Characterization of the Cassette Containing Genes for Type 3 Capsular Polysaccharide Biosynthesis in *Streptococcus pneumoniae*

By Joseph P. Dillard, Mark W. Vandersea, and Janet Yother

From the University of Alabama at Birmingham, Birmingham, Alabama 35294

Summary

The capsular polysaccharide is the major virulence factor of *Streptococcus pneumoniae.* Previously, we identified and cloned a region from the *S. pneumoniae* chromosome specific for the production of type 3 capsular polysaccharide. Now, by sequencing the region and characterizing mutations genetically and in an in vitro capsule synthesis assay, we have assigned putative functions to the products of the type-specific genes. Using DNA from the right end of the region in mapping studies, we have obtained further evidence indicating that the capsule genes of each serotype are contained in a gene cassette located adjacent to this region. We have cloned the region flanking the left end of the cassette from the type 3 chromosome and have found that it is repeated in the *S. pneumoniae* chromosome. The DNA sequence and hybridization data suggest a model for recombination of the capsule gene cassettes that not only describes the replacement of capsule genes, but also suggests an explanation for binary capsule type formation, and the creation of novel capsule types.

Streptococcus pneumoniae produces a serotype-specific capsular polysaccharide that is required for virulence. The polysaccharide protects the organism from the host's immune system and prevents phagocytosis. More than 80 different serotypes have been identified (1), and the type of capsule expressed may influence virulence (2–4). Griffith's purpose in studying pneumococci was to determine if interconversion of type occurred in patients infected with more than one serotype and determine if this transformation affected the outcome of the infection (5). Today the importance of transformation for the organism's potential ability to infect its host is still not clear. Transformation could result in an antigenic shift, allowing an organism to colonize a previously immune host. Epidemiologic studies have suggested that a significant amount of genetic exchange does occur in nature, i.e., multiple combinations of antigenically variable determinants such as capsular serotype, PspA serotype, electrophoretic type, and penicillin binding protein type have been observed (6–9).

Classic experiments demonstrated that genes specific for the production of a given capsular polysaccharide are closely linked in the chromosome and can be transferred as a unit during genetic transformation (5, 10, 11). Except in rare cases, transformation to the new type results in loss of the ability to express the original polysaccharide (12). These data suggested that the capsule genes might be replaced in the recipient's chromosome through a cassette-type recombination mechanism. Biochemical characterization of strains transformed to new types supported this hypothesis. Strains that produced a particular nucleotide sugar as an intermediate in capsule synthesis no longer produced the nucleotide sugar after being transformed to a new type that did not require it (12). These results, though suggestive, did not conclusively show exchange of the capsule genes since regulatory models could not be ruled out. Recently, we demonstrated that switching of capsular type by genetic transformation in vitro does result in replacement of the type-specific genes (13). However, the mechanism of recombination resulting in exchange of capsule type has not been fully elucidated, nor is it known how the multiple different capsule types have evolved.

Type 3 capsule synthesis has been characterized biochemically and genetically. At least two functions are necessary for its production: the synthesis of the precursors UDP-glucose (UDP-Glc)[1] and UDP-glucuronic acid (UDP-GlcA), and their polymerization into the polysaccharide. Some 25 distinct mutations resulting in a reduction or loss of capsule synthesis were mapped to a single locus thought to encode UDP-Glc dehydrogenase (UDP-Glc DH), the enzyme necessary for conversion of UDP-Glc to UDP-GlcA (14). Characterization of the nucleotide-sugar pools of several mutants revealed a loss of UDP-GlcA. In an in vitro polymerization reaction, a partially purified extract from a type 3 strain was used to produce type 3 polysaccharide (15). UDP-Glc and UDP-GlcA were the only molecules that served as substrates,

[1] *Abbreviations used in this paper:* Glc-1-P UT, glucose-1-phosphate uridylyltransferase; GlcNAc, *N*-acetyl glucosamine; HA, hyaluronic acid; IPTG, isopropyl-β-D-thiogalactoside; RBS, ribosome binding site; UDP-Glc DH, UDP-Glc dehydrogenase; UDP-Glc, UDP-glucose; UDP-GlcA, UDP-glucuronic acid.

 J. Exp. Med. © The Rockefeller University Press • 0022-1007/95/03/0973/11 $2.00
Volume 181 March 1995 973–983

and Glc and GlcA were incorporated in equal amounts. The enzyme copurified with the particulate (membrane and cell wall) fraction of cell extracts (16), and Mg^{2+} was required for activity (17).

We previously localized the chromosomal region necessary for type 3 synthesis through mutation and cloning analyses (13). We have now sequenced this region and have further characterized the genes and their products. Our results provide molecular, genetic, and biochemical evidence for the roles of the gene products involved in type 3 capsule biosynthesis. They also suggest mechanisms for the transfer of capsule type-specific cassettes and the emergence of new capsule types.

Materials and Methods

Bacterial Strains. The parent *S. pneumoniae* type 3 strain WU2 and the type 1, 5, 6B, 8, 9, and 22 *S. pneumoniae* strains have been described (4, 13, 18). Other strains are described in the table and figures. Culture conditions for *S. pneumoniae* and *Escherichia coli* have been described (13, 19).

DNA Analysis. Denatured plasmid DNA was sequenced by the Sanger dideoxy method using the Sequenase 2.0 kit (US Biochemicals, Cleveland, OH). Oligonucleotide primers (Oligos, etc., Wilsonville, OR) 5′-GCCACTATCGACTACGCG-3′ and 5′-TCATTTTGATATGCCTCCG-3′, corresponding to bp 308 to 325 and 445 to 428 of the cloning vectors pJY4163 and pJY4164 (20), respectively, were used routinely. Primers internal to the *cps* locus were used as necessary. PCR products were sequenced at least twice, from separate amplification reactions, using a PCR product-sequencing kit (US Biochem. Corp., Cleveland, OH). Greater than 97% of the sequence was obtained for each strand. Sequences were analyzed using the GCG software programs (21).

The digoxigenin labeling and chemiluminescent detection system (Boehringer Mannheim, Indianapolis, IN) was used in Southern blotting. All other DNA manipulations were performed as previously reported (13).

In Vitro Polysaccharide Synthesis. Type 3 polysaccharide was synthesized and quantitated in vitro using a modification of Smith et al. (16). Crude extracts containing membranes and cytoplasm were prepared from 200 ml of *S. pneumoniae* cultures harvested at an OD_{600} of 0.25 as described (22), except that cell material was concentrated 200-fold, and all steps were performed using a thioglycolate buffer (10 mM sodium thioglycolate, 5 mM $MgSO_4$, 100 mM Tris–HCl, pH 8.3) to stabilize the enzymes (23). Digestion of cell wall material by mutanolysin was performed in this buffer and 20% sucrose. Protoplasts were sonicated three times for 15 s at 0°C. Polysaccharide synthesis was carried out at 34°C for 2 h in a 1-ml reaction containing 100 μl of extract, 5 mM UDP-Glc, 5 mM UDP-GlcA (where indicated), and 1 mM NAD in the thioglycolate buffer. The reaction was boiled 1 min then quickly cooled to 25°C in H_2O. Following centrifugation for 30 s at 8160 *g*, the type 3-specific mAb 16.3 (24) was added in excess to the supernatant and incubation was continued at 37°C for 30 min. Specific Ag-Ab complexes were measured at 650 nm. Capsule was quantitated by comparison with a standard curve prepared using purified type 3 polysaccharide (American Type Culture Collection, Rockville, MD) (17). Reactions, done in triplicate, were standardized to protein content of the crude extract, as determined in duplicate using a protein assay kit (Bio Rad Labs., Hercules, CA).

Expression of Cps3S. A 2.1-kb Sau3AI-PstI fragment containing the 3′ end of *cps3D* and the complete *cps3S* was cloned into the expression vector pKK223-3 (25) to yield pJD424. *E. coli* TG-1 (26) or TG-1 transformants were grown to exponential phase, and isopropyl-β-D-thiogalactoside (IPTG) was added to a concentration of 1 mM to induce expression from the *tac* promoter of pKK223-3.

