The *Staphylococcus aureus* allelic genetic loci for serotype 5 and 8 capsule expression contain the type-specific genes flanked by common genes

Subrata Sau,1 Navneet Bhasin,2 Elisabeth R. Wann,3 Jean C. Lee,2 Timothy J. Foster2 and Chia Y. Lee1

THE NUCLEOTIDE SEQUENCES OF TWO GENE CLUSTERS, **cap5** AND **cap8**, INVOLVED IN THE SYNTHESIS OF *Staphylococcus aureus* TYPE 5 AND TYPE 8 CAPSULAR POLYSACCHARIDES (CPs), RESPECTIVELY, WERE DETERMINED. EACH GENE CLUSTER CONTAINED 16 ORFS, WHICH WERE NAMED **cap5A** THROUGH **cap5P** FOR TYPE 5 CP AND **cap8A** THROUGH **cap8P** FOR TYPE 8 CP. THE **cap5** AND **cap8** LOCI WERE ALLELIC AND WERE MAPPED TO THE SmaI-G FRAGMENT IN THE STANDARD SmaI MAP OF *Staph. aureus* STRAIN NCTC 8325. THE PREDICTED GENE PRODUCTS OF **cap5A** THROUGH **cap5P** AND **cap5L** THROUGH **cap5P** ARE ESSENTIALLY IDENTICAL TO THOSE OF **cap8A** THROUGH **cap8G** AND **cap8L** THROUGH **cap8P**, RESPECTIVELY, WITH VERY FEW AMINO ACID SUBSTITUTIONS. FOUR ORFs LOCATED IN THE CENTRAL REGION OF EACH LOCUS ARE TYPE-SPECIFIC. A COMPARISON OF THE PREDICTED AMINO ACID SEQUENCES OF **cap5** AND **cap8** WITH SEQUENCES FOUND IN THE DATABASES ALLOWED TENTATIVE ASSIGNMENT OF FUNCTIONS TO 15 OF THE 16 ORFs.

THE MAJORITY OF THE CAPSULE GENES ARE LIKELY TO BE INVOLVED IN AMINO SUGAR SYNTHESIS; THE REMAINDER ARE LIKELY TO BE INVOLVED IN SUGAR TRANSFER, CAPSULE CHAIN-LENGTH REGULATION, POLYMERIZATION AND TRANSPORT.

**Keywords**: polysaccharide, capsule, *Staphylococcus aureus*, **cap5** and **cap8** gene clusters

INTRODUCTION

*Staphylococcus aureus* is an important human and animal pathogen. More than 90% of clinical isolates of this bacterium produce capsular polysaccharides (CPs), which have been classified into 11 serotypes (Karakawa & Vann, 1982; Sompolinsky et al., 1985). Serotype 1 and 2 strains of *Staph. aureus* are highly encapsulated (Wilkinson, 1983) but are rarely isolated from any source. Type 1 and type 2 CPs have been shown to be important virulence factors (Lee et al., 1987; Lin et al., 1994; Melly et al., 1974; Peterson et al., 1978). Most *Staph. aureus* strains elaborate microcapsules smaller than those produced by strains of serotype 1 and 2. Strains producing type 5 and type 8 capsules account for about 80% of clinical isolates (Albus et al., 1988; Arbeit et al., 1984; Hochkeppel et al., 1987; Karakawa et al., 1985; Poutrel et al., 1988). THE ROLE OF TYPE 5 AND TYPE 8 CAPSULES IN VIRULENCE REMAINS CONTroversial (Albus et al., 1991; Baddour et al., 1992; Nemeth & Lee, 1995; Karakawa et al., 1988; Xu et al., 1992). Nonetheless, recent studies (Fattom et al., 1996; Lee et al., 1996) have shown that type 5 capsules may be the target of antibodies that protect against experimental *Staph. aureus* infections.

The biochemical structures of type 1 CP (CP1), type 2 CP (CP2), type 5 CP (CP5) and type 8 CP (CP8) have been determined; each contains hexosaminouronic acid sugars (Fournier et al., 1984; Moreau et al., 1990; Murthy et al., 1983; Hanessian & Haskell, 1964). As shown below, CP5 and CP8 are very similar and differ only in the position of O-acetyl groups and the linkages between the amino sugars. CP1, CP5 and CP8 also have a common sugar, N-acetylfucosamine (2-acetamido-2,6-dideoxygalactose). The repeating units are:

