Region A of ClfB is composed of three subdomains

Structural organization of the fibrinogen-binding region of the

Clumping factor B MSCRAMM of *Staphylococcus aureus*

Running Title: Region A of ClfB is composed of three subdomains

Samuel Perkins¹# and Evelyn J. Walsh,²# Champion C. S. Deivanayagam,³ Sthanam V. L. Narayana,³ Timothy J. Foster,² and Magnus Höök¹*.

¹Institute for Biosciences and Technology, Texas Medical Center, 2121 West Holcombe Boulevard, Houston TX 77030-303, USA

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

³Center for Biophysical Sciences and Engineering, School of Optometry, 286 BHSB, 1918 University Boulevard, University of Alabama at Birmingham, Birmingham, Al 35205, USA.

*Author to whom correspondence should be addressed

#Both authors contributed equally in this work

Keywords: *Staphylococcus aureus*, Fibrinogen, Clumping factor, Adhesin,
Region A of ClfB is composed of three subdomains

ABSTRACT

The clumping factor B (ClfB) of Staphylococcus aureus is a surface protein that binds to fibrinogen (Ní Eidhin, D., Perkins, S., Francois, P., Vaudaux, P., Hook, M., and Foster, T. J., 1998 Mol Microbiol 30, 245-257). The ligand-binding activity is located in the ~500 residue A-region (44-542) which represents the N-terminal half of the MSCRAMM protein. We now hypothesize that the ClfB A-region is comprised of three subdomains, which we have named N1, N2, and N3, respectively. To examine this hypothesis we expressed recombinant forms of the individual putative subdomains, the tandem motifs N12 and N23, and the full-length A-region N123. Far UV Circular Dichroism spectra showed that each subdomain is composed mainly of β-sheets with little or no discernable α-helices. Heat induced unfolding of individual subdomains occurred with a single state transition and was reversible, indicating that the subdomains can fold as discreet units. Gel permeation chromatography indicated that N2, N3 and N23 are globular. In contrast, domain N1 appeared to be elongated and conferred a somewhat elongated structure on segments containing this subdomain (i.e. N12 or N123). N123, N12 and N23 all bound to fibrinogen, but N23 had a higher affinity for fibrinogen than that observed for the full-length A-region; N123 or for N12. However, an extended N-terminus of N23 was required for ligand binding. A form of N23 that was generated by proteolytic processing and lacked the N-terminal extension was unable to bind fibrinogen. Recombinant forms of individual subdomains did not bind fibrinogen. Addition of recombinant N23 effectively inhibited ClfB-mediated bacterial adherence to fibrinogen, N123 caused some reduction in bacterial attachment, whereas N12 was essentially inactive. Antibodies raised against the central N2 domain of the A-region were the most effective at inhibiting bacterial adhesion to immobilized fibrinogen, although anti-N3 or anti-N1 antibodies also caused some reduction in ClfB-mediated adherence to fibrinogen.
Region A of ClfB is composed of three subdomains

**INTRODUCTION**

*Staphylococcus aureus* is an important pathogen of man and animals. It causes a variety of infections ranging from superficial skin infections to deep-seated local lesions and life-threatening systemic infections. The molecular pathogenesis of these infections are complex processes initiated by the adherence of the bacteria to tissues in the host. Bacterial adherence involves surface structures called adhesins that in specific interactions bind to ligands in the host. For *S. aureus*, which primarily causes extracellular infections, the responsible adhesins are of the MSCRAMM (microbial surface components recognizing adhesive matrix molecules) subfamily that specifically target extracellular matrix components in host tissues (1). So far, eight structurally related putative *S. aureus* MSCRAMMs have been identified and partially characterized (2). These cell wall anchored proteins all contain a ~500 amino acid long A-region with significant sequence similarity (20-61% pairwise identity) and similar predicted secondary structure compositions. Deconvolution of CD spectra of isolated recombinant MSCRAMM A-regions; Collagen-binding adhesin, CNA, (3); Clumping factor A, ClfA, (4); and Fibronectin-binding adhesin, FnbpA, (5); supports this prediction and indicate that the A-regions are composed primarily of β-sheets and coils with a small amount of α-helix. A ligand-binding subdomain of the A-region of CNA was crystallized and its structure solved. The secondary structure composition of this protein as determined by its x-ray crystal structure was 8% α-helix, 53% β-sheet, and 39% coil (6).

*S. aureus* can express at least four distinct fibrinogen-binding MSCRAMMs; the fibronectin binding MSCRAMMs FnbpA and FnbpB, which were recently shown to also bind fibrinogen (5) and the clumping factors A and B (ClfA and ClfB: [7, 8]). Both clumping factors promote adhesion to immobilized fibrinogen *in vitro*, to blood clots, to *ex vivo* biomaterial, and to damaged heart valve tissue in the rat model of endocarditis (8-12). ClfA and ClfB are structurally related and have an identical domain organization (Fig 1) with a N-
Region A of ClfB is composed of three subdomains terminal secretory signal sequence, the ligand-binding A-region, and a dipeptide repeat region R composed mainly of aspartate and serine residues. Region R is required to project the ligand-binding region from the cell surface and to transverse the cell wall (13). Sequences involved in positioning the proteins at the cell surface and anchoring them to the cell wall (W) are located towards the C-termini and include an LPXTG motif, a hydrophobic membrane-spanning domain (M), and a short positively charged cytoplasmic tail (C). The ligand-binding A-regions of ClfA and ClfB exhibit 26% residue identity. Despite sequence and structural similarity of the two clumping factors they target different sites in fibrinogen. ClfA binds to the extreme C-terminus of the γ-chain (4, 14) while ClfB primarily binds to the α-chain (8). Furthermore the expression of the two proteins in the growth curve of S. aureus differ; ClfA is expressed at all stages of growth, whereas ClfB is only found on the surface of cells in early exponential phase (8).

