 29.5-kDa molecular mass of an intact Kan^r gene product as expected. Plasmid pDDK1 also expressed a protein of approximately 28 kDa, which is consistent with the predicted mass of the product of the interrupted *rfaD* gene, suggesting that the Kan^r cassette had disrupted the *rfaD* gene. Plasmid pDHP (lane 5) expressed very little gene product. The insert in pDHP is cloned in such a manner that the *rfaF* gene is transcribed in the opposite orientation to the vector-borne promoter. In this orientation, pDHP was still able to complement both *S. typhimurium* and *H. influenzae rfaF* mutants. To better facilitate expression, the insert of pDHP was cloned downstream of the vector promoter of pGB19. The *rfaF* gene was transcribed in the same orientation as the vector promoter and therefore was transcribed from that promoter. The vector pGB19 expressed the chloramphenicol acetyltransferase (CAT) gene product (lane 6), and pDHPCm expressed the CAT gene product and the 39-kDa RfaF polypeptide. The molecular mass of the gene product is very close to the predicted mass of 38.9 kDa for the RfaF polypeptide. Finally, plasmid pD41 expressed both the RfaD and RfaF gene products. This finding suggests that expression of *rfaF* is somehow increased in the presence of the *rfaD* gene sequence.

Identification of transcription initiation sites. Primer extension studies were performed on the *rfaD* and *rfaF* genes. RNA was isolated initially from NtHi 2019 and used for the studies. With 2019 RNA, it was possible to determine a strong point of transcription initiation upstream of the *rfaD* gene (Fig. 6A). The strong band in the sample lanes corresponds with one base in the triplet CAG which is centered 27 bases upstream of the translation initiation codon of the *rfaD* gene. Further examination of this region demonstrated the presence of possible

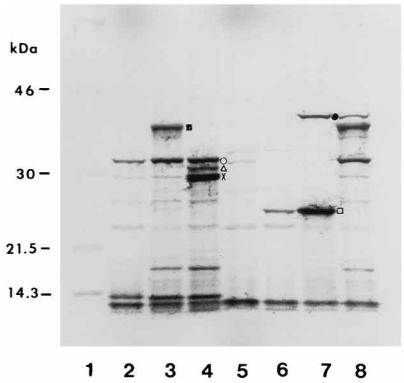

FIG. 5. In vitro transcription-translation of the NtHi 2019 *rfaD* and *rfaF* genes. Lane 1, protein molecular weight standards; lane 2, pUC19 expressed a 31.5-kDa β-lactamase band; lane 3, pDMN1 expressed a 35-kDa polypeptide (■) which is the *rfaD* gene product; lane 4, pDDK1 expressed the β-lactamase gene product (○), the 29.5-kDa kanamycin resistance gene product (△) and a truncated *rfaD* gene product (X); lane 5, pDHP, which bears the *rfaF* gene, produced no strong bands; however, the 31.5-kDa β-lactamase band is visible; lane 6, vector plasmid pGB19 expressed a 25-kDa CAT protein; lane 7, pDHPCm, which bears the *rfaF* gene under the control of a vector promoter from pGB19, produced the CAT protein (□) and also a 39-kDa polypeptide corresponding to the *rfaF* gene product (●); lane 8, pD41 expressed strong bands for both the *rfaD* and *rfaF* gene products.

