Trench-shaped Binding Sites Promote Multiple Classes of Interactions between Collagen and the Adherence Receptors, α1β1 Integrin and Staphylococcus aureus Cna MSCRAMM*

(Received for publication, April 14, 1999, and in revised form, May 20, 1999)

Rebecca L. Rich‡§, Champion C. S. Deivanayagam†, Rick T. Owens‡, Michael Carson¶, Agneta Höök‡, Dwight Moore‡, Vivian W.-C. Yang‡, Sthanam, V. L. Narayana‡, and Magnus Höök‡

From the †Center for Extracellular Matrix Biology, Institute of Biosciences and Technology, Texas A&M University, Houston, Texas 77030 and the ¶Center for Macromolecular Crystallography, University of Alabama at Birmingham, Birmingham, Alabama 35294

Most mammalian cells and some pathogenic bacteria are capable of adhering to collagenous substrates in processes mediated by specific cell surface adherence molecules. Crystal structures of collagen-binding regions of the human integrin αβ and a Staphylococcus aureus adhesin reveal a "trench" on the surface of both of these proteins. This trench can accommodate a collagen triple-helical structure and presumably represents the ligand-binding site (Emesley, J., King, S. L., Bergelson, J. M., and Liddington, R. C. (1997) J. Biol. Chem. 272, 28512–28517; 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. L. (1997) Nat. Struct. Biol. 4, 833–838). We report here the crystal structure of the α subunit I domain from the αβ integrin. This collagen-binding protein also contains a trench on one face in which the collagen triple helix may be docked. Furthermore, we compare the collagen-binding mechanisms of the human α integrin I domain and the A domain from the S. aureus collagen adhesin, Cna. Although the S. aureus and human proteins have unrelated amino acid sequences, secondary structure composition, and cation requirements for effective ligand binding, both proteins bind at multiple sites within one collagen molecule, with the sites in collagen varying in their affinity for the adherence molecule. We propose that (i) these evolutionarily dissimilar adherence proteins recognize collagen via similar mechanisms, (ii) the multisite, multiclass protein/ligand interactions observed in these two systems result from a binding-site trench, and (iii) this unusual binding mechanism may be thematic for proteins binding extended, rigid ligands that contain repeating structural motifs.

Collagen polypeptides are largely composed of repeats of the GPX tripeptide and associate to form triple-helical monomers. These monomers combine into macroscopic fibers. Prokaryotic and eukaryotic cells bind collagen via receptors on their cell surfaces (1–6). We now hypothesize that to accommodate such an unusually shaped ligand, the collagen-binding surface proteins of these cells must adopt an atypical binding-site structure.

Bacterial pathogens utilize this interaction as a means of adherence to collagenous host tissues. Some Staphylococcus aureus strains express an adhesin, Cna, of the MSCRAMM class that binds collagen (1, 7–13). Cna from S. aureus FDA 574 is depicted in Fig. 1a: it contains two major domains, A and B, in addition to features characteristic of cell-surface proteins on Gram-positive bacteria (11). The collagen-binding site has been localized within the Cna A domain (12). Binding analyses demonstrate that (i) a synthetic peptide mimicking a short sequence of the A domain can inhibit collagen binding to S. aureus (8); (ii) the A domain/collagen interaction involves more than one affinity class and multiple sites of contact within a single collagen molecule (8, 10); and (iii) the B domain does not alter the collagen binding ability of the A domain (13).

The crystal structure of a truncated form of the Cna A domain reveals a binding-site "trench" on one face of the protein. In molecular modeling studies, this trench was found to accommodate a triple-helical peptide that mimics the collagen structure. Symersky et al. (7) noted that this trench complemented well the structure of a collagen triple helix and binding studies of site-specific mutants of the S. aureus Cna truncate revealed that (i) no single residue or area within the trench was responsible for collagen binding, but rather, a number of contacts contributed to the protein/collagen interaction and (ii) this binding-domain truncate bound to multiple sites along a collagen molecule. The affinity of Cna for an individual site within collagen may be the consequence of the number of "good" and "bad" contacts within the binding trench.

Binding of eukaryotic cells to collagen serves not only as a mechanism of tissue adherence, but also may induce a complex signaling cascade in the cell. Attachment of eukaryotic cells to the extracellular matrix is primarily mediated by integrins. The integrins are transmembrane αβ heterodimeric proteins that direct cell-cell and cell-matrix interactions and are found on most mammalian cells. To date, several integrins, including αβ, αβ, αβ, αβ, αβ, αβ, and αβ, have been reported to mediate cellular adherence to collagen (14–18). Of these, αβ

* This work was supported by the Arthritis Foundation (to R. L. R.) and National Institutes of Health Grant AR44415 (to M. H. and S. V. L. N.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† The abbreviations used are: Cna, Staphylococcus aureus collagen adhesin; MSCRAMM, microorganism surface component recognizing adhesive matrix molecules; MIDAS, metal ion-dependent adhesion site; PAGE, polyacrylamide gel electrophoresis; SPR, surface plasmon resonance spectroscopy.
and $\alpha_2\beta_1$ are apparently the primary collagen-binding integrins. The $\alpha_1$, $\alpha_2$, and $\alpha_3\beta_1$ subunits each contain an “inserted” (I) domain near the N terminus (Fig. 1c). The I domains have been shown to contain a ligand-binding site and a MIDAS motif, which needs to be occupied by an appropriate cation for effective ligand binding by the integrin. Recombinant proteins duplicating these small (approximately 200 amino acids) I domain polypeptide segments effectively bind collagen, presumably because these regions are responsible for the integrins’ binding to collagens.

