Microbiology (1998), **144**, 3387–3395 Printed in Great Britain

Three new members of the serine-aspartate repeat protein multigene family of *Staphylococcus aureus*

Elisabet Josefsson,[1] Kirk W. McCrea,[2] Deirdre Ní Eidhin,[1] David O'Connell,[1] Jos Cox,[3] Magnus Höök[2] and Timothy J. Foster[1]

Author for correspondence: Timothy J. Foster. Tel: +353 1 6082014. Fax: +353 1 6799294. e-mail: tfoster@tcd.ie

[1] Department of Microbiology, Moyne Institute of Preventive Medicine, Trinity College, Dublin 2, Ireland

[2] Institute of Biosciences and Technology, Texas Medical Center, 2121 West Holcombe Boulevard, Houston, TX 77030-3303, USA

[3] Department of Biochemistry, University of Geneva, 30 quai Ernest Ansermet, 1211 Geneva 4, Switzerland

Three new genes encoding the serine-aspartate (SD) repeat-containing proteins SdrC, SdrD and SdrE were found in *Staphylococcus aureus* strain Newman. The SD repeats had earlier been found in the *S. aureus* fibrinogen-binding clumping factors ClfA and ClfB. The *clfA* and *clfB* genes encode high-molecular-mass fibrinogen-binding proteins that are anchored to the cell surface of *S. aureus*. The *sdr* genes now reported are closely linked and tandemly arrayed. The putative Sdr proteins have both organizational and sequence similarity to ClfA and ClfB. At the N-terminus, putative secretory signal sequences precede approximately 500 residue A regions. The A regions of the Sdr and Clf proteins exhibit only 20–30% residue identity when aligned with any other member of the family. The only conserved sequence is the consensus motif TYTFTDYVD. The Sdr proteins differ from ClfA and ClfB by having two to five additional 110–113 residue repeated sequences (B-motifs) located between region A and the R-region. Each B-motif contains a consensus Ca^{2} -binding EF-hand loop normally found in eukaryotic proteins. The structural integrity of recombinant SdrD(B1–B5) protein comprising the five B-repeats of SdrD was shown by bisANS fluorescence analysis to be Ca^{2} -dependent, suggesting that the EF-hands are functional. When Ca^{2} was removed the structure collapsed to an unfolded conformation. The original structure was restored by addition of Ca^{2}. The C-terminal R-domains of the Sdr proteins contain 132–170 SD residues. These are followed by conserved wall-anchoring regions characteristic of many surface proteins of Gram-positive bacteria. The *sdr* locus was present in all 31 *S. aureus* strains from human and bovine sources tested by Southern hybridization, although in a few strains it contained two rather than three genes.

Keywords: surface protein, adhesin, multigene family, EF-hand, calcium-binding protein

INTRODUCTION

Staphylococcus aureus is an important cause of nosocomial infections, particularly those of trauma and surgical patients, and infections associated with in-dwelling medical devices (Rupp, 1997; Kernodle & Kaiser, 1997). The bacterium expresses a panel of adhesins of the MSCRAMM type (Patti *et al*., 1994a) that bind extracellular matrix proteins such as fibronectin (Signas *et al*., 1989), fibrinogen (McDevitt *et al*., 1994), collagen (Patti *et al*., 1992) and elastin (Park *et al*., 1991). Many of these MSCRAMMs have been shown to act as virulence factors in animal models of staphylococcal infections (Foster *et al*., 1997).

S. aureus expresses two fibrinogen-binding MSCRAMMs that are structurally related, called clumping factor A and B (ClfA and ClfB) (McDevitt *et al*., 1994; Ni Eidhin *et al*., 1998). The *clfA* and *clfB* genes are not closely linked and are probably monocistronic.

Abbreviations: bisANS, 4,4′-dianilino-1,1′-binaphthyl-5,5′-disulphonate; SD repeat, serine-aspartate repeat.

The GenBank accession numbers for *sdrC*, *sdrD* and *sdrE* are AJ005645, AJ005646 and AJ005647, respectively.

