Molecular Microbiology (1996) **22**(5), 951–965

Use of the complete genome sequence information of *Haemophilus influenzae* strain Rd to investigate lipopolysaccharide biosynthesis

Derek W. Hood,* Mary E. Deadman, Tina Allen, Hussein Masoud,[1] Adele Martin,[1] Jean Robert Brisson,[1] Robert Fleischmann,[2] J. Craig Venter,[2] James C. Richards[1] and E. Richard Moxon
Molecular Infectious Diseases Group, University of Oxford Department of Paediatrics, Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford, OX3 9DU, UK.
[1]Institute for Biological Sciences, National Research Council of Canada, Ottawa, Ontario, Canada, K1A 0R6.
[2]The Institute for Genomic Research, 9712 Medical Centre Drive, Rockville, Maryland 20850, USA.

Summary

The availability of the complete 1.83-megabase-pair sequence of the *Haemophilus influenzae* strain Rd genome has facilitated significant progress in investigating the biology of *H. influenzae* lipopolysaccharide (LPS), a major virulence determinant of this human pathogen. By searching the *H. influenzae* genomic database, with sequences of known LPS biosynthetic genes from other organisms, we identified and then cloned 25 candidate LPS genes. Construction of mutant strains and characterization of the LPS by reactivity with monoclonal antibodies, PAGE fractionation patterns and electrospray mass spectrometry comparative analysis have confirmed a potential role in LPS biosynthesis for the majority of these candidate genes. Virulence studies in the infant rat have allowed us to estimate the minimal LPS structure required for intravascular dissemination. This study is one of the first to demonstrate the rapidity, economy and completeness with which novel biological information can be accessed once the complete genome sequence of an organism is available.

Introduction

Haemophilus influenzae is an important cause of human disease worldwide. Serotype b capsular strains cause invasive, bacteraemic infections such as meningitis, septicaemia, epiglottitis, septic arthritis and empyema, particularly in infants. Strains lacking capsule are a common cause of otitis media, sinusitis, conjunctivitis and acute lower respiratory tract infections which account for many millions of childhood deaths in developing countries (Shann *et al*., 1984). Lipopolysaccharide (LPS) is essential to the integrity and functioning of the cell wall of *H. influenzae* and is a target for host immune responses. LPS is also involved in each stage of the pathogenesis of *H. influenzae* infections: colonization of the upper respiratory tract, systemic dissemination and the cytotoxic injury to target tissues, such as the central nervous system (CNS). The biosynthesis of this complex macromolecule requires numerous functionally diverse proteins to synthesize and activate precursor lipids and sugars, to attach the sugars in specific linkages, and to transport, assemble and anchor the mature glycolipid molecules in the outer leaflet of the cell envelope. *H. influenzae* LPS is composed of a membrane-anchoring lipid A portion linked by a single 2-keto-3-deoxyoctulosonic acid (KDO) molecule to a heterogeneous oligosaccharide composed of neutral heptose and hexose sugars (Zamze and Moxon, 1987; Phillips *et al*., 1993). The molecule lacks the O-specific side chain characteristic of the LPS of the enteric bacteria. Several of the surface-exposed epitopes of terminal hexose sugars are subject to high-frequency phase variation (Weiser *et al*., 1989; Jarosik and Hansen, 1994), an adaptive mechanism which is advantageous for survival of bacteria confronted by the differing micro-environments and immune responses of the host. This heterogeneity has complicated the structural and biological analysis of *H. influenzae* LPS; therefore the construction of mutant strains, with defined core structures, has been a priority. Prior to the availability of the complete genome sequence of *H. influenzae* strain Rd, progress was dictated by classical molecular genetic analysis and relatively few genes had been characterized (Fig. 1). These included the *lic* genes (Weiser *et al*., 1989; High *et al*., 1993) and the *lex2* gene (Jarosik and Hansen, 1994) involved with LPS phase variation, the *lsg* locus (Spinola *et al*., 1990), *galE* (Maskell *et al*., 1992), and *rfaE* (Lee *et al*., 1995), *rfaD* (A. Preston, personal communication) and *isn* (Preston *et al*., 1996; Zwahlen *et al*., 1985) genes which are involved with heptose biosynthesis. The majority of biosynthetic functions remain to be

Received 6 September, 1996; revised 15 October, 1996; accepted 16 October, 1996. *For correspondence. E-mail dhood@worf.molbiol.ox.ac.uk; Tel. (01865) 222347; Fax (01865) 222626.

© 1996 Blackwell Science Ltd

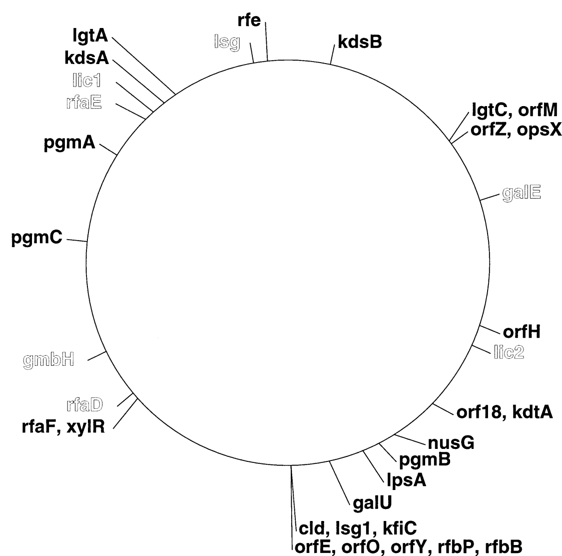

Fig. 1. Identification of genes for lipopolysaccharide biosynthesis in *H. influenzae*. The gene names shown in white outlined with black indicate those which had been identified and characterized as *H. influenzae* LPS-related genes prior to this study and which are present in the strain Rd genome database. The remaining genes were identified in this study and are shown at their relative positions on a circular 1.83 Mbp strain Rd genome map. *kdsA* and *kdsB* are KDO biosynthetic genes and the other loci are as listed in the text and Table 1.

characterized. The genome sequence of strain Rd has allowed us to address this problem and has facilitated rapid progress in identifying relevant candidate genes, constructing mutant strains, and investigating the biology of LPS in *H. influenzae*.

Results

Identification of LPS-related genes from the H. influenzae *strain Rd genome sequence*

The sequences of a wide range of LPS biosynthetic genes from Gram-negative bacteria were used as 'database probes' to screen for homologous genes, by DNA or amino acid similarity, in the *H. influenzae* strain Rd genome database prior to its annotation. Searches with amino acid sequence against translated strain Rd genome sequence proved the most useful for identification of putative LPS biosynthetic genes.

DNA sequences identified from the strain Rd genome sequence as potential homologues of LPS biosynthetic genes were translated, then searched against the combined GenBank/EMBL data banks to confirm their probable function(s). Because of a tendency in bacteria for clustering of genes with related functions, stretches of DNA sequence upstream and downstream of confirmed loci of interest were isolated and treated in a similar fashion. Table 1 lists the putative LPS-associated genes which were identified from the strain Rd genome sequence, designated by the highest homology match found after a BLAST search and their corresponding listing when the strain Rd genome sequence was published (Fleischmann *et al.*, 1995). Our directed search for candidate LPS genes allowed us to identify sequences listed either as genes of unknown function or as incomplete open reading frames (ORFs) in the published version of the strain Rd database. Several identified sequences were previously characterized *H. influenzae* genes and were not investigated further in this study.

Cloning of putative LPS-related loci, mutagenesis and construction of mutant strains

Following identification of candidate LPS genes, the relevant sequences were amplified by the polymerase chain reaction (PCR) from *H. influenzae* type d strain RM118 (Rd), the strain used for the sequencing project (Fleischmann *et al.*, 1995). In addition, we amplified the equivalent DNA tracts from two epidemiologically distinct encapsulated type b strains RM153 (Eagan) and RM7004 which have been used in our previous investigations of LPS biosynthesis. In all cases, the oligonucleotide primers designed from the strain Rd genome sequence allowed us to amplify the relevant loci from strain RM118 chromosomal DNA. Similarly, the primers enabled us to amplify most of the corresponding loci from the type b strains, with the exceptions of the *rfe*, *lgtA*, *rfbP* and *cld* loci. We suppose that this 16% failure rate reflects the divergence of the type b strains in the DNA sequences selected for the primers, but may also reflect possible differences in the genetic organization between the strains. For some loci the primers were designed to hybridize against sequence from adjacent genes or intergenic regions (data not shown).

PCR-amplified products were cloned, mutated by insertional inactivation, then transformed into the appropriate *H. influenzae* strains to obtain mutants by allelic replacement. For the majority of loci, a kanamycin-resistance cassette was inserted within the first 30% of the predicted reading frame. When tested for any given locus, the orientation of the kanamycin cassette could have an effect on the number of transformants obtained, but there was no overall preference for cassette orientation between the different loci. Some polar effect on adjacent genes could not always be eliminated, but this was minimized. Table 1 lists the putative LPS-associated genes which have been cloned and the mutant strains derived for 22 of the genes. Mutations have been made in strains RM118, RM7004 and RM153 for comparison. No differences were found by independently constructing strains mutated in the same gene but using cloned PCR products derived from a different strain. Only the disrupted *kdtA* and *rfaH*

Table 1. Candidate LPS-associated genes identified and cloned through the *H. influenzae* strain Rd genome sequence. The locus designation corresponds to the highest value match after a BLAST search of the ORF against the combined GenBank/EMBL data banks. The mutant strains obtained after transformation into *H. influenzae* strains are: Rd, RM118; E, RM153 (Eagan); 7, RM7004. The Rd database number corresponds to the entry of the equivalent locus in the published strain Rd genome sequence (Fleischmann *et al*., 1995). For *kdtA*, *orf18* (*kdtB*) and *nusG*, no transformants were obtained.