**CP1**: \(\alpha-d-GalNAcAp-(1\rightarrow4)-\alpha-d-GalNAcAp-(1\rightarrow3)-\alpha-d-FucNAcAp-(1\rightarrow\)
Genetic studies of the CPs of Staph. aureus have only recently been reported. The cap1 gene cluster required for CP1 synthesis has been cloned and sequenced (Lee, 1992; Lin et al., 1994), and the cap5 and cap8 gene clusters have been cloned and partially characterized (Lee et al., 1994; Sau & Lee, 1996; Sau et al., 1997). Southern hybridization studies have revealed that the cap5 and cap8 gene clusters have two common regions flanking a type-specific region; this finding suggests that these two gene clusters are allelic. In type 1 strain M, in addition to the cap1 gene cluster, a cap locus with extensive homology to the cap8 gene cluster has been identified, indicating that the cap1 and cap8 loci are not allelic (Sau & Lee, 1996). In this communication, we report the mapping of the cap5(8) locus on the Staph. aureus chromosome and the nucleotide sequences of the cap5 and cap8 gene clusters. We show that the DNA sequences containing 12 of the 16 ORFs of the cap5 and cap8 gene clusters are almost identical. By analogy with homologous genes from other bacteria, we discuss the possible functions of these ORFs in CP5 and CP8 biosynthesis.

### METHODS

**Strains and growth conditions.** The bacterial strains and plasmids used are listed in Table 1. *Escherichia coli* strains were grown in Luria–Bertani broth or agar (Difco) and were used for propagating and harvesting plasmids. *Staph. aureus* strains Newman and Reynolds were used to clone and sequence the cap5 gene cluster. *Staph. aureus* strain Becker, which produces CP8, was used as the source of DNA for sequencing the cap8 gene cluster.

**DNA manipulations.** DNA manipulations were performed as described by Sambrook et al. (1989). Plasmid DNA was isolated by the method of Birnboim (1983) and further purified by CsCl/ethidium bromide density gradient centrifugation or by use of either the Wizard (Promega) or the Qiagen plasmid purification kit. Restriction enzymes and other enzymes used in the study were purchased from Gibco-BRL and New England Biolabs.

**DNA sequencing and analysis.** An ~12.5 kb region of the cap5 genes was subcloned from pJCL19 (carrying an ~34 kb contiguous fragment encompassing the cap5 genes from strain Reynolds; Lee et al., 1994) in vectors pGEM7Z(+) or pLIS0. An ~7.1 kb region of the cap8 genes was also cloned from strain Newman. Nested sets of deletions were generated with an Erase-a-base kit (Promega) and sequenced with an automated DNA sequencer (model 373A) and Taq DyeDeoxy Terminator Cycle sequencing kits (Applied Biosystems). To sequence the chromosomal locus containing cap8 genes, we used the cap8 gene clusters from other bacteria, we discuss the possible functions of these ORFs in CP5 and CP8 biosynthesis.

### Table 1. Bacterial strains and plasmids

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Relevant characteristics</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Staph. aureus strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8325</td>
<td></td>
<td>J. J. Iandolo, Kansas State University, Manhattan, KS, USA</td>
</tr>
<tr>
<td>M</td>
<td>Type 1 capsule strain</td>
<td>J. H. Hash, Vanderbilt University, Nashville, TN, USA</td>
</tr>
<tr>
<td>Smith</td>
<td>Type 2 capsule strain</td>
<td>ATCC 13709</td>
</tr>
<tr>
<td>Reynolds</td>
<td>Type 5 capsule strain</td>
<td>Karakawa &amp; Vann (1982)</td>
</tr>
<tr>
<td>Newman</td>
<td>Type 5 capsule strain</td>
<td>NCTC 8178</td>
</tr>
<tr>
<td>Becker</td>
<td>Type 8 capsule strain</td>
<td>Karakawa &amp; Vann (1982)</td>
</tr>
<tr>
<td><strong>E. coli strains</strong></td>
<td>recA13 hsdS20 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 supE44 rpsL20</td>
<td>J. Lutkenhaus, University of Kansas Medical Center, Kansas City, KS, USA</td>
</tr>
<tr>
<td>HB101</td>
<td>recA1 endA1 gyrA96 thi hsdR17 supE44 relA 1A(lac–proAB)/F’ (traD36 proAB’ lacI8 lacZAM15)</td>
<td>Yanisch-Perron et al. (1985)</td>
</tr>
<tr>
<td>JM109</td>
<td>recA1 endA1 gyrA96 thi hsdR17 supE44 relA 1A(lac–proAB)/F’ (traD36 proAB’ lacI8 lacZAM15 Tn10)</td>
<td>Sambrook et al. (1989)</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td>Cloning vector</td>
<td>Promega</td>
</tr>
<tr>
<td>pGEM7Z(+)</td>
<td>Cloning vector</td>
<td>Stratagene</td>
</tr>
<tr>
<td>pBluescript</td>
<td></td>
<td>Lee et al. (1991)</td>
</tr>
<tr>
<td>pLIS0</td>
<td>E. coli–Staph. aureus shuttle cloning vector</td>
<td>Lee et al. (1994)</td>
</tr>
<tr>
<td>pJCL19</td>
<td>34 kb fragment of strain Reynolds DNA cloned in BamHI site of pH79</td>
<td></td>
</tr>
</tbody>
</table>
several overlapping fragments from a contiguous 18.5 kb DNA fragment (Sau & Lee, 1996) were generated by restriction enzymes and subcloned into suitable sites in pBluescript KS(+) . Nested sets of deletions from each subclone were generated by exonuclease III. Sequencing was performed by the method of Sanger et al. (1977) with a sequencing kit from US Biochemical.

