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

Follow this and additional works at: http://repository.usfca.edu/chem_fac

Part of the Biochemistry Commons, Chemistry Commons, and the Microbiology Commons

Recommended Citation

Identification of the ADP-1-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. APICELLA1,∗

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

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-1-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 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, 1-glycero-d-manno-heptose is added to the LOS molecule in place of the usual 1-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 t-glycerol form of heptose, which is accomplished by the ADP-1-glycero-d-manno-heptose-6-epimerase. Without the activity of this enzyme, 1-glycero-d-manno-heptose (1d-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-
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-β-D-manno-heptose-6-epimerase and the heptosyltransferase II of S. typhimurium LT2.

**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. Barak. 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. H. influenzae 2019 was cultured in brain heart infusion broth supplemented with 2% Fildes reagent (Difco Laboratories) at 37°C with agitation or on BHI agar 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 retransformed 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 Porator (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 retransformed

### Table 1. Bacterial strains and plasmids used in this study

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Genotype or characteristic(s)</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>H. influenzae</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td>Wild type</td>
<td>M. A. Apicella</td>
</tr>
<tr>
<td>DK1</td>
<td>rfaDΔ Kan^r</td>
<td>This study</td>
</tr>
<tr>
<td>FK1</td>
<td>rfaFΔ Kan^r</td>
<td>This study</td>
</tr>
<tr>
<td>3031</td>
<td>Wild type</td>
<td>M. A. Apicella</td>
</tr>
<tr>
<td>3198</td>
<td>Wild type</td>
<td>M. A. Apicella</td>
</tr>
<tr>
<td>5987</td>
<td>Wild type</td>
<td>M. A. Apicella</td>
</tr>
<tr>
<td>7052</td>
<td>Wild type</td>
<td>M. A. Apicella</td>
</tr>
<tr>
<td>Eagan</td>
<td>Wild type</td>
<td>Robert Munson</td>
</tr>
<tr>
<td>A2</td>
<td>Wild-type type b</td>
<td>M. A. Apicella</td>
</tr>
<tr>
<td>A4</td>
<td>Wild-type type b</td>
<td>M. A. Apicella</td>
</tr>
<tr>
<td>A9</td>
<td>Wild-type type b</td>
<td>M. A. Apicella</td>
</tr>
<tr>
<td>KW20</td>
<td>Wild-type Rd</td>
<td>25</td>
</tr>
<tr>
<td><strong>S. typhimurium</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LT2</td>
<td></td>
<td>SGSC</td>
</tr>
<tr>
<td>SL3770</td>
<td>rfa^- wild type</td>
<td>SGSC (14)</td>
</tr>
<tr>
<td>SGSC</td>
<td></td>
<td>SGSC (14)</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pUC19</td>
<td>Cloning vector, Amp^r</td>
<td>Gibco BRL</td>
</tr>
<tr>
<td>pSHlox-1</td>
<td>Cloning vector, Amp^r</td>
<td>Novagen</td>
</tr>
<tr>
<td>pBluescript II</td>
<td>Cloning vector, Amp^r</td>
<td>Stratagene</td>
</tr>
<tr>
<td>SK</td>
<td>Cloning vector, Cm^r</td>
<td>Gerald Barak</td>
</tr>
<tr>
<td>pGB19</td>
<td>4-kb ApoI fragment from 2019</td>
<td>This study</td>
</tr>
<tr>
<td>pDEP</td>
<td>1.6-kb EcoRI-Plr fragment from pD41 cloned into pUC19</td>
<td>This study</td>
</tr>
<tr>
<td>pDPP</td>
<td>1.4-kb Plr fragment from pD41 cloned into pUC19</td>
<td>This study</td>
</tr>
<tr>
<td>pDHP</td>
<td>1.5-kb Plr-HindIII fragment from pD41 cloned into pUC19</td>
<td>This study</td>
</tr>
<tr>
<td>pDMNI</td>
<td>2.4-kb MluI-Nsi fragment from pD41 into pBluescript SK II -</td>
<td>This study</td>
</tr>
<tr>
<td>pDDK1</td>
<td>with a Kan^r cassette cloned into the Plr site</td>
<td>This study</td>
</tr>
<tr>
<td>pDFK1</td>
<td>pDHP with a Kan^r cassette cloned into the MluI site</td>
<td>This study</td>
</tr>
<tr>
<td>pDEXO3</td>
<td>ExoII deletion of pD41</td>
<td>This study</td>
</tr>
<tr>
<td>pDEXO5</td>
<td>ExoIIII deletion of pD41</td>
<td>This study</td>
</tr>
<tr>
<td>pDEXO6</td>
<td>ExoIII deletion of pD41</td>
<td>This study</td>
</tr>
<tr>
<td>pDEXO16</td>
<td>ExoIII deletion of pD41</td>
<td>This study</td>
</tr>
</tbody>
</table>

**Abbreviations for antibiotics:** Amp, ampicillin; Cm, chloramphenicol; Kan, kanamycin.

**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 CaCl2 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 ApoI 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 SHlox 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.
formed to the mutant strain to confirm the complementation. The transformants were also tested for phase 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 mod-81A DNA synthesizer (Applied Biosystems, Inc.). DNA sequence analysis was performed on 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, and transferred to a Hybond-N membrane (Amersham) by capillary blotting overnight, and cross-linked to the membrane by using a Stratagene (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. DNA was extracted from NtHi 2019 or from E. coli HAK117 harboring recombinant plasmids. E. coli HAK117 (16) had 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.