Locus	Locus Designation		Homology	Mutant Strain	Rd Database Number
	species	homologue function			
lgtC	*N. gonorrhoeae*	Glycosyl transferase	$2.9e^{-37}$	Rd, E, 7	0258, 0259
orfM	*Mycobacterium leprae*	Unknown	$1.2e^{-34}$	Rd, E, 7	0260
orfZ		No homology		Rd, E, 7	–
opsX	*X. campestris*	Saccharide biosynthesis	$7.8e^{-51}$	Rd, E, 7	0261
rfaF	*S. typhimurium*	Heptosyl transferase	$5.7e^{-156}$	Rd, E, 7	1105
xylR (*rhaR*)	*E. coli*	Xylose regulator	$7.9e^{-156}$	Rd, E, 7	1106
cld	*E. coli*	Chain length determinant	$6.5e^{-16}$	Rd, E, 7	0866
lsg1	*H. influenzae*		$1.7e^{-178}$	Rd, E, 7	0867
kfiC	*E. coli*	Galactosyl transferase	$1.1e^{-24}$	Rd, E, 7	0868
orfE	*Erwinia*	Glycosyl transferase		Rd, E, 7	0869
orfO	*Salmonella choleraesuis*	O-antigen polymerase	$9.6e^{-07}$	Rd, E, 7	0870
orfY		No homology		Rd, E, 7	0871
rfbP	*S. typhimurium*	Galactosyl transferase (O-antigen)	$1.3e^{-193}$	E, 7	0872
rfbB	*Neisseria meningitidis*	TDP-glucose dehydratase	$8.0e^{-160}$	Rd, E, 7	0873
orf18(*kdtB*)	*E. coli*	Unknown	$1.3e^{-53}$		0651
kdtA	*E. coli*	KDO transferase	$1.5e^{-148}$		0652
pgmA	*E. coli*	Phosphoglucosamine mutase	$2.3e^{-230}$	Rd, E, 7	1463
pgmB	*M. leprae*	Phosphoglucomutase	$4.7e^{-107}$	Rd, E, 7	0740
pgmC	*E. coli*	Phosphoglucosamine mutase	$4.3e^{-231}$	Rd, E, 7	1337
lpsA	*Pasteurella haemolytica*	Glycosyl transferase	$1.6e^{-80}$	Rd, E, 7	0765
rfe	*E. coli*	GlcNAc transferase	$3.9e^{-143}$	Rd, E, 7	1716
lgtA	*N. gonorrhoeae*	Glucosyl transferase	$1.3e^{-102}$	Rd	1578
galU	*E. coli*	UDP-glucose synthetase	$2.0e^{-147}$	Rd, E, 7	0812
nusG(*rfaH*)	*E. coli*	Transcriptional terminator	$5.9e^{-91}$		0717
orfH		Heptosyl transferase	0.00024	Rd, E, 7	0523

genes and *orf18* failed to produce any mutants after repeated transformation. This was not unexpected for the *kdtA* gene, involved with the addition of KDO to lipid A, as KDO is regarded as an essential component for LPS function and is required for cell viability. The function of *orf18* (*kdtB*) is unknown but it is transcriptionally coupled to *kdtA* in a similar manner in *Escherichia coli* (Schnaitman and Klena, 1993). The *rfaH* homologue had sequence similarity to transcriptional regulators, such as *nusG*, which are essential for other cellular processes (Burova *et al*., 1995). Also, the construction of strains with a mutant *lgtA* gene has proved most difficult and may indicate that LgtA has some other function, for example, attachment of glucosamine residues to the outer portion of the lipid A molecule, or an involvement with murein biosynthesis.

Map position of LPS-related genes in the strain Rd genome sequence

The most detailed data on LPS biosynthesis has been obtained for members of the Enterobacteriaceae, where genes involved in the synthesis of the core or O-antigen portion of the LPS molecule are located together on the chromosome as large regulons (Schnaitman and Klena, 1993). Previously characterized LPS genes in *H. influenzae* are located at multiple loci scattered around the chromosome (Fig. 1). Several of the LPS-associated genes identified in this study were linked or clustered (Fig. 1). The *lgtC* and *opsX* genes were separated by two small ORFs, *orfM* and *orfZ*, with no significant homology to genes of known function. The *xylR* and *rfaF* genes were contiguous and separated by eight ORFs associated with sugar metabolism from a previously characterized LPS-related gene, *rfaD*. The *cld*, *lsg1* and *kfiC* genes were contiguous and separated by one large ORF (*orfY*) and probably two smaller ORFs, *orfO* and *orfE*, from the *rfbP* and *rfbB* genes. One of these intervening genes (*orfO*) had low homology to O-antigen polymerase from *Salmonella* and, given its proximity to the *rfb* genes, which are O-antigen specific in other organisms, this would suggest that this region of the chromosome had a function related to O-antigen biosynthesis. This was surprising as repetitive sugar side-chains analogous to O-antigen are not present in the 'rough-type' LPS of *H. influenzae.* Genes encoding the subunits of the KDO transferase (*kdtA* and *kdtB*) were adjacent to each other whilst the KDO biosynthetic genes, *kdsA* and *kdsB*, which were not investigated in this study, were unlinked. All the remaining loci identified were unlinked

(Fig. 1). *pgmA* and *pgmC* are highly related genes (99.9% DNA sequence identity) which may have arisen by recent duplication of a small segment (≈5 kb) of the strain Rd chromosome.

Analysis of LPS from mutant strains by immunoblotting

LPS molecules from wild-type and mutant strains were analysed by colony immunoblotting using a panel of monoclonal antibodies (mAbs), each specific for particular LPS epitopes (Gulig *et al*., 1987; Borrelli *et al*., 1995; Borrelli, 1996). Table 2 shows the reactivities of mAbs 4C4, 6A2, 12D9, 5G8 and 7E7 to the wild-type strain RM7004 and mutant derivatives. With the exception of the *orfO*, *orfE*, *orfY*, *rfbB*, *kfiC*, *lsg1*, *xylR*, and *rfe* mutant strains, each showed altered reactivity to one or more mAb when compared to the wild type. The first six of these eight strains were mutated in genes clustered in the 'O-antigen-specific' chromosomal segment and may indicate that these genes are cryptic, at least under laboratory growth, as has been described in *Neisseria*, a related pathogen (Robertson *et al*., 1994). A *pgmA/C* mutant strain showed altered reactivity to one of a further bank of MAHI mAbs (Borrelli, 1996) (data not shown). In our other type b strain, RM153, *orfO*, *kfiC*, *xylR* and *rfe* mutants do exhibit some changes in their reactivity with mAbs, presumably due to the differences in the most-abundant LPS molecules expressed between our closely related type b strains (see below).

Mutations in the *opsX* and *rfaF* genes and *orfH* of strains RM153 and RM118 resulted in increased binding of several mAbs after colony immunoblotting but not when purified LPS was tested by Western analysis or dot immunoblotting (data not shown).

Table 2. Reactivity of *H. influenzae* strain RM7004 and mutant strains derived from RM7004 with mAbs.

	Monoclonal Antibody				
Strain	4C4	6A2	12D9	5G8	7E7
Wild type					
RM7004	++++	++++	++++	+++	++
Mutant					
lgtC	–	√	√	–	√
orfM	√	+	√	–	√
orfZ	√	√	√	–	–
opsX	–	–	–	√	+
rfaF	–	–	–	–	+
xylR	√	√	√	√	√
cld	–	–	–	√	–
lsg1	√	√	√	√	√
kfiC	√	√	√	√	√
orfE	√	√	√	√	√
orfO	√	√	√	√	√
orfY	√	√	√	√	√
rfbP	–	–	√	√	ND
rfbB	√	ND	√	ND	ND
pgmA/C	√	√	ND	√	√
pgmB	√	–	√	√	√
lpsA	√	–	–	+	√
rfe	√	√	√	√	√
galU	–	–	–	–	–
orfH	+	–	√	+	√

The reactivities of the wild-type strain, RM7004, are given: ++++, very strong reactivity; +++, strong reactivity; ++, weaker reactivity. The relative reactivities of the mutant strains to the wild type are given as: √, unchanged reactivity; –, reduced reactivity; +, increased reactivity. ND, not determined.

Analysis of purified LPS by Tricine (T)–SDS–PAGE

LPS purified from wild-type strain RM7004 and then fractionated by T–SDS–PAGE showed a migration pattern similar to strain RM118 but of reduced mobility when compared to LPS from strain RM153 (Fig. 2). This difference between the two type b strains is due to addition of further hexose residues in the outer core region of the LPS molecules in strain RM7004 (J. C. Richards, unpublished). Figure 2A shows the LPS profiles of mutant strains derived from strain RM7004. Considering, first, the inner core of the LPS, *orfH*, *rfaF* and *opsX* mutant strains exhibited the most-truncated LPS molecules, the latter being of the highest mobility and similar to the truncated LPS of an *isn* mutant (Helander *et al*., 1988; Preston *et al*., 1996) (data not shown). The *isn* phenotype arises from a mutation in the heptose biosynthetic gene, a homologue of *gmhA* (Brook and Valvano, 1996) of *E. coli*, and is considered to be the minimal functional LPS structure comprising a single KDO linked to the lipid A. *opsX*, a gene affecting LPS and exopolysaccharide biosynthesis in *Xanthomonas campestris* (Kingsley *et al*., 1993), had some homology to *rfaQ* and heptosyl transferases (Schnaitman and Klena, 1993). This gene may encode the transferase responsible for adding the first heptose (heptose I, Fig. 3) to the KDO. LPS derived from the *rfaF* mutant would correspond to a structure containing an additional sugar, heptose I (Fig. 2B). DNA and amino acid homology comparisons would support the function of RfaF as a heptosyl-II transferase (Schnaitman and Klena, 1993). RfaF was highly homologous (80% amino acid similarity) to RfaF from *Salmonella typhimurium* and *E. coli*. Our initial search of the Rd genome sequence failed to identify a candidate gene product to add the third heptose to the growing LPS chain (Fig. 3). A further search of the sequence database for less-conserved homologues of known heptosyl transferases, or transferases which add sugars to the second heptose of LPS in other organisms, such as *rfaK* from *Neisseria* (Kahler *et al*., 1996) and *rfaG* from *E. coli* (Schnaitman and Klena, 1993), showed sequence with a very low match to *orfH*, an ORF of unknown function. Mutation of *orfH* resulted in a strain with LPS of a size corresponding to that predicted for a molecule containing two heptose sugars (Fig. 2B); therefore *orfH* may encode the heptose-III

