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. Foster[3] and Chia Y. Lee[1]

Author for correspondence: Chia Y. Lee. Tel: +1 913 588 7156. Fax: +1 913 588 7295.
e-mail: clee@kumc.edu

[1] Department of Microbiology, Molecular Genetics and Immunology, University of Kansas Medical Center, Kansas City, Kansas 66160, USA

[2] Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA

[3] Microbiology Department, Moyne Institute, Trinity College, Dublin 2, Ireland

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 *Sma*I-G fragment in the standard *Sma*I map of *Staph. aureus* strain NCTC 8325. The predicted gene products of *cap5A* through *cap5G* 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: 4)-α-D-GalNAcA*p*-(1 → 4)-α-D-GalNAcA*p*-(1 → 3)-α-D-FucNAc*p*-(1→

Abbreviation: CP, capsular polysaccharide.

The GenBank accession numbers for the nucleotide sequences of the *cap5* and *cap8* gene clusters reported in this paper are U81973 and U73374, respectively.

(A taurine residue is linked by an amide bond to every fourth D-GalNAcA*p* residue.)

CP2: 4)-*β*-D-GlcNAcA*p*-(1 → 4)-*β*-D-GlcN-(*N*-acetyl-alanyl)-AcA*p*-(1→

CP5: 4)-3-*O*-Ac-*β*-D-ManNAcA*p*-(1 → 4)-α-L-FucNAc*p*-(1 → 3)-*β*-D-FucNAc*p*-(1→

CP8: 3)-4-*O*-Ac-*β*-D-ManNAcA*p*-(1 → 3)-α-L-FucNAc*p*-(1 → 3)-*β*-D-FucNAc*p*-(1→

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 kit. Restriction enzymes and other enzymes used in the study were purchased from Gibco-BRL and New England Biolabs. Field-inversion gel electrophoresis was performed as described by Goering & Winters (1992).

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 pLI50. An ~ 7·1 kb region of the *cap5* 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,

Table 1. Bacterial strains and plasmids

Strain or plasmid	Relevant characteristics	Source or reference
***Staph. aureus* strains**		
8325		J. J. Iandolo, Kansas State University, Manhattan, KS, USA
M	Type 1 capsule strain	J. H. Hash, Vanderbilt University, Nashville, TN, USA
Smith	Type 2 capsule strain	ATCC 13709
Reynolds	Type 5 capsule strain	Karakawa & Vann (1982)
Newman	Type 5 capsule strain	NCTC 8178
Becker	Type 8 capsule strain	Karakawa & Vann (1982)
***E. coli* strains**		
HB101	*recA13 hsdS20 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 supE44 rpsL20*	J. Lutkenhaus, University of Kansas Medical Center, Kansas City, KS, USA
JM109	*recA1 endA1 gyrA96 thi hsdR17 supE44 relA* 1Δ(*lac–proAB*)/F′ (*traD36 proAB*$^+$ *lacI*^{q} *lacZ*ΔM15)	Yanisch-Perron *et al*. (1985)
XL1-Blue	*recA1 endA1 gyrA96 thi hsdR17 supE44 relA1* Δ(*lac–proAB*)/F′ (*traD36 proAB*$^+$ *lacI*^{q} *lacZ*ΔM15 Tn*10*)	Sambrook *et al*. (1989)
Plasmids		
pGEM7Z(+)	Cloning vector	Promega
pBluescript	Cloning vector	Stratagene
pLI50	*E. coli–Staph. aureus* shuttle cloning vector	Lee *et al*. (1991)
pJCL19	34 kb fragment of strain Reynolds DNA cloned in *Bam*HI site of pHC79	Lee *et al*. (1994)

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 20·5 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 20·5 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 Tn*918* mutagenesis (Lee *et al.*, 1994) and in strain Newman by Tn*917* mutagenesis (E. R. Wann & T. J. Foster, unpublished). DNA fragments flanking the Tn*918* insertion sites were used to screen a cosmid library prepared from wild-type strain Reynolds. Similarly, sequences flanking Tn*917* 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 Cap8G, 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

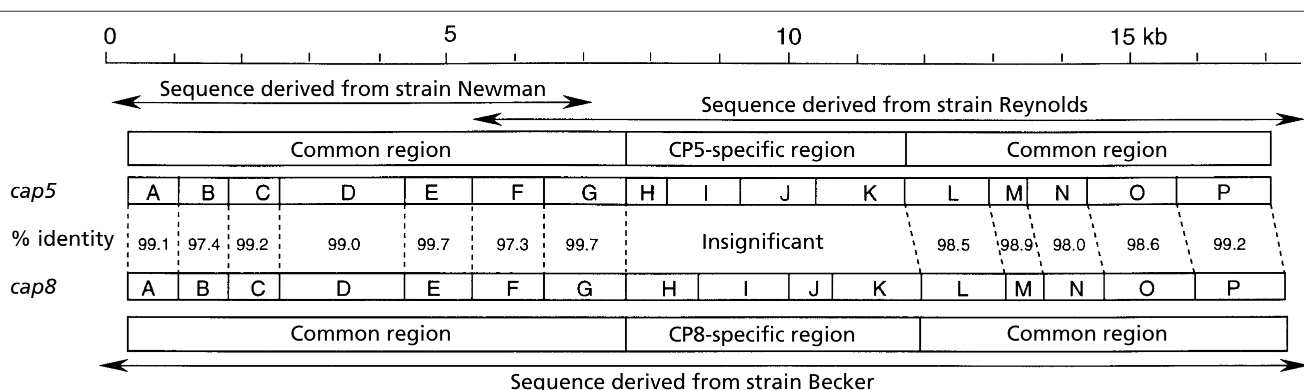

Fig. 1. Comparison of *Staph. aureus cap5* and *cap8* gene clusters. The *cap5* sequence was derived from strains Newman and Reynolds and the *cap8* sequence from strain Becker as shown. Gene designations are shown in boxes. Percentage identity indicates the amino acid identity of the deduced proteins of the two clusters. Both gene clusters are transcribed from left to right.