In this communication we report on a biophysical characterization of the ligand-binding A-region of ClfB. We propose that the A-region of ClfB is composed of 3 subdomains that we have called N1, N2, and N3 (Fig 1). The reasoning for this hypothesis is as follows. The A-region of ClfB can be cleaved at the two Alanine residues in the motif; SLAVA by metalloproteases removing residues 44-197 or 44-199 from the N-terminus (15). This suggests that the SLAVA motif is part of a sequence that separates two distinct domains. We propose that this N-terminal segment represents the N1 subdomain. Sequence and predicted secondary structure alignments of ClfB and CNA suggest that the minimal collagen binding domain of CNA corresponding to residues 197-375 show 15% identity and 45% homology with the proposed N2 subdomain of ClfB. We further propose that the C-terminal third of the A-region represents the third subdomain called N3, defined by residues 375-542. We now have expressed the putative subdomains of ClfB A-region as recombinant proteins; as individual subdomains (N1, N2, N3), tandem motifs (N12, N23), and full-length A-region
Region A of ClfB is composed of three subdomains (N123). These recombinant proteins are characterized and tested for their ability to bind fibrinogen.
Region A of ClfB is composed of three subdomains

EXPERIMENTAL PROCEDURES

Bacteria and growth conditions - Escherichia coli strains XL1-Blue (16) or JM101 (17) were used as the bacterial host for plasmid cloning and E. coli TOPP 3 (18) was used as a host for protein expression. E. coli cells harbouring plasmids were routinely grown in Luria broth (LB) or on L-agar. Ampicillin (100µg/ml) or tetracycline (10 µg/ml) was incorporated as appropriate.

Manipulation of DNA - Restriction and DNA modification enzymes were purchased from New England Biolabs Inc., Promega or Boehringer Mannheim, and were used according to the manufacturers’ instructions. DNA procedures were carried out using standard methods (19).

Amplification of clfB gene fragments – Segments encoding the different subdomains of the ligand binding A-region of ClfB shown in Fig 1 were amplified by PCR from a recombinant plasmid pAE-1EX (8) containing the clfB gene using the oligonucleotides listed in Table 1. Restriction enzyme cleavage sites were introduced at the 5’-ends of the oligonucleotides to facilitate directional cloning into the expression vector. PCR was performed with a Perkin-Elmer DNA thermocycler. Reaction mixtures (100 µl) contained 250 µM dNTPs, 5 ng pAE-1EX, 100 picomoles primers and 5 U Pfu polymerase in the standard Promega Pfu reaction buffer. The reaction mixtures were overlaid with 100 µl of mineral oil and subjected to 30 amplification cycles. Each cycle consisted of a denaturation step at 94 °C for 1 min, a 2 min annealing step at 46-52 °C depending on the primer pair, and an extension step at 72 °C for 1-2 min depending on the length of the fragment to be amplified. A final extension step for 5 min at 72 °C was carried out. The PCR products were purified using the High Pure PCR Product Purification Kit (Boehringer Mannheim) and analysed by agarose gel electrophoresis.
Region A of ClfB is composed of three subdomains

**Construction of Expression Plasmids** - Amplified fragments of the *clfB* gene were cloned into the expression plasmid pQE30 (Qiagen Inc) to generate the constructs shown in Table 1. Recombinant proteins expressed from this vector contain an N-terminal tail of six histidine residues (His-tag).

**Expression of Recombinant Proteins** - Recombinant plasmids pN1, pN2, pN3, pN23i and pN123 were transformed into *E. coli* strain JM101, while plasmids pN12 and pN23 were transformed into *E. coli* XL1-Blue cells. The integrity of recombinant plasmids were analysed by restriction enzyme digestion and sequencing at the Molecular Genetics Core Facility, UT-Houston Medical School which confirmed the expected sequence of each insert. Starter cultures of *E. coli* TOPP 3 containing the recombinant plasmids were diluted 1/20 in LB containing ampicillin and incubated with shaking until the culture reached an OD<sub>600</sub> of 0.8-1.0. Protein expression was induced by the addition of isopropyl β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.1mM and cells were incubated for an additional 2 h. Bacterial cells were harvested by centrifugation, resuspended in a minimal volume of PBS and frozen at −20 °C.

**Purification of Recombinant ClfB Proteins** - Cells were lysed by passage through a French press (1100 lb/in<sup>2</sup>) and complete EDTA-free, protease inhibitor cocktail tablets (Boehringer Mannheim) were added immediately. Cellular debris and insoluble proteins were removed by centrifugation at 20,000 x g for 20 min. The cleared, lysed cell supernatant was filtered through a 0.45 μm membrane and applied to a 5 ml HiTrap Chelating column (Pharmacia) of nickel-charged iminodiacetic acid-Sepharose. Bound protein was eluted with a continuous linear gradient of imidazole (0-200mM; total volume 200ml) in 4mM Tris-HCl, 100mM
Region A of ClfB is composed of three subdomains NaCl (pH 7.9). Eluted proteins were monitored by their absorbance at 280nm. This purification step yielded proteins that were >95% pure as estimated by SDS-PAGE. Fractions containing the recombinant proteins were pooled, diluted 10-fold with water and applied to a 5 ml HiTrap Q Sepharose column (Pharmacia). Bound protein was eluted with a continuous linear gradient of NaCl (50-500 mM; total volume 100 ml) in 20mM Tris and 2mM EDTA (pH 7.9). Eluted fractions were monitored by absorbance at 280 nm. Peak fractions were analysed by SDS-PAGE, pooled and dialyzed against 50 mM EDTA in water to remove Ni$^{2+}$, and subsequently against water to remove excess EDTA. All samples were frozen at -80°C, lyopholized for mass spectrometry or used immediately in assays. To verify the proteins’ identities electrospray ionization mass spectrometry (EI-MS) measurements and occasionally N-terminal sequencing were carried out at the Tufts Core Facility (Tufts University Department of Physiology 136 Harrison Avenue, Boston Massachusetts 02111) or the Baylor College of Medicine Core Lab, (Protein Chemistry Core Facility Baylor room 5416 Houston Texas, 77030) respectively.