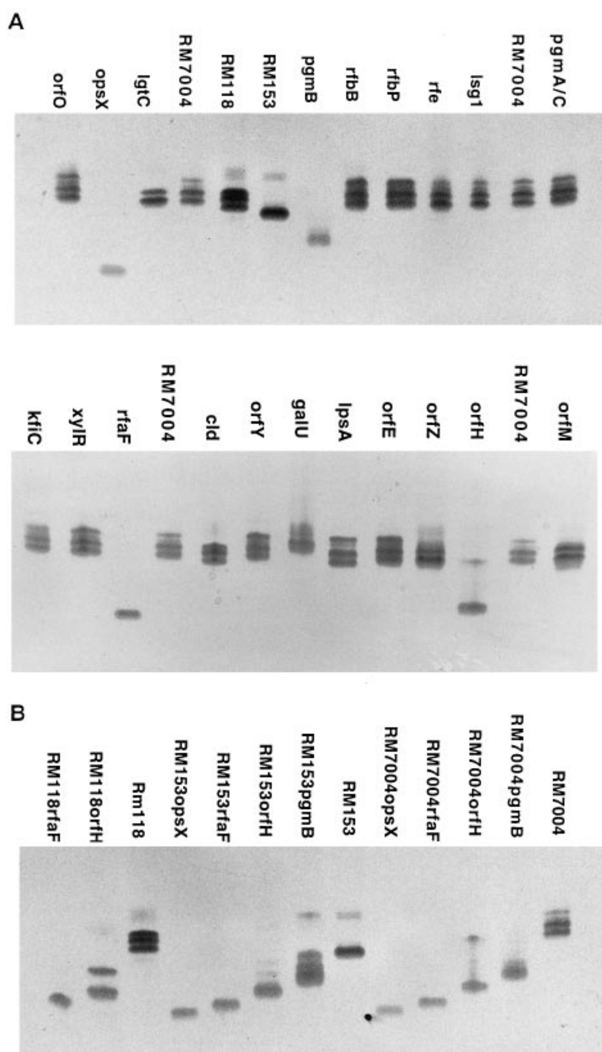

Fig. 2. The electrophoretic gel-migration patterns after T–SDS–PAGE of LPS purified from *H. influenzae* strains.
A. The upper and lower panels are LPS isolated from strains RM7004, RM153, RM118 and mutant strains derived from strain RM7004.
B. LPS isolated from strains RM7004, RM153, RM118 and selected deep-rough mutants derived from each.

transferase. The electrophoretic mobilities of LPS from the equivalent inner-core mutants (*opsX*, *rfaF* and *orfH*) in all three *H. influenzae* strains were very similar (Fig. 2B).

A heteropolymer of glucose and galactose sugars is built from the heptose backbone of the LPS molecule of *H. influenzae*. Phosphoglucomutase (*pgmB*) and the *galU* gene products are involved in the conversion of glucose-6-phosphate to UDP-glucose, the activated form of the sugar used in macromolecular biosynthesis, and thus a *pgm* mutation would limit incorporation of glucose residues and result in a truncated LPS molecule (Sandlin and Stein, 1994) (Fig. 2). *pgmB*, in common with other related genes, showed equal homology to phosphoglucomutase and phosphomannomutase, but mannose is not known to be a component of *Haemophilus* LPS. *pgmB* mutants produced an LPS of reduced molecular weight but with a complex pattern. After T–SDS–PAGE, the fastest-migrating species of LPS corresponded to molecules with no added glucose or galactose (see Table 4 later) and smaller than the major LPS species of a *galE* mutant (data not shown). LPS of higher molecular weight probably contained extensions of one, two and three hexose sugars. *galU* encodes the enzyme which catalyses the final step in the conversion of glucose-6-phosphate to UDP-glucose and in strains RM118 and RM153, the LPS derived from the relevant mutants is truncated in a similar manner to the *pgmB*-mutant-derived LPS (data not shown). However, in strain RM7004, the *galU* mutation produces some truncated LPS molecules, but the greatest proportion of LPS corresponds to the highest-molecular-weight species seen in the wild-type (Fig. 2). The reason for this difference between strain RM7004 and the others is not clear. The homology of *pgmA* and *pgmC* to *glmM* of *E. coli*, a gene encoding phosphoglucosamine mutase (Mengin-Lecreulx and van Heijenoort, 1996), suggests that these genes also encode a hexosephosphate mutase. The amino acid sequences of PgmA and PgmC contain a consensus for a hexosephosphate mutase active site with a 14/15 match to the same sequence in GlmM (PgmB has a 9/15 match to this sequence). If PgmA and PgmC are phosphoglucosamine mutases then mutations in these genes may have a pleiotropic effect on the supply of UDP-GlcNAc for the synthesis of peptidoglycan and both the lipid A and perhaps the peripheral sugars on some LPS molecules. Single mutations in either of these highly related genes, *pgmA* and *pgmC*, would have little detectable effect on LPS structure if the gene functions are effectively identical. This is confirmed by our experimental observations (Fig. 2).

The *lgtC*, *orfM*, *lpsA*, *cld*, and *orfZ* mutant strains had altered migration patterns of LPS molecules with a shift towards those of higher mobility, probably with at least one sugar less in a major LPS species than in the wild type, RM7004 (Fig. 2). This would be consistent with predicted functions for the gene products as terminal sugar transferases, or other proteins involved with the process. The simpler migration pattern of LPS from the RM7004*lgtC* mutant strain would indicate that the relevant sugar added to the basal structure was normally available for extension to form the highest-molecular-weight LPS species observed for the wild type. LgtC is a homologue of a known galactosyl transferase in *Neisseria gonorrhoeae* (Gotschlich, 1994). The *lgtC* locus had an associated tetranucleotide repeat motif; there were 23 copies of GACA in strain Rd, and 22 and 32 copies in strains RM153 and RM7004, respectively (Hood *et al*., 1996). The repeat sequences were located at the extreme 5′ end of the predicted translated reading frame in a similar

fashion to the CAAT and GCAA repeats reported previously for other LPS-associated genes (Weiser *et al.*, 1989; Jarosik and Hansen, 1994). An RM7004 strain mutated in *orfM*, the ORF adjacent to *lgtC*, produces an LPS gel-migration pattern similar, but not identical, to that from RM7004*lgtC*. In strain RM153, the *lgtC* and *orfM* mutants produce a similar or identical LPS to each other and it is therefore probable that the two genes are involved with biosynthesis of the same epitope(s). An RM7004*lpsA* mutant strain had LPS with a gel-migration pattern shifted by more than one sugar when compared to the wild-type LPS. LpsA is part of a group of presumed galactosyl transferases, typified by Lic2A and LgtB, which have been studied in *Haemophilus* and *Neisseria* species, respectively (High *et al.*, 1993; Jennings *et al.*, 1995). A correlation between the alteration of LPS structure found in strains mutated in the *cld* and *orfZ* loci with predicted function of the genes from homology comparisons cannot easily be made.

No apparent alteration of LPS structure was found for strain RM7004 mutated in the *rfbB*, *rfbP*, *xylR*, *rfe*, *lsg1*, *pgmA/C*, *kfiC*, *orfO*, *orfY* and *orfE* loci. Very minor changes in LPS structure will not always be detected by T–SDS–PAGE but, importantly, some effect of mutations in the *rfe*, *kfiC*, *rfbB*, *rfbP*, *orfY* and *orfE* loci could be seen in the LPS obtained from either strain RM153 or RM118 (data not shown). This emphasises the complexity of *H. influenzae* LPS structure in that even very closely related strains do exhibit differences in the terminal LPS epitopes elaborated.

Analysis of LPS from mutant strains by mass spectrometry

Preliminary comparative analysis, by mass spectrometry, of the LPS from selected mutants derived from strain RM153 has helped to confirm some of our predictions of gene function and the LPS biosynthetic pathway. O-deacylated samples of LPS were examined by negative-ion electrospray mass spectrometry (ES-MS), and the data are shown in Table 3. Initial structural analysis confirmed that the LPS from the *opsX* mutant contained no heptose and that the *rfaF* mutant contained one heptose molecule linked to the KDO. This agrees with the proposed functions of the gene products as heptose-I and heptose-II transferases, respectively (Fig. 3). Analysis of the LPS derived from the RM153*orfH* mutant strain showed that it contained two heptose molecules and supports the function of the *orfH* gene product as a heptose-III transferase. LPS from strain RM153*pgmB* had the lowest-molecular-weight species corresponding to only the inner core, i.e. with three heptose molecules and no added hexose sugars. The higher-molecular-weight LPS corresponded to components having one, two and three added hexose residues although their order of attachment is unknown.

