INFECTION AND IMMUNITY, Jan. 1999, p. 193–200 Vol. 67, No. 1
0019-9567/99/$04.00+0

Infection-Derived *Enterococcus faecalis* Strains Are Enriched in *esp*, a Gene Encoding a Novel Surface Protein

VISWANATHAN SHANKAR,[1] ARTO S. BAGHDAYAN,[1] MARK M. HUYCKE,[2] GUNNAR LINDAHL,[3] AND MICHAEL S. GILMORE[4]*

Department of Medicinal Chemistry and Pharmaceutics,[1] Department of Medicine and Veterans Affairs Medical Center,[2] and Departments of Microbiology and Immunology and of Ophthalmology,[4] University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73190, and Department of Medical Microbiology, Lund University, S-223 62 Lund, Sweden[3]

Received 3 September 1998/Accepted 8 October 1998

We report the identification of a new cell wall-associated protein of *Enterococcus faecalis*. Studies on the distribution of the gene encoding this novel surface protein, Esp, reveal a significant ($P < 0.001$) enrichment in infection-derived *E. faecalis* isolates. Interestingly, the *esp* gene was not identified in any of 34 clinical *E. faecium* isolates or in 4 other less pathogenic enterococcal species tested. Analysis of the structural gene among various *E. faecalis* isolates reveals the existence of alternate forms of expression of the Esp protein. The deduced primary structure of the Esp protein from strain MMH594, inferred to be 1,873 amino acids (aa) with a predicted mass of ~202 kDa, reveals a core region consisting of repeat units that make up 50% of the protein. Esp bears global organizational similarity to the Rib and C alpha proteins of group B streptococci. Identity among Esp, Rib, and C alpha proteins is strikingly localized to a stretch of 13 aa within repeats of similar length. The high degree of conservation of this 13-residue sequence suggests that it plays an important role in the natural selection for this trait among infection-derived *E. faecalis* and group B streptococcal isolates.

Enterococci have emerged as a leading cause of nosocomial infections (1, 6) and now rank among the most common nosocomial pathogens isolated from the bloodstream, surgical sites, and urinary tract infections (35). *Enterococcus faecalis* accounts for approximately 75% of all enterococcal infections, with *E. faecium* accounting for most of the rest (20). The increasing importance of enterococci as nosocomial pathogens can, in part, be attributed to their natural ability to readily exchange extrachromosomal elements encoding traits that confer survival or growth advantages. Interest in enterococcal virulence and host-parasite interaction has been stimulated to a great extent by concerns that increasing antibiotic resistance may soon render conventional therapies inadequate for treating enterococcal infections. Despite an increasing awareness of the potential for enterococci to cause serious infections, little is known about their virulence. This lack of information can be attributed partly to the fact that enterococci, which normally grow as commensal organisms of the gut, possess very subtle virulence traits that are not easily identified (20).

Several traits that may contribute to enhanced virulence have been identified in *E. faecalis*. The *E. faecalis* cytolysin (10, 23, 26) lyses a broad range of eukaryotic and prokaryotic cells, is usually plasmid encoded (22), and enhances the virulence of *E. faecalis* in animal models (5, 21, 24, 25, 27). Aggregation substance mediates adhesion to cultured renal tubular cells (28) and augments the internalization of *E. faecalis* by cultured human intestinal epithelial cells (37). Recent studies (18, 19) suggest a potential role for extracellular superoxide as a virulence factor.

In an effort to identify new factors that may contribute to enterococcal pathogenesis, we derived and systematically panned a database of nucleotide sequences compiled from random sequencing of the genome of *E. faecalis* MMH594 (21), which caused multiple infections within a hospital ward, for sequences that appeared to encode surface proteins. A chromosomal gene with localized sequence identity to *Streptococcus agalactiae rib* and *bca* (encoding C alpha antigen) was identified from partial sequence information, and its preliminary characterization was reported (42). The Rib and C alpha proteins of group B streptococci are structurally related and consist of highly repetitive structures (33, 49). These group B streptococcal surface proteins have been shown to be virulence determinants and to confer protective immunity, and they appear to contribute to immune system evasion (3, 29, 30, 32, 44). We have named the enterococcal gene *esp* (since it encodes enterococcal surface protein); in this paper we describe its structure and structural variations, localization of the gene product to the surface of the organism, and association of the gene and its product with infection-derived *E. faecalis* isolates.

MATERIALS AND METHODS

Bacteria, plasmids, and media. *E. faecalis* MMH594 (21), a clinical isolate that caused multiple infections in a hospital ward outbreak, served as the prototype for elucidation of the complete nucleotide sequence of the *esp* structural gene. *E. faecalis* strains were routinely cultivated in brain heart infusion (BHI; Difco), whereas Luria-Bertani broth (39) was used for cultivation of *Escherichia coli* strains. *E. coli* XL1-Blue and XL1-Blue MR were obtained from Stratagene (La Jolla, Calif.), and BL21-DE3 was obtained from Novagen (Madison, Wis.). Antibiotics (Sigma, St. Louis, Mo.) used for selection of *E. faecalis* strains included gentamicin (500 μg/ml) and erythromycin (50 μg/ml). For maintenance of recombinant constructs in *E. coli*, ampicillin at 100 μg/ml and kanamycin at 30 μg/ml were used where appropriate.

Enterococcal isolates. One set of enterococcal strains consisted of blood isolates drawn from a collection obtained from patients with *E. faecalis* bacteremia at the University of Wisconsin Hospital and Clinics, Madison, between June 1985 and April 1987 (21). In general, these patients had nosocomial bacteremia from various sources, had been hospitalized for more than 2 weeks, and represented all adult and pediatric intensive-care and medical-surgical wards. A second set of blood isolates were obtained from patients with endocarditis treated at the Mayo Clinic between 1973 and 1991. These patients represent both Olmsted County residents treated at the Mayo Clinic-affiliated hospitals and referral cases from elsewhere (17). A third set of blood isolates were representative of enterococcal

* Corresponding author. Mailing address: Departments of Microbiology and Immunology and of Ophthalmology, University of Oklahoma Health Sciences Center, P.O. Box 26901, Oklahoma City, OK 73190. Phone: (405) 271-1084. Fax: (405) 271-8128. E-mail: mgilmore@aardvark.ou.edu.

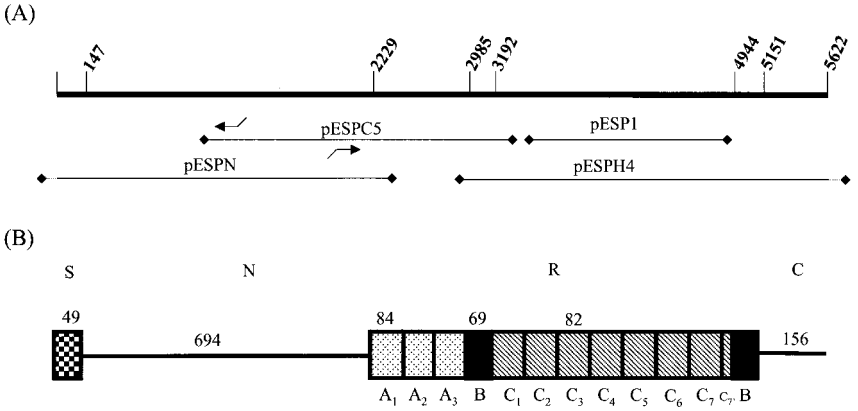


FIG. 1. Schematic of the *esp* gene and inferred protein product. (A) Structural *esp* gene and location of insert DNA from various clones used to derive the complete sequence of *esp*. Bent arrows indicate the positions of oligonucleotide primers esp15 and esp16 used for inverse PCR. (B) Deduced Esp protein showing the inferred signal (S), N-terminal (N), repeat (R), and C-terminal (C) regions. Numbers above each region denote the number of amino acid residues in each segment. The repeat blocks A, B, and C are identified, with C_7, denoting the partial C repeat.

bacteremia isolates collected from hospitals in the Buffalo, N.Y., area between January and November 1993 (38). Enterococcal stool isolates included those isolated in 1994 from stool swabs of healthy individuals between the ages of 3 and 42 years who had no history of hospitalization, did not work in health care-related areas, and lived in the greater metropolitan area of Oklahoma City (17). All strains were identified to species level with API 20S kits (bioMérieux Vitek Inc., Hazelwood, Mo.).