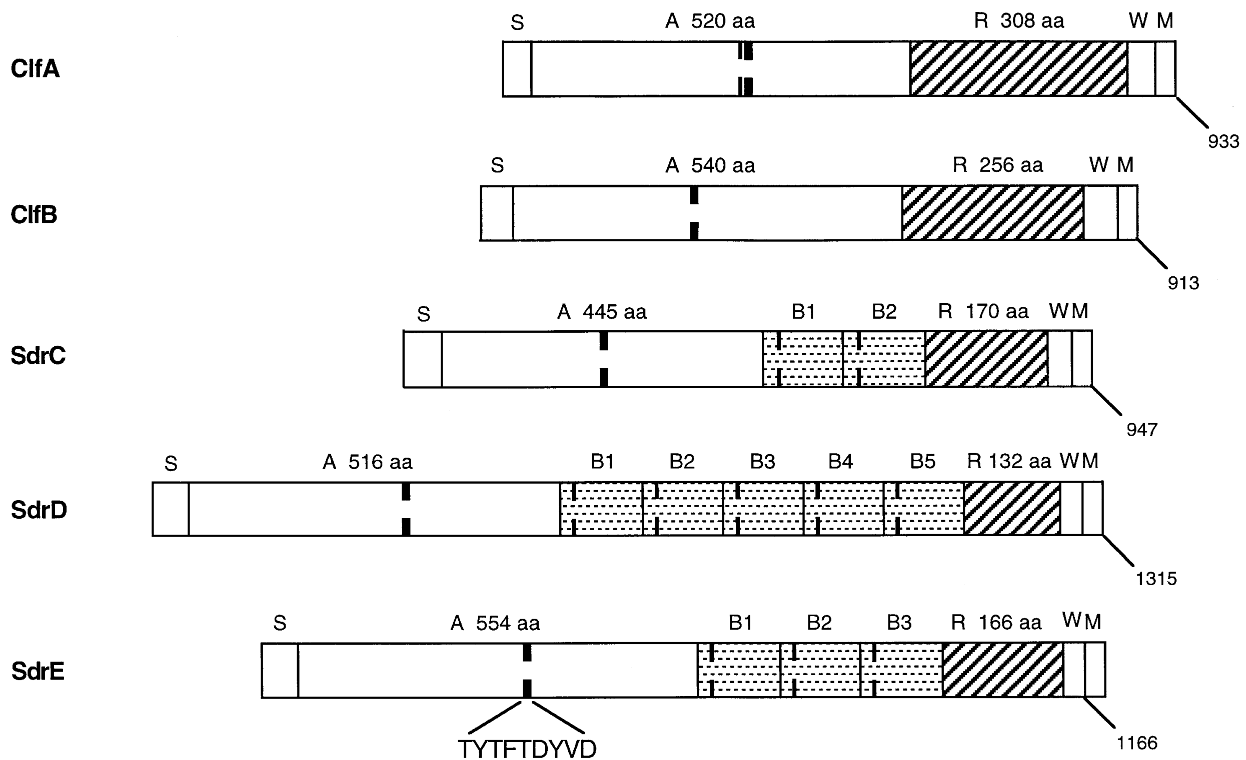

Fig. 1. Structural organization of proteins in the SD-repeat multigene family of *S. aureus* Newman. In region A the thick broken line represents the TYTFTDYVD motif. In the region A of ClfA and in region B the thin broken line represents an EF-hand loop. S, signal sequence; A, putative ligand-binding A region; B, B repeat; R, dipeptide DS repeats; W, short wall-spanning region; M, membrane-spanning segment. The LPXTG motif occurs between domains W and M.

The proteins mediate cell clumping in a solution of fibrinogen and attachment of bacterial cells to immobilized fibrinogen. ClfA and ClfB are virulence factors which mediate bacterial attachment to indwelling medical devices that have become coated with plasma proteins, to blood clots in damaged tissue, and to platelet-fibrin thrombi on damaged heart valves in a rat model of endocarditis (McDevitt *et al.*, 1994; Moreillon *et al.*, 1995; Nı Eidhin *et al.*, 1998; Vaudaux *et al.*, 1995).

The ClfA and ClfB proteins have the same domain organization (McDevitt *et al.*, 1994; Nı Eidhin *et al.*, 1998; Fig. 1), which includes features characteristic of surface proteins on Gram-positive bacteria. A long signal sequence (S) is followed by an approximately 500 residue region (A) encompassing a unique sequence. Region R, which is composed of serine-aspartate (SD) repeating dipeptides, links region A to the cell-wall-anchoring region W. The C-terminal end includes a transmembrane region (M) and a short cytoplasmic tail composed largely of positively charged residues. The ligand-binding A domains in ClfA and ClfB are not closely related, with only 27% identical residues. In contrast, the C-terminal sequences are quite similar. The R-domain in ClfA from strain Newman is composed of 154 repeats of the SD dipeptide (308 residues) and in ClfB it is 128 repeats (256 residues). The length of the R-domains of ClfA and ClfB varies in different strains (McDevitt & Foster, 1995). The dipeptide is encoded by an 18 bp repeat motif GAY TCN GAY TCN GAY AGY. At least 70 residues from region R of ClfA are required for functional expression of the ligand-binding A region on the bacterial cell surface (Hartford *et al.*, 1997). ClfA and ClfB are anchored to the cell wall, via an LPXTG motif present in the W-domains (Schneewind *et al.*, 1995).

The clumping factor ClfA binds to two distinct sites in the-γ chain of fibrinogen that are also recognized by the fibrinogen-binding mammalian integrins, αIIbβ3 and αMβ2 (O'Connell *et al.*, 1998, and unpublished). ClfA and the platelet integrin αIIbβ3 bind to the extreme C-terminus of the fibrinogen γ chain (McDevitt *et al.*, 1997; Kloczewiak *et al.*, 1984), whereas ClfA and the leucocyte intergrin αMβ2 share a binding site located in the central part of the γ chain (Altieri *et al.*, 1993; D. O'Connell and others, unpublished). The binding of ClfA and αIIbβ3 to fibrinogen is regulated by divalent cations, and both proteins contain sequence motifs that could form EF-hands. An EF-hand loop is formed by a stretch of 12 residues that displays six or seven peptide oxygens in the correct geometry for the coordination of divalent cations (da Silva & Reinach, 1991). Also, the binding of ClfA and αMβ2 to the internal site in the γ chain appears to involve the newly discovered cation-binding MIDAS

motif (DXSXS) which is present in both molecules (Lee *et al*., 1995; D. O'Connell and others, unpublished).

Southern hybridization analysis of *S. aureus* Newman genomic DNA with a labelled DNA probe comprising the part of the clumping factor *clfA* gene encoding the region R SD dipeptide revealed three *Hin*dIII fragments in addition to the 6·8 kb fragment carrying *clfA* itself (McDevitt *et al*., 1994). We recently showed that the 2·7 kb *Hin*dIII fragment carries the *clfB* gene, which encodes ClfB (Ní Eidhin *et al*., 1998). Here we set out to analyse the other *clfA*-region R-hybridizing loci to determine if these also encode surface proteins. Three new genes were cloned and sequenced. Each has the potential to encode a surface protein with structural and sequence similarity to ClfA and ClfB. The three genes are closely linked and are distinct from *clfA* and *clfB* by encoding a novel repeated motif (B) located between regions A and R. Each B motif has a cation-binding EF-hand and we show that the structural integrity of the B-motif is cation dependent.

METHODS

***S. aureus* strains.** These are shown in Table 1.

DNA manipulations. These were performed by standard techniques (Ausubel *et al*., 1987; Sambrook *et al*., 1989).

Cloning the *sdr* locus. A library of *S. aureus* Newman genomic DNA made in λGEM-12 was screened by plaque hybridization using a region R specific probe derived from *clfA* (Ní Eidhin *et al*., 1998). One chimeric phage from this library, λA6-2, was used in this study. A 3·8 kb *Hin*dIII fragment was cloned directly from genomic DNA into pBluescript SK(+) (Stratagene), forming pC1.

DNA sequencing. Automated DNA sequencing was performed by primer walking of the 3·9 kb *Hin*dIII fragment clone in pC1 (*sdrC* and the 5′ end of *sdrD*) and pEJ3 (*sdrD*), and with nested deletions generated by Erasa-Base (Promega) of plasmids pEJ1 and pEJ2 (*sdrE*).

