Journal of Endotoxin Research (1994) 1, 165–174

JOURNAL OF
ENDOTOXIN
RESEARCH

Characterization and sequence analysis of the *lsg* (LOS synthesis genes) locus from *Haemophilus influenzae* type b

R. McLaughlin, N.-G. Lee, Y. Abu Kwaik, S. M. Spinola, M. A. Apicella

Department of Microbiology, The University of Oklahoma Health Sciences Center, OK, USA, Department of Microbiology, University of Iowa, Iowa City, IA, USA, Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI, USA and Department of Medicine, Division of Infectious Diseases, University of Indiana, Indianapolis, IN, USA

SUMMARY. **Analysis of the *lsg* (LOS synthesis genes) cluster in *Escherichia coli* strain K12 and mutations in the *lsg* locus in *Haemophilus influenzae* type b indicated the presence of 3 regions responsible for sequential modifications of *E. coli* lipopolysaccharide (LPS). Sequencing of the *lsg* region yielded 7,435 bp that encompassed 7 complete and 1 partial open reading frames (ORFs 1–8). The predicted product of ORF1 had homology to the consensus sequence of cytochrome b proteins (21% identity, 51% similarity) and to other transmembrane proteins. The products of ORF5 and ORF6 share overall 23% identity and 49% similarity with each other. The ORF6 protein had high homology with the product of ORF275 of the *E. coli rfb* gene cluster (40% identity, 58% similarity), whose function is not known. Multiple sequence alignment of the ORF5 and ORF6 proteins with the RfbB, RfbJ and RfbX proteins revealed conserved motifs over the N-terminal half region of all these proteins. The products of ORF7 and ORF8 are homologous with *Azotobacter vinelandii* MolA protein (30% identity, 51% similarity) and MolB protein (26% identity, 48% similarity), respectively. The promoter regions of ORF1, 7 and 8 were determined by primer extension analysis and found to be similar to bacterial σ^{70}-dependent promoters. ORF7 and ORF8 are transcribed into diverse orientation. At least 5 of the encoded proteins have been identified using coupled *E. coli* transcription/translation system and labeling with [^{35}S]-methionine. We conclude that the genetic organization of the *lsg* biosynthesis pathway involves multiple operons that lead to the assembly of an *H. influenzae* LOS structure.**

The lipooligosaccharides (LOSs) of *Haemophilus influenzae* type b (Hib) are important virulence factors for this human pathogen.[1–3] The oligosaccharide region of Hib LOS is composed of a complex array of stable epitopes and epitopes that phase vary at frequencies from 10^{-2} to 10^{-3}.[1,2,4] Elucidation of the regulatory and biosynthetic genes responsible for LOS oligosaccharide biosynthesis and phase variation has recently begun.[5–9]

Our laboratory has described a cloned Hib gene cluster (*lsg*: LOS synthesis genes) that is responsible for assembly of a series of saccharide components on the *Escherichia coli* K-12 lipopolysaccharide (LPS).[4,5,7,10] The *lsg* locus is contained within a 7.4 kb *Bam*HI–*Pst*I fragment isolated from Hib strain A2. Three overlapping subclones of the *lsg* locus produce sequential LPS modifications in *E. coli* K-12.[5] Initial modification of the *E. coli* LPS requires the presence of the 2.8 kb *Pst*I–*Sph*I fragment that assembles a 0.4 kDa oligosaccharide on a 4.1 kDa LPS structure. The *Pst*I–*Sph*I fragment combined with the adjacent 2.7 kb *Sph*I–*Hin*dIII fragment leads to assembly of an additional 0.6 kDa oligo-

R. McLaughlin PhD, Department of Microbiology, The University of Oklahoma Health Sciences Center, OK 73190, USA, **N.-G. Lee PhD, M. A. Apicella MD**, Department of Microbiology, University of Iowa, Iowa City, IA 52242, USA, **Y. Abu Kwaik PhD**, Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48109, USA, **S. M. Spinola MD**, Department of Medicine, Division of Infectious Diseases, University of Indiana, Indianapolis, IN 46202, USA.
Correspondence to Michael A. Apicella MD, Department of Microbiology, University of Iowa, Bowen Science Building Rm 3-401, 51 Newton Road, Iowa City, Iowa 52242, USA.

saccharide on the modified structure. A 1.4 kDa modification to *E. coli* LPS is generated by addition of the 1.7 kb *Hin*dIII–*Bam*HI fragment to the *Pst*I–*Hin*dIII fragment. Generation of isogenic Hib mutants of the *lsg* locus has verified its involvement in LOS biosynthesis.[10] The chemical structure of *H. influenzae* LOS is now being elucidated.[11] The deep core region contains a branched tri-heptose structure. It appears that each heptose can be substituted by an oligosaccharide chain. This complexity has made analysis of the terminal LOS sugars difficult and has hindered assignment of specificities to the biosynthesis genes involved in their production.

In this study, we present the DNA sequence analysis and further characterization of the *lsg* locus. The sequence contained 7 complete and 1 partial open reading frames (ORFs), some of which encode proteins homologous to enterobacterial Rfb proteins. The promoters of ORF1, ORF7 and ORF8 were localized by promoter analysis and found to be similar to bacterial σ^{70}- dependent promoters. In vitro translation studies confirmed at least 5 of the gene products predicted by the ORFs.

MATERIALS AND METHODS

Bacterial strains and growth conditions

H. influenzae type b (Hib) strain A2 was grown on chocolate agar supplemented with isovitalex or brain heart infusion agar supplemented with 4% Fildes reagent (sBHI) (Difco Laboratories) at 35°C in a 5% CO_2 atmosphere or in sBHI broth at 37°C with agitation. Chloramphenicol (2 μg/ml) or kanamycin (20 μg/ml) was added to sBHI medium for selection and maintenance of Hib isolates containing the mini-Tn3(Cm) transposon or kanamycin resistance (kan^r) cassette, respectively. *E. coli* strains were routinely cultured at 37°C using LB agar or broth with appropriate antibiotics.

Monoclonal antibodies

Monoclonal antibody (MAb) 6E4 has been previously described.[7] MAb 2F2 is an IgG murine monoclonal antibody generated against the LOS from *H. influenzae* non-typable strain 2019. Both MAbs were detected after binding with protein A conjugated to horseradish peroxidase (Zymed Laboratories). Monoclonal antibody binding in ELISA assay was classified as normal (> 75%), reduced (between 25–75%) or negative (< 10%) as compared with the parental strain. All values are the average of at least 10 replicate samples from a single assay. Each assay was performed at least twice.