Table 2. Homology of putative gene products with related sequences in the databases

ORF	Size*	Homologous protein†	References
cap5(8)A	222	*Staphylococcus aureus* Cap1A (63/167), type 1 CP synthesis	Lin *et al.* (1994)
		Streptococcus thermophilus EpsC (33/183), probable chain-length regulator	Stingele *et al.* (1996)
		Streptococcus pneumoniae CpsC (32/183), probable chain-length regulator	Guidolin *et al.* (1994)
cap5(8)B	228	*Staph. aureus* Cap1B (62/228), type 1 CP synthesis	Lin *et al.* (1994)
		Strep. pneumoniae CpsD (41/117), probable chain-length regulator	Guidolin *et al.* (1994)
		Strep. thermophilus EpsD (40/117), probable chain-length regulator	Stingele *et al.* (1996)
		Rhizobium meliloti ExoP (30/113), chain-length regulator	Becker *et al.* (1995)
cap5(8)C	254	*Staph. aureus* Cap1C (59/254), type 1 CP synthesis	Lin *et al.* (1994)
		Strep. thermophilus EpsB (31/165), exopolysaccharide synthesis	Stingele *et al.* (1996)
		Strep. pneumoniae CpsB (31/165), type 19F CP synthesis	Guidolin *et al.* (1994)
cap5(8)D	607	*Staph. aureus* Cap1D (72/578), type 1 CP synthesis	Lin *et al.* (1994)
		Yersinia enterocolitica TrsG (59/192), lipopolysaccharide outer core synthesis	Skurnik *et al.* (1995)
		Vibrio cholerae ORF11 (62/114), O139 antigen synthesis	Comstock *et al.* (1996)
		Bordetella pertussis Bp1L (55/195), lipopolysaccharide synthesis	Allen & Maskell (1996)
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	Bult *et al.* (1996)
		Bo. pertussis Bp1D (38/75), lipopolysaccharide synthesis	Allen & Maskell (1996)
		Escherichia coli RffE (35/64), UDP-GlcNAc 2-epimerase	Meier-Dieter *et al.* (1990)
		Salmonella enterica RfbC (35/60), UDP-GlcNAc 2-epimerase	Keenleyside & Whitfield (1996)
		Pseudomonas solanacearum EpsC (33/68), exopolysaccharide synthesis	Huang & Schell (1995)
		Bacillus subtilis ORFX (31/70)	Soldo *et al.* (1993)
cap5H	208	*E. coli* Cat4 (42/88), chloramphenicol *O*-acetyltransferase	Parent & Roy (1992)
cap8H	360	None	
cap5I	369	None	
cap8I	464	None	
cap8J	185	*R. meliloti* NodL (54/58), *O*-acetyltransferase	Ardourel *et al.* (1995)
		E. coli LacA (48/45), thiogalactoside *O*-acetyltransferase	Hediger *et al.* (1985)
cap5J	338	None	
cap5K	394	None	
cap8K	412	None	
cap5(8)L	401	*E. coli* WcaI (39/41), probable glycosyltransferase	Stevenson *et al.* (1996)
cap5(8)M	185	*Sal. enterica* WbaP (65/41), galactosyltransferase	Wang *et al.* (1996)
		Xanthomonas campestris GumD (64/34), galactosyltransferase	Vanderslice *et al.* (1989)
		Streptococcus agalactiae CpsD (68/41), galactosyltransferase	Rubens *et al.* (1993)
cap5(8)N	295	*Salmonella typhimurium* GalE (32/75), UDP-Glc 4-epimerase	Houng *et al.* (1990)
		E. coli GalE (31/74), UDP-Glc 4-epimerase	Lemaire & Muller-Hill (1986)
cap5(8)O	420	*M. jannaschii* MJ0428 (62/126), UDP-ManNAc dehydrogenase	Bult *et al.* (1996)
		E. coli RffD (51/157), UDP-ManNAc dehydrogenase	Daniels *et al.* (1992)
		P. solanacearum EpsD (51/208), exopolysaccharide synthesis	Huang & Schell (1995)
		Pseudomonas aeruginosa AlgD (22/68), GDPmannose dehydrogenase	Deretic *et al.* (1987a)
cap5(8)P	391	*B. subtilis* ORFX (59/373)	Soldo *et al.* (1993)
		E. coli RffE (57/176), UDP-GlcNAc 2-epimerase	Meier-Dieter *et al.* (1990)
		P. solanacearum EpsC (48/220), exopolysaccharide synthesis	Huang & Schell (1995)
		Sal. enterica RfbC (54/178), UDP-GlcNAc 2-epimerase	Keenleyside & Whitfield (1996)

* 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 *Strep. thermophilus* (Stingele *et al*., 1996), CpsD of *Strep. 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 *Strep. 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 (S. Sau & C. Y. Lee, unpublished). This result is consistent with the proposed function of Cap8B as a chain-length regulator. Studies are in progress to confirm the function of *cap8B*.