Southern hybridization. Genomic DNA isolated by the method of Lindberg *et al*. (1972) was cleaved with *Hin*dIII, fractionated on a 1% agarose gel and transferred to a Nylon N+ membrane (Amersham) by the method of Southern (1975). Filters were hybridized with probes specific for region A of *sdrC*, *sdrD* and *sdrE*, and for region B1–B5 of *sdrD*. Fragments were amplified by PCR using primers shown in Table 2 and cloned into pQE30 (Qiagen). They were excised and labelled by the random primer method incorporating fluorescein (Amersham). Hybridization was recognized with anti-fluorescein–horseradish peroxidase conjugate and Enhanced Chemiluminescence (Amersham).

Expression of recombinant protein. DNA encoding the five B repeats of SdrD was amplified from pEJ3 DNA by PCR using the primers listed in Table 2 and VENT polymerase (New England Biolabs) and cloned into the expression vector pQE30 (Qiagen). The strategy employed was to place unique restriction sites in 5′ extensions to the primers allowing directional cloning into the multiple cloning site of the expression vector to form in-frame fusions with sequences encoding the hexa-histidine tags. In addition codons for glutamine (CAA) and glutamate (GAA) were introduced into the forward primer between the *Bam*HI site and the first codon for repeat B1 (GTA). This was done in order to facilitate removal of the His-tag from the recombinant protein. The recombinant plasmid was transformed into *Escherichia*

Table 1. Bacterial strains

Strain	Properties	Source* or reference
Newman	Archetypal clumping factor	Duthie & Lorenz (1952)
Phillips	Osteomyelitis isolate	Patti *et al*. (1994b)
8325-4	Laboratory strain	Novick (1967)
R19826	Bovine mastitis isolate	A. J. Guidry[1]
PSA10.6, 20-1, 3, 4.2, 5	Bovine mastitis isolates	V. Kapur[2]
AB, RRPL, JG, DA, NJ	Endocarditis isolates	L. Baddour[3]
FB, JB	Osteomyelitis isolates	L. Baddour
SA1048, SA1066, MSA932	Human isolates (disease?)	V. Kapur
SA128, SA240	Dublin MRSA	Coleman *et al*. (1985)
GH13	Guy's Hospital MRSA	Poston & Li Saw Hee (1991)
Cowan 1	Archetypal adherent strain	ATCC 12598, NCTC 8530
Webb	Endocarditis isolate.	Tompkins *et al*. (1990)
P1	Rabbit virulent derivative of ATCC 25923	Sherertz *et al*. (1993)
M60	Bovine mastitis isolate	Anderson (1976)
V8	V8 protease producer	ATCC 49775
Todd555	TSST-1 producer	Clyne *et al*. (1988)
RN4282	TSST-1 producer	Kreiswirth *et al*. (1983)
TC192 (= FRI1169), 413	TSST-1 producer	de Azevedo *et al*. (1985)

* 1, Immunology and Disease Resistance Laboratory, Livestock and Poultry Science Institute, Agricultural Research Service, USDA, Beltsville, MD, USA; 2, Dept of Veterinary Pathobiology, University of Minnesota, St Paul, USA; 3, Dept of Medicine, University of Tennessee Medical Center at Knoxville, USA.

Table 2. Primers

Region to be amplified	F/R*	Primer sequence
sdrC region A	F	CCC.GGA.TCC.GCA.GAA.CAT.ACG.AAT.GGA.G
sdrC region A	R	CCC.AAG.CTT.ACT.TTT.GGT.CGC.CAT.TAG.CAG
sdrD region A	F	CGC.GGA.TCC.CAG.GCA.GAA.AGT.ACT.AAT.AAA.GAA.TTG
sdrD region A	R	CC.GCG.GTC.GAC.TTC.TTG.ACC.AGC.TCC.GCC.ACT.TTG
sdrE region A	F	C.GCG.GGA.TCC.CAG.GCT.GAA.AAC.ACT.AGT.ACA.GAA.AAT.GCA
sdrE region A	R	CGC.AAG.CTT.CTT.TTC.TTC.AGG.TTT.AAC.AGT.ACC
sdrD B1–B5	F	C.GCG.GGA.TCC.CAA.GAA.GTA.TAT.AAA.ATT.GGT.AAC.TAC
sdrD B1–B5	R	CGC.AAG.CTT.TGA.TGT.TTC.TTC.TTC.GTA.GTA.GCC

* F, forward; R, reverse.

Restriction sites in the 5′ extensions of primers are underlined. The double-underlined codons are those that specify additional amino acids at the N-terminus and C-terminus of B1–B5.

coli XL-1 Blue (Stratagene). Expression of the recombinant protein and its purification was as described by O'Connell *et al*. (1998). The His-tag was removed by the TAGZyme system (Unizyme Laboratories), after which the SdrD(B1–B5) protein had additional amino acids GSQE at the N-terminus and SKLN at the C-terminus. The N-terminal sequence of the protein was verified by amino acid sequencing.

bis-ANS fluorimetry. The Ca^{2+}- and Mg^{2+}-dependent changes in the hydrophobic core of the SdrD(B1–B5) protein were followed by monitoring the fluorescence properties of 4,4′-dianilino-1,1′-binaphthyl-5,5′-disulphonate (bisANS) as described by Pan & Johnson (1996). Solutions of 30 μM bisANS and 2 μM metal-free SdrD(B1–B5) protein in 50 mM Tris/HCl, pH 7·5, 150 mM KCl were excited at 370 nm and the emission spectra recorded with slits of 5 nm. EGTA (20 μM), $MgCl_2$ (2 mM) or $CaCl_2$ (20 μM and 2 mM) were added to obtain the metal-free, Mg^{2+} or Ca^{2+} forms, respectively. The estimated accuracy of the method is $\pm 4\%$.

RESULTS

Identification of the *sdr* locus

Recombinant phage λA6-2 was cloned from a library of *S. aureus* Newman genomic DNA constructed in the λ replacement vector λGEM-12 (Nı Eidhin *et al*., 1998). The phage was recognized by plaque hybridization with a probe specific for the region-R-encoding sequence of *clfA*. Mapping, subcloning and sequence analysis revealed that λA6-2 contained two complete ORFs (*sdrD* and *sdrE*) and one partial ORF (*sdrC*) with sequence similarity to *clfA* and *clfB* (Fig. 2).

