A short sequence within subdomain N1 of region A of the *Staphylococcus aureus* MSCRAMM clumping factor A is required for export and surface display.

Niamh McCormack, Timothy J. Foster* and Joan A. Geoghegan**

Department of Microbiology, Moyne Institute of Preventive Medicine, School of Genetics and Microbiology, Trinity College Dublin, Dublin 2, Ireland.

Running Title: Export and surface display of ClfA.

Contents category: Cell and Molecular Biology of Microbes

Key words: *Staphylococcus aureus*, clumping factor A, MSCRAMM, protein secretion, fibronectin binding protein B.

Word counts: Summary 248, main text 5108.

*These authors contributed equally

*Corresponding author: E-mail: geoghegj@tcd.ie. Telephone: +35318961188.

Fax: +35316799294
Summary

Clumping factor A (ClfA) is the archetypal fibrinogen-binding surface protein of *Staphylococcus aureus* and a member of the Microbial Surface Component Recognising Adhesive Matrix Molecules (MSCRAMM) family. An N-terminal signal sequence directs export of the MSCRAMM by the Sec pathway and the C-terminal cell wall-anchoring domain allows covalent attachment of ClfA to peptidoglycan by sortase. Region A of ClfA comprises three independently folded subdomains N1, N2 and N3. Subdomains N2N3 comprise IgG-like folds and promote fibrinogen binding. Nothing is known about the structure or function of subdomain N1. Here we demonstrate an unexpected role for N1 in the export and surface localization of ClfA. Attempted expression of a ClfA variant lacking subdomain N1 resulted in impaired growth of *S. aureus* and accumulation of ClfA protein in the cytoplasm and cytoplasmic membrane. The presence of residues 211-228 of N1 was required to allow display of ClfA on the bacterial surface. The importance of this region was confirmed when a ClfA variant lacking residues 211-220 was also mislocalized to the cytoplasm and cytoplasmic membrane. However these residues were not required for export of ClfA lacking the Ser-Asp repeats that links region A to the wall-anchoring domain. Similarly, subdomain N1 of a related MSCRAMM fibronectin binding protein B was required for export and surface display of the full length protein but not a derivative lacking fibronectin binding repeats. In summary we demonstrate that residues in the N1 subdomain are required for export and cell wall localization of *S. aureus* MSCRAMMs.
**Introduction**

*Staphylococcus aureus* is a commensal bacterium that colonises the anterior nares of about 20% of the population (van Belkum *et al*., 2009). The organism is an important opportunistic pathogen that can cause both superficial skin lesions and more serious invasive infections such as endocarditis and sepsis (Lowy, 1998).

The ability of *S. aureus* to colonise the host and cause infection is due in part to the expression of secreted virulence factors and cell wall-anchored surface proteins. The microbial surface component recognizing adhesive matrix molecules (MSCRAMM) family comprises proteins with related N-terminal ligand-binding A domains (Foster *et al*., 2013). These contain two tandemly arrayed subdomains, N2 and N3, that are composed of IgG-like folds (Deivanayagam *et al*., 2002) and the N-terminal N1 subdomain which, in the case of the related protein ClfB, is elongated and is not compact (Perkins *et al*., 2001). The A domains are linked to the cell wall by an extended flexible region comprising repeats of the dipeptide Ser-Asp (ClfA, ClfB, SdrC, SdrD, SdrE, Bbp) or 10/11 fibronectin binding domains (FnBPA, FnBPB) (Hartford *et al*., 1997; Schwarz-Linek *et al*., 2003). MSCRAMMs are also found in coagulase-negative staphylococci and *S. pseudintermedius* (Bannoehr *et al*., 2011; Foster *et al*., 2013).

MSCRAMMs play a pivotal role during colonization and establishment of *S. aureus* infection yet very little is known about how these proteins are directed to the secretion apparatus (Schneewind & Missiakas, 2012; Sibbald *et al*., 2006). Members of the MSCRAMM family possess an N-terminal signal sequence that is predicted to mediate their translocation across the cytoplasmic membrane by the general secretory (Sec) pathway (DeDent *et al*., 2008; Sibbald *et al*., 2006). The Sec system in *S. aureus* consists of homologues of the canonical SecYEG translocation channel and the SecA ATPase (Sibbald *et al*., 2006). A ‘YSIRK/GS’ motif present in the signal sequence directs secretion of the
pre-proteins to sites of peptidoglycan biosynthesis at the region of septum formation (DeDent et al., 2008). However, the mechanisms involved in trafficking MSCRAMMs to this cellular compartment are yet to be elucidated (DeDent et al., 2008; Schneewind & Missiakas, 2012).

Surface display of S. aureus YSIRK/GS-motif containing surface proteins is reduced in mutants lacking the spdA, spdB or spdC genes but the molecular basis of this is not understood (Frankel et al., 2010).

A sorting signal is located at the C-terminus of the protein. It comprises an LPXTG motif, a hydrophobic membrane-spanning domain and at the extreme C-terminus, a stretch of positively charged residues. The last two elements delay secretion across the membrane and facilitate recognition and cleavage of the LPXTG sequence by sortase A, a protein anchored to the cytoplasmic membrane (Schneewind et al., 1993). Sortase cleaves between threonine and glycine forming an acyl-enzyme intermediate which allows transfer of the MSCRAMM to the pentaglycine crossbridge of lipid II. Following transglycosylation the MSCRAMM becomes covalently anchored to peptidoglycan (Ton-That et al., 2000).

