

Genetic Basis for the Structural Difference between *Streptococcus pneumoniae* Serotype 15B and 15C Capsular Polysaccharides

Saskia van Selm,¹ Lisette M. van Cann,¹ Marc A. B. Kolkman,²
Bernard A. M. van der Zeijst,³ and Jos P. M. van Putten^{1*}

Department of Infectious Diseases and Immunology, Utrecht University, 3584 CL Utrecht,¹ Genencor International B.V., 2300 AE Leiden,² and Netherlands Vaccine Institute, 3720 AL Bilthoven,³ The Netherlands

Received 7 April 2003/Returned for modification 26 June 2003/Accepted 5 August 2003

In a search for the genetic basis for the structural difference between the related *Streptococcus pneumoniae* capsular serotypes 15B and 15C and for the reported reversible switching between these serotypes, the corresponding capsular polysaccharide synthesis (*cps*) loci were investigated by keeping in mind that at the structural level, the capsules differ only in O acetylation. The *cps* locus of a serotype 15B strain was identified, partially PCR amplified with primers based on the related serotype 14 sequence, and sequenced. Sequence analysis revealed, among other open reading frames, an intact open reading frame (designated *cps15bM*) whose product, at the protein level, exhibited characteristics of previously identified acetyltransferases. Genetic analysis of the corresponding region in a serotype 15C strain indicated that the same gene was present but had a premature stop in translation. Closer analysis indicated that the serotype 15B gene contained a short tandem TA repeat consisting of eight TA units. In serotype 15C, this gene contained nine TA units that resulted in a frameshift and a truncated product. Genetic analysis of 17 serotype 15B and 15C clinical isolates revealed a perfect correlation between the serotype and the length of the short tandem repeat in the putative O-acetyltransferase gene. The number of TA repeating units varied between seven and nine in the various isolates. Together, the data strongly suggest that the structural difference between serotypes 15B and 15C is based on variation in the short tandem TA repeat in the O-acetyltransferase gene and that the transition between serotypes is due to slipped-strand mispairing with deletion or insertion of TA units in the *cps15bM* gene.

The human pathogen *Streptococcus pneumoniae* (pneumococcus) is a major cause of respiratory tract infections, bacteremia, and meningitis, particularly in young children and the elderly (32, 35). One of the prime virulence determinants of this bacterium is the polysaccharide capsule (CPS). This structure is thought to protect the bacterium against harmful environmental conditions and exhibits antiphagocytic properties (8). The CPS is composed of saccharide repeating units that are polymerized into a polysaccharide chain. Thus far, 90 different capsule serotypes have been identified. The diversity is based on variation in the carbohydrate structure of the oligosaccharide units or the attached side groups (18).

The genes encoding the enzymes involved in CPS biosynthesis are clustered on the bacterial genome in the capsular polysaccharide synthesis (*cps*) locus. The *cps* locus is typically flanked by the genes *dexB* and *aliA*. At this time, the *cps* loci of 16 different serotypes have been sequenced (3, 12, 13, 16, 20, 23, 24, 26–29, 33, 34, 42, 50). Nearly all loci have the same genetic organization (38). The first four genes of the loci are conserved in almost all serotypes, and it has been demonstrated that three of these genes encode enzymes involved in the regulation of capsule production (5, 7, 30, 54). The central parts of the loci contain serotype-specific genes that encode glycosyltransferases and the repeating unit transporter and polymerase. Near the 3' end, the loci may include genes that

encode transferases that allow substitution of side groups and/or enzymes that enable synthesis of specific substrates.

In general, the structure of the CPS appears to be a stable characteristic of a strain, although genetic exchange between loci can lead to serotype changes (4, 9–11, 37, 41). For serotypes 15B and 15C (Fig. 1 shows chemical structures) a change of serotype has been suspected to occur in vivo during natural infection as both of these serotypes have been repeatedly isolated from middle ears of children with otitis media (51). The switching of serotypes in vitro was demonstrated to be a reversible event with an estimated transition frequency of up to 1 in 250 (51).

In the present study, we aimed to determine the genetic basis of the structural difference between capsule serotypes 15B and 15C and the mechanism that drives the transition between these serotypes. Extensive structural analysis has indicated that the principal difference between serotypes 15B and 15C is the presence or absence of an O-acetyl group attached to the pentasaccharide repeating unit (14). To identify a putative O-acetyltransferase gene(s), we genetically analyzed the serotype 15B *cps* locus, which has not been characterized previously. Our search strategy involved identification and comparison of disparate regions in the *cps* loci of serotype 15B and the structurally related (but non-O-acetylated) serotype 14 (20) (Fig. 1). After this, we compared the relevant regions in serotypes 15B and 15C and determined the genetic basis for the difference between serotypes 15B and 15C and for the switching between these serotypes.

* Corresponding author. Mailing address: Department of Infectious Diseases and Immunology, Utrecht University, Yalelaan 1, 3584 CL Utrecht, The Netherlands. Phone: (31) 30 2534888. Fax: (31) 30 2540784. E-mail: j.vanputten@vet.uu.nl.