In addition a 3·9 kb hybridizing *Hin*dIII fragment from *S. aureus* Newman was cloned directly into the plasmid vector pBluescript SK(+), forming pC1. Mapping and sequencing revealed that the 3·9 kb *Hin*dIII fragment overlapped the 'left' end of the λA6-2 clone, providing a complete *orf* (*sdrC*) and the 5′ part of another gene (Fig. 2).

In summary, the locus carries three tandemly linked genes which we have called *sdrC*, *sdrD* and *sdrE* (*S-D r*epeat-containing proteins) with the potential to encode surface proteins related to ClfA and ClfB. Along with *clfA* and *clfB* they form a multigene family (Fig. 1).

DNA sequence of the *sdr* locus

The *sdr* locus comprises three ORFs, *sdrC*, *sdrD* and *sdrE*, of 2841, 3945 and 3498 bp, respectively. The 5′ parts of the *sdr* genes encode the putative signal sequences of the Sdr proteins and are preceded by ribosome-binding sites, GGAG. Extending 5′ from the ribosome-binding site of each ORF are 70 bp that are approximately 54 % identical, preceded by 48 bp conserved sequences (75 % identity between the three sequences). It is possible that the conserved 5′ regions contain transcriptional regulatory signals. DNA sequence similarity was also evident in the 3′ parts of each ORF encoding the SD repeats (region R) and C-terminal wall-anchoring motifs. The sequence similarity extended 40 bp downstream from each ORF and included a putative transcription-termination signal. The *sdrC* and *sdrD* genes are separated by 369 bp while *sdrD* and *sdrE* are separated by 397 bp. The relatively large intergenic regions may suggest that *sdrC*, *sdrD* and *sdrE* are transcribed independently.

Amino acid sequence of the Sdr proteins

The overall organization of the Sdr proteins is very similar to that of the ClfA and ClfB proteins (Fig. 1). Each Sdr protein has a putative signal sequence of 50–52 residues followed by an A region of 445–554 residues. The Sdr proteins differ from ClfA and ClfB by the presence of additional B motifs of 110–113 residues which are present as two, three or five tandem repeats in SdrC, SdrE and SdrD, respectively. The C-terminal domain organization of the SdrC, SdrD and SdrE proteins is very similar to that of ClfA and ClfB, with R domains of 170, 132 and 166 SD residues, respectively, followed by C-terminal cell-wall-anchoring motifs, viz. an LPXTG motif, a hydrophobic membrane-spanning domain and positively charged residues at the extreme C-terminus.

The putative signal sequences of the Sdr proteins are 75–90 % identical. The ClfA and ClfB signal sequences are markedly less related to those of the Sdr proteins (37–46 % identity).

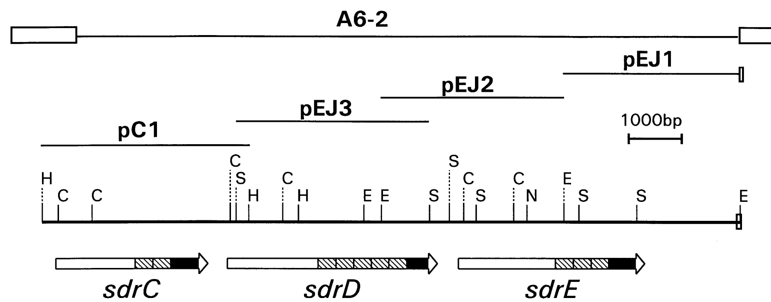

Fig. 2. The *sdr* locus of *S. aureus* strain Newman. Shown above the restriction map is the extent of the locus carried by recombinant plasmids and by the-λGEM clone A6-2. The boxes at the extremity of the λA6-2 clone represent the left and right arms of the vector. The filled boxes within the *sdr* genes indicate the positions of the sequences encoding the SD repeats; the hatched boxes indicate those encoding B repeats. Restriction enzyme cleavage sites are abbreviated as follows: E, *Eco*RI; C, *Cla*I; H, *Hin*dIII; N, *Nco*I; S, *Sca*I.

Fig. 3. Alignment of the B repeats of the Sdr proteins by CLUSTAL W. Three conserved stretches are underlined. Asterisks (*) indicate identical residues; colons (:) indicate functionally conserved residues; dots (.) indicate functionally similar residues. The broken line above the alignment indicates the position of the potentially functional EF-hand I in ClfA.

In the Sdr proteins the SD dipeptide array in the R regions rarely breaks down (two variant residues in SdrC, two residues in SdrD and none in SdrE). In the R regions of ClfA and ClfB there is some drift away from S and D residues, particularly at the end of the regions (McDevitt *et al.*, 1994, Nı Eidhin *et al.*, 1998).

The A regions of SdrC, SdrD and SdrE were aligned with each other and with the A regions of ClfA and ClfB. In any pairwise combination the sequence identity was 21–30% (data not shown). Only one stretch of amino acids is present in all five proteins. The consensus sequence for this motif is T/I Y/F T/V F T D/N Y V D/N, with the first residue of alternatives being present in 4/5 cases except the last D/N, where D is present in 3/5 cases. The motif occurs in the C-terminus of the EF-hand I of ClfA which is implicated in cation-regulated ligand binding (O'Connell *et al.*, 1998). The MIDAS-like motif (DXSXS) in ClfA and ClfB is not apparent in SdrC, SdrD or SdrE.