Chromosome Crawling and PCR. To isolate the 5′ end of *cps3D* and upstream DNA, *S. pneumoniae* WU2 chromosomal DNA was first digested with Ecl136 II (an isoschizomer of SacI that results in blunt ends) and separated on a 0.6% agarose gel. Purified fragments from 6 to 7 kb were ligated to a 35-bp XbaI UniAmp adaptor (Clontech, Palo Alto, CA). The desired fragment was amplified by using a primer for the adaptor and a primer corresponding to the predicted active site sequence (bp 1802 to 1781) of *cps3D*. A 1.8-kb PCR product extending from the active site to the SacI site upstream of *cps3D* (see Fig. 1) was obtained. PCR amplifications were performed using AmpliTaq DNA polymerase (Perkin-Elmer Corp., Norwalk, CT). In a similar manner, the 0.9-kb fragment from the *cps3D* active site to the EcoRV site upstream of *cps3D* was amplified from a 3.5-kb EcoRV fragment from the WU2 chromosome.

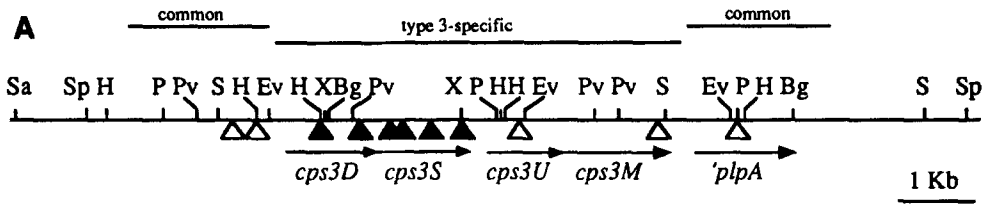
Nomenclature. The capsule locus of *S. pneumoniae* has been designated *cps* (13). To distinguish between loci of different capsular serotypes, the locus name will be followed by the number of the serotype, e.g., type 3 is indicated as *cps3*. The genes of the type 3 locus are named based on expected function (13).

Results

We previously described a region of the *S. pneumoniae* chromosome that contains genes involved in the production of type 3 capsular polysaccharide, and that is specific to type 3 strains. An ~1-kb fragment of DNA flanking this region and common to apparently all capsular serotypes was also identified (13). A genetic and physical map of the region is presented in Fig. 1 *A*. The DNA and deduced amino acid sequences of the region containing the genes *cps3D*, *cps3S*, *cps3U*, and upstream flanking DNA were determined in the present study, and are presented in Fig. 1 *B*.

Cps3D Is Homologous to UDP-glucose Dehydrogenase. Our previous genetic data indicated that we had cloned the 3′ end of *cps3D*, the gene encoding UDP-Glc DH (13). The DNA sequence and derived amino acid sequence support this assignment. The amino acid sequence is highly homologous (56% identity, 73% similarity) to that of the UDP-Glc DH (HasB) from *Streptococcus pyogenes* (Fig. 2) (27), and to open reading frames from the *E. coli* and *Salmonella enteritica rfb* clusters (28). Although not shown biochemically or genetically to be UDP-Glc DH, these latter sequences share a high degree of homology with HasB and Cps3D.

The NH_2-terminal amino acid residues 2–29 of CpsD have all the characteristics of an NAD-binding site (29), and this sequence is very homologous to regions from HasB, AlgD (the GDP-mannose dehydrogenase of *Pseudomonas aeruginosa* [30]), and the potential UDP-Glc DH from *E. coli* and *S. enteritica*. The homology with AlgD was previously noted by García et al. in the deduced amino acid sequence of the *S. pneumoniae* gene *cap3-1* (31). They suggested that Cap3-1 was the type 3 UDP-Glc DH. From the EcoRV site to our ScaI site (bp 883 to 1377, containing amino acids 1–117), our



B

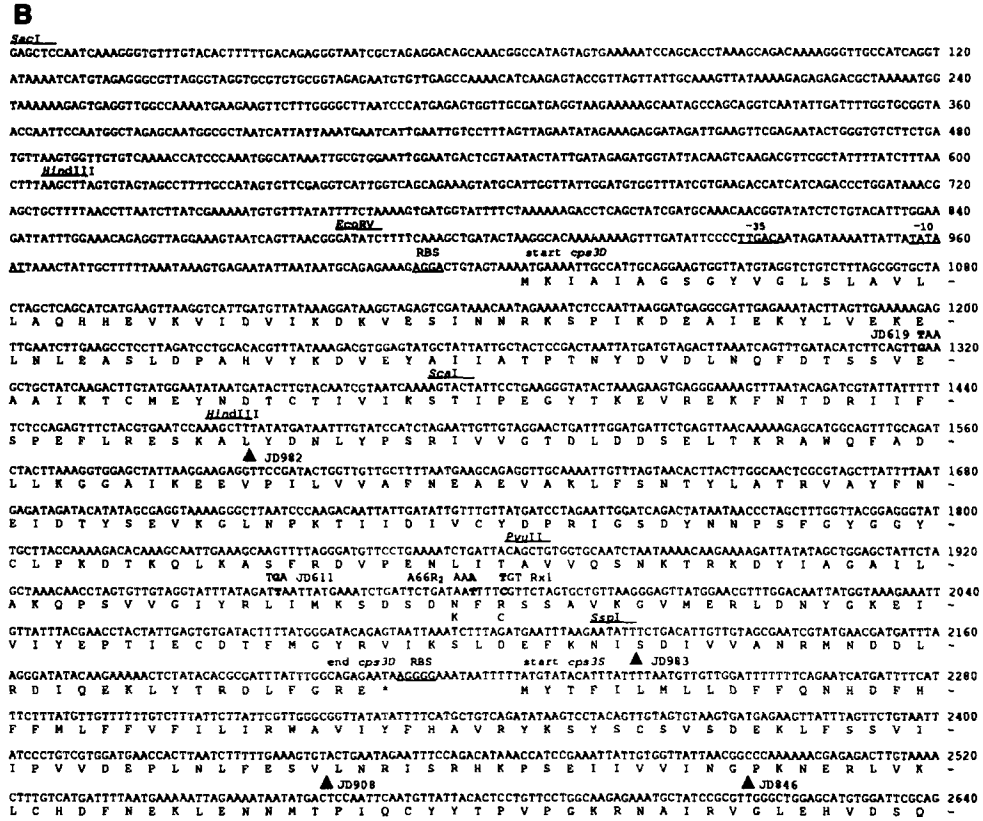


Figure 1. (*A*) Map of the type 3 capsule locus. Triangles indicate the endpoints of insertion mutations: filled, loss of capsule production; open, no apparent effect on capsule production. Restriction sites: *Bg*, BglII; *Ev*, EcoRV; *H*, HindIII; *P*, PstI; *Pv*, PvuII; *S*, SacI; *Sa*, SalI; *Sp*, SphI. (*B*) DNA sequence of the region containing *cps3D*, *cps3S*, and *cps3U*, and upstream flanking DNA. Putative promoters were identified using the FIND program and scored as in (52). A region of dyad symmetry upstream of *cps3U* is overlined. Endpoints of insertion mutations shown in *A* are indicated by triangles and are labeled with the name of the strain containing the mutation. Point mutations in *cps3D* are labeled with the sequence of the mutation and the name of the strain containing the mutation. Sequencing of the PvuII-SspI fragment of $A66R_2$ began at bp 1921, thus additional mutations between the PvuII site and this point are possible. Selected restriction sites are shown. These sequence data are available from EMBL/GenBank/DDBJ under accession number U15171.