Sequences were assembled and analysed with the Wisconsin Genetics Computer Group software package. Protein homology searches from the databases were conducted with the BLAST network service at the National Center for Biotechnology Information according to the method of Altschul et al. (1990). Protein hydrophobicity was calculated by the method of Kyte & Doolittle (1982), with windows set at nine amino acids. Alignment of amino acids was performed with the CLUSTAL program (Higgins & Sharp, 1988). The method of Klein et al. (1985) was used for predicting potential transmembrane segments. In cases for which the analogous gene products from cap5 and cap8 genes were virtually identical, only the gene products from the cap5 gene cluster were included in protein alignments and hydrophobicity profile comparisons.

RESULTS AND DISCUSSION

Nucleotide sequences of the cap5 and cap8 gene clusters

We previously reported that all 28 Cap8- mutants derived from Staph. aureus strain Becker were complemented by plasmid subclones from a contiguous 26.5 kb DNA fragment of strain Becker (Sau & Lee, 1996). This result suggests that the majority of the cap8 genes affecting CP8 biosynthesis are clustered in this region of the chromosome. The nucleotide sequence of ~ 18.5 kb within the 26.5 kb DNA region was determined. Sixteen ORFs in a 17.5 kb region, designated cap8A through cap8P (Fig. 1), were found to be tightly clustered and transcribed in one orientation.

The cap5 locus was initially identified in strain Reynolds by Tn918 mutagenesis (Lee et al., 1994) and in strain Newman by Tn917 mutagenesis (E. R. Wann & T. J. Foster, unpublished). DNA fragments flanking the Tn918 insertion sites were used to screen a cosmids library prepared from wild-type strain Reynolds. Similarly, sequences flanking Tn917 in strain Newman were cloned and used as probes against a lambda library prepared from strain Newman. A 7.1 kb DNA segment encompassing cap5A through cap5G was sequenced from strain Newman, and a 12.6 kb DNA segment including cap5F through cap5P was sequenced from strain Reynolds. The region of overlapping sequence between the two strains was 1.55 kb in length and included most of cap5F and half of cap5G. Six nucleotide differences between the two strains were found in the overlapping segment, and five of these differences resulted in no change in the amino acid sequences of the deduced proteins. One variation between strains Reynolds and Newman was in amino acid 196 of Cap5G, which was shown to be serine in the former case and proline in the latter. At the equivalent position in Cap5G, the deduced amino acid was proline.

Sequence analysis of the combined 18.1 kb region from serotype 5 strains revealed 16 contiguous ORFs that were transcribed in the same orientation and were named cap5A through cap5P (Fig. 1). Comparison of the nucleotide sequences of the cap5 and cap8 genes revealed that each of the gene clusters could be divided into three regions. Regions 1 and 3 were highly homologous with 98.1% and 98.6% identities, respectively. In contrast, the central regions showed insignificant homology (less than 43%). The overall G+C content in the coding regions of the cap5 and cap8 loci was 32.6 mol% and 32.9 mol%, respectively — values typical of the Staph. aureus genome (Oeding, 1983). However, a lower G+C content of the sequences of the four ORFs located in the type-specific regions (average of 29.1 mol% for cap5H to cap5K and 28.5 mol% for cap8H to cap8K) suggested that these type-specific genes could have been derived from other organisms with a lower G+C DNA content.