Isolation of enterococcal DNA. For preparation of total DNA from *E. faecalis*, each isolate taken from a BHI agar plate was inoculated into 4 ml of BHI containing 2.5% glycine and incubated overnight at 37°C. The cell pellet obtained after centrifugation at 5,000 × *g* for 5 min was resuspended in TES buffer (50 mM Tris-HCl [pH 7.5], 10 mM EDTA, 30 mM NaCl) containing 1 mg of lysozyme per ml and incubated at 37°C for 30 min. The cells were then lysed by addition of an equal volume of 2% SDS, treated with RNase and proteinase K, and extracted with phenol-chloroform twice, and the DNA was precipitated with 0.7 volume of isopropanol. The DNA pellet was resuspended in 200 μl of TE (10 mM Tris-HCl [pH 8.0], 1 mM EDTA) buffer, and the DNA concentration was determined by measuring the absorbance at 260 nm.

Restriction fragment analysis of enterococcal DNA. DNA was prepared from enterococcal strains grown in BHI supplemented with 2.5% glycine as described previously (34). Agarose plugs containing lysed cells were digested with either *Sfi*I or *Sma*I and electrophoresed with a CHEF DRII contour-clamped homogeneous electric field (CHEF) device (Bio-Rad Laboratories, Hercules, Calif.) with the pulse time ramped from 5 to 35 s over 24 h at 200 V. The gels were stained with ethidium bromide and photographed under UV illumination.

Cloning of *esp*. To retrieve a clone containing the entire *esp* gene and flanking regions, a cosmid library was generated in the SuperCos 1 vector (Stratagene). The cosmid library was constructed as specified by the manufacturer with MMH594 DNA partially digested with *Sau*3A as the starting material. The library was subjected to one round of amplification in host strain XL1-Blue MR. The amplified library was screened by colony hybridization with [α-^{32}P]dCTP-labeled DNA probes (11).

DNA manipulations and nucleotide-sequencing strategy. DNA manipulations such as ligation and restriction endonuclease digestions were performed by standard methods (39). Plasmid DNA from *E. coli* for cloning and sequence analysis was purified with a Wizard Preps DNA purification kit (Promega Corp., Madison, Wis.). Restriction and modifying enzymes were purchased from New England Biolabs Inc. (Beverly, Mass.) and Life Technologies Inc. (Gaithersburg, Md.). Custom oligonucleotides were obtained from Integrated DNA Technologies (Coralville, Iowa).

Initial detection of a portion of the *esp* gene was made from randomly sequencing a bank generated by cloning size-selected partial *Sau*3A fragments from the genome of *E. faecalis* MMH594 into M13mp18 (39). To facilitate subsequent sequence analysis, the entire insert from clone pESP was subcloned as an *Eco*RI and *Hin*dIII fragment into pBluescript II (SK−) to generate pESP1 (Fig. 1). To obtain a clone containing the entire *esp* gene and flanking sequences, the cosmid library was screened by using the insert from pESP1 as a probe. DNA was purified from a single cosmid clone, pESPC2, and restriction analysis indicated that the size of insert DNA was approximately 35 kb. A Southern blot of the restricted DNA probed with the *Eco*RI-*Hin*dIII fragment from pESP1 identified a ~4.5-kb *Hin*dIII fragment from within the 35-kb insert of pESPC2 that presumably contained the complete *esp* gene and flanking sequences. The *Hin*dIII fragment was gel purified and cloned into *Hin*dIII-restricted pBluescript II (SK−) to generate pESPH4. Nested deletions were generated from the forward direction after restricting with *Xba*I, protecting the ends with α-thio-deoxynucleoside triphosphates and then restricting with *Eco*RI to generate a exonuclease III-sensitive end. For the reverse direction, pESPH4 was cut first with *Kpn*I and then with *Xho*I.

To obtain the complete 5′ end of the *esp* gene, a 518-bp *Hin*dIII-*Cla*I fragment of pESPH4 was used to probe a Southern blot of *Cla*I-restricted DNA from the cosmid clone pESPC2. A 2.4-kb *Cla*I fragment that hybridized to the probe was gel purified and cloned into *Cla*I-restricted pBluescript II (SK−) to generate pESPC5. Nested deletion subclones were generated from pESPC5 in both directions and sequenced to obtain 1,904 bp of additional *esp* gene sequence 5′ to the region contained in pESPH4.

Inverse PCR (45) was used to obtain the sequence of the extreme 5′ end of the *esp* gene directly from MMH594 genomic DNA. Approximately 1 μg of MMH594 DNA was restricted with *Pst*I, and the resulting fragments were self-ligated. Ligated DNA (50 to 100 ng) was used in a typical inverse PCR with the Takara LA PCR kit as recommended by the manufacturer (Panvera Corp., Madison, Wis.). A 3.2-kb inverse PCR product was obtained by using the esp15 and esp16 primers (Table 1) located within the 5′ region of pESPC5. The reaction product was electrophoretically purified, recovered with the GeneClean Kit (BIO 101 Inc., Vista, Calif.), and cloned into a modified T-vector (Invitrogen Corp., Carlsbad, Calif.) to generate pESPN. Overlapping nested deletion subclones were generated from pESPN after restriction with *Pst*I and *Cla*I and sequenced on both strands.

To verify that the deduced sequence of *esp* represented the native gene without additions or deletions during manipulation of the various clones, Southern blots of MMH594 DNA cut with *Bgl*II and *Eco*RI (neither of which cuts within the *esp* gene) were probed with *esp* gene fragments from the various clones. In all cases, the probes hybridized to a single *Bgl*II-*Eco*RI fragment of ~7.5 kb. In addition, sequence information was verified by PCR amplification of select regions of the *esp* gene from MMH594 DNA and restriction mapping.

Sequencing reactions were carried out by the standard chain termination method with a fluorescein- or indodicarbocyanine (Cy5)-labeled primer and a T7 DNA polymerase-based Autoread sequencing kit (Amersham Pharmacia Biotech Inc., Piscataway, N.J.). Automated sequence information was obtained from a Pharmacia LKB A.L.F. *Express* DNA Sequencer. Custom fluorescein- or Cy5-labeled sequencing primers were also synthesized where necessary to prime reactions from within known sequence regions and resolve sequence ambiguities. BLAST (2, 48) searches were conducted with Wisconsin Package (version 9) software (GCG, Madison, Wis.).