DNA manipulation

Plasmid and chromosomal DNA isolation and Southern hybridization mapping were performed as previously described.[5] T4 DNA ligase and restriction enzymes were used according to the manufacturer's suggestions (GIBCO BRL and Promega). *Haemophilus* cells were made competent and transformed according to the M-IV medium procedure.[12] Transformation of *E. coli* strains with plasmid DNA was carried out by the $CaCl_2$ method.[13]

Mutagenesis

Transposon mutagenesis of the *lsg* locus was accomplished using shuttle mutagenesis as previously described.[5,14] The location of transposon insertion sites was determined by genomic Southern hybridization.

Phenotypic analysis

The immunochemical phenotypes of *lsg* mutants were determined by ELISA and Western blot as previously described.[10] The LOS banding patterns of the mutant strains were determined by digestion of whole cells with proteinase K,[15] separation by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) using tricine as the trailing ion[16] and detection by silver staining.[17]

DNA sequencing and sequence analysis

Double-stranded DNA was sequenced by the dideoxynucleotide termination method[18] labeled with [α-^{35}S]-dATP using the Sequenase II kit (United States Biochemicals). Sequence of the entire 7.4 kb region of the *lsg* locus was accomplished by generating plasmids containing specific restriction fragments and by exonuclease III digestion of various subclones.[5] Specific oligonucleotides were also used to sequence several regions.

The complete 7.4 kb sequence was assembled using the Fragment Assembly System contained in the Genetics Computer Group (GCG) sequence analysis software package.[19] Analysis of the *lsg* sequence was performed using various programs of the GCG package. Homology search with the deduced amino acid sequences within the SWISS-PROT and Genpept databases was done with GENBANK Online Service using the FASTA algorithm.[20] Protein sequence alignment was done using the BESTFIT program of the GCG package.

Primer extension analysis

RNA was extracted from Hib A2 grown in sBHI broth to OD_{600} 0.6 by sonication, digestion with proteinase K and phenol/chloroform extraction followed by ethanol precipitation as described.[13] The purified RNA was quantitated spectrophotometrically and the quality of RNA was confirmed on a formaldehyde-agarose gel stained with ethidium bromide.

Primer extension analysis was carried out using the

Primer Extension Kit (Promega) following the manufacturer's instructions except for annealing, which was done by heating the reaction mixture to 70°C, incubating at 60°C for 20 min and then slowly cooling down to room temperature. 40 μg 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 made with the primer was used as a marker to confirm the position of the primer extended products.

In vitro transcription/translation analysis

Plasmids containing the entire *lsg* locus or various restriction fragments were purified by cesium chloride buoyant density gradient[13] and analyzed in a coupled transcription/translation system (Promega) using translation grade [^{35}S]-methionine (Amersham). Translation products were separated on a 20% acrylamide gel by SDS-PAGE.[21] Low molecular weight [^{14}C]-labeled Rainbow standards (Amersham) were used to generate a linear regression curve between 14.3–46 kDa for determination of relative molecular weights.

Nucleotide sequence accession number

The DNA sequence described has been deposited at GenBank with the accession number M94855.

RESULTS

Nucleotide sequence of the *lsg* locus

The 7.4 kb *lsg* locus was completely sequenced from both strands. The fragment had a total length of 7435 bp that encompassed 7 complete and 1 partial open reading frames (ORFs1–8) determined with the FRAMES program.[19] All ORFs were transcribed from left to right with the exception of ORF7 (Fig. 1A). All ORF coding regions had the high AT content characteristic of the Hib genome.

The sequence data suggested that ORFs 1–3 and 4–6 comprised 2 operons. The predicted ATG start of ORF2 overlapped the TAA stop codon of ORF1 by 2 bases. Likewise, ORF2 and ORF3 were separated by only 1 base. ORF1 and ORF3 were translated in the same reading frame, while ORF2 was translated in an alternate frame. Possible ribosomal binding sites (RBS), based on homology with known RBS and the Shine-Dalgarno consensus sequence,[22] were identified upstream of the predicted start of the 3 ORFs.

ORFs 4–6 were separated from the first operon by an intercistronic region of 359 bases. 11 bases separated ORF4 and ORF5 and only 1 base separated ORF5 from ORF6. A RBS was not identified upstream of the first start codon for ORF4. However, a potential RBS was located upstream of a second ATG codon 33 bp from the first start. This suggested that the second ATG may be the true start codon for ORF4. ORF5 and ORF6 each had a possible RBS located upstream of their predicted start codons, although the ORF6 RBS consensus was weak. ORFs 4–6 were translated in different reading frames.

ORF7 and ORF8 were divergently transcribed from a 137 bp intercistronic region. Potential ribosomal binding sites were identified 7 bases upstream of each ORF.

Identification of sequences having homology with the *lsg* ORFs

Search for homologous sequences with the deduced amino acid sequences of the *lsg* ORFs within the SWISS-PROT and Genpept databases was done with GENBANK Online Service using the FASTA algorithm.[20] Homologies were found with the predicted products of ORFs 1, 5, 6,7 and 8 but no significant homologies to any sequences were obtained to the products of ORFs 2, 3 or 4.

The sequence having the highest homology with the ORF1 protein was a 403 amino acid (AA) consensus sequence compiled from 37 complete sequences of cytochrome b present in the SWISS-PROT database. The consensus closely matched the predicted length of the ORF1 protein (402 AA) and shared 21% identity and 51% similarity. The homology was not restricted to any particular region but evenly distributed throughout the protein sequences. A hydrophobicity profile of the ORF1 protein using Kyte-Doolittle algorithm[23] showed a striking similarity to that of the cytochrome b consensus sequence (Fig. 2). Other homologous proteins included the NADH-dehydrogenase (ubiquinone) chain 4 (24% identity, 53% similarity,), the formate hydrogenase subunit HycD (16% identity, 51% similarity) and the photosynthesis II D2 protein (17% identity, 46% similarity). All homologous sequences were transmembrane proteins, the majority of which may act as or in conjunction with electron carriers.

