Targeted Protein Engineering Provides Insights into Binding Mechanism and Affinities of Bacterial Collagen Adhesins

Caná L. Ross‡1, Xiaowen Liang‡1, Qing Liu‡, Barbara E. Murray§, Magnus Höök‡, and Vannakambadi K. Ganesh‡2
From the ‡Center for Inflammatory and Infectious Disease, Institute for Biosciences and Technology, Texas A&M Health Science Center, Houston, Texas 77030 and the §Division of Infectious Diseases, University of Texas Medical School, Houston, Texas 77030

The collagen-binding bacterial proteins, Ace and Cna, are well characterized on the biochemical and structural level. Despite overall structural similarity, recombinant forms of the Ace and Cna ligand-binding domains exhibit significantly different affinities and binding kinetics for collagen type I (CI) in vitro. In this study, we sought to understand, in submolecular detail, the bases for these differences. Using a structure-based approach, we engineered Cna and Ace variants by altering specific structural elements within the ligand-binding domains. Surface plasmon resonance-based binding analysis demonstrated that mutations that are predicted to alter the orientation of the Ace and Cna N1 and N2 subdomains significantly affect the interaction between the MSCRAMM (microbial surface components recognizing adhesive matrix molecule) and CI in vitro, including affinity, association/dissociation rates and binding ratio. Moreover, we utilized this information to engineer an Ace variant with an 11,000-fold higher CI affinity than the parent protein. Finally, we noted that several engineered proteins that exhibited a weak interaction with CI recognized more sites on CI, suggesting an inverse correlation between affinity and specificity.

Bacterial colonization of host tissues is a critical step in the infection process, and many pathogenic bacteria produce proteins that promote bacterial adhesion. One such family of proteins, the MSCRAMMs3 (microbial surface components recognizing adhesive matrix molecules), interacts with components of the host extracellular matrices (1–3). MSCRAMMs found on Gram-positive bacteria typically belong to a family of cell wall anchored proteins that have similar structural organizations. In general, the N-terminal portion of these proteins contains the ligand-binding domain in the so-called A region, whereas the C-terminal half contains repeated motifs often known as the B domains that appear to extend and project the ligand-binding domain away from the bacterial surface (4). The C terminus also contains the cell wall sorting motifs, including the LPXTG sequence, which is essential for sortase-dependent anchoring to the cell wall, followed by a stretch of hydrophobic residues and a short segment of positively charged amino acids (5, 6).

Despite a conserved structural organization, MSCRAMMs are largely species-specific, and homologues with substantial sequence similarities within the ligand-binding domains are not found in other bacteria. One exception is a family of collagen-binding MSCRAMMs, which includes Cna from Staphylococcus aureus (7), Ace from Enterococcus faecalis (8), Acn from Enterococcus faecium (9), Acm from Streptococcus equi (10), Cnm from Streptococcus mutans (11), RspA and RspB from Erysipelothrix rhusiopathiae (12), and BA0871 and BA5258 from Bacillus anthracis (13).

Cna and Ace are the most extensively studied members of this family of collagen-binding MSCRAMMs. The Ace and Cna ligand-binding regions are composed of N1 and N2 subdomains that each adopt an IgG fold variant (14–16). Structural comparisons of Cna31–344 in the apo form and in complex with a synthetic collagen peptide led to the proposal of a multistep ligand-binding mechanism called the Collagen Hug (15). In this model, ligand binding is initiated by a low affinity interaction between the ligand and residues present within a shallow “trench” located on the N2 subdomain. Next, the interdomain linker wraps around the triple helical collagen and repositions the N1 subdomain. Finally, the complex is stabilized when the C-terminal N2 extension, acting as a “latch,” inserts into a cleft and complements a β-sheet within the neighboring N2 subdomain (15). In the resulting complex, the MSCRAMM forms an interdomain hole through which the collagen projects. Structural comparisons and mutagenesis studies indicate that Ace employs a similar ligand-binding mechanism (16).

The crystal structures of Cna31–344 and Ace12–367 indicate that the hole (consisting of the interdomain linker and the N1
and N₂ subdomains) created by ligand hugging can only accommodate a single collagen triple helix. Therefore, it is likely that Cna and Ace interact with monomeric triple helix collagen but not with fibrillar collagen (15, 16). Furthermore, Cna₃₁₋₃₄₄ does not recognize gelatin or denatured collagen, indicating that the rod-like shape of the triple helix is critical for binding (15).

Expression of Ace and Cna on the surface of *E. faecalis* and *S. aureus*, respectively, results in increased bacterial adherence to immobilized collagen type I (CI) (8, 17–19). In addition, surface expression of Ace mediates *E. faecalis* adherence to collagen type IV (CIV), laminin, and dentin (20, 21), whereas the presence of the *cna* gene has been associated with adherence of *S. aureus* clinical isolates to laminin and CIV (22). Furthermore, several studies demonstrate that Ace and Cna are virulence factors in different animal models. *S. aureus* strains harboring the *cna* gene exhibit increased virulence in animal models of staphylococcal infection, including endocarditis, arthritis, osteomyelitis, mastitis, and keratitis (18, 23–26). It has also been demonstrated that ace-null strains of *E. faecalis* are attenuated in a murine model of urinary tract infection and a rat endocarditis model (27). In addition, recombinant forms of Cna can be used as effective vaccine components, whereas passive immunization with Cna antibodies is protective in a *cna* knockout model (28). All together, these studies demonstrate that Ace and Cna are virulence factors in different animal models. *S. aureus* strains harboring the *cna* gene exhibit increased virulence in animal models of staphylococcal infection, including endocarditis, arthritis, osteomyelitis, mastitis, and keratitis (18, 23–26). It has also been demonstrated that ace-null strains of *E. faecalis* are attenuated in a murine model of urinary tract infection and a rat endocarditis model (27).