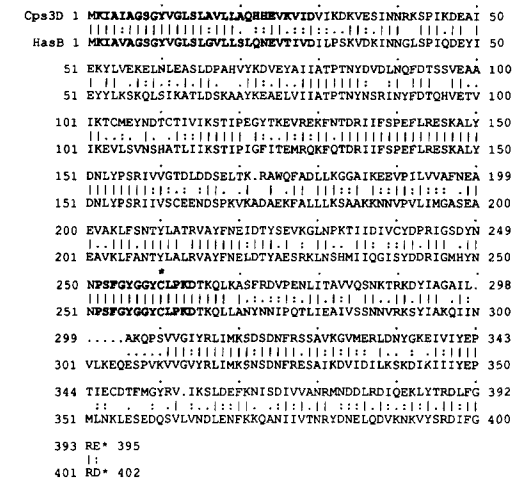


Figure 2. Homology of the Cps3D deduced amino acid sequence with HasB. The NAD-binding region (residues 2 to 29) and the active site (residues 251 to 263) are indicated in bold. An asterisk marks the essential cysteine of the active site.

sequence is in complete agreement with theirs. However, no other homology was seen, suggesting that these investigators had cloned only the 5′ end of the gene.

The Cps3D sequence at residues 251–263 is consistent with this being the active site of the enzyme. This region is identical to that of HasB (Fig. 2) and the putative *E. coli* and *S. enteritica* enzymes. The homology of the active site region of HasB with that of bovine UDP-Glc DH and AlgD has been described (27). The cysteine at residue 259 of Cps3D contains the essential thiol group of the reactive site (32). The predicted size of Cps3D (45 kD) is also similar to that of the *E. coli* enzymes (47 kD) (33).

Sequences of cps3S and cpsU. The region downstream of *cps3D* contains a second gene, *cps3S*, that is required for type 3 capsule synthesis (13). A 1,248–bp open-reading frame is transcribed in the same direction as *cps3D* and is in the same reading frame. This direction is in agreement with that previously determined using *cat* insertions (13). Only 15 bp separate a potential start codon for Cps3S from the stop codon of Cps3D. The sequence AGGGG just upstream of the putative start codon (Fig. 1 *B*) may serve as a ribosome binding site (RBS), or due to the close proximity of *cps3D*, no RBS may be necessary. The predicted Cps3S is 48 kD in size, if the first start codon at bp 1 (bp 2227 in Fig. 1 *B*) is used. Other potential start codons are located at bp +19 (2245) and +61 (2287), however neither of these are positioned near a RBS.

A short region of dyad symmetry was detected downstream of *cps3S* at bp 3718–3738. The scores for primary and secondary structure (p = 3.95, s = 22; GCG TERMINATOR program) suggest that this region could function as a weak rho-independent terminator. However, this sequence is 241 bp past the Cps3S stop codon and closer to the start of Cps3U. A potential *cps3U* promoter was detected upstream of the region of dyad symmetry, suggesting the structure could serve as an attenuator of *cps3U* expression. The *cps3U* open reading frame, 918 bp in length, is transcribed in the same direction as *cps3D* and *cps3S*, and is predicted to encode a protein of 34 kD.

Cps3S Is Homologous to Polysaccharide Synthases. The predicted Cps3S protein has significant homology to polysaccharide synthases, including HasA, the hyaluronic acid (HA) synthase from *S. pyogenes* (23% identity, 50% similarity) (34, 35), and NodC from *Rhizobium meliloti* (21% identity, 47% similarity) (Fig. 3). HA and the pneumococcal type 3 capsule are composed of β(1-4)-linked repeating disaccharide units containing GlcA. HA consists of alternating *N*-acetyl glucosamine (GlcNAc) and GlcA residues. Both contain β(1-3) and β(1-4) linkages, however the linkage to GlcA is β(1-4) in HA but β(1-3) in type 3 capsule (36). NodC is necessary for the

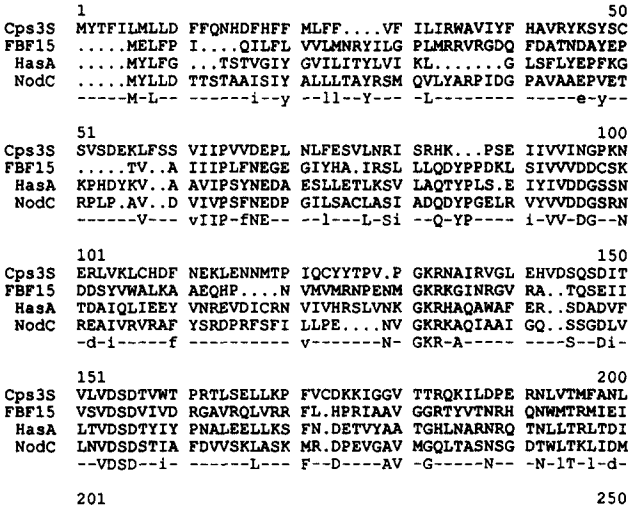


Figure 3. Bacterial polysaccharide synthases. The alignment was performed using PILEUP. The consensus sequence was prepared using PRETTY and a plurality of 3/4. Lower case letters indicate that the consensus amino acid is representative of a class.

synthesis of nodulation factor, a substituted oligosaccharide consisting of β(1-4) linked GlcNAc residues (37). It has been noted that HasA and NodC are homologous to polysaccharide synthases, including FBF15 of *Stigmatella aurantiaca*, pDG42 of *Xenopus laeveis*, and chitin synthases from both *Saccharomyces cerevisiae* and *Candida albicans* (34, 35, 38, 39). Cps3S is also homologous to these proteins. The results suggest that Cps3S is the type 3 capsular polysaccharide synthase.

Four hydrophobic stretches identified in Cps3S are found in all four bacterial synthases. As indicated for NodC, these regions may span the cell membrane (40, 41). Earlier studies indicated that the type 3 synthesizing activity also has a membrane location (16). The last hydrophobic stretch in Cps3S may be required for function since the insertion in JD897 which eliminated this region (the last 45 amino acids of the protein) resulted in loss of capsule production (Fig 1 *B*). High level expression of Cps3S in *E. coli* was, like that of NodC, lethal to the host (data not shown).

Cps3U Is Homologous to Glucose-1-phosphate Uridylyltransferases. The amino acid sequence of Cps3U has a high degree of homology with glucose-1-phosphate uridylyltransferases (Glc-1-P UT) from several other bacterial species, including GtaB from *Bacillus subtilis* (55% identity, 73% similarity) (Fig. 4 *A*). The active sites of bacterial Glc-1-P UT have not been characterized, however the active site in the enzyme from potato tuber has 5 lysine residues (42). One of these residues is important for function, and a second is absolutely required (43). Cps3U contains 24 lysines, six of which are conserved among the six bacterial Glc-1-P UT in the database (Fig. 4 *A*). Only one region from Cps3U containing a conserved lysine can be aligned well with the potato tuber enzyme sequence. It is homologous to the region containing the required lysine (Fig. 4 *B*).

Figure 4. (*A*) GAP comparison of Cps3U with GtaB. The six lysines conserved among the bacterial Glc-1-P UT are indicated in bold. (*B*) Alignment of a region of Cps3U with the region from the potato tuber UDP-Glc pyrophosphorylase (Glc-1-P UT) required for enzyme function.

Mutations Affecting Type 3 Capsular Polysaccharide Production. Two mutations resulting in the capsule-negative phenotype were previously mapped to *cps3D* (13). To determine the nature of these mutations, the regions were amplified from the chromosomes of the mutant strains and sequenced. Each mutant (JD611 and JD619) contained a single base pair transversion resulting in a premature stop codon (Fig. 1 *B*).