**SDS-PAGE** - Recombinant proteins were analysed by SDS-PAGE performed according to standard procedures (20) on gradient gels containing 10-15% acrylamide. Gels were stained with Coomassie blue.

**Gel Permeation Chromatography** - Stokes radii were measured at 25 °C using gel-permeation chromatography (GPC). Protein samples were diluted in 10mM Na$_2$HPO$_4$, 100 mM NaCl, pH 7.4, applied to a HR 10/30 Superose 12 and eluted with the same buffer at a flow rate of 0.3 ml/min. The column was calibrated using a gel filtration standard protein kit (Bio-Rad). The proteins contained in the kit were bovine thyroglobulin (670,000 Da), bovine gamma globulin (158,000 Da), chicken ovalbumin (44,000 Da), horse myoglobin (17,000 Da), and
Region A of ClfB is composed of three subdomains.

human vitamin B-12 (1,350 Da). Blue Dextran 2000 (2,000,000 Da) (Amersham Pharmacia) and Tyrosine (181 Da) (Sigma) were used to indicate the void volume and the bed volume, respectively. Protein radii were calculated from formulas outlined by Uversky (21).

Circular Dichroism - The secondary structural composition of recombinant proteins was examined by circular dichroism (CD) spectroscopy. Far-UV CD data were collected using a Jasco J720 spectropolarimeter calibrated with d-10-camphorsulfonic acid, employing a bandpass of 1 nm and integrated for 4 s at 0.2 nm intervals. All sample concentrations were less than 0.5 mg/ml in 20 mM Tris-HCl buffer, pH 7.4. Spectra were recorded at ambient temperature in cylindrical 0.2 mm path length cuvettes. Thirty scans were averaged for each spectrum, the contribution from the buffer was subtracted, and quantitation of secondary structural elements was performed by deconvolution software provided by UMDNJ-Robert Wood Johnson Medical School, (Piscataway, NJ) and D. Greenwood (Softwood Co., Brooksfield, CT). These deconvolution programs (SELCON and VARSCLC1) are derived from databases of known protein structures (22, 23). These two programs were chosen based on their ability to accurately predict the secondary structure of the N2 domain of CNA (3).

Unfolding/refolding monitored by CD – Recombinant proteins were monitored by CD at 205 nm. A one cm pathlength jacketed cell was used connected to a temperature bath. Three unfolding scans were measured for each of the recombinant proteins: N1, N2, N3, N12, N23, and N123. Refolding scans were also measured for individual subdomains as well as N123. The concentration of each of the proteins was less than 100 µg/ml in 10mM Na2HPO4, 100 mM NaCl, pH 7.4 (PBS).

Proteolytic cleavage of recombinant ClfB N23 by S. aureus culture supernatant-

Recombinant N23 (1 mg/ml) containing a N-terminal His-tag was incubated with
Region A of ClfB is composed of three subdomains concentrated culture supernatant (80 µl) from S. aureus strain DU5966 prepared as described by (15). The sample was incubated at 37 °C for 90 min and then passed through a Ni²⁺ activated HiTrap column, unbound protein (missing the His-tag) was collected and concentrated.

**Solid phase binding assay** – Immulon-1 microtiter plates were coated overnight at 4 °C with 1µg of fibrinogen in 100µl PBS per well. Fibrinogen-coated wells were then washed twice with PBS and incubated at 37 °C for 2 h with 5% (w/v) bovine serum albumin (Oxoid) in PBS (200µl/well). Wells were washed again with PBS and increasing concentrations of recombinant proteins labelled with EZ-link Biotin according to the manufacturer’s recommendation (Pierce) were added at 100µl PBS/well. The plates were incubated for 2 h at 37 °C, and unbound protein was subsequently removed by washing with PBS. Bound protein was detected by the addition of Avidin conjugated with alkaline phosphatase (Pierce, 1:10,000 dilution). Relative binding was measured by monitoring in a microplate reader (Molecular Devices) the sample absorbance at 405 nm following the addition of p-nitrophenyl phosphate (Sigma) in 1M diethanolamine, 0.5 mM MgCl₂, pH 9.0.

**Preparation of antisera**- Antibodies to N3 were raised in young New Zealand white rabbits (2 kg) whose preimmune sera did not react with S. aureus antigens as determined by Western blots. The protein (25 µg) in PBS was emulsified with an equal volume of Freund’s complete adjuvant (500 µl), and injected subcutaneously. Three subsequent injections given at 2-week intervals contained Freund’s incomplete adjuvant. The rabbits were bled, serum recovered and fractionated (24), and IgG was purified by affinity chromatography on Protein
Region A of ClfB is composed of three subdomains

A Sepharose (Sigma). Antibodies to N1 and N2 were as described by McAleese et al. (15). Antibodies to rClfA A-region were as described by McDevitt et al. (7).

Comparison of antibody reactivity- ELISA plates were coated overnight at 4 °C with recombinant N123 in PBS (1 µg/well). Plates were blocked with BSA (5% w/v) at room temperature for 2 h. After three washes with PBS IgG isolated from antisera raised to N1, N2, and N3, respectively; was diluted in BSA (1% w/v) and added to the plates (100 µl/well) for 1 h at room temperature. The plates were then washed three times with PBS and horseradish peroxidase-labeled goat anti-rabbit IgG was added (1:2000 dilution, 100 µl/well) for 1 h at room temperature. After further washes with PBS, 100 µl of a chromogenic substrate solution (580 µl/ml tetramethylbenzidine and 0.0001% H2O2 in 0.1 M sodium acetate buffer pH 5.2) was added per well and color development was allowed to proceed for 10 min in the dark. The reaction was stopped by the addition of 2 M H2SO4 (50 µl/well) and the absorption at 450 nm was quantitated using an ELISA plate reader (Labsystems Multiskan Plus).