Despite the high degree of structural similarity between Ace and Cna (Protein Data Bank (PDB) id: 37Z1P) created by ligand hugging can only accommodate a single collagen triple helix. Therefore, it is likely that Cna and Ace interact with monomeric triple helix collagen but not with fibrillar collagen (15, 16). Furthermore, Cna₃₁₋₃₄₄ does not recognize gelatin or denatured collagen, indicating that the rod-like shape of the triple helix is critical for binding (15).

**DNA Manipulation and Plasmid Construction**—Sequence changes to the DNA regions encoding the interdomain linker and the C-D loop of ace and cna were introduced using site-directed mutagenesis (29) or overlap extension PCR (30) using primers listed in supplemental Table S1 in a mixture containing Phusion polymerase (Thermo Scientific, Vantaa, Finland), HF buffer (Thermo Scientific), 50 ng of DNA template, and 200 – 250 nm dNTP mix (Invitrogen). Following digestion with DpnI (New England Biolabs, Ipswich, MA), PCRs generated from site-directed mutagenesis were transformed into *E. coli* TOP10 using standard procedures. Purified overlap PCRs were incubated at 72 °C with *Taq* polymerase (New England Biolabs) and 200 nm dATP (Invitrogen) and purified using DNA clean and concentrator 5 (Zymo Research, Carlsbad, CA) prior to ligation into pGEMTeasy (Promega, Madison, WI).

**Protein Purification**—The His-tagged recombinant proteins were purified with a 5-ml nickel-charged HiTrap chelating column (GE Healthcare, Uppsala, Sweden) and a 5-ml anion exchange Sepharose column (GE Healthcare) as described by Barbu et al. (31). Fractions obtained from the anion exchange column were concentrated and dialyzed against PBS, 10 mM EDTA, pH 7.4. Recombinant proteins were >95% pure. The concentration of the recombinant protein was determined by absorbance measured at 280 nm and calculated using the extinction coefficient for each protein.

**Molecular Modeling**—Molecular modeling studies were performed using Coot (32) and Insight II (Accelrys, Inc., San Diego, CA). To build a model of Ace in the Cna conformation, individual N₁ and N₂ domains of Ace (Protein Data Bank (PDB) id: 2Z1P) were superimposed on corresponding N₁ and N₂ domains of the crystal structures of Cna (PDB id: 2F68). Superposition based on secondary structure matching (33) available in Coot (32) was used for the superposition. One hundred and eighteen Ca atoms in the N₁ subdomain and 137 Ca atoms in the N₂ subdomain aligned with root mean square deviations of 1.67 and 1.50 Å, respectively. The model of the Ace molecule in the Cna conformation showed steric clashes between the N₁ and N₂ subdomains. To build a molecular model of the AceC₄ harboring the Cna N₁ region (C-D loop, D-strand, and D-D loop), a homology model was built using the HOMOLOGY module in the Insight II package (Accelrys, Inc.). After building a model of the Ace N₁ domain with the Cna N₁ region, this domain was superimposed on the Ace with Cna conformation to build a model of AceC₄.

**SPR Analysis**—The interactions of Cna₃₁₋₃₄₄, Ace₃₂₋₃₆₇, and the mutant derivatives with immobilized rat tail CI (R&D Systems, Inc., Minneapolis, MN) were characterized using a Bia-Sensor.
Insights on Collagen-hugging MSCRAMMs

core 3000 (GE Healthcare/Biacore) at 25 °C. Sensor chip C1 and amine-coupling kits were obtained from the same company and used to covalently attach CI onto the sensor surface using the amine coupling procedure as recommended by the manufacturer. The matrix-free C1 chip was chosen because it can achieve low density immobilization and avoid mass transfer limitation. Also, the flat surface allows interactions to take place closer to the surface, which is beneficial when working with multivalent interaction partners and large molecules. Briefly, after the surface was activated for 2 min, 10 μl of 5 μg/ml CI solution (stock of 5 mg/ml in 0.1 N HCl was diluted in 10 mM sodium acetate, pH 5.5) was injected into the flow cell at a flow rate of 5 μl/min. Approximately 440 response units (RU) of collagen were immobilized. Lower (∼150 and 180 RU) density collagen surfaces were also prepared by adjusting the activation time and volume of CI injected. A Cna surface was prepared on a separate sensor chip using 5 μg/ml Cna31–344 in 10 mM sodium acetate buffer, pH 5.5. The reference flow cells were prepared with activation and deactivation steps where no protein was coupled. Phosphate-buffered saline (8 mM Na2HPO4, 1.5 mM KH2PO4, pH 7.4, 2.7 mM KCl, 137 mM NaCl, and 0.005% Tween 20) was used as running buffer. Binding was performed at a relatively high flow rate of 30 μl/min, although no mass transfer was observed in the system. To regenerate the sensor surface, bound protein was removed by flowing 10 mM glycine (pH 1.5) over the surface for 1 min.

Base line-corrected SPR response curves (with buffer blank run further subtracted) were used for affinity determination. For steady-state analysis, equilibrium response (R eq) of each injection was collected and plotted against the concentration (A) of injected protein. A binding isotherm was fitted to the data using equation 1 (GraphPad Prism 4, GraphPad Software, Inc., La Jolla, CA) to obtain the equilibrium dissociation constant (K_D) and binding response maximum (R max).

\[ R_{eq} = R_{max} \times A/(K_0 + A) \]  
(Eq. 1)