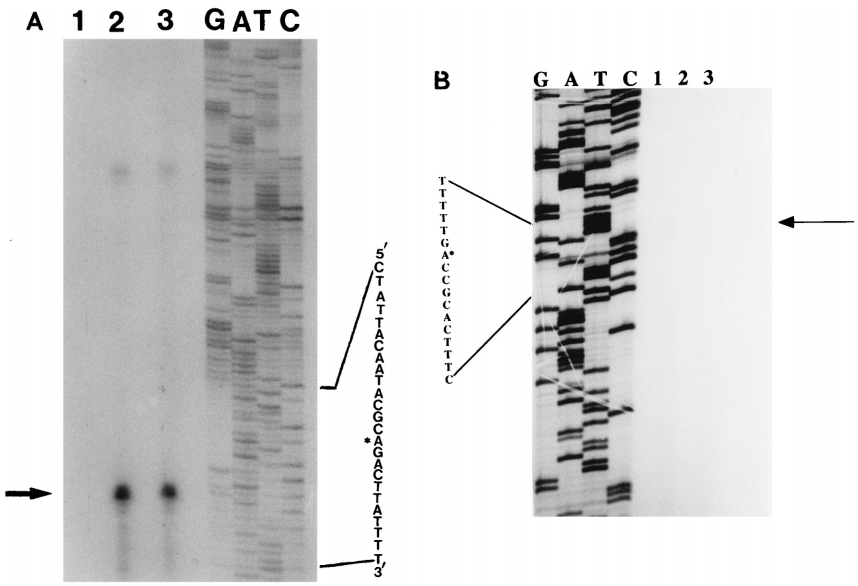

FIG. 6. Primer extension gels of the *rfaD* (A) and *rfaF* (B) genes of *H. influenzae*. (A) Primer extension analysis of the *rfaD* gene. Lane 1, no-RNA control; lane 2, RNA isolated from cells grown at 30°C; lane 3, RNA isolated from cells grown at 37°C. The transcription start site is denoted by an asterisk in the sequence on the right and by the lower arrow on the left. A faint, second band which could suggest a weak secondary start site is indicated by the upper arrow. The DNA sequencing ladder was obtained by using the same primer as that in the primer extension study. (B) Primer extension analysis of the *rfaF* gene. Lane 1, RNA isolated from cells grown at 30°C; lane 2, RNA isolated from cells grown at 37°C; lane 3, no-RNA control. Only a very weak band was visible on this gel, which suggests that *rfaF* may be transcribed from the *rfaD* promoter. It is also possible that the *rfaF* promoter is very weak.

−10 and −35 promoter sequences. The putative regulatory region of the *rfaD* gene is shown in Fig. 7A.

Primer extension studies using NtHi 2019 RNA to identify the *rfaF* promoter yielded no bands. In an attempt to purify a larger quantity of mRNA, RNA was isolated from the *E. coli* HAK117 bearing plasmid pDHP. HAK117 is a mutant strain which degrades mRNA at a rate dramatically lower than the wild-type strain. RNA isolated from HAK117/pDHP was used to examine possible transcription initiation sites of the *rfaF* gene (Fig. 6B). These studies yielded only one very weak band which corresponds to the nucleotide triplet GAC centered 27 bases upstream of the ATG codon for *rfaF*. However, examination of the nucleotide sequence upstream of the putative transcription initiation sequence yielded little in the way of candidates for −10 or −35 promoter sequences. The DNA nucleotide sequence of this region is shown in Fig. 7B. While the evidence from primer extension studies does not strongly support the presence of an independent promoter for the *rfaF* gene, the complementation studies using pDHP demonstrate that the *rfaF* gene is capable of expressing a functional gene product even when a vector-borne promoter is not present upstream of the gene. This finding suggests an independent promoter is present for the *rfaF* gene on plasmid pDHP.

Analysis of genetic organization of *rfaD* and *rfaF* on the *H. influenzae* genome. As mentioned previously, the *rfaD*, *rfaF*, and *rfaC* genes are contiguous on the *S. typhimurium* and *E. coli* genomes, and in both cases all three genes are transcribed in the same orientation. Nucleotide sequence analysis of the ORF immediately downstream of *rfaF* suggested that this ORF is not the rfaC gene; in addition, this ORF is transcribed in the opposite orientation to rfaF. Independent work in our lab has yielded a likely candidate for the *H. influenzae rfaC* gene, and this gene is not adjacent to the *rfaD-rfaF* cluster in 2019. The recently published *H. influenzae* Rd genome database (2) has mapped the *rfaD* and *rfaF* genes to the same general region of the Rd chromosome; however, there are eight genes present between the two. These differences suggest that the organization of the *rfaD*, *rfaF*, and *rfaC* genes appears to differ in the *E. coli*, *S. typhimurium*, *H. influenzae* Rd, and NtHi 2019 genomes. To better characterize the presence or absence of genes located between *rfaD* and *rfaF* from other *H. influenzae* strains, PCR amplification and DNA Southern hybridization studies were performed. PCR amplification of DNA between the *rfaD* and *rfaF* genes demonstrated that all strains tested, excluding the Rd strain KW20, have a small region of approximately 85 bases separating the two genes. The Rd strain KW20 possesses dramatically larger intervening sequences approximately 11 kb in length. Results of these experiments are shown in Table 2. Southern blotting experiments demonstrated that the PCR products all hybridized to a 530-bp DNA probe which

A)

-35 -10 * S.D.

rfaD TTTTTTGGCTAAGCGGTAGAAAATCTATTACAATACGCAGACTTATTTTAAAGAGGACAAGAATT**ATG**

B)

* S.D.

rfaF TTCCCATCAACAAAGTGCGGTTAAAAATCTCGGTGTTTTTGACCGCACTTTCATCTGAAAGACAGAAA**ATG**

FIG. 7. Upstream regions of the *rfaD* (A) and *rfaF* (B) genes. Putative regulatory regions are indicated: −35 and −10 promoter sequences are underlined; transcription initiation sites are indicated by asterisks; ribosomal binding (Shine-Dalgarno) sites are underlined and indicated by S.D.; translation initiation codons are in boldface.