We also identified three mutations upstream of the JD611 and JD619 mutations, near a site where a mutation apparently affecting type 3 capsule polymerization had previously been mapped (13, 14). These mutants produced capsular material that was detectable in cell lysates, but the cells could not be agglutinated with type 3-specific antisera, and they produced non-mucoid colonies. We tentatively identified the locus containing these mutations as *cpsB*, and, based on the phenotypes and map positions, suggested that *cpsB* might encode a polysaccharide polymerase (i.e., be *cpsP*) (13). To better localize the mutations, we amplified fragments from the parent type 3 chromosome that contained either the 5′ end of the *cps3D* coding sequence (bp 1027–1802), the promoter and the 5′ end of *cps3D* (bp 885–1802), or the 5′ end of *cps3D* plus ~1 kb of upstream DNA (bp 1–1802). Each of the fragments was used to transform the capsule-deficient mutants JD614 and JD692. JD692 was transformed to encapsulation using the 5′ end coding sequence, whereas JD614 was not. JD614 was restored by the fragment containing the 5′ end plus 141 bp of upstream DNA, including the promoter. Both mutants were restored by the 1.8-kb fragment containing the 5′ end of *cps3D* and upstream DNA, and neither was restored with a fragment containing the 3′ end of *cps3D* (bp 1759–2385, data not shown). Thus, these upstream mutations are not located in a separate gene but are in either the *cps3D* structural gene or its promoter. Since some capsule material is produced by these mutants, a mutation within the coding region (as in JD692) must be a missense mutation or an in-frame deletion or insertion that reduces the activity of the enzyme. The mutation in JD614 may be in the promoter, and thus, a promoter down mutation, or it may be in the structural gene but too close to the beginning of the gene for recombination and repair to occur with the fragment used.

Two other mutations in *cps3D* resulting in reduced amounts of type 3 capsule were restored to normal encapsulation with the 250-bp PvuII-SspI fragment from the parent strain (13). Amplification and sequencing of this same fragment from the mutant strains A66R$_2$ and Rx1 showed that each contained a missense mutation in the *cps3D* coding sequence (Fig. 1 *B*).

Use of fragments subcloned from the *cps3DSU* region to direct insertion-duplication mutations in the parent type 3 chromosome (13) resulted in several mutants that produced no detectable capsule and exhibited the extremely rough phenotype described by Taylor (44). DNA sequencing revealed that the mutations are in *cps3S* (Fig. 1). The lack of capsule in these mutants must be due to loss of *cps3S* expression, rather

than to a polar effect on downstream genes, since insertions within *cps3U* or *cps3M* had no apparent effect on capsule production, as judged by growth on blood agar medium.

cps3S and cps3D Are Transcribed as an Operon. Sequence analysis revealed no potential promoter sequences in the region upstream of *cps3S* (Fig. 1 *B*). The phenotypes of several insertion-duplication mutants also suggest that no promoter is located in the 3′ end of *cps3D* and that *cps3S* is transcribed from the *cps3D* promoter. The sites of these insertions are shown in Fig. 1 and their structures are illustrated in Fig. 5. Insertion of the plasmids results in a duplication of the cloned fragment. Therefore, mutant strains such as JD908, in which the duplicated fragment contains both the 5′ end of *cps3S* and the 3′ end of *cps3D*, have a full-length copy of *cps3S* downstream of the plasmid insertion. In addition, the full-length copy is contiguous to the 3′ end of *cps3D*. Therefore, if *cps3S* had its own promoter, or if one were located in the 3′ end of *cps3D*, these insertions should not result in loss of *cps3S* expression. However, four such insertions have been made in the WU2 chromosome (JD846, JD897, JD898, and JD908), and even with a duplication of 450 bp of the 3′ end of *cps3D*, a loss of capsule production was observed.

Two internal insertions in *cps3D* also eliminated capsule production (Fig. 5 *A*). However, since *cps3D* and *cps3S* are transcribed as an operon, this result does not prove that *cps3D* is required for capsule synthesis. That fact is demonstrated by the lack of capsule production seen in strains containing non-polar point mutations in *cps3D* (see below).

In Vitro Polymerization Assay. It is not possible to perform the standard UDP-Glc DH assay on extracts of *S. pneumoniae* due to the presence of a NADH oxidase, that copurifies with the enzyme (23, 45). Therefore, the ability of the mutants to synthesize type 3 capsule was examined in an in vitro polymerization assay. Mutants JD611 and JD619, which contain stop mutations in *cps3D*, produce no detectable capsular material (13). However, both synthesized high molecular weight type 3 polysaccharide in vitro when provided with UDP-Glc and UDP-GlcA (Table 1). No capsule was produced when UDP-GlcA was omitted from the reaction. These results support the conclusion that Cps3D is the UDP-Glc DH, and confirm that stop mutations in *cps3D* are not polar on *cps3S*.

Mutants containing insertions in *cps3S* (JD902), or between the full-length copies of *cps3D* and *cps3S* (JD908, JD897) were unable to synthesize significant amounts of capsule even with

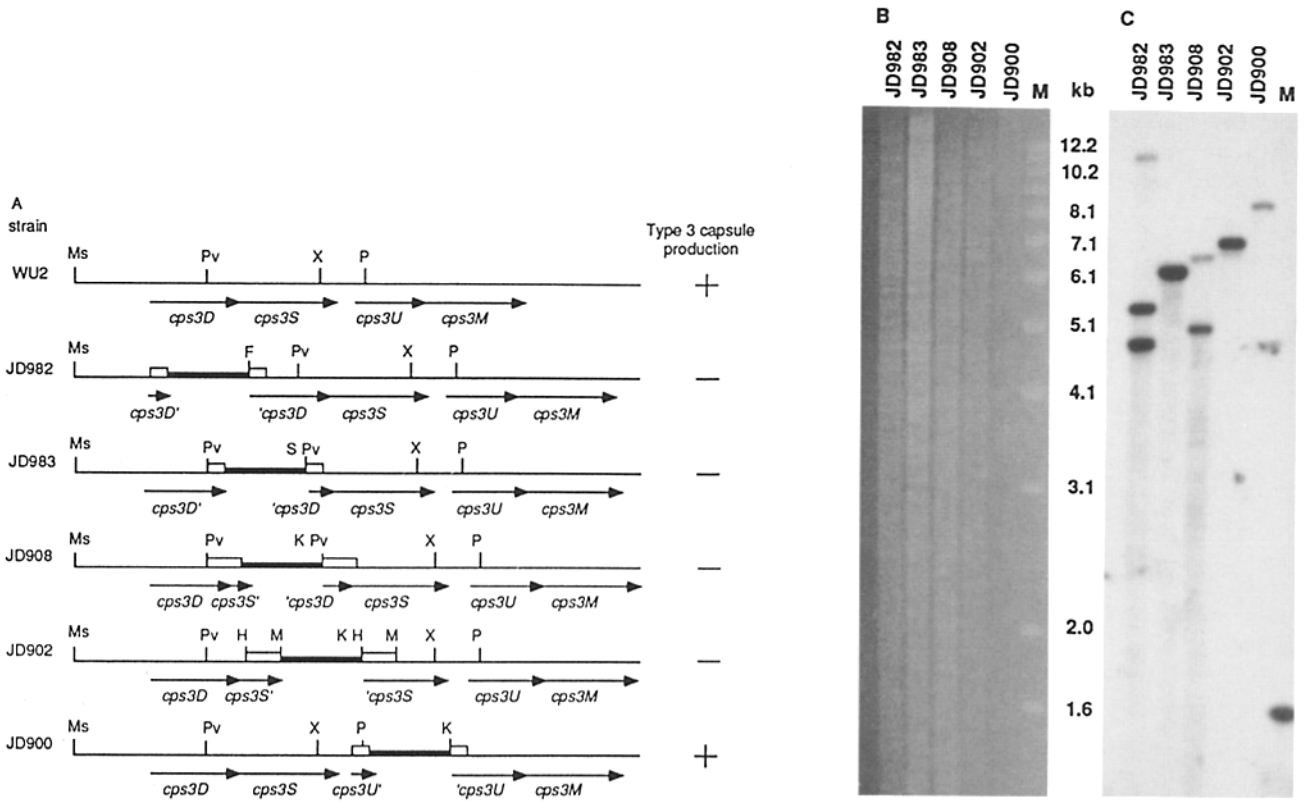


Figure 5. Location of insertion mutations in the type 3-specific region. (*A*) Schematic illustration of the insertions. The schematic was derived from Southern blot analysis such as that shown in *B* and *C*. Restriction sites: *F*, FspI; *H*, HindIII; *K*, KpnI; *Ms*, MscI; *P*, PstI; *Pv*, PvuII; *X*, XbaI. (*B*) Ethidium bromide-stained agarose gel of chromosomal DNA from insertion mutants digested with MscI/FspI for JD982, MscI/SalI for JD983, and MscI/KpnI for JD908, JD902, and JD900. (*C*) Southern blot of gel in part B probed with the vector pJY4164. Increasing distance from the MscI site to the end of the vector is demonstrated by the increase in size of the upper band. The faint band in the JD982 lane is likely a result of partial digestion. The 4.7- and 4.8-kb bands in JD982 and JD908, respectively, indicate that these mutants contain a duplication of the inserted plasmid. The vector is homologous to the 1.6-kb fragment of the molecular size standards, *M*.