The ORF5 protein had 22% identity and 46% similarity to the undefined ORF14.1 protein of the *Salmonella typhi* Ty2 *rfb* gene cluster and the equivalent ORF from *S. typhimurium* LT2. The ORF14.1 protein has been identified by mutational analysis as involved in assembly of the species specific O antigen oligosaccharide.[24] The ORF6 protein had high homology with the product of ORF275 of the *E. coli rfb* gene cluster (40% identity, 58% similarity), whose function is not known yet.[25] The ORF5 and ORF6 proteins shared homology with each other (23% identity, 49% similarity). Homology search revealed that the ORF5 and ORF6 proteins were homologous with the RfbB, RfbJ and RfbX proteins encoded in the *rfb* gene cluster, but the overall homology was rather low (18–24% identity, 41–48% similarity). Multiple sequence alignment, however, revealed several motifs conserved in the N-terminal half of these proteins (Fig. 3).

The products of ORF7 and ORF8 shared homology with the MolA protein (30% identity, 51% similarity) and the MolB protein (26% identity, 48% similarity),

respectively, encoded in the gene cluster which is involved in molybdenum transport in *Azotobacter vinelandii* (Fig. 4).[26]

Localization of the promoter regions of the *lsg* ORFs

In order to determine the promoter regions for the *lsg* ORFs, primer extension analysis was carried out using oligonucleotides complementary to the regions of DNA just downstream of the putative translation initiation codon ATGs (whose sequences are shown in Figure 6). The transcription start point of ORF1 was mapped to 54 to 56 and 51 nucleotides (nt) upstream of the putative ATG codon (Fig. 5A). It is not clear whether the transcription initiates at 2 closely spaced sites or whether the downstream site is an artifact caused by early fall off of reverse transcriptase. The promoter region contained a sequence (TAAAAT) homologous to the consensus –10 element conserved among the bacterial σ^{70}-dependent promoters (Fig. 6). The further upstream region also contained a sequence which agrees with 4 of 6 nt for the consensus –35 element, but the distance to the –10 element was only 13 nt. Primer extension with the primer for ORF7 yielded a major band at 38 nt upstream of the putative ATG codon and several minor bands. The promoter region contained the –10 element with the distance of 8 nt to the transcription start site, but no –35 like sequence was found (Figs 5B & 6). For

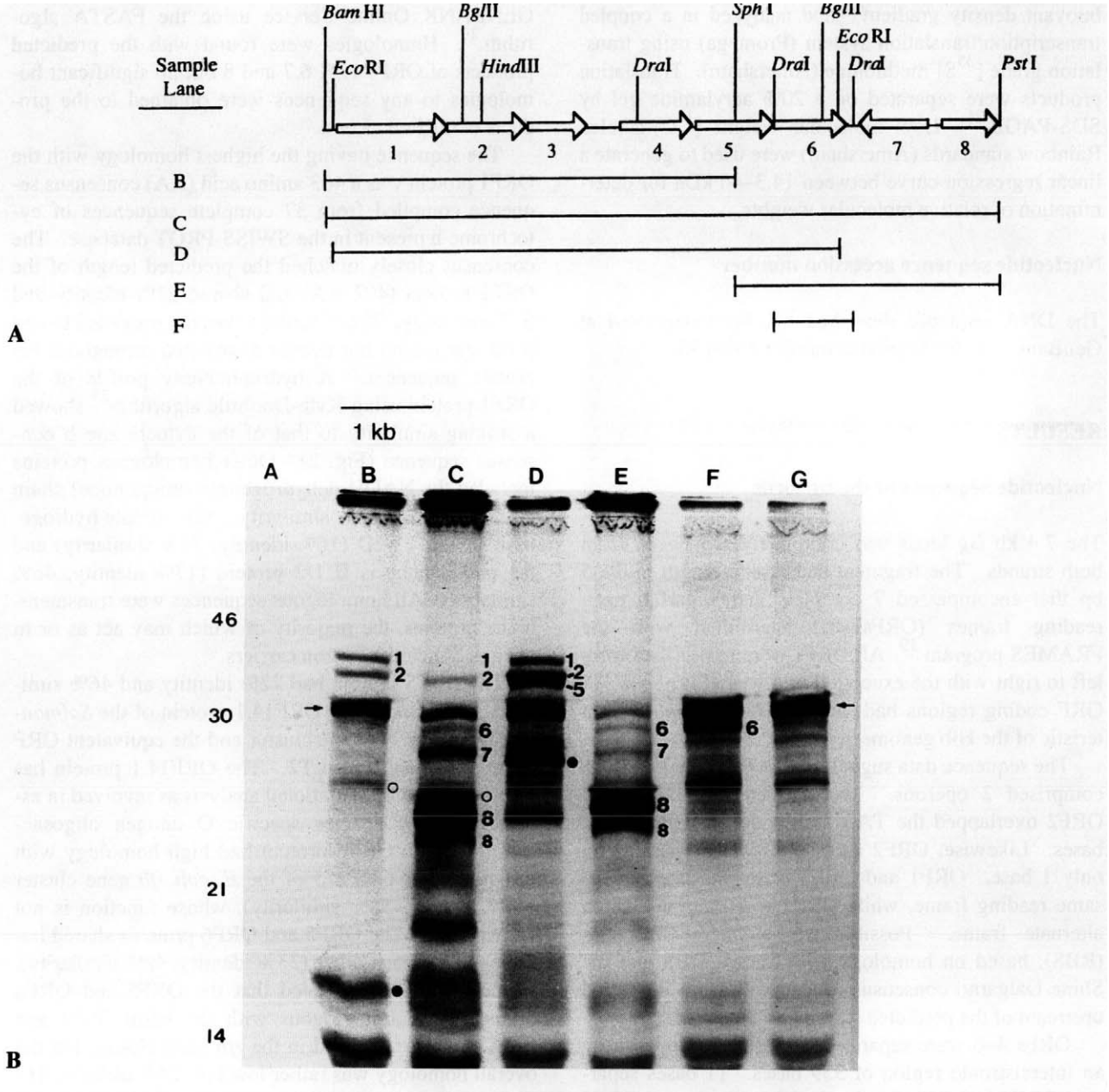

Fig. 1 — **(A)** Restriction map of the *lsg* locus showing deduced open reading frames (ORFs). Plasmids containing the restriction fragments indicated were used for in vitro transcription/translation assay. Numbers depicted below the map correspond to the respective ORFs. **(B)** In vitro transcription/translation of the *lsg* locus. Lane A, low molecular weight protein standards; lanes B–F, translation products of *lsg* subclones corresponding to those depicted in 1A; lane G, pGEM-3zf+ (Promega). Numbers appearing next to protein bands correspond to translation products of ORFs 1–8, respectively. The open circles denote weak bands that may correspond to ORF 3. The solid circles indicate truncated peptides, or peptides fused with lacZ α peptide. The band indicated by the arrow represents the β-lactamase precursor and possibly the ORF 4 translation product.