TABLE 2. Distance between rfaD and rfaF genes of several *H. influenzae* strains

Strain	Size (kb) of PCR product[a]	Distance (bp) by sequence[b]
2019	1.1	85
3031	1.1	ND
3198	1.1	85
5987	1.1	85
7052	1.1	85
Eagan	1.1	ND
A2	1.1	85
A4	1.1	ND
A9	1.1	ND
KW20	12	~11,000[c]

[a] Size of amplified PCR product between primers from rfaD and rfaF genes.
[b] Distance as determined from sequences of region between rfaD and rfaF genes. ND, not determined.
[c] The distance between rfaD and rfaF was calculated from the *H. influenzae* Rd database (2).

is comprised of DNA from strain 2019. The probe contains sequences from *rfaD* and *rfaF* genes and the intergenic sequences; however, the probe does not contain the sequences used in the construction of the PCR primers. Hybridization of the probe demonstrated that the amplified fragments were specific and that the fragments amplified from all of the strains bore sequence similarity with the 2019 sequences.

Analysis of LOS of *H. influenzae* DK1 and FK1. SDS-PAGE analysis of the LOS isolated from *H. influenzae* DK1 and FK1 was performed. The LOS of these strains migrated at similar rates, suggesting similar molecular masses of the LOS molecules. It was expected that the LOS from the *rfaD* mutant strain DK1 would possess one fewer heptose and would thus have a lower molecular mass. To verify the findings of the SDS-PAGE analysis, MS analysis was performed. MS analyses of the O-deacylated preparations from DK1 and FK1 yielded very similar spectra (Fig. 8). In both cases, the dominant peaks were seen at *m/z* 721 and 783. These two peaks correspond to the doubly deprotonated molecular ions, $(M - 2H)^{2-}$, for LOS glycoforms with average molecular masses of 1445.6 or 1445.8 (LOS-A) and 1568.4 or 1568.6 (LOS-B) from DK1 and FK1, respectively (Table 3). These masses differ by the mass of phosphoethanolamine (PEA) (Δm = 123 Da) and are in good agreement for LOS species containing a conserved oligosaccharide moiety consisting of a single heptose linked to a phospho-KDO moiety which is in turn attached to the expected diphosphoryl *O*-deacyl lipid A. After deacylation, the lipid A is converted to a *N,N′*-diacyl form in which each of the two glucosamine sugars contains a single N-linked β-hydroxymyristic acid group:

$$\text{Hep}\rightarrow\text{KDO(P)(PEA)}_{0\text{-}1}\rightarrow\text{(P)GlcNH}\text{———}\text{GlcNH(P)}$$
$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad |\qquad\qquad |$$
$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad C_{14}H_{27}O_2\qquad C_{14}H_{27}O_2$$

where M_r = 1,445.3 (LOS-A) and 1,568.4 (LOS-B). Other peaks are also present in these two spectra and can be readily

