Molecular and Genetic Characterization of the Capsule Biosynthesis Locus of *Streptococcus pneumoniae* Type 19B

JUDY K. MORONA,1 RENATO MORONA,2 AND JAMES C. PATON1*

Molecular Microbiology Unit, Women’s and Children’s Hospital, North Adelaide, South Australia 5006,1 and Department of Microbiology and Immunology, University of Adelaide, Adelaide, South Australia 5005,2 Australia

Received 27 March 1997/Accepted 27 May 1997

We have previously reported the nucleotide sequence of the *Streptococcus pneumoniae* type 19F capsular polysaccharide synthesis locus (cps19f), which consists of 15 open reading frames (ORFs) designated cps19fA to -O. Hybridization analysis indicated that close homologs for cps19fA to -H and cps19fK to -O were found in type 19B, but there were no homologs for cps19fI and -J. In this study we used long-range PCR to amplify and clone a 10.5-kb section of the *S. pneumoniae* type 19B capsule locus (cps19b) between cps19bH and cps19bK. This region of the cps19b locus is 4 kb larger than that in the cps19f locus and replaces cps19fI and cps19fJ with five new ORFs, designated cps19bP, -I, -Q, -R, and -J. We have proposed functions for four of the protein products, including functional homologs of Cps19fI and Cps19fJ. Transformation of a *S. pneumoniae* mutant containing an interrupted type 19F capsule locus with the 10.5-kb cps19b PCR product converted the recipient strain to type 19B. Southern hybridization analysis indicated that cps19bP, -I, -Q, -R, and -J are unique to type 19B and the closely related type 19C.

*Streptococcus pneumoniae* (the pneumococcus) is an important cause of invasive disease in human populations throughout the world, resulting in high morbidity and mortality. Control of pneumococcal disease is being complicated by the increasing prevalence of antibiotic-resistant strains and the suboptimal clinical efficacy of existing vaccines. *S. pneumoniae* produces a polysaccharide capsule, which is essential for virulence because it protects the pneumococcus from the nonspecific immune defenses of the host during an infection (2). All fresh isolates from patients with pneumococcal infection are encapsulated, and spontaneous nonencapsulated (rough) derivatives of such strains are almost completely avirulent.

There are now 90 recognized serotypes of *S. pneumoniae* (9), each of which produces a structurally distinct capsular polysaccharide (CPS). Classical genetic studies carried out by Austrian et al. (3) demonstrated that the *S. pneumoniae* genes required for biosynthesis and expression of CPS are closely linked on the pneumococcal chromosome. This fact enabled us to clone and sequence the capsule locus from *S. pneumoniae* type 19F (designated cps19f) (7, 21). Our studies were concentrated on *S. pneumoniae* type 19F because it is one of the commonest causes of invasive disease in children and the type 19F CPS is one of the poorest immunogens in this group (6). Group 19 consists of the immuno-cross-reactive *S. pneumoniae* types 19F, 19A, 19B, and 19C. In one study, group 19 pneumococci accounted for 7% of the isolates from cases of invasive disease (26). Of these, 65% were caused by type 19F, 34% were caused by type 19A, and 1% were caused by type 19B; type 19C was a very rare cause of disease in this study.

The structures of the immunologically cross-reactive type 19F and type 19B CPSs are quite distinct (Fig. 1), with the latter having an extra sugar in the backbone and a disaccharide side chain. Thus, the type 19B capsule locus (designated cps19b) would be predicted to contain extra genes required for biosynthesis of the more complicated type 19B CPS repeat unit, as well as a different polysaccharide repeat unit transporter and polysaccharide polymerase. In the present study, we report the isolation and characterization of the portion of cps19b which encodes these type-specific functions.

**Bacterial strains.** The *S. pneumoniae* strain Rx1-19F-I, an unencapsulated insertion-duplication mutant (in which the cps19f gene has been interrupted) of Rx1-19F (a derivative of Rx1 expressing the type 19F capsule), was constructed as described elsewhere (21). A clinical isolate of *S. pneumoniae* type 19B was obtained from Chi-Jen Lee, Center for Biologics, Food and Drug Administration, Bethesda, Md.; other clinical isolates were from the Women’s and Children’s Hospital, Adelaide, Australia. Pneumococci were routinely grown in Todd-Hewitt broth with 0.5% yeast extract or on blood agar. Where appropriate, erythromycin was added to the media at a concentration of 0.2 μg/ml. The production of capsules by pneumococci was assessed by quellung reaction, using factor-specific antisera obtained from Statens Seruminstitut, Copenhagen, Denmark. This was performed by M. Gratton, Acute Respiratory Infections Reference and Research Unit, Centre for Public Health Sciences, Queensland Health, Brisbane, Australia.