Clumping factor A is an important MSCRAMM that promotes virulence during invasive infection (Josefsson et al., 2001; McAdow et al., 2011; Moreillon et al., 1995). This involves both coating the bacterium with plasma fibrinogen and capturing and activating the complement regulator factor I which in turn results in cleavage of the complement opsonin C3b (Hair et al., 2010; Higgins et al., 2006). ClfA is also an adhesin which promotes bacterial attachment to immobilized fibrinogen and to platelet-fibrin thrombi on heart valves during the initiation of endocarditis (Moreillon et al., 1995; Que et al., 2001).

The A domain of ClfA binds to a peptide sequence comprising the extreme C-terminal residues of the γ-chain of fibrinogen (McDevitt et al., 1997). The X-ray crystal structure of ClfA N2N3 has been solved both as an apo-protein and with the fibrinogen
peptide ligand bound (Ganesh et al., 2008). The fibrinogen γ-chain peptide binds in a
hydrophobic trench formed between the separately folded N2 and N3 subdomains by the
‘dock, lock, and latch’ mechanism (Ponnuraj et al., 2003). Docking of the peptide in the
binding trench induces a conformational change in the flexible extension of the C-terminus of
the N3 subdomain which is reoriented to cover the ligand peptide and “locks” it in place.
Latching is completed by formation of an extra β-strand which is complementary to a β-sheet
of the N2 domain. The MSCRAMMs ClfB, Bbp, FnBPA and FnBPB bind to fibrinogen by a
similar mechanism (Burke et al., 2011; Ganesh et al., 2011; Keane et al., 2007; Vazquez et
al., 2011). In all cases the ligand binding site appears to be confined to subdomains N2N3
(Burke et al., 2011; Ganesh et al., 2011; Keane et al., 2007; Vazquez et al., 2011).

The boundaries of the N1, N2 and N3 subdomains were originally defined for ClfB
before structures of any of the proteins were solved. The S. aureus zinc-metalloprotease
aureolysin was found to cleave ClfB at a ‘SLAVA’ motif and the cleavage site was
designated as the boundary between N1 and N2 (McAleese et al., 2001). Virtually nothing is
known about the structure or function of subdomain N1 of the MSCRAMM family of surface
proteins. The objective of this study was to determine if subdomain N1 of ClfA plays a role
in export of the protein and its subsequent elaboration on the cell surface.
Methods

Bacterial culture conditions and reagents

Strains are listed in Table 1. *S. aureus* was grown on trypticase soy agar or broth (TSA, TSB, Oxoid) with shaking (200 rpm) at 37 °C. For expression studies stationary phase cultures were diluted 1:100 in brain heart infusion broth (BHI, Difco), grown to an OD₆₀₀ of 0.5 and induced with anhydrotetracycline (ATc) for 2 h at 37 °C. *E. coli* strains were cultured on Luria agar or broth (Difco) at 37 °C. Antibiotics were incorporated into media where appropriate at the following concentrations: 100 µg ampicillin (Ap) ml⁻¹, 10 µg chloramphenicol (Cm) ml⁻¹, 10 µg erythromycin (Em) ml⁻¹, 50 µg kanamycin (Ka) ml⁻¹, 100 – 1200 ng anhydrotetracycline (ATc) ml⁻¹. Unless otherwise stated all reagents were obtained from Sigma.

Plasmid and strain construction

Bacterial strains and plasmids used in this study are listed in Table 1. Strain NM1 was constructed by transduction of *spa::Ka*R into strain Newman *clfA5 clfB::EmR*. Deletion of *clfB::EmR* from strain NM1 to yield NM2 was achieved by allelic replacement using pIMAYΔclfB (Table 1) as previously described (Monk *et al.*, 2012). Strain BH1CC ΔfnbAfnbB *spa::Ka*R was constructed by transduction of *spa::Ka*R into strain BH1CC ΔfnbAfnbB. The complete *clfA* gene from strain Newman was amplified from genomic DNA by PCR (primers study are listed in Table S1) and cloned between EcoRI and BglII sites of pRMC2 or between the EcoRI and KpnI sites of pALC2073 to generate pRMC2clfA and pALC2073clfA, respectively. The *fnbB* gene from strain 8325 was amplified from genomic DNA by PCR and cloned between the EcoRI and BglIII sites of pRMC2 to generate pRMC2fnbB. Inverse PCR was used to generate derivatives of pRMC2clfA, pALC2073clfA and pRMC2fnbB. Alanine substitutions and FLAG epitopes were incorporated by primer
extension during inverse PCR. PCR products were treated with DpnI to eliminate parental template DNA and following blunt-end ligation plasmids were transformed into \textit{E. coli} DC10B. Plasmids were purified from DC10B and transformed into \textit{S. aureus} made electrocompetent as previously described (Lofblom \textit{et al.}, 2007; Monk \textit{et al.}, 2012). Plasmid DNA (5 µg) was introduced into \textit{S. aureus} by electroporation and transformants were recovered by plating on TSA containing Cm.