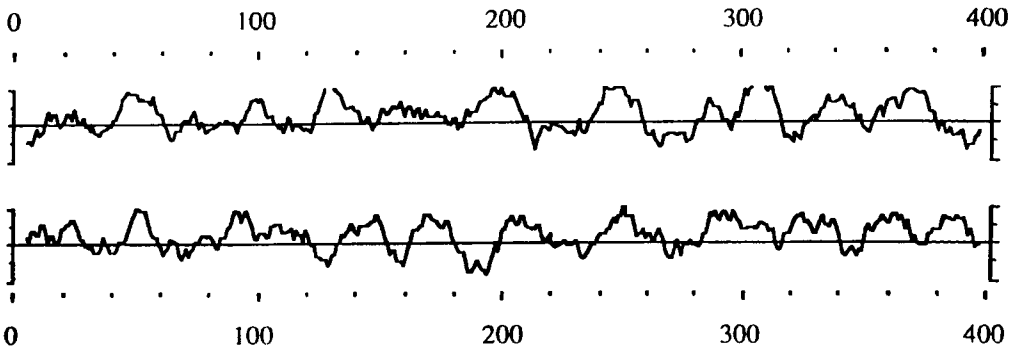

Fig. 2 — Hydrophobicity plot of the cytochrome b consensus sequence and the ORF 1 protein derived with the Pepplot program using the Kyte-Doolittle algorithm. The upper graph is the plot of the cytochrome b consensus, the lower graph is that of ORF1 protein. Values above the x-axis indicate hydrophobic regions of the proteins.

ORF8 2 major transcription initiation sites were found at 28 and 20 nt upstream of the putative ATG codon and the upstream sequence contained a perfect –10 element. Further upstream of this putative –10 element are 2 possible –35 sequences, 12 and 23 nt from the –10 region, which agree with 4 of 6 nt of the consensus –35 motif (Figs 5C & 6). The primer extension reactions for all 3 ORFs with RNA extracted from *E. coli* transformed with a plasmid containing the whole 7.4 kb *lsg* locus yielded the same band pattern as those with Hib A2 RNA, though the transcription levels were lower in *E. coli* (Fig. 5).

Translational analysis of the *lsg* locus

A coupled transcription/translation system was used to confirm the production of proteins from the deduced ORFs of the *lsg* locus. Various plasmids containing fragments of the *lsg* locus were constructed (Fig. 1A) and used for in vitro translation. Labeled proteins having M_r values approximating that of the deduced proteins for ORFs 2 and 5 through 8 were identified (Fig. 1B, Table). A weak band approximating the M_r of ORF3 was also observed. A protein equivalent to ORF4 was not identified due to probable co-migration with the β-lactamase precursor protein.

The deduced protein for ORF1 had a molecular weight of approximately 46 kDa, but the protein resulting from translation of ORF1 had a M_r of approximately 40 kDa. Examination of the primary sequence failed to reveal possible internal start codons, and the secondary structural prediction for ORF1 gave no indication of a leader sequence. ORF8 encoded a peptide of approximately 22 kDa. The secondary structure prediction for ORF8 indicated a highly hydrophobic NH_2-terminal region that contained 35 amino acids and was

```
HiORF6    M.........kfSVLMslYiKenpqfLrecfESLgaQTrqa...dEIVLVf  39
EcORF275  M...sddtpkfSVLMaiYiKdsplfLseaLqSIykNTvap...dEVIIIr  44
HiORF5    M.........LSIIVpSYNRka..eVpalLESLtqQTssn...fEVIIV.  35
RfbX      M.........kITLIIpTYNags..lwpnvLDaIkqQTiyp...dkLIVI.  36
RfbJ      M....skkplLTIaIpTYNRss..cLarlLDSIiqQenychdelEVIVc.  43
RfbB      MnlkkdkspsvaILMgTYNaek..fLeqqLESIqsQTvtn...wsLwVs.  44

HiORF6    DgvvTpElesVVtefetkLpLklVk....lpqNrGlgkalNegllhcdyD  85
EcORF275  DgkvTsElnsVIdswrryLnIkdft....lEkNmGlgaalNfglnqcmhD  90
HiORF5    DdySk.E..rVVveqrysfpVtvIr....nEtNqGAaesrNigaraskgD  78
RfbX      DsgSkdEt.vpLasdlknIsIfnIds...kDfNhG.gtrnlavaktldaD  81
RfbJ      DnaSTdEtarIaksgldkIrnstyhl...nEeNlGmdgnfQkcfelsngk  90
RfbB      DdgSSdntlaIIdkykkkwgLgrLnvrrgaqkgfahnfmslvcdieitaD  94

HiORF6    wVfrMDtDDIcvpdRfEKqVafIeqhPesiifggqiaefgknvndivayr 135
EcORF275  lViraDsDDInrtnRfEciLdfMtkngdVhiLsswveefefnpgdkgIik 140
HiORF5    wLlfLDdDDrfmpeKcEKiLqvIeqnPdInfIyhpakcemvnegftyVtq 128
RfbX      vIifLtqDaIladsdaiKnLvyyfsdPlIaaVcgrqlphkdanplavhar 131
RfbJ      yLwmIgdDDLivkngisKvfsiLksrPaLdmVyvnsaaktelnynadVrt 140
RfbB      yfsfsDqDDIweedKlEraLnaLsvfPasqpslycsrtrlvdennveIgf 144
```

Fig. 3 — Multiple sequence alignment of proteins homologous to the products of the ORF 5 and 6. Gaps inserted for optimal alignments are indicated by dots. The C-terminal sequences have been omitted because of lack of sequence similarity among *Hi*ORFs and Rfb proteins. Residues conserved in all 6 proteins are in bold, residues identical or similar (I–L–M–V, D–E, K–R, N–Q, S–T, F–Y) in more than 5 proteins are shaded and residues identical or similar in more than 4 proteins are capitalized. *Hi*ORF6, Hib A2 strain *lsg* locus ORF6 protein; *Ec*ORF275, the product of ORF275 of an unknown function present in *E. coli rfb* region (L04596); *Hi*ORF5, Hib A2 strain *lsg* locus ORF5 protein; RfbX, *S. typhimurium* RfbX protein (P26403); RfbJ, *S. choleraesuis* RfbJ protein (S22615); RfbB, *Y. enterocolitica* RfbB protein (S28578). The codes in the parentheses are the accession numbers for DNA or protein sequence libraries.

characteristic of a signal peptide. Analysis of the translation products from clones containing the ORF8 region revealed 2 bands with M_r values of 23 and 22 kDa. Cleavage of the proposed signal peptide of ORF8 would result in a protein of approximately 21 kDa.