A binding site-trench in the I domain of the human $\alpha_2\beta_1$ integrin was also suggested by the crystal structure of the $\alpha_2$ integrin I domain. Molecular modeling of this protein complexed with a collagen triple-helical peptide (19) demonstrated favorable ligand docking encompassing about 10 residues of the collagen sequence within a trench spanning the MIDAS motif. From this work Liddington and co-workers (19) suggested that the divalent cation is involved in direct ligand binding via coordination of an amino acid residue (most probably, glutamate) within the collagen molecule.

Our previous work has shown that the full-length A domain of the S. aureus Cna protein binds collagen more efficiently than the binding-domain truncate does (7, 8, 10, 12). The behavior of this construct has not been investigated to date. In addition, detailed analysis of collagen-binding activity of the human $\alpha_1$ integrin I domain, which binds Type I collagen more efficiently than the $\alpha_2$ integrin I domain does (14), has not been performed. The questions we seek to answer here include: (a) can the gross structure of I domains and the detailed topology of their MIDAS-centered binding site be determined from modeling experiments based on known structures? (b) Is a trench similar to that found on Cna the binding-site motif employed by the $\alpha_2\beta_1$ integrin? (c) Does the $\alpha_1$ integrin I domain bind to a single or multiple class(es) of sites within a collagen macromolecule? To address these questions, we compare the structures and collagen-binding characteristics of the S. aureus Cna A and the human $\alpha_1$ integrin I domains.

**EXPERIMENTAL PROCEDURES**

**Construction of Expression Plasmids**—The expression plasmid pQE-$\alpha_1$I was constructed based on the vector pQE-30 (Qiagen Inc., Chatsworth, CA) using standard molecular biology protocols (20, 21). cDNA encoding the I domain of the human $\alpha_1$ integrin was obtained by polymerase chain reaction using a human hepatoma cDNA library as a template and the oligonucleotide primers, 5'-CGATTCCTCCGCATTTC-CAAGTCGTGAAT-3' and 5'-GCTGACGTCAATTCTTTCTCCCCAGA-GTTT-3'. The amplified gene fragment was digested with the BamHI and PstI restriction endonucleases, purified by agarose gel electrophoresis (Genclean kit, ISC BioExpress), and ligated into the vector pQE-30 (previously linearized by digestion with the same endonucleases). Ligation mixtures were subsequently transformed into Escherichia coli strain JM101.

Plasmids from isolated transformants were analyzed by restriction digestions and automated DNA sequencing analysis (Molecular Genetics Core Facility, University of Texas Medical School, Houston, TX) to confirm the expected open reading frame. The $\alpha_1$ integrin I domain sequence examined here corresponds to the sequence published by Briesewitz et al. (22) except for nucleotide substitutions resulting in Lys74→Glu and Thr293→Ile.

The $\alpha_1$ integrin I domain cDNA was also cloned into the glutathione $S$-transferase expression vector pGEX-KG (23). Recombinant GST-$\alpha_1$I fusion proteins were purified by chromatography over glutathione-agarose and cleaved by digestion with thrombin as described in Ref. 23.

The construction of a plasmid for the expression of the Cna A domain has been previously described and consists of the gene segment encoding the S. aureus collagen MSCRAMM amino acids Ala30-Glu 531 cloned between the BamHI and SaI restriction sites of the pQE-30 expression vector (13).

**Expression and Purification of Recombinant Proteins**—Large-scale preparations of recombinant protein were prepared and purified as follows. Overnight cultures (40 ml) of stationary-phase bacteria were seeded to 1 ml ofuria broth and the cells were allowed to grow for 2.5 h at 37 °C (OD600 nm ~ 0.6). Protein expression was induced by addition of isopropyl-$\beta$-thiogalactopyranoside to a final concentration of 0.2 mM and the culture was incubated for an additional 3 h at 37 °C. Bacteria were then collected by centrifugation and resuspended in a minimal volume of 4 ml Tris, 100 mM NaCl, pH 7.9, before being frozen at −80 °C.

