Identification and preliminary characterization of cell-wall-anchored proteins of Staphylococcus epidermidis

M. Gabriela Bowden, Wei Chen, Jenny Singvall, Yi Xu, Sharon J. Peacock, Viviana Valtulina, Pietro Speziale and Magnus Höök

1Center for Extracellular Matrix Biology, Institute of Biosciences and Technology, Texas A&M University Health Science Center, Houston, TX 77030-3303, USA
2Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand
3Department of Biochemistry, University of Pavia, 27100 Pavia, Italy

Staphylococcus epidermidis is a ubiquitous human skin commensal that has emerged as a major cause of foreign-body infections. Eleven genes encoding putative cell-wall-anchored proteins were identified by computer analysis of the publicly available S. epidermidis unfinished genomic sequence. Four genes encode previously described proteins (Aap, Bhp, SdrF and SdrG), while the remaining seven have not been characterized. Analysis of primary sequences of the Staphylococcus epidermidis surface (Ses) proteins indicates that they have a structural organization similar to the previously described cell-wall-anchored proteins from S. aureus and other Gram-positive cocci. However, not all of the Ses proteins are direct homologues of the S. aureus proteins. Secondary and tertiary structure predictions suggest that most of the Ses proteins are composed of several contiguous subdomains, and that the majority of these predicted subdomains are folded into β-rich structures. PCR analysis indicates that certain genes may be found more frequently in disease isolates compared to strains isolated from healthy skin. Patients recovering from S. epidermidis infections had higher antibody titres against some Ses proteins, implying that these proteins are expressed during human infection. Western blot analyses of early-logarithmic and late-stationary in vitro cultures suggest that different regulatory mechanisms control the expression of the Ses proteins.

INTRODUCTION

Staphylococcus epidermidis is a Gram-positive human skin commensal that has emerged as an important pathogen, causing infections in immunocompromised patients and in individuals that carry biomaterial implants.

The virulence factors of S. epidermidis are not completely characterized. This bacterium can secrete β-haemolysin and α-haemolysin, but the importance of these toxins in the disease process is unclear. In addition, the ability of S. epidermidis to attach to fibrinogen (Fg), fibronectin (Fn), vimentin, collagen and laminin has been reported (Bowden et al., 2002; Herrmann et al., 1988; Yu et al., 1994). However, only a few S. epidermidis cell-wall-associated proteins with potential adhesin function have been identified, namely SdrG (Fbe) (Davis et al., 2001; Nilsson et al., 1998), Embp (Williams et al., 2002), GehD (Bowden et al., 2002), AtE (Heilmann et al., 1997) and Ebps (Park et al., 1996). These adhesins bind to fibrinogen, fibronectin, collagen, vimentin and elastin, respectively.

The cell-wall-anchored (CWA) proteins of Gram-positive pathogens constitute a family of surface-exposed proteins that often interact with targets in the host (Schneewind et al., 1995). These interactions are frequently important for bacterial adherence and evasion from the host immune system. In general, the CWA proteins contain an N-terminal ~40 aa secretion signal sequence (S) required for secretory secretion, followed by a non-repeat A domain that frequently contains the ligand-binding site(s). The A regions of many S. aureus CWA proteins are arranged into 2-3 subdomains; each one folds into an IgG-like structure (Deivanayagam et al., 2002). A number of tandemly

Abbreviations: CWA, cell-wall anchored; Fg, fibrinogen; Fn, fibronectin; MSCRAMM, microbial surface component recognizing adhesive matrix molecule.

The secondary and tertiary structural prediction of the A domains of Staphylococcus epidermidis cell-wall-anchored proteins is available in Supplementary Table S1 with the online version of this paper at http://mic.sgmjournals.org/.
repeated sequences, often called B repeats, are frequently found C-terminal of the A-region. C-terminal of the repeated regions are the features that are required for sorting and anchoring the proteins to the cell wall, including a proline-rich wall-spanning region (W), the wall anchoring LPXTG motif (Schneewind et al., 1993), a hydrophobic transmembrane region and a positively charged cytoplasmic tail (Comfort & Clubb, 2004). Many S. aureus surface proteins are attached to the cell wall by a transpeptidase, called sortase A (SrtA), that recognizes the LPXTG motif, cleaves this sequence between the threonine and the glycine residue and then covalently links the processed polypeptide to the peptidoglycan wall. In a comparative analysis of sortase proteins from 241 bacterial genomes, the S. aureus SrtA was considered as the archetyp of the sortase A subfamily (Comfort & Clubb, 2004).

S. epidermidis RP62A encodes only one sortase, SrtA, that belongs to the sortase A subfamily, and that presumably plays a similar role in cell-wall protein chemistry. The genome of S. epidermidis ATCC12228 encodes an additional unique sortase, SrtC, not found in other staphylococci (Gill et al., 2005) and unclassified in the phylogenetic distribution of sortase homologues (Comfort & Clubb, 2004).

Until recently, S. epidermidis was known to express at least four CWA proteins: the S. aureus Bap homologue proteins Bhp, Aap, SdrG and SdrF. The Bhp protein has been proposed to promote primary attachment to abiotic surfaces, as well as intercellular adhesion during biofilm formation (Cucarella et al., 2001). Aap is necessary for the accumulative growth on polymer surfaces that results in visible biofilm formation (Hussain et al., 1997). SdrG and SdrF are members of the Sdr family of proteins; SdrG is necessary and sufficient for the attachment of S. epidermidis to surfaces coated with human Fg, whereas the function of SdrF has yet to be determined. The ligand-binding activity of SdrG has been localized to the N-terminal A region, specifically to residues 273–597 (Davis et al., 2001). We have recently described the X-ray crystal structure of the Fg-binding region of SdrG in complex with its ligand, a synthetic peptide mimicking the N-terminus of the human Fg β-chain. Analysis of this structure allowed us to propose a ligand-binding mechanism described as the ‘dock, lock and latch’ model (Ponnuraj et al., 2003). In this model, the ligand ‘docks’ between two IgG-like subdomains; subsequently, the protein undergoes a conformational change that leads to the ‘locking’ of the peptide within its ligand pocket and further ‘latching’ of a C-terminal strand that stabilizes the complex structure. It is possible that the dock, lock and latch mechanism is also employed by other proteins that display tandem IgG-like subdomains, such as those found in SdrF.