Inhibition of bacterial cell adherence to immobilized fibrinogen - Adherence of S. aureus cells to immobilized fibrinogen was performed according to the method of Wolz et al. (25) with some modifications. Briefly, microtiter wells were incubated overnight at 4 °C with 100 µl of 5 µg/ml fibrinogen (Calbiochem) in PBS. Bovine serum albumin (200 µl/well of a 5 mg/ml solution) was added and the plates were incubated for 2 hrs at 37 °C. The plates were washed three times with PBS and recombinant inhibitory proteins were added for 2 hrs at room temperature. A bacterial cell suspension (50 µl, 2 x10^8 cfu/ml) was added to the wells and incubated for 90 minutes at 37 °C. The wells were washed with PBS and adherent cells
Region A of ClfB is composed of three subdomains

were fixed with formaldehyde (25% v/v) for 30 minutes at room temperature. The wells were washed again, and adherent cells were stained with crystal violet (5% w/v) for one minute, washed three times with PBS. Acetic acid (10% v/v) was then added (100 µl/well) for 10 minutes at room temperature. Absorbance was measured at 570 nm in an ELISA plate reader.

For inhibition by polyclonal antibodies, bacterial cells (50 µl, 2 x 10^8 cfu/ml) were pre-incubated with various concentrations of antibody (1-1000 µg/ml IgG). The incubation mixture was then added to the wells (100 µl/well) and bacteria were allowed to adhere to the substrate. Plates were washed and adherent cells were fixed with formaldehyde, stained with crystal violet and quantitated by measuring absorbance at 570 nm as above.
Region A of ClfB is composed of three subdomains

**RESULTS**

*Recombinant ClfB region A.* - SDS-PAGE analysis of the purified recombinant ClfB A-region comprising residues 44-542 followed by staining with Coomassie brilliant blue revealed a single band with an apparent molecular weight of 68 kDa (Fig. 2, lane 1). This is higher than the molecular mass of 56,607 Da predicted from the deduced amino acid sequence and also higher than the electrospray ionization mass spectroscopy (EI-MS) value of 56,601 Da determined for the intact region A. SDS-PAGE analysis of the protein after storage at 4 °C for several days showed that the intact recombinant A-region was degraded to a component with an apparent Mₘ of 39,000 Da (Fig. 2, lane 3). N-terminal sequence analysis of the A-region truncate showed that cleavage occurred between residues Leu 198 and Ala 199 (26) or between residues Val 200 and Ala 201. EI-MS gave a molecular mass of 38,956 Daltons for the larger truncate compared to 38,964 Daltons predicted for the segment corresponding to residues 199-542, indicating that cleavage of ClfB A-region occurred only at the N-terminus. Storage of the recombinant A-region in EDTA prevented processing to the 39 kDa form suggesting that a contaminating *E. coli* metalloprotease was responsible for the processing. In a separate study, the *S.aureus* metalloprotease aureolysin subsequently was shown to cleave both native and rClfB (44-542) at the SLAVA sequence (15).

*Domain organization of ClfB region A.* – The argument outlined in the introduction section led us to propose that the ClfB A-region is composed of three subdomains (Fig 1). A set of recombinant proteins were constructed and isolated with a tag at the N-terminus composed of six His residues to facilitate purification. The following recombinant proteins were made: N1 corresponding to residues (44-196), N2 (197-375), N3 (376-542), combinations of subdomains (i.e. N12 and N23) and the full-length A-region (N123) (Fig 1).
Region A of ClfB is composed of three subdomains

Analysis of these recombinant proteins by SDS-PAGE showed that N2, N3 and N23 migrated at their predicted molecular weights, whereas the full-length recombinant A-region, N123 (Fig 2), and the N1 and N12 proteins migrated slower than predicted (data not shown). A hydropathy index analysis (27) of the A-region of ClfB revealed that the N-terminal segment corresponding to N1 is much more hydrophilic than the central and C-terminal A-domain segments comprising subdomains N2 and N3.

**Gel Permeation Chromatography (GPC)** - Recombinant proteins were analysed by GPC on a HR Superose 12 column with a resolution range from 10-500 kDa, and compared with known globular protein standards (Fig. 3). Using the equation $K_{av} = (V_x - V_o)/(V_t - V_o)$ (28) the $K_{av}$, or partition coefficient, can be calculated based on the sample elution volume ($V_x$), the void volume ($V_o$) and the total column bed volume ($V_t$). From these measurements, we calculated the “effective radius” (the protein’s radius assuming it folds as a fully hydrated sphere) (3).

Proteins N2, N3, and N23 eluted at the expected positions for globular proteins of equal mass, and the effective radii are similar to those calculated for globular proteins of equal mass (Table 2). All proteins containing the N1 subunit eluted earlier than expected for a globular protein of similar size. The complete ClfB A-region (N123) eluted earlier than expected for a globular protein of 56 kDa, and behaved as a globular protein with an effective radius of 39.5 Å (Table 2). The N12 protein despite its smaller size eluted at a position similar to that of the intact A-region (Fig 3) (Table 2). The effective radius for N12 was 40 Å compared to a radius of 27 Å calculated for a globular protein with the mass of N12. The N1 domain had an effective radius of 29 Å compared to 20.9 Å for a globular protein of the same mass (Fig. 3 and Table 2). These properties of N1-containing proteins can be explained if N1 has an extended structure.
Region A of ClfB is composed of three subdomains

Circular Dichroism analysis - Far-UV CD analysis was performed on all of the recombinant proteins to determine their secondary structure composition. The CD spectra of N123 and the three subdomain proteins; N1, N2, and N3 are shown in Fig. 4. Deconvolution of the spectra (using the programs Selcon and Varselec) revealed that the full-length A-domain is primarily composed of β-sheets and loops with a small percentage α-helix (Table 3). The composition of secondary structures in each protein is calculated in Table 3. It is noteworthy that the CD spectra of N1, N2, and N3 are different but that each domain has a substantial amount of secondary structure. Furthermore deconvolution of the CD spectra suggests that the secondary structure composition of each subdomain is similar and composed mainly of β-sheets and loops.

