Organization and characterization of the capsule biosynthesis locus of *Streptococcus pneumoniae* serotype 9V

Saskia van Selm,¹ Marc A. B. Kolkman,² Bernard A. M. van der Zeijst,³ Kornelijse A. Zwaagstra,¹ Wim Gaastra¹ and Jos P. M. van Putten¹

Author for correspondence: Jos P. M. van Putten. Tel: +31 30 2534888. Fax: +31 30 2540784. e-mail: j.vanputten@vet.uu.nl

The capsular polysaccharide (CPS) synthesis locus of *Streptococcus pneumoniae* serotype 9V was amplified by long-range PCR and sequenced. The locus was 17368 bp in size and contained 15 ORFs. The genetic organization of the cluster shared many features with other *S. pneumoniae* capsule loci, including the presence of four putative regulatory genes at the 5’ end. Comparative sequence analyses allowed putative functions to be assigned to each of the gene products. The ORFs appeared to encode, besides the four regulatory genes, five glycosyltransferases, two O-acetyltransferases, an N-acetylglucosamine 2-epimerase, a glucose 6-dehydrogenase, an oligosaccharide transporter protein and a polysaccharide repeating unit polymerase. These functions covered the steps proposed in the CPS biosynthesis of serotype 9V. TLC of carbohydrate intermediates formed after incubation of bacterial membrane preparations with ¹⁴C-labelled precursors demonstrated that the fifth ORF (cps9vE) encoded a UDP-glucosyl-1-phosphate transferase. This function was confirmed with the help of a cps9vE mutant that carried a deletion of a guanine residue located adjacent to a stretch of adenines. The identification and characterization of the serotype 9V locus is a major step in unravelling the 9V capsule biosynthesis pathway and broadens the insight into the genetic diversity of the *S. pneumoniae* capsule loci.

**Keywords:** capsular polysaccharide biosynthesis, glucosyltransferase

INTRODUCTION

The human pathogen *Streptococcus pneumoniae* (pneumococcus) can cause a variety of infectious diseases, most notably otitis media, pneumonia and meningitis (Mulholland, 1999; Musher, 1992). Adequate treatment is increasingly complicated by the emergence of strains that are resistant to commonly used antibiotics (Carpentier & Tuomanen, 2000; Tomasz, 1999). Many of the vaccines that have been developed provide protection by eliciting antibodies against the pneumococcal capsule polysaccharide. The efficacy of these vaccines, however, is still not optimal as they cover only a number of capsule serotypes (Butler et al., 1999; Obaro, 2000). At this time, 90 different serotypes of *S. pneumoniae* are known, based on antigenic differences in polysaccharide capsule (Kamerling, 1999), and this number may increase further as genetic exchange between serotypes has been demonstrated (Barnes et al., 1995; Coffey et al., 1991, 1998a, b; Nesin et al., 1998; Ramirez & Tomasz, 1999).

Structurally, the capsular polysaccharide (CPS) of *S. pneumoniae* consists of a polymer of repeating oligosaccharide units. These basic building blocks are usually composed of two to eight monosaccharides that are linked via glycosidic bonds. Heterogeneity among capsules is largely established through variation in the type and linkage of the different monosaccharides that constitute the repeating unit, and their variable substitution with non-carbohydrate residues (Kamerling, 1999). Polymerization of the repeating units results in the high-molecular-mass capsule structure that surrounds the pneumococcus and protects the organism against hostile environments.

**Abbreviation:** CPS, capsular polysaccharide.

The GenBank accession number for the sequence reported in this paper is AF402095.
For a number of pneumococcal serotypes, the genes that encode the capsule biosynthesis machinery have been identified. The genes are clustered within the chromosome and in all but one serotype (serotype 37) are located between \textit{dexB} and \textit{aliA} (Llull et al., 1999). At this time the CPS synthesis (\textit{cps}) loci of 15 serotypes (1, 2, 3, 4, 6B, 8, 14, 18C, 19F, 19A, 19B, 19C, 23F, 33F and 37) have been sequenced (Arrecubieta et al., 1995; Dillard et al., 1995; Iannelli et al., 1999; Jiang et al., 2001; Kolkman et al., 1997b; Llull et al., 1998, 1999; Morona et al., 1997a, 1999a; Munoz et al., 1997, 1999; Ramirez & Tomasz, 1998). Sequence analysis and function predictions suggest that the loci contain the complete repertoire of genes specifically required for capsule biosynthesis and its regulation. In a few cases, the function of a gene has been experimentally confirmed (Arrecubieta et al., 1995, 1996; Kolkman et al., 1996, 1997a, b; Morona et al., 1997a, 2000; Munoz et al., 1997).