Expression of the N-terminal region of Esp in *E. coli*. A 1,860-bp DNA sequence corresponding to nucleotides 441 through 2300 of the *esp* gene was amplified by PCR with primers esp22 and esp23 (Table 1). The amplified product was restricted with *Bcl*I and *Hin*dIII to release a 1,151-bp internal fragment. This restriction fragment was gel purified and ligated into *Bam*HI- and *Hin*dIII-cut expression vector pET28b(+) (Novagen) containing an N-terminal oligohistidine (His-Tag) domain. The ligation mixture was used to transform nonexpression host XL1-Blue cells, and transformants were selected on Luria-Bertani agar plates containing 30 μg/ml kanamycin. Plasmid DNA was purified from appropriate recombinants, and the construct, pET28CA, was verified by sequence analysis. Purified pET28CA DNA was used to transform the expression host BL21(DE3). Fusion protein was produced as specified by the manufacturer (Novagen). The solubilized protein was purified by column chromatography on His-Trap (Pharmacia) metal chelation resin columns (14). Eluted protein was dialyzed against successive changes of phosphate-buffered saline overnight at 4°C and concentrated. Purified fusion protein was then subjected to limited N-terminal sequencing to confirm that the desired product had been obtained.

TABLE 1. Oligonucleotide primers used in this study

Name	Sequence (5′ to 3′)	Strand	Position (nt)[a]
esp2	CAGATGGATCATCTGATGAAGT	+	3254–3275
esp5	GTAACGTTACTGTTACATCTGC	−	5359–5338
esp11	TTGCTAATGCTAGTCCACGACC	+	1217–1238
esp12	GCGTCAACACTTGCATTGCCGAA	−	2171–2149
esp15	AACAACGATAGGTCGTGGACTAGCATTAGC	−	1248–1219
esp16	CAGCTAAAGGTGTAGGAGAATCGGAGCCGAT	+	2198–2228
esp22	ATTAGATGAACGATTGGCCCAG	+	441–462
esp23	ACAGTTACTGCTAAATCG	−	2300–2283
esp46	TTACCAAGATGGTTCTGTAGGCAC	+	2256–2279
esp47	CCAAGTATACTTAGCATCTTTTGG	−	3192–3169

[a] Nucleotide sequence positions refer to the *esp* sequence deposited in the GenBank database under accession no. AF034779.

Production of rabbit polyclonal antiserum. Polyclonal antiserum to purified fusion protein was raised by immunization of New Zealand White rabbits by standard methods (12). For the initial dose (day 1), 100 μg of antigen in complete Freund's adjuvant was injected subcutaneously. Booster doses (50 μg) were administered intramuscularly on days 14, 42, and 56 in incomplete adjuvant. Blood was collected from the marginal ear vein at 2-week intervals after the booster doses, and serum was checked for antibody titer by an enzyme-linked immunosorbent assay (7) with preimmune serum as a control. The rabbits were exsanguinated by cardiac puncture, and the serum was collected, separated, and stored at −20°C.

Analysis of strain MMH594 for cell surface expression of Esp. Cell surface localization of Esp was verified as described previously for the Rib protein of group B *Streptococcus* (44). Bacteria from an overnight culture were washed, resuspended to 1% in phosphate-buffered saline containing 0.05% Tween 20, and incubated with specific rabbit antiserum to the N-terminal region of Esp diluted as indicated. Bound immunoglobulin G was detected and quantitated following incubation with a known amount of ^{125}I-labeled protein G and determination of radioactivity in the washed cell pellet. Experimental controls which included strains lacking the *esp* gene as well as incubations with preimmune rabbit serum were negative in all cases. To determine if Esp was secreted into the growth medium or present in the cytoplasm, concentrated cell culture supernatants and cytosolic fractions from cell lysates were also tested for the presence of Esp.

The reactivity of antiserum raised against the N-terminal region of Esp toward Rib- and C alpha-bearing group B streptococcal cells was similarly assessed, except that the antiserum was tested at a single dilution (1:500). In addition, rabbit polyclonal antiserum to Rib and C alpha was evaluated for its ability to detect Esp on the enterococcal cell surface.

Epidemiology of the *esp* gene and evidence for repeat number variation. In each instance, three different primer combinations were used in PCR amplifications to detect the presence of the *esp* gene in DNA purified from enterococcal isolates and simultaneously assess repeat number variation. Primers esp46 and esp47 (Table 1), corresponding to nucleotides (nt) 2256 to 2279 and 3169 to 3192, respectively, were used to amplify across the A repeat region (Fig. 1). Primers esp2 and esp5 (Table 1), corresponding to nt 3254 to 3275 and 5338 to 5359 of the *esp* gene, respectively, were used for amplifications across the C repeat region (Fig. 1). Primers esp11 and esp12 (Table 1) correspond to nt 1217 to 1238 and 2149 to 2171, respectively, within the N-terminal region of *esp* (Fig. 1).

Briefly, 50 μl of PCR mixture consisted of 250 ng of DNA, 0.2 μM each forward and reverse primers, 200 μM each dATP, dCTP, dGTP, and dTTP, 2.5 mM $MgCl_2$ and 2.5 U of *AmpliTaq* DNA polymerase (Perkin Elmer Corp., Foster City, Calif.) in 1× reaction buffer. Samples were overlaid with mineral oil and, after an initial denaturation step at 95°C for 2 min, subjected to 30 cycles of denaturation (94°C for 45 s), annealing (63°C for 45 s), and extension (72°C for 4 min). One-tenth of the amplified reaction mixture was mixed with gel-loading buffer and electrophoresed in a 1% agarose gel, and the reaction products were visualized by ethidium bromide staining. Positive and negative controls were included with each set of amplifications. To reduce the possibility of false-negative results arising from lack of primer binding as a result of potential point mutations in the primer binding sites, Southern blots of *Eco*RI-restricted DNA from *esp*-negative strains, as determined by PCR, were also probed with an *esp*-specific probe generated from strain MMH594 by using primer pairs esp11 and esp12 (data not shown).

Stability of the repeat units in serially collected clonal infection isolates. A representative sample of outbreak isolates from the University of Wisconsin Hospitals and Clinics (21) was selected from a consecutive, nonduplicative collection of bacteremia isolates by choosing every third strain. These 20 isolates spanned a 2-year outbreak period from July 1985 to October 1987. All isolates appeared identical to MMH594 based on CHEF gel banding patterns (data not shown). The extent of variation in repeat number was assessed for both the A and C repeats by PCR as described above.

Statistical analysis. The statistical significance of associations among *E. faecalis* isolates was calculated with JMP software (version 3.1; SAS Institute Inc.).

Nucleotide sequence accession number. The DNA sequence reported in this article has been deposited in the GenBank nucleotide sequence database under accession no. AF034779.

RESULTS

Structural analysis of the *esp* gene and deduced protein. The nucleotide sequence of the *esp* gene from MMH594 shows an unusually large structural gene consisting of 5,622 nt capable of encoding a primary translation product of 1,873 amino acids (aa), with a theoretical molecular mass of ~202 kDa. A schematic of the structural gene and location of specific regions within subclones (as described in experimental procedures) used for nucleotide sequence determination is shown in Fig. 1. The extremely long open reading frame encodes a protein that, to our knowledge, is the largest yet reported for *E. faecalis*. Translation appears to initiate from a TTG start codon that is preceded 9 bases upstream by a putative ribosome binding site, GGAGC.