\textbf{Fractionation of} \textit{S. aureus} \textbf{and Western immunoblotting.}\n
\textit{S. aureus} cultures were grown as described above, washed in phosphate buffered saline (PBS) and adjusted to an OD$_{600}$ of 10. Cell wall-associated proteins were solubilised by resuspending bacteria in 250 µl digestion buffer (50 mM Tris-HCl, 20 mM MgCl$_2$, 30 % (w/v) raffinose, pHi 7.5) containing protease inhibitors (Roche) and lysostaphin (AMBI Products LLC, 200 µg ml$^{-1}$) at 37 °C for 10 min. Protoplasts were removed by centrifugation at 12,000 x g for 10 min and the supernatant containing solubilised cell wall proteins was transferred to a clean tube and boiled for 5 min in Laemmli sample buffer. Protoplast pellets were washed with digestion buffer and resuspended in ice-cold Tris-HCl (50 mM, pH 7.5) containing protease inhibitors (Roche) and DNase (80 µg ml$^{-1}$). Protoplasts were lysed on ice by vortexing. Complete lysis was confirmed by phase contrast microscopy. Lysed cell fractions were separated by centrifugation at 44,500 x g for 1 h at 4 °C. The supernatant was retained as the cytoplasmic fraction. The pellet was washed with ice-cold Tris-HCl (50 mM, pH 7.5) and resuspended in Tris-HCl (50 mM, pH 7.5). Fractions were separated by 7.5 % sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, transferred onto PVDF and blocked in 10 % (w/v) skimmed milk proteins. Blots were probed with ClfA mAb, a murine monoclonal antibody raised against ClfA subdomains N2N3 (a kind gift from Inhibitex Inc, 1:2000) and bound antibody was detected using horseradish peroxidise (HRP)-
conjugated rabbit anti-mouse IgG (1:3000, Dako). FnBPB blots were probed with biotinylated fibronectin and bound biotin was detected using horseradish peroxidise (HRP)-conjugated streptavadin. FLAG-tagged constructs were probed with murine monoclonal anti-FLAG IgG (1:2000, Pierce) and detected with horseradish peroxidise (HRP)-conjugated rabbit anti-mouse IgG (1:2000, Dako) Reactive bands were visualised using the LumiGLO reagent and peroxide detection system (Cell Signalling Technology). Antibodies were removed from blots by incubation at 70 °C for 30 min in stripping buffer (2 % SDS, 100 mM β-mercaptoethanol and 50 mM Tris). Stripped blots were washed twice in PBS, blocked in 10 % (w/v) skimmed milk proteins and then probed with polyclonal rabbit anti-SdrE IgG (1:2000, O'Brien et al., 2002) or rabbit anti-Micrococcus luteus F_{1}/F_{0} ATPase serum (Downer et al., 2002). Bound antibody was detected using HRP-conjugated goat anti-rabbit IgG (1:10,000, Dako) or protein A peroxidase (1:500).

**Dot immunoblotting**

Bacteria were washed twice in PBS and adjusted to an OD_{600} of 1.0. Doubling dilutions were spotted onto a nitrocellulose membrane (Protran) membrane and allowed to dry. For whole cell lysates bacteria were incubated in PBS containing protease inhibitors (Roche) and lysostaphin (AMBI Products LLC, 200 μg ml⁻¹) at 37°C for 15 mins. Lysates were incubated with DNase (80 μg/ml) for 10 min at room temperature and heated to 85°C for 5 min. Cellular debris was sedimented by centrifugation at 12,000 x g. Lysates were spotted onto a nitrocellulose membrane (Protran) membrane and allowed to dry. Membranes were blocked for 1 h in 10% (w/v) skimmed milk proteins and immunoblotting was performed as above.
Competitive growth assay

For competitive growth experiments stationary phase cultures were adjusted to an OD$_{600}$ of 0.025 in BHI. Equal volumes of two *S. aureus* strains were mixed and grown in BHI incorporating Cm to an OD$_{600}$ of 0.5. Where appropriate ATc (1200 ng ml$^{-1}$) was added and cultures were incubated at 37 °C for a further 2 h before being plated on TSA incorporating Cm. Colonies which were resistant to Cm were replica plated onto TSA containing Em and the percentage of Em resistant bacteria was calculated.

Statistical analysis

Statistical analysis was performed using Prism Graphpad 5 software. P values were calculated using Students t-test.
Results

Expression of ClfA variants lacking subdomain N1.

The subdomain boundaries of the A region of ClfA were originally defined based on amino-acid sequence similarity with ClfB where the position of a cleavage site (SLAVA) for the *S. aureus* metalloprotease aureolysin defined the junction between subdomains N1 and N2 (Perkins *et al.*, 2001). ClfA contains a similar sequence (SLAAVA). Based on this motif the amino-acid coordinates of the N2N3 subdomains of ClfA were originally defined as residues 221-559 (McDevitt *et al.*, 1995; O'Connell *et al.*, 1998). More recently the N2N3 subdomains of ClfA have been crystallized (Deivanayagam *et al.*, 2002; Ganesh *et al.*, 2008) allowing the precise definition of boundaries of the N2N3 subdomains as comprising residues 229-545 (Fig. 1).

This study set out to determine if subdomain N1 of ClfA plays a role in protein secretion and elaboration on the cell surface. To facilitate this, the clfA gene of strain Newman was cloned into and expressed from the anhydrotetracycline (ATc)-inducible vector pRMC2 (Corrigan & Foster, 2009) to yield pRMC2clfA. Plasmid pRMC2clfA was manipulated to create derivatives lacking DNA encoding the N1 subdomain of ClfA (residues 40-220 or residues 40-228) yielding plasmids pRMC2clfAΔN140-220 and pRMC2clfAΔN140-228, respectively (Fig. 1). Two pRMC2clfAΔN1 constructs were generated to reflect the different definitions of the boundary of subdomain N1. ClfAΔN140-220 represents a deletion up to the SLAAVA motif while ClfAΔN140-228 removes the entire N1 sequence defined by the X-ray crystal structure of N2N3 (Fig. 1). Plasmids were transformed into strain NM1, a ClfA-, ClfB-, Spa-deficient derivative of *S. aureus* Newman (Table 1).