Knowledge of the composition of the various \textit{cps} loci and the function of the encoded proteins is important as it provides insight into the dynamics of the capsule biosynthesis pathway and the evolution of serotypes. Genetic exchange between \textit{cps} loci occurs frequently and can lead to changes in serotype (Barnes et al., 1995; Coffey et al., 1991, 1998a, b; Nesin et al., 1998; Ramirez & Tomasz, 1999). The recombination events can involve large pieces of DNA (15–25 kb) with crossovers over points often located outside the \textit{cps} locus, as well as smaller fragments with exchange between regions of sequence similarity within the various loci (Morona et al., 1999b). In the present study, we investigated the genetic organization of the \textit{cps} locus of serotype 9V. This serotype is one out of 12 most commonly isolated serotypes worldwide (Butler, 1997; Nielsen & Henrichsen, 1992) and is included in the 7-valent conjugated vaccine (Hausdorff et al., 2000; Shinefield & Black, 2000) and the 23-valent polysaccharide vaccine (Robbins et al., 1983). We have identified and sequenced the complete 9V locus and assigned putative functions to the genes. For one of the genes, its proposed function has been experimentally confirmed.

\section*{METHODS}

\subsection*{Bacterial strains and growth conditions.} \textit{S. pneumoniae} serotypes 9V (strain 961793) and 4 were clinical isolates kindly provided by Dr A. van de Ende (Amsterdam, The Netherlands). Strain 9V\textit{Acps9E} was a derivative of strain 961793 (\textit{C} collection of Type Cultures, Central Public Health Laboratory, London, UK). All \textit{S. pneumoniae} strains were grown (37°C, 5% CO\textsubscript{2}) in Todd–Hewitt broth supplemented with 0.5% yeast extract in polypropylene tubes without shaking, or on Columbia (sheep) blood agar. Subclones were made in pBluescript II KS (Stratagene) using \textit{Escherichia coli} DH5\textsubscript{a} as a host. \textit{E. coli} was grown in LB medium containing the appropriate antibiotics (Sambrook et al., 1989).

\subsection*{DNA techniques.} Chromosomal DNA was isolated as described by Ausubel et al. (1987). Long-range PCR was performed with the Expand Long Template PCR system (Roche Diagnostics) using buffer 3, according to the manufacturer’s instructions. PCR involved a 2 min denaturation at 94°C, followed by 10 cycles of 30 s denaturation, 30 s annealing with a touchdown from 60 to 55 or 50°C (depending on the melting temperature of the primers) and 20 min extension at 68°C, then 30 additional cycles consisting of 30 s denaturation, 30 s annealing at 55 or 50°C and 20 min extension at 68°C, lengthened by 20 s every cycle. The primers used to amplify the \textit{cps} locus of serotype 9V were DEXBPR (5'-CCATGGGAGTCTTCTGTTG-3') and ALIAST2 (5'-CAAATGGTGAAGTTATACAGGCTGTCTC-3'). PCR products were purified from agarose gels with the QiAquick gel extraction kit (Qiagen). Other DNA techniques were performed as described by Sambrook et al. (1989). Restriction endonucleases and T4 DNA ligase were purchased from Amersham Pharmacia Biotech.

\subsection*{DNA sequencing and analysis.} DNA sequencing was performed on an ABI Prism 310 Genetic Analyser (Applied Biosystems) using the ABI Dye Terminator Cycle Sequencing Kit. DNA and protein data were analysed by using Lasergene software (DNASTar). The algorithm \textsc{blast} (Altschul et al., 1997) was used to compare sequences at the DNA and deduced amino acid levels to database sequences available at the National Center for Biotechnology Information (NCBI). Transmembrane segment predictions were done using the program \textsc{dAS} (http://au.expasy.org).

\subsection*{Membrane preparations.} Pneumococcal membranes for use in glycosyltransferase assays were prepared by the method of Osborn et al. (1972) with some modifications. Bacteria, grown in 10 ml Todd–Hewitt broth supplemented with 0.5% yeast extract in 15 ml tubes (15 h, 37°C, no shaking), diluted 10-fold in 250 ml medium and re-grown to mid-exponential phase (OD\textsubscript{600} 0.3), were collected by centrifugation (10 000 g, 15 min, 4°C), resuspended in 10 ml Sol1 (0.7 M sucrose, 50 mM Tris/HCl, 1 mM EDTA, pH 8.0), centrifuged again and resuspended in 10 ml Sol1 supplemented with 20 mg lysozyme and 100 U mutanolysin. After 2–4 h incubation at 4°C, 15 ml Sol1 was added and the protoplasts were collected by centrifugation and lysed in 20 ml Sol2 (20 mM Tris/HCl, 1 mM EDTA, 200 μM PMSF, pH 8.0). This suspension was sonicated (3 × 20 s) and stored overnight at −80°C. Debris was removed by centrifugation (5500 g, 20 min, 4°C) and membranes were collected by high-speed centrifugation (30000 g, 1 h, 4°C). Membranes were washed once in 20 ml Sol3 (50 mM Tris/acetate, 1 mM EDTA, 200 μM PMSF, pH 8.3), centrifuged again and finally resuspended in 0.5 ml Sol3. Membrane preparations could be stored at −80°C for a few months without apparent loss of enzyme activity.

