J. Biochem. 123, 937-945 (1998)

Diversity of Capsular Polysaccharide Synthesis Gene Clusters in *Streptococcus pneumoniae*[1]

Marc A.B. Kolkman,[2,3] Bernard A.M. van der Zeijst,[4] and Piet J.M. Nuijten[2,5]

Department of Bacteriology, Institute of Infectious Diseases and Immunology, School of Veterinary Medicine, Universiteit Utrecht, P.O. Box 80.165, 3508 TD Utrecht, The Netherlands

Received for publication, December 13, 1997

Streptococcus pneumoniae **comprises 90 serotypes, each one having its own specific polysaccharide capsule. In order to explore the diversity of capsular polysaccharide synthesis (*cps*) gene clusters in *S. pneumoniae*, we performed cross-hybridizations between the 12 *cps* genes of *S. pneumoniae* serotype 14 and chromosomal DNA of 26 strains comprising 26 different capsule types. Large variations in the hybridization patterns were observed. The genes *cps14A* to *cps14D* are conserved in most serotypes. Sequences homologous to *cps14I* to *cps14L* were only observed in the four types of serogroup 15, which all have a capsule structure similar to that of type 14. By using a *cps14E* knock-out construct, *cpsE* mutants of the pneumococcal types 9N, 13, and 15B were obtained. These mutants were unencapsulated and showed reduced glycosyltransferase activity, indicating that the pneumococcal types 9N, 13, and 15B express a glucosyl-1-phosphate transferase which is homologous to Cps14E. Glycosyltransferase assays showed that among 21 pneumococcal types which contain glucose in the core of their capsule polysaccharide, 19 types express glucosyl-1-phosphate transferase activity. However, not all of these types hybridized strongly with *Cps14E*, the type 14 glucosyl-1-phosphate transferase gene. Thus, pneumococci possess glucosyltransferase genes distinct from *cps14E*, but encoding enzymes with identical activity. All serotypes which synthesized lipid-linked lactose intermediates in glycosyltransferase activity assays (type 11B, 13, 15F, 15A, 15B, 15C) hybridized with *cps14G*. This gene encodes a galactosyltransferase which catalyzes the addition of 1,4-linked β-galactose to lipid-linked glucose. The *cps14G* homologues in type 11B, 13, 15F, 15A, 15B, and 15C may encode a similar β-galactosyltransferase activity as *cps14G* in type 14.**

Key words: capsule, glycosyltransferases, polysaccharides, *Streptococcus pneumoniae*.

Polysaccharide capsules are ubiquitous structures found on the cell surface of a broad range of bacterial species. Capsular polysaccharides (CPs) are generally composed of repeating oligosaccharides, consisting of 2 to 10 monosaccharides, sometimes complemented with other components. Any 2 monosaccharides may be joined in a number of configurations as a consequence of the multiple hydroxyl groups within each monosaccharide that may be involved in the formation of a glycosidic bond. As a result of this, CPs are an incredibly diverse range of molecules that may differ not only by the type and number of monosaccharides within a subunit, but also in how the monosaccharides are joined together. Biosynthesis of heteropolymeric CPs with a more or less complex oligosaccharide subunit structure requires a complex pathway, starting with the uptake or synthesis of the monosaccharide components and their conversion into nucleotide derivatives. A membrane-associated transferase complex would catalyze the successive linkage of monosaccharides to a lipid carrier molecule, which is followed by polymerization of the subunits, transport and attachment of the complete CP to the cell surface.

Streptococcus pneumoniae, a human pathogen causing serious invasive diseases, produces a polysaccharide capsule that is required for virulence. The capsule protects the bacterium from the host's immune system and prevents phagocytosis (*1*). Currently, 90 different serotypes have been identified (*2*), each one expressing its own distinct CP. Although the chemical structures of many of these CPs have been established (*3*), studies on the genes involved in the capsular polysaccharide synthesis (*cps*) and the function of their translation products started only recently (*4*).

[1] Nucleotide sequence accession number: The nucleotide sequence of the *S. pneumoniae* serotype 14 capsular polysaccharide synthesis locus is available under Genbank accession number X85787.

[2] M.A.B.K. was supported by the Dutch Organization for Scientific Research (NWO), grant number 900-514-150. P.J.M.N. was supported by the Utrecht Institute for Infection and Immunity.

[3] Present address: Department of Immunohaematology and Bloodbank, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, The Netherlands.

[4] To whom correspondence should be addressed. Phone: +31-30-253-4888, Fax: +31-30-254-0784, E-mail: zeijst@vetmic.dgk.ruu.nl

[5] Present address: Intervet International BV, P.O. Box 31, 5830 AA Boxmeer, The Netherlands.

Abbreviations: CP, capsular polysaccharide; cps, capsular polysaccharide synthesis; CSP, competence stimulating peptide; ORF, open reading frame.

Classic experiments demonstrated that genes specific for the synthesis of a given CP can be tranferred as a unit during genetic transformation (*5*), indicating that the *cps* genes are closely linked in the pneumococcal chromosome. This has been confirmed in recent studies for type 1 (*6*), 3 (*7-9*), 14 (*10-12*), 19B (*13*), and 19F (*14, 15*).