Nonequilibrium data were globally fit to a predefined two-state model using the BIAevaluation software (Version 4.1). The R max association and dissociation rate constants (k a1, k d1) for the binding state (A + B ↔ AB), and forward and backward rate constants (k a2, k d2) for the conformational change state (AB ↔ AB*) were obtained from the fitting with T-value ≥100 as criteria for acceptable fit. This model describes 1:1 binding of analyte to immobilized ligand followed by a conformational change in the complex (34). It simply assumes that the only way AB* can dissociate to release free A is through prior conversion to the form AB, where A is Cna or Ace protein in solution, B is immobilized collagen, AB is the complex formed by binding A to B, and AB* is the complex formed by conformational change from AB. The apparent dissociation constant (K_D ^ app) was calculated from the kinetic parameters obtained from the curve fitting.

\[ K_D ^{app} = 1/((k_{a1}/k_{d1}) \times (1 + k_{a2}/k_{d2})) \]  
(Eq. 2)

Theoretical maximum response for 1:1 binding (R max) was calculated using Equation 3, where R immob is the SPR response of the protein immobilized on the sensor surface and MW sol and MW immob are molecular weight of soluble and immobilized protein, respectively. Molecular weights for Ace and Cna are ∼39,000 and ∼36,000, respectively. Because the collagen molecule was kept at very low concentration in a low salt and low pH buffer during immobilization, the monomer (∼300 kDa) conformation was maintained.

\[ R_{max}^{1:1} = (MW_{sol}/MW_{immob}) \times R_{immob} \]  
(Eq. 3)

Molar binding ratio (N) is defined as the number of soluble proteins bound to a single immobilized molecule and is determined by comparing the experimental value R max with R max.

\[ N = R_{max}/R_{max}^{1:1} \]  
(Eq. 4)

RESULTS AND DISCUSSION

Comparisons of the Cna31–344 and Ace32–367 Structures and Interactions with Collagen—Superpositions of apo crystal structures of Cna31–344 and Ace32–367 are shown in Fig. 1A. Structural comparisons of the two structures revealed that the interdomain orientations of the Ace32–367 (PDB id: 2Z1P) N1 and N2 subdomains are altered by about 36° when compared with that of Cna31–344 (PDB id: 2F68) (16) with the interdomain linker residues disordered in both the apo structures. As a result, the apo-Ace32–367 structure is in a partially open conformation with the latch partially closed, whereas Cna31–344 exhibits a latch-closed conformation (16). Consistent with previous studies, Ace32–367 and Cna31–344 exhibit different binding affinities and kinetics in SPR analysis of MSCRAMM binding to immobilized CI (Table 1, Fig. 1B). (8, 13, 15, 16). Cna31–344 has a higher affinity for collagen (K_D ^ app ~66.0 nM) than Ace32–367 (K_D ^ app ~43.2 μM). Also, the association and complete dissociation of Ace32–367 with immobilized CI are rapid, and the reaction follows a one-step steady-state binding model (Fig. 1, B and C, Table 1). However, the interaction of CI with Cna31–344 is more complex, consistent with a two-state conformational change model (Fig. 1, B and C, Table 1).

This simplified two-state reaction model involves two steps: 1) the formation of the CI-Cna31–344 complex (k a1, k d1) and 2) a conformational change that may occur after formation of the complex (k a2 and k d2) (Fig. 1C, Table 1). The interaction of Cna31–344 with collagen involves several steps. We propose that after the initial interaction between collagen and the N2 subdomain of a Cna31–344 molecule in the closed form, additional conformational changes occur, including opening of the N1-N2 interdomain and reorientation of the N3 subdomain, which allows the interdomain linker to wrap around the collagen, and the latching event that stabilizes the collagen-Cna31–344 complex (supplemental Fig. S1). Although we believe that this model is most likely, we cannot exclude a mechanism where the open form and closed form of Cna31–344 are in an equilibrium and only the open form is capable of binding the collagen molecule (15).

In an effort to understand the molecular bases for these biochemical differences, we compared the structures of the Ace and Cna ligand-binding domains and found significant variation in the interdomain linkers and interdomain orientation (Fig. 1A). The interdomain hole of Ace32–367, formed by the interdomain linker and the N1-N2 subdomains, is more spa-
cious than that of Cna31–344. Notably, structure-based sequence alignment showed that the Ace32–367 interdomain linker is 40% longer than that of Cna31–344 and likely contributes to the larger interdomain space (Fig. 1D). In addition, the crystal structure of Cna31–344 in complex with a collagen peptide shows the interdomain linker tightly wrapped around the ligand, suggesting that 8 amino acids is the optimal interdomain linker length. Therefore, it is possible that the Ace32–367 interdomain linker may be longer than required for monomeric collagen binding. Furthermore, the sequences of the Ace and Cna interdomain linkers share little sequence homology (Fig. 1D). Considering that structural data demonstrate specific contacts between the Cna31–344 interdomain linker and residues within a synthetic collagen peptide, it is possible that sequence differences within this structural element could influence ligand binding (15).

Structural comparisons also revealed that the interdomain orientations of the Ace32–367 N1 and N2 subdomains are altered by about 36° when compared with that of Cna31–344 (16). As a result, the apo-Ace structure is in a partially open conformation with the latch partially closed (16). Although it is possible that this altered orientation represents apo-Ace32–367 in equilibrium between the open and closed conformation (16), molecular modeling predicts that Ace32–367 cannot adopt the interdomain orientation exhibited by Cna31–344. Specifically, the C-D loop at the N terminus of strand D located within the Ace32–367 N1 subdomain is longer than the corresponding C-D loop of Cna31–344 (Fig. 1D). Considering that structural data demonstrate specific contacts between the Cna31–344 interdomain linker and residues within a synthetic collagen peptide, it is possible that sequence differences within this structural element could influence ligand binding (15).

