INFECTION AND IMMUNITY, Dec. 1994, p. 5384–5396
0019-9567/94/$04.00+0

Vol. 62, No. 12

Nucleotide Sequence Analysis of Genes Essential for Capsular Polysaccharide Biosynthesis in *Streptococcus pneumoniae* Type 19F

ANGELO GUIDOLIN,[1] JUDY K. MORONA,[1] RENATO MORONA,[2] DAVID HANSMAN,[1] AND JAMES C. PATON[1]*

Department of Microbiology, 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 15 June 1994/Returned for modification 15 August 1994/Accepted 7 September 1994

Previous studies have shown that the capsular polysaccharide synthesis (*cps*) locus of the type 19F *Streptococcus pneumoniae* strain SSZ was closely linked to a copy of the insertion sequence IS*1202* (J. K. Morona, A. Guidolin, R. Morona, D. Hansman, and J. C. Paton, J. Bacteriol. 176:4437–4443, 1994). In the present study, we used plasmid insertion and rescue and inverse PCR to clone 6,322 bp of flanking DNA upstream of IS*1202*. Sequence analysis indicated that this region contains six complete open reading frames (ORFs) and one partial ORF that are arranged as a single transcriptional unit. Chromosomal disruption of any of these ORFs in a smooth-type 19F strain leads to a rough (unencapsulated) phenotype, indicating that this operon is essential for capsule production. The ORFs have therefore been designated *cps19fA* to *cps19fG*, where *cps19fA* is the first gene of the type 19F *cps* locus. Furthermore, many of the gene products from this incomplete operon exhibit strong similarities to proteins known to be involved in the production of capsular polysaccharide, exopolysaccharide, teichoic acid, enterobacterial common antigen, and lipopolysaccharide from numerous other bacterial species. This has allowed us to propose functions for many of the type 19F *cps* gene products. Southern hybridization studies reveal that *cps19fA* and *cps19fB* are conserved among all 12 pneumococcal serotypes tested, whereas genes downstream of *cps19fB* are conserved among some, but not all, of the serotypes tested.

Streptococcus pneumoniae (the pneumococcus) is an important human pathogen, causing invasive diseases such as pneumonia, meningitis, and bacteremia. Morbidity and mortality from pneumococcal infections remain high, even in regions where effective antibiotic therapy is freely available. In developing countries, in excess of 3 million children under the age of 5 years die each year from pneumonia, and *S. pneumoniae* is the commonest causative agent. *S. pneumoniae* also causes less serious, but highly prevalent infections such as otitis media and sinusitis, which have a significant impact on health-care costs in developed countries.

Pneumococci are frequently isolated from the nasopharynx of healthy people. Virtually all humans are colonized by pneumococci at some stage, and pneumococcal carriage rates are higher in young children and where people are living in crowded conditions. In a proportion of carriers (approximately 15% in one pediatric study carried out in the United States [16]), the pneumococcus invades tissues and causes disease. This scenario is more likely in persons who have only recently become colonized (2).

An important feature of *S. pneumoniae* is its capacity to produce a polysaccharide capsule, which is structurally distinct for each of the 84 known serotypes of the organism. The polysaccharide capsule is considered to be the sine qua non of pneumococcal virulence (2). This is based on the observation that all fresh isolates from patients with pneumococcal infection are encapsulated, and spontaneous nonencapsulated (rough) derivatives of such strains are almost completely avirulent. Moreover, an early study demonstrated that enzymatic depolymerization of the capsular polysaccharide (CPS) of a type 3 pneumococcus increased its 50% lethal dose approximately 10^6-fold (4). More recently, a similar effect on virulence of type 3 *S. pneumoniae* was achieved by transposon mutagenesis of a gene essential for CPS production (45). The precise manner in which the pneumococcal capsule contributes to virulence is not fully understood, although it is known to have strong antiphagocytic properties in nonimmune hosts (26, 34). Clearly, some CPS serotypes are more effective than others, as certain types are far more commonly associated with human disease. Moreover, within a given serotype, virulence appears to be related to the amount of CPS produced (2), which also implies that the level of CPS expression can be regulated. In the immune host, however, binding of specific antibody to the CPS results in opsonization and rapid clearance of the invading pneumococci. For this reason, vaccines based on purified CPS have been developed. However, protection is serotype specific and the present formulation contains CPS from the 23 commonest types. This vaccine is highly protective in healthy adults against invasive infections caused by those serotypes included in the vaccine, but protection is very poor in high-risk groups such as young children, who mount a poor immune response to several important CPS types (11).

Although the chemical structures of a number of CPS types have been determined (26), very little is known about the genes encoding CPS biosynthesis and expression in *S. pneumoniae*. This process would require a complex pathway including transport into the cell and/or synthesis of the component monosaccharides, activation of each to a nucleotide precursor, coordinated transfer of each sugar, in sequence, to the repeat-

* Corresponding author. Mailing address: Department of Microbiology, Women's and Children's Hospital, North Adelaide, S.A. 5006, Australia. Phone: 61-8-204 6302. Fax: 61-8-204 6051.

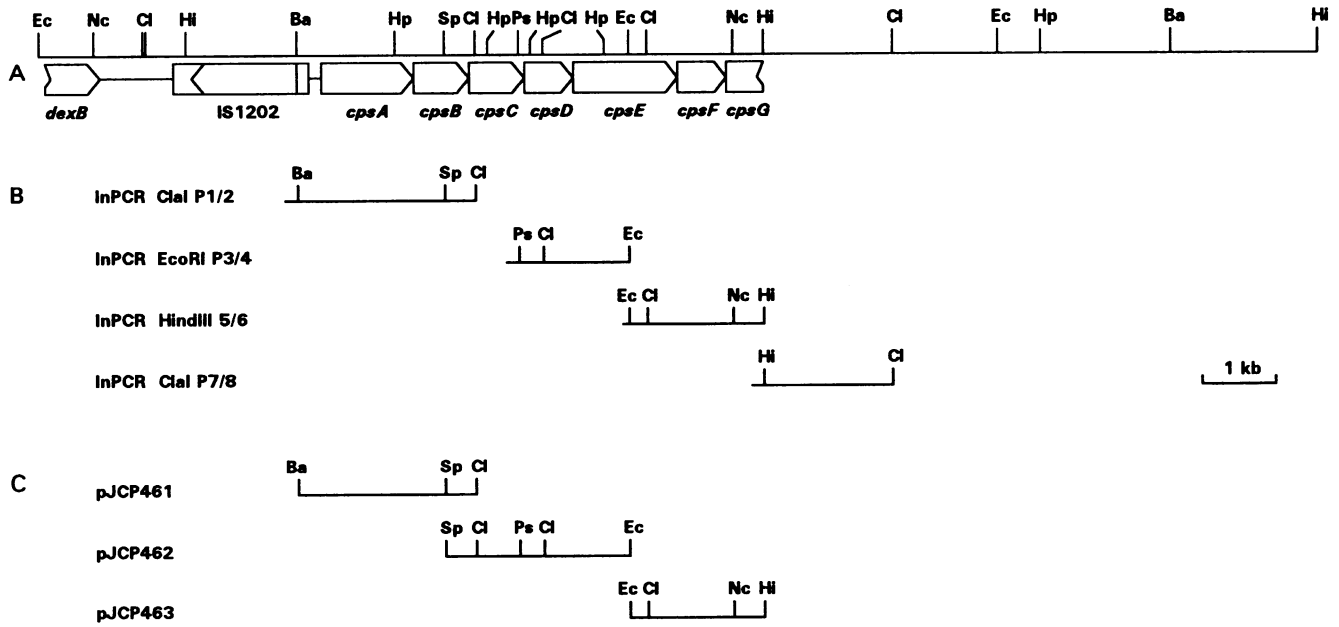

FIG. 1. (A) Physical map of the chromosome of *S. pneumoniae* Rx1-19F in the vicinity of the 19F *cps* locus. Arrows represent potential ORFs, and the box represents the insertion element IS*1202*. Gene designations are indicated below the map; genes *cps19fA-G* are abbreviated to *cpsA-G*, respectively. Restriction sites are as follows: Ba, *Bam*HI; Cl, *Cla*I; Ec, *Eco*RI; Hi, *Hin*dIII; Hp, *Hpa*I; Nc, *Nco*I; Ps, *Pst*I; Sp, *Sph*I. (B) Segments of DNA amplified by InPCR specifying the restriction enzyme and primer pair used. (C) Regions of DNA subcloned into pJCP461, pJCP462, and pJCP463 by excision and rescue of pVA891 and flanking sequences from various insertion-duplication mutants.

ing oligosaccharide, and subsequent polymerization, export, and attachment to the cell surface.

Classical genetic studies carried out by Austrian et al. (3) demonstrated that all of the *S. pneumoniae* genes required for biosynthesis and expression of CPS are closely linked on the pneumococcal chromosome. Indeed, clustering of CPS genes is a common feature of all bacterial polysaccharide loci studied so far. Thus, it has been shown that the CPS genes of *Escherichia coli* K1, K5, K7, and K-12, *Haemophilus influenzae*, and *Neisseria meningitidis* are all clustered on segments of DNA 10 to 20 kb in length (12, 23, 39). Within these gram-negative loci there appears to be a considerable degree of sequence homology and a conserved genetic organization. Each locus consists of a central region encoding enzymes for the biosynthesis of the CPS itself, flanked by regions encoding proteins involved in translocation of the CPS across the cytoplasmic membrane (proteins similar to members of the ABC superfamily of ATP-dependent transport proteins) and transport through the periplasm onto the cell surface (outer membrane proteins with porin properties) (12, 23, 39). Also, regions with putative regulatory functions have been identified in the meningococcal *cps* locus (12). In contrast, less is known of the structure and organization of gram-positive *cps* loci, and for pneumococci, only part of a single CPS-related gene (encoding a putative UDP-glucose dehydrogenase from serotype 3) has been sequenced (13).