\subsection*{Glycosyltransferase activity assays.} Glycosyltransferase activity was essentially determined as described previously (Kolkman et al., 1996). For each reaction, 40 μl membrane preparations was incubated (1 h, 10°C) with 0.025 μCi uridine diphospho-[\textsuperscript{3}H]glucose (Amersham Pharmacia) and 10 mM MgCl\textsubscript{2} in a final volume of 50 μl. Reactions were stopped by the addition of 1 ml chloroform/methanol (2:1). The solution was extracted (1 min, 22°C) with 0.2 ml PSUB (1:5 ml chloroform, 25 ml methanol, 23.5 ml water, 0.183 g KCl). The upper phase was discarded and the remaining solution was re-extracted twice. The incorporation of [\textsuperscript{3}H]-labelled glucose into the glycolipid fraction was measured in a Beckman LS3801 scintillation counter (Beckman Coulter).

\subsection*{Analysis of lipid-linked intermediates by TLC.} Lipid-linked intermediates were hydrolysed from the lipid carriers by mild acid hydrolysis. In this procedure one-fifth of the glycolipid fraction was dried in a Speed-Vac at 65°C and was resus-
putative promoter of this locus is also marked (dexB, polymerase; aliA, glycosyltransferases; Cps9vE, F, G, J and L are putative glycosyltransferases; Cps9vH and Cps9vN are involved in nucleotide sugar biosynthesis; Cps9vK and Cps9vL showed some similarity (37%) with WciU of S. pneumoniae serotype 18C. Cps9vM showed similarity to WciU of S. pneumoniae serotype 9V; Fig. 2) carries a galactose attached in an α-1,3-glycosidic linkage to a lipid-linked glucose. Cps9vN showed similarity to several putative glycosyltransferases, including WciL of serotype 4 (47%) which (as serotype 9V; Fig. 2) carried an α-1,3-glycosidic linkage to a ManNAc residue. Cps9vJ and Cps9vL were most similar to putative glycosyltransferases in other species, although Cps9vL showed some similarity (37%) with WciU of S. pneumoniae serotype 18C.

Cps9vH and Cps9vN are involved in nucleotide sugar biosynthesis

Cps9vH showed similarity with UDP-GlcNAc 2-epimerases (Table 1). These enzymes convert UDP-GlcNAc to UDP-ManNAc (Campbell et al., 2000a). Cps9vN resembled UDP-glucose-6-dehydrogenases (Table 1) were located on this strand directly adjacent to the dexB and aliA genes (Fig. 1).

Putative functions Cps9vA–D

Database searches using the BLAST algorithm indicated that the first four ORFs of the cps9V locus (Cps9vA–D) were highly similar (>87%) to the corresponding ORFs in other S. pneumoniae cps loci (Table 1). For serotype 2 and 19F, three of these ORFs (designated CpsB–D) have been found to play a role in capsule regulation and are suggested to be involved in enhancing the activity of the polymerase (Bender & Yother, 2001; Morona et al., 2000). The CpsA proteins showed similarity with a protein in group B Streptococcus type 1a that may act as a transcriptional activator of the capsule gene operon (Cieslewicz et al., 2001). On the basis of the apparent conservation of the A–D genes among serotypes, we expect the 9V homologues to have similar functions.

Cps9vE, F, G, J and L are putative glycosyltransferases

Five ORFs of the 9V locus showed similarity with putative glycosyltransferases (Table 1). Cps9vE was homologous (>83% similarity) to proteins in several other S. pneumoniae serotypes, including Cps14E which we previously demonstrated to function as a UDP-glucosyl-1-phosphate transferase (Kolkman et al., 1996, 1997a). Cps9vF resembled (>90% similarity) serogroup 19 β-1,4-N-acetylmannosaminyltransferases, which transfer UDP-ManNAc in a β-1,4-glycosidic linkage to a lipid-linked glucose. Cps9vG showed similarity to several putative glycosyltransferases, including WciL of serotype 4 (47%) which (as serotype 9V; Fig. 2) carries a galactose attached in an α-1,3-glycosidic linkage to a ManNAc residue. Cps9vJ and Cps9vL were most similar to putative glycosyltransferases in other species, although Cps9vL showed some similarity (37%) with WciU of S. pneumoniae serotype 18C.