Iodinated bacteria were thawed and passed through a French press (11,000 p.s.i) twice to lyse the cells. Insoluble debris was removed by centrifugation at 14,000 rpm for 20 min and the supernatant was filtered through a 0.45 µm membrane. Imidazole was added to a final concentration of 0.67 mM and the lysates were applied to a 10 × 100-mm column of Ni2+–charged iminodiacetic acid/Sephrose. The column was washed with 30 ml of 4 mM Tris, 100 mM NaCl, 5 mM imidazole, pH 7.9, and bound protein eluted with a 200-ml linear gradient of 0–200 mM imidazole in 4 ml Tris, 100 mM NaCl, pH 7.9. Fractions containing the desired protein, as determined by SDS-PAGE, were pooled and concentrated using an Amicon ultrafiltration system. The isolated proteins were essentially pure and appeared as single bands on an overloaded SDS-PAGE gel. The isolated recombinant proteins were dialyzed against 3 × 1-liter changes of 1 mM EDTA, 50 mM Hepes, 150 mM NaCl, pH 7.4, to remove all cations, and then dialyzed against 3 × 1-liter changes of 50 mM Hepes, 150 mM NaCl, pH 7.4, to remove EDTA. All buffers for the $\alpha_1$ integrin I domain protein also contained 5 mM $\beta$-mercaptoethanol; the justification for adding a reducing agent to this sample solution is discussed below.

During our initial analyses of the recombinant His6 tag $\alpha_1$I integrin I domain protein, we observed gradual precipitation of the protein within several days post-purification when the solution was kept at 4 °C. The presence of dimeric and higher-order multimers of the recombinant $\alpha_1$I integrin I domain protein in the solution was apparent by SDS-PAGE (data not shown). Addition of 5 mM $\beta$-mercaptoethanol delayed the protein precipitation for several weeks. All studies discussed here were performed on recombinant $\alpha_1$I integrin 1-domain protein within 2 weeks of expression and purification and in buffer containing 5 mM $\beta$-mercaptoethanol, unless noted otherwise. The far-UV CD spectra of freshly purified $\alpha_1$I integrin I domain in the presence and absence of 5 mM $\beta$-mercaptoethanol are identical (data not shown). The sensorgrams measured over a
period of days for the α₁ integrin I domain protein flowed over collagen in buffer containing the reducing agent remain unchanged; repeating this experiment in the absence of the reducing agent, however, revealed the gradual increase in association and decrease in dissociation of the protein-collagen complex over time. After approximately 2 weeks, the sensorgrams for the α₁ integrin I domain in the absence of β-mercaptoethanol duplicated those published previously (24). The increase in apparent affinity of the α₁ integrin I domain protein after storage for collagen may be due to the contribution of multiple I domain elements in the protein aggregate binding at one location within the collagen macromolecule. The addition of the reducing agent is therefore necessary to preserve the monomeric state of the recombinant protein and does not alter its structure or function.

**Surface Plasmon Resonance Spectroscopy (SPR)**—Analyses were performed using the BIAcore system as described in Ref. 13, with 5 mM β-mercaptoethanol and 0.25% octyl-β-D-glucopyranoside included in the buffer for α₁ integrin I domain analyses. No mass transport effects were observed in these measurements. The data for the construction of the Scatchard plots was obtained from the equilibrium portion of the SPR sensorgrams and analyzed as described.2

**Enzyme-linked Immunosorbent Assays**—Assays were performed as described in Ref. 13. For wells in which the buffer included MgCl₂, all washes and incubations were performed in the presence of 1 mM MgCl₂.

**Equilibrium Dialysis**—The equilibrium dialysis experiments were carried out in a double acrylic microdialysis module (Hoffer, San Francisco, CA) as described by Yang et al. (26). Aliquots of 150 μl of threbin-cleaved α₁ integrin I domain protein in 10 mM Tris-HCl, 150 mM NaCl, pH 7.0, were added to the inner compartments. The same volume of 0–5 mM ultrapure MgCl₂ (Sigma) in 10 mM Tris-HCl, 150 mM NaCl, pH 7.0, was added to the outer compartments. After incubation, the concentration of MgCl₂ in the outer compartments was determined using a Mg²⁺ detection kit (Sigma). An aliquot of 10 μl from each outer compartment and 100 μl of each kit component were mixed. The reaction was immediate and sample absorbance was measured at 525 nm using a Molecular Devices plate-reading visible spectrophotometer. Calculation of the Mg²⁺-complexed α₁ integrin I domain fraction was performed as described in Ref. 27.

**Crystallography**—The crystal structure of the recombinant His₆ tag α₁ integrin I domain protein in 10 mM HEPES, 200 mM NaCl, 5 mM MgCl₂, 5 mM β-mercaptoethanol, pH 7.0, was further purified using a 300 × 7.5 Bio-sil-TSK125 gel-filtration column. The protein solution was then concentrated to 20 mg/ml using an Amicon ultrafiltration system and dialyzed against 10 mM HEPES, 200 mM NaCl, 5 mM MgCl₂, 5 mM β-mercaptoethanol, pH 7.0. Crystallization trials were set up using the hanging-drop vapor-diffusion method. High quality crystals were obtained from a droplet made by mixing 2 μl of protein solution and 2 μl of 31% PEG2000, 50 mM HEPES, 200 mM NaCl, 5 mM MgCl₂, 5 mM β-mercaptoethanol, pH 7.5 (solution A), and equilibrating it against 1 ml of solution A. Crystals were prism-shaped, with the largest having dimensions of 0.3 × 0.2 × 0.1 mm.

