**Functional organization of the gene cluster involved in the synthesis of the pneumococcal capsule**

**Summary** *Streptococcus pneumoniae* is a major human pathogen and its capsular polysaccharide has been shown to be the main virulence factor. The molecular organization of the genes governing the formation of this capsule was not studied until the 1990s. The capsular clusters (*cap*) of eight of the 90 known pneumococcal types have now been studied. The *cap* operon, located between the *dexB* and *aliA* genes, is arranged as a central region comprising the genes coding for the specific-type polysaccharide, flanked by open reading frames that are mostly common to all of the serotypes. The biochemical functions of 24 genes required for capsular polysaccharide biosynthesis have been elucidated but the precise role of the flanking regions in capsular formation is unknown. The natural genetic transformation characteristic of pneumococci, the arrangement of the *cap* locus and the abundance of transposable elements at this locus favor the genetic variability of the capsule in this microorganism. These well-documented observations together with the finding that some genes located outside the *cap* cluster may also participate in capsule formation increase the complexity of pneumococcal infection control.

**Key words** *Streptococcus pneumoniae* · Capsular polysaccharide · Virulence capsular switching · Natural transformation

**Introduction**

*Streptococcus pneumoniae* (pneumococcus) is the principal agent of community-acquired pneumonia which may be accompanied by bacteremia. In addition, this microorganism causes meningitis, otitis media, sinusitis, conjunctivitis, and endocarditis. Nasopharyngeal carriage of pneumococci is very common and newborns are colonized rapidly. Despite advances in chemotherapy and, in particular, the use of β-lactam antibiotics, morbidity and mortality rates for pneumococcal infections remain high worldwide. For example, pneumococcal pneumonia causes over one million deaths each year in children younger than 5 years [45] and, among the elderly, the case fatality rate for bacteremia approaches 40 percent. Management of pneumococcal infections used to be relatively straightforward, and penicillin was generally the antibiotic of choice. However, the emergence of antibiotic resistance among *S. pneumoniae* isolates has led to a change in approach [23]. The development of resistance in vitro to penicillin G of pneumococci following serial passages in the presence of the antibiotic and the stepwise transfer of resistance were reported soon after the introduction of penicillin. Penicillin-resistant pneumococci were first isolated in 1967 from individuals living in Australia and Papua-New Guinea. In 1978, fully resistant and multiple-drug resistant pneumococci were discovered in South Africa and, shortly afterwards, in many other countries [3] including Spain [16].

The pathogenicity of *S. pneumoniae* has been attributed to various factors such as the capsule, the autolytic LytA amidase and its cell wall target, the pneumococcal surface protein A (PspA), the pneumolysin (Pnl), and the IgA1 protease, among others [1]. Nevertheless, the capsule has long been known to be the major virulence factor of the pneumococcus. Griffith first reported that unencapsulated pneumococcal variants were virtually avirulent [21]. Moreover, a major reduction in virulence (by a factor of 10^5) is associated with capsule loss, and addition of a type-specific polysaccharide antiserum increases susceptibility to phagocytosis. It is recognized that the virulence of *S. pneumoniae* depends on the amount [32] and composition of the capsular polysaccharide [25] as well as the genetic background of the pneumococcal strain in which the capsule is synthesized [24].

The main function of the capsule is to prevent phagocytosis.
corresponding capsular polysaccharide as well as the
responsible for the synthesis of the sugar components of the
exchange of large pieces of DNA containing the genes
one capsular type to another) should, at least in principle, involve
transformation. Inter-type transformation (transformation from
chromosome and can be transferred together during
that capsular genes are closely linked in the
inducing capsular transformation [9]. It was subsequently shown
Avery and co-workers showed that DNA was responsible for
from which the vaccine had been prepared. Sixteen years later,
due to encapsulated pneumococci of the same type as the strain
inoculated mice subcutaneously with a mixture of live
formation of the pneumococcal types in 1928 [21] when he
the biosynthetic process. F. Griffith first described trans-
an in-depth knowledge of the genes (and proteins) involved in
predominate [42].