Cps9vH and Cps9vN are involved in nucleotide sugar biosynthesis

Cps9vH showed similarity with UDP-GlcNAc 2-epimerases (Table 1). These enzymes convert UDP-GlcNAc to UDP-ManNAc (Campbell et al., 2000a). Cps9vN resembled UDP-glucose-6-dehydrogenases (Table 1) were located on this strand directly adjacent to the dexB and aliA genes (Fig. 1).

Putative functions Cps9vA–D

Database searches using the BLAST algorithm indicated that the first four ORFs of the cps9V locus (Cps9vA–D) were highly similar (>87%) to the corresponding ORFs in other S. pneumoniae cps loci (Table 1). For serotype 2 and 19F, three of these ORFs (designated CpsB–D) have been found to play a role in capsule regulation and are suggested to be involved in enhancing the activity of the polymerase (Bender & Yother, 2001; Morona et al., 2000). The CpsA proteins showed similarity with a protein in group B Streptococcus type 1a that may act as a transcriptional activator of the capsule gene operon (Cieslewicz et al., 2001). On the basis of the apparent conservation of the A–D genes among serotypes, we expect the 9V homologues to have similar functions.

Cps9vE, F, G, J and L are putative glycosyltransferases

Five ORFs of the 9V locus showed similarity with putative glycosyltransferases (Table 1). Cps9vE was homologous (>83% similarity) to proteins in several other S. pneumoniae serotypes, including Cps14E which we previously demonstrated to function as a UDP-glucosyl-1-phosphate transferase (Kolkman et al., 1996, 1997a). Cps9vF resembled (>90% similarity) serogroup 19 β-1,4-N-acetylmannosaminyltransferases, which transfer UDP-ManNAc in a β-1,4-glycosidic linkage to a lipid-linked glucose. Cps9vG showed similarity to several putative glycosyltransferases, including WciL of serotype 4 (47%) which (as serotype 9V; Fig. 2) carries a galactose attached in an α-1,3-glycosidic linkage to a ManNAc residue. Cps9vJ and Cps9vL were most similar to putative glycosyltransferases in other species, although Cps9vL showed some similarity (37%) with WciU of S. pneumoniae serotype 18C.

Cps9vH and Cps9vN are involved in nucleotide sugar biosynthesis

Cps9vH showed similarity with UDP-GlcNAc 2-epimerases (Table 1). These enzymes convert UDP-GlcNAc to UDP-ManNAc (Campbell et al., 2000a). Cps9vN resembled UDP-glucose-6-dehydrogenases (Table 1) were located on this strand directly adjacent to the dexB and aliA genes (Fig. 1).
### Table 1. Properties of the ORFs in the cps locus of *S. pneumoniae* serotype 9V