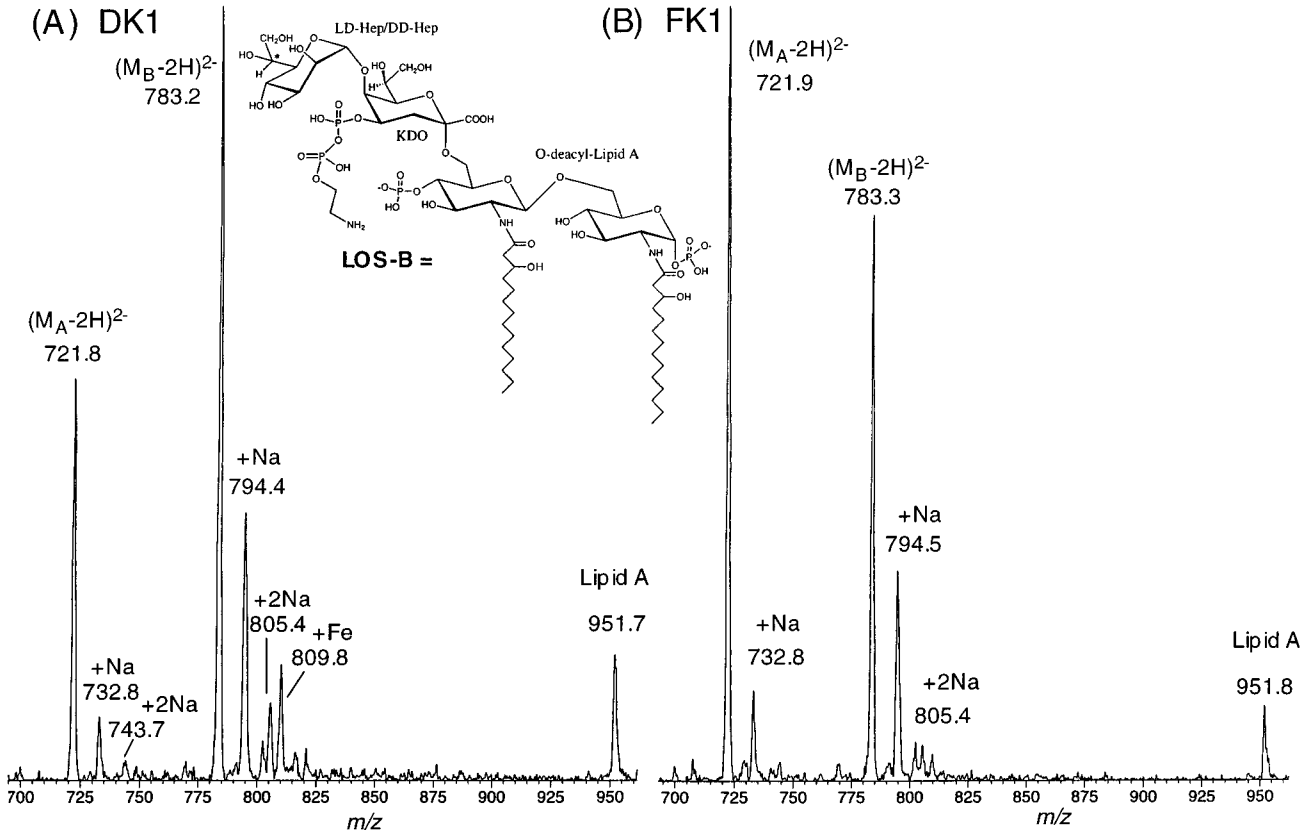

FIG. 8. Negative-ion ESI-MS spectra of O-deacylated LOS from *H. influenzae* mutants DK1 (A) and FK1 (B). The doubly charged LOS glycoforms LOS-A and LOS-B are labeled, and their cationized adducts are indicated. See Table 3 for masses and proposed compositions of these LOS. The peak at *m/z* 952 in both spectra is a gas-phase fragment corresponding to the mass of the singly charged *O*-deacyl lipid. The structures shown in insets are the likely configuration and linkages of the two major O-deacylated LOS-B glycoforms as determined from data in this study and from previous work on the parent strain *H. influenzae* A2 (18, 20) and mutant strain I-69 Rd−/b+ (8). The position of the variable PEA moiety is likely on the phosphate of KDO by homology to *H. ducreyi* LOS (unpublished data and reference 4).

TABLE 3. ESI-MS analyses of O-deacylated LOS[a]

Mutant strain	M_r		Relative % abundance	Proposed composition
	Observed	Calculated		
DK1	951.7	952.0	ND[b]	*N*,*N*′-diacyl lipid A[c]
	1,445.6	1,445.3	37	Hep-KDO(P)-O-deacyl-lipid A
	1,568.4	1,568.4	100	Hep-KDO(P)(PEA)-O-deacyl-lipid A
FK1	951.8	1,253.2	ND	KDO(P)-O-deacyl-lipid A
	1,445.8	1,445.3	95	Hep-KDO(P)-O-deacyl-lipid A
	1,568.6	1,568.4	100	Hep-KDO(P)(PEA)-O-deacyl-lipid A

[a] All average molecular weights and relative abundances for the different O-deacylated LOS-glycoforms are based on the doubly deprotonated charged molecular ions, $(M - 2H)^{2-}$, and in the case of relative ion abundances, the abundances of the sodiated molecular ion species, i.e., $(M - 2H + Na)^-$, $(M - 3H + 2Na)^-$, etc., were included.

[b] ND, not determined (the lipid A peak at m/z ~952 is a gas-phase fragment and not a molecular ion).

[c] After O-deacylation, the lipid A moiety is converted into diphosphoryl *N*,*N*′-diacyl lipid A containing two N-linked β-hydroxymyristic acid chains with an average M_r of 953.0.

assigned as salt adducts to the two major LOS glycoforms, primarily sodiated species, $(M + Na - 3H)^{2-}$ and $(M + 2Na - 4H)^{2-}$, although small amounts of potassium and iron adducts were also found for the higher-mass LOS-B glycoform. At m/z ~952, a peak is seen in both spectra for the single-charged *N*,*N*′-diacyl diphosphoryl lipid A species (calculated m/z = 952.0), and the abundance of this peak was highly dependent on the source voltages. Such dependence is usually indicative of an in-source gas-phase degradative product. Indeed, the fragmentation pathway leading to this peak has been reported previously and was assigned as originating from the cleavage of the labile KDO-lipid A glycosidic bond (6). Moreover, the absence of a peak at m/z 1052, which would be expected if the PEA moiety was on the lipid A moiety in the higher-mass LOS-B glycoforms, was not observed, therefore placing this nonstoichiometric PEA group in the oligosaccharide region (either on the phospho-KDO or on the heptose).