Table 1. *In Vitro Capsule Synthesis Assay*

Strain	Cps phenotype‡	UDP-GlcA§	CPS (μg/mg protein)
JD611	Cps3D⁻S⁺	+	9.8 ± 0.6
		–	0.9 ± 0.2
JD619	Cps3D⁻S⁺	+	5.7 ± 0.3
		–	0.2 ± 0.1
JD614	Cps3D*S*	NA‖	5.4 ± 0.4 (t_o)¶
		+	5.9 ± 0.5 (0.5)¶
JD692	Cps3D*S*	NA	4.8 ± 0.3 (t_o)
		+	7.0 ± 1.0 (2.2)
JD902	Cps3D⁺S⁻	+	1.7 ± 0.3
JD908	Cps3D⁺S⁻	+	1.5 ± 0.1
JD897	Cps3D⁺S⁻	+	1.1 ± 0.1
WU2	Cps3D⁺S⁺	NA	3.8 ± 0.2 (t_o)
		+	16.6 ± 0.3 (12.8)
		–	16.3 ± 0.8
D39**	Cps2⁺	+	0.5 ± 0.3

‡ Capsule phenotypes are based on the *cps3D* and *cps3S* genotypes. * Indicates either a missense or in-frame deletion or insertion in *cps3D* that apparently also affects *cps3S*. – indicates either a stop or insertion mutation (see Fig. 1 *B* and 5 *A* for locations of mutations).
§ The presence or absence of UDP-GlcA in the reaction is indicated by a + or –.
‖ NA, not applicable.
¶ For strains that produce capsule in vivo, the amount of polysaccharide present at the start of the assay (t_o) is given, and the amount of polysaccharide produced during the assay is indicated in parentheses.
** D39 is a type 2 strain and thus produces no type 3 capsule.

both precursors present. These results emphasize the role of Cps3S in capsule synthesis and support the conclusion that *cps3D* and *cps3S* are transcribed as an operon.

The capsule-deficient mutants JD614 and JD692 synthesized only small amounts of additional polysaccharide in the in vitro assay. This result is somewhat surprising since JD692, which contains a missense mutation in the *cps3D* coding region, should still make a functional Cps3S (i.e., the *cps3D* mutation must not be polar since intact cells synthesize some polysaccharide). The result may suggest that the defective UDP-Glc DH interferes with the ability to synthesize normal polysaccharide. Alternatively, the stability of the *cps3DS* transcript may be altered by the mutation, resulting in a reduced amount of Cps3S.

Biochemical Pathway. Based on genetic analysis, amino acid homologies, biochemical and immunochemical characterization of mutants, and previous biochemical analyses (12, 13, 15–17), we propose a pathway for the biosynthesis of type 3 capsular polysaccharide (Fig. 6). The last of the type 3-specific genes, *cps3M*, is homologous to phosphoglucomutases from several bacterial species and is described in a forthcoming publication (Caimano, M., J. P. Dillard, and J. Yother, manuscript in preparation).

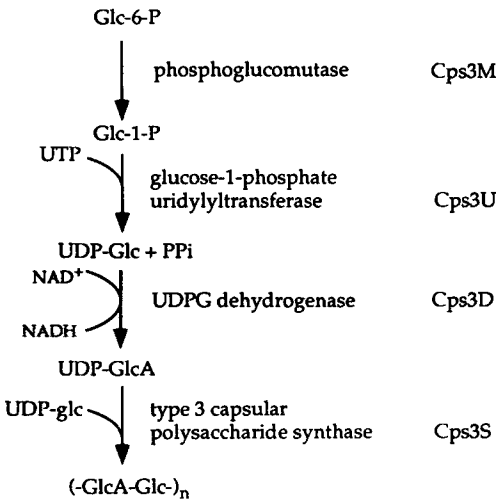


Figure 6. Biosynthetic pathway for type 3 capsular polysaccharide. Functions of proteins encoded by the type 3-specific genes are shown. Additional functions may be necessary for capsule transport or attachment.

Mapping Other Capsule Types. We previously showed that a 1.2-kb SacI-HindIII fragment flanking the type 3 capsule locus contains sequences common to apparently all capsule types (13). Sequence analysis showed that the fragment contained the 3′ end of *cps3M* and the 5′ half of a gene with 50% identity to the *S. pneumoniae amiA*. The *amiA*-like sequence has recently also been identified by Pearce et al. and named *exp1* (46), and subsequently renamed *plpA* (47). Further Southern hybridizations showed that the non–type-specific homologous DNA in the 1.2-kb SacI-HindIII fragment is *plpA* (data not shown).

The homologous fragment is closely linked not only to the type 3-specific capsule genes, but also to the type-specific genes of types 2, 5, and 6B (reference 13 and our unpublished data). Mapping studies using this fragment showed that, as in type 3, it is directly adjacent to the type-specific genes of other serotypes. The chromosome maps of the capsule regions in strains of types 2, 3, and 6B are highly conserved for at least 4 kb to the right of *plpA* (Fig. 7). The type 3 strain differs slightly in this region due to a deletion of the 5′ end of *plpA*. The sites located to the left of *plpA* are divergent among the capsule types, further suggesting that this region contains the type-specific genes in all three capsule types.

The Region Upstream of the Type 3-specific Genes Is Common to All Capsule Types and Is Repeated in the Pneumococcal Chromosome. To isolate DNA 5′ of the biosynthetic genes, a 1.8-kb fragment extending from the upstream SacI site to the *cps3D* active site was amplified from the type 3 WU2 chromosome. When this fragment was used to probe HindIII-digested chromosomal DNA from seven *S. pneumoniae* serotypes, multiple bands were detected in all strains (Fig. 8). When chromosomal DNAs of types 2, 3, and 6B were digested with PstI, PvuII, or SacI/HindIII, and probed with the cloned 610-bp SacI-HindIII fragment upstream of *cps3D* (Fig. 1), 4-10 bands were detected in each (data not shown).