**Diffraction Data Collection**—Crystals were soaked in a synthetic mother liquor containing 15% glycerol as cryoprotectant and subsequently cryo-cooled using an Oxford cryosystem (Oxford Cryosystems, Oxford, United Kingdom). X-ray diffraction data were collected using a RAXIS IV imaging plate system mounted on a RIGAKU RU-HBR rotating-anode generator (50 kV, 100 mA). A complete native data set was collected over 99 frames (oscillation of 2°, exposure time of 20 min, and crystal-to-image plate distance of 150 mm). The frames were indexed and scaled using DENZO and SCALEPACK (28). The scaled data had 99% completeness, where 71.5% of the data in the last shell was above 2σ level with an R̄max value of 6.1%. The calculated Matthews coefficients, V₁/ₐ, was 1.9 Å³ Da⁻¹, suggesting two molecules exist in the asymmetric unit with an estimated solvent content of 35%. Data collection details are presented in Table I.

**Structure Determination**—The crystal structure of the recombinant His₆ tag α₁ integrin I domain was determined by the molecular replacement method using the CCP4 integrated version of AmoRe (29, 30). We used the molecular model of the homologous component factor B middle domain as the initial molecular replacement unit (31). The C-terminal helix and all the connecting loops were removed from the starting molecular replacement search model, which had 109 residues. Repeated rounds of rigid-body refinement and checking for acceptable crystal packing helped us to identify two solutions having high correlation factors and the lowest R-factors (0.62 and 41%, respectively, for 8.0–4.5 Å resolution data). The two molecules in the asymmetric unit were not related by an exact 2-fold non-crystallographic axis. Next, the side chains of the correctly positioned model were replaced with the corresponding homologous side chains of the α₁ integrin I domain. Rigid body refinement to 3.0 Å resolution, where the individual secondary structural elements were treated as independent units in XPLOR (32), resulted in an R-factor of 39% and Rfree (calculated on 10% of the reflections) of 44%. Several rounds of manual refining to 2Fo – Fc maps using the graphics program O (33) and positional refinement in XPLOR were done while extending the resolution to 2.5 Å in small increments. At this stage, the R-factor was 31%, Rfree value was 41%, and a 2Fo – Fc map calculated had visible density for the missing C-terminal helix and for most of the deleted loop regions. At this juncture, the resolution was extended to the final 2.0 Å and two rounds of simulated annealing and model rebuilding led to the tracing of the complete C-terminal end. After one refinement cycle of individual B-factors (r = 26% and Rfree = 30%), water molecules were added to the model by picking the peaks above 3σ level in a calculated (Fo – Fc) difference map. Two of these water molecules were identified as metal ions based on their bonding geometry. The OOPS program (34) was used throughout the cycles of rebuilding for quality checks. The cross-validated maximum likelihood refinements were performed with CNS-0.4 (35). Bulk solvent corrections were applied in the last few cycles of
refinement. The final refinement yielded 229 water molecules, two Mg²⁺ ions, 3008 non-hydrogen atoms, and four cis-prolines. For 24,537 reflections (out of 24,807 reflections) between 100.0 and 2.0-Å resolution, the final R-factor was 20.6% and R_free was 24.3%. The final structure was checked using PROCHECK (36, 37) and WHAT_CHECK (38). The complete refinement statistics are presented in Table I. The α₁ integrin I domain structure was aligned with other integrin I domains and von Willebrand factor A3 domain crystal structures taken from the protein data bank (1AO3 for von Willebrand factor, 1AOX for the α₂ integrin, 1JLM for the α₃ integrin, and 1ZON for the α₃ integrin) using the program MODELER (39).

Non-crystallographic constraints were removed when refining the models at 2.0-Å resolution. The two molecules present in the asymmetric unit were refined independently and the root mean square deviation between molecules A and B for main chain atoms was 0.23 Å. The C-terminal ends were identical, while two extra residues could be traced at the N-terminal end of molecule A. Both molecules were almost identical, especially around the metal-binding site. Solvent molecules around the MIDAS site were conserved between the two non-crystallographically related α₁ integrin I domain molecules in the asymmetric unit.

Docking Search—The collagen peptide mimic used in the docking simulations was similar to the one used in the studies of *S. aureus* Cna (7). It was obtained from the Protein Data Bank crystal structure entry 1Iag (40) and shortened to the C-terminal [G-P-P⁺]₃. The docking target was the molecule B in the refined crystal structure of the 1cag (40) and shortened to the C-terminal [G-P-P⁺]₃. The docking (7). It was obtained from the Protein Data Bank crystal structure entry (GPP⁺) unit. The one solution that extended symmetrically across the MIDAS was selected for energy-minimization in XPLOR.

Among the eight best-fit solutions, four showed the collagen triple-helical peptide bound in the trench on the MIDAS face. The other four solutions were eliminated because the peptide mimic bound at the face opposite the MIDAS, where the I domain is proposed to interface with the α₁ integrin repeat units (42). The four solutions selected were superimposable, except that each was translated along the triple helical axis by one (GFP⁺) unit. The one solution that extended symmetrically across the MIDAS was selected for energy-minimization in XPLOR.