A possible alternative to antcapsular pneumococcal vaccines, is the use of drugs capable of inhibiting the synthesis of the polysaccharide. A rational approach to this aim requires an in-depth knowledge of the genes (and proteins) involved in the biosynthetic process. F. Griffith first described transformation of the pneumococcal types in 1928 [21] when he inoculated mice subcutaneously with a mixture of live unencapsulated pneumococci and a vaccine of heat-killed encapsulated pneumococci. The mice developed an infection due to encapsulated pneumococci of the same type as the strain from which the vaccine had been prepared. Sixteen years later, Avery and co-workers showed that DNA was responsible for inducing capsular transformation [9]. It was subsequently shown that capsular genes are closely linked in the S. pneumoniae chromosome and can be transferred together during transformation. Inter-type transformation (transformation from one capsular type to another) should, at least in principle, involve the exchange of large pieces of DNA containing the genes responsible for the synthesis of the sugar components of the corresponding capsular polysaccharide as well as the
transferases involved in the polymerization step (for a review, see [33]). Another interesting finding of the classic studies is that the capsule is synthesized from activated sugars, usually in the form of UDP-derivatives [34].

**Molecular organization of pneumococcal capsular genes**

The modern and powerful techniques of molecular biology were not used on pneumococcal capsule genes until the 1990s [20]. Type 3 was selected for molecular study for several reasons: (i) this serotype is the most frequent cause of invasive diseases in adults worldwide; (ii) type 3 polysaccharide is chemically simple and composed of alternating units of glucose (Glc) and glucuronic acid (GlcA) (Table 1); (iii) the colonial morphology of S3- mutants is easily distinguishable from that of encapsulated strains; and (iv) most of the classic genetic studies were carried out with this serotype. In 1993, we first succeeded in cloning a DNA fragment from a type 3 pneumococcus able to transform an unencapsulated (S3-) spontaneous mutant (strain M24) such that it produced the S3- capsule [18]. The sequence of the cloned fragment and surrounding regions showed that the unencapsulated M24 strain (as well as other spontaneous S3- mutants) harbored a mutation in a gene, cap3A, putatively encoding a UDP-glucose dehydrogenase (UDP-GlcDH) [6, 18]. This enzyme is responsible for the synthesis of UDP-GlcA from UDP-Glc, both sugar nucleotides being the precursors for type 3 capsule biosynthesis. Further studies provided new insight into the organization of the type 3 capsular gene cluster [5, 14, 15] (Fig. 1). The main conclusions raised by these studies were: (i) this cluster consists of three genes, namely cap3A, cap3B, and cap3C, that appear to be type-specific because they do not hybridize with DNAs prepared from any other serotype; (ii) the cap3 cluster is located in the S. pneumoniae chromosome between the genes dexB and aliA, which putatively encode, a glycosidase and an oligopeptide permease respectively, and are not required for capsular synthesis; (iii) the cap3ABC genes form an operon that is transcribed from a promoter located immediately upstream from cap3A; and (iv) between dexB and cap3A there is a region putatively coding for two open reading frames (ORF) (labeled as orf1 and orf2 in Fig. 1). In addition, two incomplete ORFs were also found in this region. Nevertheless, none of these ORFs appear to be involved in type 3 capsular polysaccharide biosynthesis because they are not transcribed in vivo [5].

Independently, Paton and co-workers worked out the organization of the genes involved in the biosynthesis of capsular polysaccharides of three different types belonging to serogroup 19, namely, 19F, 19A, and 19B (Table 1 and Fig. 1). In particular, pneumococci belonging to type 19F frequently cause pediatric infections (see above), and capsular
polysaccharides of types 19F and 19A are included in the 23-valent anti-pneumococcal vaccine. As shown in Fig. 1, the organization of the three capsular clusters as a whole presents a series of conserved features similar to those highlighted above for the cap3 operon. The genes responsible for type 19F and 19A biosynthesis (and presumably, also those for type 19B capsule formation) are located between \( eg B \) and \( aliA \) (sequence accession number AF094575, and references [22, 36–38]). However, the four ORFs (two incomplete) located downstream from \( dexB \), that are cryptic in type 3, together with the so-called specific genes, appear to form part of a single operon (in these cases, \( cps19f, cps19a, \) or \( cps19b \)). The four ORFs of the putative \( cps19 \) operons (genes \( A \) to \( D \)) are essentially identical (90–95% identical nucleotides) in the various capsular clusters sequenced so far, with the exception of one of the two clinical isolates of type 23F reported (sequence accession number AF057294, and reference [43]). Downstream from the fourth ORF of the cluster is a variable number of genes that have been shown to be more or less type-specific according to results obtained by Southern blot hybridization and direct sequencing of the DNA of various serotypes [29, 37]. A similar arrangement of capsular genes has also been reported for types 1 [39], 14 [28], 23F [43], and 33F [31]. However, direct experimental evidence for the presence of an active promoter located just upstream from the first gene of the \( cap/cps \) operon has only been provided for types 1 [39], 14 and 33F [31].