**Isolation of the type 19B cps genes.** In a recent study (21) we used Southern hybridization to demonstrate that all but 2 of the 15 genes in the *S. pneumoniae* type 19F capsule locus hybridized to the DNA of a type 19B pneumococcal strain. These two genes, cps19fI and cps19fJ, are located together near the middle of the cps19f locus. DNA from the two flanking genes, cps19fH and cps19fK, hybridized weakly to 19B DNA, whereas all other cps19f genes hybridized strongly at high stringency (21).

These data suggested that the extra genes required for 19B CPS biosynthesis could be closely linked and were probably located between cps19bH and cps19bK in the cps19b locus. Two PCR primers were designed, one homologous to cps19fJ (bases 5224 to 5243 of the cps19f sequence) and the other complementary to cps19fL (bases 11787 to 11811 of the cps19f sequence), to amplify this region of the 19B chromosome with the Expand long-template PCR system (Boehringer, Mannheim, Germany). The resultant PCR product was approxi-
approximately 10.5 kb, 4 kb larger than the equivalent region of \( \text{cps19f} \).

A map of the 10.5-kb PCR product was generated, using the restriction enzymes \( \text{BamHI}, \text{ClaI}, \text{HindIII}, \) and \( \text{EcoRI} \) (Fig. 2). Various restriction fragments were then cloned into \( \text{Escherichia coli} \) K-12 DH5\( \alpha \) (Bethesda Research Laboratories, Gaithersburg, Md.) with the vector pBluescript KS (Stratagene, La Jolla, Calif.), generating four recombinant plasmids with overlapping inserts, as shown in Fig. 2.

Nested deletions of the various pneumococcal DNA inserts were constructed by the method of Henikoff (8) with an Erase-a-base kit (Promega, Madison, Wis.), and the sequence of both strands was then determined with dye-labelled primers on an Applied Biosystems model 373A automated DNA sequencer. The sequence was compiled and analyzed with DNASIS and PROSIS version 7.0 software (Hitachi Software Engineering, San Bruno, Calif.). Examination of the compiled 10,549-bp \( \text{cps19b} \) sequence revealed, as expected, that the first 2.3 kb of the sequence at the 5' end has a high degree of homology to the \( \text{cps19f} \) sequence. This region contains the homologs to \( \text{cps19fG} \) and \( \text{cps19fH} \) (\( \text{cps19bG} \) and \( \text{cps19bH} \)), which exhibited 98.3 and 91.1% identity, respectively. The sequences then diverge suddenly (at base 2319 of the \( \text{cps19b} \) sequence) just prior to the end of \( \text{cps19bH} \), truncating the \( \text{cps19bH} \) gene product by two amino acids in comparison to \( \text{Cps19fH} \), as shown in Fig. 3A. There are five new potential open reading frames (ORFs), between \( \text{cps19bH} \) and \( \text{cps19bK} \), which we have designated \( \text{cps19bP}, \text{-I}, \text{-Q}, \text{-R}, \) and \( \text{-J} \), as shown in Fig. 2. Each ORF is preceded by a ribosome binding site, and the majority are very closely linked. The only potentially significant intergenic gap, of 204 nucleotides, occurs between \( \text{cps19bQ} \) and \( \text{cps19bR} \). However, no potential stem-loop structures or obvious promoter sequences were found in this region. As predicted, the 3' end of the \( \text{cps19b} \) sequence again shows homology to the \( \text{cps19f} \) sequence, starting from base 9233 (Fig. 3B); this is immediately before the start of the \( \text{cps19bK} \) gene, which has 93% identity to \( \text{cps19fK} \).

Characterization of the \( \text{cps19b} \) genes. The locations and several properties of each of the type 19B-specific ORFs, \( \text{cps19bP}, \text{-I}, \text{-Q}, \text{-R}, \) and \( \text{-J} \), are summarized in Table 1. Significant similarities with other known proteins, revealed by using the program BLASTX (1), are described below.

The \( \text{cps19bF} \) gene encodes a putative 43.3-kDa protein. Database searches with \( \text{Cps19bF} \) did not reveal significant homology to any other proteins.

The \( \text{cps19bI} \) gene encodes a putative 48.7-kDa protein. Database searches with \( \text{Cps19bI} \) also failed to identify significant homology to any other proteins. However, \( \text{Cps19bI} \) is a very hydrophobic protein, and the hydropathy plot (11, 14) exhibited marked similarity to \( \text{Cps19fI} \). Although \( \text{Cps19fI} \) is slightly larger (51.7 kDa) (21), and the hydropathy plots are not actually superimposable, both proteins have at least 12 hydrophobic, potentially membrane-spanning domains (result not presented). This is a typical hydropathy profile for Rfc-like proteins, and it suggests that \( \text{Cps19bI} \), like \( \text{Cps19fI} \), may be a polysaccharide polymerase.