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 GalNAc*p* or FucNAc*p*. 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-FucNAc*p* and since Cap1D, Cap5D and Cap8D are highly homologous, these three proteins could be epimerases or dehydratases involved in the synthesis of D-FucNAc*p*. 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-FucNAc*p*.

Cap5F and Cap8F show highest homology to a protein from *Acholeplasma 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), *Sh. 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 GalE (UDP-Glc 4-epimerase) proteins of *E. coli* (Lemaire & Müller-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*p* to UDP-ManNAc*p* 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*p* 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.

(a)

```
SaCap1B SEKSASGKSILSANIAVTYAQAGYKTLIIDGDMRKPTQHYIFDLPNN-SGLSNLIINKTTYSDSIKETRVENLNVLTAGPTP--- 129
SaCap5B SEKPGAGKSTVVSNVAITYAQAGYKTLVIDGDMRKPTQNYIFNEQNN-NGLSSLIIGRTTMSEAITSTEIENLDLLTAGPVP--- 129
SpCpsD  SVNPGEGKTTTSVNIARSFARAGYKTLLIDGDTRNSVMSGFFKSREKITGLTEFLSGTADLSHGLCDTNIENLFVVQSGTVS--- 124
StEpsD  SVEAGEGKSMISVNLAISFASVGLRTLLIDAETRNSVLSGTFKSNEPYKGLSNFLSGNADLNETICQTDISGLDVIASGPVP--- 124
RmExoP  SALPDEGKSIIAANFAALLAASGKRTLLIDADIRKPGLTQMITPAPR-TGLVETLIGEASWPAGIKVDQRTKLAILPAGGASHQR 124
             GXGKST
ATP Binding domain-A site

SaCap1B PNPSELIASSKFATIFNELLNHYDFIVIDTPPINTVTDAQVYARIVKNCVLVIDAEKNNKSEVKKAKGLLTKAGGKVLGAVLNK 212
SaCap5B PNPSELIGSERFKELVDLFNKRYDIIIVDTPPVNTVTDAQLYARAIKDSLLVIDSEKNDKNEVKKAKALMEKAGSNILGVILNK 212
SpCpsD  PNPTALLQSKNFNDMIETLRKYFDYIIVDTAPIGIVIDAAIITQKCDASILVTATGEVNKRDVQKAKQQLEQTGKLFLGVVFNK 207
StEpsD  PNPTSLLQNDNFRHLMEVARSCYDYVIIDTPPVGLVIDAVIIAHQADASLLVTEAGKIKRRFVTKAVEQLVESGSQFLGVVLNK 207
RmExoP  HQSNELLASPAMANLIENARNAFDYVVVDLAALAPVVDAKAFAPLADGILFVVEWGRTPSRLVRDLLHSEPLINSKVLGVILNK 208
                              ILILD
                  ATP Binding domain-B site
```

(b)

```
SaCap1D KTILVTGAGGSIGSEICRQVSKFDPQKIILLGHGENSIYSIHQELS---KTYGNRIEFVPVIADVQNKTRILEVMNEFKPYAVYHAAAHKHVPLMEY 375
SaCap5D KTILVTGAGGSIGSEICRQVCNFYPERIILLGHGENSIYLINRELR---NRFGKNVDIVPIIADVQNRARMFEIMETYKPYAVYHAAAHKHVPLMED 379
YeTrsG  KVVMVTGAGGSIGSELCRQIIVEKPSLLILFDISEFSLYSIENEMAAICKKNKIETEFVALLGSVQSEKRLVQIMSNFHVNTVYHAAAYKHVPLVEN 375
BpBplL  RVVMVTGAGGSIGSELCRQILALRPRKLVLFEIAEPALYAIEQDLRQ--RIGERNIEIAGVLGSVRDAAHCLAQLQEHGVQTIYHAAAYKHVPIVEH 380
SaCap5E KILLITGGTGSFGNAVMKQFLDSNIKEIRIFSRDE-------KKQDDIRKKYN-NSKLKFYIGDVRDSQSVETAMRD--VDYVFHAAALKQVPSCEF 91
Con(19)    VLVTGGAGYIGS

SaCap1D NPHEAIRNNILGTKNVAESAKEGEVSKFVMISTDKAVNPSNVMGATKRIAEMVI--QSLNEDNSKTSFVAVRFGNVLGSRGSVIPLFKNQIESGGP 469
SaCap5D NPEEAVRNNILGTKNTAEAAKNAEVKKFVMISTDKAVNPPNVMGASKRIAEMII--QSLNDETHRTNFVAVRFGNVLGSRGSVIPLFKSQIEEGGP 473
YeTrsG  NVIEGVRNNIFGTLYCAKAAIKSGVEKFVLISTDKAVRPTNTMGATKRMAELVL--QALSTEQNKTKFCMVRFGNVLGSSGSVVPLFKKQIAEGGP 469
BpBplL  NVSEGIRTNAFGTLNMAETAIQAGVLDFVLISTDKAVRPTNVMGASKRLAELIL--QAHAQIQDKTRFSMVRFGNVLGSSGSVVPLFRRQILEGGP 474
SaCap5E FPVEAVKTNIIGTENVLQSAIHQNVKKVICLSTDKAAYPINAMGISKAMMEKVFVAKSRNIRSEQTLICGTRYGNVMASRGSVIPLFIDKIKAGEP 187
Con(19)                           VXXFVXISTD
```