The N terminus of the deduced Esp protein reveals that the first 49 aa could serve as a signal sequence directing export of the mature protein outside the cytoplasm (46, 47). Following the 5′ end, encoding a putative signal sequence, is a 2,082-nt sequence specifying a 694-residue N-terminal domain. BLAST searches of the deduced amino acid sequence revealed no significant similarity scores among GenBank sequences.

The central part of the *esp* gene has a unique architecture and is made up of two distinct tandem repeating units. The first repeating unit is located immediately downstream of the region encoding the N-terminal domain and consists of three nearly identical 252-nt tandem repeats (A repeats [Fig. 2]) extending from nt 2311 to 3066. The A-repeat sequences exhibit no significant similarity to any sequence in the GenBank database. Adjacent 3′ to the A repeat region is a 207-nt spacer region (B repeat; nt 3067 to 3273) that precedes the core 1,722-nt sequence (nt 3274 to 4995), which consists of seven nearly identical 246-nt tandem repeating units (C repeats) encoding reiterations of an 82-aa sequence (Fig. 2). The reiterated C repeat region accounts for 31% of the *esp* gene, with each repeat being very highly conserved. Thr, Val, and Gly are predicted to constitute 44% of the encoded C repeat domain. A partial, eighth C repeat extends from nt 4996 to 5025 and is identical to the first 30 nt of repeat units 2 to 7 (Fig. 2). This partial repeat is flanked on the 3′ side by a second 207-bp B-repeat sequence. The two B repeats have 74% sequence identity at the amino acid level (Fig. 2).

The 156-residue C terminus of Esp is encoded by the sequence extending from nt 5233 to 5700. The amino acid sequence of this region is consistent with that for a membrane-

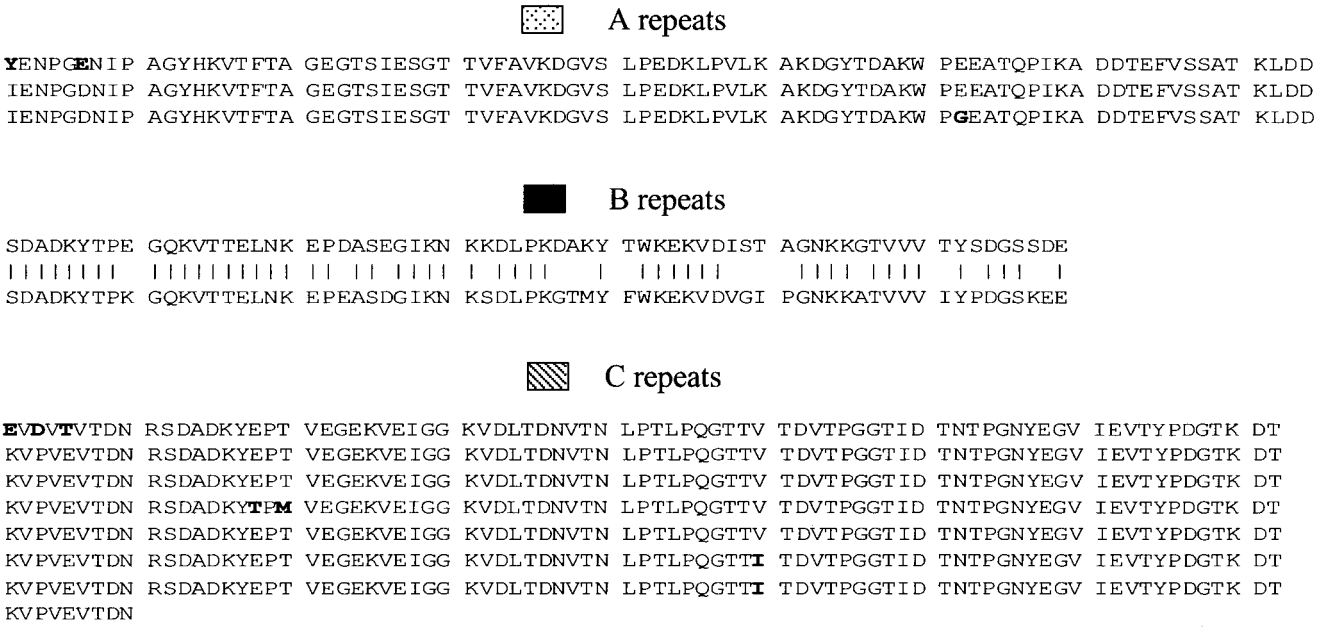
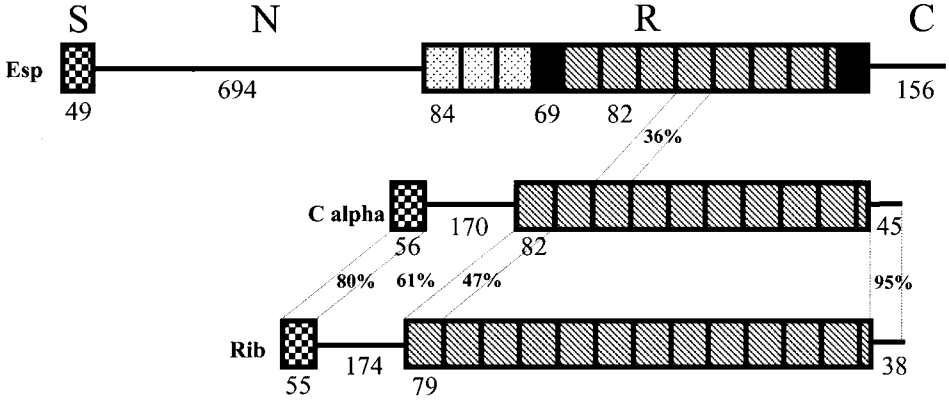


FIG. 2. Alignment of amino acid residues from within repeat blocks A, B, and C of the Esp protein, showing the extremely high degree of intramolecular sequence conservation. Substitutions are shown in boldface type.

spanning hydrophobic region and includes a YPKTGE motif, a slight variation of the LPXTGX consensus cell wall anchor motif found in most wall-associated surface proteins of gram-positive bacteria (36, 40). A charged tail ending in glutamic acid presumably extends into the cytoplasm of the cell. A region of dyad symmetry that could serve as a potential transcription terminator is situated 26 nt downstream of the TAG stop codon (4).

Conserved sequences among Esp, Rib, and C alpha. Esp exhibits global organizational similarity to two group B streptococcal proteins, Rib and C alpha (33, 49), as shown in Fig. 3. However, extensive sequence identity between the predicted




```
Esp        KVPVEVTDN RSDADKYEPT VEGEKVEIGG KVDLTDNVTN LPTLPQGTTV TDVTPGGTID
C alpha    ---------- VPDKDKYDPT GGETTVPQG- TPVSDKEITD LVKIPDGSKG VPTVVGDRPD
Rib        ---------- --DADKNDPA GKDQQVNVGE TPKAEDSIGN LPDLPKGTTV AFETP---VD

           TNTPGNYEGV IEVTYPDGTK DTV------- ----
           TNVPGDHKVT VEVTYPDGTK DTVEVTVHVT PKP-
           TATPGDKPAK VVVTYPDGSK DTVDVTVKVV DPRT
```

FIG. 3. Comparison of the Esp, Rib, and C alpha proteins, highlighting the global organizational similarity of the three proteins. Regions with high degree of sequence homology are marked by dotted lines, with the percent identity indicated. The Esp structure is derived from strain MMH594 (this study), the Rib structure is derived from group B streptococcal strain BM110 (49), and the C alpha structure is derived from group B streptococcal strain A909 (33). The bottom panel shows alignment between the 82-residue C repeat of Esp and the corresponding repeats in Rib and C alpha. Identical residues are shown in boldface type. The highly conserved 13-residue region within the repeats is underlined.