### TABLE 1
Kinetic parameters for MSCRAMM/CI interaction

| Protein | $k_1$ | $k_2$ | $k_3$ | $k_4$ | $K_{D,app}$ | $K_D$ | $N$ |
|---------|-------|-------|-------|-------|-------------|-------|-----|
| Cna31−344 | $7.33(0.47)$ | $1.51(0.14)$ | $3.28(0.07)$ | $1.52(0.08)$ | $66.0(12.0)$ | $16.0(0.4)$ | $11.7(0.4)$ |
| Cna31−344 | $0.30(0.04)$ | $3.86(0.05)$ | $14.5(0.1)$ | $0.73(0.01)$ | $625(115)$ | $5.0(0.4)$ | $12.8(1.1)$ |
| Ace32−367 | $7.33(0.47)$ | $1.51(0.14)$ | $3.28(0.07)$ | $1.52(0.08)$ | $66.0(12.0)$ | $16.0(0.4)$ | $11.7(0.4)$ |
| Ace32−367 | $0.30(0.04)$ | $3.86(0.05)$ | $14.5(0.1)$ | $0.73(0.01)$ | $625(115)$ | $5.0(0.4)$ | $12.8(1.1)$ |
| Cna31−344 | $11.6(1.5)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ |
| Cna31−344 | $37.7(2.0)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ |
| Cna31−344 | $26.2(10.5)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ |
| Cna31−344 | $43.2(8.6)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ |
| Cna31−344 | $29.0(3.0)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ |
| Cna31−344 | $30.1(0.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ |
| Cna31−344 | $63.0(3.0)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ |
| Cna31−344 | $51.7(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ |
| Cna31−344 | $36.8(0.4)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ | $15.5(1.8)$ |
| Cna31−344 | $34.4(3.2)$ | $0.93(0.05)$ | $6.68(0.30)$ | $0.62(0.16)$ | $2.2(0.3)$ | $5.0(0.3)$ | $5.0(0.3)$ |
| Cna31−344 | $1.52(0.15)$ | $1.65(0.17)$ | $7.34(0.49)$ | $0.28(0.01)$ | $41.5(12.5)$ | $5.6(0.3)$ | $5.6(0.3)$ |
lizes a more open conformation than Cna\textsubscript{31–344} to avoid an apparent steric conflict between the two subdomains.

The altered interdomain orientation of the Ace\textsubscript{32–367} N\textsubscript{1}–N\textsubscript{2} subdomains and/or the more spacious interdomain hole of Ace\textsubscript{32–367} could result in a loose or “slippery” binding region, which might contribute to the lower binding affinity and less stable binding, reflected by a faster dissociation when compared with Cna\textsubscript{31–344} (Fig. 1B). To understand how these structural differences contribute to the observed differences in ligand binding, we evaluated the binding kinetics of several recombinant Ace and Cna proteins containing mutations that are predicted to alter the possible orientation of the N\textsubscript{1} and N\textsubscript{2} subdomains as well as the sequence and length of the interdomain linkers (Fig. 3).

**Linker Length Affects Binding to Type I Collagen**—To address the contribution of interdomain linker length to collagen binding, we engineered Cna and Ace proteins containing elongated and/or truncated interdomain linkers (Fig. 3) and compared the affinity and binding kinetics to immobilized CI using SPR (Figs. 4 and 5). We determined that the interdomain linker length influenced multiple aspects of \textit{in vitro} CI binding, including affinity and binding ratio (number of MSCRAMMs bound to a single collagen molecule).

Cna proteins containing interdomain linkers harboring insertions or deletions demonstrated significantly lower CI binding affinities than Cna\textsubscript{31–344}, indicating that the length of the Cna interdomain linker is optimal for CI affinity \textit{in vitro} (Fig. 4, A–C, Table 1). The addition of 3 amino acids to the Cna interdomain linker resulted in a protein (Cna\textsuperscript{∧QIE}) with an \sim 40-fold decrease in binding affinity (Table 1, Fig. 4B), whereas deleting three residues from the interdomain linker yielded a protein (Cna\textsuperscript{∧EAG}) with an \sim 9-fold decrease in affinity for CI (Table 1, Fig. 4C). The length of the interdomain linker also influenced the number of sites in CI recognized by the MSCRAMM; Cna\textsuperscript{∧QIE} showed an increased binding ratio of 16, and Cna\textsuperscript{∧EAG} exhibited a decreased binding ratio of 5. A binding ratio of 13 was calculated for Cna\textsubscript{31–344} (Table 1). Taken together, these data demonstrate that the native Cna interdomain linker length is optimal for CI binding \textit{in vitro} as revealed by affinity calculations. However, a longer interdomain linker

---

**FIGURE 2. Implications of the Ace\textsubscript{32–367} interdomain orientation.** A, structure of Ace\textsubscript{32–367}. The open conformation exhibited by Ace\textsubscript{32–367} allows the A-B loop within the N\textsubscript{2} subdomain (yellow ribbon) to avoid steric conflicts with the C-D loop of the N\textsubscript{1} subdomain (green ribbon). For clarity, the shorter Cna\textsubscript{31–344} C-D loop (red) is superimposed on the Ace N\textsubscript{1} subdomain. B, structural model of Ace\textsubscript{32–367} in the Cna\textsubscript{31–344} conformation. The long C-D loop within the N\textsubscript{2} subdomain (green ribbon) makes severe clashes with the A-B loop of the N\textsubscript{1} subdomain (yellow ribbon), suggesting that Ace\textsubscript{32–367} cannot adopt a Cna-like conformation.

**FIGURE 3.** Description of parent and mutant recombinant Ace and Cna proteins. A, crystal structure of Ace\textsubscript{32–367} demonstrating the location of interdomain linker (blue dotted line) and C-D loop (red) within the N\textsubscript{2} subdomain where sequences were altered. B, list describing the parent and mutant recombinant proteins and corresponding sequences of the interdomain linker (blue) and C-D loop region (red). The Cna sequences are in uppercase, and the Ace sequences are lowercase and \textbf{bolded}. C, diagram of N\textsubscript{1} subdomain and the latching segment following the N\textsubscript{2} subdomain (yellow) with labeled \textit{β}-strands. The segment of the N\textsubscript{1} subdomain that is altered in the protein chimeras is shown in red.
may allow the protein to recognize a greater number of binding sites within the collagen molecule (Table 1).