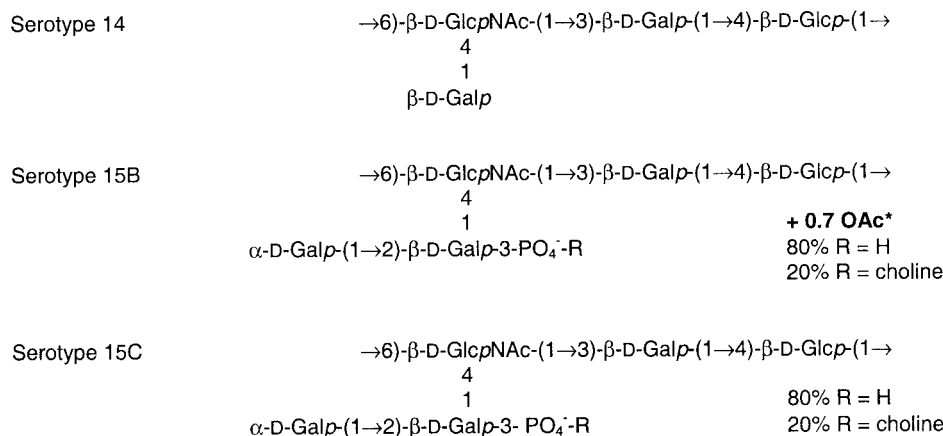


FIG. 1. Diagram showing the structures of the repeating units of the capsular polysaccharides of *S. pneumoniae* serotypes 14, 15B, and 15C (14, 22). + 0.7 OAc* indicates that on average 70% of the repeating units is O acetylated. The site of the O acetylation is unknown (14), although Venkateswaran et al. (51) suggested that it is linked to a galactose moiety.

MATERIALS AND METHODS

Bacterial strains. *S. pneumoniae* strain SSISP 15B/1 was generously provided by U. B. Skov Sørensen (University of Aarhus, Aarhus, Denmark). All other strains were clinical isolates and were kindly provided by A. van der Ende (Academic Medical Center, Amsterdam, The Netherlands). These strains were collected over a 9-year period and were isolated from different patients in different regions of The Netherlands. Pneumococci were grown in Todd-Hewitt broth (Oxoid) supplemented with 0.5% yeast extract (Oxoid) or on Columbia blood agar (prepared with 5% sheep blood) in the presence of 5% CO₂. Serotypes were determined by using the quellung reaction with pneumococcal antisera from the Statens Serum Institut (Copenhagen, Denmark).

DNA techniques. Most DNA techniques were performed as described previously (44). Chromosomal DNA was isolated with a Qiagen genomic DNA kit (Qiagen GmbH, Hilden, Germany). Long-range PCR was performed with Elongase (Invitrogen BV, Breda, The Netherlands) by following to the manufacturer's instructions. PCR products were purified with a Qiagen PCR purification kit or from agarose gels with a Qiaquick gel extraction kit (Qiagen).

PCR amplification of part of the serotype 15B cps locus. Primers BamII (5'-AGGAAAGGATCCGATGATAGAAGTATCA-3') and aliAstI (5'-CAAA TAGTTGAGGTTATCAGGGTCTGTCTC-3') were used to PCR amplify the C-terminal region of the serotype 15B cps locus. Primers 15BJF-1 (5'-AACGA CAGTTTAGAAAGTTCGC-3') and aliAstI were used to generate PCR products from serotype 15B chromosomal DNA. These PCR products were used for cycle sequencing by primer walking.

DNA sequencing and analysis. DNA sequencing was performed with an ABI Prism 310 genetic analyzer (Applied Biosystems Inc., Foster City, Calif.) by using

an ABI dye terminator cycle sequencing kit. DNA and protein data were analyzed with the software program DNASTar. The BLAST (2) and FASTA (39) algorithms were used to compare sequences with sequences available in the National Center for Biotechnology Information database. Transmembrane segment predictions were performed by using the program DAS (<http://au.expasy.org>). The putative O-acetyltransferase genes were amplified with primers 15BLR-1 (5'-ATTGATTTCTGCTATGTCTCCG-3') and 15BMR-4 (5'-AGCG GTAACATCAATTATGTCC-3'). Sequencing of the putative genes was done with primers 15BgenXF (5'-ATTTTGTAAATAGGTAGGAAAG-3') and 15BgenXR (5'-TTCTTCTTATCCGAACAGGC-3').

Nucleotide sequence accession number. The nucleotide sequence of part of the serotype 15B cps locus has been deposited in the GenBank database under accession number AY250187.

RESULTS

Genetic organization of part of the serotype 15B cps locus.

As a first step towards identification of the O-acetyltransferase gene(s) in serotype 15B, we identified parts of the cps locus that were not present in the related, non-O-acetylated serotype 14 (19). By analogy with the genetic organization of the serotype 14 locus (20), these parts were assumed to be located in the C-terminal region of the locus. Long-range PCR performed with chromosomal DNA of serotype 15B strain 910165

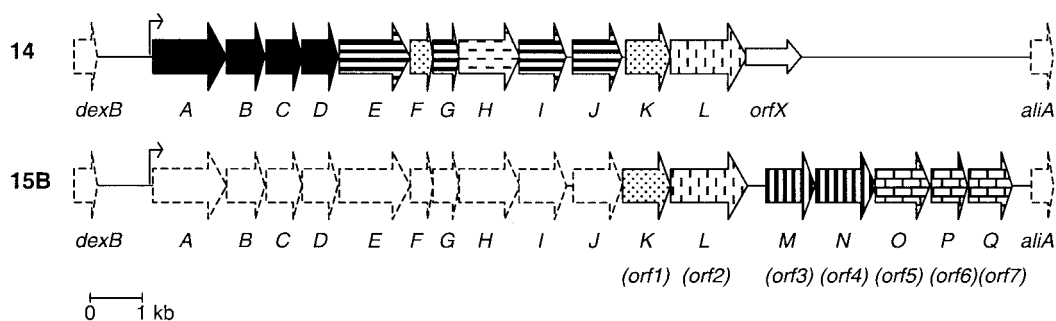


FIG. 2. Comparison of the organizations of the cps loci of *S. pneumoniae* serotypes 14 and 15B. The genes that are present in the loci are indicated by different arrows, as follows: solid arrows, 5' conserved genes; arrows with horizontal stripes, glycosyltransferases; arrow with horizontal dashes, polymerase; arrows with vertical dashes, repeating unit transporters; arrows with vertical stripes, nonsugar transferases; arrows with tessellated pattern, nucleotide sugar biosynthesis genes; arrows with dots, genes with unknown functions. The dashed arrows represent genes that were previously suggested to be present in serotype 15B (19). The small open arrow is not involved in CPS synthesis in serotype 14. The putative promoters of the loci are indicated by bent arrows. The putative genes cps14A to -K and cps15A to -Q are labeled A to K and A to Q, respectively.