(c)

```
SaCap5G KLMTIVGTRPEIIRLSSTIKACDQY--FNQILVHTGQNYDYTLNQIFFDDLELRQPDHYLEAV--GSNLGETMGNIIAKTYDVLLREQPD 90
BpBplD  KILTVLGARPQFIKASVVSAAIAQHPELTEVVVHTGQHFDANMSDVFFDELGMQTPAHQLDIH--GGGHGDMTGRMLVALEQVMQAEKPD 91
PsEpsC  KVLVVFGTRPEAIKMAPLVKALQADASLQCGVCVTAQHRE--MLDQVLRLFDIR-PDYDLNVMKPGQDLYELTSNILTGVKSVLESFEPD 110
BsOrfX  KVMTVFGTRPEAIKMAPLVLELKKYPEIDSYVTVTAQHRQ--MLDQVLDAFHIK-PDFDLNIMKERQTLAEITSNALVRLDELFKDIKPD 91
EcRffE  KVLTVFGTRPEAIKMAPLVHALAKDPFFEAKVCVTAQHRE--MLDQVLKLFSIV-PDYDLNIMQPGQGLTEITCRILEGLKPILAEFKPD 88
SeRfbC  KVLFVFGTRPEAIKMAPLVIEFKNNPAIEVKVCVTGQHRE--MLDQVLDFFEIE-PDYDLNIMKQKQSLGSITCSILTRLDEILASFMPA 89
SaCap5P KIMVIFGTRPEAIKMAPLVKEIDHNGNFEANIVITAQHRD--MLDSVLSIFDIQ-ADHDLNIMQDQQTLAGLTANALAKLDSIINEEQPD 99

SaCap5G ALLILGDTNSCLA-AVSAKRLKIPVFHMEAGNRCFDQ--NVPEEINRKIVDHVSDVNLPYTEHSRRYLLDEGFNKANIFVTGSPMTEVI 176
BpBplD  VVLVYGDTNSTLAGALAAVKLHIPVAHVEAGLRSFNL--RMPEEVNRILTDRISRWLFTPTDSATRHLAAEGQSGEKVVQVGDVMYDVA 178
PsEpsC  LVLVHGDTSTTLATTLAAYYKQVPVGHIEAGLRTGNLYSPWPEEVNRKVTGSLAALHFAPTERSRRNLLNEGVPADAVVVTGNTVIDAL 199
BsOrfX  IVLVHGDTTTTFAGSLAAFYHQIAVGHVEAGLRTGNKYSPFPEELNRQMTGAIADLHFAPTGQAKDNLLKENKKADSIFVTGNTAIDAL 180
EcRffE  VVLVHGDTTTTLATSLAAFYQRIPVGHVEAGLRTGDLYSPWPEEANRTLTGHLAMYHFSPTETSRQNLLRENVADSRIFITGNTVIDAL 177
SeRfbC  HIFVHGDTTTTFAASLAAFYQNIKVWHIEAGLRTWNMNSPFPEEGNRQLTSKLAFFHAAPTLQAKDNLLRESVKEKNIIVTGNTVIDAL 178
SaCap5P MILVHGDTTTTFVGSLAAFYHQIPVGHVEAGLRTHQKYSPFPEELNRVMVSNIAELNFAPTVIAAKNLLFENKDKERIFITGNTVIDAL 188
```

(d)

```
SaCap5L REQ---YD--KILLYSGNMGQLQDMETLISFLKLNKDQPQTLTILCGHGKKFADVKT-AIEDHRIENVKMFEF-LTGTDYADVL-K 297
EcWcaI  RNQLDLPDNKKIILYSGNIGEKQGLENVIEAADRLRDEPLIFAIV-GQGGGKARLEKMAQQRG-LRNMQ-FFP-LQSYDALPALLK 303
EcMtfC  RETVG--DNFF--LFVGAFRYYKGLHILLDAAERSR-LPVV--IV-GGGPLEAEVRREAQQRG-LSNVV-FTGMLNDED-KYIL-F 163
```

(e)

```
SaCap5M SPGPAIFKQKRPTINNELFNIYKFRSMKIDTPNVATDLM---------------DSTSYITKTGKVIRKTSIDELPQLLNVLKGEMSIVGPRP 109
SeRfbP  DGGPAIYGHQRVGRHGKLFPCYKFRSMVMNSQEVLKELLANDPIARAEWEKDFKLKNDPRITAVGRFIRKTSLDELPQLFNVLKGDMSLVGPRP 402
SaCpsD  DGGPAIFSQNRVGRNGRIFRFYKFRSMRVDAEQIKKDLLVHNQMT----GLMFKLEDDPRITKIGKFIRKTSIDELPQFYNVLKGDMSLAGTAS 272
XcGumD  SPGPVFFRQRRHGLGGREFYMFKFRSMRVHDDHGTTIQQAT--------------KNDTRITRFGSFLRRSSLDELPQIFNVLGGSMSIVGPRP 402
```

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* Cap5B (SaCap5B) and Cap1B (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* TrsG (YeTrsG), *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* Cap5G (SaCap5G), *Bo. pertussis* BplD (BpBplD), *P. solanacearum* EpsC (PsEpsC), *B. subtilis* OrfX (BsOrfX), *E. coli* RffE (EcRffE), *Sal. enterica* RfbC (SeRfbC) and *Staph. aureus* Cap5P (SaCap5P). (d) *Staph. aureus* Cap5L (SaCap5L), *E. coli* WcaI (EcWcaI) and *E. coli* MtfC (EcMtfC). (e) *Staph. aureus* Cap5M, (SaCap5M), *Sal. enterica* RfbP (SeRfbP), *Strep. agalactiae* CpsD (SaCpsD) and *X. campestris* GumD (XcGumD).