Removing residues from the Ace interdomain linker yielded proteins with CI affinities that were only marginally higher than the parent protein, Ace32–366 (\(K_D \approx 43.2 \mu M\)); Ace\(^{\Delta QIE}\) exhibited a \(K_D\) of \(\approx 29.0 \mu M\), and Ace\(^{\Delta EFGQIE}\) exhibited a \(K_D\) of \(\approx 30.1 \mu M\) (Fig. 5, A–C, Table 1). However, deleting residues within the Ace interdomain linker negatively affected the number of sites

---

**FIGURE 4.** Biacore analysis of the interactions between CI and the recombinant Cna proteins. 2-fold increasing concentrations of Cna proteins were injected over the immobilized CI (180 RU for A, D, and E; and 400 RU for B, C, F, G, and H). Sensorgrams are shown in black with lower curve corresponding to lower concentration of Cna protein injected. Kinetic analysis for A, C, D, and E (the fitted curves are shown in red) was performed to obtain the rate constants (shown in Table 1) \(K_D^{\text{app}}\) and \(R_{\text{max}}\). For steady-state interactions (B, F, G, and H), a binding isotherm was created (inset) to determine equilibrium \(K_D\) and \(R_{\text{max}}\). Drawings in each graph indicate the general location of the mutations. The presence of the mutations within the interdomain linker is represented as a filled arc, and the mutations affecting the interdomain orientation are represented as a filled circle.
recognized on the collagen molecule. Although the parent protein exhibited a binding ratio (N) of 14, binding ratios of 11 and 10 were calculated for Ace\textsuperscript{ΔQIE} and Ace\textsuperscript{ΔETGQIE}, respectively (Table 1). These results indicate that decreasing the length of the Ace interdomain linker does not strongly influence affinity of Ace for CI, but as with Cna, the longer interdomain linker may allow the protein to recognize a larger number of binding sites within CI.
Because of the repetitive nature and linear structure of the CI molecule, the multiple MSCRAMM-binding sites observed in SPR likely represent independent binding sites. However, to confirm that the data are not due to aggregation of the injected proteins, we examined binding of solubilized CI to immobilized Cna31–344 (supplemental Fig. S2). The resulting binding ratio of 1:11 demonstrates that one CI molecule was recognized by ~11 Cna31–344 molecules on the sensor surface, which is close to the N of 11.7 observed in experiments utilizing soluble Cna31–344 and immobilized CI (Table 1). Differences in binding ratio are expected because immobilization is random. Therefore, a portion of the immobilized Cna31–344 will not be able to interact with the CI molecule. Furthermore, the extremely slow off-rate observed in the SPR sensograms in supplemental Fig. S2 is due to the effects of valency, which occurs when a single CI molecule is anchored by multiple immobilized Cna31–344.

**Linker Sequence Affects Affinity and Binding Site Recognition for Type I Collagen—** The Ace and Cna interdomain linker sequences share little homology, and structural analysis of the MSCRAMM-ligand complex revealed that amino acids within the Cna interdomain linker make specific contacts with the collagen peptide (15). Specifically, Val-172, located near the C terminus of the interdomain linker, interacts with a proline residue from the leading chain of the collagen peptide. This interaction may facilitate repositioning of the N1 subdomain, which is an important step in the ligand binding process of Cna (15). Currently, a structure for the Ace ligand-binding domain in complex with collagen has not been resolved. Therefore, specific residues within the interdomain linker that make contact with the collagen ligand are unknown. However, based on position within the Ace interdomain linker, it is likely that the C-terminal residues Tyr-176 and/or Pro-177 play a role similar to that of Val-172.

To understand how the sequence of the interdomain linker contributes to collagen binding, we replaced the 8-amino acid interdomain linker from Cna31–344 with 1) the full-length, 14-amino acid Ace32–367 linker, 2) a 12-amino acid Ace linker lacking two N-terminal residues, and 3) a 10-amino acid Ace linker lacking two N- and two C-terminal residues, creating chimeric proteins CnaA1, CnaA2, and CnaA3, respectively (Fig. 3). Similarly, the Ace32–367 interdomain linker was replaced with 1) the 8-amino acid Cna linker, 2) a 10-amino acid sequence containing the Cna interdomain linker plus two C-terminal Ace linker residues, and 3) a 12-amino acid sequence containing the Cna interdomain linker flanked by 2 amino acids each from the N and C termini of the Ace linker to generate the chimeric proteins AceC1, AceC2, and AceC3, respectively (Fig. 3). We found that changes to the length and sequence of the interdomain linkers impacted the binding kinetics of the proteins and underscored the importance of specific residues within the interdomain linker sequence.

SPR analysis revealed that the sequence of the interdomain linker influenced the CI affinity and binding ratio. CnaA1, which contains the full-length Ace interdomain linker, exhibited a higher affinity for CI ($K_D^{app}$ ~ 19.0 nM) than the native Cna31–344, indicating that the sequence of the longer Ace interdomain linker can facilitate high affinity binding to CI in vitro (Fig. 4, A and D, Table 1). Additionally, removing the two N-terminal Ace interdomain linker residues, as in CnaA2, yielded a protein with an affinity ($K_D^{app}$ ~ 88.5 nM) similar to that of Cna31–344 (Fig. 4, A and E, Table 1), whereas CnaA3, which is lacking the N- and C-terminal Ace interdomain linker residues, exhibited a much lower affinity for CI ($K_D$ ~ 11.6 μM) (Fig. 4, A and F, Table 1). The binding ratio for CnaA3 was the highest (N ~ 16), whereas the binding ratios of CnaA1 (N ~ 13) and CnaA2 (N ~ 10) were more similar to Cna31–344 (N ~ 12) (Table 1).