The structure of LPS from the parent strain was recently determined (Fig. 3) (Masoud *et al.*, 1996). Detailed nuclear magnetic resonance (NMR) analysis of LPS-derived oligosaccharide from the RM153*lgtC* mutant showed it to be truncated with no galactose added by extension from heptose II. ES-MS analysis indicated that LPS from an RM153*lpsA* mutant strain was more truncated than LPS from an RM153*lgtC* mutant with no galactose (by sugar analysis) added to heptose II or to heptose III (Fig. 3). This would be consistent with functions for LgtC and LpsA as galactosyl transferases. Preliminary analysis of the LPS from the RM153*orfM* mutant strain showed truncation of the molecule equivalent to that of RM153*lgtC*, and again supports a related function for the *orfM* product in elaboration of the terminal digalactoside structure from

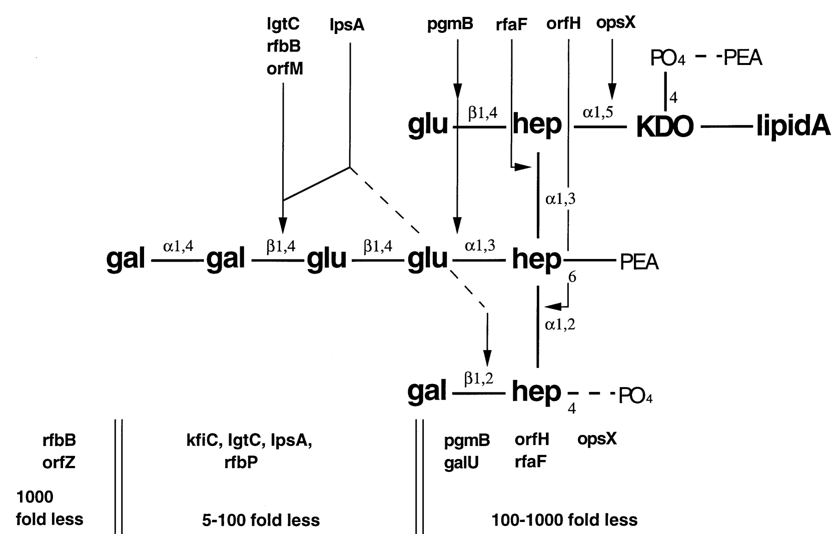

Fig. 3. A schematic representation of the structure of LPS from *H. influenzae* strain RM153 based on the results of the analysis of Masoud *et al.* (1996). The proposed sites of action in LPS biosynthesis of loci characterized after the complete genome sequence of strain Rd was elucidated are shown above the structure, based on the combined results from T–SDS–PAGE and comparitive ES-MS analysis of purified LPS. The grouping of mutant strains shown below the structure is from our virulence experiments. Represented in the LPS structure are: KDO, 2-keto-3-deoxyoctulosonic acid; hep, L-glycero-D-manno-heptose; glu, D-glucose; gal, D-galactose; PEA, phosphoethanolamine; and PO_4, phosphate. The heptose residues (hep) are listed from top to bottom as heptose I, heptose II then heptose III.

Table 3. Negative-ion ES-MS analysis and proposed composition for the major components for O-deacylated LPS of *H. influenzae* RM153 mutant strains.

Mutant Strain	Observed Ions (*m/z*)		Molecular Weight (Da)		Proposed Composition
	$(M\text{-}3H)^{3-}$	$(M\text{-}2H)^{2-}$	observed	calculated	
opsX	ND	625.4*	1252.7	1253.2	KDO–P–LipidA–OH
	ND	687.0	1375.9	1376.2	PEA$_1$–KDO–P–LipidA–OH
rfaF	ND	721.3*	1444.8	1445.4	Hep$_1$–KDO–P–LipidA–OH
	ND	782.8	1568.2	1568.4	Hep$_1$–PEA$_1$–KDO–P–LipidA–OH
orfH	ND	817.5*	1637.0	1637.6	Hep$_2$–KDO–P–LipidA–OH
	ND	878.8	1759.6	1760.6	Hep$_2$–KDO–PEA–P–LipidA–OH
pgmB	649.7	974.9*	1952.0	1952.9	Hep$_3$–PEA$_1$–KDO–P–LipidA–OH
	690.5	1036.5	2074.7	2075.9	Hep$_3$–PEA$_2$–KDO–P–LipidA–OH
	703.8	1056.3	2114.5	2115.0	Hex$_1$–Hep$_3$–PEA$_1$–KDO–P–LipidA–OH
	757.9	1137.4	2276.4	2277.1	Hex$_2$–Hep$_3$–PEA$_1$–KDO–P–LipidA–OH
lpsA	812.1	1218.3	2438.9	2439.2	Hex$_3$–Hep$_3$–PEA$_1$–KDO–P–LipidA–OH
	838.9	ND	2519.7	2519.2	Hex$_3$–Hep$_3$–PEA$_1$–KDO–P$_2$–LipidA–OH
	852.9*	1279.8	2562.0	2562.3	Hex$_3$–Hep$_3$–PEA$_2$–KDO–P–LipidA–OH
lgtC	865.8	1299.1	2600.7	2601.3	Hex$_4$–Hep$_3$–PEA$_1$–KDO–P–LipidA–OH
	892.5	ND	2681.5	2681.3	Hex$_4$–Hep$_3$–PEA$_1$–KDO–P$_2$–LipidA–OH
	907.8*	1360.8	2723.6	2724.4	Hex$_4$–Hep$_3$–PEA$_2$–KDO–P–LipidA–OH
	933.5	ND	2803.8	2804.4	Hex$_4$–Hep$_3$–PEA$_2$–KDO–P$_2$–LipidA–OH
rfbB	865.9*	1299.2	2600.7	2601.3	Hex$_4$–Hep$_3$–PEA$_1$–KDO–P–LipidA–OH
	892.6	ND	2680.8	2681.3	Hex$_4$–Hep$_3$–PEA$_1$–KDO–P$_2$–LipidA–OH
	907.0	ND	2723.6	2724.4	Hex$_4$–Hep$_3$–PEA$_2$–KDO–P–LipidA–OH
orfM	865.9*	1299.1	2600.5	2601.3	Hex$_4$–Hep$_3$–PEA$_1$–KDO–P–LipidA–OH
	896.6	ND	2680.8	2681.3	Hex$_4$–Hep$_3$–PEA$_1$–KDO–P$_2$–LipidA–OH
	907.0	1360.7	2723.6	2724.4	Hex$_4$–Hep$_3$–PEA$_2$–KDO–P–LipidA–OH
	933.6	ND	2804.0	2804.4	Hex$_4$–Hep$_3$–PEA$_2$–KDO–P$_2$–LipidA–OH

Samples were analysed by direct infusion in the negative-ion mode on a VG Quattro mass spectrometer. The base peak is depicted by an asterisk in each LPS preparation; only signals having peak heights greater than 15% of base peaks are listed. Average mass units were used for calculation of molecular-weight values based on the proposed compositions of Gibson *et al.* (1993). Sugar analysis indicates Hep corresponds to L-glycero-D-manno-heptose; Hex corresponds to D-glucose for *lpsA* and to D-glucose or D-galactose for *lgtC*, *rfbB* and *orfM* in the ratios of 3:1. For *rfbB*, a minor component corresponding to a Hex$_3$-containing species gave rise to $(M\text{-}3H)^3$ signal at an *m/z* of 812.1. Wild-type RM153 LPS contains a mixture of species comprising between 2 and 6 hexose residues (Masoud *et al.*, 1966). ND, not detected.

heptose II. An RM153 *rfbB* mutant had LPS that was similarly truncated. The major LPS species from an RM153 *orfE* mutant was the same or similar to that of wild-type LPS.

Virulence of mutant strains in the infant-rat model

The range of LPS mutants constructed in this study has allowed us, for the first time, to undertake a comprehensive study to correlate LPS structure with the virulence of *H. influenzae* strains in the infant-rat model of infection (Smith *et al.*, 1973). Table 4 shows the results obtained for intravascular dissemination after intraperitoneal inoculation with mutants derived from strain RM153 (Eagan). Strains which were attenuated and resulted in a level of bacteraemia greater than two logs below that of the wild type were those mutated in the *opsX*, *rfaF*, *pgmB*, *galU* and *rfbB* genes and *orfZ* and *orfH*. *opsX*, *rfaF*, *orfH*, *galU*, *pgmB* and *orfZ* mutants produced the most truncated LPS molecules in ascending order of complexity (Fig. 3). Mutant strains with more subtle changes to the LPS gel migration profile (Fig. 2), namely those mutated in the *lgtC*, *lpsA* and *rfbP* genes, gave rise to levels of bacteraemia 5–100-fold less than that of the wild type, a statistically significant difference in each case (Table 4). Apart from *orfM*, strains mutant in the *rfe*, *cld*, *pgmA/pgmC*, *kfiC*, *lsg1* and *xylR* genes and *orfO* and *orfY* showed no detectable alteration of LPS profile in strain RM153 and no significant difference in virulence when compared to the wild type. Thus, a correlation between LPS structure and virulence can be drawn (Fig. 3). The exceptions from the rat-model data are the strains mutated in the *rfbB* gene and *orfZ*. Although there was only an apparently minor change to the LPS profile of these mutants when compared to the wild type, there was a decrease of approx. 1000-fold in the level of bacteraemia attained. This strain with a mutation in *rfbB* was constructed twice to ensure that the observed attenuation was related to the *rfbB* mutation and not to a fortuitous secondary mutation, and similar results were obtained each time.

No gross change in other virulence determinants such as capsule or outer-membrane proteins has yet been found for any of the mutants, but minor changes have not yet been ruled out for all mutations in all strain backgrounds.

Table 4. Virulence of isogenic mutants of *H. influenzae* strain RM153 (Eagan) in the infant-rat model of infection. The numbers indicated in parentheses are those for the wild-type control in the same experiment. The levels of bacteraemia are those after 48 h following intraperitoneal inoculation with 50–150 organisms. A minimum bacteraemia of 200 organisms ml^{-1} of blood is given for any animal in which no organisms were detected during the experiment. A geometric mean value of 2.30 is the minimum value obtainable with no animals exhibiting bacteraemia during the experiment.