Whole cell dot immunoblotting was performed to measure the amount of ClfA on the bacterial surface. Serial dilutions of bacteria were applied to a nitrocellulose membrane and a
monoclonal antibody that recognizes an epitope in subdomains N2N3 of ClfA (ClfA mAb) was used to detect the protein. Expression of ClfA by NM1 (pRMC2clfA) could not be detected in the absence of inducer (Fig. 2a) in agreement with previous studies showing that gene expression from pRMC2 is tightly repressed (Corrigan & Foster, 2009). Strain NM1 carrying the pRMC2 empty vector was used as a negative control. Addition of increasing concentrations of ATc resulted in increasing amounts of ClfA on the surface of *S. aureus* (Fig. 2a). A high level of surface expression similar to that achieved by the single chromosomal gene in strain Newman was achieved at inducer concentrations of 800 ng ml\(^{-1}\) and 1.2 µg ml\(^{-1}\). In contrast neither ClfAΔN1\(_{40-220}\) nor ClfAΔN1\(_{40-228}\) could be detected on the surface of *S. aureus* even at the highest concentration of inducer tested (Fig. 2a). All proteins could be detected when whole cell lysates were probed with ClfA mAb (Fig. 2b). Thus the failure to detect ClfAΔN1\(_{40-220}\) and ClfAΔN1\(_{40-228}\) on the surface of *S. aureus* suggested that N1 might have a role in protein export. As a control, the same bacteria were spotted onto a nitrocellulose membrane and probed for expression of SdrE, a member of the MSCRAMM family which is exposed on the surface of *S. aureus*. All bacteria were recognized equally by antibodies directed against SdrE (Fig. S1). This demonstrates that similar amounts of bacteria were immobilised on the nitrocellulose membrane.

Variants of pRMC2 clfA were constructed which retained differing amounts of N1 in an attempt to determine the minimum sequence that supported surface display. A mutant lacking residues 40-210 but retaining C-terminal residues of N1 (ClfAΔN1\(_{40-210}\)) was displayed on the surface of *S. aureus*, albeit at a slightly lower level than full-length ClfA expressed from pRMC2clfA (Fig. 2a). However when cultures of NM1 (pRMC2clfA) and NM1 (pRMC2clfAΔN1\(_{40-210}\)) were induced with 800 ng and 1.2 µg ATc ml\(^{-1}\), respectively, the amount of ClfA and ClfAΔN1\(_{40-210}\) on the surface of *S. aureus* was found to be the same. These inducer concentrations were used for all further experiments.
Subdomain N1 is required for ClfA to be localized to the *S. aureus* cell wall.

Whole cell immunoblotting indicated that ClfAΔN1\(^{40-220}\) was not elaborated on the surface of *S. aureus* following induction with ATc (Fig. 2a) but ClfAΔN1\(^{40-220}\) could be detected in whole cell lysates (Fig. 2b). Therefore the subcellular location of ClfAΔN1\(^{40-220}\) following induction was investigated. Fractions corresponding to the cytoplasm, the cytoplasmic membrane and proteins solubilised from the cell wall by lysostaphin treatment during protoplast formation (cell wall fraction) were analyzed by Western immunoblotting probing with the ClfA mAb. In the wild type ClfA sample a band of 150 kDa was detected only in the cell wall fraction. This is consistent with normal sorting. In contrast, ClfAΔN1\(^{40-220}\) was not detected in the cell wall fraction but was instead located in the membrane and cytoplasmic fractions (Fig. 3a). This protein migrated more slowly than expected and may represent translation products that had not been cleaved by signal peptidase and sortase A.

Similar results were obtained for ClfAΔN1\(^{40-228}\) (Fig. S2). The failure to localize to the cell wall explains why ClfAΔN1\(^{40-220}\) and ClfAΔN1\(^{40-228}\) were not detected on the surface of intact bacterial cells. In contrast ClfAΔN1\(^{40-210}\) was only detected in the cell wall fraction in a similar fashion to full-length ClfA indicating that this truncate had been correctly secreted and sorted. As a control for secretion and sorting activity in cells attempting to express the ΔN1 constructs, samples were also probed for the expression of SdrE, a member of the MSCRAMM family which is known to be sorted to the cell wall (O’Brien *et al.*, 2002). This protein was detected only in the cell wall fraction indicating that it had been exported and sorted correctly in all strains (Fig. 3b). Fractions were also probed with antibodies recognizing the integral membrane protein F\(_1\)/F\(_0\) ATPase as a control for purity of cytoplasmic membrane preparations. This protein was detected only in the membrane fraction of all strains (Fig. 3c). Cytoplasmic fractions were stained with Coomassie to show equal loading of protein (Fig. 3d). In summary, these data indicate that ClfAΔN1\(^{40-220}\) is...
mislocalised to the membrane and cytoplasmic fractions while ClfAΔN1<sub>40-210</sub> is only found in the cell wall fraction in a similar fashion to full-length ClfA.

**Identification of residues in subdomain N1 required for ClfA to be localized to the cell wall.**

The N1 truncate ClfAΔN1<sub>40-210</sub> was sorted efficiently to the cell wall of *S. aureus* while ClfAΔN1<sub>40-220</sub> and ClfAΔN1<sub>40-228</sub> were found in the cytoplasm and membrane (Fig. 3a, Fig S3) suggesting that residues 211 - 220 may be required for efficient translocation and cell wall localization. In order to investigate this further, plasmid pRMC2clfA was manipulated to create a derivative lacking DNA encoding amino-acids 211–220 of ClfA (plasmid pRMC2clfAΔ<sub>211-220</sub>). ClfAΔ<sub>211-220</sub> was not detected in the cell wall but was present in the membrane and cytoplasm fractions (Fig. 4a) indicating that residues 211-220 are necessary for cell wall localization.

While ClfAΔN1<sub>40-210</sub> was detected in the cell wall fraction only (Fig. 3), a variant where residues 211-220 were substituted with ten alanines (ClfAΔN1<sub>40-220+10 Ala</sub>) was not (Fig 5a). These data show that residues 211-220 are necessary for cell wall localization of ClfA. SdrE was detected only in the cell wall fraction indicating that it had been exported and sorted correctly in all strains (Fig. 4b).