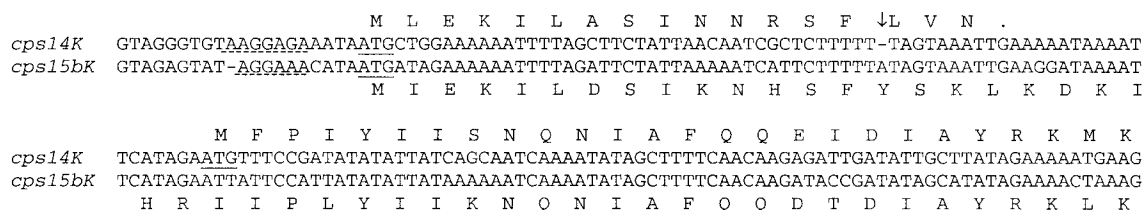


FIG. 3. Comparison of the nucleotide sequences at the 5' ends of the *cps14K* and *cps15bK* genes. The amino acid sequences encoded by the ORFs are shown above the sequence for *cps14K* and below the sequence for *cps15bK*. Start codons of the ORFs are underlined. Putative ribosome binding sites are indicated by dashed underlining. The arrow indicates the position of the deletion in serotype 14.

and primers based on the sequence of the *cps14I* gene and the flanking *ali4* gene yielded an approximately 9.5-kb DNA fragment. Sequencing of this fragment with a primer based on the sequence of the *cps14I* gene revealed a short sequence which served as a basis for further sequencing by primer walking. Computer-assisted sequence analysis of the entire fragment revealed seven open reading frames (ORFs), designated ORF1 to ORF7 (Fig. 2). There was a 365-bp intergenic gap between ORF2 and ORF3. The other intergenic regions were much smaller. All ORFs were preceded by a putative ribosome binding site. The G+C content of ORF3 was remarkably low (24%). The G+C contents of the other ORFs varied between 30 and 37% (Table 1).

Identification of putative glycerophosphate biosynthesis genes and a glycosyl transferase. Comparative sequence analysis of the corresponding regions of serotypes 15B and 14 revealed that at the deduced amino acid level Cps15bK (encoded by ORF1) was 88% similar to Cps14K (20), 60% similar to the putative glycosyltransferase WciW of serotype 18C (16), and 54% similar to an α -1,2-galactosyltransferase (EpsI) of *Lactobacillus delbruekii* subsp. *bulgaricus* (21). Alignment of Cps14K and Cps15bK genes revealed that the Cps14K gene had a single adenine deletion at position 44 of the serotype 15B sequence. This caused a frameshift and a premature stop of translation after 17 amino acids (Fig. 3). Translation of the ORF from the next start codon was considered unlikely as no ribosome binding site in close proximity could be identified (Fig. 3). This apparent defect in *cpsK* is of particular interest as its similarity to the EpsI gene of *L. delbruekii* suggests that it may encode a galactosyltransferase. The deletion may thus explain the difference between the carbohydrate compositions of serotypes 14 and 15B.

ORF2 was 89% similar at the deduced amino acid level to the serotype 14 Cps14L gene. This protein is assumed to be a putative repeating unit transporter (20). Homologues of ORF3

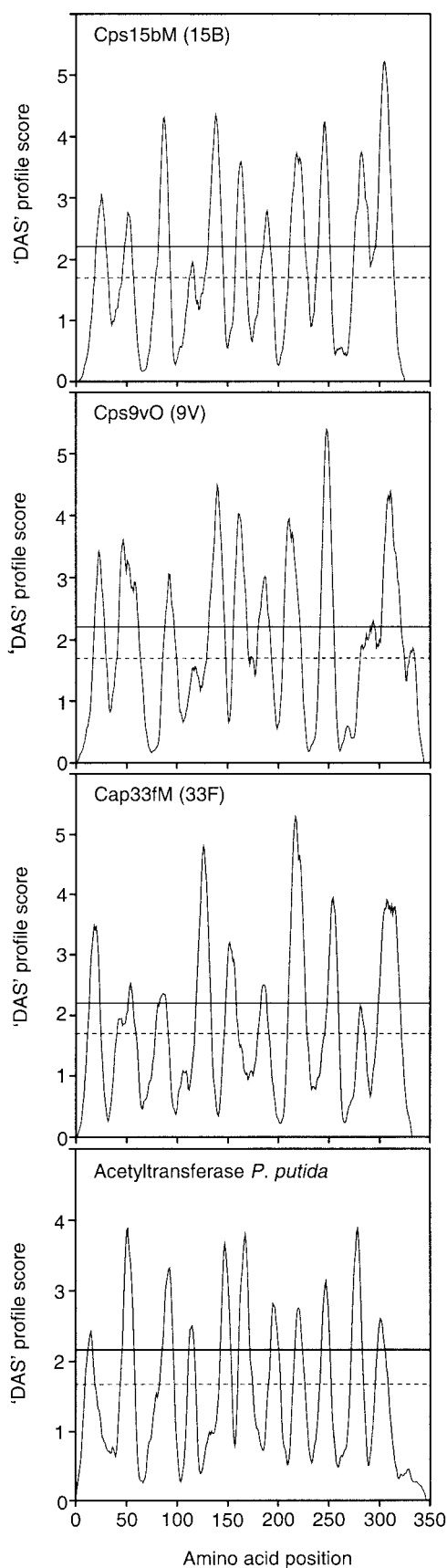
to ORF7 were not found in the *cps* locus of serotype 14. Functions for these genes were proposed on the basis of amino acid similarities (Table 1) with protein sequences available in the GenBank database. BLASTX analysis (2) indicated that Cps15bN (encoded by ORF4) was 91% similar to Cps23fK (corresponding to Cps23fW) of serotype 23F (26, 42) and 60% similar to WciY of serotype 18C (16). These proteins are putative glycerophosphotransferases. The proteins encoded by ORF5, ORF6, and ORF7 (Cps15bO, Cps15bP, Cps15bQ) were virtually identical to Cps23fL, Cps23fM, and Cps23fN (corresponding to Cps23fX, Cps23fY, and Cps23fZ) of serotype 23F (26, 42), respectively. These proteins are predicted to be involved in the biosynthesis of CDP-2-glycerol, the precursor required for addition of the glycerol 2-phosphate side chain in serotype 23F CPS.