Although strain FK1 would be expected to contain a single core LD-heptose, strain DK1, which is defective in the heptosyl epimerase, would not be expected to contain any heptose at all. Nonetheless, the MS data clearly support the inclusion of a single heptose moiety based on the observed masses. It is possible, then, that the mutant strain is capable of inserting the biosynthetic precursor DD-heptose in lieu of any available LD-heptose at this position. To examine this latter possibility, both strains were subjected to monosaccharide composition analysis using high-pH anion-exchange chromatography. These data identified an abundant peak for D-glucosamine in both the DK1 and FK1 hydrolysates as expected from the conserved lipid A moieties (spectrum not shown). The LOS hydrolysate from FK1 also showed the presence of a later-eluting LD-heptose whereas no such peak was apparent in the hydrolysate of the LOS from strain DK1. As previously reported for *H. ducreyi* LOS, peracetylated derivatives of both LD- and DD-heptose yield stable derivative that are readily separated by GC. Therefore, to determine the nature of the heptose in the DK1 LOS, we converted a portion of the monosaccharide hydrolysates from both strains into their corresponding preacetylated derivatives and analyzed them by GC-MS using the LOS from *A. liquefaciens* SJ19, which contains roughly equal amounts of these two heptose isomers, as a control. As shown in Fig. 9, both the LOS DK1 and FK1 hydrolysates yielded two peaks: the expected peracetylated form of glucosamine and a smaller later-eluting peak at ca. 30 to 40% abundance. In FK1, the second component eluted at the retention time expected for alditol acetate of LD-heptose. In contrast, the analogous peak in the DK1 sample eluted much earlier and corresponded to the elution position for the alditol acetate of DD-heptose (Fig. 9).

DISCUSSION

The structural similarities of the inner core structures of the LPS of *S. typhimurium* and the LOS of *H. influenzae* have provided us with an excellent system for the selection of inner core biosynthesis genes from *H. influenzae*. The *rfaD* and *rfaF* genes of NtHi 2019 were selected from a genomic library by complementation of *S. typhimurium* deep rough LPS mutants. The recombinant plasmid pD41, which contained a 4.8-kb insert, was able to complement both *rfaD* strain SA3600 and *rfaF* strain SA3789 of *S. typhimurium* to novobiocin resistance. This finding suggested that both genes were present on the insert and that both genes function in *trans*. Derivative subclones and ExoIII deletions were constructed from pD41 and tested in the complementation system. A gene localized on plasmid pDMN1 was able to complement SA3600, and a gene on the plasmid pDHP was able to complement SA3789. In addition to the novobiocin resistance study, LPS characterization on SDS-polyacrylamide gels demonstrated that the complemented strains expressed LPS which appeared identical to the wild-type molecule. It was noted, however, that plasmid pDHP was unable to completely complement strain SA3789.

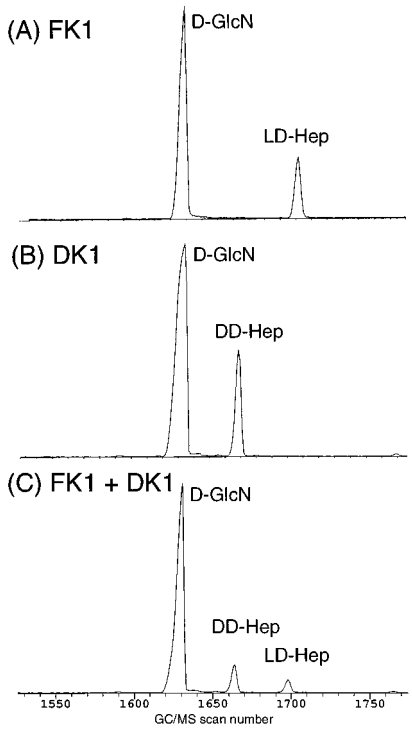

FIG. 9. Analysis of alditol acetates from the derivatized hydrolysates of LOS from strains FK1 (A) and DK1 (B) and coinjection of FK1 and DK1 (C). Chromatograms show GC-MS traces of the reduced and peracetylated monosaccharide derivatives. Note the presence of a peak corresponding to LD-heptose in the FK1 LOS hydrolysate and its absence in DK1, with an earlier-eluting peak in DK1 corresponding to the elution position of DD-heptose as determined by an external standard derived from the LPS of *A. liquefaciens* SJ19.