TABLE 2. Frequency of the *esp* gene among *E. faecalis* isolates

Source	No. of isolates with[a]:		Total no. of isolates
	esp present	*esp* absent	
Endocarditis	14	19	33
Blood	29	71	100
Stool	1	33	34

[a] The presence of the *esp* gene was assessed by PCR as described in Materials and Methods ($P < 0.001$ by chi-square testing).

protein sequences occurs only within the highly reiterated repeats. Moreover, within the repeats, sequence identities are localized and most pronounced for an inferred nearly identical 13-residue motif with a sequence consisting of (I/V)(E/V)VTYPDG(T/S)KDTV. As a result of the similar size of the Rib, C alpha, and Esp repeats, this highly conserved 13-residue motif occurs repeatedly with a near-identical periodicity in each gene product. However, Esp differs from Rib and C alpha by virtue of having three different kinds of repeat units and much longer unrelated amino- and carboxy-terminal regions.

Distribution of the *esp* gene among clinical and commensal isolates. PCR amplifications of enterococcal DNA with *esp* gene-specific primer pairs revealed an enrichment for the *esp* gene among infection-derived *E. faecalis* isolates compared to commensal stool isolates. As shown in Table 2, 29 (29%) of 100 blood isolates and 14 (42%) of 33 endocarditis isolates were positive for the *esp* gene, compared to only 1 (3%) of 34 stool isolates ($P < 0.001$ by the chi square test). The *esp* gene was not identified among any of 34 clinical isolates of *E. faecium* or 2 isolates each of *E. avium*, *E. gallinarum*, *E. casseliflavus*, and *E. raffinosus* by either PCR or Southern analysis.

Repeat number variation in the A and C repeats of *esp*. The presence of tandem, highly conserved sequences within the A and C repeat regions of the *esp* gene suggested that different enterococcal isolates may exhibit size and repeat number variations as a result of homologous recombination within identical repeat units. PCR amplifications across the A repeat (primer pair esp46 and esp47 [Table 1]) and C repeat (primer pair esp2 and esp5 [Table 1]) regions revealed substantial variation in the size and number of both repeats among *esp*-positive isolates (Fig. 4). This difference in size corresponded to multiples of either 252 bp (A repeats) or 246 bp (C repeats). To verify that variation occurred in the number of A repeat units, the amplified product in each instance was restricted with *Hin*dIII (which cuts once within each A repeat unit) and the

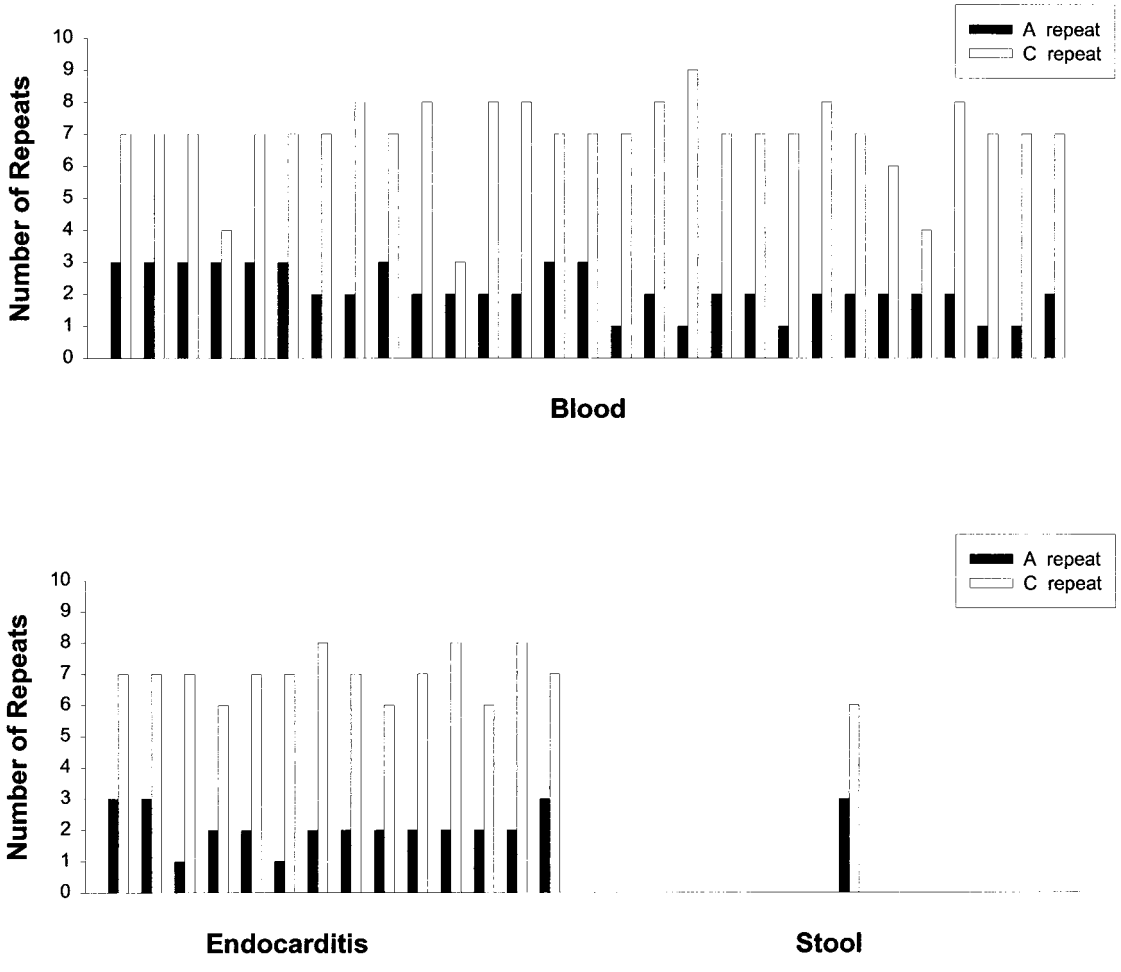


FIG. 4. Graphical representation of the variation in the number of A repeat (252-bp) and C repeat (246-bp) units among *esp*-positive blood, endocarditis, and stool isolates of *E. faecalis*. Repeat number variation was assessed by PCR as described in Materials and Methods, with primer pairs esp46 and esp47 (A repeats) and esp2 and esp5 (C repeats).

restriction products were analyzed on a 2% agarose gel. In all cases, only three restriction fragments of the expected sizes (173, 252, and 261 bp) were observed, with an overrepresentation of the repeating 252-bp fragment based on its intensity in ethidium bromide-stained gels (data not shown). The number of A repeat units varied from one to three among strains (Fig. 4).