Temperature dependent denaturation and refolding of putative subdomains observed by CD - Temperature dependent denaturation of recombinant ClfB subdomains, N1, N2, and N3, demonstrate that these proteins unfold within a relatively narrow temperature range (Fig. 5 and Table 3). The midpoints of denaturation were: N1 51.6 ± 1.1 °C, N2 50.4 ± 0.5 °C, and N3 38.5 ± 1.4 °C. After unfolding, individual subdomain proteins were refolded by cooling. All three individual subdomains; N1, N2, and N3; were able to refold (Fig. 5). Taken together these results indicate that each recombinant segment can fold individually and represents a subdomain.

When N12 and N23 subdomains were analysed in a similar manner both proteins unfolded in one step within a very narrow temperature range with melting temperatures (T_m) of 51.0 ± 0.2 °C and 50.9 ± 0.3 °C, respectively. Also the recombinant full-length A-region (N123) unfolds in one step within a very narrow temperature range with a T_m of 49.1 ± 0.2 °C (Table 3). This process was reversible when the full-length A-region was immediately cooled.
Region A of ClfB is composed of three subdomains back to 20 °C indicating that no aggregation or precipitation events are occurring. Recombinant N23 does not unfold at the T_m (38.5 °C) of the isolated N3, indicating the presence of stabilizing interactions between the N2 and N3 subdomains.

**Binding of recombinant proteins to fibrinogen** - An ELISA-type ligand-binding assay was used to characterize the interaction of the recombinant proteins with immobilized fibrinogen. The full-length A-region N123 and two truncates, N23 and N12, each bound to immobilized fibrinogen in a concentration dependent manner (Fig. 6a). In contrast, individual subdomains (N1, N2, and N3) did not bind to the fibrinogen coated wells (data not shown). The recombinant N23 protein bound fibrinogen with highest affinity. The fibrinogen-coated wells were saturated with N23 and an apparent binding coefficient (K_D) of 2 x 10^{-8} M was determined for this interaction. Proteins N123 and N12 bound with lower affinity. The fibrinogen-coated wells were not saturated with bound recombinant N123 and N12 at the concentrations used and the apparent K_Ds for these interactions are >1 x 10^{-7} M and >2 x 10^{-7} M, respectively.

Processing of the recombinant N123 occurred if metalloprotease activity was not inhibited by EDTA. Cleavage by *E. coli* proteases generated a protein corresponding to residues 199-542 but without a His-Tag as described above. This protein did not bind to fibrinogen (data not shown). Furthermore, loss of the His-tag from N23 (residues 197-542) as a result of treatment with *S. aureus* metalloprotease completely abolished the ability of the processed protein to bind immobilized fibrinogen (Fig 6b and (15)). These results suggest that an intact N-terminus of N23 is critical for the fibrinogen-binding activity of this protein. However, a second N23 construct containing the second and third subdomains called N23i, starting at A199 bound to fibrinogen when it had an N-terminal His-tag present. The residues Ser 197 and Leu 198, present in N23 but not N23i, are not necessary for ligand-binding but
Region A of ClfB is composed of three subdomains the presence of a short N-terminal extended segment, which in N23i is represented by the unrelated His-tag, maintains fibrinogen binding activity of the protein.

*Inhibition of Staphylococcus aureus cells adhering to fibrinogen by recombinant proteins* – A strain of *S. aureus*, DU5874, in which the ClfA gene has been inactivated, adheres to adsorbed fibrinogen in a ClfB-mediated process (8). The different recombinant subdomain constructs of ClfB A-region were examined for their ability to inhibit DU5874 adherence to fibrinogen (Fig 7). In this assay N23 effectively inhibited bacterial attachment in a concentration dependent manner resulting in 85% inhibition at 0.1 mM N23. The full-length A-region N123 also caused some reduction in bacterial attachment at the highest concentration tested whereas N12 was essentially inactive. Thus the differences in apparent affinities seen between the domain constructs in their binding to fibrinogen are also reflected by a difference in their ability to inhibit ClfB dependent bacterial attachment to fibrinogen.

*Inhibition of Staphylococcus cells binding to fibrinogen by antibodies against individual domains* – Rabbit antisera were raised against the individual ClfB A-region subdomains, N1, N2, and N3, and IgG was purified by affinity chromatography on a Protein A column. The relative reactivity of these antibodies to recombinant ClfB N123 was examined in an ELISA. The absorbed N123 was saturated with antibody from each of the preparations when these were added at a concentration of 2 µg or higher (Fig. 8A). The maximal amount of IgG bound to ClfB was highest from the anti-N3 preparation followed by the anti-N1 and the anti-N2 derived IgG, respectively. These antibody preparations were tested for their ability to inhibit ClfB dependent *S. aureus* adherence to a fibrinogen substrate using the DU5874 strain (Fig 8B). Our results indicate that anti-N2 IgG was the most effective inhibitor of cell adhesion causing 50% inhibition of cell adhesion at 15 µg/ml. Anti-N3 antibodies also
Region A of ClfB is composed of three subdomains effectively inhibited bacterial adhesion with a 50% reduction at 50 µg/ml. Additionally, anti-N1 antibodies caused some reduction in adherence, but were less potent than anti-N2 and anti-N3. IgG antibodies raised against recombinant ClfA did not inhibit ClfB-mediated bacterial adherence.
DISCUSSION

ClfB is a member of a family of structurally related surface proteins present on S. aureus cells that currently includes seven additional members; ClfA, FnbpA, FnbpB, SdrC, SdrD, SdrE, and CNA. These proteins are all of similar size (Mw~ 100,000 Da) and contain at the C-terminus, structural features required for cell wall anchoring including an LPXTG sequence recognized by the transpeptidase sortase, a hydrophobic domain which may span the membrane and a short cytoplasmic tail composed primarily of positively charged residues. Although the amino acid sequences immediately outside the cell wall anchoring domain vary substantially among the different members, some specific regions are found in several proteins and smaller subgroups can be defined within the family based on these shared regions (for a review see [2]). However, the N-terminal halves of all of the proteins, called the A-region, are similar in size and composed of significantly similar amino acid sequences. Furthermore, deconvolution of the CD-spectra of recombinant versions of these A-regions suggests that they contain a similar composition of secondary structures.