**Expression of a ClfA variant lacking subdomain N1 inhibits growth of *S. aureus.*

Expression of ClfAΔN1<sub>40-220</sub> or ClfAΔN1<sub>40-228</sub> could not be detected on the surface of *S. aureus* when NM1 (pRMC2clfAΔN1<sub>40-220</sub>) and NM1 (pRMC2clfAΔN1<sub>40-228</sub>) were grown in the presence of a high concentration of inducer (Fig. 2a). To determine if expression of ClfA
ΔN1_{40-220} is deleterious, the *clfA* gene was cloned into pALC2073, the precursor of pRMC2 which allows constitutive expression in *S. aureus* due to weak TetR repression (Bateman *et al.*, 2001; Corrigan & Foster, 2009). Plasmid pALC2073*clfAΔN1_{40-220}* was constructed in and isolated from *E. coli* DC10B. Equal amounts of the parental plasmid pALC2073*clfA* and pALC2073*clfAΔN1_{40-220}* DNA were used to transform *S. aureus* NM1. Transformation with pALC2073*clfA* yielded ~2 x 10^{3} Cm resistant colonies. However, pALC2073*clfAΔN1_{40-220}* yielded fewer than 20 transformants. Both plasmids could be introduced into *E. coli* with equal efficiency (~ 3 x 10^{4} CFU per transformation, Fig. S3) confirming the integrity of the DNA. The failure to establish transformants with pALC2073*clfAΔN1_{40-220}* in *S. aureus* suggested that this plasmid may be unstable or that constitutive expression of a variant of ClfA lacking subdomain N1 inhibits growth.

This prompted us to investigate if inducing expression of ClfAΔN1_{40-220} in *S. aureus* would inhibit bacterial growth. Gene expression from pRMC2 is tightly repressed in the absence of inducer. Plasmids pRMC2*clfA* and pRMC2*clfAΔN1_{40-220}* transformed into *S. aureus* with equal efficiency (data not shown). Competitive growth experiments were performed with bacteria harbouring pRMC2*clfA* or pRMC2*clfAΔN1_{40-220}*. In order to distinguish between cells harbouring the different plasmids, pRMC2*clfAΔN1_{40-220}* was introduced into strain NM2, a variant of NM1 where the *clfB::EmR* mutation was replaced by a deletion of the *clfB* gene (*ΔclfB*) which is Em sensitive. Equal numbers of NM1 (pRMC2*clfA*) and NM2 (pRMC2*clfAΔN1_{40-220}*) were mixed and co-cultured either in the presence or absence of the inducer ATc until the OD_{600} of the culture reached 2.0. Total viable counts were obtained by plating on agar containing Cm. Replica plating onto agar incorporating Em allowed the number of Em resistant bacteria NM1 (pRMC2*clfA*) to be counted.
When ClfA expression was induced and the bacteria were allowed to grow for two generations the number of NM2 (pRMC2clfAΔN140-220) recovered was 1.5 times lower than for NM1 (pRMC2clfA) (Fig. 5). This indicates that expression of ClfA lacking subdomain N1 confers a growth disadvantage compared to expression of full-length ClfA. In the absence of inducer when ClfA expression was repressed (Fig. 2a), equal numbers of NM1 (pRMC2clfA) and NM2 (pRMC2clfAΔN140-220) were recovered (Fig. 5). This indicates that the fitness disadvantage is due to the expression of ClfA ΔN140-220. To investigate if expression of wild-type ClfA affected the fitness of S. aureus, control experiments were conducted with NM1 (pRMC2clfA) and NM2 carrying the empty vector pRMC2. The proportion of EmS to EmR bacteria was the same for cultures grown in ATc and the inducer-free control (Fig. S4). These data indicate that expression of ClfAΔN140-220 in S. aureus inhibits bacterial growth.

Subdomain N1 is not required for cell wall localization of a ClfA variant lacking residues comprising the serine-aspartate repeat region.

The A domains of MSCRAMMs are linked to the cell wall by an extended flexible region. In the case of ClfA this comprises Ser-Asp dipeptide repeats. We hypothesised that the N1 subdomain might be required for the export of the flexible repeat region. To investigate this, plasmids pRMC2clfA and pRMC2clfAΔN140-220 were manipulated in order to create variants lacking DNA encoding the Ser-Asp dipeptide repeat region (residues 559-875) generating plasmid pRMC2clfAΔSD559-875 and plasmid pRMC2clfAΔN140-220ΔSD559-875, respectively. Both ClfAΔSD559-875 and ClfAΔN140-220ΔSD559-875 were detected only in the cell wall (Fig. 6). This demonstrates that the N1 subdomain is not required for export or cell wall
localization of ClfA lacking the flexible Ser-Asp repeat region and is only required for expression of the repeat containing protein.

Subdomain N1 is required for cell wall localization of another MSCRAMM with a long unstructured repeat region FnBPB. Region A of FnBPB shares ~25% amino-acid identity with the corresponding region of ClfA and is organised into N1, N2 and N3 subdomains (3). FnBPB does not have a Ser-Asp repeat region and instead harbours fibronectin binding repeats at the C-terminus of the protein. In order to determine if subdomain N1 was required for localisation of other S. aureus MSCRAMMs, we studied the subcellular localisation of FnBPB variants.

The fnbB gene from strain 8325 was cloned into pRMC2 to generate pRMC2fnbB. Plasmid pRMC2fnbB was manipulated to remove the entire N1 subdomain spanning residues 38-162 to yield pRMC2fnbBΔN138-162. Plasmids were transformed into a strain of the MRSA isolate BH1CC lacking both FnBPA and FnBPB as well as protein A (BH1CC ΔfnbAfnbB spa::KaR, Table 1). This was chosen as the host strain since FnBPB is not subject to proteolytic processing and remains intact on the surface of BH1CC during all stages of growth (Geoghegan et al., 2013).