Characterization of the *cps* loci would undoubtedly improve our understanding of the mechanisms of CPS production and regulation in pneumococci, which would in turn contribute to an understanding of the pathogenesis of pneumococcal disease. To date, our studies have concentrated on *S. pneumoniae* type 19F, because it is one of the commonest causes of invasive disease in children. Moreover, type 19F CPS, a linear polysaccharide with a repeat unit consisting of →4)-β-D-Man*p*NAc-(1→4)-α-D-Glc*p*-(1→2)-α-L-Rha*p*-(1-PO_4^-→ (26), is one of the poorest immunogens in this group (11). We report here the complete nucleotide sequence of the first six genes of an operon essential for type 19F capsule production. We have been able to assign functions to several of the open reading frames (ORFs) on the basis of similarities with a variety of proteins which function in polysaccharide synthesis in a range of bacterial species. Furthermore, we demonstrate that some of these genes are highly conserved among different pneumococcal serotypes, while others are serotype specific.

(A preliminary report of this work was presented at the American Society for Microbiology Fourth International Conference on Streptococcal Genetics, Santa Fe, N.M., May 1994 [17a].)

MATERIALS AND METHODS

Bacterial strains and plasmids. The *S. pneumoniae* strains used were Rx1, a nonencapsulated, highly transformable derivative of type 2 strain D39 (41), and SSZ, a type 19F strain obtained from Chi-Jen Lee, Center for Biologics, Food and Drug Administration, Bethesda, Md. A derivative of Rx1 expressing type 19F capsule (designated Rx1-19F) was constructed by transformation of Rx1 with DNA from strain SSZ, as described elsewhere (32). Rx1-19F-B1 is a derivative of Rx1-19F in which the major ORF in the single copy of IS*1202* (which is closely linked to the *cps* locus) has been interrupted by insertion-duplication mutagenesis, using pVA891 (32). Clinical isolates belonging to types 19A, 19B, and 19C were also obtained from Chi-Jen Lee; other clinical isolates were from the Women's and Children's Hospital, Adelaide, South Australia. Pneumococci were routinely grown in Todd-Hewitt broth with 0.5% yeast extract or on blood agar. When appropriate, erythromycin was added to media at a concentration of 0.2 μg/ml.

E. coli K-12 DH1 (18) and DH5α (Bethesda Research

*Bam*HI -35

1 GGATCCTTTTAATATGACTATTCTACCAAATGGGACATTTTCACGTTCGATTTACTAAAGACATTATCACATTCGAATTACATTAAAGATGCAGATAGTAAAAAAAATGTAGACATTACC

cps19fA→

-10 M S R R F K K S R S Q K V K R S V N I V L L T I

121 GTAAAAAAGTGATATAATCGTATGATGTTCAAGGTATAGGTGTTAATCATGAGTAGACGTTTTAAAAAATCACGTTCACAGAAAGTGAAGCGAAGTGTTAATATCGTTTTGCTGACTATT

Y L L L V C F L L F L I F K Y N I L A F R Y L N L V I T A L V L L V A L L G L L

241 TATTTATTGTTAGTTTGTTTTTTATTGTTCTTAATCTTTAAGTACAATATCCTTGCTTTTAGATATCTTAACCTAGTGATAACTGCGTTAGTCCTACTAGTTGCCTTGCTAGGGCTACTC

L I I Y K K A E K F T I F L L V F S I L V S S V S L F A V Q Q F V G L T N R L N

361 TTGATTATCTATAAAAAAGCTGAAAAGTTTACTATTTTCCTGTTGGTGTTCTCTATCCTTGTCAGCTCTGTGTCGCTCTTTGCAGTACAGCAGTTTGTTGGACTGACCAATCGTTTAAAT

A T S N Y S E Y S I S V A V L A D S D I E N V T Q L T S V T A P T G T D N E N I

481 GCGACTTCTAATTACTCAGAATATTCAATCAGTGTCGCTGTTTTAGCAGATAGTGATATCGAAAATGTTACGCAACTGACGAGTGTGACAGCACCGACTGGGACTGATAATGAAAATATT

Q K L L A D I K S S Q N T D L T V N Q S S S Y L A V Y K S L I A G E T K A I V L

601 CAAAAACTACTAGCTGATATTAAGTCAAGTCAGAATACCGATTTGACGGTCAACCAGAGTTCGTCTTACTTGGCAGTTTACAAGAGTTTGATTGCAGGGGAGACTAAGGCCATTGTCCTA

N S V F E N I I E S E Y P D Y A S K I K K I Y T K G F T K K V E A P K T S K N Q

721 AATAGTGTCTTTGAAAACATCATCGAGTCAGAGTATCCAGACTACGCATCGAAGATAAAAAAGATTTATACCAAGGGATTCACTAAAAAAGTAGAAGCTCCTAAGACGTCTAAGAATCAG

S F N I Y V S G I D T Y G P I S S V S R S D V N I L M T V N R D T K K I L L T T

841 TCTTTCAATATCTATGTTAGTGGAATTGACACCTATGGTCCTATTAGTTCGGTGTCGCGATCAGATGTCAATATCCTGATGACTGTCAATCGAGATACCAAGAAAATCCTCTTGACCACA

T P R D A Y V P I A D G G N N Q K D K L T H A G I Y G V D S S I H T L E N L Y G

961 ACGCCACGTGATGCCTATGTACCAATCGCAGATGGTGGAAATAATCAAAAAGATAAATTAACCCATGCAGGCATTTATGGAGTTGATTCGTCCATTCACACCTTAGAAAATCTCTATGGA

V D I N Y Y V R L N F T S F L K M I D L L G G V D V H N D Q E F S A L H G K F H

1081 GTGGATATCAATTACTATGTGCGATTGAACTTCACTTCTTTCTTGAAAATGATTGACTTATTGGGAGGGGTAGATGTTCATAATGATCAAGAGTTTTCAGCTCTACATGGGAAGTTCCAT

F P V G N V H L D S E Q A L G F V R E R Y S L A D G D R D R G R N Q Q K V I V A

1201 TTCCCAGTAGGGAATGTCCATCTAGACTCTGAGCAGGCTCTAGGTTTTGTACGTGAACGCTACTCACTAGCCGATGGAGACCGTGATCGTGGTCGCAACCAACAAAAGGTCATTGTAGCA

I I Q K L T S T E V L K N Y S S I L Q G L Q D S L Q T N M P I E T M I D L V N T

1321 ATTATTCAGAAGTTAACTTCTACAGAGGTTTTGAAAAACTATAGTAGTATTCTTCAAGGATTGCAGGATTCTCTTCAAACAAATATGCCGATTGAGACTATGATAGATTTAGTGAATACT

Q L E S G G N Y K V N S Q D L K G T G R M V L P S Y A M P D S N L Y V M E I D D

1441 CAGTTGGAAAGTGGGGGGAATTATAAAGTAAATTCTCAAGATTTAAAAGGGACAGGTCGGATGGTTCTTCCTTCTTATGCAATGCCAGACAGTAACCTCTATGTGATGGAAATAGATGAT

cps19fB→

S S L A V V K A A I Q D V M E G R * M I D I H S H I V F D V D D G P K S R E E S

1561 AGTAGTTTAGCTGTAGTTAAAGCAGCTATACAGGATGTGATGGAGGGTAGATGAAATGATAGACATCCATTCGCATATCGTTTTTGATGTAGATGACGGTCCCAAGTCAAGAGAGGAAAG

K A L L A E S Y R Q G V R T I V S T S H R R K G M F E T P E E K I A E N F L Q V

1681 CAAGGCTCTCTTGGCAGAATCCTACAGGCAGGGGGTGCGAACCATTGTCTCTACCTCTCACCGTCGCAAGGGCATGTTTGAAACTCCGGAAGAGAAGATAGCAGAAAACTTTCTTCAGGT

R E I A K E V A D D L V I A Y G A E I Y Y T L D A L E K L E K K E I P T L N D S

1801 TCGGGAAATTGCAAAAGAAGTGGCAGATGATTTAGTCATTGCTTATGGCGCAGAGATATACTATACTCTGGATGCTCTAGAAAAGCTAGAAAAAAAAGAAATTCCTACCCTTAATGATAG

R Y A L I E F S M H T S Y R Q I H T G L S N I L M L G I T P V I A H I E R Y D A

1921 TCGTTATGCCTTGATTGAGTTTAGCATGCATACTTCCTATCGTCAGATTCATACGGGATTGAGCAATATTTTGATGTTGGGAATCACGCCAGTAATTGCTCATATTGAACGTTATGATGC

L E N N E K R V R E L I D M G C Y T Q I N S Y H V S K P K F F G E K Y K F M K K

2041 TTTGGAGAATAACGAAAAACGTGTTCGTGAACTGATTGATATGGGGTGCTATACTCAGATAAATAGTTATCATGTTTCAAAACCTAAGTTCTTTGGTGAAAAATATAAATTCATGAAAAA

R A R Y F L E R D L V H V V A S D M H N L D S R P P Y M Q Q A Y D I I A K K Y G

2161 GAGAGCTCGGTATTTTTTGGAACGTGATTTAGTTCATGTAGTTGCAAGTGACATGCACAATTTAGACAGTAGACCTCCATATATGCAACAGGCATATGATATCATTGCTAAGAAATATGG

cps19fC→

A K K A K E L F V D N P R K I I M D Q L I * M K E Q N T L E I D V L Q L F

2281 AGCGAAAAAAGCAAAAGAACTTTTTGTAGATAATCCCAGGAAAATTATAATGGATCAATTAATTTAGGAGAAAATATGAAGGAACAAAACACTTTGGAAATCGATGTATTGCAACTATTC

R A L W K R K L V I L L V A I I T S S V A F A Y S T F V I K P E F T S M T R I Y

2401 AGAGCTTTATGGAAAAGAAAGTTGGTCATTTTATTAGTGGCAATTATAACTTCTTCAGTTGCTTTTGCCTACAGTACTTTTGTTATCAAACCTGAGTTTACTAGTATGACTCGGATTTAT

V V N R D Q G E K S G L T N Q D L Q A G S S L V K D Y R E I I L S Q D V L E E V

2521 GTAGTTAACCGTGATCAGGGAGAGAAGTCTGGTTTAACCAATCAAGACTTGCAGGCAGGATCATCCTTGGTTAAAGACTATCGTGAAATTATCCTATCGCAGGATGTTTTGGAGGAAGTT

V S D L K L D L T P K D L A N K I K V T V P V D T R I V S V S V S D R V P E E A

2641 GTTTCTGATTTGAAACTAGATTTGACGCCAAAAGATTTGGCTAATAAAATTAAAGTAACAGTACCAGTTGATACCCGTATTGTCTCTGTTTCAGTTAGTGATCGAGTTCCTGAAGAGGCA