Mutant Strain	No. of rats	No. of rats bacteraemic	Geometric Mean (log_{10}) of Bacteraemia	Mann–Whitney *p* Value
lgtC	50 (49)	42 (45)	5.15 (6.07)	0.0045
orfM	13 (12)	11 (12)	5.36 (5.60)	1.0000
orfZ	12 (12)	1 (12)	2.61 (6.83)	0.0001
opsX	23 (22)	1 (22)	2.47 (6.61)	0.0001
rfaF	23 (22)	2 (22)	2.67 (6.61)	0.0001
xylR	12 (11)	11 (8)	5.52 (5.16)	0.8771
cld	12 (12)	11 (12)	5.62 (5.70)	0.5831
lsg1	12 (12)	10 (6)	5.93 (4.29)	0.1327
kfiC	12 (12)	7 (12)	5.31 (6.50)	0.5252
orfE	12 (12)	10 (12)	5.73 (6.83)	0.1573
orfO	12 (12)	5 (7)	4.04 (4.45)	0.4160
orfY	12 (12)	11 (6)	6.02 (4.29)	0.0831
rfbP	12 (12)	9 (12)	4.72 (6.50)	0.0027
rfbB	24 (24)	3 (22)	2.60 (5.69)	0.0001
pgmA/C	12 (12)	12 (12)	6.00 (5.70)	0.5067
pgmB	24 (22)	0 (22)	2.30 (6.61)	0.0001
lpsA	24 (22)	10 (22)	3.87 (6.24)	0.0003
rfe	11 (11)	8 (8)	4.58 (5.16)	0.2494
galU	12 (12)	0 (8)	2.30 (4.45)	0.0135
orfH	12 (12)	0 (12)	2.30 (6.50)	0.0001

Molecular modelling by Metropolis Monte Carlo (MMC) calculations

Molecular modelling was used to generate a three-dimensional model and to calculate conformational flexibility of the core oligosaccharide region of LPS from strain RM153 (Masoud *et al*., 1996). The LPS consists of a conserved heptose-containing inner-core trisaccharide unit attached via a phosphorylated KDO residue to a lipid A component. In the major LPS population group, each heptose is substituted with further chain elongation from the central unit where molecules containing four, five and six residues are most prevalent (Fig. 3). The inner-core element is also a common feature of LPS from strain RM7004 (Schweda *et al*., 1993). Space-filling and ball-and-stick models of the core oligosaccharide are shown in Fig. 4. As noted previously for LPS from *E. coli* (Kastowsky *et al*., 1992), the MMC calculations indicated that the heptose-containing inner-core region forms a compact structural unit. The galabiose-containing side-chain attached to heptose II adopts an elongated shape and it shows considerable conformational flexibility. It was found that even though these outer-core regions of the molecule are flexible, the heptose III and the β-glucose in the side-chain were spatially arranged on the same side of the molecule. This modelling is being extended to investigate the conformation of the LPS from the mutants derived from strain RM153 and strains RM7004 and RM118.

Discussion

The complete genome sequence of *H. influenzae* strain Rd has allowed us to identify 25 candidate LPS genes. Detection of homologues of LPS biosynthetic genes in the *H. influenzae* genome database using DNA sequences as probes was much more rapid and reliable than could have been achieved by hybridization experiments. However, the major advantage of the genome sequence information was to identify candidate loci, found only, or most reliably, by amino acid homology. This is particularly important when studying biological systems such as LPS where it is not uncommon for proteins of related function, such as the sugar transferases, to be encoded by genes of divergent sequence. A total of 60% of the genes described in Table 1 could not have been reliably identified by DNA sequence alone, and therefore would not have been found by hybridization experiments using the relevant heterologous probe.

The 25 LPS-related genes identified in this study were dispersed around the genome of strain Rd either singly or in small groups. The largest cluster of novel genes were the eight contiguous ORFs with an apparent bias of function directed towards O-antigen biosynthesis. LPS from *H. influenzae* lacks an O-antigen-specific side-chain but it is found in some related species, for example *Actinobacillus pleuropneumoniae* (Altman *et al*., 1988). This region of DNA may have some function in other macromolecular synthesis, such as elaboration or modification of capsule polysaccharide, but as yet we have found no change in capsule production in the type b mutant strains. The alteration of LPS seen in some of these mutants might indicate that less typical sugars, *rfbB* is rhamnose specific, or modifications of existing sugars can be incorporated into the LPS as minor components under certain, as yet uncharacterized, conditions. Rhamnose is a component of the O-antigen of many pathogenic *Salmonella* strains

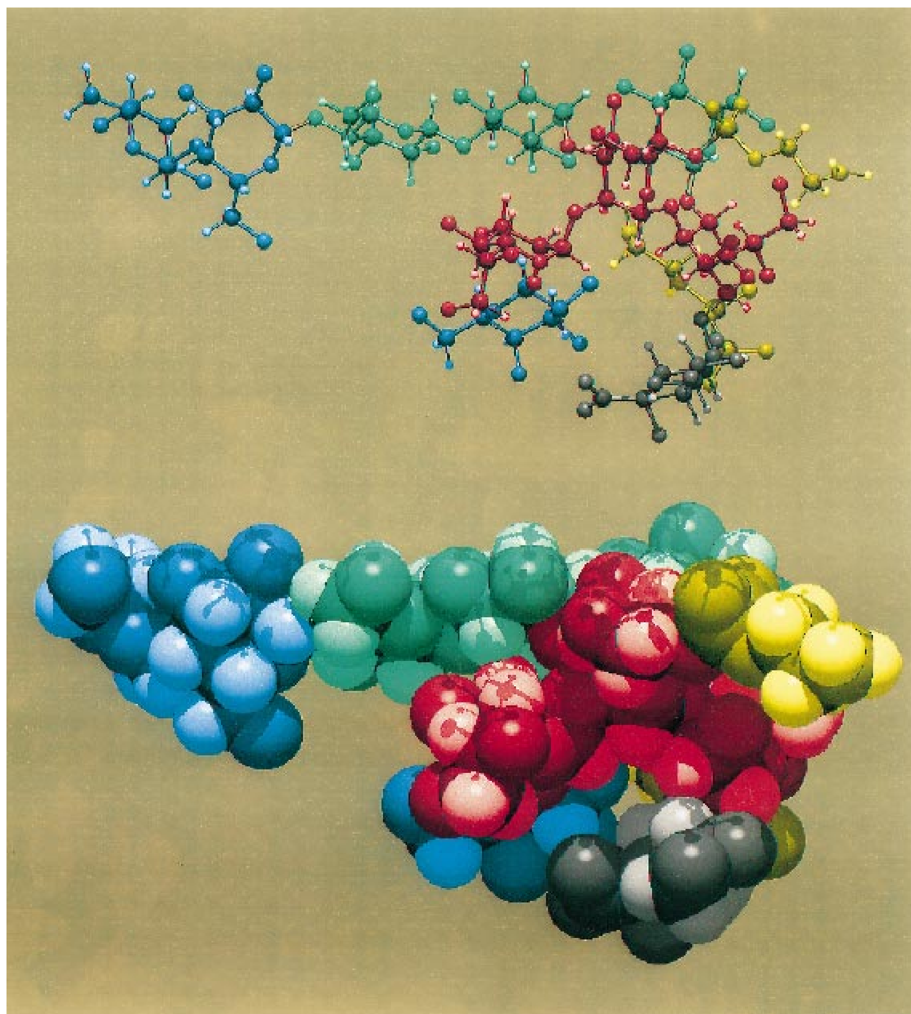

Fig. 4. Space-filling and ball-and-stick models of an extended conformation of the galabiose-containing core oligosaccharide region of LPS from *H. influenzae* strain RM153 generated from the MCC calculations. At the bottom of model, the KDO moiety (grey) is substituted at the O-4 position by a pyrophosphoethanolamine group (yellow, behind). The heptose-containing trisaccharide (red) is attached to the O-5 position of KDO. Heptose I is substituted at O-4 by a glucose (green), while heptose III bears a galactose (blue) residue at the O-2 position. The main oligosaccharide chain (tetrasaccharide) is attached to O-3 of heptose II which also carries a phosphoethanolamine group (yellow, in front) at O-6. This tetrasaccharide chain, containing the terminal galabiose moiety (blue), shows considerable flexibility.

(Jann and Jann, 1984) and is synthesized from glucose by the *rfbA–D* gene products. Homologues of the other genes are not known to be present in *H. influenzae*. There may be changes in the pattern of LPS expressed *in vivo* when compared to that analysed after repeated laboratory culture. Transcription studies must confirm whether adjacent genes are functionally linked and co-regulated.

By T–SDS–PAGE analysis and mAb reactivity we have shown an effect on LPS structure in strains derived from RM7004 which are mutated in 11 of the 25 identified loci (*lgtC*, *orfM*, *orfZ*, *opsX*, *rfaF*, *cld*, *rfbP*, *pgmB*, *lpsA*, *galU* and *orfH*). Analysis of the mutants derived from strains RM153 and RM118 showed an association with LPS structure for a further seven of these loci (*xylR*, *kfiC*, *orfE*, *orfO*, *orfY*, *rfbB* and *rfe*). Other genes, such as *kdtA* and *nusG*, are probably essential to the cells' viability and have LPS-related function implied by homology comparisons alone. The electrophoretic pattern and preliminary structural analysis of LPS isolated from the mutant strains allows us to make some predictions as to the initial steps involved in the biosynthesis of the *H. influenzae* LPS molecule. It is known that the LPS from the type b strains RM153 and RM7004 have identical inner-core structures although the latter has a predominance of higher-molecular-weight species (J. Richards, unpublished). LPS from the type d-derived strain, RM118, apparently has a very similar basal core structure to the type b LPS, and deep-rough mutants of all three strains had equivalent electrophoretic mobility. The LPS inner-core

structure is synthesized by the apparent sequential addition of three heptose molecules to the KDO, possibly catalysed by the gene products of *opsX*, *rfaF* and *orfH*, respectively (Fig. 3). Once complete, the hexose sugars can be added to the heptose backbone. A small amount of extension can be seen from the *orfH* mutant LPS in strain RM7004 (Fig. 2) and may be due to a leaky mutation allowing some addition of heptose III rather than addition of hexoses to a two-heptose inner core. This observation will be confirmed through detailed structural analysis of the LPS species. The addition of hexose sugars to each heptose may be concurrent rather than sequential. No single mutation has been characterized which blocks the addition of a hexose sugar to any of the three heptoses and prevents all subsequent additions.