In this study, we searched the available S. epidermidis genome from strain RP62A for sequences encoding putative CWA proteins. We analysed the primary structures of the S. epidermidis CWA proteins and predicted their secondary and tertiary structures. We expressed recombinant versions of the non-repeated region of each protein and detected their expression on the S. epidermidis cell wall. In addition, we determined the presence of the corresponding genes among several S. epidermidis isolates and the expression of the proteins during staphylococcal infection in humans.

METHODOLOGY

Prediction of LPXTG proteins. The S. epidermidis RP62A unfinished genomic sequence was searched for ORFs with potential to encode for LPXTG proteins using a similar method to that published by Sillanpää et al. (2004). Briefly, preliminary sequence data were obtained from The Institute for Genomic Research (TIGR) website at http://www.tigr.org. New LPXTG motif proteins were only accepted if they were adjacent to a transmembrane domain followed by a positively charged C-terminal amino acid sequence. Initially, the genome sequence of S. epidermidis RP62A was downloaded from the TIGR web site (http://www.tigr.org) to a local Silicon Graphics computer. ORFs with a minimum of 500 nucleotides were predicted using Glimmer 2 (obtained from TIGR and installed locally). The ORFs were then translated into amino acid sequences using a custom-designed translation program capable of translating batch sequences. The process was automated using Unix C shell scripts. The peptide sequences were formatted into a BLAST-searchable database using Formatdb obtained from NCBI and installed locally. Amino acid sequences in the Gram-positive cell wall anchor family (Gram-pos_anchor) were obtained from the protein family database of alignments and hidden Markov methods (HMMs) (http://www.sanger.ac.uk/cgi-bin/ Pfam/). These sequences were analysed for the occurrence of amino acid residues at the LPXTG location. The pattern LPXTG[AIGANS] covers 95% of all the sequences in the family and was used as the search pattern in PHI-BLAST using locally installed stand-alone BLAST obtained from NCBI. The amino acid sequences of several proteins were used as templates in PHI-BLAST, including CNA (a collagen-binding protein) and protein A, two known CWA proteins of S. aureus with different sequences and structures. The output from the PHI-BLAST searches was combined and analysed to select for proteins that contain typical features of CWA proteins: a signal peptide at the N terminus (predicted using SignalP server at http://www.cbs.dtu.dk/services/ SignalP/), the LPXTG motif close to the C terminus (visual examination), followed by a hydrophobic transmembrane segment (predicted using the TMHMM server at http://www.cbs.dtu.dk/services/TMHMM-2.0/) and several positively charged residues at the C terminus (visual examination).

Protein structure analysis. Online bioinformatics tools were used to characterize protein sequences from the genomic analysis. The repeat domains were identified visually and using the RADAR (http://www.ebi.ac.uk/Radar/) algorithm. Similarity searches were carried out using NCBI BLAST (Altschul et al., 1997; Schaffer et al., 2001; Zhang et al., 1998). Protein secondary structure was predicted using the Predict Protein service (http://cubic.bioc.columbia.edu/predictprotein/). Secondary and tertiary structure was modeled using the 3D-PSSM Fold recognition server (http://www.csbio.ic.ac.uk/~3dpssm/). The 3D-PSSM server is a web-based method for protein fold recognition using primary and secondary sequence profiles coupled with information about secondary structure solvation potential. The PSSM E values represent a percentage of certainty in the folding prediction as follows: 0–0.05, 95%; 0.05–0.1, 90%; 0.1–0.25, 80%; 0.25–0.35, 70%; 0.35–0.7, 50%; 0.7–1, no certainty. Additional sequence analysis was performed using the Network Protein Sequence analysis from Pole BioInformatique Lyonnois (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_spped.html).
Western Blotting of lysostaphin-solubilized cell-wall fractions. *S. epidermidis* RP62A cultures were grown in tryptic soy broth (Difco) with aeration. Culture samples were taken at four different time points in the growth curve, corresponding to the logarithmic, late-logarithmic, early-stationary and stationary phase (OD<sub>600</sub> values of 0-7, 1-10, 2-05 and 2-33, respectively). The cells were harvested by centrifugation at 7000 g at 4 °C for 10 min and washed with PBS, pH 7-4 (140 mM NaCl, 270 µM KCl, 430 µM Na<sub>2</sub>HPO<sub>4</sub>, 147 µM KH<sub>2</sub>PO<sub>4</sub>). Cells were resuspended in digestion buffer [50 mM Tris/HCl, pH 7-5, 145 mM NaCl, 30 % (w/v) raffinose, 100 µg lysostaphin ml<sup>-1</sup>, 100 µg lysozyme ml<sup>-1</sup>, 10 µg DNase ml<sup>-1</sup>, 1 µg iodoacetamide ml<sup>-1</sup> and 1 mM PMFS] to an OD<sub>600</sub> value of 30. Proteins were released from the bacterial cell wall by incubation at 37 °C for 30 min with gentle rocking. Proteoplasts were removed by centrifugation at 3000 g for 20 min at room temperature, and the supernatant was taken as the ‘lysostaphin extract’. Samples were prepared for electrophoresis by boiling for 5 min in sample buffer [125 mM Tris/HCl, pH 7-5, 4 % (w/v) SDS, 20 % (v/v) glycerol, 10 % (v/v) β-mercaptoethanol, 0-002 % (w/v) bromophenol blue] and analysed in 10 % (w/v) acrylamide gels. The gels were transferred to nitrocellulose Western blotting membranes (Osmonics) by the semidy mid system (Bio-Rad). The membranes were blocked for 15 h at 4 °C with 5 % (w/v) BSA in PBS. Mouse or rabbit polyclonal antibodies were used at the dilutions indicated in Results. Bound antibody was recognized by incubation with anti-mouse or anti-rabbit IgG conjugated to alkaline phosphatase. The membrane was washed, and the CWA proteins were visualized with BCIP and 300 μl nitro blue tetrazolium chloride (NBT) with 5 % (w/v) BSA in PBS. Mouse or rabbit polyclonal antibodies was quantified by absorbance at 280 nm, as measured on a Beckman Du-70 UV–visible spectrophotometer. The molar extinction coefficient of the proteins was calculated using the method of Pace et al. (1995).