S R I A N S L R E V A A Q K I I S I T R V S D V T T L E E A R P A T S P S S P N

2761 AGCCGTATCGCTAACTCTTTGAGAGAAGTAGCTGCTCAAAAAATTATCAGTATTACTCGTGTTTCTGATGTGACAACACTGGAGGAGGCAAGACCGGCGACATCACCGTCTTCGCCAAAT

I K R S T L I G F L A G V I G T S V I V L I L E L L D T R V K R P K D I E D T L

2881 ATTAAACGCAGTACACTAATTGGTTTTTTGGCAGGAGTGATTGGAACTAGTGTTATAGTTCTTATTCTTGAACTTTTGGACACTCGTGTGAAACGTCCGAAAGATATCGAAGATACACTG

cps19fD→

Q M T L L G I V P N L N K L K * M P T L E I A Q K K L E F I K K A E E Y Y

3001 CAGATGACACTTTTGGGAATTGTACCAAACTTGAATAAGTTGAAATAGGAGAGAGGAATGCCGACATTAGAAATAGCACAAAAAAAACTGGAGTTCATTAAGAAGGCAGAAGAATATTAC

N A L C T N I Q L S G D K L K V I S V T S V N P G E G K T T T S V N I A R S F A

3121 AATGCCTTGTGTACAAATATACAGTTGAGCGGAGATAAACTAAAAGTAATTTCCGTTACTTCTGTTAACCCTGGGGAAGGAAAAACAACTACTTCCGTAAATATAGCAAGGTCGTTTGCG

FIG. 2. Nucleotide and amino acid sequences of the 19F CPS locus. The nucleotide sequence from the *Bam*HI site upstream of *cps19fA* to the *Hin*dIII site within *cps19fG* is shown. The amino acid translation for each ORF is represented by a single-letter code above the first nucleotide of each codon. Possible ribosome binding sites are underlined, and putative −10 and −35 promoter regions upstream from *cps19fA* are doubly underlined.

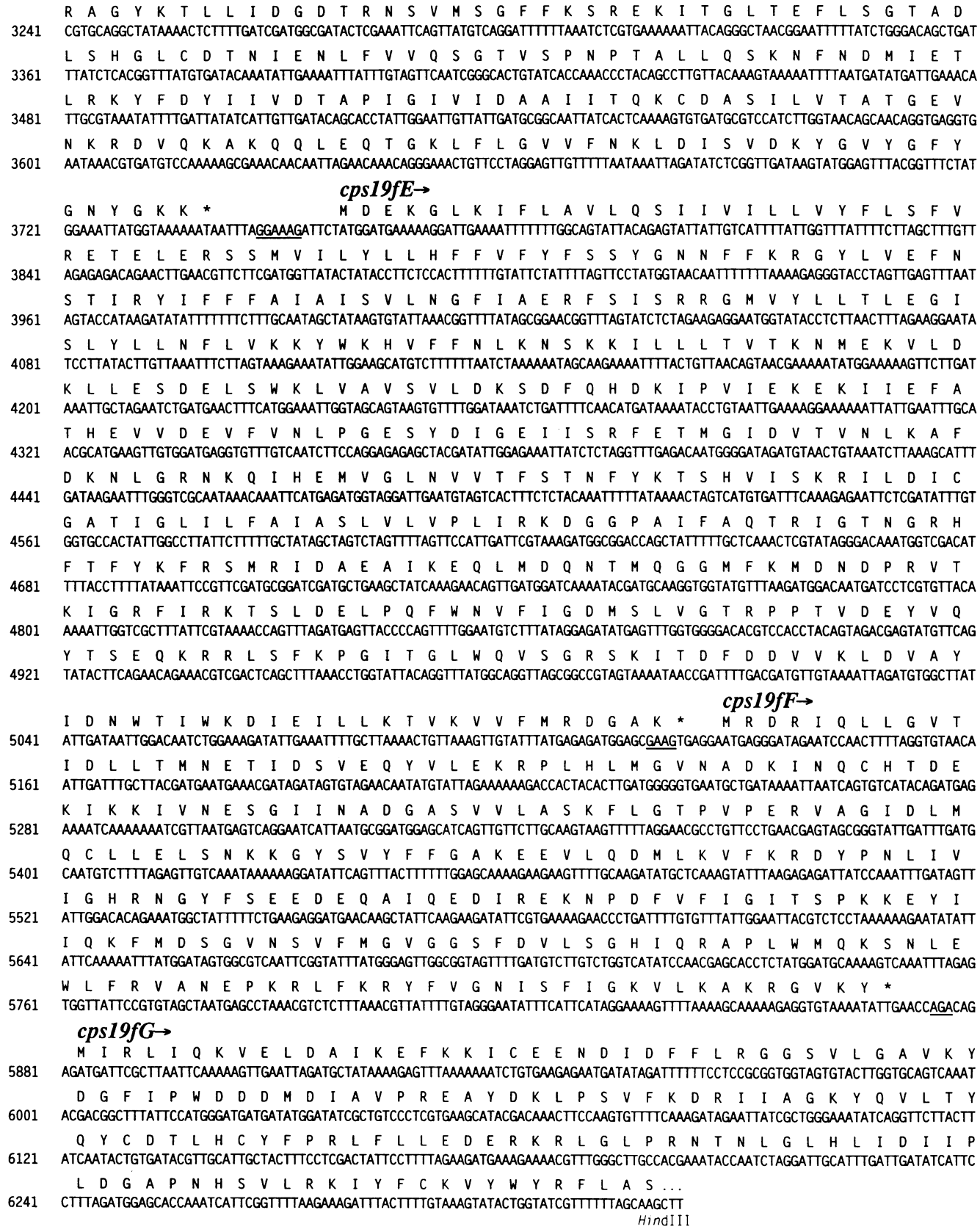

FIG. 2—*Continued.*

TABLE 1. Summary of *S. pneumoniae cps19f* ORFs

ORF	Nucleotide position in sequence	Amino acids (no.)	Predicted molecular size	Predicted pI	Mean hydropathy index[a]	% G+C[b]
Cps19fA	169–1615	481	53,572	8.86	0.04	38.1
Cps19fB	1616–2347	243	28,352	7.58	−0.46	38.0
Cps19fC	2356–3048	230	25,497	9.34	0.11	38.2
Cps19fD	3058–3741	227	24,947	8.69	−0.13	34.5
Cps19fE	3751–5124	455	52,595	9.51	0.14	33.2
Cps19fF	5128–5871	247	28,155	8.93	−0.19	33.6
Cps19fG[c]	5883–end	146	17,183	7.08	−0.12	35.7

[a] According to Kyte and Doolittle (24), as implemented in PROSIS.
[b] Percent guanine plus cytosine (G+C) of coding region.
[c] ORF is truncated.

Laboratories, Gaithersburg, Md.) were grown in Luria-Bertani broth (29) with or without 1.5% Bacto Agar (Difco Laboratories, Detroit, Mich.). When appropriate, chloramphenicol, ampicillin, or erythromycin was added to the growth medium at a concentration of 25, 50, or 10 μg/ml, respectively.

Bacterial transformation. Transformation of *E. coli* with plasmid DNA was carried out with $CaCl_2$-treated cells as described by Brown et al. (7). The encapsulated strain Rx1-19F was transformed, with chromosomal or plasmid DNA, as described previously for D39 (5, 47). Transformants were selected on blood agar containing 0.2 μg of erythromycin per ml.

Assessment of encapsulation. Production of capsule by pneumococci was assessed by quellung reaction, using monospecific antisera obtained from Statens Seruminstitut, Copenhagen, Denmark.

DNA manipulations. *S. pneumoniae* chromosomal DNA used in Southern hybridization experiments was extracted and purified as previously described (36). Plasmid DNA was isolated from *E. coli* by the alkaline lysis method (31). Analysis of recombinant plasmids was carried out by digestion of DNA with one or more restriction enzymes under the conditions recommended by the supplier. Restricted DNA was electrophoresed in 0.8 to 1.5% agarose gels with a Tris-borate-EDTA buffer system, as described by Maniatis et al. (29).

InPCR. The method used for inverse PCR (InPCR) was that described by Ochman et al. (35). Briefly, chromosomal DNA (5 μg) was digested to completion with an appropriate restriction enzyme. The cleaved DNA was then self-religated at 16°C for 16 h at a concentration of 10 μg/ml. An amplification reaction (35 cycles, 50°C annealing temperature) was performed with 1 μg of ligated DNA and 50 pmol of each appropriate primer. Each InPCR product was cloned, and the ends were sequenced to permit design of primers for further rounds of InPCR.

Southern hybridization analysis. Chromosomal DNA (2.5 μg) was digested with appropriate restriction enzymes, and the digests were electrophoresed on agarose gels in Tris-borate-EDTA buffer. DNA was then transferred to a positively charged nylon membrane (Hybond N^+; Amersham International) as described by Southern (42), hybridized to digoxigenin-labelled probe DNA, washed, and then developed using anti-digoxigenin-alkaline phosphatase conjugate (Boehringer, Mannheim, Germany) and 4-nitroblue tetrazolium–X-phosphate substrate according to the manufacturer's instructions. Digoxigenin-labelled lambda DNA, restricted with *Hin*dIII, was used as a DNA molecular size marker.

Plasmid insertion and rescue. The first stage of plasmid insertion and rescue involves cloning an appropriate fragment of pneumococcal DNA into pVA891 (which encodes chloramphenicol and erythromycin resistance but cannot replicate in *S. pneumoniae* [27]) and using this construct to transform Rx1-19F. Recombination between the pneumococcal DNA insert and the homologous region of the pneumococcal chromosome results in integration of the plasmid sequences. Plasmid sequences along with flanking DNA were then recovered by digestion of chromosomal DNA (5 μg) with an appropriate restriction enzyme and self-religation of the cleaved DNA (at a concentration of 10 μg/ml), followed by transformation into *E. coli* DH5α, selecting for either chloramphenicol or erythromycin resistance. Pneumococcal DNA was subsequently subcloned into pGEM7Zf(+) (Promega Corp., Madison, Wis.) for further analysis.