Mutations in the *pgmB* and *galU* genes, which affect the supply of activated glucose precursor, have similar effects on LPS biosynthesis as have been reported previously for *galE* and *galK* mutants (Maskell *et al*., 1992). The major LPS species from a strain with a mutation in the *pgmB* gene comprises only the inner-core structure of three heptose residues. As galactose is not included in the major LPS species, it might indicate that the normal extension of hexose sugars from the inner core (Fig. 3) requires glucose to be added to heptose I and/or heptose II before galactose is added to heptose III. The higher-molecular-weight species of LPS from the *pgmB* mutant must indicate that either alternative pathways to make glucose-1-phosphate exist within the cell, or other sugars, such as galactose, may substitute in the LPS under atypical conditions. UDP-glucose is an important biosynthetic precursor and has been postulated as a potential intracellular signal molecule in *E. coli* (Bohringer *et al*., 1995). It can be replenished from UDP-galactose via the *galE* gene product in the presence of exogenous galactose (Maskell *et al*., 1992). Strains mutated in both the *pgmB* and *galE* genes are being constructed to test this and detailed structural analysis of the LPS species from the relevant mutant strains will be required to resolve this issue. Structural analysis of LPS will also help to resolve the anomaly of why LPS derived from equivalent mutations, such as *galU*, in our three test strains is apparently different. Further work, including the construction of a *pgmA/pgmC* double-mutant strain, is required to confirm the role of the other hexosephosphate mutase genes, *pgmA* and *pgmC*. *N*-acetyl glucosamine is not a major constituent of *H. influenzae* LPS but has been reported in the type b strain A2 as a minor component (Philips *et al*., 1993) and in the related species *H. ducreyi* (Schweda *et al*., 1994).

The *lgtC* and *lpsA* genes and *orfM* have predicted functions in the elaboration of the outer core of the LPS molecule but the role of other gene products is less clear. The *rfbP* gene in *Salmonella* is known to encode a product of dual function which helps transfer UDP-galactose linked to acyl-carrier lipid across the cytoplasmic membrane as the first step in O-antigen biosynthesis (Wang *et al*., 1996). As *Haemophilus* has no O-antigen, the observed effect of mutation in RM7004 may be on the transfer of galactose to the outer-most portion of the LPS core through a similar mechanism involving an acyl-carrier intermediate. The genes for *cld* and *orfZ* have no homology to known glycosyl transferases but their influence is similar to that of *lgtC* and *orfM*. Not all of these gene products would be expected to have a direct function as sugar transferases but might be involved with accessory functions such as presentation or supply of the activated sugars or regulation of the transfer process. The complex pattern of LPS expression in *Haemophilus* makes it difficult to assign putative transferase functions to genes solely on the basis of electrophoretic profiles and mAb reactivities of LPS from the mutants. The complex LPS-migration patterns observed on gels may mask any minor effect of mutating some of these loci, a problem which is compounded by some differences in the expression pattern of terminal LPS epitopes both between the wild types of the two type b strains RM153 and RM7004 and the type d strain, RM118, and their mutant derivatives. It remains to be clarified whether similar sugar linkages at different positions on the LPS molecule are always synthesized by completely independent transferases or whether other transferases can compensate for functions lost by mutation. Complete structural analysis of LPS from mutant strains, the construction of multiply-mutant strains and biochemical characterization will be required to identify the relevant function and contribution of each gene product. It also remains to be proved that all the major direct LPS biosynthetic functions have been identified from the strain Rd genome database and that these functions are sufficient for LPS production in the type b strains. This will always present a potential problem when applying information based on the genome content of the sequenced strain to other closely related strains and species. But, we have identified sufficient potential transferase functions to account for all the known variations in type b strain LPS structure.

A major interest of our laboratory is in the phase variation of LPS epitopes exerted through nucleotide repeats situated within biosynthetic genes. These repeat regions, normally composed of tetranucleotides, are subject to loss or gain of one or more 4-bp repeats through slipped-strand mispairing during DNA replication, and result in translational frame-shifts and non-functional gene products. This reversible on-off switching of LPS biosynthetic functions at high frequency underlies the phase-variable phenotype characteristic of *H. influenzae* LPS. A further gene contributing to phase variation of LPS, *lgtC*, has been identified from the strain Rd genome

sequence. The loss of reactivity to mAbs 4C4 and 5G8 of the *lgtC* mutant strains indicated that LgtC influences the same region of LPS core structure as does Lic2 A (CAAT repeats) (High *et al*., 1993). The mAb 4C4 recognizes the galα(1–4)βgal-containing epitope which is expressed in the LPS of related bacteria and also in human cells (Virji *et al*., 1990). This epitope is expressed as the terminal extension from heptose II in the LPS from strain RM153 (Masoud *et al*., 1996), and loss of 4C4 reactivity is consistent with the truncation of the LPS structure found by structural studies of the LPS from an RM153*lgtC* mutant strain. A contribution of LgtC to the phase-variable expression of the digalactoside epitope has been shown (Hood *et al*., 1996). The importance of the galα(1–4)βgal structure for molecular mimicry may explain why several genes, at least two of which promote phase variation, control its expression. Interestingly, the *lgtC* gene in *Neisseria* has a homopolymeric tract of guanosine nucleotides located within the reading frame which appears to be responsible for the phase-variable expression of a terminal digalactoside epitope of the LPS (Gotschlich, 1994). Knowledge of the complete repertoire of repeated motifs in strain Rd (Hood *et al*., 1996) allows us to propose that the CAAT- and GACA-associated loci are the only structural genes contributing directly to a phase-variable LPS phenotype in that strain. The LPS biosynthetic locus (*lex2*) associated with a GCAA-repeat motif in type b strains (Jarosik and Hansen, 1994) was not present in the strain Rd genome sequence.

The range of mutant strains obtained in this study has not allowed us, thus far, to define unambiguously the specific LPS epitopes recognized by other mAbs. Mutants of strains RM153 and RM118 (*opsX*, *rfaF* and *orfH*) with severely truncated LPS show increased reactivity to some of the mAbs 6A2, 12D9, 4C4 and 5G8, when compared to the wild type, upon colony immunoblotting. It is possible that minor high-molecular-weight LPS species, which can still be synthesized, are contributing to significant mAb reactivity. However, the majority of this reactivity was lost when purified LPS was investigated by immunoblotting, and thus an element of mAb-binding specificity must be determined by deep-core structure presentation or other factors on the bacterial cell surface. It seems evident that in many instances minor differences in LPS structure between strains has a profound effect on mAb binding and that for at least some of the mAbs used in this study, the LPS epitopes recognized appear to be more complex than may have been assumed. The reasons for this become more apparent when considering the proposed three-dimensional structure of the LPS molecule (Fig. 4). The conformational flexibility implied in this model of the core oligosaccharide does not fully explain the orientation of the oligosaccharide in relation to the bacterial cell membrane, but provides a starting point for understanding the accessibility of oligosaccharide epitopes and their reactivities with mAbs. The contribution of individual structures within LPS populations is being investigated by Western blot studies and further structural analysis and modelling of LPS derived from mutant strains should allow a better insight into mAb-binding specificity.

Identification of the LPS epitopes critical to virulence has allowed us to fulfil one of the main aims of our biological investigation of *H. influenzae* LPS. Results from our experiments on virulence combined with data from other previously identified LPS-related mutants allow us to make some predictions as to the minimal LPS structure required for efficient intravascular dissemination of *H. influenzae* in the infant rat. Three heptose molecules and at least two hexose sugars are required for maintenance of high levels of bacteraemia in the infant rat after intraperitoneal inoculation. Under these conditions there is good correlation of structure with function, with mutants with a majority of LPS molecules with less than two hexose sugars being severely attenuated and strains with higher-molecular-weight LPS being only mildly or not at all attenuated. The exception is RM153*rfbB* where the minor change in the LPS, but dramatic attenuation in the infant rats, is unexplained. Possibilities include a role for the gene in some more subtle modifications which may affect only a portion of the LPS molecules or may be an *in vivo* specific effect not detected after laboratory culture. *rfbB* is involved with the production of rhamnose, a sugar not routinely associated with LPS from *Haemophilus* but which has been identified in one type b strain as a minor component (J. C. Richards, unpublished). It is therefore possible that some subtle additions or modifications to LPS structure are host specific and are not detected after laboratory culture. Mutants with truncated LPS tested to date are not defective in other virulence determinants, nor are they impaired for growth *in vitro*, and it is reasonable to conclude that this attenuation is LPS specific. It is evident that factors influencing maintenance of bacteria in the bloodstream are very different to those affecting colonization and invasion, and a further important correlation between LPS size, structure and virulence may be evident after experiments following intra-nasal inoculation of the infant rats have been completed. The accumulated information will allow us to identify key LPS genes and to extend our study to the LPS structure in other strains. Studies are underway to confirm the extent of conservation of LPS-related genes and the core LPS structure across all *H. influenzae* types and should help evaluate the use of LPS core structure as a potential candidate for broad-range vaccine development.

In summary, this study has allowed us to rapidly advance our understanding of *H. influenzae* LPS biology. We have identified genes involved in the crucial stages of LPS biosynthesis: precursor supply, sugar transferase,

export and potential regulatory functions. Mutations in these genes have allowed us to fulfil our main objective and to compile a bank of mutant strains elaborating a range of LPS molecules. These have helped to elucidate some of the steps in LPS biosynthesis and to estimate a minimal structure required for systemic infection of the infant rat. This study has demonstrated the potential use of whole genome sequencing to extend biological knowledge. The speed and ease of detection of genes is significantly greater than that by classical molecular genetic analysis and, in particular, allows the identification of genes found even under circumstances of weak amino acid homology. In other organisms, the availability of sequence from random clones giving almost complete coverage of the genome should still allow successful gene identification.