Mutagenesis of the *Eco*RI–*Pst*I fragment

We have previously reported the generation and analysis of transposon insertion mutants within the internal *Eco*RI fragment of the *lsg* locus.[5,10] These insertions led to alterations to the Hib LOS defined by reduced or negative binding of MAb 6E4. We now report the analysis of additional transposon insertion mutants within the terminal 1.7 kb *Eco*RI–*Pst*I fragment. In addition, we have generated several kan^r cassette insertions and replacement/insertions within the *lsg* locus.

Transposon insertions within the 1.7 kb *Eco*RI–*Pst*I fragment of the *lsg* locus resulted in alterations to LOS epitopes that were detected with MAbs 6E4 and 2F2. Transposon insertions that were localized to the 3′ end of ORF6 reduced the MAb 6E4 and 2F2 binding to LOS from these mutants more than 25% compared to that of the parental strain (Fig. 7A). These values were

A

```
HiORF7    MKNTEILLTIKLQQAL..FIDPKRVRLLKEIQQCGSINQAAKNAKVSYKS  48
          |..| :|  :.|: .:   :...|:|||..|:. ||||.||| ..:|||.
AvMolA    MTATRFLARMSLDTDVGTALSDTRIRLLEAIEREGSINRAAKVVPLSYKA  50

HiORF7    AWDHLEAMNKISPRPLLERNTGGKNGGGTVLTTYAERLLQLYDLLERTQE  98
          ||| ::.||.:.| ||: | .||:.|||| ||.|: |::.:|  || . :
AvMolA    AWDAIDTMNNLAPEPLVVRVAGGRQGGGTQLTDYGRRIVAMYRALEIEYQ 100

HiORF7    HAFHILQD.....ESVPLDSLLTATARFSLQSSARNQFFGRVAQQRIIDS 143
           |:. | :     .:..::.:     .:|:..|||||| | |.. |: :
AvMolA    SALDRLSERLNEVTGGDIQAFQRLMHSMSMKTSARNQFAGIVTGLRVGGV 150

HiORF7    RCVVDVNVQGLPTPLQVSITTKSSARLKLITEKEVMLMFKAPWVKISEQP 193
          : | :.::: ...:.. ||. |...|.| .:|||: :.|...|.:..:|
AvMolA    DYEVRIRLDA.ENEIAAVITKASAENLELAIGKEVFALVKSSSVMLTTEP 199

HiORF7    .....LENQ..SNQFPVNIKSLNEEEAILQFAESNIEFCATVHQPNQWQI 236
                ||  :: :.::  .:|:| .:   .:..:. ..|.. ...:.:
AvMolA    SLKLTARNQLWGEVIDIHEGPVNNEVTLALPSGRSVTCVVTADSCKALGL 249

HiORF7    GQQVW..IHIDQEQIILATLG*                              255
          :..|     :. . :|||..|
AvMolA    APGVAACAFFKSSSVILAVYG*                              270
```

B

```
HiORF8     KLTKISTALLIAGLGFSFAASAKVTVFAAASMTDALQQVAKDYAKQNPKN  52
           .|.:: .| | ||| : |..|.| | .||.:|.::.::|.::.|:.. :
AvMolB     SLSRLVVA.LGAGLLACAAQAAEVQVAVAANFTAPMKDIASQFEKDTG.H  51

HiORF8     EVVFSFASSSTLAKQVEEGAPADIFVSASNKWMKYLSEKDLTVKETEKVL 102
           .|: ||:..:.: .|:::||| ::|:.|.:.  . |...: || :.  ..
AvMolB     KVITSFGPTGGFYSQIQNGAPFEVFLAADDTTPEKLEKEGGTVAGSRFTY 101

HiORF8     VGNDLVLIAPAKSAVNSVDIAKGEWINALKDSYLSVGDPAHVPAGQYAAE 152
           ....||| .: .: |:  | : . . ||:|.  ||:::|  .| |. |.:
AvMolB     AVGKLVLWSAKPGYVD..DQGAVLKKNAFKH..LSIANPKTAPYGAAAVQ 147

HiORF8     ALTKLNLWDKVQDRLARAKDVRGALALVERAEAPYGIVYSTDAKVSQQVK 202
           .|.||.| : ...:|. : .: .| .:|. ::|..|:|  .:.  ...:.
AvMolB     VLAKLGLTEATKSKLVEGASIAQAHQFVATGNAELGFVALSQVYKDGKLT 197

HiORF8     TVAVF..PADSHKPV.VYAVSIVKVTTMR.ILAILEY*              235
           ....:  |:| ..|:   || :.|... .   |:::|
AvMolB     GGSGWNVPGDLYEPIRQDAVILTKGKDNPAAQALVDYLKGPKATEVIKAY 248
```

Fig. 4 — Comparison of the predicted amino acid sequences of the ORF7 product (*Hi*ORF7) and the *A. vinelandii* MolA protein (*Av*MolA) (**A**) and the ORF8 product (*Hi*ORF8) and the *A. vinelandii* MolB protein (*Av*MolB) (**B**). The DNA sequence of ORF8 past *Pst*I site was completed by single stranded sequencing. Identical residues are denoted by bars and similar residues by dots. Gaps were inserted for maximal alignments. Asterisks indicate stop codons.