RffD has UDP-ManNAc*p* dehydrogenase activity, catalysing the conversion of UDP-ManNAc*p* to UDP-ManNAcA*p* (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 Cap5O and Cap8O are dehydrogenases involved in synthesis of ManNAcA*p*, which is a component of both CP5 and CP8. In fact, *cap5O* 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 Cap5O and Cap8O 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 local 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. 2(e). In addition, the C-terminal half of WhaP contains a potential transmembrane domain; a similar hydropathic profile was found in Cap5M and Cap8M (not shown). On the basis of these findings, we propose that Cap5M and Cap8M are glycosyltransferases.

The *cap5I* 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 Cap5I and Cap8H are hydrophilic, and the hydrophobicity profiles of the two resemble each other (not shown). Because *cap5I* and *cap8H* are type-specific genes, they may be transferase genes that provide either the serotype-specific linkage between ManNAcA*p* and L-FucNAc*p* or between D-FucNAc*p* and the adjacent ManNAcA*p* 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-LacA-CysE acetyltransferase family (Downie, 1989) including Cap1G, which is required for staphylococcal CP1 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.*, 1995). 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 CP5(8) synthesis. Moreover, the failure to find an ABC transporter (containing an ATP-binding cassette and multiple transmembrane domains) in the

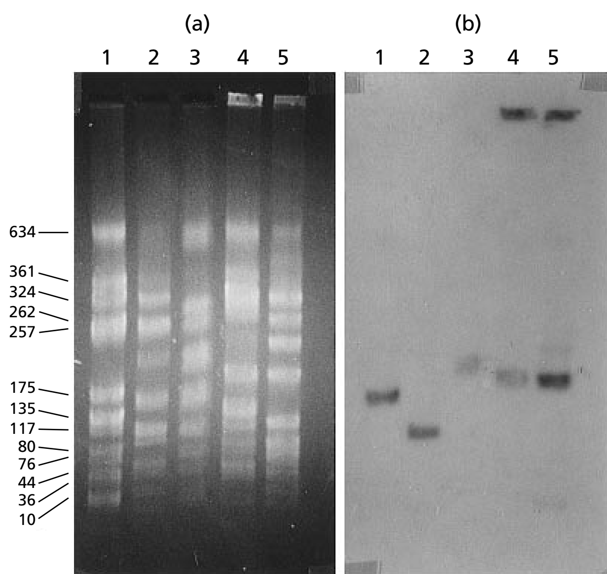

Fig. 3. Mapping of the *cap5*(*8*) locus. (a) Chromosomal DNAs were digested with *Sma*I and resolved in 0·8% agarose gel as described by Goering & Winters (1992). (b) The resolved DNA fragments from (a) were transferred to nitrocellulose paper and probed with an ~8·5 kb *cap8*-specific DNA fragment (from 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 Becker. Numbers to the left of the figure indicate the sizes of the *Sma*I fragment from NCTC 8325 in kb. Note that strains M (serotype 1) and Smith (serotype 2) also contain the *cap8*-related DNA (Sau & Lee, 1996).

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* (Stingele *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 *Sma*I-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 *Sma*I-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 *Sma*I-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-ManNAcA*p*, L-FucNAc*p* and D-FucNAc*p*. We propose that these sugars are derived from a nucleotide precursor of D-GlcNAc*p* by a number of epimerases and dehydratases encoded by *cap5*(*8*)*D*, *E*, *F*, *G*, *N*, *O* and *P*. One of the sugars, D-ManNAcA*p*, is further O-acetylated by Cap5H(Cap8J). The three sugar monomers are transferred by three transferases, Cap5(8)L, Cap5(8)M and Cap5I(Cap8H), 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(Cap8I) 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

Albus, A., Fournier, J. M., Wolz, C., Boutonnier, A., Ranke, M., Hoiby, N., Hochkeppel, H. & Doring, G. (1988). *Staphylococcus aureus* capsular types and antibody response to lung infection in patients with cystic fibrosis. *J Clin Microbiol* **26**, 2205–2209.

Albus, A., Arbeit, R. D. & Lee, J. C. (1991). Virulence of *Staphylococcus aureus* mutants altered in type 5 capsule production. *Infect Immun* **59**, 1008–1014.

Allen, A. & Maskell, D. (1996). The identification, cloning and mutagenesis of a genetic locus required for lipopolysaccharide biosynthesis in *Bordetella pertussis*. *Mol Microbiol* **19**, 37–52.

Altschul, A. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Arbeit, R. D., Karakawa, W. W., Vann., W. F. & Robbins, J. B. (1984). Predominance of two newly described capsular polysaccharide types among clinical isolates of *Staphylococcus aureus*. *Diagn Microbiol Infect Dis* **2**, 85–91.

Ardourel, M., Lortet, G., Maillet, F., Roche, P., Truchet, G., Promé, J. C. & Rosenberg, C. (1995). In *Rhizobium meliloti*, the operon associated with the *nod* box n5 comprises *nodL*, *noeA* and *noeB*, three host-range genes specifically required for the nodulation of particular *Medicago* species. *Mol Microbiol* **17**, 687–699.