The *cps* locus of *S. pneumoniae* serotype 14 (*cps14*) comprises 12 genes (Fig. 1), four of these genes (*cps14E*, *cps14G*, *cps14I*, and *cps14J*) encode the glycosyltransferases required for the synthesis of the type 14 tetrameric subunit (*11, 12*). Based on sequence comparison, other putative functions have been deduced for the gene products of *cps14A* (regulation), *cps14C* (chain length determination), *cps14D* (export), *cps14H* (polysaccharide polymerization), and *cps14L* (transport of the repeating unit) (*11, 12*). The *cps19f* gene cluster consists of 15 genes, all required for CP biosynthesis in *S. pneumoniae* type 19F (*14, 15*). Besides genes encoding (putative) functions as described above for serotype 14, the *cps19f* locus also contains genes involved in the synthesis of the nucleotide sugars, needed for assembly of type 19F CP. The *cps* gene cluster of type 19B (*cps19b*) resembles that of type 19F, but the *cps19b* locus contains three additional *cps* genes (*13*). Furthermore, the potential polysaccharide polymerase and the repeating unit transporter of type 19B are distinct from those of type 19F. The *cps* locus of type 1 (*cap1*) contains eleven genes, the *cap1K* gene of this gene cluster encodes a UDP-glucose dehydrogenase (*6*). The *cps* locus of *S. pneumoniae* type 3 is considerably less complex than those of type 1, 14, 19B, and 19F. This locus contains three type-specific genes, which encode a UDP-glucose dehydrogenase, a polysaccharide synthase, and a glucose-1-phosphate uridylyltransferase (*4, 7, 9*). An additional type 3-specific gene, encoding a phosphoglucomutase, was reported by Dillard *et al.* (*7*). However, the corresponding sequence in another type 3 strain is interrupted by an IS*1167* sequence (*4*), indicating that this gene is dispensable for type 3 CP synthesis. Furthermore, the type 3 *cps* locus contains two additional ORFs, almost identical to *cps14C*/*Cps19fC* and *cps14D*/*cps19fD*, however these ORFs are not functional in type 3 pneumococci (*4*).

In this study we explored the genetic diversity of *cps* loci among several pneumococcal strains from various serogroups, using our detailed knowledge of the *cps14* gene cluster. We used individual *cps14* genes for cross-hybridization with chromosomal DNA from 26 different pneumococcal serotypes. Furthermore, we discuss these findings in relation to the type-specific glycosyltransferase activities detected in many of these pneumococcal types.

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions—*S. pneumoniae* serotype 14 was NCTC 11902. The pneumococcal strain 5MC (*nov-r1*) has been described elsewhere (*16*). All other pneumococcal strains were obtained from A.J.W. van Alphen (Academic Medical Center Amsterdam, The Netherlands). Pneumococci were grown at 37°C in Todd-Hewitt broth supplemented with 0.5% yeast extract (THY-medium) in a static culture, or on blood agar plates under aerobic conditions. Tetracycline and novobiocin were used at a concentration of 10 μg/ml for selection of *S. pneumoniae* mutants.

Plasmids and Antibodies—The plasmid pKOM03, containing *cps14E* with a *tetM* insertion mutation, has been described before (*10*). Monoclonal antibody HASP8 (reacting to pneumococcal C-polysaccharide) was obtained from U.B. Skov Sørensen (Statens Seruminstitut, Copenhagen, Denmark).

DNA Techniques—Most DNA techniques were performed as described previously (*17*). Chromosomal DNA was isolated as described by Ausubel *et al.* (*18*). Transformation of the encapsulated pneumococci with chromosomal DNA from strain 5MC by using the competence-stimulating peptides CSP-1 and CSP-2 (kindly provided by D.A. Morrison, University of Illinois at Chicago, Chicago) was essentially performed as described by Pozzi *et al.* (*19*). Mutants were selected for novobiocin resistance. Similarly, transformations were performed with plasmid pKOM03. In this case, transformants were selected for tetracycline resistance and mutations were confirmed by Southern hybridization analysis.

Southern Hybridization Analysis—Chromosomal DNA (approximately 3 μg) was digested with *Eco*RI, electrophoresed in a 1% agarose gel in 0.5 × TBE buffer (1 × TBE = 89 mM Tris/89 mM boric acid/2 mM EDTA) and transferred to a nylon membrane (Hybond N, Amersham) (*17*). The following DNA fragments, either restriction fragments of previously published subclones of cMK02 (*10-12*) or obtained by PCR amplification, were labeled with ^{32}P by using a random priming labeling kit (Amersham) and used as a probe: nucleotides (Genbank accession number X85787) 1419-2916, 3126-3631, 3631-4252, 4235-4934, 4965-6393, 6396-6868, 6766-7272, 7279-8514, 8465-9448, 9450-10501, 10668-11365, 11379-12899 for *cps14A* to *cps14L*, respectively. Labeled DNA fragments of nucleotides 12810-13595 and 13173-13471 were used as an entire *orfX* probe, and as an internal *orfX* probe lacking

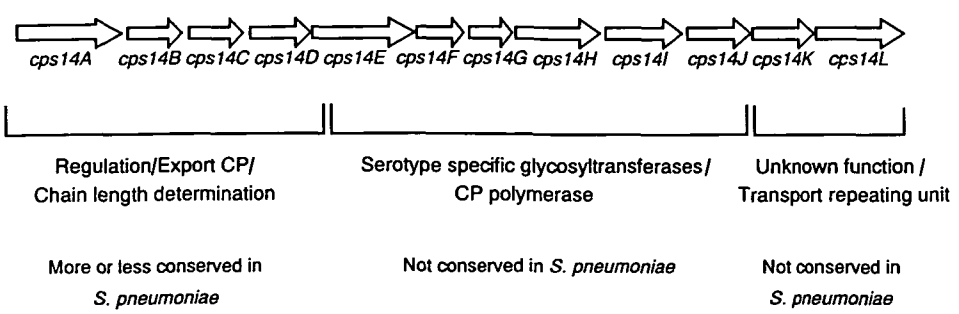

Fig. 1. **Organization of the *cps* locus of *S. pneumoniae* serotype 14.** A central region containing the type-specific glycosyltransferases and the putative CP polymerase is flanked by two regions presumed to encode functions such as regulation and transport (see text for details). Only the genes *cps14A* to *cps14D* are more or less conserved in *S. pneumoniae*. The nucleotide sequence of the *cps14* gene cluster is available in the GenBank under accession number X85787.

the IS*1167* sequence at the 3′ end of *orfX*, respectively. Hybridizations were carried out at 60°C without formamide. Washes were performed three times with 2×SSPE (1×SSPE=0.18 M NaCl/10 mM NaH_2PO_4/1 mM EDTA [pH 7.7]); 0.1% SDS at 60°C for 15 min.