**Biochemical function of cap genes**

As illustrated in Fig. 2, biochemical evidence concerning the precise role of most cap genes is still lacking and the functions attributed to these genes have been suggested based solely on sequence comparisons (Table 2). The only capsular cluster that has been completely characterized is \( cap3 \). In addition to genetic evidence [4, 15, 18], the \( cap3A \) gene has been cloned in *Escherichia coli*, and the overproduced protein identified as a UDP-GlcDH [6]. \( Cap3B \) is the type 3 polysaccharide synthase, and it has been demonstrated that this is the only protein required to synthesize high molecular weight type 3 capsular polysaccharide in *S. pneumoniae* or *E. coli* strains, provided that they synthesize both UDP-Glc and UDP-GlcA [7]. However, mutational analysis of the \( cap3 \) genes revealed that only the \( cap3C \) gene was dispensable for capsule synthesis in all conditions tested [5, 15]. In addition, \( cap3C \) complements a \( galU \) mutation in *E. coli* and, therefore, codes for a UDP-Glc pyrophosphorylase [5].

Experimental evidence concerning the role of the four genes involved in the biosynthesis of the type 14 capsule has recently been provided. The \( cps14E \) gene has been demonstrated to code for a glucosyl-1-phosphate transferase [26] responsible for the first step of subunit synthesis for the type 14 capsule, the addition of glucose to a lipid carrier [27]. Experimental evidence suggesting that this step may also be the first in the biosynthesis of most of the capsular polysaccharides of *S. pneumoniae* containing glucose has recently been provided [29]. It is likely that the genes \( cps19f \) and \( cps19e \), which are almost identical to \( cps14e \), play the same role in the biosynthesis of their respective capsular polysaccharides. The \( cps14G, cps14I, \) and \( cps14J \) genes code for a \( \beta1,4-\)
galactosyltransferase, a β-1,3-N-acetylglucosaminyl transferase, and a β-1,4-galactosyltransferase, respectively, implicated in the second, third, and fourth steps of type 14 capsule biosynthesis [27, 28]. The gene cps19fK (and its homolog cps19bK) is likely to be a UDP-N-acetylglucosamine-2-epimerase involved in the synthesis of UDP-N-acetylmannosamine [37], the precursor of one of the sugar components of the serogroup 19 repeating unit (Table 1). The genes cps19fLMNO and their homologs in other capsular clusters (see Fig. 1 and 2) encode enzymes involved in dTDP-rhamnose (dTDP-Rha) biosynthesis (Glc-1-P-thymidylyl transferase; dTDP-4-dehydrorhamnose 3,5-epimerase; dTDP-Glc 4,6-dehydratase; and dTDP-4-dehydrorhamnose reductase) as deduced from the finding that they can substitute for *Shigella flexneri* rfbBDAC [37].

The gene cps19fK (and its homolog cps19bK) is likely to be a UDP-N-acetylglucosamine-2-epimerase involved in the synthesis of UDP-N-acetylmannosamine [37], the precursor of one of the sugar components of the serogroup 19 repeating unit (Table 1). The genes cps19fLMNO and their homologs in other capsular clusters (see Fig. 1 and 2) encode enzymes involved in dTDP-rhamnose (dTDP-Rha) biosynthesis (Glc-1-P-thymidylyl transferase; dTDP-4-dehydrorhamnose 3,5-epimerase; dTDP-Glc 4,6-dehydratase; and dTDP-4-dehydrorhamnose reductase) as deduced from the finding that they can substitute for *Shigella flexneri* rfbBDAC [37].

Genetic complementation of cap3A mutants and biochemical characterization of the gene expressed in *E. coli* have shown that Cap1K is a UDP-GlcDH [39]. It should be emphasized that cap1K is only 65% identical to cap3A, which encodes the type 3-specific UDP-GlcDH, suggesting that the two genes may have been introduced into *S. pneumoniae* from different sources because no more than 5% divergence has been found for other pneumococcal genes. In fact, phylogenetic analysis has indicated that Cap1K is equally distant from Cap3A and from other UDP-GlcDH such as, for instance, *Streptococcus pyogenes* HasB, *E. coli* KfiD, and *S. flexneri* Udg [40]. Very recently, we have characterized Cap1J as a UDP-galacturonate 4-epimerase, the first enzyme of this class to be characterized at the molecular level [41]. This enzyme is involved in the interconversion of UDP-GlcA and UDP-galacturonic acid (UDP-GalA). It should be kept in mind that UDP-GalA is the precursor of GalA, one of the components of the type 1 capsular polysaccharide (Table 1). However, four ORFs virtually identical to cps19fLMNO have been found in all type 1 *S. pneumoniae* clinical isolates tested (Fig. 1) despite
the fact that type 1 polysaccharide does not contain Rha (Table 1) [39]. As none of these genes, orf1 through orf4, (Fig. 1) are needed for capsular biosynthesis in type 1 pneumococci they may be a nonfunctional remnant of the capsular gene cluster of the ancestor of type 1 strains. This observation provides clues to the clonal origin of capsular types in S. pneumoniae.