Identification of a putative O-acetyltransferase gene in serotype 15B. The predicted protein encoded by ORF3 (Cps15bM) exhibited 39% similarity to an acetyltransferase of *Pseudomonas putida* (17) and 40% similarity to an integral membrane acetyltransferase of *Actinobacillus actinomycetemcomitans* (36). This protein also exhibited 69% similarity to WciX encoded by the *cps* locus of serotype 18C (16), 41% similarity to the putative acetyltransferase Cap33fM present in serotype 33F (24, 38), and 39% similarity to the putative O-acetyltransferase Cps9vO of serotype 9V (50). Transmembrane segment predictions for Cps15bM, WciX, Cap33fM, and Cps9vO and the putative acetyltransferases of *P. putida* and *A. actinomycetemcomitans* yielded very similar profiles (Fig. 4). These profiles resembled those of a series of proteins defined as a family of membrane proteins involved in the acetylation of carbohydrate moieties on extracytoplasmic molecules (46; data not shown). Based on the observed similarities we propose that Cps15bM is a putative O-acetyltransferase.

Comparative sequence analysis of the putative O-acetyltransferase genes in serotypes 15B and 15C. To assess the

TABLE 1. Properties of the additional ORFs in the *cps* locus of *S. pneumoniae* serotype 15B compared to the serotype 14 *cps* locus

ORF	Gene size (bp)	G+C content (%)	No. of amino acids	Predicted molecular mass (kDa)	Predicted localization of protein in membrane	Similar gene product (% similarity)	Proposed function of gene product
<i>cps15bK</i>	909	30	302	35.9	Yes	<i>S. pneumoniae</i> Cps14K (88)	α -1,2-Galactosyltransferase
<i>cps15bL</i>	1,464	30	487	55.3	Yes	<i>S. pneumoniae</i> Cps14L (89)	CPS repeating unit transporter
<i>cps15bM</i>	978	24	325	38.4	Yes	<i>S. pneumoniae</i> serotype 18C WciX (69)	O-Acetyltransferase
<i>cps15bN</i>	1,149	30	382	45.2	Yes	<i>S. pneumoniae</i> Cps23fK (91)	Glycerol-2-phosphotransferase
<i>cps15bO</i>	1,065	37	354	37.6	Yes	<i>S. pneumoniae</i> Cps23fL (100)	Glycerol-2-phosphate dehydrogenase
<i>cps15bP</i>	705	32	234	26.1	No	<i>S. pneumoniae</i> Cps23fM (99)	Glycerol-2-phosphate cytidyltransferase
<i>cps15bQ</i>	834	33	277	31.1	No	<i>S. pneumoniae</i> Cps23fN (100)	Glyceraldehyde-2-phosphotransferase



genetic basis for the difference between serotypes 15B and 15C, the *cps15cM* gene of serotype 15C strain 910162 was PCR amplified from genomic DNA by using primers based on the serotype 15B *cps15bM* sequence. This yielded a DNA fragment whose size was similar to the size of the *cps15bM* gene. Sequencing of the PCR product and comparative sequence analysis revealed that the sequences of the *cps15bM* and *cps15cM* genes were identical, except for the apparent insertion of two nucleotides (TA) in the central region of the serotype 15C gene. These nucleotides were part of a stretch of repetitive TA nucleotides (Fig. 5). The *cps15bM* gene of serotype 15B contained eight repetitive TA units, which resulted in an intact ORF. In serotype 15C, the apparent insertion of one TA unit interrupted the ORF (Fig. 5). This caused a premature stop in translation and most likely resulted in a loss of enzyme activity, which in turn may explain the absence of *O*-acetyl groups in the CPS of serotype 15C.

Correlation between serotypes 15B and 15C and the length of the TA repeat in the putative *O*-acetyltransferase gene. The apparent relationship between the status of the putative *O*-acetyltransferase gene and the serotype 15B and 15C capsule phenotypes was further investigated by assessing the lengths of the short tandem TA repeats in seven serotype 15B and seven serotype 15C clinical isolates collected from different patients in different regions of The Netherlands over a 9-year period and in the serotype 15B reference strain SSISP 15B/1 (Fig. 5). The putative *O*-acetyltransferase genes were amplified by PCR from these strains, and the lengths of the TA repeats were determined by sequencing. As shown in Fig. 5, the number of repetitive TA units in the putative *O*-acetyltransferase genes of the various strains ranged from seven to nine. All serotype 15B strains had the same sequence with eight repetitive TA units, which resulted in an intact ORF. For the serotype 15C isolates, six of seven strains were identical except for the presence of either seven or nine TA units, which resulted in a nonfunctional gene. Unexpectedly, strain 981519, which had been identified as a serotype 15C strain, contained eight repetitive TA units, which is typical of serotype 15B strains. Retyping of this strain, however, demonstrated that it had a serotype 15B capsule and thus that either the serotype identification had been erroneous or the strain had switched to serotype 15B after the initial typing. These data strongly suggest that the number of TA units in the *cps15bM/cps15cM* gene determines whether a strain is serotype 15B or 15C and that the reported switching of these serotypes (51) may have been caused by variation in the number of TA repeats.