Glycosyltransferase Assays and Analysis of Lipid-Linked CP Intermediates by Thin Layer Chromatography—Preparation of pneumococcal membranes and glycosyltransferase activity assays were essentially performed as described before (*10*). For each reaction, pneumococcal membrane preparation (approximately 25 μg protein) was incubated at 10°C for 1 h with 0.05 μCi UDP-[^{14}C]glucose (Amersham, 296 mCi/mmol) or 0.05 μCi UDP-[^{14}C]galactose (Amersham, 305 mCi/mmol), and 10 mM $MgCl_2$ in a final volume of 50 μl. Reactions were stopped by the addition of 1 ml chloroform-methanol (2:1). This solution was extracted three times with 0.2 ml of PSUP (1.5 ml of chloroform, 25 ml of methanol, 23.5 ml of water, and 0.183 g of KCl). The incorporation of ^{14}C label into the glycolipid fraction in the organic phase was measured in a scintillation counter (Beckman). Lipid-linked intermediates were hydrolyzed from the lipid carrier by mild acid hydrolysis (*11*), dried in a Speed-Vac, and resuspended in 40% 2-propanol containing 5 mg/ml unlabeled glucose, galactose, and lactose, serving as carriers and internal standards. Thin layer chromatography (TLC) was carried out on HPTLC silica gel (Merck), developed in 1-butanol-ethanol-water (5:3:2), sprayed with En^{3}hance spray (DuPont) and autoradiographed for 1 to 2 days. Unlabeled sugar standards were visualized by spraying the TLC plate with 5% H_2SO_4 in ethanol and heating at 100°C for 10 min.

RESULTS

Distribution of cps14 Genes in Other Pneumococcal Serotypes—In previous studies, we identified twelve genes (*cps14A-L*) which are involved in CP synthesis in *S. pneumoniae* serotype 14 (*10-12*). In this study, we performed cross-hybridization experiments in order to examine the presence of *cps* genes in other pneumococcal serotypes. DNA fragments encoding individual *cps14* genes were ^{32}P labeled, and used to probe Southern blots of digested chromosomal DNA of 26 strains comprising 26 capsule types belonging to 20 different serogroups. Large variations in the hybridization patterns were observed (Table I).

All pneumococcal types tested hybridized with the *cps14A* probe (Table I). The *cps14A* gene is almost identical to *cps19fA* of *S. pneumoniae* type 19F (*12, 14*), which also hybridized with DNA of all 20 pneumococcal strains, belonging to 16 different serogroups, tested by Morona *et al.* (*15*). Cps14A/Cps19fA are possibly regulatory proteins, that have homology with LytR of *Bacillus subtilis*, a transcriptional regulator of the *lytABC* operon (*12, 14*). All pneumococcal types used by Morona *et al.* (*15*) also hybridized with *cps19fB*, which is almost identical to *cps14B* of type 14. However, we observed that the types 11A and 11B, which were not tested with the *cps19fB* probe, did not hybridize with *cps14B*. Sequences closely related to *cps14C* and *cps14D* were also found in many serotypes, but not all (Table I). The function of Cps14B (and its homologues) is not known, but Cps14C and Cps14D may have a role in chain length determination and export of CP (*11*). The functions encoded by *cps14A-D* are not type-specific, since *cps14A-D* are found in several, structurally unrelated capsule types.

The *cps14E* gene, encoding a glucosyl-1-phospate transferase catalyzing the transfer of glucose (Glc) to a lipid carrier (*10*), hybridized with DNA of seven members of five different serogroups (Table I). These types (9N, 13, 15B, 15C, 18A, 18F, 19F) contain a Glc residue in the backbone of their CP structure (*3*). These data indicate that a glucosyltransferase closely related to Cps14E may be responsible for the first step in the CP biosynthesis in these serotypes (see below).

Sequences homologous to *cps14G*, a gene encoding a β-1,4-galactosyltransferase that catalyzes the addition of β-1,4-linked galactose (Gal) to lipid-linked Glc (*11*), were only observed in type 13, 15F, 15A, 15B, and 15C. The subtypes 11A and 11B showed a weak hybridization signal with the *cps14G* probe. As in serotype 14, all these subtypes contain Galβ(1→4)βGlc as part of the repeating unit, except type 11A which has Galβ(1→4)αGlc in its CP structure (*3*). In addition, analysis of the intermediates in the synthesis of the repeating subunits in type 11B, 13, 15F, 15A, 15B, and 15C showed that, in the second biosynthesis step, galactose is coupled to lipid-linked glucose (see below). These data suggest that these six pneumococcal types possess a *cps14G* homologue which encodes the galactosyltransferase activity required for the linkage of β-1,4-Gal to lipid-linked Glc. The *cps14G* homologue in type 11A may be responsible for the addition of β-Gal in a 1,4-linkage to α-Glc. Interestingly, *cps14F* and *cps14G* seem to be coupled, since the *cps14F* probe showed a similar hybridization pattern as *cps14G*. Although the precise function of Cps14F is unclear, glycosyltransferase assays indicated that Cps14F has an enhancing role in glycosyltransferase activity (*11*).

The *cps14H* probe only hybridized with DNA of type 15B and 15C. The *cps14H* gene probably encodes the type 14 polysaccharide polymerase (*11*). Type 15B and 15C have a core structure identical to that of serotype 14, whereas type 15F and 15A have a different core antigen (Fig. 2). Thus, the polymerization reaction in type 15F and 15A significantly differs from that in type 14, 15B, and 15C. This may explain the presence of *cps14H*-like sequences in type 15B and 15C, and the absence in type 15F and 15A.

Sequences homologous to *cps14I* to *cps14L* were only found in the four members of serogroup 15 (Table I). *Cps14I* and *cps14J* encode a β-1,3-*N*-acetylglucosaminyltransferase and a β-1,4-galactosyltransferase, respectively (*12*). These enzymes are required for the last two steps in the biosynthesis of the type 14 repeating unit, the β-1,3 linkage of GlcNAc to lipid-linked lactose, and the β-1,4 linkage of Gal to lipid-linked Lac-GlcNac. Both enzyme activities are also required for the synthesis of the type 15F, 15A, 15B, and 15C repeating units (Fig. 2) and thus it is very likely that the *cps14I* and *cps14J* homologues present in these subtypes encode similar enzyme activities.

The function of the gene product encoded by *cps14K* is not clear, but Cps14L, which is homologous to several RfbX-like proteins, is probably involved in transport of the type 14 tetrameric repeating unit (*12*). The activities encoded by *cps14K* and *cps14L* seem to be related to the CP structure, since only type 15F, 15A, 15B, and 15C contain *cps14K*- and *cps14L*-like sequences.