Experimental procedures

Bacterial strains and culture conditions

The *H. influenzae* type d strain RM118 (strain KW-20 obtained from H. O. Smith, the same source for the strain used in the sequencing project of Fleischmann *et al.* (1995)) (Alexander and Leidy, 1951), and the type b strains RM153 (Eagan) and RM7004 (both are disease isolates from the USA (Anderson *et al.*, 1972) and the Netherlands (van Alphen *et al.*, 1983), respectively) were used in this study. *H. influenzae* strains were grown at 37°C in brain–heart infusion (BHI) broth supplemented with haemin ($10\,\mu g\,ml^{-1}$) and NAD ($2\,\mu g\,ml^{-1}$). For selection after transformation, kanamycin ($10\,\mu g\,ml^{-1}$) was added to the growth medium.

E. coli strain DH5α was used to propagate cloned PCR products and was grown at 37°C in Luria–Bertani (LB) broth supplemented with ampicillin ($100\,\mu g\,ml^{-1}$) (Sambrook *et al.*, 1989).

Identification of LPS-related genes from the H. influenzae *genome sequence*

Data bank DNA or amino acid sequences of LPS biosynthetic genes from a wide range of organisms were searched against our database of the *H. influenzae* strain Rd genomic DNA sequence with FASTA or BLAST searches (Devereux *et al.*, 1984). Searches were carried out using early versions of the strain Rd genome sequence prior to its completion and annotation. Minimum cut-off values of 40% homology for a DNA match and a smallest sum probability of e^{-10} for amino acid similarity (BLAST) over at least 50% of any locus were used for initial identification and selection. The DNA sequences of loci of interest were isolated and searched against the combined GenBank/EMBL databases to predict the position and function of identified ORFs. Regions of DNA sequence, typically 1.5 kb, on either side of each target locus were treated in a similar manner.

Recombinant DNA methodology and nucleotide sequence analysis

Restriction endonucleases and DNA-modifying enzymes were obtained from Boehringer Mannheim and used according to the manufacturer's instructions.

Plasmid DNA was prepared from *E. coli* strains by the alkaline-lysis method (Sambrook *et al.*, 1989). Chromosomal DNA was prepared from *Haemophilus* by the method described previously (High *et al.*, 1993). The nucleotide sequence of double-stranded DNA was analysed by the Sanger dideoxy chain-termination method using Sequenase (United States Biochemical Corp.). Oligonucleotide primers were synthesized on an Applied Biosystem Model AB1392 synthesizer. PCR amplification was performed using Taq polymerase in 50 μl reaction volumes in a Perkin Elmer DNA thermal cycler. Southern blotting and hybridization analysis were performed as described by Sambrook *et al.* (1989).

Cloning of putative LPS-related genes, mutagenesis and construction of mutant strains

Oligonucleotide primers were designed from the strain Rd genome sequence to amplify loci of interest by PCR. A total of 1–1.5 kb of DNA around each target locus was amplified by PCR for 1 min periods of denaturation (94°C), annealing (45°C) and polymerization (72°C) for 30 cycles. A total of 1 μl of PCR product was ligated with 50 ng of plasmid pT7 Blue (Novagen) or pCRII (Invitrogen) vector and transformed into *E. coli* strain DH5α. Plasmid DNA was prepared from transformants and cloned inserts were confirmed by restriction endonuclease digestion and sequencing from plasmid-specific primers. Target genes were then mutated by insertional inactivation using a kanamycin-resistance cassette (Km^R; from pUC4Kan; Pharmacia) placed into suitable restriction sites obtained from the strain Rd genome sequence or by restriction mapping. Restriction sites were routinely selected within the first half of any given reading frame. A total of 2–3 μg of linearized plasmid DNA containing the mutated cloned DNA of interest was transformed into *H. influenzae* strains RM118, RM153 and RM7004 by the method of Herriot *et al.* (1970) and transformants were selected on kanamycin. The type b strains were transformed with equal efficiency by DNA originating from either type b strain or strain Rd. Transformants were checked by re-culturing on BHI/kanamycin plates, then were confirmed as mutants by PCR amplification and Southern blotting/hybridization of endonuclease-digested chromosomal DNA.

Analysis of lipopolysaccharide by immunoblotting

The reactivity of wild-type and mutant strains of *H. influenzae* to a panel of LPS-specific monoclonal antibodies was analysed as described by Roche *et al.* (1994). Aliquots of 1 μl of purified LPS were also blotted to nitrocellulose filters, dried at room temperature, then reacted with mAbs as described above.

Analysis of lipopolysaccharide by electrophoresis

The patterns of LPS isolated from wild-type and mutant strains were determined after fractionation by T–SDS–PAGE (Lesse *et al.*, 1990) essentially as described by Roche *et al.* (1994), but using 17.7% acrylamide gels.

Virulence of mutant strains in the infant-rat model

The virulence of *H. influenzae* strain RM153 and RM153-derived LPS mutants was determined in infant rats, a well-validated model of *H. influenzae* infection (Smith *et al.*, 1973). Five-days-old Sprague-Dawley rats from natural litters, reduced to 12 at birth, were inoculated intraperitoneally with an inoculum of 50–150 organisms in 0.1 ml of phosphate-buffered saline. Bacteraemia was assessed by diluting and plating 0.01 ml of blood sampled from all surviving rats after 48 h by tail-vein puncture. Mortality was assessed daily up to 5 d. The statistical significance of the data was determined by Mann–Witney and Student *t*-tests.

Mass spectrometric analysis of purified LPS

The complete structure of the LPS molecules from *H. influenzae* strain RM153 has previously been determined by ES-MS and detailed NMR analysis of samples of partially and completely deacylated LPS (Masoud *et al.*, 1996). O-deacylated LPS from mutant strains derived from RM153 was prepared for comparative analysis with the wild-type structure by ES-MS on a VG Quattro mass spectrometer (Micromass) as detailed by Masoud *et al.* (1996).

Cells from 10 l batch cultures (10 lots of 1 l) were harvested after overnight growth then LPS was extracted by the hot phenol–water method (Westphal and Jann, 1965) as described by Masoud *et al.* (1996). The LPS was O-deacylated and then 0.5 mg was analysed by ES-MS as described by Masoud *et al.* (1996). LPS from RM153*lgtC* and RM153*lpsA* was further analysed by NMR (Masoud *et al.*, 1996).

Molecular modelling

The starting geometry for all sugars was submitted to a complete refinement of bond lengths, valence and torsion angles by using the molecular mechanics program MM3(92) (QPCE) (Burket and Allinger, 1982). All calculations were performed using the minimized co-ordinates for the methyl glycoside. For a given disaccharide, the conformational analysis was evaluated using the potential for oligosaccharide (PFOS) approach which included the contributions arising from van der Waals interactions, torsional and exo-anometric potential. Rotational barriers of 1.0 kcal mol^{-1} and 0.5 kcal mol^{-1} were used for ϕ and φ, respectively. No electrostatic contribution was considered. The ring geometry was treated as invariant, and hydroxylic hydrogen atoms were not taken into account. The linkage conformation was determined from ϕ = (H1–C1–O1–Cx), and φ = (C!–O1–Cx–Hx), where x refers to the aglycon-linkage site. All angles are expressed in degrees. The glycosidic bridge angle was set to 117°. The phosphorus groups were generated from standard co-ordinates (Alchemy, Tripos software) and minimum energy conformations found in crystal structures (Saenger *et al.*, 1984). The MMC calculations were performed as described previously (Peters *et al.*, 1993). All pendant groups were treated as invariant except for the phosphorus groups which were allowed to rotate about the Cx–Ox and Ox–P bonds. The starting angles for the oligosaccharide were taken from the minimum energy conformers calculated for each disaccharide unit present in the molecule. A 24-dimensional MMC calculation of the decasaccharide with the PO_4 and PPO_4 groups attached was carried out with 5000 macro moves. The graphics were generated using the Schakal software (Egbert Keller, Kristallographisches Institut der Universitaet, Freiberg, Germany).

Acknowledgements

Derek Hood, Mary Deadman and Tina Allen were supported by a programme grant from the Medical Research Council, UK. Grateful thanks to Eric Hansen and Alan Lesse for providing monoclonal antibodies and to Don Krajcarski for assistance in ES-MS analysis.

References

Alexander, H.E., and Leidy, G. (1951) Determination of inherited traits of *H. influenzae* by deoxyribonucleic acid fractions isolated from type specific cells. *J Exp Med* **93:** 345–359.

Altman, E., Brisson, J.R., and Perry, M.B. (1988) Structure of the O-antigen polysaccharide of *Haemophilus pleuropneumoniae* serotype 3 (ATCC 27090) lipopolysaccharide. *Carbohydr Res* **179:** 245–248.

Anderson, P., Johnston, R., and Smith, D. (1972) Human serum activities against *Haemophilus influenzae* type b. *J Clin Invest* **51:** 31–38.

Bohringer, R., Fischer, D., Mosler, G., and Hengge-Aronis, R. (1995) UDP-glucose is a potential intracellular signal molecule in the control of expression of σ^S and σ^S-dependent genes in *Escherichia coli*. *J Bacteriol* **177:** 413–422.

Borrelli, S. (1996) Monoclonal antibodies against *Haemophilus* lipopolysaccharides. PhD Thesis. Dept of Immunology, Microbiology, Pathology, and Infectious Disease, Karolinska Institute, Sweden.

Borrelli, S., Hegedus, O., Shaw, D.H., Jansson, P.-E., and Lindberg, A.A. (1995) The tetrasaccharide L-α-D-heptose 1-3-L-α-D-heptose 1-(3-deoxy-D-manno-octulosonic acid) and phosphate in lipid A define the conserved epitope in *Haemophilus* lipopolysaccharides recognised by a monoclonal antibody. *Infect Immun* **63:** 3683–3692.

Brook, J.S., and Valvano, M.A. (1996) Molecular cloning of the *Haemophilus influenzae gmhA* (*lpcA*) gene encoding a phosphoheptose isomerase required for lipooligosaccharide biosynthesis. *J Bacteriol* **178:** 3339–3341.