We now propose that each of the A-regions are composed of three subdomains that we have called N1, N2, and N3. Using ClfB as a model protein we have attempted to provide support for this hypothesis. The border between the N1 and N2 subdomains was defined by a proteolytically sensitive SLAVA motif. Furthermore, comparison of amino acid sequence and secondary structure prediction between the minimal collagen-binding motif of CNA and ClfB defines N2. The different subdomains expressed as recombinant proteins have a defined secondary structure composition as shown by CD analysis and they refold after heat denaturation suggesting that they could represent individual folding units. Recombinant N2 and N3 behaved on gel permeation chromatography as globular proteins and it appears likely that these regions represent distinct subdomains. The N1 protein behaves as an extended, non-globular protein in gel permeation chromatography. Another possibility is that the N1
subdomain forms a dimer, which would account for the larger Stokes radii ($R_s$) for proteins containing N1 in Table 2. However, neither non-reducing SDS PAGE analysis nor analytical ultracentrifugation analysis showed the presence of N1-containing dimers (data not shown). Furthermore, N1 contains a large percentage of hydrophilic amino acid residues and N1 migrates abnormally under reducing conditions on SDS PAGE. Clearly the N1 sequence is not a globular structure. However, it is not clear if N1 folds as one subdomain or if N1 is composed of several distinct domains.

Even though the different putative domains can fold independently it is clear that at least N2 and N3 interact if they are expressed together as can be concluded from the denaturation experiments. The recombinant isolated N3 domain denatures at $\sim$37 °C. However, temperature denaturation of the N23 tandem motif does not reveal any partial unfolding at this low temperature. Instead the N23 tandem motif remains intact up to $\sim$50 °C when it unfolds in apparently one step indicating that in the N23 tandem motif the structure of N3 is stabilized by the presence of N2. Furthermore, none of the putative subdomains are able to bind fibrinogen when expressed and analysed individually. However, the N23 tandem motif binds fibrinogen in a specific, concentration dependent and saturable manner with an apparent $K_D$ of $2 \times 10^{-8}$ M. Also the N12 tandem motif binds fibrinogen but with an apparent lower affinity. Taken together these results suggest that the protein segment required for forming a fibrinogen-binding site is not contained in an isolated subdomain but that two subdomains (N12 or N23) are needed to form an active protein. This does not imply that sequences from the two subdomains combine to form the actual binding region. It is possible that sequences within a specific subdomain are found in the binding region, but that the folding of the subdomain and the formation of an active site is influenced by neighbouring sequences. This possibility appears to be supported by the finding that both N12 and N23 bind fibrinogen and that proteolytically processed N23 with an exposed N-terminus at residue
Region A of ClfB is composed of three subdomains

A199 is unable to bind fibrinogen but when this protein construct contains a short unrelated N-terminal extension it acquires fibrinogen-binding activity. These observations also suggest that the N2 subdomain plays a key role in fibrinogen binding. In apparent support of this hypothesis we found that antibodies raised against the N2 subdomain were the most efficient in inhibiting the binding of ClfB-mediated adherence to fibrinogen although antibodies to both the N3 and the N1 subdomains also had some inhibitory activity.

The proteolytic processing of ClfB, which results in the removal of the N1 domain, occurs both on intact *S. aureus* cells (8) and during the isolation of recombinant A-region produced in *E. coli* (15). In both cases the protease is a metalloprotease and this processing can be largely prevented by addition of EDTA. The N1 segment does not appear to be released as an intact subdomain but a set of shorter peptides can be detected from the processing of the recombinant protein (unpublished observation). This observation supports the idea that N1 is not a tightly packed globular subdomain but contains many sites that are susceptible to proteolytic attack.

The biological implications of an MSCRAMM composed of several subdomains, which may be proteolytically processed, are unclear. A regulated processing of N1 could greatly affect the fibrinogen binding activity of the remaining N23 domains of ClfB. Thus, a recombinant N23 appears to bind fibrinogen with a higher affinity compared to that observed with the intact A-region. A partial removal of the N1 subdomain may lead to an activation of fibrinogen-binding activity whereas a complete removal of the N1 subdomain results in loss of fibrinogen-binding activity since cleaved N23 does not bind fibrinogen.

Does the proposed model of the subunit organization and processing of the ClfB A-region also apply to other structurally related proteins on *S. aureus*? We believe that the ClfA A-region is very similar to the ClfB A-region. The minimal fibrinogen-binding activity is located to a subregion composed of a N23 tandem motif. The putative N1 subdomain is
Region A of ClfB is composed of three subdomains hydrophilic and readily processed (7, 29). The N1-N2 interface contains the site SLAVA in both ClfB and ClfA, which can be targeted by metalloproteases. Analysis of the FnbpA and FnbpB A-regions both of which we recently showed to contain a fibrinogen-binding site, suggest that these have a similar organization. In fact, a N23 construct of FnbpA was shown to bind fibrinogen (5). Subdomains in FnbpA and FnbpB are hydrophilic and proteolytic processing of FnbpA has been reported (30). However a SLAVA motif is not present in either of these MSCRAMMs. Also in the CNA A-region we can identify three subdomains and the recombinant A-region behaves as a non-globular extended protein (3). However, the N1 subdomain is less hydrophilic compared to this domain in other members of this protein family and no partial processing has been observed. Furthermore the ligand (collagen) binding site is functional in an isolated N2 subdomain of CNA (31). Taken together these observations point to a common structural subdomain organization of the A-regions of these proteins.