TABLE I. **Hybridization of serotype 14 *cps* genes with chromosomal DNA of other pneumococcal types.**

Type	DNA probes												Presence of glucose in the core structure and monosaccharide linkages in CP			
	cps14A	*cps14B*	*cps14C*	*cps14D*	*cps14E*	*cps14F*	*cps14G*	*cps14H*	*cps14I*	*cps14J*	*cps14K*	*cps14L*	Glc	β1,4 Gal-Glc	β1,3 GlcNac-Gal	β1,4 Gal-GlcNac
1	+	+	+	+	−	−	−	−	−	−	−	−	+	−	−	−
2	+	±	−	−	−	−	−	−	−	−	−	−	+	−	−	−
3	+	+	+	+	−	−	−	−	−	−	−	−	+	−	−	−
4	+	±	+	+	−	−	−	−	−	−	−	−	+	−	−	−
6A	+	±	−	−	−	−	−	−	−	−	−	−	+	−	−	−
9N	+	+	+	+	+	−	−	−	−	−	−	−	+	−	−	−
10A	+	±	−	−	−	−	−	−	−	−	−	−	+	−	−	−
11A	+	−	−	−	−	±	±	−	−	−	−	−	+	+	−	−
11B	+	−	−	−	−	±	±	−	−	−	−	−	+	+	−	−
12F	+	±	−	−	−	−	−	−	−	−	−	−	+	−	−	−
13	+	+	+	+	+	+	+	−	−	−	−	−	+	+	−	−
14	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
15F	+	+	+	−	−	+	+	−	+	+	+	+	+	+	+	+
15A	+	+	+	−	−	+	+	−	+	+	+	+	+	+	+	+
15B	+	+	+	+	+	+	+	±	+	+	+	+	+	+	+	+
15C	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
17F	+	±	−	−	−	−	−	−	−	−	−	−	+	−	−	−
18F	+	+	+	+	+	−	−	−	−	−	−	−	+	−	−	−
18A	+	+	+	+	+	−	−	−	−	−	−	−	+	−	−	−
19F	+	+	+	+	+	−	−	−	−	−	−	−	+	−	−	−
19A	+	±	−	−	−	−	−	−	−	−	−	−	+	−	+[a]	−
20	+	+	+	+	−	−	−	−	−	−	−	−	+	−	−	−
22F	+	±	−	−	−	−	−	−	−	−	−	−	+	−	−	−
27	±	±	−	−	−	−	−	−	−	−	−	−	+	−	−	−
31	±	±	−	−	−	−	−	−	−	−	−	−	−	−	−	−
33F	+	±	−	−	−	−	−	−	−	−	−	−	+	−	−	−
34	+	+	−	−	−	−	−	−	−	−	−	−	+	−	−	−

[a]Present in one of two distinct putative CP structures of type 19A depending on culture conditions (*3*).

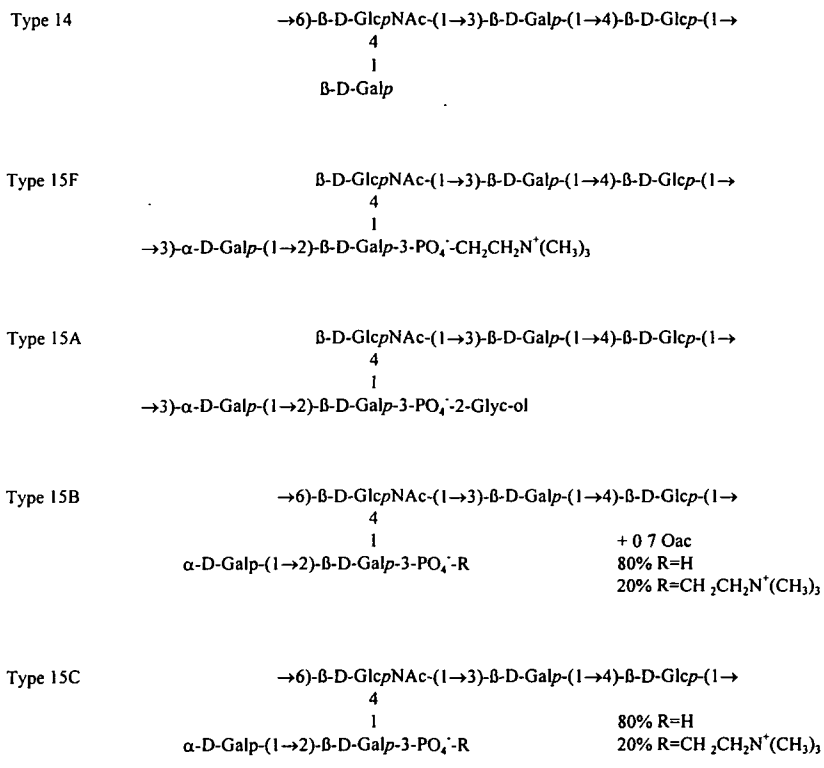

Fig. 2. **Repeating structures of the capsular polysaccharides of *S. pneumoniae* serotype 14, type 15F, 15A, 15B, and 15C.**

TABLE II. **Competence induction in pneumococcal strains of different serogroups and genetic transformation with a *cps14E* knock-out construct (pKOM03) by synthetic peptide pheromones (CSP-1/CSP-2).**

Capsule type	Competence induction by CSP-1/CSP-2	Hybridization with *cps14E*	Transformable with pKOM03	Unencapsulated *cps14E* mutants by double crossing-over
9N	+	+	+	+
13	+	+	+	+
14	+	+	+	+
15F	+	−	−	−
15B	+	+	+	+
15C	+	+	−	−
18A	−	+	−	−
18F	−	+	−	−
19F	−	+	−	−

It was shown previously that an ORF, designated *orfX*, located immediately downstream *cps14L*, is not involved in capsule biosynthesis in serotype 14 (*12*). The entire reading frame of *orfX* hybridized with several chromosomal *Eco*RI fragments of different sizes, of almost all pneumococcal types tested (data not shown). This suggested that the 150 bp long IS*1167* sequence at the 3′ end of *orfX* (*12*) was involved in the hybridization reactions. Therefore, also an internal fragment of *orfX*, lacking the IS*1167* sequence, was used as a probe. This fragment hybridized with type 14 DNA only, indicating that the actual *orfX* ORF is present in the serotype 14 strain exclusively (data not shown).