Isolation of IgG from convalescent patients’ sera. Sera were collected from 26 individuals convalescent from *S. epidermidis* infection, including peri-tonsillitis, infection at sites of surgery and infection of implanted medical devices. IgG was purified by chromatography on Vectron-SA (Biovectora). The concentration of the purified antibodies was quantified by absorbance at 280 nm with human IgG as the standard. Control IgG was obtained from eight healthy adults representing the same age group as the patients tested.

PCR screening of ses genes in clinical isolates. The clinical isolates screened in this analysis have been described previously (de Silva et al., 2001). The presence of ses genes was analysed by PCR amplification using primers designed to the non-repeated regions within each gene (Table 1). Genomic DNA from each bacterial strain was isolated using the Wizard Genomic DNA purification kit (Promega) with the addition of 30 µg lysostaphin ml<sup>-1</sup> at the lysis step. The PCR mixes for all nine sets of primers were as follows: 1 × reaction buffer, 1-5 mM MgCl<sub>2</sub>, 100 pmol forward and reverse primers, 100 ng template DNA, 200 µM dNTP mix and 2-5 U Taq polymerase. The PCR cycling conditions were as follows:

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Start (aa)</th>
<th>Restriction site</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SdrFA F</td>
<td>35</td>
<td>BamHI</td>
<td>5′-CCGGATCCGTTAGAATCATCAATTAG-3′</td>
</tr>
<tr>
<td>SdrFA R</td>
<td>677</td>
<td>HindIII</td>
<td>5′-CCGGATCCGTTAGAATCATCAATTAG-3′</td>
</tr>
<tr>
<td>SdrGA F</td>
<td>50</td>
<td>BamHI</td>
<td>5′-CCGGATCCGTTAGAATCATCAATTAG-3′</td>
</tr>
<tr>
<td>SdrGA R</td>
<td>597</td>
<td>KpnI</td>
<td>5′-CCGGATCCGTTAGAATCATCAATTAG-3′</td>
</tr>
<tr>
<td>SesA F1</td>
<td>40</td>
<td>Spil</td>
<td>5′-TGACCCGGGTTAGATACCTCTTATAGCT-3′</td>
</tr>
<tr>
<td>SesA R1</td>
<td>570</td>
<td>Smal</td>
<td>5′-TGACCCGGGTTAGATACCTCTTATAGCT-3′</td>
</tr>
<tr>
<td>SesC F1</td>
<td>38</td>
<td>BamHI</td>
<td>5′-TGACCCGGGTTAGATACCTCTTATAGCT-3′</td>
</tr>
<tr>
<td>SesC R1</td>
<td>676</td>
<td>Sal</td>
<td>5′-TGACCCGGGTTAGATACCTCTTATAGCT-3′</td>
</tr>
<tr>
<td>Bhp (SesD) F1</td>
<td>46</td>
<td>Spil</td>
<td>5′-TGACCCGGGTTAGATACCTCTTATAGCT-3′</td>
</tr>
<tr>
<td>Bhp (SesD) R1</td>
<td>689</td>
<td>Smal</td>
<td>5′-TGACCCGGGTTAGATACCTCTTATAGCT-3′</td>
</tr>
<tr>
<td>SesE F1</td>
<td>48</td>
<td>BamHI</td>
<td>5′-TGACCCGGGTTAGATACCTCTTATAGCT-3′</td>
</tr>
<tr>
<td>SesE R1</td>
<td>300</td>
<td>Sal</td>
<td>5′-TGACCCGGGTTAGATACCTCTTATAGCT-3′</td>
</tr>
<tr>
<td>Aap (SesF)F1</td>
<td>55</td>
<td>BamHI</td>
<td>5′-TGACCCGGGTTAGATACCTCTTATAGCT-3′</td>
</tr>
<tr>
<td>Aap (SesF) R1</td>
<td>600</td>
<td>Sal</td>
<td>5′-TGACCCGGGTTAGATACCTCTTATAGCT-3′</td>
</tr>
<tr>
<td>SesG F1</td>
<td>49</td>
<td>BamHI</td>
<td>5′-TGACCCGGGTTAGATACCTCTTATAGCT-3′</td>
</tr>
<tr>
<td>SesG R1</td>
<td>630</td>
<td>Sal</td>
<td>5′-TGACCCGGGTTAGATACCTCTTATAGCT-3′</td>
</tr>
<tr>
<td>SesH F1</td>
<td>47</td>
<td>BamHI</td>
<td>5′-TGACCCGGGTTAGATACCTCTTATAGCT-3′</td>
</tr>
<tr>
<td>SesH R1</td>
<td>200</td>
<td>Sal</td>
<td>5′-TGACCCGGGTTAGATACCTCTTATAGCT-3′</td>
</tr>
<tr>
<td>SesI F1</td>
<td>40</td>
<td>BamHI</td>
<td>5′-TGACCCGGGTTAGATACCTCTTATAGCT-3′</td>
</tr>
<tr>
<td>SesI R1</td>
<td>570</td>
<td>Sal</td>
<td>5′-TGACCCGGGTTAGATACCTCTTATAGCT-3′</td>
</tr>
</tbody>
</table>

Table 1. Synthetic oligonucleotides used in this study

Italic type, stop codon; F, oligonucleotide primer in the forward direction; R, in the reverse direction. The underlined sequences correspond to the restriction sites used.
94 °C for 3 min, followed by 30 cycles of 94 °C for 1 min, 50 °C for 1 min and 72 °C for 1 min, which was suitable for amplifying the PCR products for each gene. Aliquots of the reaction mixtures were analysed by 1 % agarose gel electrophoresis. Positive and negative controls (reaction mixture without DNA) were included in each PCR run.