Mutagenesis. The Kan’ cassettes were cloned into restriction sites of the rfaD and rfaF genes. The Kan’ cassette used was isolated from plasmid pUC4K (Amersham). It contains its own promoter, which was deleted 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 5% 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 [35S]methionine (Amersham). The reaction mixtures were separated on an SDS–15% polyacrylamide gel (12), which was dried and exposed to an X-ray film. [35S]-labeled protein molecular weight markers (Amersham) were used to generate a linear range between 14.3 and 200 kDa for determination of relative molecular weights.

FIG. 2. Physical maps of parental plasmid pD41 and derivative subclones. The insert of pD41 is a 4.8-kb Apol fragment. Also shown are the results of complementation studies of S. typhimurium rfaD and rfaF mutants. Triangles indicate the relative positions of the Kan’ cassette insertions to construct mutants in rfaD and rfaF genes. Abbreviations: E, E.coli; M, Mshl; N, NdiI; P, PvuII; S, StyI.

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-glycerool-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...
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 EcoRI-PstI and 1.4-kb PstI fragments, respectively, from pD41 into pUC19. Plasmid pDHP was constructed by cloning the 1.5-kb HindIII-PstI fragment into PstI-HindIII sites of pUC19. Plasmid pDMN1 was constructed by cloning the 2.4-kb MluI-NsiI fragment from pD41 into pBluescript SK II−. The MluI 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 exonucleases 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 H11114 and H11105 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 encode 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. 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.
in NtHi 2019. This was accomplished by the incorporation of an intact Kan\textsuperscript{r} gene into the rfaD and rfaF genes. The locations of the Kan\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{r} ampicillin-susceptible colonies. The strain which incorporated the Kan\textsuperscript{r} gene in the rfaD gene was called DK1, and the strain which incorporated the Kan\textsuperscript{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 plasmids pDMN1 and pDFK1 and screening for Kan\textsuperscript{r} ampicillin-susceptible colonies. The strain which incorporated the Kan\textsuperscript{r} gene in the rfaD gene was called DK1, and the strain which incorporated the Kan\textsuperscript{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 rfaD and rfaF genes. 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\textsuperscript{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\textsuperscript{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.
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 strain 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 and −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

![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.](http://iai.asm.org/)

![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.](http://iai.asm.org/)
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, \((M - 2\text{H})^{2-}\), 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) (\(\Delta m = 123\) 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 \(\text{N,N'-diacyl}\) form in which each of the two glucosamine sugars contains a single N-linked \(\beta\)-hydroxymyristic acid group:

\[
\text{Hep} \rightarrow \text{KDO}(\text{PEA})_{\text{O} \rightarrow \text{P}} \text{GlcNH} \rightarrow \text{GlcNH}(\text{PEA})
\]

where \(M_r = 1445.3\) (LOS-A) and 1568.4 (LOS-B). Other peaks are also present in these two spectra and can be readily

![Diagram](http://iai.asm.org/fig8.png)

**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 1-69 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

<table>
<thead>
<tr>
<th>Mutant strain</th>
<th>M&lt;sub&gt;i&lt;/sub&gt; (Observed)</th>
<th>M&lt;sub&gt;i&lt;/sub&gt; (Calculated)</th>
<th>Relative % abundance</th>
<th>Proposed composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>DK1</td>
<td>951.7</td>
<td>952.0</td>
<td>ND</td>
<td>N,N'-diacyl lipid A&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>1,445.6</td>
<td>1,445.3</td>
<td>37</td>
<td>Hep-KDO(P)-O-deacyl-lipid A</td>
</tr>
<tr>
<td></td>
<td>1,568.4</td>
<td>1,568.4</td>
<td>100</td>
<td>Hep-KDO(P)(PEA)-O-deacyl-lipid A</td>
</tr>
<tr>
<td>FK1</td>
<td>951.8</td>
<td>1,253.2</td>
<td>ND</td>
<td>KDO(P)-O-deacyl-lipid A</td>
</tr>
<tr>
<td></td>
<td>1,445.8</td>
<td>1,445.3</td>
<td>95</td>
<td>Hep-KDO(P)-O-deacyl-lipid A</td>
</tr>
<tr>
<td></td>
<td>1,568.6</td>
<td>1,568.4</td>
<td>100</td>
<td>Hep-KDO(P)(PEA)-O-deacyl-lipid A</td>
</tr>
</tbody>
</table>

*All average molecular weights and relative abundances for the different O-deacylated LOS-glycoforms are based on the doubly deprotonated charged molecular ion, (M + 2Na)<sup>2-</sup>, and in the case of relative ion abundances, the abundances of the sodiated molecular ion species, i.e., (M + 2H + Na)<sup>2+</sup>, (M + 3H + 2Na)<sup>2+</sup>, etc., were included.

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

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<sub>i</sub> of 953.0.

Although strain FK1 would 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 heptose moiety based on the observed masses. 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).

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 pDHF 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 pDHF 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 conjugate 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 co-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 *NiHi* 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 cassette into the genes of interest on a recombinant plasmid. The insert DNA was then excised and transformed into *NiHi* 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 size of 34.8 kDa. The nucleotide sequences upstream of the transcription initiation site of *rfAD* and *rfaF* genes, respectively.

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 *NiHi* *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 *rfaF* 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 *NiHi* 2019. While these results provide strong evidence that the *rfaD* and *rfaF* genes of *NiHi* 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 R 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 genomics 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 HepKDO(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 L-deoxy-heptose molecule. These studies showed that while FK1 had the L-deoxy-heptose in the LOS, DK1 utilized the DD-heptose. The presence of the L-deoxy-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 L-deoxy-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 α-D-heptose in the LOS of strain DK1 suggests that the heptosyltransferase II does not recognize the α-D-heptose at this position in lieu of α-L-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


Editor: P. E. Orndorff