Baddour, L. M., Lowrance, C., Albus, A., Lowrance, J. H., Anderson, S. K. & Lee, J. C. (1992). *Staphylococcus aureus* microcapsule expression attenuates bacterial virulence in a rat model of experimental endocarditis. *J Infect Dis* **165**, 749–753.

Bastin, D. A., Stevenson, G., Brown, P. K., Haase, A. & Reeves, P. (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.

Batchelor, R. A., Alifano, P., Biffali, E., Hull, S. & Hull, R. (1992). Nucleotide sequences of the genes regulating O-polysaccharide antigen chain length (*rol*) from *Escherichia coli* and *Salmonella typhimurium*: protein homology and functional complementation. *J Bacteriol* **174**, 5228–5236.

Becker, A., Niehaus, K. & Pühler, A. (1995). Low-molecular-weight succinoglycan is predominantly produced by *Rhizobium meliloti* strains carrying a mutated ExoP protein characterized by a periplasmic N-terminal domain and a missing C-terminal domain. *Mol Microbiol* **16**, 191–203.

Birnboim, H. C. (1983). A rapid alkaline extraction method for the isolation of plasmid DNA. *Methods Enzymol* **100**, 243–255.

Bult, C. J., White, O., Olsen, G. J. & 37 other authors (1996). Complete genome sequence of the methanogenic Archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058–1073.

Comstock, L. E., Johnson, J. A., Michalski, J. M., Morris, J. G., Jr & Kaper, J. B. (1996). Cloning and sequence of a region encoding a surface polysaccharide of *Vibrio cholerae* O139 and characterization of the insertion site in the chromosome of *Vibrio cholerae* O1. *Mol Microbiol* **19**, 815–826.

Daniels, D. L, Plunkett, G., III, Burland, V. & Blattner, F. R. (1992). Analysis of the *Escherichia coli* genome: DNA sequence of the region from 84·5 to 86·5 minutes. *Science* **257**, 771–778.

Deretic, V., Gill, J. F. & Chakrabarty, A. M. (1987a). *Pseudomonas aeruginosa* infection in cystic fibrosis: nucleotide sequence and transcriptional regulation of the *algD* gene. *Nucleic Acids Res* **15**, 4567–4581.

Deretic, V., Gill, J. F. & Chakrabarty, A. M. (1987b). Gene *algD* coding for GDP-mannose dehydrogenase is transcriptionally activated in mucoid *Pseudomonas aeruginosa*. *J Bacteriol* **169**, 351–358.

Downie, J. A (1989). The *nodL* gene from *Rhizobium leguminosarum* is homologous to the acetyl transferases encoded by *lacA* and *cysE*. *Mol Microbiol* **3**, 1649–1651.

Fath, M. & Kolter, R. (1993). ABC transporters: bacterial exporters. *Microbiol Rev* **57**, 995–1017.

Fattom, A. I., Sarwar, J., Ortiz, A. & Naso, R. (1996). A *Staphylococcus aureus* capsular polysaccharide (CP) vaccine and CP-specific antibodies protect mice against bacterial challenge. *Infect Immun* **64**, 1659–1665.

Fournier, J. M., Vann, W. F. & Karakawa, W. W. (1984). Purification and characterization of *Staphylococcus aureus* type 8 capsular polysaccharide. *Infect Immun* **45**, 87–93.

Goering, R. V. & Winters, M. A. (1992). Rapid method for epidemiological evaluation of gram-positive cocci by field inversion gel electrophoresis. *J Clin Microbiol* **30**, 577–580.

Guidolin, A., Morona, J. K., Morona, R., Hansman, D. & Paton, J. (1994). Nucleotide sequence analysis of genes essential for capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* type 19F. *Infect Immun* **62**, 5384–5396.

Hanessian, S. & Haskell, T. H. (1964). Structural studies on staphylococcal polysaccharide antigens. *J Biol Chem* **239**, 2758–2764.

Hediger, M. A, Johnson, D. F., Nierlich, D. P. & Zabin, I. (1985). DNA sequence of the lactose operon: the *lacA* gene and the transcriptional termination region *Proc Natl Acad Sci USA* **82**, 6414–6418.

Higgins, D. G. & Sharp, P. M. (1988). CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. *Gene* **73**, 237–244.

Hochkeppel, H. K., Braun, D. G., Vischer, W., Imm, A., Sutter, S., Staeubli, U., Guggenheim, R., Kaplan, E. L., Boutonnier, A. & Fournier, J. M. (1987). Serotyping and electron microscopy studies of *Staphylococcus aureus* clinical isolates with monoclonal antibodies to capsular polysaccharide type 5 and type 8. *J Clin Microbiol* **25**, 526–530.

Houng, H. H.-S., Kopecko, D. & Baron, L. S. (1990). Molecular cloning and physical and functional characterization of the *Salmonella typhimurium* and *Salmonella typhi* galactose utilization operons. *J Bacteriol* **172**, 4392–4398.

Huang, J. & Schell, M. (1995). Molecular characterization of the *eps* gene cluster of *Pseudomonas solanacearum* and its transcriptional regulation at a single promoter. *Mol Microbiol* **16**, 977–989.

Ielpi, L., Couso, R. O. & Dankert, M. A. (1993). Sequential assembly and polymerization of the polyprenol-linked pentasaccharide repeat unit of the xanthan polysaccharide in *Xanthomonas campestris*. *J Bacteriol* **175**, 2490–2500.