Chain-length determination genes

Cap5A and Cap8A show significant homology to EpsC of Streptococcus thermophilus (Stingele et al., 1996) and...
<table>
<thead>
<tr>
<th>ORF</th>
<th>Size*</th>
<th>Homologous protein†</th>
<th>References</th>
</tr>
</thead>
</table>
| cap5(8)A    | 222   | *Staphylococcus aureus* Cap1A (63/167), type 1 CP synthesis  
             |       | *Streptococcus thermophilus* EpsC (33/183), probable chain-length  
             |       | regulator  
             |       | *Streptococcus pneumoniae* CpsC (32/183), probable chain-length  
             |       | regulator | Lin et al. (1994)  
             |       |           | Stingele et al. (1996)  
| cap5(8)B    | 228   | *Staph. aureus* Cap1B (62/228), type 1 CP synthesis  
             |       | *Strep. pneumoniae* CpsD (41/117), probable chain-length regulator  
             |       | *Strep. thermophilus* EpsD (40/117), probable chain-length regulator  
             |       | *Rhizobium meliloti* ExoP (30/113), chain-length regulator | Lin et al. (1994)  
             |       |           | Guidolin et al. (1994)  
             |       |           | Stingele et al. (1996)  
             |       |           | Becker et al. (1995)  
| cap5(8)C    | 254   | *Staph. aureus* Cap1C (59/254), type 1 CP synthesis  
             |       | *Strep. thermophilus* EpsB (31/165), exopolysaccharide synthesis  
             |       | *Strep. pneumoniae* CpsB (31/165), type 19F CP synthesis | Lin et al. (1994)  
             |       |           | Stingele et al. (1996)  
             |       |           | Guidolin et al. (1994)  
| cap5(8)D    | 607   | *Staph. aureus* Cap1D (72/578), type 1 CP synthesis  
             |       | *Yersinia enterocolitica* TrsG (59/192), lipopolysaccharide outer core synthesis  
             |       | *Vibrio cholerae* ORF11 (62/114), O139 antigen synthesis  
             |       | *Bordetella pertussis* BpIL (55/195), lipopolysaccharide synthesis | Lin et al. (1994)  
             |       |           | Skurnik et al. (1995)  
| cap5(8)E    | 342   | *Methanococcus jannaschii* Prot D (43/239), CP synthesis | Bult et al. (1996)  
| cap5(8)F    | 371   | *Acholeplasma laidlawii* (58/369), putative nucleotide-binding protein | GenBank accession Z22875  
| cap5(8)G    | 374   | *M. jannaschii* Bp1D (46/70), lipopolysaccharide synthesis  
             |       | *Bo. pertussis* Bp1D (38/75), lipopolysaccharide synthesis  
             |       | *Escherichia coli* RHE (35/64), UDP-GlcNAc 2-epimerase  
             |       | *Salmonella enterica* RfB (35/60), UDP-GlcNAc 2-epimerase  
             |       | *Pseudomonas solanacearum* EpsC (33/68), exopolysaccharide synthesis  
             |       | *Bacillus subtilis* ORFX (31/70) | Allen & Maskell (1996)  
| cap5H       | 208   | *E. coli* Cat4 (42/88), chloramphenicol O-acetyltransferase | GenBank accession Z22875  
| cap8H       | 360   | None | Parent & Roy (1992)  
| cap8I       | 369   | None | Parent & Roy (1992)  
| cap8J       | 464   | None | Parent & Roy (1992)  
| cap8J       | 185   | *R. meliloti* NodL (54/58), O-acetyltransferase  
             |       | *E. coli* LacA (48/45), thiogalactoside O-acetyltransferase | Ardourel et al. (1995)  
| cap5J       | 338   | None | Hediger et al. (1985)  
| cap5K       | 394   | None | Parent & Roy (1992)  
| cap5K       | 412   | None | Parent & Roy (1992)  
| cap5(8)L    | 401   | *E. coli* WcaI (39/41), probable glycosyltransferase | Parent & Roy (1992)  
| cap5(8)M    | 185   | *Sal. enterica* WhaP (65/41), galactosyltransferase  
             |       | *Xanthomonas campestris* GumD (64/34), galactosyltransferase  
             |       | *Streptococcus agalactiae* CpsD (68/41), galactosyltransferase | Wang et al. (1996)  
             |       | Vanderslice et al. (1989)  
             |       | Rubens et al. (1993)  
| cap5(8)N    | 295   | *Salmonella typhimurium* GaLE (32/75), UDP-Glc 4-epimerase  
             |       | *E. coli* GaLE (31/74), UDP-Glc 4-epimerase | Houng et al. (1990)  
             |       | Lemaire & Muller-Hill (1986)  
| cap5(8)O    | 420   | *M. jannaschii* MJ0428 (62/126), UDP-ManNAc dehydrogenase  
             |       | *E. coli* RfH (51/157), UDP-ManNAc dehydrogenase  
             |       | *P. solanacearum* EpsD (51/208), exopolysaccharide synthesis  
             |       | *Pseudomonas aeruginosa* AlgD (22/68), GDPmannose dehydrogenase | Bult et al. (1996)  
             |       | Daniels et al. (1992)  
             |       | Huang & Schell (1995)  
             |       | Deretic et al. (1987a)  
| cap5(8)P    | 391   | *B. subtilis* ORFX (59/373) | Soldo et al. (1993)  
             |       | *E. coli* RfH (57/176), UDP-GlcNAc 2-epimerase | Meier-Dieter et al. (1990)  
             |       | Huang & Schell (1995)  
             |       | Keenleyside & Whitfield (1996)  
| cap5(8)Q    | 402   | *P. aeruginosa* MexE (65/219), exopolysaccharide synthesis  
             |       | *E. coli* WcaI (65/219), probable glycosyltransferase | Parent & Roy (1992)  
             |       | Parent & Roy (1992)  