RESULTS

Identification of the LPXTG proteins

The unfinished *S. epidermidis* RP62A genome sequences were searched for ORFs with the potential to encode surface protein genes belonging to the CWA protein family. The identified ORFs encoded polypeptides predicted to have an N-terminal secretion sequence along with a C-terminal sorting signal, the LPXTG motif, followed by a putative membrane-spanning hydrophobic domain and several positively charged residues at the C terminus. The described domain organization is common in proteins that are anchored on the cell wall by sortases such as SrtA (Comfort & Clubb, 2004; Pallen et al., 2001). From the eleven identified LPXTG-predicted proteins, four had been previously described (Aap, Bhp, SdrF and SdrG). For the remaining seven predicted proteins, we have adopted the terminology proposed by Mazmanian et al. (2001) Roche et al. (2003) and Gill et al. (2005), designating them as *Staphylococcus epidermidis* surface (Ses) proteins. The GenBank accession numbers for the sequences described in this work (SERPXXXX) are indicated within the text.

*S. epidermidis* surface proteins: deduced amino acid sequences and predicted folds

Initially, to characterize the predicted proteins, we examined their amino acid primary sequences and predicted their secondary and tertiary structures using computer-modelling programs (see Supplementary Table S1, available with the online version of this paper at http://mic.sgmjournals.org/). These models are sometimes useful to define domains that may function as modular units and predict their arrangement within each protein. The primary sequence organization of these proteins is represented in Fig. 1 and their most prominent features are summarized in Table 2. In general, the primary sequence organization of the deduced Ses proteins is similar to those found in the *S. aureus* Sas proteins, having typical features of cell-surface proteins that are covalently anchored to the cell surface of Gram-positive bacteria. Based on published data and on our own structural analysis, we grouped these proteins into subclasses.

Serine-aspartate-repeat (Sdr) proteins (SdrG and SdrF)

The *S. epidermidis* Sdr subclass includes SdrF (SERP0026), SdrG (SERP0207) and SdrH (SERP1487). SdrF is truncated in RP62A, but is present in its entirety in many other
S. epidermidis isolates. The specific ligands for SdrF and SdrG have not been identified; SdrG is a fibrinogen-binding protein. At their N terminus, SdrF and SdrG contain signal sequences (52 and 50 residues, respectively), predicted to facilitate the secretion of the proteins across the bacterial membrane. The predicted molecular masses of the mature SdrF and SdrG are 175 and 93.7 kDa, respectively. The SD repeat regions are present in the SdrG and SdrF proteins, as well as in the S. aureus Sdr proteins, proximal to the LPXTG motifs. Structurally distinct, non-repeat-containing domains called A regions are present at the N termini of the S. epidermidis Sdr proteins. The A regions of SdrF and SdrG are 22% identical to each other and 20–35% identical to the A regions of the S. aureus Cna, FnbpA, FnbpB, SasA and other Sdr proteins. Topologically, the A regions of SdrF and SdrG can be subdivided into three subdomains: N1, N2 and N3. The N1 subdomain is frequently degraded by proteases; the region encompassing N2–N3 is proteolytically stable both in vivo and in vitro (Ni Eidhin et al., 1998; Perkins et al., 2001). We have solved the crystal structure of the N2–N3 SdrG region (amino acids 273–597) in complex with its ligand, a fibrinogen-derived peptide. The solved structure shows that each N2 and N3 subdomain folds into two β-sheets to form ‘sandwich’-like structures (Deivanayagam et al., 2002), and that the ligand-binding region is formed by a cleft present between N2 and N3. A detailed analysis of the SdrG, ClfA and ClfB crystal structures, along with computer-generated structural models of other CWA proteins, allowed us to predict that this type of N2–N3 structural arrangement is present in many other Gram-positive CWA proteins, including S. epidermidis SdrF and SesG (Ponnuraj et al., 2003). Immediately following the SdrF and SdrG A regions, a number of repetitive amino acid sequences are present, termed B repeats. The B repeat regions of the SdrF and SdrG proteins are 110–119 residues long, and their amino acid composition is 43–85% identical. These B regions also show a high degree of identity to those found in the S. aureus Sdr proteins SdrC, SdrD and SdrE (39–73%) (McCrea et al., 2000). Downstream from the B repeats, a region composed of repetitive Ser and Asp alternating residues is found; these SD-repeat regions are 558 and 56 residues long in SdrF and SdrG, respectively. The SD-repeat region is proposed to fold as a coil that traverses the cell wall, extending the A region and B repeats outward from the cell surface (Kehoe, 1994). The SD-repeat region is required for the proper display of the fibrinogen-binding domain on the S. aureus clumping factor ClfA (Hartford et al., 1997; Kehoe, 1994), and has been proposed to have a similar function in other Sdr proteins. The SD-repeat region is followed by the cell-sorting motif LPXTG, a hydrophobic membrane-spanning region, and a stretch of positively charged residues.

The SdrH protein has an unusual organization, with a short 60-residue A region at its N terminus followed by an SD-repeat region, a unique 277-residue C region and a C-terminal hydrophobic segment. SdrH lacks an LPXTG motif and has not been localized to the cell wall fraction; for these reasons it is not included in this study.

### Biofilm-associated proteins (Aap and Bhp)

#### Accumulation-associated protein (Aap)

Aap was first described by Hussain et al. (1997) as a 140 kDa extracellular protein necessary for the accumulation of...
S. epidermidis strains onto polymer surfaces. In the RP62A sequence posted by TiGR, it is also named SesF (SERP2398) (Gill et al., 2005). As observed in other CWA proteins, the Aap amino acid sequence is organized into discrete domains. Immediately downstream from the predicted signal sequence, the Aap protein is composed of several regions: an N-terminal domain composed of short, 16 aa repeats, the all-β A region, 13 repeats of 128 aa, and 19 repeats of 6 aa that are rich in Pro residues. At the C terminus, the sortase-recognition sequence LPDTG and a predicted transmembrane region are found, followed by a region rich in positively charged residues.