DNA sequencing and analysis. Nested deletions of pneumococcal DNA cloned into pGEM7Zf(+) were constructed by the method of Henikoff (19), using an Erase-a-base kit (Promega). This DNA was transformed into *E. coli* DH5α, and the resulting plasmid DNA was characterized by restriction analysis. Double-stranded template DNA for sequencing was prepared as recommended in the Applied Biosystems sequencing

```
Cps19fA TVNQSSSYLAVYKSLIAGETKAIVL----NSVFENIIESEYPDYASKIKKIYTKGFTKKV
BsLytR  MRNERRK---KKKTLLLTILTIIGLLVLGTGGYAYYLWHKAASTVASIHESIDKS-KKRD

Cps19fA EAPKTSKNQSFNIYVSGIDTYGPISSVSRSDVNILMTVNRDTKKILLTTTPRDAYVPIAD
BsLytR  KEVSINKKDPFSVLIMGVD--ERDGDKGRADTLIYMTVNPKTNTTDMVSIPRDTYTKII-

Cps19fA GGNNQKDKLTHAGIYG-VDSSIHTLENLYGVDINYYVRLNFTSFLKMIDLLGGVDVHNDQ
BsLytR  -GKGTMDKINHSYAFGGTQMTVDTVENFLDVPVDYFVKVNMESFRDVVDTLGGITVNSTF

Cps19fA EFSALHGKFHFPVGNVHLDSEQALGFVRERYSLADGDRDRGRNQQKVIVAIIQKLTSTEV
BsLytR  AFS--YDGYSFGKGEITLNGKEALAYTRMRKEDPRGDFGRQDRQRQVIQGIINKGANISS

Cps19fA LKNYSSILQGLQDSLQTNMPIETMIDLVNTQLESGGNYKVNSQDLKGTGRMVLPSYAMPD
BsLytR  ITKFGDMFKVVENNVRTNLTFDNMWDIQSDY--KGARKHIKQHELKGTGTKI--------

Cps19fA SNLYVMEIDDSSLAVVKAAIQDVMEGR
BsLytR  NGIYYYQADESALSDITKELKESLE-K
```

FIG. 3. Alignment of Cps19fA, from amino acid position 160 to the C-terminal end, with the *B. subtilis* LytR protein (BsLytR) (25), as determined using the default settings of the program CLUSTAL (20). Identical residues are boxed; similar residues are shown in boldface. –, absence of a residue.

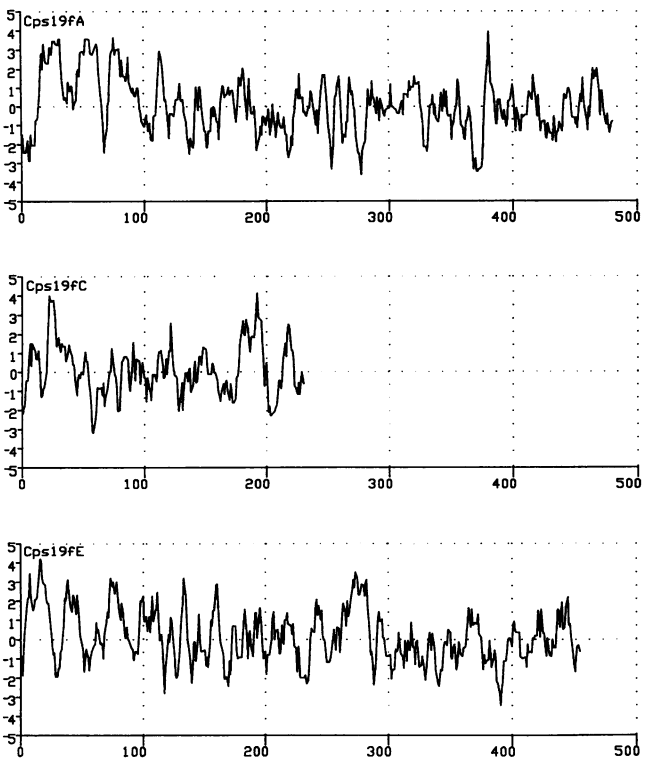

FIG. 4. Hydropathy plots of Cps19fA, Cps19fC, and Cps19fE, generated by the method of Kyte and Doolittle (24). Numbers on the *x* axis indicate amino acid position. Positive numbers on the *y* axis indicate hydrophobic regions.

manual. The sequences of both strands were then determined using dye-labelled primers on an Applied Biosystems model 373A automated DNA sequencer. The sequence was analyzed by using DNASIS and PROSIS Version 7.0 software (Hitachi Software Engineering, San Bruno, Calif.). The program BLASTX (1) was used to translate DNA sequences and conduct homology searches of the protein databases available at the National Center for Biotechnology Information, Bethesda, Md. Amino acid sequence alignments were performed with the program CLUSTAL (20).

Nucleotide sequence accession number. The nucleotide sequence described in this paper has been deposited with GenBank under accession number U09239.

RESULTS AND DISCUSSION

Genomic mapping of the type 19F *cps* locus. We have previously shown by cotransformation studies that type 19F capsule production was closely linked to a copy of IS*1202* in *S. pneumoniae* Rx1-19F. This copy of IS*1202* appears to have inserted in an intergenic region of the chromosome (32), bordered at one end by the putative pneumococcal *dexB* gene, as shown in Fig. 1A. This finding, coupled with the very close linkage observed between CPS production and IS*1202*, suggested that the *cps* locus is located upstream of the insertion sequence.

Numerous attempts to isolate low-copy-number cosmid clones from this region proved unsuccessful, indicating that this region of the pneumococcal chromosome is very unstable in *E. coli*. Therefore, as a preliminary step in isolating this region, we generated a physical map that would enable us to choose appropriate enzymes for plasmid rescue of pneumococcal DNA (see below) and to verify the identity of any subsequent DNA fragments obtained in this manner. Figure 1B shows the restriction enzymes and primer pairs used to obtain the four indicated InPCR products. Sequential rounds of InPCR and genomic mapping by Southern hybridization, using the InPCR products as probes (data not shown), generated the physical map spanning approximately 17 kb of the Rx1-19F chromosome, as shown in Fig. 1A.

Isolation of CPS genes. Since large chromosomal fragments from this portion of the pneumococcal genome appear to be unstable in *E. coli*, we developed an alternative strategy to obtain DNA for sequence analysis. This involved rescue by transformation into *E. coli* of the pVA891 replicon, from insertion-duplication mutants of *S. pneumoniae* (see below), such that it transferred with it small regions of flanking chromosomal DNA (2 to 3 kb). This approach permitted the isolation of the three consecutive portions of the Rx1-19F chromosome, covering a region of 6.3 kb, shown in Fig. 1C. The first clone was obtained by rescue of pVA891 from strain Rx1-19F-B1, in which plasmid pVA891 had been inserted into the ORF of IS*1202*, as previously described (32). These segments of pneumococcal DNA were subsequently subcloned into pGEM7Zf(+) to give plasmids pJCP461, pJCP462 and pJCP463, from which nested deletions were generated to obtain the complete sequence of 6,322 bp shown in Fig. 2. The

```
Cps19fB MIDIHSHIVFDVDDGPKSREESKALLAESYRQGVRTIVSTSHRRKGMFETPEEKIAENFL
SaCpsA  MIDIHSHIVFDVDDGPKTLEESLSLIEESYRQGVRIIVSTSHRRKGMFETPEDIIFRNFS

Cps19fB QVREIAKEVADDLVIAYGAEIYYTLDALEKLEKKEIPTLNDSRYALIEFSMHTSYRQIHT
SaCpsA  IVKHETEKRFEHLQIILYGGELYYTSDMLEKLKLKQIPTLNNTKFALIEFSMQTSWKDIHT

Cps19fB GLSNILMLGITPVIAHIERYDALENNEKRVRELIDMGCYTQINSYHVSKPKFFGEKYKFM
SaCpsA  ALSNVLMLGITPVVAHIERYNALENQKERVRELINMGCYTQINSFHILKQKLFNDKEKRF

Cps19fB KKRARYFLERDLVHVVASDMHNLDSRPPYMQQAYDIIAKKYGAKKAKELFVDNPRKIIMD
SaCpsA  KKRARYFLEENLVHFVASDMHNLDVRPPFLAEAYKIICRDFGKERANQLFIENAQSILKN

Cps19fB QLI
SaCpsA  HYI
```

FIG. 5. Alignment of Cps19fB with *S. agalactiae* CpsA (SaCpsA) (40), as determined using the default settings of the program CLUSTAL (20). Identical residues are boxed; similar residues are shown in boldface.

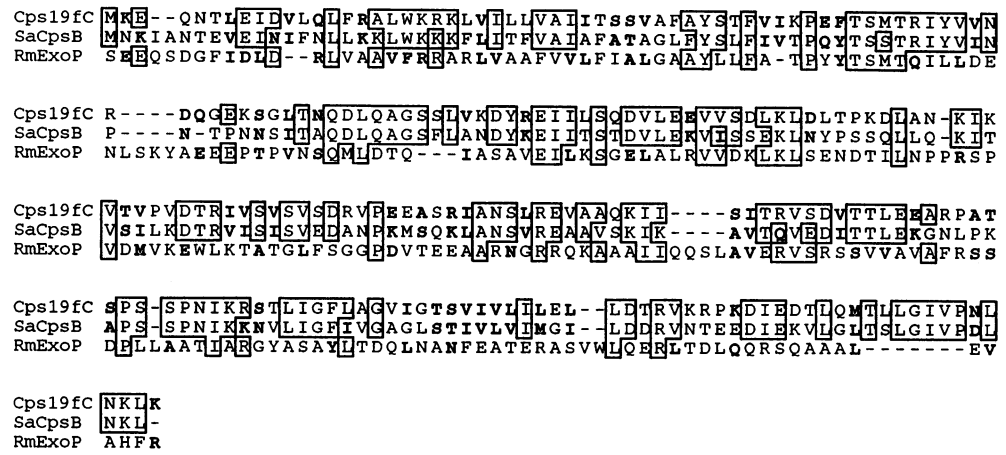

FIG. 6. Alignment of Cps19fC with *S. agalactiae* CpsB (SaCpsB) (40) and *R. meliloti* ExoP (RmExoP) from amino acid positions 20 to 250 (15), as determined using the default settings of the program CLUSTAL (20). Identical residues are boxed; similar residues are shown in boldface. –, absence of a residue.

sequence across the *Eco*RI site was obtained by sequencing the cloned InPCR *Hin*dIII P5/6 product (Fig. 1B).