The B motifs of SdrC, SdrD and SdrE were aligned (Fig. 3) and compared. The B motifs lying adjacent to region R in each protein were the most highly related (95–96% residue identity). The middle B motif of SdrE, the three central B motifs of SdrD and the first B motif of SdrC form a second cluster with 65–85% residue identity. The first B motifs of SdrD and SdrE were quite different from each other (42% identity) and from their neighbours (48 and 52% identity). In each B motif three distinct conserved regions, with 36%, 32% and 20% identity, were noted (Fig. 3). The first contains a

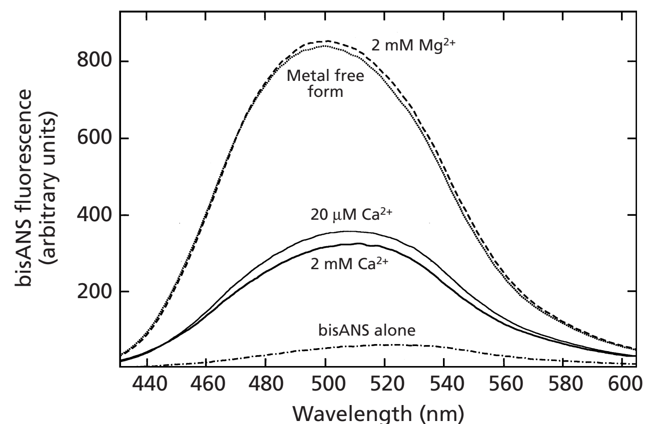

Fig. 4. Hydrophobic exposure as monitored by fluorescence enhancement of bisANS: spectra of 40 μM bisANS in the presence of 2 μM recombinant SdrD(B1–B5) repeat domain in the absence of divalent cations (dotted lines), or in the presence of 20 μM Ca^{2+} (thin solid line), 2 mM Ca^{2+} (thick solid lines) or 2 mM Mg^{2+} (dashed line) as described in Methods. The dashed-dotted line represents the fluoresence of 40 μM bisANS alone.

potential cation-binding EF-hand loop with strong similarity to the consensus (Krestinger, 1987).

The wall-anchoring regions C-terminal to region R are very similar for the Sdr group of proteins (83–97% identical) while similarities to the Clf proteins are much lower (37–44%).

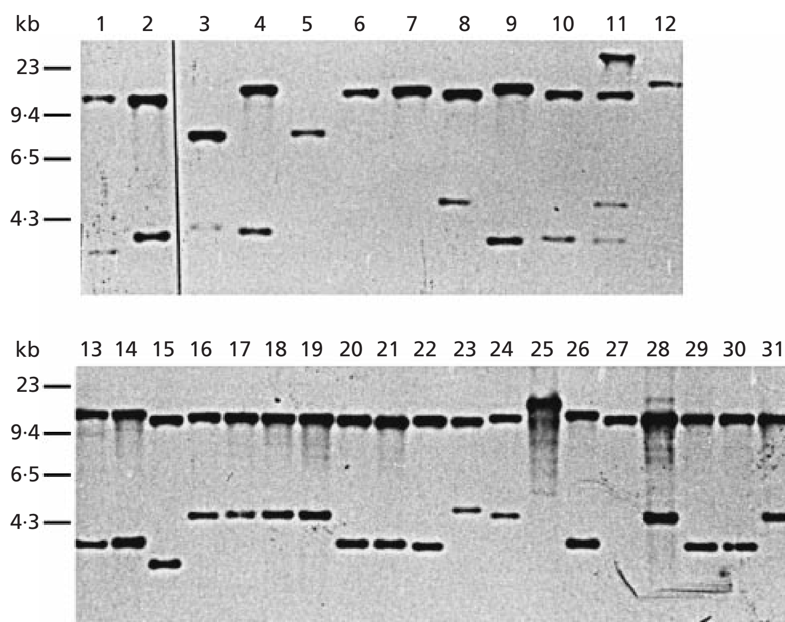

Fig. 5. Southern hybridization analysis. Genomic DNA of *S. aureus* strains was cleaved with *Hin*dIII, fractionated on a 1% agarose gel, transferred to a Nylon membrane and probed with labelled *sdrD* B1–B5 DNA. Lane 1, Phillips; lane 2, Newman; lane 3, 8325-4; lane 4, PSA 10.6; lane 5, PSA 20-1; lane 6, PSA 3; lane 7, PSA 4.2; lane 8, PSA 5; lane 9, SA1048; lane 10, SA1066; lane 11, R19826; lane 12, MSA932; lane 13, AB; lane 14, RRPL; lane 15, JG; lane 16, DA; lane 17, NJ; lane 18, FB; lane 19, JB; lane 20, SA128; lane 21, SA240; lane 22, GH13; lane 23, Cowan 1; lane 24, Webb; lane 25, P1; lane 26, M60; lane 27, V8; lane 28, Todd 555; lane 29, TC192; lane 30, RN4282; lane 31, 413.

The structural integrity of the B motifs is cation dependent

The hydrophobic core of a protein can be monitored with particular fluorescent probes such as bisANS (Vanderheeren & Hanssens, 1994; Gibbons & Horowitz, 1995). Fig. 4 shows that, whereas bisANS alone fluoresced weakly with a λ_{max} of 523 nm, addition of a recombinant protein encompassing the five B motifs of SdrD [SdrD(B1–B5)] enhanced the fluorescence and led to a pronounced blue shift. The metal-free protein shows the highest enhancement, with a λ_{max} of 500 nm, and addition of 2 mM Mg^{2+} did not change this enhancement. Addition of micromolar Ca^{2+} concentrations to the metal-free SdrD(B1–B5) protein led to a 2·5-fold ($\pm$0·1) decrease in fluorescence, with a λ_{max} of 510 nm; addition of millimolar concentrations of Ca^{2+} caused a further small decrease. Together with other conformational studies (not shown) our data indicate that the metal-free protein displays an unfolded conformation with a weak hydrophobic core. Mg^{2+} does not modify this conformation significantly, but binding of Ca^{2+} to high-affinity sites in the protein leads to the formation of a compact conformation with a solvent-shielded hydrophobic core. The effects of the metal-free and Ca^{2+}-saturated protein on bisANS fluorescence are reminiscent of *Nereis* sarcoplasmic Ca^{2+}-binding protein, where the Ca^{2+} form is well structured, whereas the metal-free form has characteristics of an early folding intermediate (Precheur *et al.*, 1996).