The biophysical/biochemical analysis of the recombinant putative subdomains of ClfB supports the proposed multi-domain hypothesis for the A-regions. However, the ultimate proof of this hypothesis will come from a structural analysis of these proteins. We are currently involved in trying to solve the crystal structures of several MSCRAMMs. As additional structural information becomes available we will undoubtedly refine our understanding of the organization and mechanism of action of these important proteins.

ACKNOWLEDGEMENTS

This work was supported in part by NIH Grant number AI20624, The Wellcome Trust Grant number 052320, Inhibitex Inc., and The Wellcome Burroughs Fund Travel Grant.
Region A of ClfB is composed of three subdomains.

REFERENCES


Region A of ClfB is composed of three subdomains


Region A of ClfB is composed of three subdomains


Region A of ClfB is composed of three subdomains

FIGURE LEGENDS

Fig. 1. Schematic representation of the domains of ClfB. S, signal sequence; A, fibrinogen-binding region A; R, Serine-Aspartate repeat region; W, wall spanning domain; M, membrane anchor; C, cytoplasmic domain. The LPETG cell wall anchoring motif is indicated. The protease-sensitive SLAVA motif is indicated by an open box, the Proline rich region between the A and R regions is indicated by a hashed box, and the N-terminal His-tag is designated by the horizontal dash (8). Shown are the proposed subdomains (N1, N2, and N3) of the A-region. Recombinant proteins used in this study are indicated by the putative subdomain compositions and comprising residues. All constructed recombinant proteins contained a N-terminal His-tag as indicated.

Fig. 2. SDS-PAGE Analysis of recombinant 6xHis-ClfB region A (N123). SDS gel with N123 after purification (lane 1), after storage at 4 °C (lanes 3), and molecular weight standards (lane 2).

Fig. 3. Gel Permeation chromatography of standard and recombinant proteins.
Proteins were analyzed on a HR 10/30 column of Superose 12 eluted with PBS. Shown is a plot of partition coefficient (K_{av}) verses the logarithm of M_w for individual proteins. Recombinant ClfB proteins are represented as: (X) N123, (●) N12, (◆) N23, (○) N1, (△) N2, and (▽) N3. The elution position of standard proteins: BD, Blue Dextran 2000 (2,000,000 Da) (Amersham Pharmacia), bovine thyroglobulin (670,000 Da), bovine gamma globulin (158,000 Da), chicken ovalbumin (44,000 Da), horse myoglobin (17,000 Da), human vitamin B-12 (1,350 Da) (Bio-Rad Gel Filtration Standard kit), and Tyr, Tyrosine (181 Da) (Sigma) (■) are plotted along with their corresponding linear fit.
Region A of ClfB is composed of three subdomains.

Fig. 4. **Far-UV CD spectra of ClfB N123, N1, N2, and N3.** Shown are the CD spectra of the three individual domains, N1, N2, and N3, as well as the full-length N123 from 260nm to 185nm. The (X) N123 spectrum is dashed so it stands out from the individual domains, (●) N1, (▲) N2, and (▼) N3. Secondary structure predictions are presented in Table 3.

Fig. 5. **Temperature dependent folding and unfolding of recombinant ClfB proteins monitored by Far UV CD spectroscopy.** Changes in the secondary structure of recombinant subunits N1, N2, N3, N12, N23, and the full-length A-region (N123) were monitored by changes in CD at 205nm. The upper and lower rows show the one-step unfolding of the ClfB recombinant. The refolding of the individual subdomains N1, N2, and N3 is given in the central panels.

Fig. 6A. **Binding of biotinylated recombinant proteins to immobilized fibrinogen.** Recombinant ClfB constructs (●) N123, (□) N12 and (●) N23, were biotinylated and incubated in wells coated with human fibrinogen (100 ng/well). Fig 6B. Recombinant N23 constructs (●) N23, (○) N23i, and (X) N23 treated with concentrated staphylococcal culture supernatant (15) were also biotinylated and incubated in wells coated with human fibrinogen (100ng/well). Bound protein was detected by the addition of streptavidin-HRP followed by a chromogenic substrate. The plates were read at 450 nm. Background binding to the blocking agent (5% w/v BSA) was subtracted from the values obtained for the fibrinogen-coated wells.

Fig. 7. **Inhibition of bacterial adherence to immobilized fibrinogen by recombinant ClfB constructs.** Unlabelled recombinant ClfB proteins (●) N123, (◊) N12 and (●) N23, were preincubated in wells coated with fibrinogen (500 ng/well). Suspensions (~1 x 10⁸ colony-forming units) of *S. aureus* strain DU5874 (7) expressing ClfB (but not ClfA) were
Region A of ClfB is composed of three subdomains then added to the wells and bacterial adherence was measured using crystal violet staining. Values are the mean ± SD of triplicate from two experiments.

Fig. 8A. The titer of polyclonal IgG antibodies raised against the individual ClfB A-region subdomains N1, N2, and N3 and their ability to inhibit ClfB-mediated bacterial adherence to immobilized fibrinogen. Purified IgG antibodies isolated from antisera raised against (O) N1, (●) N2, and (▼) N3 were incubated with immobilized recombinant N123. Binding was detected using horseradish peroxidase-labeled goat anti-rabbit IgG and quantitated by absorbance at 450nm. Fig 8B. Polyclonal IgG antibodies raised against the individual domains (O) N1, (△) N2, and (▼) N3 as well as (●) anti-ClfA antibodies were preincubated in wells coated with fibrinogen (500 ng/well). Suspensions (~1 x 10^8 colony-forming units) of S. aureus strain DU5874 (7) expressing ClfB (but not ClfA) were then added to the wells and bacterial adherence was measured using crystal violet staining. Values are the mean ± SD of triplicate from two experiments.
Region A of ClfB is composed of three subdomains.