Variation in C repeat units was similarly verified by gel electrophoresis of *Cla*I restriction fragments. In all cases, three restriction fragments of the expected sizes (196, 246, and 434 bp) were observed, with an overrepresentation of the 246-bp repeating fragment based on its intensity in ethidium bromide-stained gels (data not shown). The number of C repeat units varied from three to nine, with a majority of the isolates possessing seven complete repeat units (Fig. 4).

Stability in repeat number among clonal infection-derived isolates. Analysis of 20 clonal blood isolates exhibiting genetic identity to MMH594 and collected over a 2-year period (21) revealed that 19 of the isolates possessed *esp* determinants with the same number of A repeats (three) and C repeats (seven) as MMH594. The single deviant strain exhibited the same number of 252-bp A repeats but showed a reduction in the size of the 246-bp C repeat region consistent with the presence of only two repeat units (data not shown). This isolate was collected approximately 6 months after the initial identification and isolation of strain MMH594. In separate experiments, DNA prepared from strain MMH594 passaged in laboratory medium and compared to a frozen archived strain showed no changes over a similar 2-year period (data not shown).

Esp is expressed on the surface of *E. faecalis*. Rabbit antiserum to a purified N-terminal region of Esp expressed in *E. coli* was raised to assess Esp localization on the surface of *E. faecalis*. Strains harboring the *esp* gene, as well as those that lacked the *esp* determinant, were evaluated for binding of radiolabeled protein G following incubation of bacterial suspensions with specific antiserum to the recombinant N-terminal portion of Esp. As shown in Fig. 5, greater than 60% of the added radiolabeled protein G was detected in the pellet of cells harboring the *esp* gene, compared to less than 1% above background exhibited by control strain lacking *esp*. Controls with preimmune rabbit serum showed no nonspecific binding of radiolabeled protein G. Also, concentrated culture supernatants as well as cytosolic fractions from *esp*-positive cell lysates were negative for the presence of Esp. These experiments unequivocally demonstrated that Esp is a surface protein and that the amino-terminal region is accessible to antibodies. In all cases, the detection of Esp correlated precisely with the presence of the *esp* gene (data not shown).

No cross-reactivity against group B streptococcal strains expressing Rib or C alpha was detected with antiserum to the N-terminal region of Esp (Fig. 5). In similar experiments, antiserum to purified Rib and C alpha failed to detect Esp on the enterococcal cell surface.

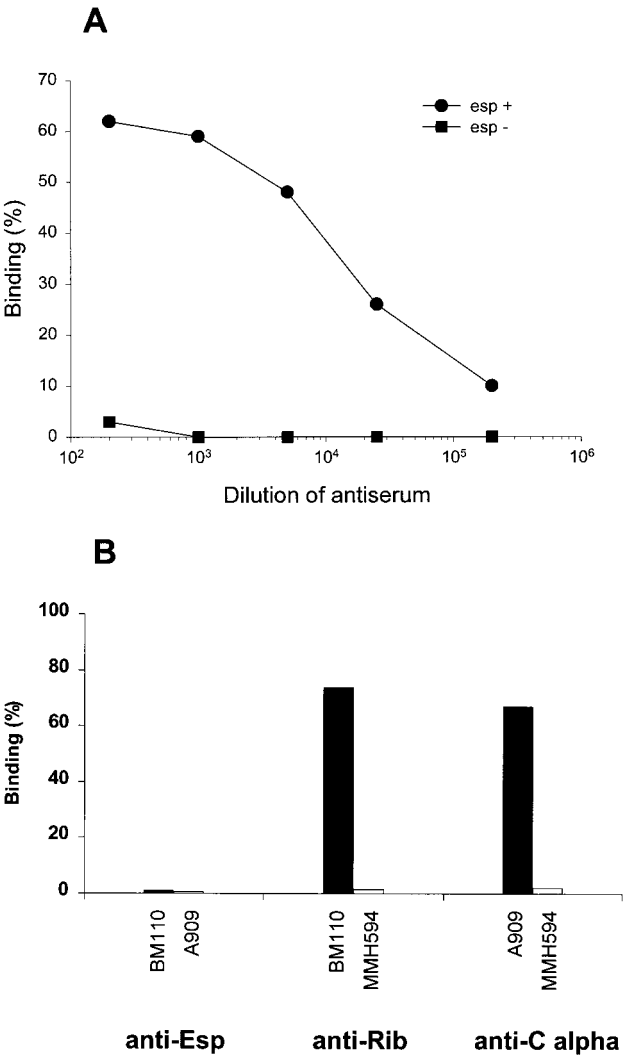


FIG. 5. Expression of Esp on the surface of *E. faecalis* and lack of cross-reactivity with GBS proteins Rib and C alpha. (A) Bacterial suspensions (1%) were incubated with dilutions of rabbit antiserum as indicated, and bound antibodies were detected after incubation with radiolabeled protein G. (B) Bacterial suspensions were incubated with antiserum (diluted 1:500), and bound antibodies were detected after incubation with radiolabeled protein G. The strains evaluated were *E. faecalis* MMH594 (this study), Rib-positive group B streptococcal strain BM110 (49), and C alpha-positive group B streptococcal strain A909 (33). The results are expressed as a percentage of the total radioactivity added. Values represent the mean of duplicate tests with similar results. Controls with preimmune rabbit serum were included in all experiments and gave completely negative results in all cases.

DISCUSSION

Although the ability of *E. faecalis* to cause serious disease is well recognized, not much is known about enterococcal virulence factors that contribute to its pathogenesis. For instance, factors that may influence the ability of enterococci to colonize host tissues, translocate across epithelial barriers, or survive in grossly different host environments are poorly understood. By using a genome-panning approach to identify new factors that may play a role in enterococcal virulence, we have identified and characterized, from an infection-derived *E. faecalis* isolate, a chromosomal gene, *esp*, which encodes a novel surface protein. The *esp* gene and the inferred protein product possess unique structural features, with an enrichment for the gene among disease-causing *E. faecalis* isolates. A statistically significant ($P < 0.001$) association with infection-derived *E. faecalis* isolates compared to isolates from healthy individuals provides indirect evidence for a contributory role for the Esp protein in virulence. In further support of this deduction, Esp was observed to be absent from less pathogenic *Enterococcus* species.

The high degree of conservation of A and C repeats at the nucleotide sequence level in *esp* suggests that the repeats are structurally important features. Specifically, repeats that exist as constant features of a gene would be expected to accumulate mutations, many of which may be silent at the amino acid level

but nonetheless would be detectable at the nucleotide sequence level. The repeats in *esp* are very nearly perfectly conserved over a large region. This provides evidence that the repeats are not static. Confirmatory evidence for the hypothesis that these repeat units may be hot spots for homologous recombination comes from the detection of considerable variation in the number of A and C repeat units among nonclonal infection-derived *E. faecalis* isolates, which appears to result from the precise addition or deletion of repeat blocks. "Shuffling" between the repeat units may lead to the expression of variant proteins that are identical at the amino and carboxy termini but differ in the number of repeat motifs. Such a phenomenon has been shown to occur in both Rib and C alpha proteins (31, 49) and is thought to be related to evasion of the immune response (32). Interestingly, none of the *esp*-positive isolates exhibited a complete loss of either the A or C repeats, suggesting that these regions are essential for maintaining an overall stable configuration of Esp or are otherwise of selective value.