Cell wall, membrane and cytoplasmic fractions were probed for FnBPB expression in a ligand affinity blot using biotinylated fibronectin. Similarly to ClfA, FnBPB is covalently anchored to the pentaglycine crossbridge of peptidoglycan and is released by lysostaphin treatment of S. aureus. FnBPBΔN138-162 was not detected in the cell wall fraction. Reactive bands with a molecular weight of ~100 kDa were observed in the membrane and cytoplasm fractions (Fig. 7a). This indicates that the N1 subdomain of FnBPB is required for cell wall localization of the protein.
To determine if the N1 subdomain is necessary for export of FnBPB in the absence of the fibronectin binding repeat regions, pRMC2\textit{fnbB}ΔN1\textsubscript{38-162} was manipulated to yield pRMC2\textit{fnbB}ΔN1\textsubscript{38-162}ΔFnBR\textsubscript{481-811} FLAG. A FLAG epitope tag was introduced by primer extension during PCR. A reactive band with a molecular weight of ~65 kDa was observed in the cell wall fraction (Fig. 7b). This demonstrates that the N1 subdomain is not required for export or cell wall localization of FnBPB lacking the flexible fibronectin binding repeat region.
Discussion

Clumping factor A is the archetypal member of the MSCRAMM family of wall-anchored proteins of *S. aureus*. The defining features of these proteins are (i) the A region and (ii) the flexible unfolded region linking the A region to the cell wall spanning/anchoring domain. The A domains comprise two separately folded subdomains N2 and N3. MSCRAMMs related to ClfA have an N-terminal subdomain N1 with no known function. This paper reports for the first time a crucial role for subdomain N1 in export and surface expression of MSCRAMMs. We demonstrate that residues 211-220 of N1 of ClfA are required for the protein to be translocated across the membrane so that it can be anchored by sortase to cell wall peptidoglycan. Deletion mutants lacking N1 were not exported and instead accumulated in the cytoplasm and the membrane fractions. This could only be detected in bacteria where ClfA expression was tightly repressed and was only expressed after addition of an inducer. Constitutive expression of an N1 deletion mutant following transformation into *S. aureus* resulted in a very low yield of transformants. Induction of the truncated *clfA* mutant in *S. aureus* caused a significant loss of fitness compared to cells expressing wild-type ClfA which shows that attempting to express the protein lacking N1 is causing damage to the cell. Dissection of the N1 subdomain of ClfA revealed that a stretch of 10 residues close to the boundary between N1 and N2 is necessary for surface expression, and that only this region plus the 8 residues that join it to the beginning of N2 as defined by the X-ray crystal structure is necessary. As yet no function has been identified for residues 40 – 210 of ClfA and the role of this region warrants further investigation.

We hypothesised that if this is a feature of the MSCRAMM family and not just a property specific to ClfA then the N1 subdomains of other MSCRAMMs would also be required for surface expression. Accordingly deletion of the N1 subdomain of FnBPB prevented surface expression of that protein. Further support comes from an
observation that a plasmid constitutively expressing an FnBPA N1 deletion mutant could not
be established in *S. aureus* (Geoghegan et al., 2013). The level of amino acid sequence
similarity between the N1 subdomains of the MSCRAMMs is very low so that it is not
possible to determine if a similar ‘motif’ is present in this region. All MSCRAMMs with a
repeat region contain an N1 subdomain. MSCRAMMs are found in many species of
staphylococci including *S. epidermidis, S. pseudintermedius* and *S. lugdunensis* (Bannoehr et
al., 2011; Heilbronner et al., 2011; Ponnuraj et al., 2003).

Previous studies have shown that MSCRAMM proteins are exported by the Sec
system with the other major secretory pathways (twin-arginine and accessory Sec system)
playing no role (DeDent et al., 2008). However, virtually nothing is known about how
proteins of the MSCRAMM family are trafficked to the Sec apparatus (Schneewind &
Missiakas, 2012; Sibbald et al., 2006). Previous studies identified a ‘YSIRK-GS’ motif in
the signal sequence of MSCRAMMs and certain other surface proteins (DeDent et al., 2008).
This feature is required for the proteins to be targeted to the cross-wall of the cell but despite
the fact that this phenomenon has been described for streptococci and staphylococci the
molecular details of the mechanism are completely lacking (Carlsson et al., 2006; DeDent et
al., 2008; Schneewind & Missiakas, 2012). Similarly, we have herein identified a new
region of the MSCRAMM required for its export, but the mechanism involved is not yet
understood. In the absence of a fundamental understanding of pre-protein trafficking in *S.
aureus* it is difficult to speculate about the functions of N1. It is possible that the N1
subdomain acts as a cytoplasmic chaperone by directly binding to another region of the
protein. Alternatively, residues 211-220 could engage a molecular chaperone that guides the
long protein from the ribosome to the Sec translocation apparatus. Since truncates lacking
the repeat region are exported and cell wall anchored in the absence of N1 it is likely that this
region is required to stabilise the long flexible repetitive repeat regions and to assist with transport.

In conclusion this paper provides new insights into the requirements for ClfA export. Our findings suggest that secretion of large MSCRAMM proteins by Gram-positive bacteria involves mechanisms more complex than previously appreciated.