* Number of amino acids.
† Numbers in parentheses indicate percentage identity of amino acid sequence/length of the homologous region.
CpsC of Streptococcus pneumoniae (Guidolin et al., 1994) (Table 2). EpsC is thought to be involved in chain-length determination because of the similarity of amino acid sequences and hydrophobicity profiles to several proteins involved in polysaccharide chain-length determination, including Cld of E. coli (Batchelor et al., 1992; Bastin et al., 1993), Rol of Salmonella enterica (Bastin et al., 1993), Rol of Shigella flexneri (Morona et al., 1995), and the N-terminal half of ExoP of Rhizobium meliloti (Becker et al., 1995). Although pairwise alignment of Cap5A or Cap8A to Cld Rol, or the N-terminal half of ExoP revealed only limited homology to the proposed conserved motif of these chain-length determinants, their hydrophobicity plots are similar, showing two potential membrane-spanning domains, one at either end, and a hydrophilic central region (not shown). Thus cap5A and cap8A could be involved in chain-length determination.

As shown in Table 2, Cap8B and Cap5B show significant homology to EpsD of Strept. thermophilus (Stingele et al., 1996), CpsD of Strept. pneumoniae (Guidolin et al., 1994) and the C-terminal half of ExoP of R. meliloti (Becker et al., 1995). All of these proteins contain the ATP-binding motif. ExoP is a large protein with 786 amino acids. The N-terminal half of ExoP has been implicated in chain-length determination of succinoglycan synthesis, whereas the C-terminal half appears to exert a regulatory function following nucleotide binding (Becker et al., 1995). As shown in Fig. 2(a), Cap8B and Cap5B also contain a conserved ATP-binding motif found in many bacterial transporter genes (Fath & Kolter, 1993). The nucleotide-binding motif is composed of the A site, with conserved GXGKST at residues 52–57, and the B site, with a conserved aspartic acid residue at position 157. Since Cap5(8)B exhibits significant homology with the conserved nucleotide-binding motif of the C-terminal half of ExoP, and since Cap5(8)A has a hydrophobicity profile similar to that of the N-terminal half of ExoP, it is likely that Cap5(8)B forms a complex with Cap5(8)A to regulate the chain length of CP5(8). A similar suggestion was first put forth with regard to Strept. thermophilus EpsC and EpsD by Stingele et al. (1996) who speculated that epsC and epsD, located next to each other like cap5(8)A and cap5(8)B, were originally one gene and were later separated or that exoP had evolved from a gene fusion. Recently, we found that a site-specific cap8B mutant produced the same amount of CP8 as wild-type strain Becker but that the CP8 of the mutant was of lower molecular mass. As shown in Table 2, Cap8B and Cap5B show significant homology to EpsD of Strep. thermophilus (Stingele et al., 1996), CpsD of Strept. pneumoniae (Guidolin et al., 1994) and the C-terminal half of ExoP of R. meliloti (Becker et al., 1995). Although pairwise alignment of Cap5A or Cap8A to Cld Rol, or the N-terminal half of ExoP revealed only limited homology to the proposed conserved motif of these chain-length determinants, their hydrophobicity plots are similar, showing two potential membrane-spanning domains, one at either end, and a hydrophilic central region (not shown). Thus cap5A and cap8A could be involved in chain-length determination.

Amino sugar synthesis genes

The deduced proteins from cap5D and cap8D genes show a high degree of homology to several proteins in the databases that have not been characterized functionally. Nevertheless, these proteins show local homology to a number of proteins involved in the sugar synthetic pathways (most of which are UDP-Glc 4-epimerases and UDP-Glc dehydratases). Skurnik et al. (1995) aligned 19 of these proteins and identified two conserved regions that may be essential for enzymic function. They proposed that Yersinia enterocolitica TrsG might be involved in the synthesis of GalNACP or FucNACP. Fig. 2(b) shows the alignment of some of these proteins at the regions near the two consensus regions. Since staphylococcal CP1, CP5 and CP8 all contain d-FucNACP and since Cap1D, Cap5D and Cap8D are highly homologous, these three proteins could be epimerases or dehydratases involved in the synthesis of d-FucNACP. Interestingly, we also found that the amino acid sequence of Cap5(8)D was moderately similar to the deduced sequences from cap5E and cap8E (Fig. 2b). The homologous regions include the two conserved motifs identified by Skurnik et al. (1995). Thus, it is possible that Cap5E and Cap8E work in concert with Cap5D and Cap8D, respectively, in the synthesis of d-FucNACP.