FIG. 4. Transmembrane segment predictions for putative *O*-acetyltransferases. The predicted profiles were generated for *S. pneumoniae* Cps15bM (serotype 15B), Cps9vO (serotype 9V), Cap33fM (serotype 33F), and an acetyltransferase of *P. putida* by using the program DAS (<http://au.expasy.org>). There are two cutoffs indicated on the plot. The strict cutoff (represented by the solid line) at a DAS score of 2.2 is informative in terms of the number of matching segments. The loose cutoff (represented by the dashed line) gives the actual location of the transmembrane segment.

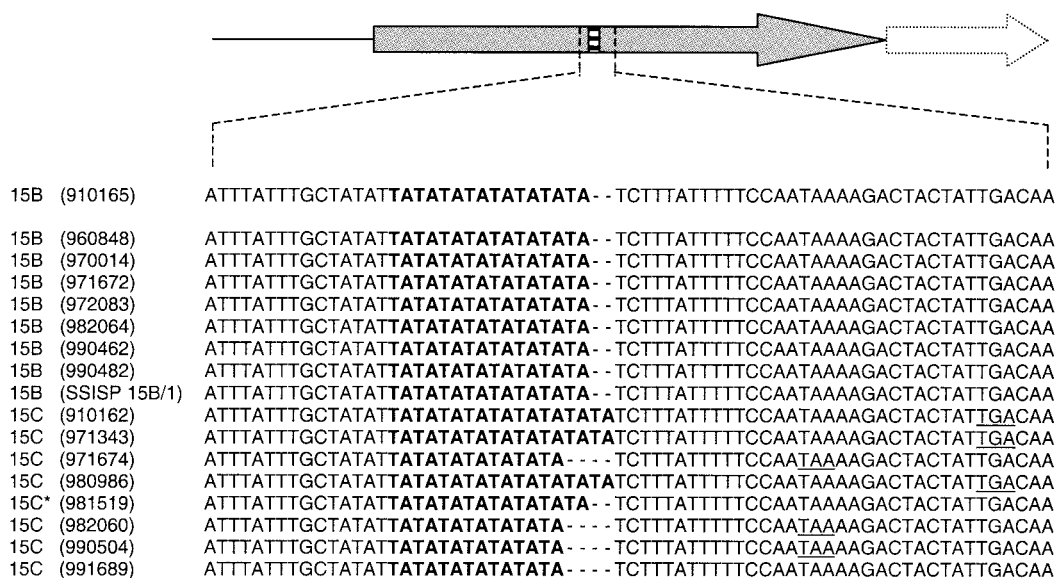


FIG. 5. Schematic representation of the putative *O*-acetyltransferase genes, showing the locations of the TA repeats and the variation in the TA repeat length in nine serotype 15B and eight serotype 15C clinical isolates (positions 398 to 463 in the gene). The strain numbers are indicated in parentheses. Strain 981519 was identified by us as a serotype 15B strain. The premature stops in the *cps15cM* ORFs of serotype 15C strains are underlined.

DISCUSSION

S. pneumoniae capsule serotypes 15B and 15C differ in terms of the presence of an *O*-acetyl group in the structure of the CPS. Here we report that the genetic basis for the difference in *O* acetylation is located in the *cps15bM/cps15cM* gene. This gene is predicted to encode an *O*-acetyltransferase. In serotype 15B this gene is intact, while in serotype 15C the gene is defective due to insertion or deletion of a TA unit in a short tandem TA repeat sequence that is present in the central region of the gene. The perfect correlation between the length of the TA repeat and serotype further supported the hypothesis that the *cps15bM/cps15cM* gene is a key determinant of the different serotypes and suggested that changes in the number of TA repeat units cause the reported reversible switching of serotypes 15B and 15C.

For identification of the serotype 15B and 15C *cps* loci, we took advantage of the available nucleotide sequence of the serotype 14 *cps* locus (20). Previous DNA hybridization experiments suggested that serotype 15B and 15C *cps* loci closely resemble the serotype 14 *cps* locus (19). The only reported structural differences between the repeating units of serotypes 14 and 15 are the presence of an additional galactose moiety and a phosphorylcholine group in serotypes 15B and 15C and the presence of an *O*-acetyl group in serotype 15B (14) (Fig. 1). After sequencing of part of the serotype 15B *cps* locus and comparison with the serotype 14 sequence, several genes unique to serotype 15B were identified. These genes encode a putative *O*-acetyltransferase, a putative galactosyltransferase, and several proteins that are probably involved in the synthesis of a glycerol 2-phosphate side chain. The putative galactosyltransferase CPS15bK likely accounts for the presence of the additional galactose residue in the pentasaccharide repeating unit of serotype 15B compared to the serotype 14 residues. In serotype 14, the gene encoding this protein is present, but the

N-terminal part of the protein is not present due to a single nucleotide deletion.

The identification of putative glycerophosphate biosynthesis genes in serotype 15B was unexpected as both serotypes 15B and 15C have been reported to contain a phosphorylcholine rather than a glycerol 2-phosphate moiety (14), although for one serotype 15B strain the presence of a glycerol 2-phosphate side chain has been reported (1). Glycerol 2-phosphate is present in the CPS of serotype 15A (6). As lateral gene transfer among *cps* loci appears to be a frequent event (10) and assuming that the structures determined are correct, our data may indicate that there is diversity in lipid side chain biosynthesis in serotype 15 strains.