The primary structure of Aap resembles those of the S. aureus SasG and Pls proteins. The overall amino acid sequence of Aap is 54.7% identical and 78.1% similar to SasG, and 36.1% identical and 62.8% similar to Pls. Both Aap and Pls sequences contain a 16 aa motif repeated 10 times at the extreme N terminus of the A region. These 16 aa motifs in Aap are 22% identical and 62-4% similar to those present in Pls. These short amino acid repeats are predicted to fold as a β-barrel sandwich hybrid, an all-β fold found in biotinyl/lipoyl carrier proteins and domains, such as the lipoyl domain of dihydrolipoamide acetyltransferase from Homo sapiens (PSSM E value = 0.012). The overall Aap A region (residues 263–608) is 34.2% identical to that of Pls (residues 283–608) and 40.9% identical to that of SasG (residues 50–428). The Aap A region is predicted to fold into an all-β structure, highly homologous to ComA-like lectins/glucanases (PDB entry 1fat, PSSM E value = 0.0507) (Dao-Thi et al., 1996), proteins that bind to sugars such as mannose, galactose, and glucose. Within the Aap A region, a 212 aa sequence (residues 396–608) is highly conserved and in a similar position in the S. aureus proteins SasG (residues 207–428) and Pls (residues 391–608) (Roche et al., 2003). These 212 residues of Aap share a 50-6% and 56-13% amino acid identity with those present in Pls and SasG, respectively.

Immediately following the A region, Aap contains 12-8 B repeats composed of 128 aa, encompassing amino acids 577–2220. These repeats are 60–67% identical to those present in the CWA proteins S. aureus SasG and Pls, and are predicted to fold into β-rich structures, such as the Strep-tococcus pneumoniae penicillin-binding protein 2X (PBP-2X, PSSM E value = 0.00544). The Aap sequence posted by Hussain et al. (1997) (GenBank accession no. CAB77251) only displays about 3-8 B repeats, in contrast to the 12-8 B repeats we found in the RP62A sequence. This variability in the number of B repeats has been documented before for other microbial surface component recognizing adhesive matrix molecules (MSCRAMMs), such as S. aureus SasG (Roche et al., 2003). Subsequent to the B repeats, Aap has a region of 19 PG(T/K)PAE tandem repeats that are predicted to form a flexible domain, and presumably span the bacterial cell wall. These short, proline-rich repeats are reminiscent of those present at the C-terminal region of S. aureus protein A (Spa, accession no. Q99X2). Residues 268–355 in Spa form 12 PGKED(N/G)(N/K)K tandem repeats, a region that has been shown to transverse the cell wall (Guss et al., 1984).

**Bap homologue protein (Bhp).** Bhp was first posted by Toromo and co-workers as the S. aureus Bap (Biofilm-associated protein) homologue from S. epidermidis RP62A that promotes biofilm formation (NCBI accession no. AAK29746). In the RP62A sequence posted by TIGR, Bhp is also named SesD (SERP2395) (Gill et al., 2005). Following the signal sequence, the Bhp protein is composed of an A region (amino acids 44–722) predicted to fold as an all-β domain topologically homologous to the extracellular domain of the human LDL receptor (Rudenko et al., 2002) and glyc an hydrolases such as bacterial sialidases (6-sheet β-propeller, PSSM E value = 0.013) (Gaskell et al., 1995). Immediately following its A region, Bhp contains multiple copies of tandem repeats. The repeats present in Bhp are 45–57% identical to those occurring in the S. aureus Bap and SasA proteins. These 18 repeats are predicted to fold as all-β structures, homologous to the domains formed in the Xenopus laevis C-cadherin ectodomain (PSSM E value = 0.0028) (Boggon et al., 2002) and the D domains of Yersinia pseudotuberu-losis invasin (PSSM E value = 0.00393) (Hamburger et al., 1999), both members of the immunoglobulin superfamily (IgSF) fold.

**Novel, repeat-containing proteins (SesA, SesE, SesG and SesH)**

**SesA.** The SesA sequence (SERP1316) encodes a protein predicted to fold mostly into α-helices and coils. The A region (amino acids 38–959) is predicted to fold into an α-β superhelix similar to those found in the constant regulatory domain of protein phosphatase 2A (PR65/A subunit H. sapiens, PSSM E value = 0.325). The repeat region (amino acids 960–3576) has 34 repeats, each of 77 aa. These repeats are 25–29% identical to those present in the Mrp (Wu & De Lencastre, 1999) and FmtB (Komatsuzawa et al., 2000) proteins of S. aureus Mu50 and N315. The repeat region of SesA appears as a single polypeptide (Fig. 2). Therefore, our results suggest that SesA is associated to the cell wall, despite deviating from the canonical LPXTG motif.

**SesE.** The SesE sequence (SERP0719) has been annotated as lipoprotein vsaC (accession no. NP_764383) in the S. epidermidis ATCC 12228 genome analysis (Zhang et al., 2003). Typically, staphylococcal lipoproteins such as
β-lactamase (BlaZ) contain a type II leader peptide (Chan, 1986; Wang & Novick, 1987) and the lipid modification consists of a glyceride thioether linked to cysteine at the N terminus of the mature polypeptide chain (Nielsen & Lampen, 1983). However, the SesE sequence does not display any of the conserved elements necessary for N-terminal lipid modification, nor has it been shown to localize to the membrane fraction or be lipid-modified. On the contrary, we showed that it has the sortase-specific sequence LPETG on its C-terminus, and we could detect SesE in the cell wall extract after *S. epidermidis* digestion with lysozyme (Fig. 2). Therefore, we concluded that SesE is a peptidoglycan-anchored protein. According to the 3D-PSSM program, the SesE A region (amino acids 46–409) is predicted to fold as an all-β structure, similar to those found in viral coat and capsid proteins (PSSM E value = 0.0909). Immediately downstream from its A region, SesE has three distinct types of repeats (B, C and D). This repeat region is also predicted to fold into an all-β structure (PSSM E value = 0.0691). However, using the consensus secondary structure prediction from PBIL (Lyon, France), this protein is predicted to be composed of up to 85% coil. The discrepancies between these two software analyses, at this point, cannot be explained.