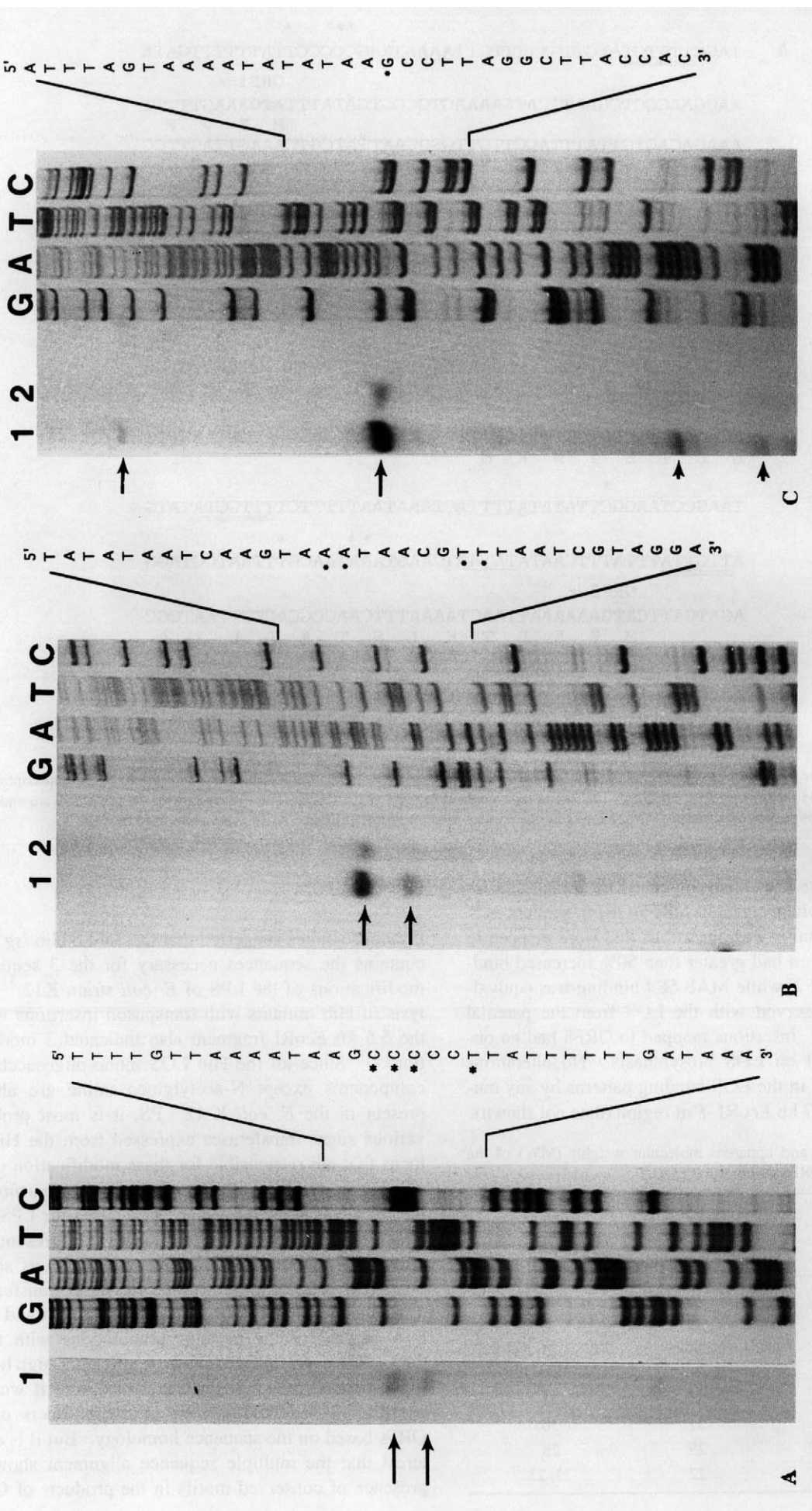

Fig. 5 — Determination of the transcriptional start sites for the ORFs of the *Isg* locus. Primer extension analyses were performed using 21-mer oligonucleotides with RNA extracted from *H. influenzae* b strain A2 (lane 1) and from *E. coli* transformed with a plasmid carrying the entire 7.4 kb *Isg* locus (lane 2). The DNA sequencing ladders shown (GATC) were obtained using the same primers as that in the primer extension reactions to confirm the positions of the extended products. (A) ORF1; (B) ORF7; (C) ORF8. The arrows show the primer extended products and the asterisks show the positions of the transcription start sites on nucleotide sequences.

Fig. 6 — Promoter regions of the ORFs 1 (A), ORF7 and ORF8 (B) of the *lsg* locus. The asterisks indicate the positions of the transcription start sites determined by primer extension analysis. The arrows indicate the directions of the ORFs. The –10 regions for the promoters are underlined and the –35 regions double-underlined. The sequences for the primers used for primer extension analysis are underlined by dots.

similar to those previously observed for transposon insertions elsewhere in the ORF5/ORF6 sequences.[10] LOS from mutants with insertions that were mapped to the ORF7 region had greater than 50% increased binding of MAb 2F2, while MAb 6E4 binding was equivalent to that observed with the LOS from the parental strain (Fig. 7). Insertions mapped to ORF8 had no observable effect on LOS biosynthesis. No alterations were observed in the LOS banding patterns by any mutation in the 1.7 kb *Eco*RI–*Pst*I region (data not shown).

Table. Deduced and apparent molecular weights (MW) of the products of the ORFs within the *lsg* locus

ORF	Deduced MW (kDa)	Apparent MW (kDa)
1	46	40
2	36	36
3	26	25
4	31	?
5	34	34
6	31	30
7	29	28
8	22	23, 22

DISCUSSION

Previous studies suggested that the 7.4 kb Hib *lsg* locus contains the sequences necessary for the 3 sequential modifications of the LPS of *E. coli* strain K12.[5] Analysis of Hib mutants with transposon insertions within the 5.6 kb *Eco*RI fragment also indicated 3 modifications.[10] Since all the Hib LOS mono-oligosaccharide components except N-acetylglucosamine are already present in the *E. coli* K-12 LPS, it is most probably various sugar transferases expressed from the Hib *lsg* locus that are responsible for these modification of the existing *E. coli* LPS. In *E. coli* and *S. typhimurium*, all the *rfa* genes encoding sugar transferases for LPS core structure, with the exception of *rfaE*, are present as a cluster.[27] Thus, it is likely that the *lsg* locus should contain a series of genes coding for sugar transferases. The sequence analysis of the 7.4 kb fragment and database search for the proteins homologous with the 8 ORFs, however, failed to show significantly high homology to any known sugar transferase, and it was not possible to deduce functions of the products of the ORFs based on the sequence homology. But it is of interest that the multiple sequence alignment show the presence of conserved motifs in the products of ORF5

and ORF6, and the RfbB, RfbJ and RfbX proteins, suggesting that they are all related. The product of ORF1 does not contain the conserved motifs shown in Figure 3, but shares homology with the RfbX protein (20% identity, 50% similarity). These *rfb* genes of *Enterobacteriaceae* encode proteins involved in O antigen synthesis. ORF275 in the *E. coli rfb* gene cluster, whose product is highly homologous to the ORF6 protein, is thought to encode a galactosyltransferase.[28] Although Hib LOS does not contain O antigen, ORF1, ORF5 and ORF6 may encode proteins with similar functions in LOS biosynthesis.