The \( \text{cps19bQ} \) gene encodes a putative 34.9-kDa protein, which has 20 to 24% homology with rhamnosyl transferases from \( \text{Shigella dysenteriae} \) (13), \( \text{Leptospira interrogans} \) (20), and \( \text{Shigella flexneri} \) (18) and a 6-deoxyaltrosyl transferase from \( \text{Yersinia enterocolitica} \) (28). The alignment of \( \text{Cps19bQ} \) with the other proteins (constructed with the program CLUSTAL [10])

![Fig. 1. Biological repeat units of pneumococcal type 19F and type 19B CPS. D-Glu\( \beta \), glucose; D-Man\( \beta \)NAc, N-acetyl mannosamine; D-Rha\( \beta \), rhamnose; D-Rib\( \beta \), ribose; PO\( 4^- \), phosphate.](http://jb.asm.org/)

![Fig. 2. Physical map of part of the \( \text{S. pneumoniae} \) type 19B capsule locus (\( \text{cps19b} \)). Black arrows represent ORFs which are also present in \( \text{cps19f} \), and hatched arrows represent type 19B-specific ORFs. Gene designations are indicated below the map; \( \text{cps19bF} \) to \( \text{-L} \) are abbreviated \( \text{F} \) to \( \text{L} \), respectively. Restriction sites: B, \( \text{BamHI} \); C, \( \text{ClaI} \); E, \( \text{EcoRI} \); H, \( \text{HindIII} \). The regions of DNA subcloned into various recombinant plasmids are shown below the map.](http://jb.asm.org/)
shows several regions of homology, including a motif previously identified in rhamnosyl and other 6-deoxyhexosyl transferases (21, 22), as shown in Fig. 4. This motif contains conserved aspartate residues and is reminiscent of the catalytic sites identified in RfbAO:54 of *Salmonella enterica* serovar Borreze (12). A second motif was also identified (Fig. 4, shaded amino acids), suggesting that these proteins form a closely related subgroup of this type of transferase. Cps19bQ is the second putative rhamnosyl transferase in the *cps19b* locus. Cps19bH, which has 91.1% identity with Cps19fH, is also a rhamnosyl transferase, and it adds rhamnose to the polysaccharide backbone (21). Thus, Cps19bQ is predicted to add rhamnose (the first sugar of the disaccharide side chain [Fig. 1]) to the distal N-acetylmannosamine (ManNAc).

The *cps19bR* gene encodes a putative 76.3-kDa protein. Cps19bR has homology (21.4% identity, 40.3% similarity) to the central portion (amino acids 366 to 1071) of hypothetical protein 3 from the capsule locus of *Haemophilus influenzae* Type b (27). Although the function of hypothetical protein 3 is not known, it is located in the serotype-specific region II of the capsule locus. The only sugar in the *H. influenzae* Type b capsule that is also found in the pneumococcal type 19B CPS is ribose. Thus, it is possible that Cps19bR and protein 3 function in the synthesis of an activated ribose precursor. How-

### Table 1. Summary of ORFs *cps19bP* to *J*

<table>
<thead>
<tr>
<th>ORF</th>
<th>Location in sequence</th>
<th>Predicted mol wt</th>
<th>No. of amino acids</th>
<th>Hydrophobicity index&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Predicted pl</th>
<th>mol% G+C content&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>cps19bP</em></td>
<td>2350–3432</td>
<td>43,334</td>
<td>361</td>
<td>−0.41</td>
<td>8.69</td>
<td>29.5</td>
</tr>
<tr>
<td><em>cps19bI</em></td>
<td>3451–4695</td>
<td>48,667</td>
<td>414</td>
<td>0.77</td>
<td>9.73</td>
<td>27.2</td>
</tr>
<tr>
<td><em>cps19bQ</em></td>
<td>4703–5605</td>
<td>34,876</td>
<td>300</td>
<td>−0.26</td>
<td>8.20</td>
<td>29.7</td>
</tr>
<tr>
<td><em>cps19bR</em></td>
<td>5809–7746</td>
<td>76,348</td>
<td>645</td>
<td>−0.30</td>
<td>8.65</td>
<td>27.2</td>
</tr>
<tr>
<td><em>cps19bJ</em></td>
<td>7736–9181</td>
<td>53,851</td>
<td>481</td>
<td>0.91</td>
<td>9.89</td>
<td>29.5</td>
</tr>
</tbody>
</table>

<sup>a</sup> According to Kyte and Doolittle (14), as implemented in PROSIS.

<sup>b</sup> Percent G+C of coding region.
ever, the nature of the activated ribose precursor that is used in CPS biosynthesis is not known (27).

