Identification and Biochemical Characterization of the Ligand Binding Domain of the Collagen Adhesin from *Staphylococcus aureus*†

Joseph M. Patti,* Jeffrey O. Boles, and Magnus Höök

Institute of Biosciences and Technology, Center for Extracellular Matrix Biology, Texas A&M University, Houston, Texas 77030

Received June 1, 1993; Revised Manuscript Received August 10, 1993®

ABSTRACT: We have recently shown that the expression of a collagen adhesin is both necessary and sufficient to mediate the attachment of *Staphylococcus aureus* to cartilage, a complex collagen-containing substrate [Switalski, L. M., Patti, J. M., Butcher, W., Gristina, A. G., Speziale, P., & Höök, M. (1993) *Mol. Microbiol. 7*, 99–107]. We now report on the localization of the ligand binding site within the 135-kDa *S. aureus* collagen adhesin. Using deletion mutagenesis in combination with Western ligand blot and direct binding assays, the collagen binding domain (CBD) was localized to a 168 amino acid long segment [CBD(151–318)] within the N-terminal portion of the adhesin. Using biospecific interaction analysis, pepsin-digested bovine type II collagen was found to contain eight binding sites for CBD(151–318); two binding sites were of "high" affinity (K_d = 3 μM) and six sites were of low affinity (K_d = 30 μM). Short truncations in the terminal flanking regions of CBD(151–318) resulted in two CBDs (180–318 and 151–297) that lacked collagen binding activity. Analysis by circular dichroism of the recombinant CBDs in the far UV revealed similar secondary structures, predominantly β-sheet, whereas the near-UV spectra indicated dramatic changes in the degree of intermolecular packing (tertiary structure). The deduced amino acid sequence of the ligand binding domain of the collagen adhesin is presented.

Collagen is the major structural protein in vertebrates. The architecture of various tissues is dependent on the supramolecular organization of collagen, which is regulated by intricate interactions between different collagen molecules and non-collagenous components (Heinegård & Oldberg, 1989). A large number of genetically distinct collagen types has now been identified, and the corresponding genes have been cloned and sequenced (Mayne & Burgeson, 1987). Some of these collagen types have a tissue-specific distribution. For example, type II collagen is preferentially located in cartilage, while type IV collagen is found almost exclusively in basement membranes (Miller & Gay, 1987). Other collagens such as type I and type III collagens have a wide distribution pattern.

Not only are the collagens of importance for the structural integrity of the tissue, but they also exert important biological activities. Perhaps the best characterized activity is related to the ability of collagen to serve as substrata for some cell surface receptors of the integrin family. Thus, collagens may affect the adhesion (Gullberg et al., 1990) and migration (Hay, 1991) of cells as well as gene expression (Horton et al., 1987).

It is apparent that many pathogenic microorganisms have taken advantage of the availability of adhesive proteins present in the extracellular matrix of the host. Microorganisms have developed adhesins that, when expressed on their cell surface, allow their adhesion to the host substrata. We have previously demonstrated the presence of a collagen receptor on the surface of some *Staphylococcus aureus* cells (Switalski et al., 1989). Isolation and characterization of the corresponding *cna* gene revealed a protein that can occur in at least two forms with M_r's of 135 000 and 110 000, respectively, in different isolates of *S. aureus* (Switalski et al., 1993). Expression of the collagen adhesin was shown to be both necessary and sufficient to mediate the attachment of *S. aureus* strains to cartilage, a complex collagen-containing substrate. Furthermore, almost all *S. aureus* strains retrieved from patients with osteomyelitis or septic arthritis where bone and joint tissues are infected were found to express a collagen adhesin, whereas only one-third of the clinical isolates obtained from patients with only soft tissue infections expressed the collagen adhesin (Switalski et al., 1993). Taken together, these data suggest that the collagen adhesin may serve as a "homing device", in that it allows bacteria to colonize bone and cartilage and as such represents an important virulence factor in the pathogenesis of septic arthritis and osteomyelitis.

The study presented here was undertaken with the goals to localize the collagen binding site within the *S. aureus* adhesin and to develop a basis for future studies investigating the interactions of collagens with non-collagenous proteins.

EXPERIMENTAL PROCEDURES

Materials. Escherichia coli strains were grown at 37 °C in Luria broth medium (LB)[1] (Gibco BRL, Grand Island, NY). Ampicillin (100 μg/mL) (Sigma Chemical Co., St. Louis, MO) was added when appropriate. Restriction and DNA modification enzymes were purchased from United

† This study was supported by a grant from the NIH (HL-47313).

* Author to whom correspondence should be addressed. Phone: (713) 677-7556. Fax: (713) 677-7576.