The identification of the Cps15bM protein as a putative *O*-acetyltransferase was based on a combination of moderate levels of sequence similarity with the putative acetyltransferases of *S. pneumoniae* serotypes 33F and 9V, *P. putida*, and *A. actinomycetemcomitans* and a strong resemblance in the predicted transmembrane segments to a family of membrane proteins involved in acetylation. The apparent presence of species- and/or serotype-specific sequences within a relatively conserved overall protein structure may indicate that during evolution the *cps15bM/cps15cM* gene was acquired by horizontal transfer from other species and subsequently acquired serotype-specific characteristics. The extremely low G+C content of the *cps15bM/cps15cM* gene (24%) compared to that of the *S. pneumoniae* genome (39.7%) (48) is consistent with this hypothesis.

A key finding of our work was that the structural difference between the serotype 15B and 15C capsules (i.e., the presence of an *O*-acetyl group in serotype 15B but not in serotype 15C) is based on the presence of an intact *cps15bM/cps15cM* gene. In serotype 15C, the corresponding ORF was interrupted due to a frameshift. This likely explains the lack of *O* acetylation of

serotype 15C CPS. Interestingly, the premature stop in translation in *cps15cM* was caused by insertion of a TA unit into a short tandem repeat sequence consisting of eight repetitive TA units. Further analysis of 15 serotype 15B and 15C strains confirmed the direct relationship between the number of repetitive TA units in the putative *O*-acetyltransferase gene and the serotype. In all serotype 15B strains that were analyzed the *cps15bM* gene had eight repetitive TA units in the central part of the gene, which resulted in an intact ORF. In contrast, serotype 15C strains always contained seven or nine repetitive TA units, which resulted in a defective gene and loss of *O*-acetyl groups from the CPS.

The presence of a short tandem TA repeat sequence rather than a more structural gene defect in serotype 15C may also explain the reversible switching between the 15B and 15C capsule serotypes that has been demonstrated in the laboratory and likely occurs during natural infection (51). Despite numerous attempts, we were unable to demonstrate capsule switching and/or variation in the number of TA repeat units in the laboratory. Colony blotting to select for capsule variants failed due to cross-reactivity of the polyclonal antisera in this assay, and construction of an antibiotic reporter gene did not yield the desired phase variants (data not shown). These data suggest that switching of the serotype is not a frequent event and escapes the detection limit under the conditions employed. Some serotype 15C strains examined by Venkateswaran et al. (51) also failed to give rise to variants *in vitro*, while others showed a relatively high frequency of variation. These data suggest that the frequency of variation may be strain dependent.

In many gram-negative pathogens, variations in the lengths of stretches of repetitive homo- or polymeric nucleotides have been demonstrated to play an important role in bacterial virulence (15, 31, 45, 47, 49, 53, 55). These DNA repeats are found within or upstream of coding sequences and can vary in length due to slippage of the DNA polymerase during replication. Slippage can occur between short direct repeats and involves pausing by the DNA polymerase within the repeat and dissociation of the polymerase from the DNA. At the same time, the terminal portion of the newly synthesized strand separates from the template and anneals to another direct repeat, and after resumption of DNA replication the number of repeats in the newly synthesized strand might differ from the number in the template DNA (52). For gram-positive microorganisms, bacterial surface variation and/or adaptation via the slipped-strand base mispairing mechanism has been demonstrated only rarely so far (25, 40, 43). In the genome of *S. pneumoniae* serotype 4, iterative DNA motifs have been identified in 18% of the genes, most of which are homopolymeric tracts. Twenty-five of these genes are directly related to virulence and include genes from the teichoic acid and capsule biosynthesis pathways. Pericone et al. (40) demonstrated for various genes that homopolymeric tracts exhibited length variation. Our data suggest that heteropolymeric tracts present in pneumococci may contribute to surface variation.