Transformation experiments were performed to examine

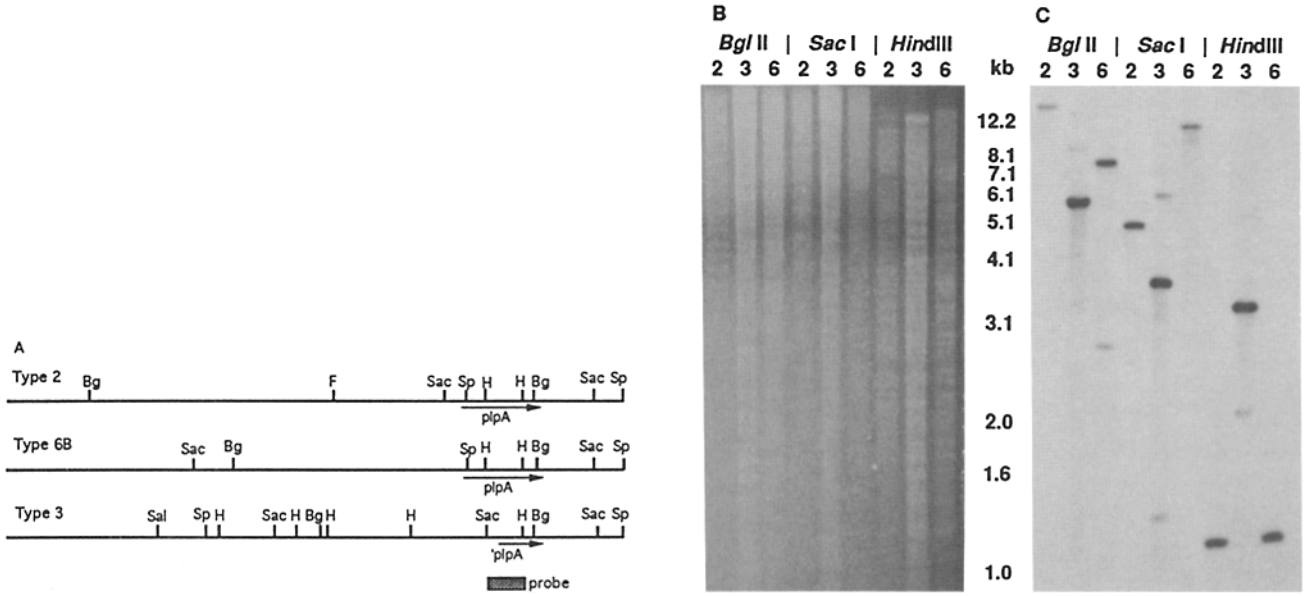


Figure 7. (*A*) Chromosome maps of the capsule regions in types 2, 3, and 6B. The 1.2-kbSacI-HindIII fragment (pJD377) from type 3 used for the probe is shown below the maps. Restriction sites are *Bg*, BglII; *F*, FspI; *H*, HindIII; *S*, SalI; *Sac*, SacI; *Sp*, SphI. (*B*) Ethidium bromide-stained agarose gel and (*C*) Southern blot showing chromosomal DNA from strains of types 2, 3, and 6B probed with pJD377. Faint bands in addition to the band of interest may be due to the detection of fragments containing the *amiA*-like genes that have homology to *plpA*.

linkage of the upstream region to the type-specific genes. A plasmid (pJD392) containing the 610 bp SacI-HindIII fragment was introduced into the chromosome of the type 3 strain. The insert, located in the 2.2-kb HindIII fragment adjacent to the type 3-specific genes, did not affect capsule production. When the resulting strain was used to transform recipients of types 2 and 6B, greater than 95% of the erythromycin-resistant isolates expressed type 3 capsule. However, when pJD392 was transformed into strains of types 2 and 6B, the plasmid inserted into an 8-kb HindIII fragment, and the type-specific genes could not be moved to strains of heterologous types (i.e., 2, 3, or 6B) by transformation and selection for linkage to the erythromycin marker in the insertions.

Discussion

Based on genetic, molecular, and biochemical data we have assigned putative functions to the type 3-specific genes. Two of the genes, *cps3D* and *cps3S*, are required for capsule synthesis. Our previous genetic data, along with the sequence and biochemical analyses reported here, provide substantial evidence that *cps3D* encodes UDP-Glc DH. Cps3D is highly homologous to HasB, and contains sequences homologous to the active and the NAD-binding sites in HasB and other known UDP-Glc DH. Extracts from Cps3D mutants could synthesize type 3 capsule in vitro if supplied with UDP-GlcA, i.e., they lacked the ability to convert UDP-Glc to UDP-GlcA and thus lack UDP-Glc DH activity.

Cps3S is new member of a family of polysaccharide synthases. All of these synthases, for which the structures of the polysaccharides are known, produce β(1-4) linked polysaccharides. Thus, Cps3S may form the β (1-4) linkage in the type 3 disaccharide cellobiuronic acid (GlcA β[1-4] Glc), and a second enzyme may create the β(1-3) linkages required to polymerize the disaccharides into full length polysaccharide. However, HasA creates both linkages in the production of HA capsule (34), and is sufficient for HA synthesis in heterologous bacteria, given the nucleotide sugars (48). Therefore, Cps3S, like HasA, may synthesize the polysaccharide by monomer addition.

Neither *cps3U* nor *cps3M* appears to be required for type 3 synthesis. Cps3M and Cps3U should function to convert Glc-6-P into Glc-1-P, and Glc-1-P into UDP-Glc, respectively (Fig. 6). Since UDP-Glc is necessary for the production of essential cell constituents, including teichoic acid and lipoteichoic acid (12), the products of other genes may complement functions lost in the mutants. However, retention of these genes in the type-specific region may indicate that part of their function cannot be duplicated by the second enzymes. Possibly, this function is the ability to be regulated under specific conditions, such as those that might be encountered during infection. The large noncoding region upstream of *cps3U* might be a site of regulation. An alternative explanation is that these genes were obtained along with the type-specific genes in a horizontal transfer from another organism and have not been lost. This theory is consistent with hybridization data indicating that none of the type 3-specific genes could be detected in strains of six other pneumococcal types, including types with related capsule structures (13). However, if these genes serve no necessary function, it is surprising that they have been maintained in the type 3 cassettes

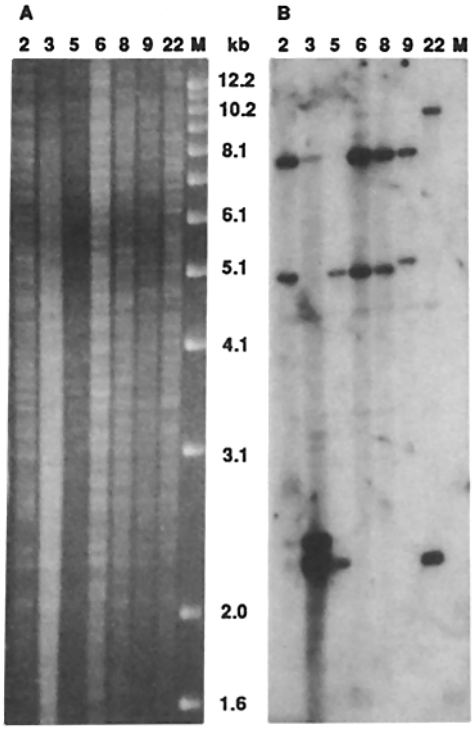


Figure 8. Repeated element. (*A*) Ethidium bromide-stained agarose gel and (*B*) Southern blot of *S. pneumoniae* chromosomal DNA from seven different serotypes digested with HindIII. The blot was probed with the 1.8-kb fragment containing the region from the SacI site upstream of the type-specific genes to just before the PvuII site in *cps3D* (bp 1 to 1802). Identical results were obtained using pJD392 (containing bp 1 to 610) as the probe.

of multiple strains; i.e., the restriction maps of the type 3 regions of five non-clonal strains are identical, and all have *cps3U* and *cps3M* (our unpublished data).

There are three requirements for a DNA region to be considered a gene cassette: (*a*) more than one copy of a gene or set of genes must exist, each specifying the production of a different, but related, product; (*b*) each copy must be flanked by DNA that is common to all the copies; and (*c*) cassettes must recombine to cause replacement of one copy by another. There is strong evidence to indicate that the type-specific genes are arranged as a cassette. First, the presence of more than 80 different serotypes implies that as many different sets of genes exist. Second, the type 2, 3, 5, and 6B type-specific genes are flanked to the right by a fragment common to apparently all types, and containing *plpA*. Although the left flanking region from type 3 is common to all capsule types we have examined, it may not flank the type-specific genes in other types. If not, then presumably other common DNA is located further upstream of these genes. The third requirement for a cassette is fulfilled by previous biochemical evidence (12) and our recent molecular evidence (13) demonstrating replacement of type-specific enzymes and genes, respectively, following transformation of capsule type.

Since the proposal was put forth that capsule genes are exchanged as cassettes, there has always been one glaring exception–binary encapsulation. At low frequency, strains of certain types transformed with DNA from strains of certain other types were found to produce both polysaccharides (12). Stable binary strains contained the second set of type-specific genes at a site unlinked to the recipient's type-specific genes. Once integrated, these genes could not be moved to the normal location in a strain of heterologous type. These observations led Bernheimer et al. to suggest that recombination involved strong homology at only one end (49, 50). Unstable binary strains frequently lost the donor type-specific genes, which were usually linked to the recipient type-specific genes (49, 51).