<table>
<thead>
<tr>
<th>ORF*</th>
<th>BPGN name†</th>
<th>Position in sequence (nt)</th>
<th>G + C (mol%)</th>
<th>No. of aa</th>
<th>Predicted mol. mass (kDa)</th>
<th>Homologue in <em>S. pneumoniae</em> serotype‡</th>
<th>Similar gene product (% similarity/no. of aa)$§$</th>
<th>Proposed function of gene product</th>
</tr>
</thead>
<tbody>
<tr>
<td>cps9vA</td>
<td>wzg</td>
<td>570–2015</td>
<td>39</td>
<td>481</td>
<td>53.3</td>
<td>All except 3 + 37</td>
<td><em>S. pneumoniae</em> Cps2A (99/481)</td>
<td>Transcriptional activator of <em>cps</em> locus†</td>
</tr>
<tr>
<td>cps9vB</td>
<td>wzb</td>
<td>2017–2748</td>
<td>42</td>
<td>243</td>
<td>28.2</td>
<td>All except 3 + 37</td>
<td><em>S. pneumoniae</em> serotype 4 Wzh (99/243)</td>
<td>Phosphotyrosine protein phosphatase</td>
</tr>
<tr>
<td>cps9vC</td>
<td>wzd</td>
<td>2757–3449</td>
<td>42</td>
<td>230</td>
<td>25.6</td>
<td>All except 3 + 37</td>
<td><em>S. pneumoniae</em> Cps8C (96/230)</td>
<td>Required for Cps9vD phosphorylation</td>
</tr>
<tr>
<td>cps9vD</td>
<td>wze</td>
<td>3459–4148</td>
<td>39</td>
<td>229</td>
<td>25.1</td>
<td>All except 3 + 37</td>
<td><em>S. pneumoniae</em> serotype 6B Wze (98/229)</td>
<td>Autophosphorylating protein-tyrosine kinase</td>
</tr>
<tr>
<td>cps9vE</td>
<td>wchA</td>
<td>4163–5530</td>
<td>38</td>
<td>455</td>
<td>52.1</td>
<td>2, 6B, 8, 14, 18C, 19F–C, 23F, 33F</td>
<td><em>S. pneumoniae</em> Cps 23fE (98/455)</td>
<td>UDP-glucosyl-1-phosphate transferase¶</td>
</tr>
<tr>
<td>cps9vF</td>
<td>wchO</td>
<td>5536–6279</td>
<td>33</td>
<td>247</td>
<td>28.2</td>
<td>19F–C</td>
<td><em>S. pneumoniae</em> Cps19aF (94/246)</td>
<td>β-1,4-N-Acetyl-mannosaminyltransferase§</td>
</tr>
<tr>
<td>cps9vG</td>
<td>wcjA</td>
<td>6254–7339</td>
<td>32</td>
<td>361</td>
<td>41.0</td>
<td></td>
<td><em>S. pneumoniae</em> putative glycosyltransferase (47/191)</td>
<td>Glycosyltransferase#</td>
</tr>
<tr>
<td>cps9vH</td>
<td>mnaA</td>
<td>7336–8424</td>
<td>39</td>
<td>361</td>
<td>40.1</td>
<td>4, 19F–C</td>
<td><em>S. pneumoniae</em> Cps19aK (89/359)</td>
<td>UDP-N-acetyl-glucosamine 2-epimerase</td>
</tr>
<tr>
<td>cps9vI</td>
<td>wxy</td>
<td>8466–9605</td>
<td>29</td>
<td>379</td>
<td>43.4</td>
<td></td>
<td><em>P. putida</em> O-antigen polymerase (43/364)</td>
<td>CPS repeating unit polymerase</td>
</tr>
<tr>
<td>cps9vJ</td>
<td>wcjB</td>
<td>9589–10458</td>
<td>29</td>
<td>289</td>
<td>33.7</td>
<td></td>
<td><em>Bacillus fragilis</em> putative glycosyltransferase (58/199)</td>
<td>Glycosyltransferase#</td>
</tr>
<tr>
<td>cps9vK</td>
<td>wzx</td>
<td>10458–11882</td>
<td>29</td>
<td>474</td>
<td>53.6</td>
<td></td>
<td><em>Lactococcus lactis</em> polysaccharide biosynthesis export protein (54/486)</td>
<td>CPS repeating unit transporter</td>
</tr>
<tr>
<td>cps9vL</td>
<td>wcjC</td>
<td>11889–13082</td>
<td>31</td>
<td>397</td>
<td>45.9</td>
<td></td>
<td><em>Actinobacillus pleurogrammonea</em> putative glycosyltransferase (43/347)</td>
<td>Glycosyltransferase#</td>
</tr>
<tr>
<td>cps9vM</td>
<td>wcjD</td>
<td>13083–13748</td>
<td>33</td>
<td>221</td>
<td>24.3</td>
<td></td>
<td><em>E. coli</em> putative O-acetyltransferase (55/208)</td>
<td>O-Acetyltransferase</td>
</tr>
<tr>
<td>cps9vN</td>
<td>ugd</td>
<td>15400–16632</td>
<td>34</td>
<td>410</td>
<td>46.1</td>
<td>1, 2, 3, 8</td>
<td><em>S. pneumoniae</em> Cap8L (85/412)</td>
<td>UDP-glucose-6-dehydrogenase</td>
</tr>
<tr>
<td>cps9vO</td>
<td>wcjE</td>
<td>16844–17878</td>
<td>31</td>
<td>344</td>
<td>39.9</td>
<td></td>
<td><em>Mesorhizobium loti</em> acetyltransferase (44/342)</td>
<td>O-Acetyltransferase</td>
</tr>
</tbody>
</table>

* ORFs were designated according to their appearance in the cps9V locus.
† ORF name according to Bacterial Polysaccharide Gene Nomenclature (Reeves et al., 1996).
‡ These homologues have at least 70% similarity with the corresponding 9V gene product. In serotype 3 CpsA–D homologues are present in a truncated or otherwise mutated form (Arrecubieta et al., 1995).
§ In the case of a similar gene product within a *cps* locus of another *S. pneumoniae* serotype, the gene with the highest scoring similarity was given. Similarity was expressed as a percentage of amino acid similarity over the no. of amino acids indicated.
¶ The function of this protein was established in this paper.
† The reactive intermediate of the monosaccharides transferred by these transferases is a UDP monosaccharide.
that convert UDP-glucose to UDP-glucuronic acid (Campbell et al., 2000b). Cps9vN homologues in pneumococcal serotypes 1 and 3 have been confirmed to possess UDP-glucose-6-dehydrogenase activity (Arrecubieta et al., 1996; Munoz et al., 1997). The functions of Cps9vH and Cps9vN suggest that the proteins indirectly participate in capsule biosynthesis by providing appropriate substrates.