RESULTS

Recombinant α₁ Integrin I Domain Protein Adopts a Dinucleotide-binding Fold and Contains an Active MIDAS Motif—Resolution of the crystallographic data of the α₂ integrin I domain revealed that this protein adopts a dinucleotide-binding (Rossman) fold, in which a central core of five parallel β-strands and one smaller anti-parallel β-strand are encased in seven α-helices (Fig. 2a). This general structural organization has been observed in the crystal structures of I domains from other integrin α-subunits and I domain-like segments of von Willebrand factor and complement factor B (43–50). The order of the β-strands, beginning at the N terminus, is β₁-β₅-β₆. Five (α₁-α₃-α₄-α₅-α₆) helices are parallel to one another and anti-parallel to the neighboring β-strands. The α₂ helix is parallel to the β₂ strand where its N-terminal end is connected to the C-terminal end of the β₁-strand through the short anti-parallel β₅-strand. The short, two-turn α₃ helix protrudes above the molecule in the carboxyl end of the β-sheet.

A MIDAS motif composed of Asp¹⁵⁴, Ser¹⁵⁶, Ser¹⁵⁸, Thr²²⁴, and Asp²⁵⁷ (the numbering of residues in the mature protein follows that given by Emsley et al. (19)) exists in the α₁ integrin I domain (Fig. 2b). The crystal structure of the α₁ integrin I domain protein in the presence of 5 mM MgCl₂ revealed that Mg²⁺ is octahedrally coordinated to Ser¹⁵⁶, Ser¹⁵⁸, and Asp²⁵⁷, and three water molecules with distances of 2.1 ± 0.1 Å. Asp¹⁵⁴ and Thr²²⁴ of the α₁ integrin I domain are hydrogen-bonded to the Mg²⁺ through one of the coordinated water molecules. The MIDAS residues are conserved between the α₁ and α₂ integrin I domains.

A Binding Site Trench in the α₁ Integrin I Domain Accommodates a Triple-helical Collagen Peptide Mimic—The α₁ integrin I domain contains a structural feature found in the α₂ integrin I domain but not observed in the other I domains: the short, two-turn α₅ helix (denoted as the C-helix in the α₂ integrin I domain (19), which defines one side of the putative ligand-binding surface. In the α₁ integrin I domain, this helix is composed of residues 287–291 (GSYNR) and protrudes out of the MIDAS-centered trench of the α₁ integrin I domain.

![Fig. 2. a, ribbon diagram of the human α₁ integrin I domain. β-Sheets are shown in green, α-helices in cyan, 310-helices in navy, and loops in orange; the Mg²⁺ ion is depicted as a orange sphere. b, structure of the MIDAS and coordination of Mg²⁺. The numbering of residues follows that given for the α₁ integrin I domain in Ref. 19. The backbone of the molecule is shown in gray, with specific carbon atoms in green and oxygen atoms in red. The cation (M) is shown in orange and the ion-coordinated water (W) molecules are depicted as magenta spheres. c, space-filled depiction of a collagen triple-helical mimic docked in the MIDAS-centered trench of the α₁ integrin I domain.](image)

3 M. Carson, unpublished data.
ing in ligand binding (Fig. 3a). The Mg\(^{2+}\) ion is located in the deep central trench pocket, which is lined with all four types of residues. From Fig. 3a, it is apparent that the cation contributes a small percentage of the surface area of the trench and is potentially involved in ligand capture. Similar results were reported for the cation in the \(\alpha_2\) integrin I domain trench (19).

Complex Protein/Collagen Interactions May Result from Ligand Binding in the Trenches of the \(\alpha_1\) Integrin I Domain and \(S. aureus\) Cna—A bacterial adhesin, Cna from \(S. aureus\), also binds collagen (1, 7–13). In crystallographic studies, collagen docked well in a trench-shaped binding site on one face of its minimal binding domain, Cna 151–318 (7). The trench in Cna 151–318 is 5 Å deep, 25 Å long, and 15 Å wide (7), and encompasses three collagen GP\(\times\) repeats, as do the trenches in the I domains of the \(\alpha_1\) and \(\alpha_2\) integrin I domains, but its topology and residue distribution are unlike that of the I domains. The trench in Cna 151–318 is dominated by polar residues, with very few acidic, basic, and hydrophobic residues possibly participating in collagen binding (Fig. 3b). This trench contains two polar pockets and one hydrophobic/polar pocket that may be amenable to the binding of bulky collagen side chains (Fig. 3d).

Divalent Cations Enhance the Collagen Binding of Human \(\alpha_1\) Integrin I Domain, but Not That of the \(S. aureus\) Cna A Domain—SPR changes were used to analyze the binding of recombinant forms of the \(\alpha_1\) integrin I and \(S. aureus\) Cna A domains to immobilized Type I collagen. In both panels of Fig. 4, the protein/collagen association occurred between 140 and 375 s, with the dissociation beginning at 375 s. For both the \(\alpha_1\) integrin I domain/Mg\(^{2+}\) and \(S. aureus\) Cna A domain, the association and dissociation with collagen was rapid and apparently quite similar.