The *sdr* genes are widespread in *S. aureus* isolates

In order to determine how widespread *sdrC*, *sdrD* and *sdrE* are in different *S. aureus* strains, a set of 31 strains from diverse human and bovine sources were examined by Southern hybridization using probes specific for the B-motif-encoding sequences of *sdrD* and the region-A-encoding sequences of *sdrC*, *sdrD* and *sdrE*. Initially, *Hin*dIII-cleaved genomic DNA was hybridized with the B motif probe (Fig. 5). In the parental strain, Newman, the A region of *sdrD* contains *Hin*dIII sites that result in two hybridizing fragments, one containing *sdrC* with the 5′ end of *sdrD*, and the other with the 3′ end of *sdrD* and *sdrE* (Fig. 2). As expected, the B-motif-specific probe recognized two *Hin*dIII fragments, of 3·9 kb and approximately 12 kb (Fig. 5, lane 2; McDevitt *et al.*, 1994). The majority of strains tested also had two *Hin*dIII fragments that hybridized with the region B probe (Fig. 5). The smaller *Hin*dIII fragments of these strains usually hybridized with the *sdrC* and *sdrD* region-A-specific probes whereas the larger fragments hybridized to both the *sdrD* and *sdrE* probes (data not shown). Two exceptions were Phillips and 8325-4, which failed to hybridize to the *sdrC* or the *sdrE* region A probes, respectively (data not shown). The region B probe hybridized to a single *Hin*dIII fragment in six strains (PSA20.1, PSA3, PSA4.2, RRPL, P1 and V8). These strains hybridized to all three *sdr* region A probes with the exception of PSA20.1 and PSA4.2, which failed to hybridize with the *sdrE* and *sdrC* region A probes, respectively (data not shown). These results indicate that at least two *sdr* genes are present in a wide range of *S. aureus* isolates from human and animal sources, which implies that the genes are expressed and that the proteins confer properties of importance to the bacteria.

DISCUSSION

A new subfamily of putative cell-wall-associated proteins in *S. aureus* is reported in this paper. The common denominator is the presence of a R region composed largely of SD repeated dipeptides. The *sdrC*, *sdrD* and *sdrE* genes form a subgroup within the family and are characterized by encoding repeated B motifs

which are not present in the ClfA and ClfB proteins. The *sdr* genes seem to be present in a broad spectrum of *S. aureus* isolates according to DNA hybridization data.

The A regions of the five SD repeat proteins are sufficiently related (21–30%) to conclude that they have a common origin, but they are diverse enough to suggest that they bind to different ligands and have different functions. This is already obvious from comparing the A regions of ClfA and ClfB. The A region of ClfA binds the fibrinogen γ-chain while ClfB, which also binds fibrinogen, recognizes sites in the α- and β-chains. It seems likely that the A domains of SdrC, SdrD and SdrE interact with host proteins, but these ligands have not yet been identified.

The A regions of the fibronectin-binding MSCRAMMs FnBPA and FnBPB of *S. aureus* (McDevitt *et al.*, 1994) and the fibronectin-binding MSCRAMM F2 of *Streptococcus pyogenes* (Caparon & Hanski, 1996) have similarity to the A regions of the Clf-Sdr proteins. A common feature of the different A domains in the Clf-Sdr family is the presence of a conserved sequence motif: TYTFTDYVD. In the Sdr and Clf proteins this motif is highly conserved, while a degenerate version occurs in fibronectin-binding MSCRAMMS, as well as the collagen-binding protein Cna (unpublished data). The importance of this motif is currently unknown.

It was recently reported that a *Staphylococcus epidermidis* strain expressed a fibrinogen-binding protein Fbe with a similar domain organization to ClfA (Nilsson *et al.*, 1998). Examination of Fbe reveals an A region typical of the A regions of the Clf-Sdr family described here, except that Fbe has 41% identity with SdrE region A, whereas its similarity to A regions of the other proteins is 24–30%, close to the level of similarity between the A regions of the *S. aureus* proteins. The Fbe A region contains a TYTFTDYVD motif. In addition Fbe has two B repeats and is thus a member of the Sdr protein subfamily reported here.

The repeated B motifs are the common denominator of the subgroup of SD repeat proteins described here. These motifs are found in different numbers in the three Sdr proteins from strain Newman. The B motifs bind Ca^{2+} as suggested by the presence of a well-defined 12 residue cation-binding EF-hand loop. Indeed, the structural integrity of the five B repeats from SdrD expressed as a recombinant protein in *E. coli* is cation dependent. The protein unfolds to a typical molten globule state when Ca^{2+} is removed. Since only micromolar concentrations of Ca^{2+} are enough to induce the folded state, it must be assumed that the cation-binding sites are always occupied and that this domain of the protein is always fully folded when expressed on the bacterial surface. It is unlikely that the structure of the B motif will be regulated by fluctuating Ca^{2+} levels, since Ca^{2+} is present in blood at 1–2 mM, although concentrations can vary more in the extracellular spaces (Brown *et al.*, 1995).

There are clear distinctions between the individual B motifs. The most conserved units are those located adjacent to the R regions (SdrC B2, SdrD B5 and SdrE B3). They differ from the rest at several sites, especially in the C-terminal half. A noteworthy structural detail is that adjacent B repeats are always separated by a proline residue present in the C-terminal region, but a proline never occurs between the last B repeats and the R region. Instead this linker is characterized by a short acidic stretch. These differences suggest that the end units have a different structural or functional role compared to the other B motifs. The N-terminal B motifs of SdrD and SdrE could be said to have drifted apart from the others, since there are numerous amino acid alterations, including small insertions and deletions, whereas the remaining internal B motifs are more highly conserved even when comparing the C-terminal end units with those in the middle. Taking these findings together, there seems to be a purpose for the existence of two different kinds of B motifs. Note that each of the three Sdr proteins has at least one B motif of each kind. Perhaps two units, one of each kind, is the minimum combination needed for function or correct folding.

The B motifs also display some homology to the B repeats of the collagen-binding protein Cna of *S. aureus* (unpublished data). Each Cna B repeat consists of two different but related halves, and one Cna B repeat corresponds to two Sdr B repeats: e.g. SdrD B2B3 has 22% identity to Cna B1 and B2.

Have the repeated B motifs a structural role in surface display or do they bind ligands? Here the number of B units in each surface protein could be of relevance. Either the number of repeated motifs reflects the magnitude of ligand binding activity, or perhaps the number is not crucial, as long as at least two are present in a structural role such as displaying the A regions. It is possible that the B motifs in conjunction with the R regions are necessary for displaying the ligand-binding domain at some distance from the cell surface. The R regions of the Sdr proteins are shorter than those of ClfA and ClfB although even the smallest is of sufficient length to mediate efficient surface expression of ClfA region A, at least in a capsule-negative host (Hartford *et al.*, 1997). Whatever the function of the B domains, it is possible that the number could vary in different *S. aureus* strains due to unequal crossing over.