Other genes involved in capsule biosynthesis

With the exception of cap3C, all of the type-specific capsular genes reported above appear to be required for capsular polysaccharide biosynthesis in S. pneumoniae. Nevertheless, other genes located outside the capsular gene cluster may also be involved in capsule formation. Watson and co-workers claimed that a region of the pneumococcal chromosome located downstream from lytA, the gene coding for the main autolytic enzyme of S. pneumoniae, is essential for the formation of the type 3 capsule [47]. However, a pneumococcal strain with a long deletion encompassing the lytA gene and its surrounding regions was easily transformed to give the type 3 capsule, which demonstrated that lytA and the DNA region located downstream from this gene are not involved in the encapsulation of S. pneumoniae [19].

As most pneumococcal capsules contain Glc and/or galactose (Gal) or various derivatives, in addition to other sugars [46], UDP-Glc must play a central role in the biosynthesis of the capsule. In addition, UDP-Glc is required for the interconversion of Gal and Glc via the Leloir pathway [17]. As reported above, cap3C codes for a UDP-Glc pyrophosphorylase that catalyzes the formation of UDP-Glc from Glc-1-P and UTP. No gene putatively coding for a similar enzyme has been found in any of the capsular gene clusters thus far investigated (Fig. 1 and 2, and Table 2).
these premises strongly suggest that another gene may also encode a UDP-Glc pyrophosphorylase that may also be common to all pneumococci. Such a gene, designated galU, has recently been cloned, sequenced, and overexpressed in *E. coli* [35]. Knockout galU mutants of type 1 and type 3 pneumococci are unable to synthesize a detectable capsule. Presumably, a similar result would be obtained for other capsule types. As eukaryotic UDP-Glc pyrophosphorylases appear to be completely unrelated to their prokaryotic counterparts, it is conceivable that GalU may be an appropriate target in the search for new drugs to control the pathogenicity of pneumococci and other bacteria.

### Genetic variability in *S. pneumoniae* capsular genes

The striking diversity in the capsular gene clusters of *S. pneumoniae* studied to date, in which high levels of similarity between genes involved in specific functions is exceptional, strongly suggests the existence of underlying mechanisms of genetic exchange. As *S. pneumoniae* is naturally transformable, capsule type changes may take place as a consequence of transformation events involving homologous recombination mediated by the conserved DNA regions that flank the type-specific genes. This has been shown to be true both in in vitro and in vivo conditions. It has been reported that type 19B transformants are recovered after transformation of a recipient type 19F strain with a 10.5-kb PCR product containing the *cps19b* genes and the flanking regions homologous to some of the *cps19f* genes (Fig. 1). However, in the last few years, several researchers have reported that some clinically relevant (multiresistant) pneumococcal strains are essentially identical in overall genotype but differ in capsule type. This finding has been interpreted as evidence that the new strains are the products of a capsular transformation event from one serotype to another. Very recently, Coffey and co-workers studied in detail eight type 19F variants that were otherwise identical to the major Spanish multiresistant 23F clone and confirmed that recombination at the *cap* locus had taken place on at least four occasions [12].

Transposition-like events may have also contributed to capsular diversity in *S. pneumoniae*. As shown in Fig. 1A, all of the capsular gene clusters of *S. pneumoniae* are flanked, on one or both sides, by insertion sequence (IS) elements. This is particularly evident in the region downstream from *dexB*, in which copies of IS1167 or IS1202 are frequently found either in active or in inactive form. An additional copy of the IS1167 element is also located immediately downstream from the *cap1G* operon followed by four cryptic ORFs that putatively code for enzymes involved in the synthesis of dTDP-Rha (see above). This finding suggests that the type 1 pneumococcal strains arose via a transposition event that took place in an ancestor strain that had Rha in its capsular polysaccharide [39]. Besides, the formation of binary encapsulated pneumococcal strains, that is, pneumococci possessing two different *cap* clusters and synthesizing two capsules, depends upon the presence of a copy of IS1167 in that this element provides the homology required for the insertion of the extra gene cluster into the *S. pneumoniae* chromosome [39]. A partial, incomplete copy of a IS1167-like element is located downstream from *cap3C* [5, 11]. Yother and co-workers [48] noted that upstream from the