The sensorgrams shown in Fig. 4a demonstrate that the presence of 1 mM Mg\(^{2+}\) in the milieu increased the \(\alpha_1\) integrin I domain’s collagen-binding capacity dramatically. A Scatchard analysis of equilibrium dialysis determination of the \(\alpha_1\) integrin I domain’s affinity for Mg\(^{2+}\) was linear and showed a single Mg\(^{2+}\)-binding site in the I domain having a \(K_D\) of approximately 10 \(\mu\)M (data not shown). This results in >99% of the \(\alpha_1\) I domain being cation-complexed in 1 mM Mg\(^{2+}\) (27).

In contrast, addition of 1 mM Mg\(^{2+}\) to the analysis buffer had little observable effect on the collagen binding capacity of the Cna A domain in the SPR measurements (Fig. 4b). The collagen binding by the recombinant A domain truncate, Cna 151–318 (of which the crystal structure is known), was also cation-independent (data not shown). These results were not surprising considering the absence of MIDAS, EF-hand, or other cation-binding motifs in the Cna protein sequence.

Multiple Binding Classes Exist for the Interaction of Collagen with Human \(\alpha_1\) Integrin I and \(S. aureus\) Cna A Domains—In an attempt to obtain kinetic and equilibrium constants for these protein/collagen interactions, we examined the SPR pro-
files over a range of concentrations of α1 integrin I and Cna A domains flowed over immobilized collagen. Fig. 5a illustrates the SPR profiles expected for a simple 1:1 immobilized ligand/mobile protein system (or 1:P, where all protein macromolecules, P, bind the immobilized ligand noncooperatively and with equal affinity) over a range of mobile protein concentrations (53). As the protein concentration exceeds the dissociation constant, saturation of sites within the immobilized ligand occurs earlier in the sensorgram, with the equilibrium plateau becoming more apparent. Fig. 5, b and c, are the profiles of the mobile Cna A and α1 integrin I domains flowed over immobilized Type I collagen. From these panels it is apparent that neither the Cna A nor the α1 integrin I domain recombinant proteins obey pseudo first-order binding kinetics; but rather, both proteins’ interactions with collagen are more complex.

To examine these interactions further, we determined the binding of each recombinant to collagen using SPR across an even wider concentration range and calculated the populations of collagen-bound and -free recombinant protein. These measurements produced the Scatchard plots shown in Fig. 6. The hypothetical one-simple-binding-class data from Fig. 5a would yield a linear Scatchard plot. The Scatchard plots of the Cna A and α1 integrin I domain recombinant proteins shown in Fig. 6 are dramatically concave upward, however. Similar concave Scatchard plots were obtained: 1) by flowing these proteins over Type II collagen; 2) by replacing Mg2+ with Mn2+ in the α1 integrin I domain analysis buffer; and 3) for recombinant Cna proteins spanning residues 151–318 and 30–721 (data not shown). The nonlinear Scatchard plots of Fig. 6 are not merely experimental artifacts, for under similar experimental conditions we obtained a linear plot for the binding of collagen by an Enterococcus faecalis MSCRAMM, Ace.

Not only do the Scatchard plots in Fig. 6 demonstrate the multiple binding classes of these proteins’ interactions with collagen, but the plots also reveal that the recombinant α1 integrin I and Cna A domains bind at a host of sites along the collagen strand. The highest affinity interactions of the α1 integrin I or Cna A domain with collagen occur at the fewest number of sites, with an increasing number of sites, n, occupied as the proteins’ affinities for collagen decreases. The n1 obtained from the linear extrapolations in Fig. 6 represent those matching the “highest” and “lowest” affinities described above. Clearly, intermediate n1 also exist, as may higher-order n1 that correlate with the very low affinity protein/collagen interactions.

**DISCUSSION**

As different as the MSCRAMM and cation-bound integrin recombinant proteins appear initially by sequence and structure comparisons, their collagen-binding mechanisms appear quite similar. The α1 integrin I (Mg2+) and Cna A domains exhibited comparable net affinities (Fig. 5, legend) and kα and kD (Fig. 4) in their interactions with Type I collagen. (Fitting
the kinetic data for each protein to one (or even two) on- and off-rates did not yield statistically acceptable results, indicating the presence of more than two on- and off-rates.) In addition, the sensorgrams of neither the α1 integrin I (Mg2+) nor Cna A domain flowed over immobilized Type I collagen approximated that of a system having one or very few binding classes (Fig. 5) and the Scatchard plots of both these proteins binding to collagen were not easily resolved into standard fitting curves (Fig. 6). We interpreted these data to be the result of multiple classes of interactions occurring between the protein and collagen (with each class, i, have a corresponding number of interactions, ni). We have considered several of the typical binding mechanism scenarios (most significantly, the possibilities of binding cooperativity (54) and overlapping adhesin-binding sites in collagen (55)), but have found that none fit the data in Fig. 6 well, for each lacks a factor to account for the microscopic heterogeneity in collagen and consequently, the possibility of multiple nonidentical adhesin-binding sites (56). From the linear regressions of the Scatchard plots in Fig. 6, we report here 1) the lowest and relative highest number of interactions and 2) the class of highest affinity, the class of lowest observed affinity, and noted that intermediate classes of undetermined affinities exist (as well as classes of progressively weaker affinities beyond the detection limits of this assay system).