Jiang, X.-M., Santiago, N. R., Lee, S. L., Romana, L. K. & Reeves, P. R. (1991). Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar *typhimurium* (strain LT2). *Mol Microbiol* **5**, 695–713.

Karakawa, W. W. & Vann, W. F. (1982). Capsular polysaccharides of *Staphylococcus aureus*. *Semin Infect Dis* **4**, 285–293.

Karakawa, W. W., Fournier, J. M., Vann, W. F., Arbeit, R., Schneerson, R. & Robbins, J. B. (1985). Method for the serological typing of the capsular polysaccharides of *Staphylococcus aureus*. *J Clin Microbiol* **22**, 445–447.

Karakawa, W. W., Sutton, A., Schneerson, R., Karpas, A. & Vann, W. F. (1988). Capsular antibodies induce type-specific phagocytosis of *Staphylococcus aureus* by human polymorphonuclear leukocytes. *Infect Immun* **56**, 1090–1095.

Keenleyside, W. J. & Whitfield, C. (1996). A novel pathway for O-polysaccharide biosynthesis in *Salmonella enterica* serovar Borreze. *J Biol Chem* **271**, 28581–28592.

Kido, N., Torgov, V. I., Sugiyama, T., Uchiya, K., Sugihara, H., Komatsu, T., Kato, N. & Jann, K. (1995). Expression of the O9 polysaccharide of *Escherichia coli*: sequencing of the *E. coli* O9 *rfb* gene cluster, characterization of mannosyl transferases, and evidence for an ATP-binding cassette transport system. *J Bacteriol* **177**, 2178–2187.

Klein, P., Kanehisa, M. & Delisi, C. (1985). The detection and classification of membrane-spanning proteins. *Biochim Biophys Acta* **815**, 468–476.

Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132.

Lee, C. Y. (1992). Cloning of genes affecting capsule expression in *Staphylococcus aureus* strain M. *Mol Microbiol* **6**, 1515–1522.

Lee, C. Y., Buranen, S. L. & Ye, Z. H. (1991). Construction of single-copy integration vectors for *Staphylococcus aureus*. *Gene* **103**, 101–105.

Lee, J. C., Betley, M. J., Hopkins, C. A., Perez, N. E. & Pier, G. B. (1987). Virulence studies, in mice, of transposon-induced mutants of *Staphylococcus aureus* differing in capsule size. *J Infect Dis* **156**, 741–750.

Lee, J. C., Xu, S., Albus, A. & Livolsi, P. J. (1994). Genetic analysis of type 5 capsular polysaccharide expression by *Staphylococcus aureus*. *J Bacteriol* **176**, 4883–4889.

Lee, J. C., Albus, A. & Livolsi, P. J. (1995). Identification of a gene essential for O-acetylation of the *Staphylococcus aureus* type 5 capsular polysaccharide. In *Abstracts of the 95th Annual Meeting of the American Society for Microbiology*, abstract no. B453. Washington, DC: American Society for Microbiology.

Lee, J. C., Park, J.-S., Shepherd, S. E. & Fattom, A. (1996). Protective efficacy of antibodies to the *Staphylococcus aureus* capsular polysaccharides to a rat model of endocarditis. In *Abstracts of the 96th Annual Meeting of the American Society for Microbiology*, abstract no. B328. Washington, DC: American Society for Microbiology.

Lemaire, H.-G. & Müller-Hill, B. (1986). Nucleotide sequences of the *galE* gene and the *galT* gene of *E. coli*. *Nucleic Acids Res* **14**, 7705–7711.

Lin, W. S., Cunneen, T. & Lee, C. Y. (1994). Sequence analysis and molecular characterization of genes required for the biosynthesis of type 1 capsular polysaccharide in *Staphylococcus aureus*. *J Bacteriol* **176**, 7005–7016.

Liu, D., Cole, R. A. & Reeves, P. R. (1996). An O-antigen processing function for Wzx (RfbX): a promising candidate for O-unit flippase. *J Bacteriol* **178**, 2102–2107.

Macpherson, D. F., Manning, P. A. & Morona, R. (1994). Characterization of the dTDP-rhamnose biosynthetic genes encoded in the *rfb* locus of *Shigella flexneri*. *Mol Microbiol* **11**, 281–292.

Marolda, C. L. & Valvano, M. A. (1995). Genetic analysis of the dTDP-rhamnose biosynthesis region of the *Escherichia coli* VW187 (O7:K1) *rfb* gene cluster: identification of functional homologs of *rfbB* and *rfbA* in the *rff* cluster and correct location of the *rffE* gene. *J Bacteriol* **177**, 5539–5546.

Meier-Dieter, U., Starman, R., Barr, K., Mayer, H. & Rick, P. D. (1990). Biosynthesis of enterobacterial common antigen in *Escherichia coli*. *J Biol Chem* **265**, 13490–13497.

Melly, M. A., Duke, L. J., Liau, D. F. & Hash, J. H. (1974). Biological properties of the encapsulated *Staphylococcus aureus* M. *Infect Immun* **10**, 389–397.

Moreau, M., Richards, J. C., Fournier, J. M., Byrd, R. A., Karakawa, W. W. & Vann, W. F. (1990). Structure of the type-5 capsular polysaccharide of *Staphylococcus aureus*. *Carbohydr Res* **201**, 285–297.

Morona, R., Mavris, M., Fallarino, A. & Manning, P. A. (1994). Characterization of the *rfc* region of *Shigella flexneri*. *J Bacteriol* **176**, 733–747.