The higher level of sequence identity in the regions encoding the signal peptides and the membrane–wall-spanning regions when comparing the Sdr proteins with each other rather than with ClfA or ClfB suggests a more recent amplification of a primordial *sdr* gene. On the other hand the A regions of all five proteins are much less conserved. The A regions of the Sdr proteins are no more similar to each other than they are to ClfA or ClfB. This could be a result of selective pressure to avoid immune responses and/or to acquire new ligand-binding activities. In contrast, the greater conservation of the B motifs and the R region may reflect more limited exposure to antibodies or the functional importance of specific sequences. This type of structural diversity is reminiscent of the M proteins of *Streptococcus*

pyogenes, where the signal sequences are very similar, the most N-terminal residues of the mature proteins the most divergent, and the C-terminal sequences are more closely related (Kehoe, 1994).

Based on the genetic analysis reported in this paper a new subfamily of *S. aureus* surface proteins with a potential to function as adhesins is described. Future work will aim to determine the conditions under which the Sdr proteins are expressed, the ligands to which they bind, and the role of the B repeats.

ACKNOWLEDGEMENTS

E. J. and K. W. McC. made equal contributions to this paper.

The authors wish to acknowledge the following for support: BioResearch Ireland, Forbairt and The Wellcome Trust (T. J. F., D. NíE. and D. O'C.); King Gustav V 80th Birthday Foundation, the Swedish Medical Research Council, the Swedish Institute and the Adlerbertska Foundation (E. J.); The Swiss National Science Foundation (J. C.); NIH grant AI 20624, the Markey Foundation and the Neva and Wesley West Foundation (M. H.); and Inhibitex Inc. (T. J. F. and M. H.). They also wish to acknowledge helpful advice from B. Kreikemeyer.

REFERENCES

Altieri, D. C., Plescia, J. & Plow, E. F. (1993). The structural motif glycine 190-valine 202 of the fibrinogen γ-chain interacts with CD11b/CD18 integrin ($\alpha_M\beta_2$, Mac-1) and promotes leukocyte adhesion. *J Biol Chem* **268**, 1847–1853.

Anderson, J. C. (1976). The contribution of the mouse mastitis model to our understanding of staphylococcal infection. *Zentbl Bakteriol Parasitenkd Infektionskr Hyg Abt 1 Orig Reihe A* **5** (**Suppl.**), 783–790.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G. & Struhl, K. (1987). *Current Protocols in Molecular Biology*. New York: Wiley.

de Azevedo, J. C. S., Foster, T. J., Hartigan, P. J., Arbuthnott, J. P., O'Reilly, M., Kreiswirth, B. N. & Novick, R. P. (1985). Expression of the cloned toxic shock syndrome toxin 1 gene (*tst*) in vivo with a rabbit uterine model. *Infect Immun* **50**, 304–309.

Brown, E. M., Vassilev, P. M. & Hebert, S. C. (1995). Calcium ions as extracellular messengers. *Cell* **83**, 679–682.

Caparon, M. G. & Hanski, E. (1996). Protein F2, a novel fibronectin-binding protein from *Streptococcus pyogenes* possesses two binding domains. *Mol Microbiol* **21**, 373–384.

Clyne, M., de Azevedo, J., Carlson, E. & Arbuthnott, J. P. (1988). Production of gamma-hemolysin and lack of production of alpha-hemolysin by strains of *Staphylococcus aureus* associated with toxin shock syndrome. *J Clin Microbiol* **26**, 535–539.

Coleman, D. C., Pomeroy, H., Estridge, J. K., Keane, C. T., Cafferkey, M. T., Hone, R. & Foster, T. J. (1985). Susceptibility to antimicrobial agents and analysis of plasmids in gentamicin- and methicillin-resistant *Staphylococcus aureus* from Dublin hospitals. *J Med Microbiol* **20**, 157–167.

Duthie, E. S. & Lorenz, L. L. (1952). Staphylococcal coagulase: mode of action and antigenicity. *J Gen Microbiol* **6**, 95–107.

Foster, T. J., Hartford, O. & O'Connell, D. (1997). Host–pathogen protein–protein interactions in *Staphylococcus*. In *Molecular Aspects of Host–Pathogen Interactions* (Society for General Microbiology Symposium no. 55), pp. 67–94. Edited by M. A. McCrea, J. R. Saunders, C. J. Smyth & N. D. Stow. Cambridge: Cambridge University Press.

Gibbons, D. L. & Horowitz, P. M. (1995). Exposure of hydrophobic surfaces on the chaperonin GroEL oligomer by protonation or modification of His-401. *J Biol Chem* **270**, 7335–7340.

Hartford, O., François, P., Vaudaux, P. & Foster, T. J. (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.

Kehoe, M. A. (1994). Cell-wall-associated proteins in Gram-positive bacteria. In *Bacterial Cell Wall*, pp. 217–261. Edited by J.-M. Ghuysen & R. Hakenbeck. Amsterdam: Elsevier Science.

Kernodle, D. S. & Kaiser, A. B. (1997). Wound infections and surgical prophylaxis. In *The Staphylococci in Human Disease*, pp. 355–377. Edited by K. B. Crossley & G. L. Archer. New York: Churchill Livingstone.

Kloczewiak, M., Timmons, S., Lukas, T. J. & Hawiger, J. (1984). Platelet receptor recognition site on human fibrinogen. Synthesis and structure–function relationship of peptides corresponding to the carboxy-terminal segment of the g chain. *Biochemistry* **23**, 1767–1774.

Kreiswirth, B. N., Löfdahl, M. S., Betley, M. J., O'Reilly, M., Schlievert, P. M., Bergdoll, M. S. & Novick, R. P. (1983). The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. *Nature* **305**, 709–712.

Krestinger, R. H. (1987). Calcium coordination and the calmodulin fold. Divergent versus convergent evolution. *Cold Spring Harbor Symp Quant Biol* **52**, 499–510.

Lee, J. O., Rieu, P., Arnaout, M. A. & Liddington, R. (1995). Crystal structure of the A-domain from the a subunit of integrin CR3 (CDIIb/CD18). *Cell* **80**, 631–638.

Lindberg, M., Sjöstrom, J. E. & Johansson, T. (1972). Transformation of chromosomal and plasmid characters in *Staphylococcus aureus*. *J Bacteriol* **109**, 844–847.