This atypical ligand-binding behavior may result from collagen binding in the trench of the α1 integrin I or the Cna A domains. Many segments of collagen may fit within the trench, but the protein’s affinity for a particular segment may be determined by the specific interactions (e.g., hydrophobic, ionic, hydrogen-bonding) between particular residues in collagen and those lining the binding-site trench. The microstructure of collagen (particularly the presence of a particular amino acid in the third position of the repeat sequence, GPX) will determine which segment is most amenable to docking in the protein’s trench. From Fig. 3 it is apparent that the topologies of the α1 integrin I and the Cna A domains’ trenches are quite different, suggesting that these two adherence receptors preferentially bind different collagen segments. It is also possible that the differences in these trenches provide for the integrin and bacterial protein to differ in their affinities for various collagen types.

A collagen triple-helical peptide composed of three GPX repeats also docked well within the α2 integrin I domain trench (19). Superposition of the accessible surface areas within the trenches of the α1 and α2 integrin I domains revealed that the α2 integrin I domain trench is: 1) much less flexible and 2) more restricted in the number of collagen triple-helical conformations it is amenable to docking than the trench of the α1 integrin I domain (data not shown). The most significant difference between the trench topologies of the α1 and α2 integrin I domain trenches is the positioning of a tyrosine residue. Tyr285 of the α0 integrin I domain was found to be pointing into the trench, but the comparable tyrosine of the α1 integrin I domain (Tyr285) is shifted to the bottom of the trench and held in place by enhanced hydrophobic interactions due to the substitution of Phe289 for the Leu286 present in the α2 integrin I domain. In addition, the hydroxyl group of Tyr285 is buried and hydrogen-bonded to a backbone nitrogen and the carboxylate side chain of Glu259. In addition, there appear to be significant differences in the contours and charge/hydrophobicity distributions within the trenches of the α1 and α2 integrin I domains, which indicates that these two proteins may bind dissimilar segments of collagen (or differ in their affinities for various collagen types).

Each of these adhesive proteins binds at multiple sites in collagen, with perhaps the most efficient interaction occurring at only one (or a very few) site. If the recombinant Cna or integrin protein recognizes various peptide sequences in the collagen macromolecule (which may contain a few required key residues and other variable residues that determine binding efficiency), the protein population may well bind at multiple locations along the collagen strand, with each binding event having a unique KD (Fig. 7). There may be indeed a particular amino acid sequence or conformation in the collagen strand that is most amenable for binding to a trench of a particular adherence molecule, but this site in the collagen strand is not dramatically more suitable than many others. Such behavior would explain the collagen-binding results we observe for the MSCRAMM and integrin ligand-binding domains: multiple protein molecules bound, with varying affinity, to a single collagen moiety, with equilibrium and site saturation not easily achieved. The sum of all these interactions could produce the spectrum of affinities that we observe in the binding analyses of these recombinants and collagen.

The high degree of amino acid sequence homology between the α1 and α2 integrin I domains would indicate that the 2-fold similarly. Hence, it is no surprise that ribbon diagrams of the two proteins appear almost identical, with a root mean square deviation of 1.45 Å for the main chain atoms. In fact, modeling the α1 integrin I domain sequence using the structural coordinates of the α2 integrin I domain suggested that the α1 integrin I domain would adopt a Rossman folding motif. This modeling, however, did not provide adequate resolution to refine the trench microstructure. Only upon solving the crystallographic structure of the α1 integrin I domain were we able to characterize its trench topology and propose which residues interact with collagen. We suggest that modeling of other I domains for which the structures have not been solved experimentally will reveal whether or not they adopt the expected Rossman fold, but the characterization of the ligand-binding site (particularly the putative trench of the collagen-binding α10 integrin I domain (17)) needs to be experimentally determined, however.

The trench-as-binding-site motif has also been reported for collagen-binding proteins that are not cell-surface proteins. These include the fiddler crab and human fibrobast collagenases. Fletterick and co-workers (57) reported that the fiddler crab collagenase-binding site is a negatively-charged, elongated, cylindrical pocket wide enough to accommodate the collagen triple helix. The human fibrobast collagenase resembles the integrin I domains in that a cation (Zn2+) resides in the center of the trench and is crucial for efficient catalysis (58). Lovejoy et al. (58) identified multiple interactions between residues within the collagenase trench and the inhibitor: 1) the zinc ion presumably coordinates an inhibitor carboxylate group; 2) eight hydrogen bonds exist between the two species; and 3) inhibitor hydrophobic residues fit in complementary pockets within the trench.