Cap5F and Cap8F show highest homology to a protein from Accholeplasma laidlawii whose function is unknown. Cap5F and Cap8F also show limited homology (not shown), especially at the regions near the N-terminal end, to several bacterial nucleotide sugar epimerases or dehydratases required for polysaccharide synthesis, including the RfbB proteins of E. coli (Marolda & Valvano, 1995), Sb. flexneri (Macpherson et al., 1994) and Sal. enterica (Jiang et al., 1991). Therefore, we propose that Cap5F and Cap8F are either nucleotide sugar epimerases or dehydratases. Similarly, Cap5N and Cap8N may be epimerases involved in sugar conversion because of limited homology to various GaIE (UDP-Glc 4-epimerase) proteins of E. coli (Lemaire & Muller-Hill, 1986) and Salmonella typhimurium (Houng et al., 1990), especially at the N-terminal end (not shown).

It is interesting that Cap5(8)G and Cap5(8)P exhibit 29–0% overall identity; this observation suggests that the two proteins may have similar functions but with different substrates. Cap5P and Cap8P show high degrees of homology to several gene products in the databases (Table 2) including RffE of E. coli (Meier-Dieter et al., 1990) and RfbC of Sal. enterica (Keenleyside & Whitfield, 1996), whereas Cap5(8)G show limited homology to these gene products. RffE has been shown to have 2-epimerase activity, catalysing the conversion of UDP-GlCNAc to UDP-ManNAc in the biosynthesis of enterobacterial common antigen, a surface glycolipid associated with all members of the Enterobacteriaceae (Meier-Dieter et al., 1990). RfbC of Sal. enterica serovar Borreze has been implicated as a UDP-GlcNAc 2-epimerase (Keenleyside & Whitfield, 1996). Thus, it is likely that Cap5(8)G and Cap5(8)P are epimerases. Fig. 2(c) shows the alignment of Cap5G, Cap5P and the related proteins at the regions near the N-terminal end. Cap5O and Cap8O are similar to several proteins, including RffD of E. coli (Daniels et al., 1992) and AlgD of Pseudomonas aeruginosa (Deretic et al., 1987a). The E. coli rffD gene is involved in the biosynthetic pathway leading to enterobacterial common antigen expression.
Fig. 2. Alignment of various *Staph. aureus* Cap proteins with related sequences in the databases. Gaps (−) were introduced to allow maximal alignment. Identical residues are boxed. Residue numbers are shown on the right. (a) *Staph. aureus* CapSB (SaCapSB) and CaplB (SaCap1B), *Strep. pneumoniae* CpsD (SpCpsD), *Strep. thermophilus* EpsD (StEpsD) and the N-terminal half of *R. meliloti* ExoP (RmExoP). The conserved ATP-binding domains (Fath & Kolter, 1993) are shown. (b) *Staph. aureus* Cap1D (SaCap1D) and Cap5D (SaCap5D), *Y. enterocolitica* TrsC (YeTrsC), *Bo. pertussis* BplL (BpBplL) and *Staph. aureus* Cap5E (SaCap5E). The consensus regions of 19 proteins [con(19)] reported by Skurnik et al. (1995) are shown. (c) *Staph. aureus* Cap5M (SaCap5M), *Sal. enterica* RfbP (SeRfbP), *Strep. agalactiae* CpsD (SaCpsD) and *X. campestris* GumD (XcGumD). (d) *Staph. aureus* Cap5L (SaCap5L), *E. coli* WcaL (EcWcaL) and *E. coli* MtLC (EcMtLC). (e) *Staph. aureus* Cap5M, (SaCap5M), *Sal. enterica* RfbP (SeRfbP), *Strep. agalactiae* CpsD (SaCpsD) and *X. campestris* GumD (XcGumD).
RffD has UDP-ManNAcP dehydrogenase activity, catalysing the conversion of UDP-ManNAcP to UDP-ManNAcAp (Meier-Dieter et al., 1990). AlgD of P. aeruginosa has been shown biochemically to be a GDP-D-mannose dehydrogenase (Deretic et al., 1987b). Thus it is likely that Cap50 and Cap80 are dehydrogenases involved in the synthesis of ManNAcAp, which is a component of both CP5 and CP8. In fact, cap50 complements an rffD mutation in an E. coli mutant defective in enterobacterial common antigen synthesis (K. Kiser & J. C. Lee, unpublished). The presence of NAD-binding domains (Wierenga et al., 1986) in the N-terminal ends of both Cap50 and Cap80 is in accordance with their proposed function as dehydrogenases that require NAD as a cofactor.