---

### Table 2: Putative functions attributed to type-specific capsular gene products of *Streptococcus pneumoniae*

<table>
<thead>
<tr>
<th>Proposed function*</th>
<th>Gene product</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycosyl transferases</td>
<td>Cap1E</td>
</tr>
<tr>
<td>GalA transferase</td>
<td>Cps19aE, Cps23fE, Cap33fE</td>
</tr>
<tr>
<td>Glucosyl transferase</td>
<td>Cap23fH, Cap33fF, Cap33fG, Cap33fJ</td>
</tr>
<tr>
<td>Galactosyl transferase</td>
<td>Cps19f/a/fF</td>
</tr>
<tr>
<td>N-acetyl-mannosamine transferase</td>
<td>Cps19f/a/fH, Cps23fF, Cps23fI</td>
</tr>
<tr>
<td>Rhamnosyl transferase</td>
<td>Cap1G, Cap33fH</td>
</tr>
<tr>
<td>Other</td>
<td>Cap1F, Cap33fO</td>
</tr>
<tr>
<td>Acetyl transferases</td>
<td>Cps14L, Cps19f/a/fJ, Cap23fJ, Cap33fL</td>
</tr>
<tr>
<td>Repeating unit transporter</td>
<td>Cps14H, Cps19f/a/fJ, Cps23fH</td>
</tr>
<tr>
<td>Polysaccharide polymerase</td>
<td>Cap33fI</td>
</tr>
<tr>
<td>Others</td>
<td>Cps19fK</td>
</tr>
<tr>
<td>UDP-galactopyranose mutase</td>
<td>Cps23fK</td>
</tr>
<tr>
<td>UDP-N-acetylglucosamine 2-epimerase</td>
<td>Cps23fL</td>
</tr>
<tr>
<td>Glycero-phosphotransferase</td>
<td>Cps23fM</td>
</tr>
<tr>
<td>Glycero-2-phosphate dehydrogenase</td>
<td>Cps23fN</td>
</tr>
<tr>
<td>Unknown</td>
<td>Cps19f/a/f</td>
</tr>
</tbody>
</table>

*Based on sequence and/or structural similarities.
cap3A gene, there is a partial ORF similar to the IS-like H-rpt (Hinc) repeat elements of E. coli. This kind of sequence is also associated with polysaccharide biosynthetic loci in Gram-negative bacteria.

Perspectives

As shown in Fig. 2 and Table 2, only 24 of the 74 type-specific genes apparently required for capsular polysaccharide biosynthesis in the eight cap clusters deciphered in S. pneumoniae so far, have been biochemically characterized. Much work is required to define clearly the role of the remaining gene products. In particular, the proteins putatively involved in the transport of the repeating unit and the polymerization of the polysaccharide were identified purely on the basis of hydrophobicity plots because these proteins have no significant amino acid similarities. Hence, the question of how the growing polysaccharide chain is transported out of the cell and polymerized remains unanswered. Based solely on partial sequence similarities, a role for cap/cpsCD gene products in chain length regulation and export, and a regulatory role in capsule gene expression for Cap/CpsA have been repeatedly proposed [22, 27, 43]. However, the only experimental evidence of a relationship between these genes and capsule biosynthesis comes from insertion-duplication mutagenesis experiments [22]. As polar effects on the genes located downstream from the insertion cannot be ruled out, the real role of cap/cpsABCD in pneumococcal capsular biosynthesis remains to be elucidated. In addition, as reported above, it has been demonstrated that the corresponding ORFs in type 3 isolates are cryptic genes that are not translated [5] and, consequently, their involvement in type 3 capsule biosynthesis can be ruled out.

To date, the cap cluster of only 8 of 90 capsular types in S. pneumoniae has been studied. Much additional work is required before we can answer the question as to whether all of the genes involved in the biosynthesis of all pneumococcal capsular polysaccharides are arranged as described here. In particular, it should be kept in mind that sites for the insertion of capsular loci different from the usual particular, it should be kept in mind that sites for the insertion of capsular polysaccharides are arranged as described here. In

References

involved in the synthesis of the capsular polysaccharide of *Streptococcus pneumoniae* type 3. Mol Gen Genet 239:188–195


44. van Dam JEG, Fleer A, Snippe H (1990) Immunogenicity and immunochromy of *Streptococcus pneumoniae* capsular polysaccharides. A van Leeuwenhoek 58:1–47