McDevitt, D. & Foster, T. J. (1995). Variation in the size of the repeat region of the fibrinogen receptor (clumping factor) of *Staphylococcus aureus* strains. *Microbiology* **141**, 937–943.

McDevitt, D., François, P., Vaudaux, P. & Foster, T. J. (1994). Molecular characterization of the clumping factor (fibrinogen receptor) of *Staphylococcus aureus*. *Mol Microbiol* **11**, 237–248.

McDevitt, D., Nanavaty, T., House-Pompeo, K., Bell, E., Turner, N., McIntire, L., Foster, T. J. & Höök, M. (1997). Characterization of the interaction between the *Staphylococcus aureus* clumping factor (ClfA) and fibrinogen. *Eur J Biochem* **247**, 416–424.

Moreillon, P., Entenza, J. M., Francioli, P., McDevitt, D., Foster, T. J., François, P. & Vaudaux, P. (1995). Role of *Staphylococcus aureus* coagulase and clumping factor in pathogenesis of experimental endocarditis. *Infect Immun* **63**, 4738–4743.

Ní Eidhin, D., Perkins, S., François, P., Vaudaux, P., Höök, M. & Foster, T. J. (1998). Clumping factor B (ClfB), a new surface-located fibrinogen-binding adhesin of *Staphylococcus aureus*. *Mol Microbiol* **29**, 245–257.

Nilsson, M., Frykberg, L., Flock, J.-I., Pei, L., Lindberg, M. & Guss, B. (1998). A fibrinogen-binding protein of *Staphylococcus epidermidis*. *Infect Immun* **66**, 2666–2673.

Novick, R. P. (1967). Properties of a cryptic high frequency transducing phage in *Staphylococcus aureus*. *Virology* **33**, 155–166.

O'Connell, D. P., Nanavaty, T., McDevitt, D., Gurusiddappa, S., Höök, M. & Foster, T. J. (1998). The fibrinogen-binding MSCRAMM (clumping factor) of *Staphylococcus aureus* has an

integrin-like Ca^{2+}-dependent inhibitory site. *J Biol Chem* **273**, 6821–6829.

Pan, B.-S. & Johnson, R. G. (1996). Interaction of cardiotonic thiadiazinone derivatives with cardiac troponin C. *J Biol Chem* **271**, 817–823.

Park, P. W., Roberts, D. D., Grosso, L. E., Parks, W. C., Rosenbloom, J., Abrams, W. R. & Mecham, R. P. (1991). Binding of elastin to *Staphylococcus aureus*. *J Biol Chem* **266**, 23399–23406.

Patti, J. M., Jönsson, H., Guss, B., Switalski, L. M., Wiberg, K., Lindberg, M. & Höök, M. (1992). Molecular characterization and expression of a gene encoding *Staphylococcus aureus* collagen adhesin. *J Biol Chem* **267**, 4766–4772.

Patti, J. M., Allen, B. A., McGavin, M. J. & Höök, M. (1994a). MSCRAMM-mediated adherence of microorganisms to host tissues. *Annu Rev Microbiol* **45**, 585–617.

Patti, J. M., Bremell, T., Krajewska-Pietrasik, D., Abdelnour, A., Tarkowski, A., Rydén, C. & Höök, M. (1994b). The *Staphylococcus aureus* collagen adhesin is a virulence determinant in experimental septic arthritis. *Infect Immun* **62**, 152–161.

Poston, S. M. & Li Saw Hee, F. L. (1991). Genetic characterisation of resistance to metal ions in methicillin-resistant *Staphylococcus aureus*: elimination of resistance to cadmium, mercury and tetracycline with loss of methicillin resistance. *J Med Microbiol* **34**, 193–201.

Prêcheur, B., Cox, J. A., Petrova, T., Mispelter, J & Craescu, C. T. (1996). Nereis sarcoplasma Ca^{2+}-binding protein has a highly unstructured apo state which is switched to the native state upon binding of the first Ca^{2+} ion. *FEBS Lett* **395**, 89–94.

Rupp, M. E. (1997). Infections of intravascular catheters and vascular devices. In *The Staphylococci in Human Disease*, pp. 379–399. Edited by K. B. Crossley & G. L. Archer. New York: Churchill Livingstone.

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

Schneewind, O., Fowler, A. & Faull, K. F. (1995). Structure of the cell wall anchor of surface proteins in *Staphylococcus aureus*. *Science* **268**, 103–106.

Sherertz, R. J., Carruth, W. A., Hampton, A. A., Byron, M. P. & Solomon, D. D. (1993). Efficacy of antibiotic-coated catheters in preventing subcutaneous *Staphylococcus aureus* infection in rabbits. *J Infect Dis* **167**, 98–106.

Signäs, C., Raucci, G., Jönsson, K., Lindgren, P. E., Anantharamaiah, G. M., Höök, M. & Lindberg, M. (1989). Nucleotide sequence of the gene for a fibronectin-binding protein from *Staphylococcus aureus*: use of this peptide sequence in the synthesis of biologically active peptides. *Proc Natl Acad Sci USA* **86**, 699–703.

da Silva, A. C. R. & Reinach, F. C. (1991). Calcium binding induces conformational changes in muscle regulatory proteins. *Trends Biochem Sci* **16**, 53–57.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**, 503–517.

Tompkins, D. C., Hatcher, V. B., Patel, D., Orr, G. A., Higgins, L. L. & Lowy, F. D. (1990). A human endothelial cell membrane protein that binds *Staphylococcus aureus* in vitro. *J Clin Invest* **85**, 1248–1254.

Vanderheeren, G. & Hanssens, I. (1994). Thermal unfolding of bovine alpha-lactalbumin. *J Biol Chem* **269**, 7090–7094.

Vaudaux, P. E., François, P., Proctor, R. A., McDevitt, D., Foster, T. J., Albrecht, R. M., Lew, D. P., Wabers, H. & Cooper, S. L. (1995). Use of adhesion-defective mutants of *Staphylococcus aureus* to define the role of specific plasma proteins in promoting bacterial adhesion to canine arteriovenous shunts. *Infect Immun* **63**, 585–590.

Received 19 May 1998; revised 3 September 1998; accepted 14 September 1998.