**SesG.** The SesG (SERP1482) non-repeated A region (amino acids 47–620) is predicted to fold into two IgG-like domains homologous to those found in SdrG, flanked by α-helices comprising its N and C termini, linking the A region to the downstream repeat region. The repeat region is composed of 12 subdomains of 105 aa (amino acids 621–1898) that are 22% identical to those of Bap from *S. aureus*. The SesG repeats are predicted to fold into all-β structures homologous to the Type III domain repeats of human fibronectin (FNIII) (PSSM E value = 0.012).

**Ebh.** The Ebh protein (SERP1011), the largest one from the family, is composed of 10 203 amino acids and contains at least 13 distinct subdomains, some of which encompass repeated sequences. Although it has been reported to be associated to the cell wall, Ebh does not have a canonical LPXTG motif. The sequence LKGTG (amino acids 10 032–10 036) seems to be the closest homologue to an LPXTG motif. Because it may not be anchored onto the peptidoglycan and is extremely large and complex, we decided not to analyse this sequence any further.

**Novel, non-repeat-containing proteins (SesH, SesI and SesC) SesH, SesI and SesC.** These three smaller (237–676 aa) proteins do not contain predicted repeats: SesH (SERP1483), SesI (SERP1654) and SesC (SERP2264). SesH and SesI are predicted to fold as α and β mixed proteins. According to the 3D-PSSM prediction, the SesH and SesI proteins may fold as structures homologous to a *Plasmodium falciparum* antigen (PSSM E value = 0.459) and the choline-binding domain of the *Streptococcus pneumoniae* autolysin (PSSM E value = 2.03), respectively. SesC is predicted to fold as 47% α-helix and 50.4% coil, with a tertiary structure homologous to that of the all-α class of proteins, such as annexins (PSSM E value = 0.965).
Expression of the Ses proteins in *S. epidermidis*

To determine whether or not the predicted surface proteins are expressed by *S. epidermidis*, we used immunological methods. Specific mouse antisera were raised to recombinant proteins representing the A regions of each different surface protein. The PCR primers used to clone the predicted A regions are described in Table 1, and their amino acid lengths are described in Table 2. We were able to clone and raise antibodies against most of the CWA proteins, with the exception of SesB. Using these A-region–specific antisera, we showed that the predicted surface proteins are expressed in the *S. epidermidis* RP62A strain (Fig. 2). Interestingly, Western blot analysis of lysozyme extracts from samples collected at different time points during bacterial growth revealed that certain surface proteins are more abundant in extracts from samples collected during the early log growth phase compared to those from the stationary phase, and vice versa. These data suggest that certain surface proteins are differentially expressed during different stages of growth. For example, SdrG, SdrF, SesG and Bhp are abundant in lysozyme extracts obtained from early exponential growth cultures, but are hardly detectable in stationary-phase extracts from the same culture. In the stationary-phase extracts, SdrG, SdrF and SesG are apparently degraded, since in our Western assays we detected low-molecular-mass polypeptides that reacted with the respective anti-SdrG, anti-SdrF and anti-SesG antibodies (Fig. 2). These exponential-phase–expressed proteins may be important in the early stages of the infection process. On the other hand, a protein involved in biofilm formation, such as the Aap protein, was more abundant in extracts obtained during stationary phase compared to the same culture sampled in the early exponential growth phase. It has been previously reported that Aap is necessary during the accumulation phase of biofilm formation and, in certain aspects, this phase can imitate the environment present during the stationary phase of liquid growth. In our Western blot analysis, Aap could be detected as two polypeptides, one of the predicted molecular mass and another of smaller molecular mass. It is possible that the smaller-molecular-mass band represented a degradation product; this smaller polypeptide seemed to be more abundant in the stationary-phase samples. Other proteins that were abundant in stationary-phase samples were SesA and SesI. These data suggest that SesA and SesI could also be expressed during biofilm formation.

The SesH, SesI and SesG genes may be associated with disease isolates

To determine if any of the novel surface-protein genes are associated with normal skin colonization or disease, the presence or absence of the 10 genes was determined for clinical isolates obtained from pre-term infants in the neonatal intensive care unit at the John Radcliffe Hospital, Oxford, UK (de Silva et al., 2001). The *S. epidermidis* strains analysed in this study included 11 isolates obtained from bacteraemic patients. Thirty-four isolates obtained from the blood of patients who did not show evident symptoms of sepsis were considered blood contaminants, and 71 skin isolates obtained from patients with invasive disease and healthy babies were considered to be colonizing skin flora (Table 3). Genomic DNA was analysed by PCR with primers specific for the region encoding the A domains in each gene. There was no significant difference among the groups for the prevalence of genes, so it is not clear whether the presence of certain genes is linked to disease or colonization isolates. A group of genes is found in all the strains tested.