We determined that changes to the length and sequence of the Ace interdomain linkers resulted in modest changes in CI affinity (Table 1, Fig. 5, A, D, and E). However, the numbers of CI-binding sites were significantly altered (Table 1). Specifically, Ace32–367 and AceC3, which harbor interdomain linkers containing the N-terminal residues, Val-164 and Thr-165, and the C-terminal residues, Tyr-176 and Pro-177, exhibited binding ratios of 14. Proteins containing interdomain linkers lacking the N- and C-terminal residues, AceC1 (lacks Ace residues Val-164, Thr-165, Tyr-176, and Pro-177) and AceC2 (lacks Tyr-176 and Pro-177), exhibited binding ratios ranging from 4 to 5.

Taken together, the SPR data obtained from the engineered Cna and Ace proteins harboring interdomain linkers with altered sequences and lengths suggest that the N- and C-terminal residues of the Ace interdomain linker positively influence binding to CI in vitro. Moreover, it is possible that the C-terminal residues, YP, present within the Ace interdomain linker sequence stack better with the CI proline residues than the corresponding SV amino acids from the Cna interdomain linker (Fig. 3). As a result, proteins containing the YP residues may be able to anchor the collagen molecule within the N2 domain, suggesting that the ligand affinity is affected more by the sequence than by the length of the interdomain linker.

The Interdomain Orientation Affects Affinity and Binding Kinetics for Type I Collagen—Changes to the Ace interdomain sequence and length did not result in recombinant proteins that demonstrate stable interactions with immobilized CI, as was observed for the engineered Cna proteins, indicating that another aspect of the Ace32–367 structure is responsible for the low affinity and fast on- and off-rate. Analysis of the structure of apo-Ace32–367 reveals a protein with a partially open conformation. The Ace32–367 N1 subdomain contains a short D-D loop, a short D strand, and a large C-D loop, whereas the Cna31–344 N1 subdomain contains a large D-D’ loop, a longer D strand, and a short C-D loop (Figs. 3 and 2C). The larger C-D loop likely precludes the Ace32–367 protein from adopting a conformation that is optimal for ligand binding, thus preventing a tight interaction with collagen (Fig. 2B). In addition, the larger D-D’ loop found in Cna31–344 forms part of the latching trench and may yield a stronger latching event than the shorter Ace32–367 D-D’ loop. The open conformation and/or weaker latching event may explain why Ace32–367 has a lower affinity and more rapid off-rate for immobilized CI when compared with Cna31–344.

To understand how structural regions that do not interact with the ligand can influence orientation of the N1 and N2 subdomains, we engineered a protein that was predicted to adopt a more closed conformation than Ace32–367. We replaced the amino acid sequence spanning the Ace32–367 C-D loop, D-strand, and D-D’ loop with the corresponding residues from Cna, creating AceC4 (Fig. 3). In addition, we replaced the
Insights on Collagen-hugging MSCRAMMs

Cna\textsubscript{31–344} C-D loop, D-strand, and D-D’ loop with the corresponding sequence from Ace to create CnaA4 (Fig. 3), which should adopt a conformation that is more open than that exhibited by Cna\textsubscript{31–344}.

SPR analysis demonstrated that changes predicted to influence the N\textsubscript{1} and N\textsubscript{2} interdomain orientation affect several aspects of the MSCRAMM/CI interaction. The CI affinity of AceC4 was \(~11,000\)-fold higher than Ace\textsubscript{32–367} (Fig. 5, A and G, Table 1), whereas CnaA4 exhibited an \(~25\)-fold decrease in affinity for CI when compared with Cna\textsubscript{31–344} (Fig. 4, A and F, Table 1). These results suggest that a more closed conformation allows a tighter interaction with CI.

The partially open conformation observed for apo-Ace\textsubscript{32–367} may explain why the purified Ace proteins containing the shortest interdomain linkers (AceC1 and Ace\textsuperscript{ATEGQIE}) exhibited the lowest binding ratios for CI. In an effort to assess the relationship between the interdomain linker and the interdomain orientation, we created AceC1A4, which contains the shorter Cna interdomain linker and the shorter Cna C-D loop and CnaA1A4 that harbors the full-length Ace interdomain linker and the larger Ace C-D loop (Fig. 4). AceC1C4 showed a 600-fold increase in CI affinity when compared with AceC1, which harbors the Cna linker but not the smaller C-D loop (Fig. 5, A and F, Table 1). However, CnaA1A4 (\(K_D\) \(~26.2\) \(\mu M\)) contains the Ace interdomain linker sequence and the larger C-D loop, showed an affinity for CI that was comparable with CnaA4 (\(K_D\) \(~37.7\) \(\mu M\)). Together, these data suggest that removing the apparent steric hindrance from the Ace N\textsubscript{1} and N\textsubscript{2} subdomains allows the Ace protein to accommodate a shorter interdomain linker.

We also found that mutations to the C-D loop region, which would likely influence the interdomain orientation of Ace\textsubscript{32–367} and Cna\textsubscript{31–344} proteins, significantly altered the overall binding kinetics of the engineered proteins, resulting in an “Ace-like” Cna and a “Cna-like” Ace. The Cna proteins (CnaA4 and CnaA1A4) containing the larger C-D loop exhibited similar affinities and binding profiles (a rapid association and complete dissociation) to CI as Ace\textsubscript{32–367} (Fig. 4, A, G and H, Table 1). On the other hand, SPR analysis of the AceC4 and AceC1C4 interaction with immobilized CI revealed kinetic data that were more similar to results obtained with Cna\textsubscript{31–344} (Figs. 4A and 5, G and H, Table 1). This is most evident when comparing the dissociation of the purified proteins. Dissociation from CI by AceC4 and AceC1C4 is slow and incomplete (Fig. 5, G and H, Table 1) and suggests an interaction with CI that is more complex than the parent protein Ace\textsubscript{32–367} and may involve a conformational change.