The unusually large *esp* gene encodes a protein with inferred global structural similarity to Rib and C alpha proteins. Amino acid sequence identity, however, is localized to a stretch of 13 residues, within large repeat blocks that occur in each protein. The high degree of conservation of this core motif, together with its highly conserved periodicity, suggests an important functional role within proteins of this family. With no evidence so far to suggest that the repeat motifs in both Rib and C alpha mediate binding to host factors, it is tempting to speculate that this region serves only to maintain an elongated conformation of the protein at the cell surface. It is possible that the extreme amino-terminal end participates in interactions with the host. In a recent report (13), it was shown that the dipeptide repeat region of clumping factor in *Staphylococcus aureus* serves only to project the amino-terminal fibrinogen-binding domain well clear of the cell surface and the peptidoglycan layer. However, the high degree of conservation of the 13-residue core sequence argues in favor of a conserved function with limited permissible drift.

Typical bacterial transport signal sequences include a basic N-terminal region followed by a hydrophobic core and a potential signal peptidase cleavage site (46, 47). Fairly long signal peptides have been reported for gram-positive bacteria previously (15), and signal peptide sequences in *E. faecalis* and *Bacillus subtilis* are known to exhibit large hydrophobic segments (16, 43). The length of the inferred 49-residue signal peptide in Esp is somewhat unusual, but it is consistent with those reported for Rib and C alpha, which are 55 and 56 residues, respectively (49).

The consensus hexapeptide sequence LPXTGX is a highly conserved C-terminal motif in at least 50 gram-positive bacterial surface proteins studied to date (8) and is regarded as a universal cell wall-anchoring motif (40) in gram-positive bacteria. Substitutions are known to occur at positions 3 and 6 of the hexapeptide, but they rarely occur only at positions 4 and 5 and (so far) not at all at positions 1 and 2 (8). While deletion of the entire LPXTGX motif or mutation of the proline has been shown to abrogate anchoring, mutation of the conserved threonine residue seemed to have little effect (41). The inferred cell wall anchor motif of YPKTGE in Esp is the first instance, to our knowledge, where the conserved leucine at position 1 is replaced by a tyrosine residue. Despite this variation, the detection of Esp at the cell surface by using antibodies to the recombinant N-terminal region of Esp indicates that the amino terminus of Esp is displayed on the cell surface and suggests that the protein is anchored by its carboxy terminus.

Enterococcal surface components responsible for target cell or serum factor binding have not been identified; however, the motif RGD has been found in enterococcal aggregation substance and has been suggested to mediate binding to eukaryotic cells via integrins (9). A recent study (50) investigating the ability of enterococci isolated from human infections to interact with selected extracellular matrix and serum proteins found that *E. faecalis* specifically bound thrombospondin, lactoferrin, and vitronectin. Although no specific enterococcal cell surface structures that mediated the binding were identified, it was suggested that increased cell surface hydrophobicity in *E. faecalis*, compared to *E. faecium*, played a role in binding. The contribution of Esp to cell surface hydrophobicity, as well as its role in binding any of the extracellular matrix proteins, remains to be evaluated.

In conclusion, this study has shown that infection-derived *E. faecalis* isolates are significantly enriched for the *esp* gene, which encodes a potential virulence determinant. The structural gene is unique in that it allows for alternate forms of expression of the Esp protein. The Esp protein is surface localized, and such variations in structure at the bacterial cell surface may contribute to the ability of *E. faecalis* to evade detection by the immune system in an infected host or otherwise persist at sites of infection.

ACKNOWLEDGMENTS

We gratefully acknowledge the assistance of Richard Facklam, Barbara Murray, Thomas Russo, and Mark Zervos in providing clinical isolates of *E. faecalis* and Daniel Sahm in providing isolates representing different enterococcal species.

This work was supported, in part, by Public Health Service grants AI 40651 (V.S.) and AI 41108 (M.S.G.) from the National Institutes of Health and by Research to Prevent Blindness, Inc.

REFERENCES

1. **Aliberti, L. C.** 1995. Enterococcal nosocomial infection: epidemiology and practice. Gastroenterol. Nursing **18:**177–181.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Bevanger, L., and A. I. Naess.** 1985. Mouse-protective antibodies against the Ibc proteins of group B streptococci. Acta Pathol. Microbiol. Immunol. Scand. Sect. B **93:**121–124.
4. **Brendel, V., and E. N. Trifonov.** 1984. A computer algorithm for testing potential prokaryotic terminators. Nucleic Acids Res. **12:**4411–4427.
5. **Chow, J. W., L. A. Thal, M. B. Perri, J. A. Vazquez, S. M. Donabedian, D. B. Clewell, and M. J. Zervos.** 1993. Plasmid-associated hemolysin and aggregation substance production contribute to virulence in experimental enterococcal endocarditis. Antimicrob. Agents Chemother. **37:**2474–2477.
6. **Emori, T. G., and R. P. Gaynes.** 1993. An overview of nosocomial infections, including the role of the microbiology laboratory. Clin. Microbiol. Rev. **6:**428–442.
7. **Engvall, E.** 1980. Enzyme immunoassay, ELISA and EMIT. Methods Enzymol. **70:**419–439.
8. **Fischetti, V. A.** 1996. Gram-positive commensal bacteria deliver antigens to elicit mucosal and systemic immunity. ASM News **62:**405–410.
9. **Galli, D., F. Lottspeich, and R. Wirth.** 1990. Sequence analysis of *Enterococcus faecalis* aggregation substance encoded by the sex pheromone plasmid pAD1. Mol. Microbiol. **4:**895–904.
10. **Gilmore, M. S., R. A. Segarra, M. C. Booth, C. P. Bogie, L. R. Hall, and D. B. Clewell.** 1994. Genetic structure of the *Enterococcus faecalis* plasmid pAD1-encoded cytolytic toxin system and its relationship to lantibiotic determinants. J. Bacteriol. **176:**7335–7344.
11. **Grunstein, M., and D. Hogness.** 1975. Colony hybridization: a method for the isolation of cloned DNA's that contain a specific gene. Proc. Natl. Acad. Sci. USA **72:**3961.
12. **Harlow, E. D., and D. Lane.** 1988. Antibodies: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
13. **Hartford, O., P. Francois, P. Vaudaux, and T. J. Foster.** 1997. The dipeptide repeat region of the fibrinogen-binding protein (clumping factor) is required for functional expression of the fibrinogen-binding domain on the *Staphylococcus aureus* cell surface. Mol. Microbiol. **25:**1065–1076.
14. **Hoffmann, A., and R. G. Roeder.** 1991. Purification of His-tagged proteins in non-denaturing conditions suggests a convenient method for protein interaction studies. Nucleic Acids Res. **19:**6337–6338.