**Cps9vI and Cps9vK are involved in capsule assembly**

On the basis of similarities to an O-antigen polymerase of *Pseudomonas putida* (43%) (Rodriguez-Herva et al., 1999) and a repeating unit polymerase of *Streptococcus agalactiae* (42%) (GenBank accession no. AF337958), Cps9vI was considered to be the CPS repeating unit polymerase. This idea was supported by transmembrane segment analysis using the program DAS which predicted that Cps9vI contained 12 transmembrane domains and showed a similar DAS profile to other putative repeating unit polymerases (data not shown).

The sequence of Cps9vK was most similar to polysaccharide export proteins and O-antigen repeating unit transporters. This suggests that this protein acts as the CPS repeating unit transporter that transports the repeating oligosaccharide unit across the cell wall.

**Cps9vM and Cps9vO are putative O-acetyltransferases**

Cps9vM and Cps9vO both showed similarity at the amino acid level to putative O-acetyltransferases. Cps9vM was 53% similar to Cap1F of *S. pneumoniae* serotype 1, which is a putative O-acetyltransferase (Munoz et al., 1997). Cps9vO showed 41% similarity to WciX of *S. pneumoniae* serotype 18C. WciX is a protein of unknown function, but since serotype 18C contains one O-acetyl group in its structure (Jiang et al., 2001) and no other gene in the type 18C locus was predicted to encode an O-acetyltransferase, this protein may likely act as the O-acetyltransferase. Transmembrane segment predictions suggest that Cps9vO has various transmembrane segments, while Cps9vM was predicted to be cytosolic.

**Insertion sequences**

Four sequences were found in the *cps9V* region that were similar to known insertion sequences (Table 2). An insertion sequence similar to RUPA, but with a deletion of 19 nt was found in front of the locus. *orf1* (582 bp) and *orf2* (783 bp), located between the genes *cps9vO* and *aliA* on the opposite strand, were similar to insertion sequence IS1167 with a frameshift in the ORF. This insertion element has also been found in other *cps* loci and at other positions within the *S. pneumoniae* chromosome (Coffey et al., 1998a; Kolkman et al., 1997b; Munoz et al., 1997; Tettelin et al., 2001). At the protein level, ORF3 (1044 bp) showed 52% similarity to a putative transposase of *Mycobacterium tuberculosis* (Cole et al., 1998). ORF4 (348 bp) is almost identical (99%) to the first part of the insertion sequence IS630-Spn1, which is also found in several other pneumococcal *cps* loci and elsewhere on the chromosome (Coffey et al., 1998a; Iannelli et al., 1999; Lulli et al., 1999; Munoz et al., 1997, 1999; Tettelin et al., 2001). We cannot exclude that the entire transposon is present in serotype 9V as the region between *dexB* and *cps9vA* was not fully sequenced.

**Identification of a spontaneous *cps9vE* mutant**

Repeated sequencing of regions of the 9V locus indicated that the chromosomal DNA used for amplifying DNA of the *cps9V* locus appeared to be a mixture of intact *cps9V* DNA and DNA with a single base deletion in *cps9vE*. At position 977, a guanidine residue adjacent to a stretch of seven adenines was deleted which resulted in a frameshift in the ORF. This finding was unexpected, as the DNA used for sequencing was isolated from an overnight culture derived from a single colony. When a single colony of *S. pneumoniae* serotype 9V strain 961793 was grown in broth, it was noticed that subcultures on Columbia agar plates appeared as a mixture of large and small colonies. This feature has previously been demonstrated to be associated with a variation in capsule production (Griffith, 1928). The *cps9vE* mutant 9VΔ*cps9vE* was used to further analyse the function of the Cps9vE protein (see below).