### Table 3. Distribution of genes encoding CWA proteins among *S. epidermidis* clinical isolates

<table>
<thead>
<tr>
<th>Gene</th>
<th>Bacteraemia*</th>
<th>Blood culture contaminant†</th>
<th>Colonizing skin flora‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>sdrF</td>
<td>6/11 (54%)</td>
<td>22/34 (65%)</td>
<td>48/71 (67%)</td>
</tr>
<tr>
<td>sdrG</td>
<td>11/11 (100%)</td>
<td>34/34 (100%)</td>
<td>71/71 (100%)</td>
</tr>
<tr>
<td>sesA</td>
<td>11/11 (100%)</td>
<td>34/34 (100%)</td>
<td>71/71 (100%)</td>
</tr>
<tr>
<td>sesC</td>
<td>11/11 (100%)</td>
<td>34/34 (100%)</td>
<td>71/71 (100%)</td>
</tr>
<tr>
<td>bhp</td>
<td>1/11 (9%)</td>
<td>0/34 (0%)</td>
<td>9/71 (13%)</td>
</tr>
<tr>
<td>sesE</td>
<td>11/11 (100%)</td>
<td>34/34 (100%)</td>
<td>71/71 (100%)</td>
</tr>
<tr>
<td>aap</td>
<td>3/11 (27%)</td>
<td>17/34 (50%)</td>
<td>31/71 (43%)</td>
</tr>
<tr>
<td>sesG</td>
<td>2/11 (18%)</td>
<td>3/34 (9%)</td>
<td>5/71 (7%)</td>
</tr>
<tr>
<td>sesH</td>
<td>8/11 (72%)</td>
<td>26/34 (76%)</td>
<td>44/71 (62%)</td>
</tr>
<tr>
<td>sesI</td>
<td>5/11 (45%)</td>
<td>10/34 (29%)</td>
<td>24/71 (34%)</td>
</tr>
</tbody>
</table>

*Two or more independent blood cultures positive for an identical *S. epidermidis* isolate, in association with other symptoms consistent with sepsis.
†Blood culture isolates from infants who did not meet the criteria for bacteraemia.
‡Skin (axilla and earlobe) isolates from healthy infants and patients with invasive disease.
(sdrG, sesA, sesE, sesC), whereas other genes are not (sdrF, bhp, aap, sesI, sesH, sesG). These observations may be explained by the fact that *S. epidermidis* clinical isolates are genetically diverse. This genetic diversity is reflected in their dissimilar genomic profiles, which can be detected by PFGE (de Silva et al., 2001).

**ELISA-based screen using convalescent patients’ sera**

To determine whether or not the surface proteins were expressed during infection in the host, recombinant predicted A regions from SdrG, SdrF, Bhp, Aap, SesA [FmtB (factor that affects the methicillin resistance levels in the presence and absence of Triton X-100)-like] (Komatsuzawa et al., 2000), SesG, SesE [VsaC (variable surface antigen C)-like] (Shen et al., 2000), SesH and SesI were purified. These purified recombinant proteins were used in ELISA-based assays to screen for antibodies in sera isolated from patients convalescing from *S. epidermidis* infections. Sera were collected from 25 donors (ages 20–65 years old), IgGs were purified from each sample, and their reactivity with each recombinant protein was tested (Fig. 3). Sera 1–8 were obtained from healthy donors. The reactivity of the sera to the recombinant proteins was highly variable; it is possible that *S. epidermidis* expresses certain surface proteins depending on the type of tissue infected. We therefore examined the type of infection that each patient had suffered, collecting as much information as possible. Sera 9, 17, 18, 19, 21, 22, 23 and 24 were obtained from patients diagnosed with *S. epidermidis* infections following the insertion of intravenous central catheters, serum 20 was obtained from a patient with an infected orthopaedic device. Biofilm formation was observed in all the explanted medical devices. Sera 14 and 15 were obtained from patients affected with native valve endocarditis, and the clinical complications in these patients were heart blockage, heart failure and stroke. Sera 10–13 were collected from patients undergoing continuous ambulatory peritoneal dialysis. Serum 16 was from a patient affected with prosthetic valve endocarditis. Thrombi and vegetations were observed on the surface of the explanted prosthetic valve. We observed higher titres of antibodies to SdrG, SesH, SesI and SesG compared to the control IgG, suggesting that these proteins are expressed during *S. epidermidis* infections and that they elicit an immune response. The rest of the proteins reacted above background levels occasionally; SesA, Aap, Bhp and SdrF were recognized by a few serum samples from the whole group. The low reactivity of this group of proteins could be attributed to their decreased expression during infection in the host, or to their poor ability to elicit the production of specific antibodies.

**DISCUSSION**

The adherence properties of *S. epidermidis* suggest that this organism expresses MSCRAMMs. For example, *S. epidermidis* expresses a fibrinogen-binding protein (SdrG) that has a structural organization similar to that of the *S. aureus* fibrinogen-binding proteins ClfA and ClfB. In contrast to *S. aureus, S. epidermidis* is much less toxigenic. *S. epidermidis* can secrete a few extracellular exoenzymes, such as a metalloprotease with elastase activity (Teufel & Gotz, 1993); a serine protease that catalyses the processing of the epidermin precursor peptide (Geissler et al., 1996) and two lipases proposed to be involved in skin colonization (Bowden et al., 2002; Farrell et al., 1993; Longshaw et al., 2000). *S. epidermidis* can also secrete β-haemolysin and δ-haemolysin, proteins that form pores in mammalian cell membranes and lead to the lysis of cells such as erythrocytes. However, with the exception of native valve endocarditis, *S. epidermidis* rarely causes pyogenic infections in a healthy host, and there is little evidence to suggest a role for this organism in toxin-mediated syndromes. Consequently, it is likely that CWA proteins and other adhesins...
play a major role during the early stages of *S. epidermidis* colonization and the later establishment of disease. In an attempt to understand the events that lead to biomaterial colonization and the development of infection, we have searched the publicly available *S. epidermidis* RP62A unfinished genome for ORFs encoding proteins predicted to be anchored to the cell wall and exposed to the bacterial cell surface. We found eleven ORFs predicted to encode for CWA proteins, four that had already been identified and seven that were novel. Comparing our data to those obtained in the analysis of the *S. epidermidis* ATCC 12228 genome, we found that eight of the eleven ORFs present in the RP62A strain were also identified in the ATCC 12228 strain \( (Zhang \ et \ al.,\ 2003) \). These ORFs encode the following proteins: SdrF \( (SERP2395) \), SdrG \( (SERP0331) \), Aap \( (SERP0175) \), SesA \( (SERP1429) \), SesB \( (SERP2162) \), SesC \( (SERP2232) \), SesE \( (SERP0828) \) and SesH \( (SERP1628) \). Two hypothetical proteins are present in strain ATCC 12228 but not in RP62A \( (SE1500 \ and \ SE1501) \), whereas the ORFs encoding Bhp, SesG and SesI are only found in the *S. epidermidis* RP62A genome.