DNA nucleotide sequence analysis of plasmid pD41 yielded 4,751 bp of insert DNA. Four ORFs were present on the insert. Two ORFs were localized to positions consistent with the findings of the complementation studies. One intact ORF was found on pDMN1, and this gene was called *rfaD*. Another ORF was localized to the insert of pDHP, and this gene was named *rfaF*. The predicted molecular mass of the RfaD polypeptide was 34.8 kDa, and the predicted amino acid sequence of *rfaD* possesses a strong identity (74.5%) to the RfaD polypeptide of *E. coli*, which is the ADP-L-*glycero*-D-*manno*-heptose-6-epimerase involved in LOS biosynthesis. The predicted molecular mass of RfaF was 38.9 kDa, and the predicted amino acid sequence was 64% identical to the sequence of the RfaF protein of *E. coli*, which is the heptosyltransferase II enzyme. The nucleotide sequences of the *H. influenzae rfaD* and *rfaF* genes were 96 and 95% identical to the HI1114 and HI1105 genes from the Rd database. The HI1114 and HI1105 ORFs most resembled the *rfaD* and *rfaF* genes from *E. coli.*

While the results of the complementation studies and sequence analysis strongly suggested that the *rfaD* and *rfaF* genes of NtHi 2019 encoded the ADP-L-*glycero*-D-*manno*-heptose-6-epimerase and heptosyltransferase II enzymes, it was still important to construct isogenic mutants of these genes to examine the resultant phenotype. Such mutants were constructed by the insertion of a Kan^r cassette into the genes of interest on a recombinant plasmid. The insert DNA was then excised and transformed into NtHi 2019, where it was integrated into the bacterial chromosome. Strains DK1 and FK1 contained insertion mutations within the *rfaD* and *rfaF* genes, respectively.

In complementation studies performed with these isogenic mutants, plasmid pD41 was able to complement both strains to wild type. Plasmid pDMN1 was able to complement strain DK1, and pDHP complemented strain FK1, thus providing further evidence that these plasmids bear genes which can act in *trans.*

To better characterize the gene products of these genes, in vitro transcription-translation analysis was performed on plasmids bearing the *rfaD* and *rfaF* genes. The *rfaD* gene located on pDMN1 encodes a polypeptide of approximately 35 kDa, very similar to the predicted size of 34.8 kDa. The *rfaF* gene was not expressed well on plasmid pDHP; however, when the insert DNA was cloned downstream of a vector-borne promoter, the gene expressed large amounts of a 39-kDa protein. The size of this protein is also very close to the predicted molecular mass of 38.9 kDa. The parental plasmid pD41 expressed both RfaD and RfaF polypeptides. The ability of the *rfaF* gene to express large amounts of protein from pD41 while it is unable to do so from pDHP suggested that some sequences absent on pDHP are essential for high-level expression of *rfaF.*

Primer extension analysis demonstrated the presence of a strong transcriptional initiation sequence 27 bases upstream of the translation start site of *rfaD*. Further examination of the nucleotide sequences upstream of this transcription initiation site demonstrated the presence of sequences which resemble the −10 and −35 promoter sequences in both composition and location. Primer extension of the *rfaF* gene was much less conclusive. RNA isolated from *E. coli* HAK117 bearing pDHP was used for primer extension studies that suggested the presence of a very weak transcription initiation site 27 bases upstream of the ATG codon of *rfaF*. Examination of upstream sequences yielded no apparent −10 or −35 sequences.

While the primer extension studies were inconclusive for the presence of an independent promoter for the *rfaF* gene, complementation studies did suggest the existence of such a promoter. pDHP was able to complement *S. typhimurium* and NtHi *rfaF* mutant strains; however, in *Salmonella*, plasmid pD41 provided a more complete complementation. Also, pDHP expressed little if any RfaF protein in the in vitro transcription-translation studies, while pD41 expressed large amounts of the protein. We believe that these data suggest the presence upstream of the *rfaF* gene of a very weak promoter which allows low levels of expression alone. We also believe that sequences present in the *rfaD* gene are responsible for increasing the level of *rfaF* expression. It has previously been demonstrated (22) that in *E. coli* and *S. typhimurium*, the *rfaF* gene is transcribed from the *rfaD* gene promoter, and therefore it is possible that a similar operon functions in NtHi 2019. While these results provide strong evidence that the *rfaD* and *rfaF* genes of NtHi 2019 express the ADP-L-*glycero*-D-*manno*-heptose-6-epimerase and heptosyltransferase II enzymes, molecular analysis of the LOS purified from the isogenic strains would provide direct evidence of the functions of the gene products.