The cps19bJ gene encodes a putative 53.9-kDa protein with low-level homology to RfbX proteins from *E. coli*, *S. dysenteriae*, and *Y. enterocolitica*, to the CapF protein of *Staphylococcus aureus*, and to Cps19fJ, as shown in Table 2. The RfbX proteins are known to be involved in the export of O-antigen repeat units (17, 19). The hydropathy plots for RfbX-like proteins are all very similar, with 10 to 12 hydrophobic, membrane-spanning domains, and those for Cps19bJ and Cps19fJ are almost superimposable (result not presented). Thus, Cps19bJ is likely to be the polysaccharide repeat unit transporter.

Specificity of the cps19b genes. To examine the relationship between cps19b and encapsulation loci of other *S. pneumoniae* serotypes, we labelled individual cps19b genes (bases 2305 to 3188 for cps19bP, 3412 to 4545 for cps19bI, 4540 to 5718 for cps19bQ, 5713 to 7766 for cps19bR, and 7615 to 9046 for cps19bJ) with digoxigenin. These were used to probe (at high stringency) Southern blots of restricted chromosomal DNA (extracted and purified as previously described [23]) from representative pneumococci belonging to the following types/groups: 2, 3, 4, 6, 7F, 7B, 8, 9, 12, 14, 16, 17, 18, 19F, 19A, 19C, 22, 23, and 24. None of the probes hybridized to DNA from any serotype tested, except to the closely related type 19C, which has homologs for all five genes.

### Table 2. Similarity of Cps19bJ to other proteins

<table>
<thead>
<tr>
<th>Protein</th>
<th>% Identity*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cps19bJ</td>
<td>100</td>
</tr>
<tr>
<td>K12RfbX</td>
<td>23.3 [404]</td>
</tr>
<tr>
<td>YeTrsA</td>
<td>21.6 [402]</td>
</tr>
<tr>
<td>SdRfbX</td>
<td>22.3 [394]</td>
</tr>
<tr>
<td>SaCapF</td>
<td>21 [395]</td>
</tr>
<tr>
<td>YeRfbX</td>
<td>19 [420]</td>
</tr>
<tr>
<td>Cps19bJ</td>
<td>18.2 [406]</td>
</tr>
</tbody>
</table>

*Percentage of identical amino acids determined using PROSIS. Numbers in brackets indicate the number of amino acids over which the % identity occurs.

---

* S. pneumoniae Cps19bJ.
* E. coli K-12 RfbX (25).
* Y. enterocolitica TrsA (24).
* S. dysenteriae RfbX (13).
* S. aureus CapF (16).
* Y. enterocolitica RfbX (Genbank/EMBL accession no. U46859).
* S. pneumoniae Cps19fJ (21).
Capsule type switching by transformation. To determine if the genes sequenced were sufficient for type 19B CPS biosynthesis, we transformed the 10.5-kb PCR product (as described previously [4]) into an unencapsulated, erythromycin-resistant derivative of Rtx-19F, designated Rtx-19F-I, in which the cps19fI gene had been disrupted by insertion-duplication mutagenesis with pVAA891 (21). Several smooth transformants were checked for erythromycin sensitivity, which indicates loss of the pVAA891 sequence. Southern hybridization was used to confirm the absence of both pVAA891 and the cps19fI gene and the presence of each of the cps19fJ-, I-, Q-, R-, and J genes. The production of a type 19B capsule by one of these smooth transformants, designated Rtx-19B, was then confirmed by quellung reaction. This shows that it is possible to alter capsule production from type 19F to type 19B by replacing part of the capsule locus and that the region of cps19f described in this study determines the 19B serotype.

Conclusions. The chemical structure of the S. pneumoniae type 19B capsule is considerably more complex than that of the type 19F capsule locus and that the region of the cps19fJ gene cluster has been demonstrated in vitro by the transformation of Rtx-19F-I with the 10.5-kb cps19f PCR product to produce a transformant, Rtx-19B, which expresses the type 19B capsule. Also of interest is the low G+C content of this type 19B-specific gene cluster, which ranges from 27.2 to 29.7 mol% for the individual genes (Table 1). This is comparable to the G+C content of 29.7 mol% for the cps19fJ-cps19fI gene cluster but differs considerably from that of the remainder of the locus, which ranges from 30.3 to 42.3 mol% in the cps19fJ locus (7, 21). This is consistent with the acquisition of this gene cluster from a source distinct from that of other parts of the capsule locus.

Nucleotide sequence accession number. The nucleotide sequence described in this paper has been deposited with GenBank under accession no. AF004325.

We are grateful to Mike Gratten for factor typing of pneumococcal transformants and to James C. Richards for helpful discussions. This work was supported by a grant from the National Health and Medical Research Council of Australia.

REFERENCES

17. Liu, D., R. A. Cole, and P. R. Reeves. 1996. An O-antigen processing function...