Morona, R., van den Bosch, L. & Manning, P. A. (1995). Molecular, genetic, and topological characterization of O-antigen chain length regulation in *Shigella flexneri*. *J Bacteriol* **177**, 1059–1068.

Murthy, S. V. K. N., Melly, M. A., Harris, T. M., Hellerqvist, C. G. & Hash, J. H. (1983). The repeating sequence of the capsular polysaccharide of *Staphylococcus aureus*. *Carbohydr Res* **117**, 113–123.

Nemeth, J. & Lee, J. C. (1995). Antibodies to capsular polysaccharides are not protective against experimental *Staphylococcus aureus* endocarditis. *Infect Immun* **63**, 375–380.

Oeding, P. (1983). Taxonomy and identification. In *Staphylococci and Staphylococcal Infections*, pp. 1–32. Edited by C. S. G. Easmon & G. Adlam. New York: Academic Press.

Parent, R. & Roy, P. H. (1992). The chloramphenicol acetyltransferase gene of Tn*2424*: a new breed of cat. *J Bacteriol* **174**, 2891–2897.

Pattee, P. A., Lee, H. C. & Bannantine, J. P. (1992). Genetic and physical mapping of the chromosome of *Staphylococcus aureus*. In *Molecular Biology of the Staphylococci*, pp. 41–67. Edited by R. P. Novick. New York: VCH.

Peterson, P. K., Wilkinson, B. J., Kim, Y., Schmeling, D. & Quie, P. G. (1978). Influence of encapsulation on staphylococcal opsonization and phagocytosis by human polymorphonuclear leukocytes. *Infect Immun* **19**, 943–949.

Poutrel, B., Boutonnier, A., Sutra, L. & Fournier, J. M. (1988). Prevalence of capsular polysaccharide types 5 and 8 among *Staphylococcus aureus* isolates from cow, goat, and ewe milk. *J Clin Microbiol* **26**, 38–40.

Rubens, C. E., Heggen, L. M., Haft, R. G. & Wessels, M. R. (1993). Identification of *cpsD*, a gene essential for type III capsule expression in group B streptococci. *Mol Microbiol* **8**, 844–855.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**, 5463–5467.

Sau, S. & Lee, C. Y. (1996). Cloning of type 8 capsule genes and analysis of gene clusters for the production of different capsular polysaccharides in *Staphylococcus aureus*. *J Bacteriol* **178**, 2118–2126.

Sau, S., Sun, J. & Lee, C. Y. (1997). Molecular characterization and transcriptional analysis of type 8 genes in *Staphylococcus aureus*. *J Bacteriol* **179**, 1614–1621.

Skurnik, M., Venho, R., Toivanen, P. & Al-Hendy, A. (1995). A novel locus of *Yersinia enterocolitica* serotype O:3 involved in lipopolysaccharide outer core biosynthesis. *Mol Microbiol* **17**, 575–594.

Soldo, B., Lazarevic, V., Margot, P. & Karamata, D. (1993). Sequencing and analysis of the divergon comprising *gtaB*, the structural gene of UDP-glucose pyrophosphorylase of *Bacillus subtilis* 168. *J Gen Microbiol* **139**, 3185–3195.

Sompolinsky, E., Samra, Z., Karakawa, W. W., Vann, W. F., Schneerson, R. & Malik, Z. (1985). Encapsulation and capsular types in isolates of *Staphylococcus aureus* from different sources and relationship to phage types. *J Clin Microbiol* **22**, 828–834.

Stevenson, G., Andrianopoulos, K., Hobbs, M. & Reeves, P. R. (1996). Organization of the *Escherichia coli* K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid. *J Bacteriol* **178**, 4885–4893.

Stingele, F., Neeser, J.-R. & Mollet, B. (1996). Identification and characterization of the *eps* (exopolysaccharide) gene cluster from *Streptococcus thermophilus* Sfi6. *J Bacteriol* **178**, 1680–1690.

Vanderslice, R. W., Doherty, D. H., Capage, M. A., Betlach, M. R., Hassler, R. A., Henderson, N. M., Ryan-Graniera, J. & Techlenberg, M. (1989). Genetic engineering of polysaccharide structure in *Xanthomonas campestris*. In *Biomedical and Biotechnological Advances in Industrial Polysaccharides*, pp. 145–156. Edited by V. Crescenzi, I. C. M. Dea, S. Paoletti, S.

Stivala & I. Sutherland. New York: Godon & Breach Science Publishers.

Wang, L., Liu, D. & Reeves, P. R. (1996). C-terminal half of *Salmonella enterica* WbaP (RfbP) is the galactosyl-1-phosphate transferase domain catalyzing the first step of O-antigen synthesis. *J Bacteriol* **178**, 2598–2604.

Whitfield, C. (1995). Biosynthesis of lipopolysaccharide O antigens. *Trends Microbiol* **3**, 178–185.

Wierenga, R. K., Terpstra, P. & Hol, W. G. J. (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.

Wilkinson, B. J. (1983). Staphylococcal capsules and slime. In *Staphylococci and Staphylococcal Infections*, pp. 481–523. Edited by C. S. G. Easmon & G. Adlam. New York: Academic Press.

Xu, S., Arbeit, R. D. & Lee, J. C. (1992). Phagocytic killing of encapsulated and microencapsulated *Staphylococcus aureus* by human polymorphonuclear leukocytes. *Infect Immun* **60**, 1358–1362.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.

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