CpsE Mutants of the Serotypes 9N, 13, and 15B Lack Glucosyltransferase Activity—We have previously shown that a *cps14E* mutant of *S. pneumoniae* serotype 14 lacks glucosyltransferase activity (*10, 11*). In this study, we tried to obtain *cpsE* mutants of the strains belonging to the serotypes 9N, 13, 15B, 15C, 18F, 18A, and 19F, which all hybridized with the *cps14E* probe. Plasmid pKOM03, which contains *cps14E* with a *tetM* insertion mutation was used to transform the pneumococcal strains. Since pKOM03 is unable to replicate in *S. pneumoniae*, tetracycline-resistant mutants are the result of a homologous recombination event directed by the flanking *cps14E* sequences of *tetM* in pKOM03. Transformation of these encapsulated strains can only be successful when cells are competent for genetic transformation, and when there is sufficient sequence homology between *cps14E* of pKOM03 and target sequences on the chromosome of the recipients. Only the serotypes 9N, 13, 15B, and 15C responded to the competence inducing peptides CSP-1 and CSP-2 (Table II), and were transformed to novobiocin resistance with 5MC DNA. These data agree well with those of Pozzi *et al.* (*19*) who found that approximately 48% of 42 encapsulated strains of different serotypes became competent after addition of either CSP-1 or CSP-2. Transformation with pKOM03 yielded only tetracycline resistant mutants of the serotypes 9N, 13, and 15B (Table II). Type 15F, used as a negative control, became also competent after the addition of the competence peptides, but it lacks sequences homol-

ogous to *cps14E* and could not be transformed with pKOM03 (Table II). It is not clear why we did not obtain any type 15C mutants with pKOM03, since this strain could be induced to competence and DNA of type 15C showed a strong hybridization signal with the *cps14E* probe. Southern hybridization analysis of the type 9N, 13, and 15B mutants revealed that a double cross-over event led to the integration of *tetM* into a DNA fragment on the pneumococcal chromosome, which hybridized with the *cps14E* probe (data not shown). These mutants were all unencapsulated as judged by agglutination with Hasp8, a C-polysaccharide-specific monoclonal antibody, which reacts with unencapsulated pneumococcal mutants but not with encapsulated wild-type strains (*10*). Membranes of wild-type bacteria and the *cpsE* mutants of these serotypes were prepared and incubated with radioactively labeled UDP-glucose. The *cpsE* mutants of the serotypes 9N, 13, and 15B all showed reduced glycosyltransferase activity of approximately 25% of the activity of the wild type strains (Table III), as previously observed for the *cps14E* mutant of *S. pneumoniae* serotype 14 (*10*). In addition, analysis of the intermediates formed in the wild-type membranes showed that the subunit synthesis in type 9N, 13, and 15B starts with the addition of a glucose residue to a lipid carrier (see below). These observations also suggest that these three pneumococcal types possess a *cps14E* homologue encoding the glucosyl-1-phosphate transferase activity, required for the addition of glucose to a lipid carrier.

Synthesis of Capsule-Specific Intermediates in Several Pneumococcal Serotypes—We have previously shown that membrane fractions of *S. pneumoniae* type 14 contain capsule-specific glycosyltransferase activity which can be measured in glycosyltransferase assays (*10, 11*). In these assays, intermediates in the synthesis of the oligosaccharide subunit are formed on a lipid carrier. After extraction and mild acid hydrolysis, these intermediates can be analyzed by thin layer chromatography (TLC) (*11*). In this study, membrane fractions of all pneumococcal serotypes used in the hybridization experiments, were incubated with ^{14}C-labeled UDP-Glc or ^{14}C-labeled UDP-Gal. The formed ^{14}C-labeled intermediates were characterized by TLC (Fig. 3). An UDP-glucose-4-epimerase activity, which is present in this system, converts UDP-Glc to UDP-Gal, and *vice versa*. This explains why similar intermediates were formed when either labeled UDP-Glc (Fig. 3A) or UDP-Gal (Fig. 3B) was added to the pneumococcal membranes. All types tested, except type 1, 4, 10A, 12F, and 31 contain glucose in the backbone of their CP structure (*3*). We observed glucosyltransferase activity in all these glucose containing serotypes, except in serotype 3 and 11A. In

TABLE III. **Incorporation of [^{14}C]glucose by pneumococcal membranes into glycolipids.**

Membrane preparation	Incorporation (cpm/μg protein)
Type 9N wild-type	140
Type 9N *cpsE* mutant	47
Type 13 wild-type	346
Type 13 *cpsE* mutant	97
Type 14 wild-type	592
Type 14 *cpsE* mutant	82
Type 15B wild-type	430
Type 15B *cpsE* mutant	64