Acknowledgements

This work was funded by a Science Foundation Ireland Programme Investigator grant 08/IN/B1845 and a Health Research Board grant RP/2009/19.
References


### Table 1

<table>
<thead>
<tr>
<th>Strain</th>
<th>Relevant Features</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>S. aureus</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Newman</td>
<td>Human clinical isolate, NCTC 8178</td>
<td>(Duthie &amp; Lorenz, 1952)</td>
</tr>
<tr>
<td>Newman clfA5 clfB::EmR</td>
<td>Derivative of strain Newman deficient in ClfA, ClfB. Frameshift mutation in the clfA gene, clfB::EmR</td>
<td>(Fitzgerald et al., 2006)</td>
</tr>
<tr>
<td>Newman spa</td>
<td>Derivative of strain Newman deficient in protein A. spa::KaR</td>
<td>(Higgins et al., 2006)</td>
</tr>
<tr>
<td>NM1</td>
<td>Derivative of strain Newman deficient in ClfA, ClfB and protein A. Frameshift mutation in the clfA gene, clfB::EmR spa::KaR</td>
<td>This study</td>
</tr>
<tr>
<td>NM2</td>
<td>Derivative of NM1. Replacement of clfB::EmR with a deletion of the clfB gene isolated by allele exchange.</td>
<td>This study</td>
</tr>
<tr>
<td>BH1CC ΔfnbAfnbB</td>
<td>Derivative of MRSA strain BH1CC deficient in FnBPA and FnBPB</td>
<td>(Geoghegan et al., 2013)</td>
</tr>
<tr>
<td>BH1CC ΔfnbAfnbB spa::KaR</td>
<td>Derivative of BH1CC ΔfnbAfnbB deficient in protein A. spa::KaR</td>
<td>This study</td>
</tr>
<tr>
<td><strong>E. coli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DC10B</td>
<td>dam^{+}Δdcm ΔhsdRMS endA1 recA1</td>
<td>(Monk et al., 2012)</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pALC2073</td>
<td>Vector allowing high level gene expression</td>
<td>(Bateman et al.,</td>
</tr>
<tr>
<td>Plasmid</td>
<td>Description</td>
<td>Reference</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>pALC2073clfA</td>
<td>pALC2073 containing full-length clfA gene.</td>
<td>A. Loughman, Trinity College Dublin. This study</td>
</tr>
<tr>
<td>pALC2073clfAΔN140-220</td>
<td>pALC2073 clfA lacking region encoding residues 40-220 of ClfA</td>
<td>This study</td>
</tr>
<tr>
<td>pRMC2clfA</td>
<td>Anhydrotetracycline-inducible expression vector with tight repression in the absence of inducer. Ap&lt;sup&gt;R&lt;/sup&gt; Cm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Corrigan &amp; Foster, 2009)</td>
</tr>
<tr>
<td>pRMC2clfAΔN140-220</td>
<td>pRMC2 clfA lacking DNA encoding residues 40-220 of ClfA</td>
<td>This study</td>
</tr>
<tr>
<td>pRMC2clfAΔN140-228</td>
<td>pRMC2 clfA lacking DNA encoding residues 40-228 of ClfA</td>
<td>This study</td>
</tr>
<tr>
<td>pRMC2clfAΔN140-210</td>
<td>pRMC2 clfA lacking DNA encoding residues 40-210 of ClfA</td>
<td>This study</td>
</tr>
<tr>
<td>pRMC2clfAΔ211-220</td>
<td>pRMC2 clfA lacking region encoding residues 211-220 of ClfA</td>
<td>This study</td>
</tr>
<tr>
<td>pRMC2clfAΔN140-220+10Ala</td>
<td>pRMC2 clfA lacking DNA encoding residues 40-220 and inserting DNA encoding 10 alanine codons.</td>
<td>This study</td>
</tr>
<tr>
<td>pRMC2clfAΔSD559-875</td>
<td>pRMC2 clfA lacking DNA encoding residues 559-875</td>
<td>This study</td>
</tr>
<tr>
<td>pRMC2clfAΔN140-220</td>
<td>pRMC2</td>
<td>This study</td>
</tr>
<tr>
<td><strong>ΔSD&lt;sub&gt;559-875&lt;/sub&gt;</strong></td>
<td><em>clfA</em> lacking DNA encoding residues 40-220 and 559-875</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------------------------------------</td>
<td>---</td>
</tr>
<tr>
<td><strong>pRMC2*fnbB</strong></td>
<td>pRMC2 containing full-length <em>fnbB</em> gene from strain 8325-4</td>
<td>This study</td>
</tr>
<tr>
<td><strong>pRMC2*fnbBΔN&lt;sub&gt;138-162&lt;/sub&gt;</strong></td>
<td>pRMC2<em>fnbB</em> lacking DNA encoding residues 38-167</td>
<td>This study</td>
</tr>
<tr>
<td><strong>pRMC2*fnbBΔN&lt;sub&gt;138-162ΔFnBR&lt;sub&gt;481-811&lt;/sub&gt;&lt;/sub&gt;</strong></td>
<td>pRMC2<em>fnbB</em> lacking DNA encoding residues 38-167 and 481-811</td>
<td>This study</td>
</tr>
<tr>
<td><strong>pIMAYΔ<em>clfB</em></strong></td>
<td>Plasmid for creating <em>clfB</em> deletion of Newman by allelic exchange. Carries 500 bp fragment of DNA from upstream of and 500 bp downstream from <em>clfB</em>. Cm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Mulcahy <em>et al.</em>, 2012)</td>
</tr>
</tbody>
</table>
**Figure Legends.**

**Fig. 1.** Schematic representation of ClfA derivatives. The N terminal signal sequence (SS) is followed by three independently folded subdomains, N1, N2 and N3 comprising the A domain. A flexible repeat region (R) links the A domain to the C-terminal Wall (W) and Membrane (M) spanning region which contain the LPXTG motif that allows anchoring of the protein to cell wall peptidoglycan by sortase A. The amino-acid coordinates of the signal sequence (SS), N1, N2 and N3 subdomains, SD repeat region (R) and cell wall (W) and membrane spanning (M) regions are indicated. The crystal structure of residues 229-545 of subdomains N2N3 is shown (Protein Data Bank code 2VR3). Residues 211-229 comprising the C- terminal region of subdomain N1 and the first residue of N2 as defined by the crystal structure are highlighted. Amino acid coordinates of the predicted cleavage site of the metalloprotease aureolysin are indicated.