Altering the C-D loop also affected the number of CI-binding sites recognized by the proteins. The binding ratios (\(N\)) for the purified Ace proteins containing the smaller C-D loop were significantly lower (AceC4 \(~5\), AceC1C4 \(~4\)) than the parent protein (Ace\textsubscript{32–367} \(~14\)), whereas the binding ratios for CnaA4 (\(~15.5\)) and CnaA1A4 (\(~13.6\)) were higher than the binding ratio for Cna\textsubscript{31–344} (\(~11.7\)) (Table 1), suggesting that a more open conformation may allow the protein to recognize a greater number of binding sites within the CI molecule (Table 1).

Our data demonstrate that the lower affinity and unstable binding kinetics for immobilized CI exhibited by Ace\textsubscript{32–367} are in large part due to the presence of a longer loop that connects the C and D strands of the Ace\textsubscript{32–367} N\textsubscript{1} subdomain, which likely results in an altered orientation of the N\textsubscript{1} and N\textsubscript{2} subdomains when compared with Cna\textsubscript{31–344}. Furthermore, these results indicate that the longer Ace interdomain linker may be necessary to accommodate the altered interdomain orientation observed in the apo crystal structure of Ace (16). Surprisingly, the SPR data revealed that a more open conformation may allow the proteins to recognize a greater number of binding sites on the CI molecule as all proteins containing the Ace N\textsubscript{1} sequence, and presumably a more open conformation, exhibited higher binding ratios.

Concluding Remarks—Our studies point to an inverse correlation between affinity and binding site specificity. However, it is not clear whether a high affinity interaction or the ability to recognize multiple binding sites would be more advantageous for mediating bacterial attachment to a collagenous surface. A high affinity interaction between a surface-anchored protein and a host ligand would facilitate attachment and proliferation. On the other hand, recognition of a greater number of binding sites would result in increased avidity, and disassociation at low affinity motifs by a fraction of the MSCRAMMs may not result in detachment of the bacterium from the substrate. We are currently working to understand how alterations within the ligand-binding domains of Ace and Cna affect adherence of \textit{E. faecalis} and \textit{S. aureus} to CI, CIV, and \textit{L. in vitro}. We are also examining how the presence of the ligand-binding domains affects virulence in different animal models. These studies may also yield insight into how affinity and binding site specificity affect the host/pathogen interaction.

This study demonstrates that the MSCRAMM/CI interaction is complex as several structural features influence ligand binding affinity, specificity, and kinetics. For the Cna-like MSCRAMMs, the interdomain orientation appears to be most important followed by the interdomain linker sequence and finally interdomain linker length. Our data indicate that although the regions within Ace\textsubscript{32–367} that are predicted to interact with collagen should facilitate a tighter interaction, the more open orientation and larger interdomain hole adopted by Ace\textsubscript{32–367} are not optimal for high affinity CI binding. This structural variation may allow Ace to bind additional rod-shaped ligands with a relatively larger diameter or recognize different collagen arrangements. Therefore, it is possible that small variations in the structure could dictate ligand specificity, underscoring the need to study key MSCRAMM/ligand interactions individually in submolecular detail.

Acknowledgment—We thank Jose Rivera for editing the manuscript.

REFERENCES

1. Patti, J. M., Allen, B. L., McGavin, M. J., and Höök, M. (1994) MSCRAMM-mediated adherence of microorganisms to host tissues. \textit{Annu. Rev. Microbiol.} \textbf{48}, 585–617