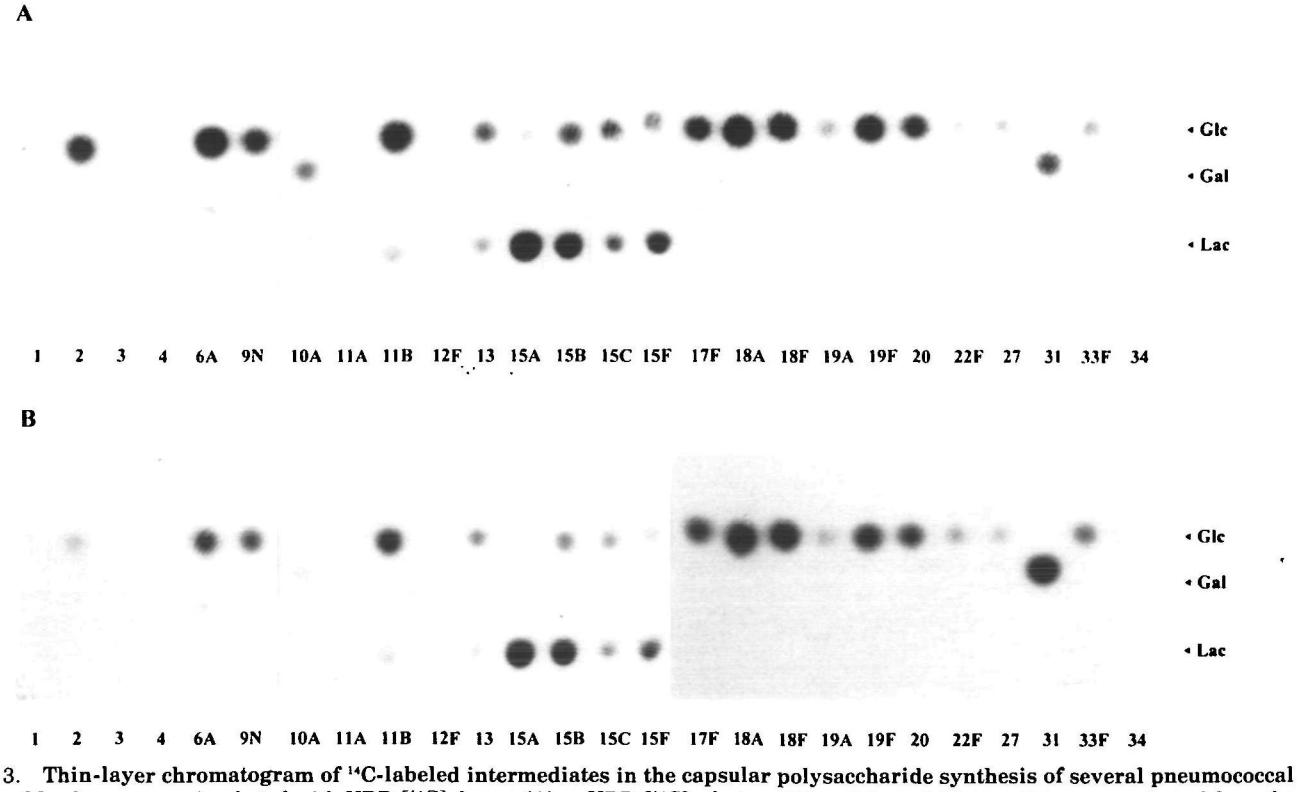

Fig. 3. **Thin-layer chromatogram of ^{14}C-labeled intermediates in the capsular polysaccharide synthesis of several pneumococcal types.** Membranes were incubated with UDP-[^{14}C]glucose (A) or UDP-[^{14}C]galactose (B), and formed intermediates were released from the lipid carriers by mild acid hydrolysis. Glc, glucose; Gal, galactose; Lac, lactose.

addition, as expected no glycosyltransferase activity was detected in type 1, 4, and 12F, but galactosyltransferase activity was observed in the serotypes 10A and 31. These two serotypes lack glucose, but do contain a galactose residue in their CP structure (*3*). These data show that in most pneumococcal serotypes, the CP biosynthesis starts with the addition of glucose to a lipid carrier. In the serotypes 10A and 31, the synthesis of the repeating unit starts with the linkage of a galactose residue to the lipid carrier. As observed in serotype 14 (*11*), lipid-linked lactose is formed in membrane fractions of type 11B, 13, 15F, 15A, 15B, and 15C. In fact, these types do have Galβ(1→4)Glc as part of their repeating unit structure, and thus in these serotypes galactose is coupled to lipid-linked glucose in the second step of the CP biosynthesis. Type 6A shows a faint spot migrating between galactose and lactose (Fig. 3). This product is probably Galα(1→3)Glc, which is part of the type 6A CP structure.

DISCUSSION

To examine the presence of *cps14*-like genes in other pneumococcal serotypes, we performed cross-hybridization experiments in which individual *cps14* genes were used to probe Southern blots of digested chromosomal DNA of pneumococci belonging to 26 different capsule types. Similar experiments were performed by Morona *et al.* (*15*) with the *cps19f* genes of *S. pneumoniae* type 19F. It is important to emphasize that the presence of homologous sequences not necessarily implies the existence of similar genes. A homologous gene in another serotype may not be expressed or may encode a different enzyme activity (*e.g.*, homologous glycosyltransferases may have different substrate specificities). The first case is illustrated by the observation that the *cps3* (*cap3*) locus of *S. pneumoniae* type 3 contains sequences with extensive homology to *cps19fA*/*cps14A* and *cps19fB*/*cps14B*, however no open reading frames could be detected in this region due to two separate frame-shifts (*4*). In addition, Arrecubieta *et al.* (*8*) reported that insertion-duplication mutagenesis of this DNA region did not affect encapsulation of the type 3 strain. On the other hand, these cross-hybridizations revealed that the *cpsE* gene of serotype 14 has a homologue in type 19F, and *vice versa*. Sequence analysis of the *cps14* and *cps19f* gene clusters showed that the *cps14E* and *cps19fE* genes are almost identical (*10*). In addition, functional characterization showed that both genes encode a glucosyl-1-phosphate transferase, an enzyme required for the first step in the biosynthesis of the type-specific repeating unit (*11, 15*). This example shows that these hybridization experiments can give useful clues about the CP biosynthesis in other pneumococcal serotypes.

Morona *et al.* (*15*) showed that the region upstream of *cps19fA* hybridized with DNA of 20 different pneumococcal types tested. Sequence data revealed that this conserved region is also present in front of the *cps14* locus. However, the region downstream *cps14L*, which is the last gene of the *cps14* locus, contains an ORF (*orfX*) which is not present downstream the *cps* loci of type 19F, 1, or 3 (*12*). In addition, cross-hybridizations showed that *orfX* is only present in serotype 14, and not in all other types used in this study. Furthermore, we previously presented data indicating that *orfX*, which is homologous to the teichoic acid synthesis genes *tagB* and *tagF* of *Bacillus subtilis*, is not involved in the synthesis of the species-specific (lipo-) teichoic acids in *S. pneumoniae* (*12*). *orfX* may be a nonfunctional relic of the capsule gene cluster of the ancestor of the type 14 strain(s). Alternatively, *orfX* itself might have been transferred to the ancestral type 14 strain by an exchange of polysaccharide biosynthesis genes, but subsequently lost its function.