Examination of the compiled sequence, shown in Fig. 2, reveals the presence of six potential ORFs (nucleotides 169 to 5871) that are arranged as an operon. A seventh, incomplete ORF (nucleotides 5883 to 6322) also forms part of this operon and lies at the end of the sequence data available. This genetic organization is represented in Fig. 1A. An almost perfect consensus promoter sequence (TAGACA-17 bp-TATAAT) is situated 30 bp upstream of the first ORF. Each of the ORFs is preceded by a ribosome binding site (Shine-Dalgarno sequence), and they are all closely coupled, being separated by 1 to 15 nucleotides. Five of the ORFs are in the same reading frame. The terminus of IS*1202* (nucleotide position 82) is located 87 bp upstream of the start codon of the first ORF.

Insertion-duplication mutagenesis of CPS genes. In order to determine whether these ORFs are involved in type 19F capsule production, the chromosomal copies of the potential genes were individually disrupted in the encapsulated strain Rx1-19F and the phenotype of each construct was scored for type 19F capsule production by the quellung reaction. To achieve this, a small internal segment of each ORF (nucleotides 895 to 1221 for *cps19fA*, 1944 to 2258 for *cps19fB*, 2380 to 2998 for *cps19fC*, 3126 to 3519 for *cps19fD*, 3941 to 4544 for *cps19fE*, and 5225 to 5725 for *cps19fF*) was cloned into plasmid pVA891. Subsequent clones were transformed into Rx1-19F, and the cells were plated onto blood agar plates containing erythromycin. As the pVA891 replicon cannot function in pneumococci, erythromycin-resistant transformants are the result of a homologous recombination event directed by the cloned fragment of pneumococcal DNA that leads to the integration of the pVA891 plasmid into the host chromosome and consequent disruption of the gene of interest. Furthermore, the cloned segment of pneumococcal DNA is duplicated and flanks the integrated copy of pVA891.

Disruption of each of the six complete potential ORFs at the correct location in the chromosome was confirmed by Southern hybridization analysis of each insertion-duplication mutant generated (data not shown). All six mutants exhibited a rough phenotype and did not produce a type 19F capsule, as judged

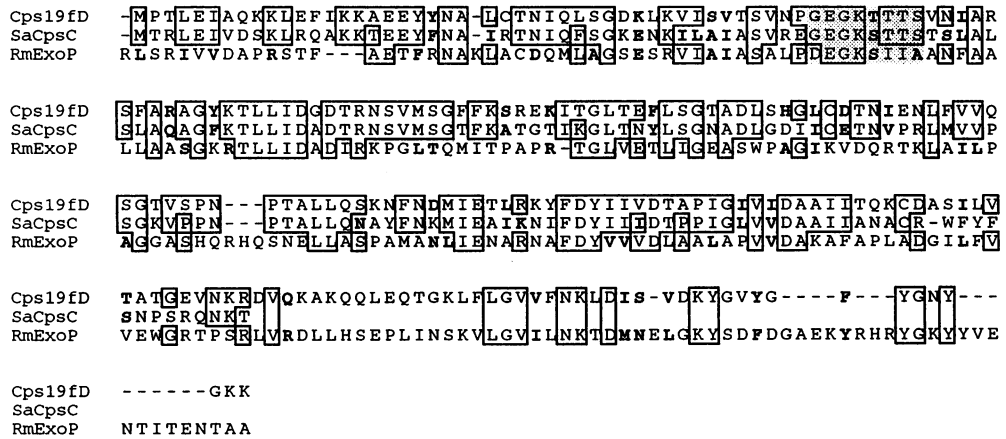

FIG. 7. Alignment of Cps19fD with *S. agalactiae* CpsC (SaCpsC) (40) and *R. meliloti* ExoP (RmExoP) from amino acid positions 542 to 786 (15), as determined using the default settings of the program CLUSTAL (20). Identical residues are boxed; similar residues are shown in boldface. –, absence of a residue. The shaded amino acids frequently align with a region in the ORFs available under the following accession numbers: PIR | S18080, A32812, S18148, B42465; SWISSPROT | P05682, P31856, P18197, P21590; GENPEPT | X75356.

TABLE 2. Homology of Cps19fE with other proteins

Protein	% Identity[a]					
	SaCpsD[b]	LT2RfbP[c]	XcXps2a[d]	RmExoY[e]	RnExoY[f]	RlPss[g]
Cps19fE	45.6 (270)	31.4 (486)	24.1 (464)	37.0 (200)[h]	37.0 (200)[h]	34.0 (206)[h]
SaCpsD		29.2 (274)	36.0 (125)	31.7 (123)	33.3 (123)	38.1 (126)
LT2RfbP			26.7 (490)	50.5 (198)[h]	51.0 (198)[h]	34.3 (210)[h]
XcXps2a				33.5 (197)[h]	32.5 (197)[h]	42.2 (199)[h]
RmExoY					86.7 (226)	33.8 (207)
RnExoY						33.0 (209)

[a] Determined using FASTA, as implemented in PROSIS. Numbers in parentheses indicate amino acid sequence length over which the similarity occurs and is in most cases over the entire length of the protein.
[b] *S. agalactiae* CpsD (40).
[c] *Salmonella enterica* serovar typhimurium strain LT2 RfbP (22).
[d] *X. campestris* Xps2a protein = GumD (14).
[e] *R. meliloti* ExoY (33).
[f] *Rhizobium* species ExoY (17).
[g] *R. leguminosarum* Pss (6).
[h] The alignment occurs at the C terminus of the ORF indicated on the left.

by the quellung reaction. This operon is thus essential for the synthesis of type 19F CPS, and we have therefore designated the ORFs *cps19fA* to *cps19fG*. We have previously demonstrated that insertion-duplication mutagenesis of the ORF within IS*1202* (immediately upstream of *cps19fA*) does not interfere with type 19F capsule production (32).

Characterization of *cps19fA-G*. An analysis of each ORF detected in the DNA sequence presented above is described below. The locations and several properties of each ORF are given in Table 1.

Residues 160 to 481 of Cps19fA exhibit significant homology (27.6% identity) to the entire *Bacillus subtilis lytR* gene product (25). An alignment of the two protein sequences is shown in Fig. 3. LytR is a basic protein that acts as a transcriptional regulator of *lytABC* (autolysin) expression and is thought to be membrane bound via an N-terminal anchoring domain. Cps19fA is also basic and, as can be seen from the hydrophobicity plot in Fig. 4, contains three hydrophobic segments (amino acid positions 16 to 44, 46 to 68, and 74 to 97) near its N terminus which may be membrane-anchoring domains. Interestingly, the predicted LytR protein lacks this hydrophobic region of the sequence but has a single hydrophobic region (amino acid positions 12 to 35) near its N terminus. We therefore suggest that Cps19fA may play a role as a regulator of capsule gene expression.

The *cps19fB* gene product has 63.8% identity with the product of the *Streptococcus agalactiae* (group B streptococcus) *cpsA* gene (40), and their alignment can be seen in Fig. 5. The *S. agalactiae cpsA* gene has been shown to be involved in the production of type III CPS in group B streptococci (40). No other significant sequence similarities were detected, and the functions of these gene products are not known.

The *cps19fC* gene product has 46% identity with the *S. agalactiae* CpsB protein (40), and 22.5% identity with the N-terminal portion of the *Rhizobium meliloti* ExoP protein (15). Figure 6 shows an amino acid sequence alignment of the *cps19fC* gene product with the other two proteins. Inspection of the Cps19fC protein's hydropathy plot indicates that this protein has two hydrophobic segments located at its N and C termini (Fig. 4). The ExoP protein functions in the export of succinoglycan, the major exopolysaccharide of *R. meliloti* (15). Hence, we suggest that Cps19fC is membrane associated and functions in the export of type 19F CPS.

The *cps19fD* gene product has strong homology (59.2% identity) with the *S. agalactiae cpsC* gene product (40). In addition, the Cps19fD protein exhibits homology (29.9% identity) to the C-terminal portion of the *R. meliloti* ExoP protein (15). Figure 7 shows the alignment of Cps19fD with the amino acid sequences of the two other proteins. The *S. agalactiae cpsC* gene product is truncated with respect to Cps19fD. The Cps19fD protein is relatively hydrophilic (Table 1) and has no major hydrophobic segments (results not shown). As stated above, the ExoP protein functions in the export of an exopolysaccharide, and so we infer that Cps19fD functions in the export of the type 19F CPS. Cps19fC and Cps19fD are homologous to the first and last thirds of ExoP, respectively. This may indicate that ExoP has multiple functional domains which are encoded by separate genes in streptococci. We have also detected limited amino acid sequence homology between Cps19fD and a large number of other ORFs in the databases. This is centered on the motif GxGKTTTS (Fig. 7, shaded region). These ORFs have a number of functions, but they do not assist in clarifying the significance of the motif. The motif is similar to that associated with ATP- and nucleotide-binding proteins, although we have not detected similarity to members of those families of proteins.

The *cps19fE* gene product is relatively hydrophobic (Table 1), and it exhibits strong homology with two families of proteins (Table 2). Three are known glycosyltransferases. The first of these is the galactosyltransferase encoded by the *S. agalactiae cpsD* gene (40), which shows 45.6% identity to the *cps19fE* gene product. The second is the *Salmonella enterica* serovar typhimurium strain LT2 *rfbP* gene product (22), which is an undecaprenyl-phosphate galactosephosphate transferase and catalyzes the initial step in O-antigen biosynthesis. The third is the *Xanthomonas campestris xps2a*(*gumD*) gene product, which is a glucose transferase and catalyzes the first step in xanthan gum synthesis (21). The alignment of the amino acid sequences of these proteins is shown in Fig. 8. The N-terminal halves of the proteins have little similarity. The *S. agalactiae* CpsD ORF is truncated at both the N and C termini relative to Cps19fE, although examination of the CpsD coding region indicates that several frameshifts would increase its size (unpublished observations). Examination of the hydrophobicity plot of Cps19fE (Fig. 4) revealed that the N-terminal portion of the protein has three hydrophobic segments which may be potential membrane-spanning domains; this suggests that Cps19fE is anchored to the bacterial membrane. This region of the hydropathy plot is very similar to that of Cps19fA (Fig. 4). Interestingly, although the N-terminal portions of Cps19fE, RfbP, and Xps2a have little amino acid sequence homology, the hydropathy plots for these regions are very similar (Fig. 9).