2. Hawiger, J., Timmons, S., Strong, D. D., Cottrell, B. A., Riley, M., and Doolittle, R. F. (1982) Identification of a region of human fibrinogen interacting with staphylococcal clumping factor. \textit{Biochemistry} \textbf{21}, 585–617
1407–1413
3. Greene, C., McDevitt, D., Francois, P., Vaudox, P., LeW, D. P., and Foster, T. J. (1995) Adhesion properties of mutants of Staphylococcus aureus defective in fibronectin-binding proteins and studies on the expression of fab genes. Mol. Microbiol. 17, 1143–1152
4. Foster, T. J., and Höök, M. (1998) Surface protein adhesins of Staphylococcus aureus. Trends Microbiol. 6, 484–488
5. Schneewind, O., Model, P., and Fischetti, V. A. (1992) Sorting of protein A to the staphylococcal cell wall. Cell 70, 267–281
6. Schneewind, O., Fowler, A., and Faull, K. F. (1995) Structure of the cell wall anchor of surface proteins in Staphylococcus aureus. Science 268, 103–106
7. Patti, J. M., Boles, J. O., and Höök, M. (1993) Identification and biochemical characterization of the ligand-binding domain of the collagen adhesin from Staphylococcus aureus. Biochemistry 32, 11428–11435
8. Rich, R. L., Kreikemeyer, B., Owens, R. T., LaBrenz, S., Narayana, S. V., Weinstock, G. M., Murray, B. E., and Höök, M. (1999) Ace is a collagen-binding MSCRAMM from Enterococcus faecalis. J. Biol. Chem. 274, 26939–26945
9. Nallapareddy, S. R., Weinstock, G. M., and Murray, B. E. (2003) Clinical isolates of Enterococcus faecium exhibit strain-specific collagen binding mediated by Acm, a new member of the MSCRAMM family. Mol. Microbiol. 47, 1733–1747
10. Lanneråsgård, J., Frykberg, L., and Guss, B. (2003) CNE, a collagen-binding protein of Streptococcus equi. FEMS Microbiol. Lett. 222, 69–74
11. Sato, Y., Okamoto, K., Kagami, A., Yamamoto, Y., Igarashi, T., and Kizaki, H. (2004) Streptococcus mutans strains harboring collagen-binding adhesin. J. Dent. Res. 83, 534–539
12. Shimooji, Y., Ogawa, Y., Osaki, M., Kabeya, H., Maruyama, S., Mikami, T., and Sekizaki, T. (2003) Adhesive surface proteins of Erysipelothrix rhusio- pathiae bind to polystyrene, fibronectin, and type I and IV collagens. J. Bacteriol. 185, 2739–2748
13. Xu, Y., Liang, X., Chen, Y., Koehler, T. M., and Höök, M. (2004) Identification and biochemical characterization of two novel collagen-binding MSCRAMMs of Bacillus anthracis. J. Biol. Chem. 279, 51760–51768
14. Symersky, J., Patti, J. M., Carson, M., House-Pompeo, K., Teale, M., Moore, D., Jin, L., Schneider, A., DeLucas, L. J., Höök, M., and Narayana, S. V. (1997) Structure of the collagen-binding domain from a Staphylococcus aureus adhesin. Nat. Struct. Biol. 4, 833–838
15. Zong, Y., Xu, Y., Liang, X., Keene, D. R., Höök, A., Gurusiddappa, S., Höök, M., and Narayana, S. V. (2005) A “Collagen Hug” model for Staphylococcus aureus CNA binding to collagen. EMBO J. 24, 4224–4236
16. Liu, Q., Ponnuraj, K., Xu, Y., Ganesh, V. K., Sillanpää, J., Murray, B. E., Narayana, S. V., and Höök, M. (2007) The Enterococcus faecalis MSCRAMM ACE binds its ligand by the Collagen Hug model. J. Biol. Chem. 282, 19629–19637
17. Switalski, L. M., Speziale, P., and Höök, M. (1989) Isolation and characterization of a putative collagen receptor from Staphylococcus aureus strain Cowan I. J. Biol. Chem. 264, 21080–21086
18. Patti, J. M., Bremell, T., Krajewska-Pietrasik, D., Abdelnour, A., Tarkowski, A., Rydén, C., and Höök, M. (1994) The Staphylococcus aureus collagen adhesin is a virulence determinant in experimental septic arthritis. Infect. Immun. 62, 152–161
19. Xu, Y., Rivas, J. M., Brown, E. L., Liang, X., and Höök, M. (2004) Virulence potential of the staphylococcal adhesin CNA in experimental arthritis is determined by its affinity for collagen. J. Infect. Dis. 189, 2323–2333
20. Nallapareddy, S. R., Qin, X., Weinstock, G. M., Höök, M., and Murray, B. E. (2000) Enterococcus faecalis adhesin, Ace, mediates attachment to extracellular matrix proteins collagen type IV and laminin as well as collagen type I. Infect. Immun. 68, 5218–5224
21. Hubble, T. S., Hatton, J. F., Nallapareddy, S. R., Murray, B. E., and Gillespie, M. J. (2003) Influence of Enterococcus faecalis proteases and the collagen-binding protein, Ace, on adhesion to dentin. Oral Microbiol. Immunol. 18, 121–126
22. de Bentzmann, S., Tristan, A., Etienne, J., Brousse, N., Vandenesch, F., and Lina, G. (2004) Staphylococcus aureus isolates associated with necrotizing pneumonia bind to basement membrane type I and IV collagens and laminin. J. Infect. Dis. 190, 1506–1515
23. Hienz, S. A., Schennings, T., Heimdahl, A., and Flock, J. I. (1996) Collagen binding of Staphylococcus aureus is a virulence factor in experimental endocarditis. J. Infect. Dis. 174, 83–88
24. Mamo, W., Fröman, G., and Müller, H. P. (2000) Protection induced in mice vaccinated with recombinant collagen-binding protein (CnBP) and α-toxoid against intramammary infection with Staphylococcus aureus. Microbiol. Immunol. 44, 381–384
25. Rahm, M. N., Lech, E. M., Patti, J. M., McDevitt, D., Höök, M., Jones, D. B., and Wilhelmsen, K. R. (2000) The collagen-binding adhesin is a virulence factor in Staphylococcus aureus keratitis. Infect. Immun. 68, 3776–3779
26. Elasri, M. O., Thomas, J. R., Skinner, R. A., Blevins, J. S., Beenken, K. E., Nelson, C. L., and Smeltzer, M. S. (2002) Staphylococcus aureus collagen adhesin contributes to the pathogenesis of osteomyelitis. Bone 30, 275–280
27. Singh, K. V., Nallapareddy, S. R., Sillanpää, J., and Murray, B. E. (2010) Importance of the collagen adhesin Ace in pathogenesis and protection against Enterococcus faecalis experimental endocarditis. PLoS Pathog. 6, e1000716
28. Nilsson, I. M., Patti, J. M., Bremell, T., Höök, M., and Tarkowski, A. (1998) Vaccination with a recombinant fragment of collagen adhesin provides protection against Staphylococcus aureus-mediated septic death. J. Clin. Invest. 101, 2640–2649
29. Papworth C., Bauer, J. C., Braman, J., and Wright, D. (1996) Site-directed mutagenesis in one day with >80% efficiency. Strategies 9, 3–4
30. Ho, S. N., Bauer, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77, 51–59
31. Barbu, E. M., Ganesh, V. K., Gurusiddappa, S., Mackenzie, R. C., Foster, T. J., Sudhof, T. C., and Höök, M. (2010) β-Neurexin is a ligand for the Staphylococcus aureus MSCRAMM SdrC. PLoS Pathog. 6, e1000726
32. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol. Crystallogr. 60, 2126–2132
33. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol. Crystallogr. 60, 2256–2268
34. Biacore Life Sciences (2004) BIAevaluation Software Handbook, GE Healthcare, Pittsburgh, PA