The *cps14A* to *cps14E* genes of serotype 14 are almost identical to *cps19fA* to *cps19fE* of *S. pneumoniae* type 19F, respectively (*11, 12*). We did not find important differences in the hybridization data obtained with *cps14A-E* and *cps19fA-E*, between strains belonging to the same serogroups. *Cps14A* to *cps14D* encode non type-specific functions such as regulation, chain length determination and export of CP. These genes seem to be conserved in many pneumococcal serotypes.

Glycosyltransferase activity assays showed that among 21 pneumococcal types which contain a glucose residue in their CP core structure, 19 serotypes express glucosyl-1-phosphate transferase activity. In these serotypes, the synthesis of the repeating unit starts with the addition of glucose to a lipid carrier molecule. Serotype 3 did not not show glucosyl-1-phosphate transferase activity, although it has a glucose residue in its CP structure. The polysaccharide synthesis in *S. pneumoniae* type 3 seems to proceed by successive monomer addition, rather then by polymerization of lipid-linked intermediates (*7, 9*). The simple structure of type 3 CP allows this kind of capsule biosynthesis, which resembles for instance the synthesis of the *E. coli* K1 and K5 antigens (*20*). However, this mechanism is probably quite unique in *S. pneumoniae*, since most pneumococcal serotypes synthesize a more complex CP and need to co-ordinate sequential assembly of repeat units. This is confirmed by our observation that most pneumococcal types do form lipid-linked intermediates. It is not clear why we did not detect any capsular intermediates in membranes of type 11A, since the type 11A CP core only contains glucose and galactose. Since type 11A has a rather complex CP structure, it is unlikely that the CP synthesis proceeds by successive monomer addition as in type 3. Perhaps the expression level, or the activity itself, of the type 11A glycosyltransferases is much lower compared with that in other pneumococcal types.

Of the 19 subtypes with glucosyl-1-phosphate transferase activity, only seven (type 9N, 13, 15B, 15C, 18A, 18F, and 19F) hybridized strongly with the type 14 glucosyl-1-phosphate transferase gene *cps14E*. When very low stringent hybridization and washing conditions were used, most of the other 12 types with glucosyltransferase activity showed some weak hybridizing bands (data not shown). These data indicate that pneumococci have glucosyl-1-phosphate transferase genes which are distinct from *cps14E*, although they may still have some sequence similarity with *cps14E*. These genes may have diverged from an ancestral *cps14E* gene, or may have been acquired from another origin by genetic exchange.

The subunit synthesis in type 10A and 31 starts with the addition of galactose to a lipid carrier (Fig. 3), therefore these types should express a galactosyl-1-phosphate transferase. Furthermore, the *cps14G* homologues observed in type 11B, 13, 15F, 15A, 15B, and 15C probably encode the β-1,4-galactosyltransferase activity expressed in mem-

brane fractions of these strains.

Within serogroup 15, type 15B and 15C hybridized with all the *cps14* probes. The CP structures of these types (Fig. 2) indicate that they should have a number of additional enzyme activities not found in serotype 14. Type 15F and 15A also hybridized with most *cps14* genes, but not all. The different CP core structure of type 15F and 15A compared with that of type 14, 15B, and 15C necessitates an alteration in the specificity of the polysaccharide polymerase. This probably explains why sequences homologous to *cps14H* (the putative CP polymerase) were only observed in type 15B and 15C, and not in type 15F and 15A. Remarkably, type 15F and 15A also lack *cps14D* and *cps14E* homologues. Since type 15F and 15A showed glucosyltransferase activity, these types should also express a glucosyl-1-phosphate transferase which is different from Cps14E.

The organization of the *cps14* locus is similar to that of the gene clusters in other Gram-positive bacteria encoding CP biosynthesis in *Staphylococcus aureus* (*21*) and exopolysaccharide synthesis in *Streptococcus thermophilus* (*22*) and *Lactococcus lactis* (*23*). The order of the functions of the genes in these polysaccharide-synthesis loci seems to be: regulation, chain length regulation/export, biosynthesis, and polymerization of the repeating units, export. The same order of functions is found in the *cps* loci of *S. pneumoniae* type 1 and 19F. However, these pneumococcal types contain some additional *cps* genes at the 3′ end of the *cps* loci, whose gene products are involved in the synthesis of nucleotide sugar precursors (*6, 15*). The genetic organization of the *cps14* gene cluster also resembles the general organization observed in *cps* loci of many Gram-negative bacteria such as *Escherichia coli* and *Haemophilus influenzae* (*20, 24*): a central serotype-specific region (*cps14E* to *cps14J*) encoding the type 14 glycosyltransferases and the CP subunit polymerase is flanked on both sides by two regions (*cps14A* to *cps14D*; *cps14L*) presumed to encode proteins for more common functions, such as regulation and transport (Fig. 1). In general, both flanking regions are conserved in these Gram-negative bacteria. However, only *cps14A-D* is more or less conserved in *S. pneumoniae*, the genes at the 3′ part of the *cps* loci of *S. pneumoniae* are not conserved and encode enzymes with different functions.

By using a *cps14E* knock-out construct, we were able to transform the pneumococcal types 9N, 13, and 15B to unencapsulation. These type 9N, 13, and 15B mutants, in which a tetracycline resistance cassette (*tetM*) was integrated in a *cps14E* homologue, showed reduced glucosyltransferase activity. Although we cannot exclude the possibility that the *tetM* insertion causes polar effects on downstream genes, possibly encoding the glucosyl-1-phosphate transferase, it is most likely that disruption of these *cpsE* homologues itself led to reduced glucosyl-1-phosphate transferase activity in these mutants. These data support our earlier suggestion that the pneumococcal types 9N, 13, and 15B express a glucosyl-1-phosphate transferase which is homologous to Cps14E.

Several genetic events may have contributed to the capsular polysaccharide diversity in *S. pneumoniae*. Since the pneumococcus is naturally transformable, capsule type changes may occur as a consequence of transformation with donor DNA of a different capsule type. Homologous recombination between conserved sequences that flank a serotype-specific region serves as a likely model (*7*). The type-specific gene cassette might become modified by point mutations, insertions, or by gene duplication and divergent evolution as suggested for the glycosyltransferase genes *cps14I* and *cps14J* (*12*). The presence of insertion elements or vestiges of these elements at the extremities of the *cps* loci of type 1, 3, 14, and 19F suggest that transposition-like events might have been involved in the exchange of *cps* genes (*6, 12*). Moreover, insertion sequences not only provide multiple sites of homology for other recombination systems, active elements can also create gene disruption mutations, regulatory mutations and large rearrangements within the chromosome. These events might also have contributed to variability in CP expression. Further genetical and functional analysis of *cps* loci of other pneumococcal serotypes may provide answers to the question how capsule diversity in *S. pneumoniae* has been generated.