Burket, U., and Allinger, N.L. (1982) *Molecular Mechanisms*. Washington, DC: American Chemical Society, ACS Monograph 177.

Burova, E., Hung, S.C., Sagitov, V., Stitt, B.L., and Gottesman, M.E. (1995) *Escherichia coli* NusG protein stimulates transcription rates *in vivo* and *in vitro*. *J Bacteriol* **177:** 1388–1392.

Devereux, J., Haeberli, P., and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the Vax. *Nucl Acids Res* **12:** 387–395.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Butt, C.J., Tomb, J.-F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D.,

Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O., and Venter, J.C. (1995) The genome of *Haemophilus influenzae* Rd. *Science* **269:** 496–512.

Gibson, B.W., Melaugh, W., Phillips, N.J., Apicella, M.A., Campagnari, A.A., and Griffiss, J.M. (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.

Gotschlich, E.C. (1994) Genetic locus for the biosynthesis of the variable portion of *Neisseria gonorrhoeae* lipooligosaccharide. *J Exp Med* **180:** 2181–2190.

Gulig, P.A., Patrick, C.C., Hermanstorfer, L., McCracken, Jr, G.H., and Hansen, E.J. (1987) Conservation of epitopes in the oligosaccharide portion of the lipooligosaccharide of *Haemophilus influenzae* type b. *Infect Immun* **55:** 513–520.

Helander, I.M., Lindner, B., Brade, H., Altmann, K., Lindberg, K.K., Rietschel, E.T., and Zahringer, U. (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.

Herriot, R.M., Meyer, E.M., and Vogt, M.J. (1970) Defined non-growth media for stage II development of competence in *Haemophilus influenzae*. *J Bacteriol* **101:** 517–524.

High, N.J., Deadman, M.E., and Moxon, E.R. (1993) The role of the repetitive DNA motif (5′-CAAT-3′) in the variable expression of the *Haemophilus influenzae* lipopolysaccharide epitope αGal (1–4)βGal. *Mol Microbiol* **9:** 1275–1282.

Hood, D.W., Deadman, M.E., Jennings, M.P., Bisceric, M., Fleischmann, R.D., Venter, J.C., and Moxon, E.R. (1996) DNA repeats identify novel virulence genes in *Haemophilus influenzae*. *Proc Natl Acad Sci USA* **93:** 1121–1125.

Jann, K., and Jann, B. (1984) Structure and biosynthesis of O-antigen. In *Handbook of Endotoxin*. Vol. 1. Reitschel, E.T. (ed.). Amsterdam: Elsevier Science Publisher, pp. 138–186.

Jarosik, G.P., and Hansen, E.J. (1994) Identification of a new locus involved in the expression of *Haemophilus influenzae* type b lipooligosaccharide. *Infect Immun* **62:** 4861–4867.

Jennings, M.P., Hood, D.W., Peak, I.R.A., Virji, M., and Moxon, E.R. (1995) Molecular analysis of a locus for the biosynthesis and phase-variable expression of the lacto-*N*-neotetraose terminal lipopolysaccharide structure in *Neisseria meningitidis. Mol Microbiol* **18:** 729–740.

Kahler, C.M., Carlson, R.W., Rahman, M.M., Martin, L.E., and Stephens, D.S. (1996) Inner core biosynthesis of lipooligosaccharide (LOS) in *Neisseria meningitidis* serogroup B: identification and role of the α-1,2 *N*-acetylglucosamine transferase (RfaK). *J Bacteriol* **178:** 1265–1273.

Kastowsky, M., Gutberlet, T., and Bradaczek, H. (1992) Molecular modelling of three-dimensional structure and conformation flexibility of bacterial lipopolysaccharide. *J Bacteriol* **174:** 4798–4806.

Kingsley, M.T., Gabriel, D.W., Marlow, G.C., and Roberts, P.D. (1993) The *opsX* locus of *Xanthomonas campestris* affects host range and biosynthesis of lipopolysaccharide and extracellular polysaccharide. *J Bacteriol* **175:** 5839–5850.

Lee, N., Sunshine, M.G., and Apicella, M.A. (1995) Molecular cloning and characterization of the non-typeable *Haemophilus influenzae* 2019 *rfaE* gene required for lipopolysaccharide biosynthesis. *Infect Immun* **63:** 818–824.

Lesse, A.J., Campagnari, A.A., Bittner, W.E., Apicella, M.A. (1990) Increased resolution of lipopolysaccharides and lipooligosaccharides utilising tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *J Immunol Methods* **126:** 109–117.

Maskell, D.J., Szabo, M.J., Deadman, M.E., and Moxon, E.R. (1992) The *gal* locus from *Haemophilus influenzae*: cloning, sequencing and the use of *gal* mutants to study lipopolysaccharide. *Mol Microbiol* **6:** 3051–3063.

Masoud, H., Moxon, E.R., Martin, A., Krajcarski, D., and Richards, J.C. (1996) Structure of the variable and conserved lipopolysacharide oligosaccharide epitopes expressed by *Haemophilus influenzae* serotype b strain Eagan. *Biochemistry*, in press.

Mengin-Lecreulx, D., and van Heijenoort, J. (1996) Characterisation of the essential gene *glmM* encoding phosphoglucosamine mutase in *Escherichia coli. J Biol Chem* **271:** 32–29.

Peters, T., Meyer, B., Stuike-Prill, R., Somorjai, R., and Brisson, J.R. (1993) A Monte Carlo method of conformational analysis of saccharides. *Carbohydr Res* **238:** 49–73.

Phillips, N.J., Apicella, M.A., Griffiss, J.M., and Gibson, B.W. (1993) Structural studies of the lipooligosaccharides from *Haemophilus influenzae* type b strain A2. *Biochem* **32:** 2003–2012.

Preston, A., Maskell, D., Johnson, A., and Moxon, E.R. (1996) Altered lipopolysaccharide characteristic of the I69 phenotype in *Haemophilus influenzae* results from mutation in a novel gene, *isn*. *J Bacteriol* **178:** 396–402.

Robertson, B.D., Frosch, M., and van Putten, J.P.M. (1994) The identification of cryptic rhamnose biosynthetic genes in *Neisseria gonorrhoeae* and their relationship to lipopolysaccharide biosynthesis. *J Bacteriol* **176:** 6915–6920.

Roche, R.J., High, N.J., and Moxon, E.R. (1994) Phase variation of *Haemophilus influenzae* lipopolysaccharide: characterisation of lipopolysaccharide from individual colonies. *FEMS Microbiol Lett* **120:** 279–284.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbour, New York: Cold Spring Harbour Laboratory Press.

Sandlin, R.C., and Stein, D.C. (1994) Role of phosphoglucomutase in lipooligosaccharide biosynthesis in *Neisseria gonorrhoeae*. *J Bacteriol* **176:** 2930–2937.

Schnaitman, C.A., and Klena, J.D. (1993) Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol Rev* **57:** 655–682.

Schweda, E.K.M., Hegedus, O.E., Borrelli, S., Lindberg, A.A., Weiser, J.W., Maskell, D.J., and Moxon, E.R. (1993) Structural studies of the saccharide portion of cell envelope lipopolysaccharide from *Haemophilus influenzae* strain AH1-3 ($lic3^{+}$). *Carbohydr Res* **246:** 319–330.

Schweda, E.K.H., Sundstrom, A.C., Eriksson, L.M., and Lindberg, A.A. (1994) Structural studies of the cell envelope lipopolysaccharides from *Haemophilus ducreyi* strains ITM2665 and ITM4747. *J Biol Chem* **16:** 12040–12048.

Seanger, W. (1984) *Principles of Nucleic Acid Structure*. New York: Springer-Verlag.

Shann, F., Grattan, M., Germer, S., Linnsman, V., Hazlem, D., and Payne, R. (1984) Aetiology of pneumonia in children in Goroke Hospital, Papua New Guinea. *Lancet* **ii:** 537–541.

Smith, A.L., Smith, D.H., Averill, D.R., and Moxon, E.R. (1973) Production of *Haemophilus influenzae* b meningitis in infant rats by intraperitoneal inoculation. *Infect Immun* **8:** 278–290.

Spinola, S.M., Kwaik, Y.A., Lesse, A.J., Campagnari, A.A., and Apicella, M.A. (1990) 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* **58:** 1558–1564.

Van Alphen, L., Riemens, T., Poolman, J., Hopman, C., and Zanen, H.C. (1983) Homogeneity of cell envelope protein sub-types, lipopolysaccharide serotypes and biotypes among *Haemophilus influenzae* type b from patients with meningitis in the Netherlands. *J Infect Dis* **148:** 75–81.

Virji, M., Weiser, J.N., Lindberg, A.A., and Moxon, E.R. (1990) Antigenic similarities in lipopolysaccharides of *Haemophilus* and *Neisseria* and expression of a digalactoside structure also present on human cells. *Microb Pathog* **9:** 441–50.

Wang, L., Liu, D., and Reeves, P.R. (1996) C-terminal half of *Salmonella enterica* WbaP (RfbP) is the galactosyl-1 phosphate transferase domain catalaysing the first step of O-antigen biosynthesis. *J Bacteriol* **178:** 2598–2604.

Weiser, J.N., Love, J.M., and Moxon, E.R. (1989) The molecular mechanism of *Haemophilus influenzae* lipopolysaccharide epitopes. *Cell* **59:** 657–665.

Westphal, O., and Jann, K. (1965) Bacterial lipopolysaccharides. Extraction with phenol–water and further applications of the procedure. *Meth Carbohydr Chem* **5:** 83–91.

Zamze, S.E., and Moxon, E.R. (1987) Composition of the lipopolysaccharide from different capsular serotype strains of *Haemophilus influenzae. J Gen Microbiol* **133:** 1443–1451.

Zwahlen, A., Rubin, L.G., Connelly, C.J., Inzana, T.J., and Moxon, E.R. (1985) Alteration of the cell wall of *Haemophilus influenzae* type b by transformation with cloned DNA: association with attenuated virulence. *J Infect Dis* **152:** 485–492.