The products of ORF7 and ORF8 have homology with the proteins encoded by *molA* and *molB* which are suggested to be involved in molybdenum transport in *A. vinelandii*,[26] but the exact functions of the MolA and MolB proteins have not been determined. We have evidence that ORF 7 may not code for any structural proteins but may code for a regulatory protein. First, although the *Eco*RI–*Pst*I fragment containing ORF7 and ORF8 was unable to alter *E. coli* LPS, transposon insertions within ORF7 of Hib genome resulted in approximately 2-fold increased binding of MAb 2F2 without changing the binding of 6E4. These mutations did not alter the LOS band patterns on SDS-PAGE. Secondly, in vitro transcription/translation studies also showed that the deletion of ORF7 resulted in an increase in the expression of protein products from the other ORFs and that this is not due to decrease in competition for enzymes or substrates in the experimental system (data not shown).

DNA sequence analysis of the 7.4 kb *lsg* locus suggested that ORFs 1–6 are composed of 2 operons, each of which contains 3 ORFs (Fig. 1A) The transcription

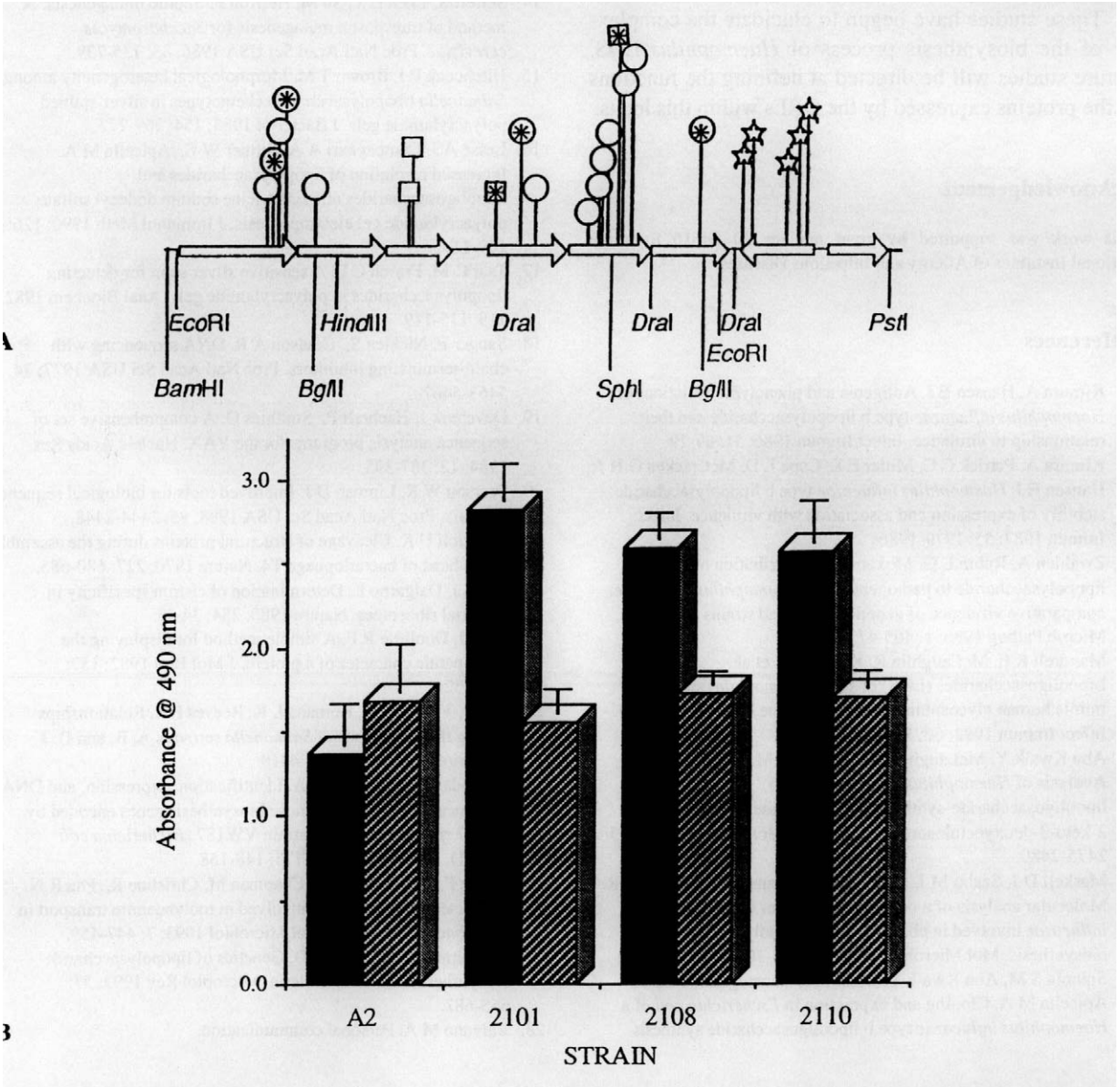

Fig. 7 — (**A**) Locations of m-Tn3(Cm) insertion sites as determined by genomic Southern hybridization. The insertion sites are marked on the map with open reading frames. Symbols indicated the effect of transposon insertion on the binding of the LOS to monoclonal antibodies 6E4 and 2F2. Symbols: circles 6E4/2F2 reduced binding; squares 6E4/2F2 negative phenotype; stars 2F2 increased binding; asterisks multiple insertions mapped to these sites. (**B**) Reactivity of *H. influenzae* type b A2 and isogenic *lsg* ORF 7 mutant strains 2101, 2108 and 2110 with monoclonal antibodies 6E4 and 2F2. Each value represents the average of 11 replicate samples ± SD from a single assay. Assays were performed at least twice giving similar results. Filled bar represents MAb 2F2 binding, hatched bar represents MAb 6E4 binding.

start site for ORF1 was determined by primer extension analysis. The primer extension analysis for the putative second operon containing ORFs 4–6, using a few oligonucleotides complementary to the region of DNA just upstream or downstream of the putative ATG codon of ORF4, yielded several faint bands at very high positions on a gel. It was, therefore, not possible to determine the exact position of the primer extended products. At this point, it is not clear whether ORFs 4–6 are transcribed as a part of the first operon with ORFs 1–3 and these faint bands are from cleaved RNA products or whether transcription of the second operon starts at a very upstream region from the ORF3 coding region. ORF7 and ORF8 are transcribed in opposite directions, sharing a promoter region. This is an interesting feature, considering the possibility that these 2 ORFs encode regulatory proteins, whose expressions, in turn, could be regulated by other factors.