The C-terminal half of Cps19fE also shows some sequence homology to the *exoY* and *pss* gene products of three *Rhizobium* species (Table 2). The ExoY protein is needed for the addition of galactose to the lipid carrier that initiates the synthesis of the succinoglycan subunit of this organism's exopolysaccharide; the ExoF protein is also needed for this step (14). These proteins also show a high degree of homology with the C-terminal halves of *S. agalactiae* CpsD, RfbP, and Xps2a proteins (Table 2). The amino acid sequence alignments for these proteins are also shown in Fig. 8. The hydropathy plots for ExoY and Pss proteins (Fig. 9) revealed the presence of a hydrophobic segment near their N termini. This is also present at the beginning of the C-terminal halves of Cps19fE, *S.*

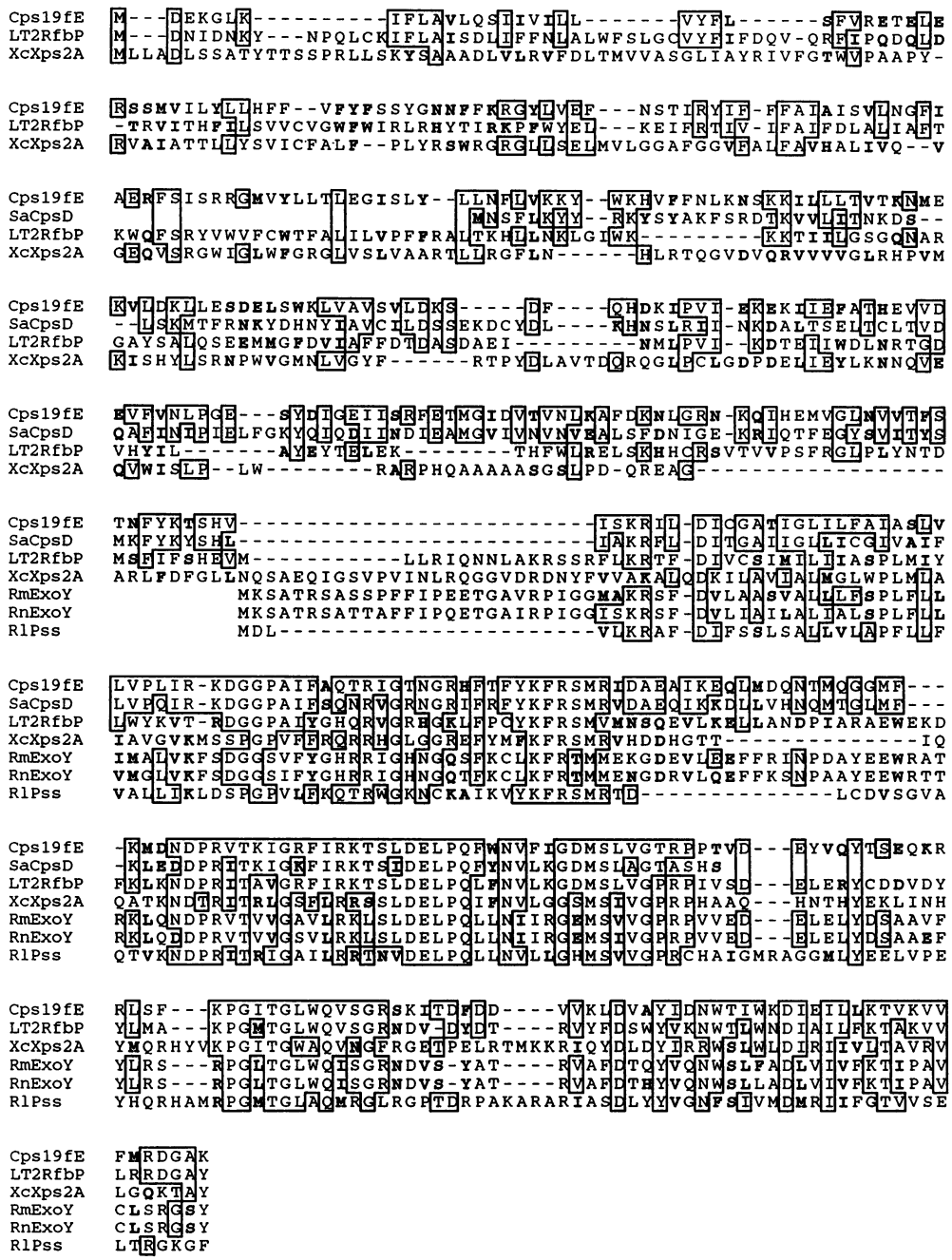

FIG. 8. Alignment of the amino acid sequences of Cps19fE, *Salmonella enterica* serovar typhimurium strain LT2 RfbP (LT2RfbP) (22), *X. campestris* Xps2a protein = GumD (XcXps2a) (14), *S. agalactiae* CpsD (SaCpsD) (40), *R. meliloti* ExoY (RmExoY) (33), *Rhizobium* species ExoY (RnExoY) (17), and *R. leguminosarum* Pss (RlPss) (6), as determined using the default settings of the program CLUSTAL (20). Identical residues are boxed; similar residues are shown in boldface. –, absence of a residue.

agalactiae CpsD, RfbP, and Xps2a. The conservation of the carboxy-terminal halves of the latter proteins with those encoded by the discrete *exoY* and *pss* genes implies that Cps19fE, *S. agalactiae* CpsD, RfbP, and Xps2a may be bifunctional proteins. Glucksmann et al. (14) have suggested that ExoF (which appears to function at the same step as ExoY [38]) may perform the role of the N-terminal portion of RfbP. Recently, the RfbP protein has been shown to have a role in O-antigen synthesis other than that provided by its galactosyl-transferase activity (44). We suggest that the nonconserved, hydrophobic, N-terminal-half Cps19fE, *S. agalactiae* CpsD, RfbP, and Xps2a proteins may function in interacting with the respective lipid carrier of each species and in transferring specific sugars to the lipid carrier. This is consistent with the absence of galactose from type 19F CPS and xanthan gum. The C-terminal halves of Cps19fE, *S. agalactiae* CpsD, RfbP, and Xps2a proteins and of ExoY and Pss proteins may represent a conserved domain which is involved in the initiation of polysaccharide subunit

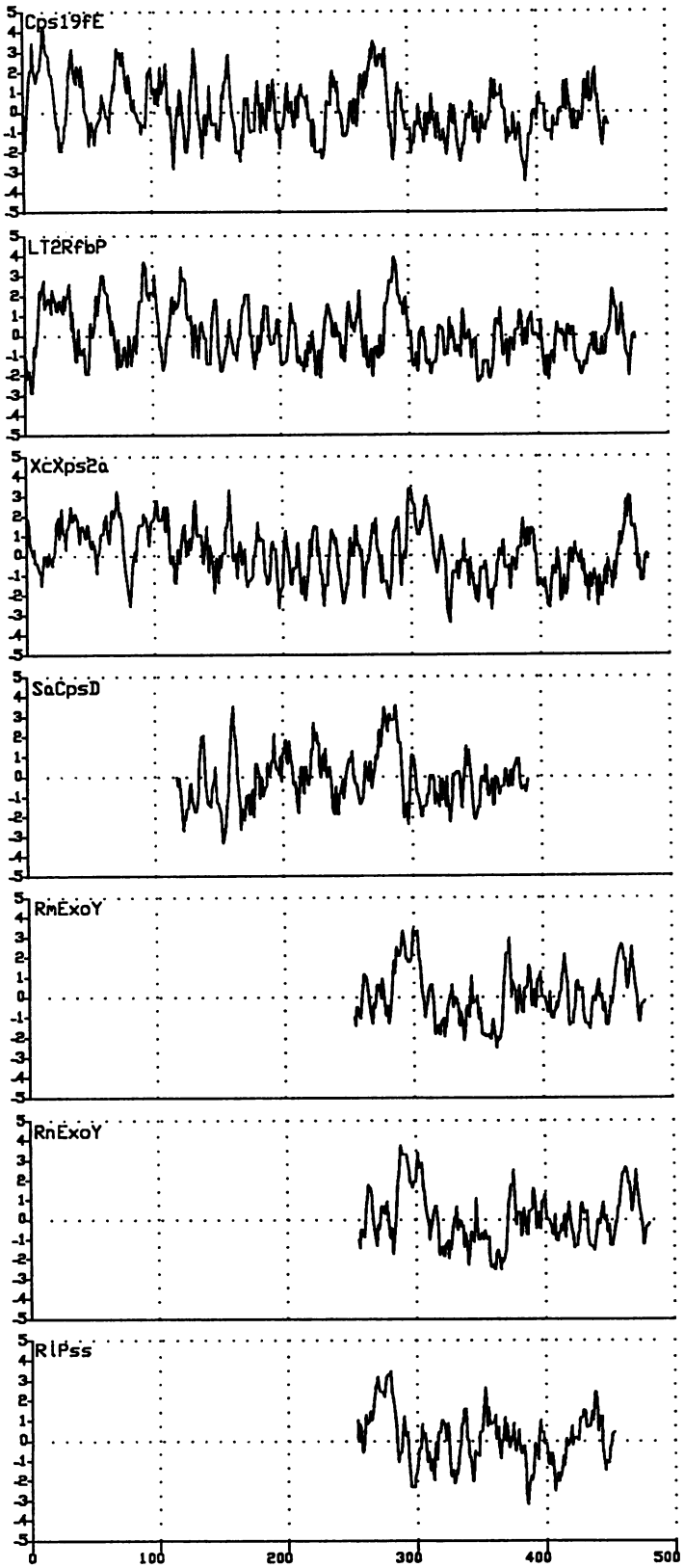

FIG. 9. Comparison of the hydropathy plots of Cps19fE-related proteins. Hydropathy plots of the indicated proteins (designated as described in the legend to Fig. 8) were generated by the method of Kyte and Doolittle (24), as implemented in PROSIS. Numbers on the *x* axis indicate amino acid positions for Cps19fA, LT2RfbP, and XcXps2a. The plots of SaCpsD, RmExoY, RnExoY, and RlPss are aligned with the homologous portion of Cps19fE. Positive numbers on the *y* axis indicate hydrophobic regions.