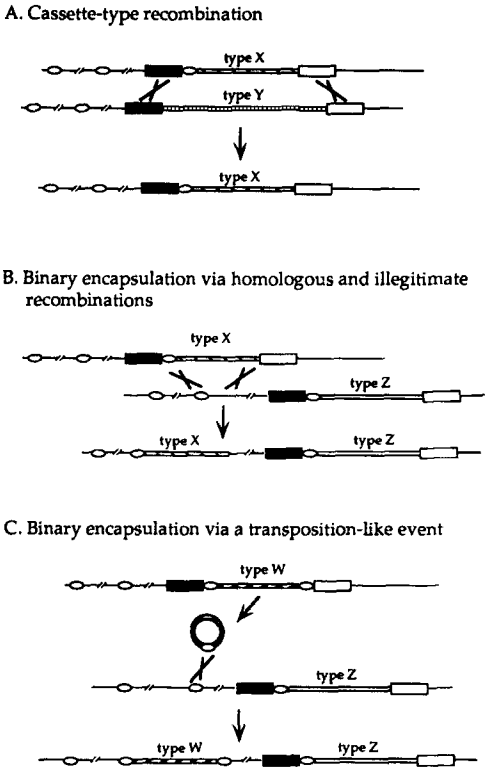


Figure 9. Models for the transfer of type-specific genes. (*A*) Cassette type-recombination. Replacement of the recipient's type-specific genes with those of the donor results from homologous recombination between regions that flank the type-specific genes and are common to all types. ⬭, sequence containing repeated element; ■, common DNA upstream of type-specific cassettes; ▭, common DNA (including *plpA*) downstream of type-specific cassettes. (*B*) Binary encapsulation by recombination involving homology at only one end. Homologous recombination at one end of the type-specific cassette would occur through the repeated element. Integration at the other end would result from an apparent illegitimate recombination. Linkage of the two type-specific cassettes would result if the integration occurred in a repeat element in or closely linked to the recipient's capsule genes. (*C*) Binary encapsulation via a transposition-like event. Type-specific cassettes flanked by the repeated element would resolve out of the chromosome and be transferred to recipient cells as circular intermediates. Recombination into the recipient chromosome could occur at a repeat element unlinked (as shown) or linked to the recipient's type-specific genes. Transfer of linear DNA could also yield binary strains as a result of recombination with two repeats that are closely linked in the recipient chromosome.

Based on our data and the extensive work of Bernheimer concerning binary encapsulation (14, 49–51), we can propose models for capsule type change and binary capsule formation. Cassette-type recombination would result from crossover events in the homologous flanking regions, leading to replacement of the type-specific genes (Fig. 9 *A*). Binary encapsulation may be mediated through the repeated element identified upstream of the type 3 capsule genes. Linkage analysis showed that at least one copy of the repeat is unlinked to the type 2 and 6B type-specific genes. In type 3, one copy is linked but, based on transformation experiments, another is not (our unpublished observation). The mechanism proposed by Bernheimer et al. for stable binary strains could involve homologous recombination at a repeat element unlinked to the capsule locus; the recombination at the other end of the capsule genes would occur by an apparent illegitimate recombination event (Fig. 9 *B*). An alternative possibility involves a transposition-like event that could result if certain type-specific genes are flanked on both sides by the repeat element (Fig. 9 C). Unstable binary strains could result from either type of integration occurring at repeated elements in, or closely linked to, the recipient's type-specific genes. Instability could result from recombination through genes common to both capsule types, as suggested by Bernheimer et al., for the UDP-Glc DH of types 1 and 3. Our present results provide the basis for examining these possibilities. Binary strains containing the two sets of genes linked are of particular interest since they might recombine to form a novel capsule type. Examination of strains producing related capsule structures may help elucidate the possible mechanisms involved in novel capsule type formation.

We thank Melissa Caimano for her assistance and helpful discussions regarding *cps3M* and *plpA*.

This work was supported by Public Health Service grants AI-28457 and T32 AI-07041 from the National Institutes of Health.

Address correspondence to Dr. Yother, 661 Bevill Biomedical Research Bldg., University of Alabama at Birmingham, Birmingham, AL 35294.

Received for publication 25 August 1994 and in revised form 1 November 1994.

References

1. van Dam, J.E.G., A. Fleer, and H. Snippe. 1990. Immunogenicity and immunochemistry of *Streptococcus pneumoniae* capsular polysaccharides. *Antonie Leeuwenhoek.* 58:1–47.
2. Walter, A.W., V.H. Guerin, M.W. Beattie, H.Y. Cotler, and H.B. Bucca. 1941. Extension of the separation of types among pneumococci: description of 17 types in addition to types 1 to 32 (Cooper). *J. Immunol.* 41:279–294.
3. Briles, D.E., M.J. Crain, B.M. Gray, C. Forman, and J. Yother. 1992. Strong association between capsular type and virulence for mice among human isolates of *Streptococcus pneumoniae. Infect. Immun.* 60:111–116.
4. Kelly, T., J.P. Dillard, and J. Yother. 1994. Effect of genetic switching of capsular type on virulence of *Streptococcus pneumoniae. Infect. Immun.* 62:1813–1819.
5. Griffith, F. 1928. The significance of pneumococcal types. *J. Hyg.* 27:119–159.
6. Crain, M.J., D. Waltman, II, J.S. Turner, J. Yother, D.F. Talkington, L.S. McDaniel, B.M. Gray, and D.E. Briles. 1990. Pneumococcal surface protein A (PspA) is serologically variable and is expressed by all clinically important capsular serotypes of *Streptococcus pneumoniae. Infect Immun.* 58:3293–3299.
7. 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.
8. Sibold, C., J. Wang, J. Henrichsen, and R. Hakenbeck. 1992. Genetic relationships of penicillin-susceptible and -resistant *Streptococcus pneumoniae* strains isolated on different continents. *Infect. Immun.* 60:4119–4126.
9. Versalovic, J., V. Kapur, E.O. Mason Jr., U. Shah, T. Koeuth, J.R. Lupski, and J.M. Musser. 1993. Penicillin-resistant *Streptococcus pneumoniae* strains recovered in Houston: identification and molecular characterization of multiple clones. *J. Infect. Dis.* 167:850–858.
10. Dawson, M.H. 1930. The transformation of pneumococcal types. II. The interconvertibility of type-specific *S. pneumococci. J. Exp. Med.* 51:123–147.
11. Effrussi-Taylor, H. 1951. Genetic aspects of transformation of pneumococci. *Cold Spring Harbor Symp Quant. Biol.* 16:445–456.
12. Austrian, R., H.P. Bernheimer, E.E.B. Smith, and G.T. Mills. 1959. Simultaneous production of two capsular polysaccharides by pneumococcus. II. The genetic and biochemical bases of binary capsulation. *J. Exp. Med.* 110:585–602.
13. Dillard, J.P., and J. Yother. 1994. Genetic and molecular characterization of capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* type 3. *Mol. Microbiol.* 12:959–972.
14. Bernheimer, H.P., I.E. Wermundsen, and R. Austrian. 1968. Mutation in Pneumococcus type III affecting multiple cistrons concerned with the synthesis of capsular polysaccharide. *J. Bacteriol.* 96:1099–1102.
15. Smith, E.E.B., G.T. Mills, H.P. Bernheimer, and R. Austrian. 1960. The synthesis of type III pneumococcal capsular polysaccharide from uridine nucleotides by a cell-free extract of *Diplococcus pneumoniae* type III. *J. Biol. Chem.* 235:1876–1880.
16. Smith, E.E.B., G.T. Mills, and H.P. Bernheimer. 1961. Biosynthesis of pneumococcal capsular polysaccharides: I. Properties of the system synthesizing type III capsular polysaccharide. *J. Biol. Chem.* 236:2179–2182.
17. Bernheimer, A.W. 1953. Synthesis of type III pneumococcal