**Fig. 2.** Expression of ClfA and ClfA derivatives. Dot immunoblot of intact bacteria (a) or whole cell lysates (b) of *S. aureus* NM1 expressing ClfA, ClfA ΔN1 40-220, ClfA ΔN1 40-228 or ClfA ΔN1 40-210 from pRMC2. Newman Δspa has been included as a positive control for ClfA expression and NM1 carrying pRMC2 empty vector serves as a negative control. The concentrations of inducer and the OD$_{600}$ of the cells are indicated. Filters were probed with ClfA mAb.

**Fig. 3.** Subcellular localization of ClfA, ClfAΔN1 40-220 and ClfAΔN1 40-210. Cellular fractions from *S. aureus* expressing ClfA, ClfA ΔN1 40-220 or ClfA ΔN1 40-210 were probed with a) ClfA mAb, b) anti-SdrE IgG or c) anti-ATPase IgG in a Western immunoblot. Fractions are
labelled cell wall (CW), membrane (CM) and cytoplasm (CP). Cytoplasm fractions were stained with Coomassie brilliant blue (d).

**Fig. 4.** Subcellular location of ClfA, ClfAΔN140-220 + 10 Ala and ClfAΔ211-220. Cellular fractions from *S. aureus* expressing ClfA induced with 800 ng ml⁻¹ ATc, ClfAΔN140-220 + 10 Ala and ClfAΔ211-220 induced with ATc (1200 ng ml⁻¹) were probed with a) ClfA mAb or b) anti-SdrE IgG in a Western immunoblot. Fractions are labelled cell wall (CW), membrane (CM) and cytoplasm (CP).

**Fig. 5.** Growth inhibitory effects of expression of ClfAΔN140-220.

*S. aureus* NM1 (pRMC2clfA) and NM2 (pRMC2clfAΔN140-220) cells were induced with ATc (1200 ng ml⁻¹) and plated on agar incorporating Cm. Colonies resistant to Cm were replica plated onto agar containing Em. Bars represent the mean percentage CFU based on the total Cm resistant CFU. Error bars represent the standard deviation on the mean of three independent experiments. Statistical analysis was performed using Students *t* test.

**Fig. 6.** Subcellular location of ClfA, ClfAΔSD559-875 and ClfAΔN140-220ΔSD559-875

Cellular fractions from *S. aureus* expressing ClfA, ClfAΔSD559-875 or ClfAΔN140-220 ΔSD559-875 probed with ClfA mAb. Fractions are labelled cell wall (CW), cytoplasmic membrane (CM) and cytoplasm (CP).
Fig. 7. Subcellular location of FnBPB, FnBPΔN138-162 and FnBPΔN138-162ΔFnBR481-811.

a) Cellular fractions from *S. aureus* expressing FnBPB or FnBPΔN138-162 probed with biotinylated fibronectin and detected with HRP conjugated streptavidin. Fractions are labelled cell wall (CW), membrane (CM) and cytoplasm (CP).

b) Cellular fractions from *S. aureus* expressing FnBPΔN138-162ΔFnBR481-811 probed with monoclonal rabbit anti-FLAG IgG. Fractions are labelled cell wall (CW), membrane (CM) and cytoplasm (CP).
<table>
<thead>
<tr>
<th>Strain</th>
<th>ATc concentration (ng ml(^{-1}))</th>
<th>OD(_{600})</th>
<th>A Tc concentration (ng ml(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newman (spa)</td>
<td>0</td>
<td>0.5</td>
<td>Newman (spa)</td>
</tr>
<tr>
<td>NM1 pRMC2</td>
<td>0</td>
<td>0.25</td>
<td>NM1 pRMC2 1200</td>
</tr>
<tr>
<td>NM1 pRMC2 (clfA)</td>
<td>0</td>
<td>0.125</td>
<td>NM1 pRMC2 (clfA)</td>
</tr>
<tr>
<td>NM1 pRMC2 (clfA)</td>
<td>100</td>
<td>0.063</td>
<td>NM1 pRMC2 (clfA)ΔN1(_{40-220})</td>
</tr>
<tr>
<td>NM1 pRMC2 (clfA)</td>
<td>200</td>
<td></td>
<td>NM1 pRMC2 (clfA)ΔN1(_{40-210})</td>
</tr>
<tr>
<td>NM1 pRMC2 (clfA)</td>
<td>400</td>
<td></td>
<td>NM1 pRMC2 (clfA)ΔN1(_{40-228})</td>
</tr>
<tr>
<td>NM1 pRMC2 (clfA)</td>
<td>800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM1 pRMC2 (clfA)</td>
<td>1200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM1 pRMC2 (clfA)ΔN1(_{40-220})</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM1 pRMC2 (clfA)ΔN1(_{40-220})</td>
<td>1200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM1 pRMC2 (clfA)ΔN1(_{40-210})</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM1 pRMC2 (clfA)ΔN1(_{40-228})</td>
<td>1200</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
170kDa
130kDa
100kDa
70kDa
55kDa
40kDa

ClfA
ClfA ΔSD_{559-875}
ClfA ΔN1_{40-220} ΔSD_{559-875}

CW  CM  CP
CW  CM  CP
CW  CM  CP

ClfA mAb