Transferases

Both Cap5L and Cap8L show limited homology (about 20% identity and 47.4% similarity over 365 amino acids) to WcaI of E. coli, a putative glycosyltransferase (Stevenson et al., 1996). The proposed function of WcaI was based on a low degree of homology to MtfC, a mannosyltransferase for E. coli O9 antigen synthesis (Kido et al., 1995). Although Cap5L and Cap8L do not have significant homology to MtfC, limited homology between Cap5L(Cap8L) and MtfC was detected (Fig. 2d). Thus, Cap5L and Cap8L are candidates for glycosyltransferases.

Cap5M and Cap8M are homologous to many similar-size gene products in the databases that are thought to be glycosyltransferases. These two proteins are also homologous to the C-terminal half of a group of larger proteins, including WbaP (RfbP) of Sal. enterica (Wang et al., 1996), CpsD of Streptococcus agalactiae (Rubens et al., 1993) and GumD of Xanthomonas campestris (Ielpi et al., 1993; Vanderslice et al., 1989), which all have galactosyltransferase activity. In the case of WbaP, the transferase activity is located in the C-terminal half of the protein to which Cap5M and Cap8M show a high degree of homology. The alignment of these proteins is shown in Fig. 2c. In addition, the C-terminal half of WbaP contains a potential transmembrane domain; a similar hydrophobic profile was found in Cap5M and Cap8M (not shown). On the basis of these findings, we propose that Cap5M and Cap8M are glycosyltransferases.

The cap5l and cap8h genes are located in the type-specific central regions of the cap5 and cap8 gene clusters, respectively. The predicted proteins show no homology to the reported proteins in the databases. Both Cap5L and Cap8H are hydrophilic, and the hydrophobicity profiles of the two resemble each other (not shown). Because cap5l and cap8h are type-specific genes, they may be transferase genes that provide either the serotype-specific linkage between ManNAcAp and L-FucNAcP or between D-FucNAcP and the adjacent ManNAcAp residues.

O-acetylation genes

Although Cap5H and Cap8J show little overall homology (27% identity over a region of 94 amino acids), they are similar at a region of about 50 amino acids that is located near the C-terminus of the protein. This consensus region is found in members of the NodL-LacC-CysE acetyltransferase family (Downie, 1989) including Cap1G, which is required for staphylococcal CPS synthesis (Lin et al., 1994). Cap5H is involved in O-acetylation of the third carbon of the mannosaminuronic acid residue of CP5 in Staph. aureus strain Reynolds (Lee et al., 1993). By analogy, Cap8J is likely to be an acetyltransferase involved in O-acetylation of the fourth carbon of mannosaminuronic acid of CP8. Although it is somewhat surprising that the three putative staphylococcal acetyltransferases are so dissimilar in terms of overall homologies, this dissimilarity is indicative of the substrate specificities of these enzymes.

Polymerization and export genes

Bacterial polysaccharides may be synthesized on the cytoplasmic side of the membrane and then transported through the membrane by an ABC transporter enzyme. Alternatively, repeating subunits may be synthesized on the inner side of the membrane, transported to the outer side, and then polymerized (Whitfield, 1995). An example of the latter mechanism is the O-antigen transport of Sal. enterica group B O-antigen (Whitfield, 1995). An enzyme termed flippase, Wzx (formerly RfbX), is required for flipping the lipid-bound O-subunit from the cytoplasmic side to the periplasmic side of the membrane, where it is polymerized by another membrane-bound O-antigen polymerase, Rfc (Liu et al., 1996; Morona et al., 1994). Wzx and Rfc are predicted to be integral membrane proteins with 12 transmembrane domains. However, little or no sequence homology was found among wzx genes or among O-antigen polymerases of different O-antigen gene clusters (Morona et al., 1994).