We suggest that the trench observed in the crystal structures...
Collagen Binding by Cell Surface Receptors

16. Wayner, E. A., and Carter, W. G. (1987) J. Cell Biol. 105, 1873–1884
17. Camper, L., Hellman, U., and Lundgren-Åkerland, E. (1998) J. Biol. Chem. 273, 20383–20389
18. Walzog, B., Schuppan, D., Heimpel, C., Hafezi-Moghadam, A., Gaethgens, P., and Ley, K. (1995) Exp. Cell Res. 218, 28–38
19. Emsley, J., King, S. L., Bergelson, J. M., and Liddington, R. C. (1997) J. Biol. Chem. 272, 26512–26517
20. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
21. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. M., Seidman, J. G., Smith, J. A., and Struhl, K. (1991) Current Protocols in Molecular Biology, Greene Publishing and Wiley-Interscience, New York
22. Brown, R., Epstein, M. R., and Marcantonio, E. (1993) J. Biol. Chem. 268, 2889–2906
23. Guan, K. L., and Dixon, J. E. (1991) Anal. Biochemistry 192, 262–267
24. Calderwood, D. A., Tuckwell, D. S., Ehle, J., Kuhn, K., and Humphries, M. J. (1997) J. Biol. Chem. 272, 12311–12317
25. Stern, L. J., Brown, J. H., Jardetzky, T. S., Gorga, J. C., Urban, R. G., Strominger, J. L., and Kyle, D. C. (1994) Nature 368, 215–221
26. Yang, V. W.-C., LaBrenz, S. R., Rosenberg, L. C., McQuillan, D., and Ho¨ök, M. (1999) J. Biol. Chem. 274, 12454–12460
27. Rich, R. (1995) Photophysics and Activity of Biological Systems Containing the Optical Probes 7-Azatryptophan and Its Analogs. Doctoral dissertation, Iowa State University, Ames, IA
28. Ottowinson, Z., and Minor, W. (1997) Methods Enzymol. 276, 307–326
29. Navaza, J. (1994) Acta Crystallogr. Sec. A 50, 157–163
30. Collaborative Computer Project No. 4 (1994) Acta Crystallogr. Sec. B 50, 1576–1576
31. Tuckwell, D. S., Xu, Y., Newham, P., Humphries, M. J., and Volanakis, J. E. (1997) Biochemistry 36, 6605–6613
32. Brunger, A. T. (1992) X-PLOR manual, Version 3.1, Yale University Press, New Haven, CT
33. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Acta Crystallogr. Sec. A 47, 110–119
34. Kleywegt, G. J., and Jones, T. A. (1997) Methods Enzymol. 276, 208–230
35. Adams, P. D., Pannu, N. S., Read, R. J., and Buggage, A. T. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 5018–5023
36. Morris, A. C., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. (1992) Proteins Struct. Funct. Genet. 12, 345–364
37. Leskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) J. Appl. Crystallogr. 26, 283–291
38. Ho¨ök, M. (1995) Methods Enzymol. 272, 560–763
39. Tuckwell, D. S., Xu, Y., Newham, P., Humphries, M. J., and Volanakis, J. E. (1997) Biochemistry 36, 6605–6613
40. Springer, T. A. (1997) Methods Enzymol. 281–296
41. Huizinga, E. G., van der Plas, R. M., Kroon, J., Sixma, J. J., and Gros, P. (1997) Structure 5, 1147–1156
42. Bieniaukswsky, J., Cruz, M., Atiemo, A., Handin, R., and Liddington, R. (1995) J. Biol. Chem. 270, 12322–13340
43. Hung, I.-C., Lee, J., and Liddington, R. (1998) J. Mol. Biol. 289, 10277–10281
44. Qu, A., and Leahy, D. J. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 65–72
45. Emsley, J., Cruz, M., Handin, R., and Liddington, R. (1998) J. Biol. Chem. 273, 10396–10401
46. Lee, J.-O., Rieu, P., Arnautov, M. A., and Liddington, R. (1995) Cell 80, 621–638
47. Qu, A., and Leahy, D. J. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 5927–5932
48. Huizinga, E. G., van der Plas, R. M., Kroon, J., Sixma, J. J., and Gros, P. (1997) Structure 5, 1147–1156
49. Bieniaukswsky, J., Cruz, M., Atiemo, A., Handin, R., and Liddington, R. (1997) J. Biol. Chem. 272, 25162–25167
50. Tuckwell, D. S., Xu, Y., Newham, P., Humphries, M. J., and Volanakis, J. E. (1997) Biochemistry 36, 6605–6613
51. Nicholls, A., Sharp, K. A., and Honig, B. (1991) Proteins Struct. Funct. Genet. 11, 281–296
52. Carson, M. (1997) Methods Enzymol. 277, 493–505
53. Schuck, P. (1997) Annu. Rev. Biophys. Biomol. Struct. 26, 541–566
54. Vaughan, D. E., and Bjorkman, P. J. (1997) Biochemistry 36, 9374–9380
55. McGhee, J. D., and von Hippeil, P. H. (1974) J. Mol. Biol. 86, 469–489
56. Cantor, C. R., and Schimmel, P. R. (1980) Biophysical Chemistry, W. H. Freeman and Co., New York, 80
57. Perona, J. A., Zhu, C.-H., Crast, C. S., and Fletterick, R. J. (1997) Biochemistry 36, 5381–5392
58. Lovejoy, B., Cleasby, A., Hassell, A. M., Longley, K., Luther, M. A., Weigl, D., McGeehan, G., McElroy, A. B., Drewry, D., Lambert, M. H., and Jordan, S. R. (1994) Science 263, 375–377

REFERENCES

1. Switalski, L. M., Patti, J. M., Butcher, W., Gristina, A. G., Spieziale, P., and Ho¨ök, M. (1993) J. Biol. Chem. 268, 2889–2906