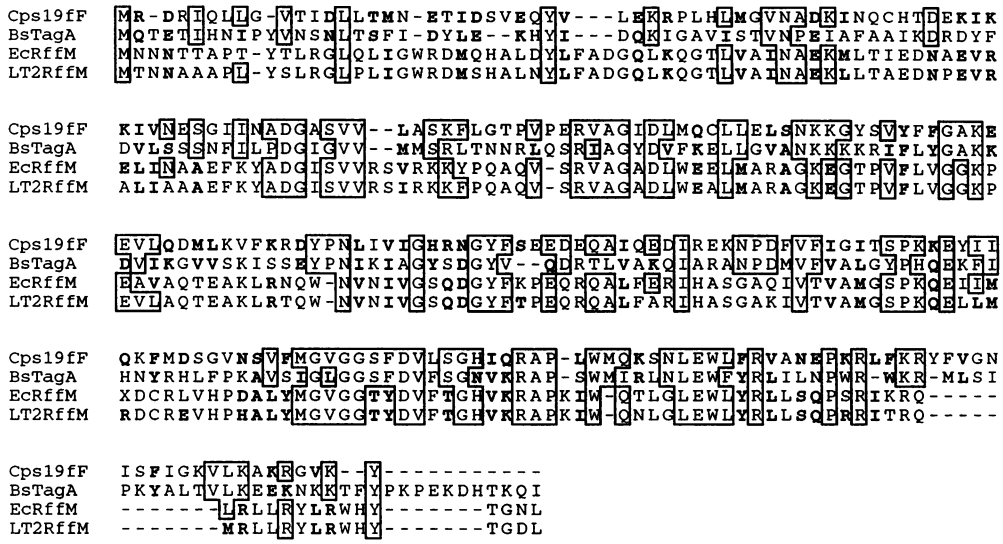

FIG. 10. Alignment of the amino acid sequences of Cps19fF, *B. subtilis* TagA (BsTagA) (30), *E. coli* K-12 RffM (EcRffM) (8), and *Salmonella typhimurium* LT2 RffM (LT2RffM) (27), as determined using the default settings of the program CLUSTAL (20). Identical residues are boxed; similar residues are shown in boldface. –, absence of a residue.

synthesis, as described by Reed et al. (37). However, there are insufficient data for any of the systems described above to allow us to assign functions to the two domains conclusively. Together, these observations suggest that Cps19fE is a glycosyltransferase with as yet undetermined sugar specificity.

The *cps19fF* gene product has homology (31 to 33% identity) with three other proteins, namely, the *B. subtilis tagA* gene product which is needed for cell wall teichoic acid biosynthesis (30) and the *E. coli* K-12 and *Salmonella enterica* serovar typhimurium strain LT2 *rffM* gene products, which are putative UDP-*N*-acetyl-D-mannosaminuronic acid transferases, involved in the synthesis of enterobacterial common antigen (8, 27). An alignment of their amino acid sequences (Fig. 10) shows several areas which are highly conserved. Since the type 19F CPS contains *N*-acetyl-D-mannosamine, we propose that Cps19fF is the transferase which catalyzes the addition of this sugar in the synthesis of the type 19F polysaccharide.

The *cps19fG* gene (incomplete, positions 5887 to 6322) encodes a truncated protein which exhibits 38% identity with the amino-terminal end of the LicD protein of *H. influenzae*, which is encoded by the *licD* gene of its lipopolysaccharide locus (46). Figure 11 shows an alignment of the truncated Cps19fG with LicD. However, the precise function of LicD is unknown.

Conservation of pneumococcal *cps* loci. In view of the level of conservation of *cps19fB*, *cps19fC*, *cps19fD*, and *cps19fE* between *S. pneumoniae* type 19F and related genes in group B streptococci, we have used various probes to look for the presence of similar genes in other *S. pneumoniae* serotypes. Each of the type 19F genes sequenced to date were used to probe (at high stringency) Southern blots of restricted chromosomal DNA from representative pneumococci belonging to several other clinically important serotypes, as well as other members of serogroup 19 (Table 3). Large variations in the hybridization patterns were obtained with the different gene-specific probes. All serotypes tested hybridized with *cps19fA* and *cps19fB*, but not all serotypes hybridized to *cps19fC*, *cps19fD*, *cps19fE*, *cps19fF*, and *cps19fG*. Within group 19, types 19B and 19C exhibited hybridization patterns identical to that of 19F. However, type 19A does not appear to have sequences closely related to those of *cps19fC*, *cps19fD*, *cps19fE*, *cps19fF*, and *cps19fG*. Of the other serotypes, type 14 appears the most similar to type 19F, while types 2, 6, and 19A are the least similar. There did not appear to be any consistent correlation between hybridization pattern and the chemical structure of the various polysaccharide serotypes.

These data support the suggestion that pneumococcal *cps* loci have similar overall genetic organizations. Within these loci, however, some functions have been highly conserved between different serotypes, while others are serotype specific, as would be expected given the differences in the polysaccharide composition and structure of each serotype.

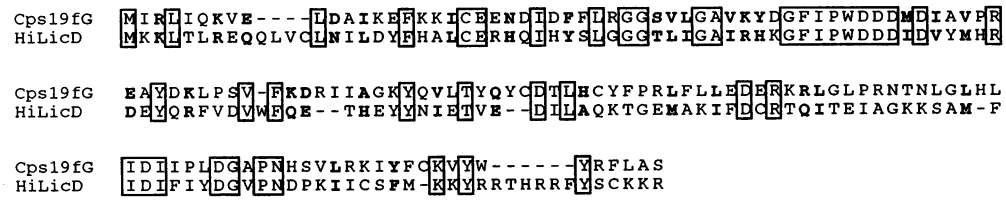

FIG. 11. Alignment of the amino-terminal portion of Cps19fG with the amino-terminal portion of *H. influenzae* LicD (HiLicD) (46), as determined using the default settings of the program CLUSTAL (20). Identical residues are boxed; similar residues are shown in boldface. –, absence of a residue.

TABLE 3. Hybridization of 19F CPS genes with other pneumococcal CPS loci

Serotype	Hybridization with probe[a]:						
	cps19fA	*cps19fB*	*cps19fC*	*cps19fD*	*cps19fE*	*cps19fF*	*cps19fG*
19F	+	+	+	+	+	+	+
19B	+	+	+	+	+	+	+
19C	+	+	+	+	+	+	+
14	+	+	+	+	+	−	−
3	+	+	+	+	−	−	−
4	+	+	+	+	−	−	−
18C	+	+	+	+	−	−	−
23F	+	+	+	+	−	−	−
2	+	+	−	−	−	−	−
6	+	+	−	−	−	−	−
19A	+	+	−	−	−	−	−

[a] DNA fragments from nucleotide positions 336 to 1468, 1571 to 2380, 2380 to 2998, 3126 to 3739, 3682 to 4979, 5225 to 5725, and 5969 to 6322 were labelled with digoxigenin and used as probes for *cps19fA*-, *cps19fB*-, *cps19fC*-, *cps19fD*-, *cps19fE*-, *cps19fF*-, and *cps19fG*-related sequences, respectively. +, hybridization at high stringency; −, no hybridization.

Conclusions. We have used sequential rounds of insertion-duplication mutagenesis, followed by excision and rescue of plasmid DNA and flanking sequences, to isolate a region of chromosomal DNA derived from a type 19F strain that is essential for capsule production. We have characterized a region of DNA containing six complete ORFs and one incomplete ORF (designated *cps19fA-G*), which are arranged as a single transcriptional unit. Insertion-duplication mutagenesis of each of these ORFs in the type 19F chromosome results in a nonencapsulated phenotype. Interruption of the ORF immediately upstream of *cps19fA*, however, has no effect on capsule production. Thus, we believe we have isolated the first six genes (and part of the seventh gene) of the pneumococcal type 19F *cps* locus. Although the size of the locus is currently unknown, we are using further rounds of pVA891 insertion and rescue to isolate additional downstream sequences. The effect of ORF interruption on capsular phenotype is being used to determine which of the additional downstream genes are involved in CPS biosynthesis, thereby delineating the *cps* locus.

Clues to the possible function of some of the gene products have been provided by comparisons with known proteins whose sequences have been deposited with databases (as described above). However, confirmation of the importance of each gene product for CPS production, and elucidation of the function of each protein, will require characterization of the phenotypic impact of in-frame deletions in the respective ORF. The insertion-duplication mutants generated in the present study are not suitable for this purpose because of the likelihood of polar effects. Nevertheless, this procedure is appropriate for determining whether or not a given ORF is part of an operon essential for capsule production.

These studies provide a first step towards a better understanding of the complex genetic and biochemical processes required for the production of CPS by *S. pneumoniae*. At present, the only other sequence accessible on GenBank for pneumococcal capsule genes is that of a portion of the UDP-glucose dehydrogenase gene from type 3 (13). During review of our manuscript, however, Dillard and Yother (9) published data on the genetic organization of a segment of *S. pneumoniae* type 3 DNA sufficient to transform an unrelated strain to type 3 capsule production. They proposed that the region contained four genes encoding the UDP-glucose dehydrogenase, a polysaccharide synthase, a UTP:glucose-1-phosphate uridyltransferase, and a phosphomutase (9). Sequence data supporting this were presented at the American Society for Microbiology Fourth International Conference on Streptococcal Genetics (10). In the present study, we demonstrated that *cps19fA*-, *cps19fB*-, *cps19fC*-, and *cps19fD*-specific probes all hybridized at high stringency to type 3 DNA. However, it is not yet possible to compare the type 3 and type 19F sequences, as the former is not accessible on GenBank. Interestingly, Dillard and Yother (10) demonstrated a significant degree of sequence homology between the type 3 locus and portions of the hyaluronic acid biosynthesis (*has*) locus of group A streptococci reported by Van de Rijn et al. (43). Thus, there appear to be close evolutionary relationships between pneumococcal *cps* loci and analogous loci of both group A and group B streptococci.

ACKNOWLEDGMENTS

We thank Andrew Lawrence and Anne Berry for assistance with the quellung reaction and pneumococcal transformation, respectively. We also thank P. Reeves for communication of data prior to publication.