**cps9vE encodes a UDP-glucosyl-1-phosphate transferase**

To confirm the proposed function of Cps9vE as the glycosyltransferase that confers the first step in capsule biosynthesis, i.e. the transfer of glucose to the lipid carrier, we performed a direct functional enzyme assay. Membrane preparations of serotype 9V, the 9VΔ*cps9vE* mutant, and of the serotypes 14 and 4 as positive and negative controls, respectively, were incubated with 14C-labelled UDP-glucose (Kolkman et al., 1996, 1997a). The synthesis of labelled intermediates was determined after extraction of the glycolipid fraction. Liquid scintillation counting of the obtained fractions showed that the

\[
\begin{align*}
&\rightarrow \alpha-D-GlcP(1 \rightarrow 3)-\beta-D-MannNAc(1 \rightarrow 4)-\beta-D-GlcP(1 \rightarrow 4)-\alpha-D-GlcP(1 \rightarrow 2) OAc (3\%) \\
&\rightarrow \alpha-D-GlcP(1 \rightarrow 3)-\beta-D-MannNAc(1 \rightarrow 4)-\beta-D-GlcP(1 \rightarrow 4)-\alpha-D-GlcP(1 \rightarrow 3) OAc (17\%) \\
&\rightarrow \alpha-D-GlcP(1 \rightarrow 3)-\beta-D-MannNAc(1 \rightarrow 4)-\beta-D-GlcP(1 \rightarrow 4)-\alpha-D-GlcP(1 \rightarrow 2) OAc (5\%) \\
&\rightarrow \alpha-D-GlcP(1 \rightarrow 3)-\beta-D-MannNAc(1 \rightarrow 4)-\beta-D-GlcP(1 \rightarrow 4)-\alpha-D-GlcP(1 \rightarrow 3) OAc (55\%)
\end{align*}
\]

**Fig. 2.** Structure of the repeating unit of the CPS of *S. pneumoniae* serotype 9V (Perry et al., 1981; Rutherford et al., 1991). Glc, glucose; GlcA, glucuronic acid; Gal, galactose; ManNAc, 2-acetamido-2-deoxymannose; OAc, O-acetyl.
Table 2. Insertion sequences found in the cps locus of S. pneumoniae serotype 9V

<table>
<thead>
<tr>
<th>Position in cps9V sequence (nt)*</th>
<th>Homologue (percentage identity†)</th>
<th>Reference</th>
<th>Homologue in cps locus in other S. pneumoniae serotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>10–357 c</td>
<td>IS630-Spn1 (99)</td>
<td>Tettelin et al. (2001)</td>
<td>1, 2, 4, 8, 19F, 37</td>
</tr>
<tr>
<td>365–452</td>
<td>RUPA 19 bp deletion (91)</td>
<td>Oggioni &amp; Claverys (1999)</td>
<td>1, 2, 3, 4, 6B, 14, 18C, 19F, 19A, 19C</td>
</tr>
<tr>
<td>13871–14914 c</td>
<td>Putative transposase of Mycobacterium tuberculosis [52 (aa similarity)]</td>
<td>Cole et al. (1998)</td>
<td></td>
</tr>
<tr>
<td>17921–19227 c</td>
<td>IS1167 truncated (97)</td>
<td>Tettelin et al. (2001)</td>
<td>1, 14, 19F</td>
</tr>
</tbody>
</table>

* c indicates sequence on complementary strand.
† Percentage identity at the nucleotide level, unless indicated otherwise.

**DISCUSSION**

Pneumococcal capsule biosynthesis requires glycosyltransferases, assembly proteins that export and polymerize the oligosaccharide units, and modifying enzymes. In addition, regulatory genes may be needed to control the amount of capsule produced. Analysis of the cps9V locus identified and characterized in this work indicates that the locus encodes proteins that cover all these functions. In addition, two genes likely to be involved in nucleotide sugar biosynthesis were identified. These genes may enable the synthesis of substrates not provided by housekeeping genes.

The genetic organization of the cps9V locus shows strong resemblance to cps loci of other serotypes (Jiang et al., 2001; Paton & Morona, 2000). The locus is located between dexB and aliA. It contains the regulatory genes at the 5′ end, the genes that confer the sequential steps in the oligosaccharide biosynthesis and the assembly in the central part and the genes encoding modifying enzymes or enzymes for the synthesis of activated monosaccharide precursors at the 3′ end of the locus. Furthermore, the locus has a conserved σ70 promoter site just in front of the first gene. The locus carries several genes with a relatively low G+C content. The conserved nature of the 9V locus compared with sequenced loci from other serotypes feeds the concept that the 9V locus is part of a gene pool that contributes to the evolution of capsule diversity. Genetic exchange of capsule genes resulting in altered antigenic properties of the capsule have been reported both at the level of (nearly) complete loci and of individual genes (Barnes et al., 1995; Coffey et al., 1991, 1998a, b; Nesin et al., 1998; Ramirez & Tomasz, 1999). The low G+C content of some of the genes, about 29–34 mol%, compared to the mean for pneumococcal genes, 39.7 mol% (Tettelin et al., 2001), suggests that some cps genes may have originally been acquired from other species. Knowledge of the cps9V sequence may help to define the extent of