The primary sequence organization of most of the predicted proteins has features typical of cell-surface proteins that are covalently anchored to the cell wall. These features include a signal sequence at the N terminus of the proteins, predicted to facilitate secretion of the proteins across the bacterial cell membrane. Frequently, a distinct, non-repeat region is found following the signal sequence. In characterized MSCRAMMs, this region encompasses about 600 aa and represents the ligand-binding domain, also termed the A region. The A region is often followed by long stretches of repeated amino acid sequences. Curiously, most surface proteins are predicted to fold into subdomains that are rich in \( \beta \)-sheets, including the N-terminal ligand-binding region called the A domain and the repeated B regions. With the exception of *S. aureus* FnbpA, where the repeat region binds to fibronectin, these long repeats have no known adhesive function. The folding of subdomains into \( \beta \)-strand-rich structures is a common feature found in previously characterized *S. epidermidis* and *S. aureus* MSCRAMMs \( (Deivanayagam \ et \ al., 1999, 2000, 2002; Ponnuraj \ et \ al., 2003) \). The long B repeats present in most proteins are predicted to fold into consecutive subunits in a 'beads on a string' arrangement. This type of organization is common in flexible protein domains, such as those present in the fibronectin type III repeat region \( (FNIII) \) and in the C-cadherin ectodomain. Many of these repeated regions are predicted to fold similarly to those present in the *S. aureus* collagen binding MSCRAMM Cna. The Cna B repeats exhibit a fold that also resembles the immunoglobulin-like (IgG-like) domains. Crystallographic studies of the Cna B repeats show that these domains pack in a zig-zag fashion, like an accordion; they might stretch and contract from the bacterial cell wall and thus aid in the projection of the A region away from the cell surface \( (Deivanayagam \ et \ al., 2000) \). Presumably, the repeat regions of the *S. epidermidis* proteins have a similar function. Toward the C terminus, the long-repeat region is often followed by a variable subdomain composed of short amino acid repeats, such as the SD region in SdrG and SdrF, or the (PG(T/K)PAE) repeat region in Aap. In the cases of the *S. aureus* Protein A and ClfA, these repeated regions span the peptidoglycan wall and are important for efficient protein secretion and assembly on the bacterial surface. We propose that in the *S. epidermidis* proteins, these regions may serve a similar function. C terminal of the repeated regions are the features that are required for sorting and anchoring the proteins to the cell wall, including a prolinerich wall-spanning region \( (W) \), the wall-anchoring LPXTG motif \( (Schneewind \ et \ al., 1993) \), a hydrophobic transmembrane region and a positively charged cytoplasmic tail.

With the exception of the fibrinogen adhesin SdrG, the specific binding activities of the CWA proteins are unknown. The structural analysis and predicted tertiary structures of the A regions may help to guide us in the search for their respective ligands. For example, the A region of Aap is predicted to fold as a lectin, so we will try to find if this region is able to bind to carbohydrates. The A region of Bhp is predicted to fold in a similar manner to the low density lipoprotein (LDL) receptor, which binds to mannose. The fact that both Aap and Bhp were found to be important during biofilm formation suggests that these proteins may play a role in intercellular adhesion. How this effect is achieved is not known, but it is possible that these proteins mediate cellular accumulation and intercellular adhesion by associating with polysaccharide intercellular adhesin (PIA), a *S. epidermidis* N-acetyl-\( \beta \)-(1,6)-glucosamine extracellular polymer required for biofilm formation \( (Joyce \ et \ al., 2003) \).

We have shown that the predicted proteins are expressed on *S. epidermidis* under *in vitro* culture conditions and *in vivo*, as assessed by the presence of antibodies against these proteins in the sera of infected patients. In culture, some of the proteins seem to be preferentially expressed during either the logarithmic or the stationary growth phase. This observation implies that the genes that encode them are under some type of cell-density regulation, such as the *agr* system. In the patients’ sera, we observed higher titres of antibodies to SdrG, SesA, SesH, SesI and SesG compared to the control IgG. Sera from individual patients varied substantially in titre towards particular proteins. For example, patient 17 had high titres for Aap, Bhp and SesI, but low titres for SdrG, SdrF, SesH and SesG. Conversely, patient 18 had high titres for SdrG, SdrF, SesI and SesG compared to the control IgG. Sera from individual patients varied substantially in titre towards particular proteins. For example, patient 17 had high titres for Aap, Bhp and SesI, but low titres for SdrG, SdrF, SesH and SesG. Conversely, patient 18 had high titres for SdrG, SdrF, SesI and SesG, and low titres for Aap and Bhp. The variation in the IgG reactivity for each protein across the patient group can be attributed to the wide range in the patients’ age, the type of infection and the time of serum collection in relation to the infection/disease progress. In spite of the reactivity variations, these data clearly show that the CWA proteins are expressed in infected humans, making them attractive candidates for the development of alternative therapies.
We have initially characterized the CWA proteins from *S. epidermidis*. We have shown that the majority of these mosaic proteins are composed of subdomains rich in β-strands, a common feature present in well-characterized Gram-positive MSCRAMMs. The newly identified proteins are expressed on the bacterial surface, and some are recognized by the sera of infected patients. Although the function of most of these proteins is still not clear, they may contribute to the pathogenic potential of *S. epidermidis*, constituting appropriate candidates for the development of novel therapies against staphylococcal infections.

**ACKNOWLEDGEMENTS**

Preliminary sequence data were obtained from The Institute for Genomic Research through the website at http://www.tigr.org. Sequencing of *Staphylococcus epidermidis* RP62A was accomplished with support from TIGR. This work was supported by National Institutes of Health award AR44415 (to M.H.) and postdoctoral fellowship award AI 10629 (to M.G.B.). This study was also supported by grants from a Wellcome Trust Project (to S. J. P.), Inhibitex Inc. (to M.H.) and Ricerca Finalizzata Ministero della Sanita (to P. S.).

**REFERENCES**


epidermidis: comparison of sequence with those of other staphylococcal lipases. Microbiology 146, 1419–1427.