This work was supported by grants from the World Health Organization Programme for Vaccine Development and the National Health and Medical Research Council of Australia.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.
2. **Austrian, R.** 1981. Some observations on the pneumococcus and on the current status of pneumococcal disease and its prevention. Rev. Infect. Dis. **3**(Suppl.):S1–17.
3. **Austrian, R., H. P. Bernheimer, E. E. B. Smith, and G. T. Mills.** 1959. Simultaneous production of two capsular polysaccharides by pneumococcus. II. The genetic and biochemical bases of binary capsulation. J. Exp. Med. **110**:585–602.
4. **Avery, O. T., and R. Dubos.** 1931. The protective action of a specific enzyme against type III pneumococcus infections in mice. J. Exp. Med. **54**:73–89.
5. **Berry, A. M., J. Yother, D. E. Briles, D. Hansman, and J. C. Paton.** 1989. Reduced virulence of a defined pneumolysin-negative mutant of *Streptococcus pneumoniae*. Infect. Immun. **57**:2037–2042.
6. **Borthakur, D., R. F. Barker, J. W. Latchford, L. Rossen, and A. W. B. Johnston.** 1988. Analysis of *pss* genes of *Rhizobium leguminosarum* required for exopolysaccharide synthesis and nodulation of peas: their primary structure and their interaction with *psi* and other nodulation genes. Mol. Gen. Genet. **213**:155–162.
7. **Brown, M. C. M., A. Weston, J. R. Saunders, and G. O. Humphreys.** 1979. Transformation of *E. coli* C600 by plasmid DNA at different phases of growth. FEMS Microbiol. Lett. **5**:219–222.
8. **Daniels, D. L., G. Plunkett III, V. D. Burland, and F. R. Blattner.** 1992. Analysis of the *Escherichia coli* genome: DNA sequence of the region from 84.5 to 86.5 minutes. Science **257**:771–778.
9. **Dillard, J. P., and J. Yother.** 1994. Genetic and molecular characterization of capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* type 3. Mol. Microbiol. **12**:959–972.
10. **Dillard, J. P., and J. Yother.** 1994. Genetic organization and molecular characterization of the casette containing genes for the production of type 3 capsular polysaccharide in *Streptococcus pneumoniae*. Program Abstr. Am. Soc. Microbiol. Fourth Int. Conf. Streptococcal Genet., abstr. M45.
11. **Douglas, R. M., J. C. Paton, S. J. Duncan, and D. J. Hansman.** 1983. Antibody response to pneumococcal vaccination in children younger than five years of age. J. Infect. Dis. **148**:131–137.
12. **Frosch, M., C. Weisgerber, and T. F. Meyer.** 1989. Molecular characterization and expression in *E. coli* of the gene complex encoding the polysaccharide capsule of *Neisseria meningitidis* group B. Proc. Natl. Acad. Sci. USA **86**:1669–1673.
13. **Garcia, E., P. Garcia, and R. Lopez.** 1993. Cloning and sequencing of a gene involved in the synthesis of the capsular polysaccharide of *Streptococcus pneumoniae* type 3. Mol. Gen. Genet. **239**:188–195.

14. **Glucksmann, M. A., T. L. Reuber, and G. C. Walker.** 1993. Family of glycosyl transferases needed for the synthesis of succinoglycan by *Rhizobium meliloti*. J. Bacteriol. **175**:7033–7044.
15. **Glucksmann, M. A., T. L. Reuber, and G. C. Walker.** 1993. Genes needed for the modification, polymerization, export, and processing of succinoglycan by *Rhizobium meliloti*: a model for succinoglycan biosynthesis. J. Bacteriol. **175**:7045–7055.
16. **Gray, B. M., G. M. Converse III, and H. C. Dillon, Jr.** 1980. Epidemiological studies of *Streptococcus pneumoniae* in infants: acquisition, carriage and infection during the first 24 months of life. J. Infect. Dis. **142**:923–933.
17. **Gray, J. X., M. A. Djordjevic, and B. G. Rolfe.** 1990. Two genes that regulate exopolysaccharide production in *Rhizobium* sp. strain NGR234: DNA sequences and resultant phenotypes. J. Bacteriol. **172**:192–203.
17a. **Guidolin, A., J. K. Morona, R. Morona, D. Hansman, and J. C. Paton.** 1994. Analysis of genes essential for the production of *Streptococcus pneumoniae* type 19F capsular polysaccharide. Program Abstr. Am. Soc. Microbiol. Fourth Int. Conf. Streptococcal Genet., abstr. M47.
18. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
19. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.
20. **Higgins, D. G., and P. M. Sharp.** 1988. CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. Gene **73**:237–244.
21. **Ielpi, L., R. O. Couso, and M. A. Dankert.** 1993. Sequential assembly and polymerization of the polyprenol-linked pentasaccharide repeating unit of the xanthan polysaccharide in *Xanthomonas campestris*. J. Bacteriol. **175**:2490–2500.
22. **Jiang, X.-M., B. Neal, F. Santiago, S. J. Lee, L. K. Romana, and P. R. Reeves.** 1991. Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). Mol. Microbiol. **5**:695–713.
23. **Kroll, J. S., S. Zamze, B. Loynds, and E. R. Moxon.** 1989. Common organization of chromosomal loci for production of different capsular polysaccharides in *Haemophilus influenzae*. J. Bacteriol. **171**:3343–3347.
24. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydrophobic character of a protein. J. Mol. Biol. **157**:105–132.
25. **Lazarevic, V., P. Margot, B. Soldo, and D. Karamata.** 1992. Sequencing and analysis of the *Bacillus subtilis lytRABC* divergon: a regulatory unit encompassing the structural genes of the *N*-acetylmuramoyl-L-alanine amidase and its modifier. J. Gen. Microbiol. **138**:1949–1961.
26. **Lee, C.-J., S. D. Banks, and J. P. Li.** 1991. Virulence, immunity and vaccine related to *Streptococcus pneumoniae*. Crit. Rev. Microbiol. **18**:89–114.
27. **Lu, C.-D., and A. T. Abdelal.** 1993. The *Salmonella typhimurium* uracil-sensitive mutation *use* is in *argU* and encodes a minor arginine tRNA. J. Bacteriol. **175**:3897–3899.
28. **Macrina, F. L., R. P. Evans, J. A. Tobian, D. L. Hartley, D. B. Clewell, and K. R. Jones.** 1983. Novel shuttle plasmid vehicles for *Escherichia-Streptococcus* transgeneric cloning. Gene **25**:145–150.
29. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
30. **Mauel, C., M. Young, and D. Karamata.** 1991. Genes concerned with synthesis of poly(glycerol phosphate), the essential teichoic acid in *Bacillus subtilis* strain 168, are organized in two divergent transcriptional units. J. Gen. Microbiol. **137**:929–941.
31. **Morelle, G.** 1989. A plasmid extraction procedure on a miniprep scale. Focus **11.1**:7–8.
32. **Morona, J. K., A. Guidolin, R. Morona, D. Hansman, and J. C. Paton.** 1994. Isolation, characterization, and nucleotide sequence of IS*1202*, an insertion sequence of *Streptococcus pneumoniae*. J. Bacteriol. **176**:4437–4443.
33. **Muller, P., M. W. Wang, J. Quandt, W. Arnold, and A. Puhler.** 1993. Genetic analysis of *Rhizobium meliloti exoYFQ* operon: ExoY is homologous to sugar transferases and ExoQ represents a transmembrane protein. Mol. Plant Microbe Interact. **6**:55–65.
34. **Musher, D. M.** 1992. Infections caused by *Streptococcus pneumoniae*: clinical spectrum, pathogenesis, immunity and treatment. Clin. Infect. Dis. **14**:801–807.
35. **Ochman, H., A. S. Gerber, and D. L. Hartl.** 1988. Genetic applications of an inverse polymerase chain reaction. Genetics **120**:621–623.
36. **Paton, J. C., A. M. Berry, R. A. Lock, D. Hansman, and P. A. Manning.** 1986. Cloning and expression in *Escherichia coli* of the *Streptococcus pneumoniae* gene encoding pneumolysin. Infect. Immun. **54**:50–55.
37. **Reed, J. W., M. Capage, and G. C. Walker.** 1991. *Rhizobium meliloti exoG* and *exoJ* mutations affect the ExoX-ExoY system for modulation of exopolysaccharide production. J. Bacteriol. **173**:3776–3788.
38. **Reuber, T. L., and G. C. Walker.** 1993. Biosynthesis of succinoglycan, a symbiotically important exopolysaccharide of *Rhizobium meliloti*. Cell **74**:269–280.
39. **Roberts, I., R. Mountford, R. Hodge, K. B. Jann, and G. J. Boulnois.** 1988. Common organization of gene clusters for production of different capsular polysaccharides (K antigens) in *Escherichia coli*. J. Bacteriol. **170**:1305–1310.
40. **Rubens, C. E., L. M. Heggen, R. F. Haft, and M. R. Wessels.** 1993. Identification of *cpsD*, a gene essential for type III capsule expression in group B streptococci. Mol. Microbiol. **8**:843–855.
41. **Shoemaker, N. B., and W. R. Guild.** 1974. Destruction of low efficacy markers is a slow process occurring at a heteroduplex stage of transformation. Mol. Gen. Genet. **128**:283–290.
42. **Southern, E.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
43. **Van de Rijn, I., D. Crater, and B. Dougherty.** 1994. Molecular analysis of the group A streptococcal hyalurinic acid capsule operon. Program Abstr. Am. Soc. Microbiol. Fourth Int. Conf. Streptococcal Genet., abstr. M20.
44. **Wang, L., and P. R. Reeves.** 1994. Involvement of the galactosyl-1-phosphate transferase encoded by the *Salmonella enterica rfbP* gene in O-antigen subunit processing. J. Bacteriol. **176**:4348–4356.
45. **Watson, D. A., and D. M. Musher.** 1990. Interruption of capsule production in *Streptococcus pneumoniae* serotype 3 by insertion of transposon Tn*916*. Infect. Immun. **58**:3135–3138.
46. **Weisner, J. N., J. M. Love, and E. R. Moxon.** 1989. The molecular mechanism of phase variation of *H. influenzae* lipopolysaccharide. Cell **59**:657–665.
47. **Yother, J., L. S. McDaniel, and D. E. Briles.** 1986. Transformation of encapsulated *Streptococcus pneumoniae*. J. Bacteriol. **168**:1463–1465.