REFERENCES

1. Cross, A.S. (1990) The biologic significance of bacterial encapsulation. *Curr. Top. Microbiol. Immunol.* **150**, 87-95
2. Henrichsen, J. (1995) Six newly recognized types of *Streptococcus pneumoniae*. *J. Clin. Microbiol.* **33**, 2759-2762
3. Van Dam, J.E.G., Fleer, A., and Snippe, H. (1990) Immunogenicity and immunochemistry of *Streptococcus pneumoniae* capsular polysaccharides. *Antonie Leeuwenhoek* **58**, 1-47
4. García, E. and López, R. (1997) Molecular biology of the capsular genes of *Streptococcus pneumoniae*. *FEMS Microbiol. Lett.* **149**, 1-10
5. Dowson, M.H. (1930) The transformation of pneumococcal types. II. The interconvertibility of type-specific *S. pneumococci*. *J. Exp. Med.* **51**, 123-147
6. Muñoz, R., Mollerach, M., López, R., and García, E. (1997) Molecular organization of the genes required for the synthesis of type 1 capsular polysaccharide of *Streptococcus pneumoniae*: formation of binary encapsulated pneumococci and identification of cryptic dTDP-rhamnose biosynthesis genes. *Mol. Microbiol.* **25**, 79-92
7. Dillard, J.P., Vandersea, M.W., and Yother, J. (1995) Characterization of the cassette containing genes for type 3 capsular polysaccharide biosynthesis in *Streptococcus pneumoniae*. *J. Exp. Med.* **181**, 973-983
8. Arrecubieta, C., García, E., and López, R. (1995) Sequence and transcriptional analysis of a DNA region involved in the production of capsular polysaccharide in *Streptococcus pneumoniae* type 3. *Gene* **167**, 1-7
9. Arrecubieta, C., López, R., and García, E. (1996) Type 3-specific synthase of *Streptococcus pneumoniae* (Cap3B) directs type 3 polysaccharide biosynthesis in *Escherichia coli* and in pneumococcal strains of different serotypes. *J. Exp. Med.* **184**, 449-455
10. Kolkman, M.A.B., Morrison, D.A., van der Zeijst, B.A.M., and Nuijten, P.J.M. (1996) The capsule polysaccharide synthesis locus of *Streptococcus pneumoniae* serotype 14: identification of the glycosyltransferase gene *cps14E*. *J. Bacteriol.* **178**, 3736-3741
11. Kolkman, M.A.B., van der Zeijst, B.A.M., and Nuijten, P.J.M. (1997) Functional analysis of glycosyltransferases encoded by the capsular polysaccharide locus of *Streptococcus pneumoniae* serotype 14. *J. Biol. Chem.* **272**, 19502-19508
12. Kolkman, M.A.B., Wakarchuk, W., Nuijten, P.J.M., and van der Zeijst, B.A.M. (1997) Capsular polysaccharide synthesis in *Streptococcus pneumoniae* serotype 14: molecular analysis of the complete *cps* locus and identification of genes encoding glycosyltransferases required for the biosynthesis of the tetrasaccharide subunit. *Mol. Microbiol.* **26**, 197-208
13. Morona, J.K., Morona, R., and Paton, J.C. (1997) Molecular and genetic characterization of the capsule biosynthesis locus of

Streptococcus pneumoniae type 19B. *J. Bacteriol.* **179**, 4953-4958
14. Guidolin, A., Morona, J.K., Morona, R., Hansman, D., and Paton, J.C. (1994) Nucleotide sequence analysis of genes essential for capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* type 19F. *Infect. Immun.* **62**, 5384-5396
15. Morona, J.K., Morona, R., and Paton, J.C. (1997) Characterization of the locus encoding the *Streptococcus pneumoniae* type 19F capsular polysaccharide biosynthetic pathway. *Mol. Microbiol.* **23**, 751-763
16. Morrison, D.A., Lacks, D.A., Guild, W.R., and Hageman, J.M. (1983) Isolation and characterization of three new classes of transformation-deficient mutants of *Streptococcus pneumoniae* that are defective in DNA transport and genetic recombination. *J. Bacteriol.* **156**, 281-290
17. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
18. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Smith, J.A., Seidman, J.G., and Struhl, K. (eds.) (1987) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York
19. Pozzi, G., Masala, L., Iannelli, F., Manganelli, R., Håvarstein, L.S., Piccoli, L., Simon, D., and Morrison, D.A. (1996) Competence for genetic transformation in encapsulated strains of *Streptococcus pneumoniae*: two allelic variants of the peptide pheromone. *J. Bacteriol.* **178**, 6087-6090
20. Roberts, I.S. (1996) The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annu. Rev. Microbiol.* **50**, 285-315
21. Lin, W.S., Cunneen, T., and Lee, C.Y. (1994) Sequence analysis and molecular characterization of genes required for the biosynthesis of type 1 capsular polysaccharide in *Staphylococcus aureus*. *J. Bacteriol.* **176**, 7005-7016
22. Stingele, F., Neeser, J.-R., and Mollet, B. (1996) Identification and characterization of the *eps* (exopolysaccharide) gene cluster from *Streptococcus thermophilus* Sfi6. *J. Bacteriol.* **178**, 1680-1690
23. Van Kranenburg, R., Marugg, J.D., van Swam, I.I., Willem, N.J., and de Vos, W.M. (1997) Molecular characterization of the plasmid-encoded eps gene cluster essential for exopolysaccharide biosynthesis in *Lactococcus lactis*. *Mol. Microbiol.* **24**, 387-397
24. Boulnois, G.J. and Jann, K. (1989) Bacterial polysaccharide synthesis, export and evolution of structural diversity. *Mol. Microbiol.* **3**, 1819-1823