It was also discovered that the *rfaD*, *rfaF*, and *rfaC* genes are organized differently in *S. typhimurium*, *E. coli*, and *H. influenzae* strains. In *E. coli* and *S. typhimurium*, the *rfaD*, *rfaF*, and *rfaC* genes are present as a contiguous gene cluster. We have demonstrated that while *rfaD* and *rfaF* are adjacent on the 2019 genome, *rfaC* is not associated with these genes. We have also demonstrated that the genetic organization of *rfaD* and *rfaF* genes varies among strains of *H. influenzae*. The genome sequence of Rd strain KW20 demonstrated that eight genes are located between *rfaD* and *rfaF*. We have verified this finding with PCR and also demonstrated that the *rfaD* and *rfaF* genes were adjacent in eight other strains which were examined. The differences of genetic organization among the clinical strains and the Rd strain KW20 suggest that while the genome sequence of the Rd strain is a useful tool for studying the genetics of *H. influenzae*, it cannot be relied on as a perfect representation of the genomes of other strains of *H. influenzae.*

The LOS of *H. influenzae* mutants DK1 and FK1 were characterized by SDS-PAGE analysis followed by silver staining. Both mutants expressed a truncated LOS molecule as expected, which demonstrated that the gene products were involved in LOS biosynthesis. The similar migration patterns of the DK1 and FK1 LOS on these gels suggested that the LOS molecules had similar molecular masses. It was expected that in the absence of ADP-L-*glycero*-D-*manno*-heptose-6-epimerase activity, the first heptose would not be added to the KDO portion of the LOS and would lead to a reduction in molecular mass of the DK1 LOS equivalent to the mass of one heptose molecule. MS studies showed essentially identical molecular masses for the two LOS species. The molecular masses measured correspond to an LOS structure that contains a conserved lipid A, a single phospho-KDO, and a single heptose unit as well as a nonstoichiometric amount of PEA. A Hep-KDO(P)-lipid A structure is consistent with the model of RfaF as the heptosyltransferase II enzyme. In the absence of heptosyltransferase II activity, the LOS structure is terminated after the first heptose. Since a heptose was also incorporated into the LOS of the *rfaD* mutant, chromatographic studies were performed to determine if the DD-heptose was utilized in the LOS in the absence of the LD-heptose molecule. These studies showed that while FK1 had the LD-heptose in the LOS, DK1 utilized the DD-heptose. The presence of the DD-heptose in the LOS of DK1 is further support for the model of RfaD as the ADP-L-*glycero*-D-*manno*-heptose-6-epimerase. This information along with that provided by the MS and composition analysis disproved the hypothesis that in the absence of epimerase activity, no heptose would be added to the exposed KDO. In the absence of LD-heptose, the cells use the DD-heptose molecule; however, this DD-heptose does not allow

further additions to the LOS molecule. The absence of the second heptose beyond the nonreducing terminal DD-heptose in the LOS of strain DK1 suggests that the heptosyltransferase II does not recognize the DD-heptose at this position in lieu of LD-heptose. Additionally, the truncated LOS structures of both mutants suggest that the glycosyltransferase responsible for the addition of a glucose to heptose I did not function with only a single heptose present in the LOS. It is therefore likely that the activity of this enzyme requires either the presence of the second heptose or the complete triheptose core.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grants AI24616 and AI31254. We also acknowledge the UCSF Mass Spectrometry Facility, which is supported by the National Center for Research Resources (grant RR01614).

We acknowledge Robert Munson and Joel Bozue for providing strains used in this study.