15. **Hollingshead, S. K., V. A. Fischetti, and J. R. Scott.** 1986. Complete nucleotide sequence of type 6 M protein of the group A streptococcus—repetitive structure and membrane anchor. J. Biol. Chem. **261:**1677–1686.
16. **Hols, P., A. Baulard, D. Garmyn, B. Delplace, S. Hogan, and J. Delcour.** 1992. Isolation and characterization of genetic expression and secretion signals from *Enterococcus faecalis* through the use of broad-host-range alpha-amylase probe vectors. Gene **118:**21–30.
17. **Huycke, M. M., and M. S. Gilmore.** 1995. Frequency of aggregation substance and cytolysin genes among enterococcal endocarditis isolates. Plasmid **34:**152–156.
18. **Huycke, M. M., and M. S. Gilmore.** 1997. In vivo survival of *Enterococcus faecalis* is enhanced by extracellular superoxide production. Adv. Exp. Med. Biol. **418:**781–784.
19. **Huycke, M. M., W. Joyce, and M. F. Wack.** 1996. Augmented production of extracellular superoxide by blood isolates of *Enterococcus faecalis*. J. Infect. Dis. **173:**743–746.
20. **Huycke, M. M., D. F. Sahm, and M. S. Gilmore.** 1998. Multiply resistant enterococci: the nature of the problem and an agenda for the future. Emerging Infect. Dis. **4:**239–249.
21. **Huycke, M. M., C. A. Spiegel, and M. S. Gilmore.** 1991. Bacteremia caused by hemolytic, high-level gentamicin-resistant *Enterococcus faecalis*. Antimicrob. Agents Chemother. **35:**1626–1634.
22. **Ike, Y., and D. B. Clewell.** 1992. Evidence that the hemolysin/bacteriocin phenotype of *Enterococcus faecalis* subsp. *zymogenes* can be determined by plasmids in different incompatibility groups as well as by the chromosome. J. Bacteriol. **174:**8172–8177.
23. **Ike, Y., D. B. Clewell, R. A. Segarra, and M. S. Gilmore.** 1990. Genetic analysis of the pAD1 hemolysin/bacteriocin determinant in *Enterococcus faecalis*: Tn*917* insertional mutagenesis and cloning. J. Bacteriol. **172:**155–163.
24. **Ike, Y., H. Hashimoto, and D. B. Clewell.** 1984. Hemolysin of *Streptococcus faecalis* subspecies *zymogenes* contributes to virulence in mice. Infect. Immun. **45:**528–530.
25. **Ike, Y., H. Hashimoto, and D. B. Clewell.** 1987. High incidence of hemolysin production by *Enterococcus* (*Streptococcus*) *faecalis* strains associated with human parenteral infections. J. Clin. Microbiol. **25:**1524–1528.
26. **Jett, B. D., M. M. Huycke, and M. S. Gilmore.** 1994. Virulence of enterococci. Clin. Microbiol. Rev. **7:**462–478.
27. **Jett, B. D., H. G. Jensen, R. E. Nordquist, and M. S. Gilmore.** 1992. Contribution of the pAD1-encoded cytolysin to the severity of experimental *Enterococcus faecalis* endophthalmitis. Infect. Immun. **60:**2445–2452.
28. **Kreft, B., R. Marre, U. Schramm, and R. Wirth.** 1992. Aggregation substance of *Enterococcus faecalis* mediates adhesion to cultured renal tubular cells. Infect. Immun. **60:**25–30.
29. **Larsson, C., M. Stalhammar-Carlemalm, and G. Lindahl.** 1996. Experimental vaccination against group B streptococcus, an encapsulated bacterium, with highly purified preparations of cell surface proteins Rib and alpha. Infect. Immun. **64:**3518–3523.
30. **Li, J., D. L. Kasper, F. M. Ausubel, B. Rosner, and J. L. Michel.** 1997. Inactivation of the alpha C protein antigen gene, bca, by a novel shuttle/suicide vector results in attenuation of virulence and immunity in group B *Streptococcus*. Proc. Natl. Acad. Sci. USA **94:**13251–13256.
31. **Madoff, L. C., S. Hori, J. L. Michel, C. J. Baker, and D. L. Kasper.** 1991. Phenotypic diversity in the alpha C protein of group B streptococci. Infect. Immun. **59:**2638–2644.
32. **Madoff, L. C., J. L. Michel, E. W. Gong, D. E. Kling, and D. L. Kasper.** 1996. Group B streptococci escape host immunity by deletion of tandem repeat elements of the alpha C protein. Proc. Natl. Acad. Sci. USA **93:**4131–4136.
33. **Michel, J. L., L. C. Madoff, K. Olson, D. E. Kling, D. L. Kasper, and F. M. Ausubel.** 1992. Large, identical, tandem repeating units in the C protein alpha antigen gene, bca, of group B streptococci. Proc. Natl. Acad. Sci. USA **89:**10060–10064.
34. **Murray, B. E., K. V. Singh, J. D. Heath, B. R. Sharma, and G. M. Weinstock.** 1990. Comparison of genomic DNAs of different enterococcal isolates using restriction endonucleases with infrequent recognition sites. J. Clin. Microbiol. **28:**2059–2063.
35. **National Nosocomial Infections Surveillance System.** 1997. National Nosocomial Infections Surveillance (NNIS) report, data summary from October 1986 to April 1997, issued May 1997. A report from the NNIS system. Am. J. Infect. Control **25:**477–487.
36. **Navarre, W. W., and O. Schneewind.** 1994. Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in gram-positive bacteria. Mol. Microbiol. **14:**115–121.
37. **Olmsted, S. B., G. M. Dunny, S. L. Erlandsen, and C. L. Wells.** 1994. A plasmid-encoded surface protein on Enterococcus faecalis augments its internalization by cultured intestinal epithelial cells. J. Infect. Dis. **170:**1549–1556.
38. **Russo, T.** 1997. Personal communication.
39. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
40. **Schneewind, O., D. Mihaylova-Petkov, and P. Model.** 1993. Cell wall sorting signals in surface proteins of gram-positive bacteria. EMBO J. **12:**4803–4811.
41. **Schneewind, O., P. Model, and V. A. Fischetti.** 1992. Sorting of protein A to the staphylococcal cell wall. Cell **70:**267–281.
42. **Shankar, V., and M. S. Gilmore.** 1997. Characterization of the *Enterococcus faecalis* alpha C protein homolog. Evidence for the expression of alternate forms in commensal and infection derived isolates. Adv. Exp. Med. Biol. **418:**1045–1048.
43. **Smith, H., A. de Jong, S. Bron, and G. Venema.** 1988. Characterization of signal-sequence-coding regions selected from the *Bacillus subtilis* chromosome. Gene **70:**351–361.
44. **Stalhammar-Carlemalm, M., L. Stenberg, and G. Lindahl.** 1993. Protein Rib: a novel group B streptococcal cell surface protein that confers protective immunity and is expressed by most strains causing invasive infections. J. Exp. Med. **177:**1593–1603.
45. **Triglia, T., M. G. Peterson, and D. J. Kemp.** 1988. A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. Nucleic Acids Res. **16:**8186.
46. **von Heijne, G.** 1984. Analysis of the distribution of charged residues in the N-terminal region of signal sequences: implications for protein export in prokaryotic and eukaryotic cells. EMBO J. **3:**2315–2318.
47. **von Heijne, G.** 1983. Patterns of amino acids near signal-sequence cleavage sites. Eur. J. Biochem. **133:**17–21.
48. **Warren, G., and D. J. States.** 1993. Identification of protein coding regions by database similarity search. Nat. Genet. **3:**266–272.
49. **Wastfelt, M., A. Delisse, T. Cabezon, and G. Lindahl.** 1996. Identification of a family of streptococcal surface proteins with extremely repetitive structures. J. Biol. Chem. **271:**18892–18897.
50. **Zareba, T. W., C. Pascu, W. Hryniewicz, and T. Wadstrom.** 1997. Binding of extracellular matrix proteins by enterococci. Curr. Microbiol. **34:**6–11.

Editor: E. I. Tuomanen