These studies have begun to elucidate the complexity of the biosynthesis process of *Haemophilus* LOS. Future studies will be directed at defining the functions of the proteins expressed by the ORFs within this locus.

Acknowledgement

This work was supported by grant number AI 24616 from the National Institutes of Allergy and Infectious Diseases.

References

1. Kimura A, Hansen E J. Antigenic and phenotypic variations of *Haemophilus influenzae* type b lipopolysaccharide and their relationship to virulence. Infect Immun 1986; 51: 69-79.
2. Kimura A, Patrick C C, Miller E E, Cope L D, McCracken G H Jr, Hansen E J. *Haemophilus influenzae* type b lipopolysaccharide: stability of expression and association with virulence. Infect Immun 1987; 55: 1979-1986.
3. Zwahlen A, Rubin L G, Moxon E R. Contribution of lipopolysaccharide to pathogenicity of *Haemophilus influenzae*: comparative virulence of genetically-related strains in rats. Microb Pathog 1986; 1: 465-473.
4. Mandrell R E, McLaughlin R, Kwaik Y A et al. Lipooligosaccharides (LOS) of some *Haemophilus* species mimic human glycosphingolipids, and some LOS are sialyated. Infect Immun 1992; 60: 1322-1328.
5. Abu Kwaik Y, McLaughlin R E, Apicella M A, Spinola S M. Analysis of *Haemophilus influenzae* type b lipooligosaccharide-synthesis genes that assemble or expose a 2-keto-3-deoxyoctulosonic acid epitope. Mol Microbiol 1991; 5: 2475-2480.
6. Maskell D J, Szabo M J, Butler P D, Williams A E, Moxon E R. Molecular analysis of a complex locus from *Haemophilus influenzae* involved in phase-variable lipopolysaccharide biosynthesis. Mol Microbiol 1991; 5: 1013-1022.
7. Spinola S M, Abu Kwaik Y, Lesse A J, Campagnari A A, Apicella M A. Cloning and expression in *Escherichia coli* of a *Haemophilus influenzae* type b lipooligosaccharide synthesis gene(s) that encodes a 2-keto-3-deoxyoctulosonic acid epitope. Infect Immun 1990; 58: 1558-1564.
8. Weiser J N, Lindberg A A, Manning E J, Hansen E J, Moxon E R. Identification of a chromosomal locus for expression of lipopolysaccharide epitopes in *Haemophilus influenzae*. Infect Immun 1989; 57: 3045-3052.
9. Weiser J N, Love J M, Moxon E R. The molecular mechanism of phase variation of *H. influenzae* lipopolysaccharide. Cell 1989; 59: 657-665.
10. McLaughlin R, Spinola S M, Apicella M A. Generation of lipooligosaccharide mutants of *Haemophilus influenzae* type b. J Bacteriol 1992; 174: 6455-6459.
11. Phillips N J, Apicella M A, Griffiss J, Gibson B W. Structural studies of the lipooligosaccharides from *Haemophilus influenzae* type b strain A2. Biochem 1993; 32: 2003-2012.
12. Herriott R M, Meyer E M, Vogt M. Defined nongrowth media for stage II development of competence in *Haemophilus influenzae*. J Bacteriol 1970; 101: 517-524.
13. Sambrook J, Fritsch E F, Maniatis T. Molecular cloning: a laboratory manual, 2nd edn. NY: Cold Spring Harbor Laboratory Press, 1989.
14. Seifert S, Chen E Y, So M, Heffron F. Shuttle mutagenesis: A method of transposon mutagenesis for *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 1986; 83: 735-739.
15. Hitchcock P J, Brown T M. Morphological heterogeneity among *Salmonella* lipopolysaccharide chemotypes in silver-stained polyacrylamide gels. J Bacteriol 1983; 154: 269-277.
16. Lesse A J, Campagnari A A, Bittner W E, Apicella M A. Increased resolution of lipopolysaccharides and lipooligosaccharides utilizing tricine sodium dodecyl sulfate polyacrylamide gel electrophoresis. J Immunol Meth 1990; 126: 109-117.
17. Tsai C-M, Frasch C E. A sensitive silver stain for detecting lipopolysaccharides in polyacrylamide gels. Anal Biochem 1982; 119: 115-119.
18. Sanger F, Nicklen S, Coulson A R. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 1977; 74: 5463-5467.
19. Devereux J, Haeberli P, Smithies O. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 1984; 12: 387-395.
20. Pearson W R, Lipman D J. Improved tools for biological sequence analysis. Proc Natl Acad Sci USA 1988; 95: 2444-2448.
21. Laemmli U K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970; 227: 680-685.
20. Shine J, Dalgarno L. Determination of cistron specificity in bacterial ribosomes. Nature 1980; 254: 34-38.
23. Kyte J, Doolittle R F. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982; 157: 105-132.
24. Liu D, Verma N K, Romana L K, Reeves P R. Relationships among the *rfb* regions of *Salmonella* serovars A, B, and D. J Bacteriol 1991; 173: 4814-4819.
25. Marolda C L, Valvano M A. Identification, expression, and DNA sequence of the GDP-mannose biosynthesis genes encoded by the O7 *rfb* gene cluster of strain VW187 (*Escherichia coli* O7:K1). J Bacteriol 1993; 175: 148-158.
26. Luque F, Mitchenall L A, Chapman M, Christine R, Pau R N. Characterization of genes involved in molybdenum transport in *Azotobacter vinelandii*. Mol Microbiol 1993; 7: 447-459.
27. Schnaitman C A, Klena J D. Genetics of lipopolysaccharide biosynthesis in enteric bacteria. Microbiol Rev 1993; 57: 655-682.
28. Valvano M A. Personal communication.