Among the predicted proteins of cap5 and cap8 loci, Cap5J, Cap5K, Cap8I, and Cap8K contain possible multiple transmembrane domains. The hydrophobicity profiles (not shown) of Cap5J and Cap8I are very similar with 10 possible transmembrane domains, though they show only 17-1% identity over 340 amino acids. Similarly, the hydrophobicity profiles of Cap5K and Cap8K are nearly identical, with 11 possible transmembrane domains, and yet the primary amino acid sequences are quite different (with 19-8% identity over 303 amino acids). Despite no similarity to known flippases or polymerases, the transmembrane domains found in these proteins are consistent with the postulation that Cap5J(Cap8I) and Cap5K(Cap8K) may be a pair of membrane-bound flippases and polymerases involved in CPS(8) synthesis. Moreover, the failure to find an ABC transporter (containing an ATP-binding cassette and multiple transmembrane domains) in the
S. SAU and OTHERS

were digested with Smal and resolved in 0.8% agarose gel as fragments from (a) were transferred to nitrocellulose paper and the Smal fragment from NCTC 8325 in kb. Note that strains M described by Goering & Winters (1992). (b) The resolved DNA about 2 to 10-5 kb coordinates in Fig. 1). Lanes: 1, strain NCTC 8325; 2, strain M; 3, strain Smith; 4, strain Reynolds; 5, strain probed with an - 8.5 kb cap8-specific DNA fragment (from Nitrocellulose paper and probed with a DNA fragment Becker, strain Reynolds and the mapping strain NCTC pneumoniae (Guidolin et al., 1994) but neither of these (serotype 1) and Smith (serotype 2) also contain the of the physical map of strain NCTC 8325 (Pattee et al., 1992). (c) The nearly identical nucleotide sequences of the cap5 or cap8 loci of Staph. aureus indicates that staphylococcal capsules of types 5 and 8 may be exported by a flippase-polymerase mechanism.

Other genes

The Cap5C and Cap8C sequences show a moderate degree of homology to EpsB of Strep. thermophilus (Stingle et al., 1996) and to Cps19fB of Strep. pneumoniae (Guidolin et al., 1994) but neither of these has been functionally characterized. We therefore cannot assign any function to cap5(8)C.

Mapping of the cap5 and cap8 gene clusters

The nearly identical nucleotide sequences of the cap5 and cap8 gene clusters at two regions that flank the central type-specific regions indicate that cap5 and cap8 loci are allelic. To determine the location of this allele on the Staph. aureus genome, we performed field-inversion gel electrophoresis of Smal-digested DNAs from strain Becker, strain Reynolds and the mapping strain NCTC 8325. The resolved DNA fragments were transferred to nitrocellulose paper and probed with a DNA fragment internal to the cap8 operon (Fig. 3). We found that cap5 and cap8 gene clusters mapped to the Smal-G fragment of the physical map of strain NCTC 8325 (Pattee et al., 1992). Interestingly, the cap1 gene cluster, which is not allelic to the cap8 locus (Sau & Lee, 1996), was also mapped to the same Smal-G fragment (S. Ouyang & C. Y. Lee, unpublished data). Thus, the cap1 and cap5(8) loci indeed map close together within a 175 kb region of the chromosome.

Conclusion

The functions that we have predicted for the cap5 and cap8 genes are based on amino acid sequence homologies with genes in the databases. Biochemical evidence is required to confirm these proposed functions. Nevertheless, our predictions are in accord with the chemical structures of the capsules, allowing us to propose a preliminary synthetic pathway for CP5 and CP8. The repeating unit of CP5(CP8) is composed of three sugars: D-ManNAcAp, L-FucNAcAp and D-FucNAcAp. We propose that these sugars are derived from a nucleotide precursor of D-GlcNAcAp by a number of epimerases and dehydratases encoded by cap5(8)D, E, F, G, N, O and P. One of the sugars, D-ManNAcAp, is further O-acetylated by Cap5H(Cap8H). The three sugar monomers are transferred by three transferases, Cap5(8)L, Cap5(8)M and Cap5I(Cap8I), to form a repeating unit possibly linked to a lipid carrier at the inner side of the cytoplasmic membrane. The lipid-linked repeating unit is then transported through the membrane and polymerized at the outer surface of the membrane by Cap5J(Cap8J) and Cap5(8)K with the aid of the chain-length determinators Cap5(8)A and Cap5(8)B. Of all the predicted gene products, Cap5(8)C is the only one to which we were unable to assign any function. However, since Cap5C and Cap8C show very strong homology with Cap1C, it is tempting to speculate that they are involved in a process common to CP5, CP8 and CP1 synthesis, such as transport. Our current efforts are directed at obtaining biochemical evidence to support the gene functions that we have proposed herein.

ACKNOWLEDGEMENTS

We thank Paul Livolsi for technical assistance and Michael Coyne for helpful suggestions. This work was supported by grants AI37027 (to C.Y.L.) and AI29040 (to J.C.L.) from the NIH, and by the Health Research Board of Ireland (to T.J.F and E.R.W.).

REFERENCES


Capsule genes of Staph. aureus


Capsule genes of *Staph. aureus*


Received 31 December 1996; revised 21 February 1997; accepted 21 March 1997.