REFERENCES

1. Abeygunawardana, C., T. C. Williams, J. S. Sumner, and J. P. Hennessey, Jr. 2000. Development and validation of an NMR-based identity assay for bacterial polysaccharides. *Anal. Biochem.* **279**:226–240.
2. Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
3. Arrecubieta, C., E. Garcia, and R. Lopez. 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.
4. Barnes, D. M., S. Whittier, P. H. Gilligan, S. Soares, A. Tomasz, and F. W. Henderson. 1995. Transmission of multidrug-resistant serotype 23F *Streptococcus pneumoniae* in group day care: evidence suggesting capsular transformation of the resistant strain *in vivo*. *J. Infect. Dis.* **171**:890–896.
5. Bender, M. H., and J. Yother. 2001. CpsB is a modulator of capsule-associated tyrosine kinase activity in *Streptococcus pneumoniae*. *J. Biol. Chem.* **276**:47966–47974.
6. Caroff, M., and M. B. Perry. 1984. The specific capsular polysaccharide of *Streptococcus pneumoniae* type 15A (American type 30). *Can. J. Biochem. Cell Biol.* **62**:151–161.
7. Cieslewicz, M. J., D. L. Kasper, Y. Wang, and M. R. Wessels. 2001. Functional analysis in type Ia group B *Streptococcus* of a cluster of genes involved in extracellular polysaccharide production by diverse species of streptococci. *J. Biol. Chem.* **276**:139–146.
8. Claverys, J. P., M. Prudhomme, I. Mortier-Barriere, and B. Martin. 2000. Adaptation to the environment: *Streptococcus pneumoniae*, a paradigm for recombination-mediated genetic plasticity? *Mol. Microbiol.* **35**:251–259.
9. Coffey, T. J., C. G. Dowson, M. Daniels, J. Zhou, C. Martin, B. G. Spratt, and J. M. Musser. 1991. Horizontal transfer of multiple penicillin-binding protein genes, and capsular biosynthetic genes, in natural populations of *Streptococcus pneumoniae*. *Mol. Microbiol.* **5**:2255–2260.
10. Coffey, T. J., M. C. Enright, M. Daniels, J. K. Morona, R. Morona, W. Hryniewicz, J. C. Paton, and B. G. Spratt. 1998. Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of *Streptococcus pneumoniae*. *Mol. Microbiol.* **27**:73–83.
11. Coffey, T. J., M. C. Enright, M. Daniels, P. Wilkinson, S. Berron, A. Fenoll, and B. G. Spratt. 1998. Serotype 19A variants of the Spanish serotype 23F multiresistant clone of *Streptococcus pneumoniae*. *Microb. Drug Resist.* **4**:51–55.
12. Dillard, J. P., M. W. Vandersea, and J. Yother. 1995. Characterization of the cassette containing genes for type 3 capsular polysaccharide biosynthesis in *Streptococcus pneumoniae*. *J. Exp. Med.* **181**:973–983.
13. Iannelli, F., B. J. Pearce, and G. Pozzi. 1999. The type 2 capsule locus of *Streptococcus pneumoniae*. *J. Bacteriol.* **181**:2652–2654.
14. Jansson, P. E., B. Lindberg, U. Lindquist, and J. Ljungberg. 1987. Structural studies of the capsular polysaccharide from *Streptococcus pneumoniae* types 15B and 15C. *Carbohydr. Res.* **162**:111–116.
15. Jennings, M. P., D. W. Hood, I. R. Peak, M. Virji, and E. R. Moxon. 1995. Molecular analysis of a locus for the biosynthesis and phase-variable expression of the lacto-*N*-neotetraose terminal lipopolysaccharide structure in *Neisseria meningitidis*. *Mol. Microbiol.* **18**:729–740.
16. Jiang, S. M., L. Wang, and P. R. Reeves. 2001. Molecular characterization of *Streptococcus pneumoniae* type 4, 6B, 8, and 18C capsular polysaccharide gene clusters. *Infect. Immun.* **69**:1244–1255.
17. Junker, F., and J. L. Ramos. 1999. Involvement of the *cis/trans* isomerase Cti in solvent resistance of *Pseudomonas putida* DOT-T1E. *J. Bacteriol.* **181**:5693–5700.
18. Kamerling, J. P. 1999. Pneumococcal polysaccharides: a chemical view, p. 81–114. In A. Tomasz (ed.), *Streptococcus pneumoniae: molecular biology and mechanisms of disease*. Mary Ann Liebert, Inc., Larchmont, N.Y.
19. Kolkman, M. A. B., B. A. M. van der Zeijst, and P. J. M. Nuijten. 1998. Diversity of capsular polysaccharide synthesis gene clusters in *Streptococcus pneumoniae*. *J. Biochem. (Tokyo)* **123**:937–945.
20. Kolkman, M. A. B., W. Wakarchuk, P. J. M. Nuijten, and B. A. M. van der Zeijst. 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.
21. Lamothe, G. T., L. Jolly, B. Mollet, and F. Stinglele. 2002. Genetic and biochemical characterization of exopolysaccharide biosynthesis by *Lactobacillus delbrueckii* subsp. *bulgaricus*. *Arch. Microbiol.* **178**:218–228.
22. Lindberg, B., J. Lonngren, and D. A. Powell. 1977. Structural studies on the specific type-14 pneumococcal polysaccharide. *Carbohydr. Res.* **58**:177–186.
23. Llull, D., R. Munoz, R. Lopez, and E. Garcia. 1999. A single gene (*its*) located outside the *cap* locus directs the formation of *Streptococcus pneumoniae* type 37 capsular polysaccharide. Type 37 pneumococci are natural, genetically binary strains. *J. Exp. Med.* **190**:241–251.
24. Llull, D., R. Lopez, E. Garcia, and R. Munoz. 1998. Molecular structure of the gene cluster responsible for the synthesis of the polysaccharide capsule of *Streptococcus pneumoniae* type 33F. *Biochim. Biophys. Acta* **1443**:217–224.
25. Lukomski, S., K. Nakashima, I. Abdi, V. J. Cipriano, B. J. Shelvin, E. A. Graviss, and J. M. Musser. 2001. Identification and characterization of a second extracellular collagen-like protein made by group A *Streptococcus*: control of production at the level of translation. *Infect. Immun.* **69**:1729–1738.
26. Morona, J. K., D. C. Miller, T. J. Coffey, C. J. Vindurampulle, B. G. Spratt, R. Morona, and J. C. Paton. 1999. Molecular and genetic characterization of