polysaccharide by suspensions of resting cells. *J. Exp. Med.* 97:591–600.
18. Briles, D.E., M. Nahm, K. Schoroer, J. Davie, P. Baker, J. Kearney, and R. Barletta. 1981. Antiphosphocholine antibodies found in normal mouse serum are protective against intravenous infection with type 3 *Streptococcus pneumoniae. J. Exp. Med.* 153:694–705.
19. Dillard, J.P., and J. Yother. 1991. Analysis of *Streptococcus pneumoniae* sequences cloned into *Escherichia coli*: effect of promoter strength and transcription terminators. *J. Bacteriol.* 173: 5105–5109.
20. Yother, J., G.L. Handsome, and D.E. Briles. 1992. Truncated forms of PspA that are secreted from *Streptococcus pneumoniae* and their use in functional studies and cloning of the *pspA* gene. *J. Bacteriol.* 174:610–618.
21. Genetics Computer Group. 1991. Program Manual for the GCG Package, Version 7.
22. Yother, J., and J.M. White. 1994. Novel surface attachment mechanism of the *Streptococcus pneumoniae* protein PspA. *J. Bacteriol.* 176:2976–2985.
23. Smith, E.E.B., G.T. Mills, R. Austrian, and H.P. Bernheimer. 1960. Uridine pyrophosphoglucose dehydrogenase in capsulated and non-capsulated strains of pneumococcus type I. *J. Gen. Microbiol.* 22:265–271.
24. Briles, D.E., J.L. Claflin, K. Schroer, and C. Forman. 1981. Mouse IgG_3 antibodies are highly protective against infection with *Streptococcus pneumoniae. Nature (Lond.).* 294:88–90.
25. Brosius, J., and A. Holy. 1984. Regulation of ribosomal RNA promoters with a synthetic *lac* operator. *Proc. Natl. Acad. Sci. USA.* 81:6929–6933.
26. Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular Cloning. A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
27. Dougherty, B.A., and I. van de Rijn. 1993. Molecular characterization of *hasB* from an operon required for hyaluronic acid synthesis in group A streptococci. *J. Biol. Chem.* 268:7118–7124.
28. Bastin, D.A., G. Stevenson, P.K. Brown, A. Hasse, and P.R. Reeves. 1993. Repeat unit polysaccharides of bacteria: a model for polymerization resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. *Mol. Microbiol.* 7:725–734.
29. Wierenga, R.K., P. Terpstra, and W.G.J. Hol. 1986. Prediction of the occurrence of the ADP-binding $\beta\alpha\beta$-fold in proteins, using an amino acid sequence fingerprint. *J. Mol. Biol.* 187:101–107.
30. Deretic, V., J.F. Gill, and A.M. Chakrabarty. 1987. *Pseudomonas aeruginosa* infection in cystic fibrosis: nucleotide sequence of the *algD* gene. *Nucleic Acids Res.* 15:4567–4581.
31. García, E., P. García, and R. López. 1993. Cloning of a gene involved in the synthesis of the capsular polysaccharide of *Streptococcus pneumoniae* type 3. *Mol. & Gen. Genet.* 239:188–195.
32. Ridley, W.P., J.P. Houchins, and S. Kirkwood. 1975. Mechanism of action of uridine diphosphoglucose dehydrogenase: evidence for a second reversible dehydrogenation step involving an essential thiol group. *J. Biol. Chem.* 250:8761–8767.
33. Schiller, J.G., F. Lamy, R. Frazier, and D.S. Feingold. 1976. UDP-glucose dehydrogenase from *Escherichia coli*. Purification and subunit structure. *Biochem. Biophys. Acta.* 453:418–425.
34. DeAngelis, P.L., J. Papaconstantinou, and P.H. Weigel. 1993. Molecular cloning, identification, and sequence of the hyaluronan synthase gene from group A *Streptococcus pyogenes. J. Biol. Chem.* 268:19181–19184.
35. Dougherty, B.A., and I. van de Rijn. 1994. Molecular characterization of *hasA* from an operon required for hyaluronic acid synthesis in group A streptococci. *J. Biol. Chem.* 269:169–175.
36. Reeves, R.E., and W.F. Goebel. 1941. Chemoimmunological studies on the soluble specific substance of pneumococcus. V. The structure of the type III polysaccharide. *J. Biol. Chem.* 139:511–519.
37. Lerouge, P., P. Roche, C. Faucher, F. Maillet, G. Truchet, J.C. Pome, and J. Dénarié. 1990. Symbiotic host specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. *Nature (Lond.).* 344:781–784.
38. Atkinson, E.M., and S.R. Long. 1992. Homology of *Rhizobium meliloti* NodC to polysaccharide polymerizing enzymes. *Mol. Plant-Microbe Interact.* 5:439–442.
39. Debellé, F., C. Rosenberg, and J. Dénarié. 1992. The Rhizobium, Bradyrhizobium, and Azorhizobium NodC proteins are homologous to yeast chitin synthases. *Mol. Plant-Microbe Interact.* 5:443–446.
40. Johnson, D., L.E. Roth, and G. Stacy. 1989. Immunogold localization of the NodC and NodA proteins of *Rhizobium meliloti. J. Bacteriol.* 171:4583–4588.
41. John, M., J. Schmidt, U. Wieneke, H.D. Krüssman, and J. Schell. 1988. Transmembrane orientation and receptor-like structure of the *Rhizobium meliloti* common nodulation protein NodC. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:583–588.
42. Kazuta, Y., Y. Omura, M. Tagaya, K. Nakano, and T. Fukui. 1991. Identification of lysyl residues located at the substrate-binding site in UDP-glucose pyrophosphorylase from potato tuber: affinity labeling with uridine di- and triphosphopyridoxals. *Biochemistry.* 30:8541–8545.
43. Katsube, T., Y. Kazuta, K. Tanizawa, and T. Fukui. 1991. Expression in *Escherichia coli* of UDP-glucose pyrophosphorylase cDNA from potato tuber and functional assessment of the five lysyl residues located at the substrate binding site. *Biochemistry.* 30:8546–8551.
44. Taylor, H.E. 1949. Additive effects of certain transforming agents from some variants of pneumococcus. *J. Exp. Med.* 89:399–424.
45. Smith, E.E.B., G.T. Mills, H.P. Bernheimer, and R. Austrian. 1958. The formation of uridine pyrophosphoglucuronic acid from uridine pyrophosphoglucose by extracts of a noncapsulated strain of pneumococcus. *Biochim. Biophys. Acta.* 28:211–212.
46. Pearce, B.J., Y.B. Yin, and H.R. Massure. 1993. Genetic identification of exported proteins in *Streptococcus pneumoniae. Mol. Microbiol.* 9:1037–1050.
47. Pearce, B.J., A.M. Naughton, and H.R. Massure. 1994. Peptide permeases modulate transformation in *Streptococcus pneumoniae. Mol. Microbiol.* 12:881–892.
48. DeAngelis, P.L., J. Papaconstantinou, and P.H. Weigel. 1993. Isolation of a *Streptococcus pyogenes* gene locus that directs hyaluronan biosynthesis in acapsular mutants and in heterologous bacteria. *J. Biol. Chem.* 268:14568–14571.
49. Bernheimer, H.P., I.E. Wermundsen, and R. Austrian. 1967. Qualitative differences in the behavior of pneumococcal deoxyribonucleic acids transforming to the same capsular type. *J. Bacteriol.* 93:320–333.
50. Bernheimer, H.P., and I.E. Wermundsen. 1972. Homology in capsular transformation reactions in *Pneumococcus. Mol. & Gen. Genet.* 116:68–83.
51. Bernheimer, H.P., and I.P. Wermundsen. 1969. Unstable binary capsulated transformants in pneumococcus. *J Bacteriol.* 98:1073–1079.
52. Harr, R., M. Häggström, and P. Gustafsson. 1983. Search algorithm for pattern match analysis of nucleic acid sequences. *Nucleic Acids Res.* 11:2943–2957.