![Fig. 3. TLC of 14C-labelled intermediates in CPS synthesis in S. pneumoniae, isolated from membranes of serotypes 14, 9V, 9VΔcps9vE and 4. Intermediates were isolated from lipid preparations incubated with UDP-[14C]glucose and were released from lipid carriers by mild acid hydrolysis prior to TLC. The incorporation of 14C label in the glycolipid fraction of the membranes is indicated at the bottom in c.p.m. (µg protein).](image)

incorporation of glucose catalysed by serotype 14- and serotype 9V-derived membranes was much higher than by the mutant 9VΔcps9vE and serotype 4 fractions (Fig. 3).

Analysis by TLC of the intermediates formed demonstrated that serotype 14 and wild-type 9V membranes both conferred the transfer of glucose to the lipid carrier (Fig. 3). The mutant 9VΔcps9vE and serotype 4 showed only traces of glucose in this assay. As serotype 4 does not contain glucose in its repeating unit structure, the low but detectable levels of glucose likely reflect the incorporation into C-polysaccharide synthesis intermediates. Overall, the data are consistent with Cps9vE being a UDP-glycosyl-1-phosphate transferase.
the pneumococcal capsule gene repertoire and, ultimately, perhaps reconstruction of the evolution of the capsule serotypes.

At one point, the cps9V locus differs in organization from other known loci. In the 9V locus, the cps9vH gene, involved in nucleotide sugar biosynthesis, is located in the central part of the locus between the serotype-specific glycosyltransferase and assembly genes. In serotypes 4 and 19 with a cps9vH homologue, the gene is located at the end of the locus, adjacent to other nucleotide sugar biosynthesis genes. A second notable difference with other cps loci is the presence of a putative transposase-encoding gene on the opposite strand between two cps genes. The functional relevance of this gene (if any) remains to be determined.

The cps9V locus contains 15 ORFs. On the basis of their putative functions and current knowledge about capsule biosynthesis, the following sequence of reactions can be proposed for the biosynthesis of the 9V CPS (Fig. 4). First, a glucose is attached to the lipid carrier, possibly undecaprenyl phosphate (Paton & Morona, 2000; Whitfield & Roberts, 1999). This reaction is catalysed by Cps9vE as confirmed by the TLC results of the enzymic assay. The repeating unit of serotype 9V CPS contains two glucose residues, but based on the structural homology between serotypes 9V and 14 (which contains only one glucose residue), this reaction likely involves the glucose residue that will be linked to the ManNac residue. After formation of the glucose-lipid intermediate, the four additional glycosyltransferases, Cps9vF, Cps9vG, Cps9vJ and Cps9vL, sequentially add monosaccharides to the growing chain. On the basis of sequence similarities and the observation that genes in the cps loci of S. pneumoniae seem to be arranged in the order that the assembly of the repeating unit is expected to occur, we anticipate that Cps9vG, Cps9vJ and Cps9vL exhibit α-1,3-galactosyltransferase, α-1,3-glucuronic acid transferase and α-1,4-galactosyltransferase activities, respectively. Some of the substrates used in these reactions are available in the pneumococcus via the activity of the GlcNac-1-phosphate uridylyltransferase (GlmU) (Sulzenbacher et al., 2001), UTP-glucose-1-phosphate uridylyltransferase (GalU) (Mollerach et al., 1998) and UDP-glucose-4-epimerase (GalE) (Tettelin et al., 2001) that are located outside the 9V locus. The substrates UDP-ManNac and UDP-glucuronic acid, which do not seem to be part of 9V housekeeping, are synthesized with the help of the Cps9vH and Cps9vN proteins encoded by the 9V locus. To the completed oligosaccharide unit, additional O-acetyl groups may be variably attached to the ManNac, glucuronic acid and glucose residues (Fig. 2) (Perry et al., 1981; Rutherford et al., 1991) by the putative O-acetyltransferases Cps9vM and Cps9vO. After transport across the membrane facilitated by Cps9vK, the repeating units are likely polymerized via the Cps9vI polymerase. Finally, the CPS is linked to the peptidoglycan. The mechanism by which this step is established has not yet been determined.

ACKNOWLEDGEMENTS

We thank Linda Heijmen-van Dijk and Fons van Asten for technical assistance with automated sequencing.

REFERENCES