- the capsule biosynthesis locus of *Streptococcus pneumoniae* type 23F. Microbiology **145**:781–789.
27. Morona, J. K., R. Morona, and J. C. Paton. 1997. Characterization of the locus encoding the *Streptococcus pneumoniae* type 19F capsular polysaccharide biosynthetic pathway. Mol. Microbiol. **23**:751–763.
 28. Morona, J. K., R. Morona, and J. C. Paton. 1999. Comparative genetics of capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* types belonging to serogroup 19. J. Bacteriol. **181**:5355–53564.
 29. Morona, J. K., R. Morona, and J. C. Paton. 1997. Molecular and genetic characterization of the capsule biosynthesis locus of *Streptococcus pneumoniae* type 19B. J. Bacteriol. **179**:4953–4958.
 30. Morona, J. K., J. C. Paton, D. C. Miller, and R. Morona. 2000. Tyrosine phosphorylation of CpsD negatively regulates capsular polysaccharide biosynthesis in *Streptococcus pneumoniae*. Mol. Microbiol. **35**:1431–1442.
 31. Moxon, E. R., P. B. Rainey, M. A. Nowak, and R. E. Lenski. 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. Curr. Biol. **4**:24–33.
 32. Mulholland, K. 1999. Strategies for the control of pneumococcal diseases. Vaccine **17**(Suppl. 1):S79–S84.
 33. Munoz, R., M. Mollerach, R. Lopez, and E. Garcia. 1999. Characterization of the type 8 capsular gene cluster of *Streptococcus pneumoniae*. J. Bacteriol. **181**:6214–6219.
 34. Munoz, R., M. Mollerach, R. Lopez, and E. Garcia. 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.
 35. Musher, D. M. 1992. Infections caused by *Streptococcus pneumoniae*: clinical spectrum, pathogenesis, immunity, and treatment. Clin. Infect. Dis. **14**:801–807.
 36. Nakano, Y., Y. Yoshida, Y. Yamashita, and T. Koga. 1998. A gene cluster for 6-deoxy-L-talan synthesis in *Actinobacillus actinomycetemcomitans*. Biochim. Biophys. Acta **1442**:409–414.
 37. Nesin, M., M. Ramirez, and A. Tomasz. 1998. Capsular transformation of a multidrug-resistant *Streptococcus pneumoniae* in vivo. J. Infect. Dis. **177**:707–713.
 38. Paton, J. C., and J. K. Morona. 2000. *Streptococcus pneumoniae* capsular polysaccharide, p. 201–213. In V. A. Fischetti (ed.), Gram-positive pathogens. American Society for Microbiology, Washington D.C.
 39. Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.
 40. Pericone, C. D., D. Bae, M. Shchepetov, T. McCool, and J. N. Weiser. 2002. Short-sequence tandem and nontandem DNA repeats and endogenous hydrogen peroxide production contribute to genetic instability of *Streptococcus pneumoniae*. J. Bacteriol. **184**:4392–4399.
 41. Ramirez, M., and A. Tomasz. 1999. Acquisition of new capsular genes among clinical isolates of antibiotic-resistant *Streptococcus pneumoniae*. Microb. Drug Resist. **5**:241–246.
 42. Ramirez, M., and A. Tomasz. 1998. Molecular characterization of the complete 23F capsular polysaccharide locus of *Streptococcus pneumoniae*. J. Bacteriol. **180**:5273–5278.
 43. Rasmussen, M., and L. Bjorck. 2001. Unique regulation of SclB—a novel collagen-like surface protein of *Streptococcus pyogenes*. Mol. Microbiol. **40**:1427–1438.
 44. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
 45. Sarkari, J., N. Pandit, E. R. Moxon, and M. Achtman. 1994. Variable expression of the Ope outer membrane protein in *Neisseria meningitidis* is caused by size variation of a promoter containing poly-cytidine. Mol. Microbiol. **13**:207–217.
 46. Slauch, J. M., A. A. Lee, M. J. Mahan, and J. J. Mekalanos. 1996. Molecular characterization of the *oafA* locus responsible for acetylation of *Salmonella typhimurium* O-antigen: OafA is a member of a family of integral membrane transacylases. J. Bacteriol. **178**:5904–5909.
 47. Stern, A., M. Brown, P. Nickel, and T. F. Meyer. 1986. Opacity genes in *Neisseria gonorrhoeae*: control of phase and antigenic variation. Cell **47**:61–71.
 48. Tettelin, H., K. E. Nelson, I. T. Paulsen, J. A. Eisen, T. D. Read, S. Peterson, J. Heidelberg, R. T. DeBoy, D. H. Haft, R. J. Dodson, A. S. Durkin, M. Gwinn, J. F. Kolonay, W. C. Nelson, J. D. Peterson, L. A. Umayam, O. White, S. L. Salzberg, M. R. Lewis, D. Radune, E. Holtzapple, H. Khouri, A. M. Wolf, T. R. Utterback, C. L. Hansen, L. A. McDonald, T. V. Feldblyum, S. Angiuoli, T. Dickinson, E. K. Hickey, I. E. Holt, B. J. Loftus, F. Yang, H. O. Smith, J. C. Venter, B. A. Dougherty, D. A. Morrison, S. K. Hollingshead, and C. M. Fraser. 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science **293**:498–506.
 49. van Ham, S. M., L. van Alphen, F. R. Mooi, and J. P. M. van Putten. 1993. Phase variation of *H. influenzae* fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. Cell **73**:1187–1196.
 50. van Selm, S., M. A. B. Kolkman, B. A. M. van der Zeijst, K. A. Zwaagstra, W. Gaastra, and J. P. M. van Putten. 2002. Organization and characterization of the capsule biosynthesis locus of *Streptococcus pneumoniae* serotype 9V. Microbiology **148**:1747–1755.
 51. Venkateswaran, P. S., N. Stanton, and R. Austrian. 1983. Type variation of strains of *Streptococcus pneumoniae* in capsular serogroup 15. J. Infect. Dis. **147**:1041–1054.
 52. Viguera, E., D. Canceill, and S. D. Ehrlich. 2001. Replication slippage involves DNA polymerase pausing and dissociation. EMBO J. **20**:2587–2595.
 53. Wang, G., Z. Ge, D. A. Rasko, and D. E. Taylor. 2000. Lewis antigens in *Helicobacter pylori*: biosynthesis and phase variation. Mol. Microbiol. **36**:1187–1196.
 54. Weiser, J. N., D. Bae, H. Epino, S. B. Gordon, M. Kapoor, L. A. Zenewicz, and M. Shchepetov. 2001. Changes in availability of oxygen accentuate differences in capsular polysaccharide expression by phenotypic variants and clinical isolates of *Streptococcus pneumoniae*. Infect. Immun. **69**:5430–5439.
 55. Weiser, J. N., M. Shchepetov, and S. T. Chong. 1997. Decoration of lipopolysaccharide with phosphorylcholine: a phase-variable characteristic of *Haemophilus influenzae*. Infect. Immun. **65**:943–950.