REFERENCES

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1987. Current protocols in molecular biology. John Wiley & Sons, New York, N.Y.
2. **Fleishmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J.-F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J. D. Gocayne, J. Scott, R. Shirley, L.-I. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. M. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, and J. C. Venter.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269:**496–512.
3. **Galanos, C., O. Luderitz, E. T. Rietschel, and O. Westphal.** 1969. A new method for the extraction of R lipopolysaccharides. Eur. J. Biol. **9:**245–249.
4. **Gibson, B. W., J. J. Engstrom, C. M. John, W. Hines, and A. M. Falick.** Characterization of bacterial lipooligosaccharides by delayed extraction matrix-assisted laser ionization time-of-flight mass spectrometry. J. Am. Soc. Mass Spectrom., in press.
5. **Gibson, B. W., W. Melaugh, N. J. Phillips, M. A. Apicella, A. A. Campagnari, and J. M. Griffiss.** 1993. Investigation of the structural heterogeneity of lipooligosaccharides from pathogenic *Haemophilus* and *Neisseria* species and of R-type lipopolysaccharides from *Salmonella typhimurium* by electrospray mass spectrometry. J. Bacteriol. **175:**2702–2712.
6. **Gibson, B. W., N. J. Phillips, W. Melaugh, and J. J. Engstrom.** 1996. Determining structures and functions of surface glycolipids in pathogenic *Haemophilus* bacteria by electrospray mass spectrometry, p. 166–184. *In* C. Fenselau (ed.), Biochemical and Biotechnological applications of electrospray ionization mass spectrometry. American Chemical Society. Washington, D.C.
7. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166:**557–580.
8. **Helander, I. M., B. Lindner, H. Brade, K. Altmann, A. A. Lindberg, E. T. Rietschel, and U. Zähringer.** 1988. Chemical structure of the lipopolysaccharide of *Haemophilus influenzae* strain I-69 Rd^{-}/B^{+}. Description of a novel deep-rough chemotype. Eur. J. Biochem. **177:**483–492.
9. **Herriot, R. M., E. M. Meyer, and M. Vogt.** 1970. Defined nongrowth media for stage II development of competence in *Haemophilus influenzae*. J. Bacteriol. **101:**517–524.
10. **Kimura, A., and E. J. Hansen.** 1986. Antigenic and phenotypic variations of *Haemophilus influenzae* type b lipopolysaccharide and their relationship to virulence. Infect. Immun. **51:**69–79.
11. **Kimura, A., C. C. Patrick, E. E. Miller, L. D. Cope, G. H. McCracken, Jr., and E. J. Hansen.** 1987. *Haemophilus influenzae* type b lipopolysaccharide: stability of expression and association with virulence. Infect. Immun. **55:** 1979–1986.
12. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
13. **Lee, N.-G., M. G. Sunshine, and M. A. Apicella.** 1995. Molecular cloning and characterization of the nontypeable *Haemophilus influenzae* 2019 *rfaE* gene required for lipopolysaccharide biosynthesis. Infect. Immun. **63:**818–824.
14. **Lehmann, V., G. Hammerling, M. Nurminen, I. Minner, E. Ruschmann, O. Luderitz, T.-T. Kuo, and B. A. D. Stocker.** 1973. A new class of heptose-defective mutant of *Salmonella typhimurium*. Eur. J. Biochem. **32:**268–275.
15. **Melaugh, W., N. J. Phillips, A. A. Campagnari, R. Karalus, and B. W. Gibson.** 1992. Partial characterization of the major lipooligosaccharide from a strain of *Haemophilus ducreyi*, the causative agent of chancroid, a genital ulcer disease. J. Biol. Chem. **267:**13434–13439.
16. **Ono, M., and M. Kuwano.** 1979. A conditional lethal mutation in an *Escherichia coli* strain with longer chemical lifetime of messenger RNA. J. Mol. Biol. **129:**343–357.
17. **Pegues, J. C., L. Chen, A. W. Gordon, L. Ding, and W. G. Coleman, Jr.** 1990. Cloning expression and characterization of the *Escherichia coli* K-12 *rfaD* gene. J. Bacteriol. **172:**4652–4660.
18. **Phillips, N. J., M. A. Apicella, J. M. Griffiss, and B. W. Gibson.** 1993. Structural studies of the lipooligosaccharides from *Haemophilus influenzae* type b strain A2. Biochemistry **32:**2003–2012.
19. **Phillips, N. J., C. M. John, L. G. Reinders, B. W. Gibson, M. A. Apicella, and J. M. Griffiss.** 1990. Structural models for the cell surface lipooligosaccharides of *Neisseria gonorrhoeae* and *Haemophilus influenzae*. Biomed. Environ. Mass Spectrom. **19:**731–745.
20. **Phillips, N. J., R. McLaughlin, T. J. Miller, M. A. Apicella, and B. W. Gibson.** 1996. Characterization of two transposon mutants from *Haemophilus influenzae* type b with altered lipooligosaccharide biosynthesis. Biochemistry **35:**5937–5947.
21. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
22. **Schnaitman, C. A., and J. D. Klena.** 1993. Genetics of lipopolysaccharide biosynthesis in enteric bacteria. Microbiol. Rev. **57:**655–682.
23. **Sirisena, D. M., P. R. MacLachlan, S.-L. Liu, A. Hessel, and K. E. Sanderson.** 1994. Molecular analysis of the *rfaD* gene, for heptose synthesis, and the *rfaF* gene, for heptose transfer, in lipopolysaccharide synthesis in *Salmonella typhimurium*. J. Bacteriol. **176:**2379–2385.
24. **Tsai, C.-M., and C. E. Frasch.** 1982. A sensitive silver stain for detecting lipopolysaccharides in polyacrylamide gels. Anal. Biochem. **119:**115–119.
25. **Wilcox, K. W., and H. O. Smith.** 1975. Isolation and characterization of mutants of *Haemophilus influenzae* deficient in an adenosine 5′-triphosphate-dependent deoxyribonuclease activity. J. Bacteriol. **122:**443–453.
26. **Zwahlen, A., L. G. Rubin, and E. R. Moxon.** 1986. Contribution of lipopolysaccharide to pathogenicity of *Haemophilus influenzae*: comparative virulence of genetically-related strains in rats. Microb. Pathog. **1:**465–473.

Editor: P. E. Orndorff