Table 1: Synthetic oligonucleotides primers used to amplify \textit{clfB} gene fragments.

<table>
<thead>
<tr>
<th>Expression constructs</th>
<th>Amino acids</th>
<th>Sequence$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>pN1 44-196</td>
<td>5' -</td>
<td>CGAGGTATCCTCACGGCAATCGAAGATACACG-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5' -</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AACGGTACACTACGACTAGCTCTCCTTTCTAAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ACT-3'</td>
</tr>
<tr>
<td>pN2 197-375</td>
<td>5' -</td>
<td>GCTGGATCCAGTTTAGCTGTTGCTGAAACC GGTA-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5' -</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-AATGGTACCACTATTTTGTTTTATCAATTTCTGC AAT-3'</td>
</tr>
<tr>
<td>pN3 376-542</td>
<td>5' -</td>
<td>AAAAGGTACCGCGCGAACAATTTTCTTCTCAAAATT-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5' -</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGGGGGTACCACCATTTTACTGCTGAAATCACCATCA GC-3'</td>
</tr>
<tr>
<td>pN12 44-196</td>
<td>5' -</td>
<td>-CGCGGTATCCTCACGAAGAATCGAAGATAC-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5' -</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-CCCAAGCTTAGAAATGTTCGCGCCATTTT-3'</td>
</tr>
<tr>
<td>pN23 197-542</td>
<td>5' -</td>
<td>-CGCGAGCTAGTTAGCTGTTGCTGAAAC-3'</td>
</tr>
</tbody>
</table>
Region A of ClfB is composed of three subdomains

pN23i 199-542

5'-'CCCAGCTTATTTACTGCTGAATCACC-3'

GCTGAGCTGCTGTTGCTGAACCGGTAGTAAAT-

3' 5'-CCCAGCTTATTTACTGCTGAATCACC-3'

pN123 44-542

5'-'CGAGGATCTTTCAGGACAACTGAACGATACAAACG-

3' 5'-'TGGGGTACCTTACTGCTGAATCACCACCATCA

GC-3'

The restriction endonuclease sites are underlined."
Region A of ClfB is composed of three subdomains

Table 2: Estimated Radii of ClfB Subunit Constructs.

<table>
<thead>
<tr>
<th>Species</th>
<th>Calculated Mw$^1$</th>
<th>EI-MS Mw$^1$</th>
<th>$R_s$ from Equation$^2$</th>
<th>$R_s$ from</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>18,611</td>
<td>18,426</td>
<td>20.9</td>
<td>29</td>
</tr>
<tr>
<td>N2</td>
<td>21,404</td>
<td>21,401</td>
<td>22.1</td>
<td>25</td>
</tr>
<tr>
<td>N3</td>
<td>21,725</td>
<td>21,714</td>
<td>22.2</td>
<td>21</td>
</tr>
<tr>
<td>N12</td>
<td>37,096</td>
<td>37,199</td>
<td>27</td>
<td>40</td>
</tr>
<tr>
<td>N23</td>
<td>40,184</td>
<td>40,213</td>
<td>27.9</td>
<td>26</td>
</tr>
<tr>
<td>N123</td>
<td>56,607</td>
<td>56,611</td>
<td>31.6</td>
<td>39.5</td>
</tr>
</tbody>
</table>

$^1$Molecular weight is given in Daltons and is calculated from Macvector (7.0) Oxford Molecular Ltd. $^2$R$_s$ (Stokes Radii) is given in Angstroms

Table 3: Summary of Secondary Structural Predictions in Percent and Melting Temperatures of ClfB Subdomains

<table>
<thead>
<tr>
<th>Species</th>
<th>Alpha Helix</th>
<th>Beta Sheet</th>
<th>Other</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>6±1</td>
<td>41±1</td>
<td>53±1</td>
<td>51.6±1.1</td>
</tr>
<tr>
<td>N2</td>
<td>0</td>
<td>49±1</td>
<td>51±1</td>
<td>50.4±0.5</td>
</tr>
<tr>
<td>N3</td>
<td>7±1</td>
<td>38±3</td>
<td>55±3</td>
<td>38.5±1.4</td>
</tr>
<tr>
<td>N12</td>
<td>7±2</td>
<td>28±4</td>
<td>65±4</td>
<td>51.0±0.2</td>
</tr>
<tr>
<td>N23</td>
<td>7±2</td>
<td>47±1</td>
<td>46±3</td>
<td>50.9±0.3</td>
</tr>
<tr>
<td>N123</td>
<td>3±2</td>
<td>37±2</td>
<td>60±4</td>
<td>49.1±0.2</td>
</tr>
</tbody>
</table>
Figure 1

- SLAVAE motif
- Proline rich region
- 6xHis Tag

N123  (rClfB44-542)
N12  (rClfB44-375)
N23  (rClfB197-542)
N23i (rClfB199-542)
N1   (rClfB44-196)
N2   (rClfB197-375)
N3   (rClfB375-542)
Figure 2
Figure 4
Figure 6 (A)
Figure 6 (B)

Absorbance (450 nm) vs. Protein Concentration (mM)
Figure 7

[Graph showing the effect of inhibitor concentration on binding to fibrinogen (%). The x-axis represents inhibitor concentration in uM, ranging from 0 to 100, and the y-axis represents binding to fibrinogen (%), ranging from 0 to 100. The graph includes data points and error bars for different inhibitor concentrations.]
Structural organization of the fibrinogen-binding region of the clumping factor B 
MSCRAMM of staphylococcus aureus
Samuel Perkins, Evelyn J. Walsh, Champion C.S. Deivanayagam, Sthanam V. L. Narayana, 
Timothy J. Foster and Magnus Höök

J. Biol. Chem. published online September 21, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M106741200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at 
http://www.jbc.org/content/early/2001/09/21/jbc.M106741200.citation.full.html#ref-list-1