® Abstract published in *Advance ACS Abstracts*, October 1, 1993.

[1] Abbreviations: AP, alkaline phosphatase; ATP, adenosine triphosphate; BCIP, 5-bromo-4-chloro-3-indolyl phosphate; BSA, bovine serum albumin; bp, base pair; CBD, collagen binding domain; dNTP, deoxynucleotide triphosphate; ECM, extracellular matrix; GST, glutathione-*S*-transferase; HBSS, Hanks' buffered salt solution; IPTG, isopropyl β-thiogalactoside; kb, kilobase; LB, Luria broth; MW, molecular weight; NBT, nitroblue tetrazolium; PCR, polymerase chain reaction; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide electrophoresis; TBS, Tris-buffered saline.

0006-2960/93/0432-11428$04.00/0

States Biochemical Corporation (USB, Cleveland, OH) or Gibco BRL. IPTG was purchased from 5 Prime–3 Prime, Inc. (West Chester, PA). Purified oligonucleotides were obtained from Oligos Etc. (Gilford, CT). All other chemicals were molecular biology grade from Sigma or USB.

Collagen. Chicken and bovine collagen type II were gifts from Dr. Richard Mayne of the University of Alabama at Birmingham. $Na^{125}I$, specific activity 15 mCi/μg, was obtained from Amersham Corp. (Arlington Heights, IL). Collagen was labeled with ^{125}I by the chloramine-T method (Hunter, 1978) as described previously (Switalski et al., 1989). Estimated specific activities of type II collagen ranged from 2×10^6 to 5×10^7 cpm/μg protein.

DNA Techniques. Standard molecular biology techniques, such as T4 DNA ligation, bacterial transformation, restriction digest analysis, plasmid purifications, and agarose gel electrophoresis, were conducted as described in Sambrook et al. (1989) or in *Current Protocols in Molecular Biology* (Ausubel (Ed.), 1989).

Isolation of the cna Gene Fragments. The polymerase chain reaction (PCR), together with the oligonucleotides listed in Table I, was used to amplify specific *cna* gene fragments from chromosomal DNA from *S. aureus* strain FDA 574. Genomic DNA was isolated from *S. aureus* cells as previously described (Marmur, 1961). To facilitate rapid cloning of the amplified *cna* gene fragments, oligonucleotides used for PCR were designed to contain a restriction enzyme cleavage site incorporated into the 5′-end of their sequences. PCR was performed with a Cetus/Perkin-Elmer DNA thermocycler (Perkin-Elmer Corp., Norwalk, CT).

Reaction mixtures (100 μL) contained 1 μM each primer, 200 μM each dNTP, 1 mM Tris-HCl (pH 8.3), 5 mM KCl, 0.15 mM $MgCl_2$, 0.001% gelatin, 3 μg of template DNA, and 2.5 units of AmpliTaq DNA polymerase (Perkin-Elmer Corp.). The reaction mixtures were overlaid with 100 μL of mineral oil and amplified for 30 cycles consisting of a 2-min denaturation period at 94 °C, a 2-min annealing period at 55 °C, and a 3-min extension period at 72 °C. After amplification, 10 μL of the PCR products was analyzed on a 1% agarose gel (SeaKem GTG, FMC Bioproducts Inc., Rockland, ME). The remaining *cna* gene PCR products were treated with proteinase K, phenol chloroform extracted, and ethanol precipitated as described (Crowe et al., 1991). The precipitated DNA was dissolved in distilled H_2O and digested with the appropriate restriction enzymes. The digested PCR products were purified by agarose gel electrophoresis according to the manufacturer's directions (Elu-Quik; Schleicher & Scheull, Keene, NH) and used in the construction of the expression plasmids described below.

Construction of Plasmids Expressing cna Gene Fragments. Recombinant *S. aureus* collagen adhesin fragments were overexpressed in *E. coli* using three different prokaryotic expression systems. The amino terminus, including the entire A domain, was amplified from *S. aureus* FDA 574 chromosomal DNA using PCR together with primers CNA 20 and CNA 21. The amplified 1.6-kb *cna* gene fragment was cleaved with *Eco*RI and *Pst*I, gel purified, and ligated to the prokaryotic expression vector pKK223-3 (Brosius & Holy, 1984), obtained from Pharmacia LKB Biotechnology (Piscataway, NJ), to create plasmid pKK1.6 (Figure 1A). Expression vector pKK223-3 contains an IPTG-inducible *tac* promoter adjacent to a consensus Shine–Dalgarno ribosomal binding site (DeBoer et al., 1982). However, this vector lacks an initiation codon; therefore, the DNA to be expressed must contain an appropriate start codon. In order to express an internal *cna* gene fragment, a DNA linker sequence containing

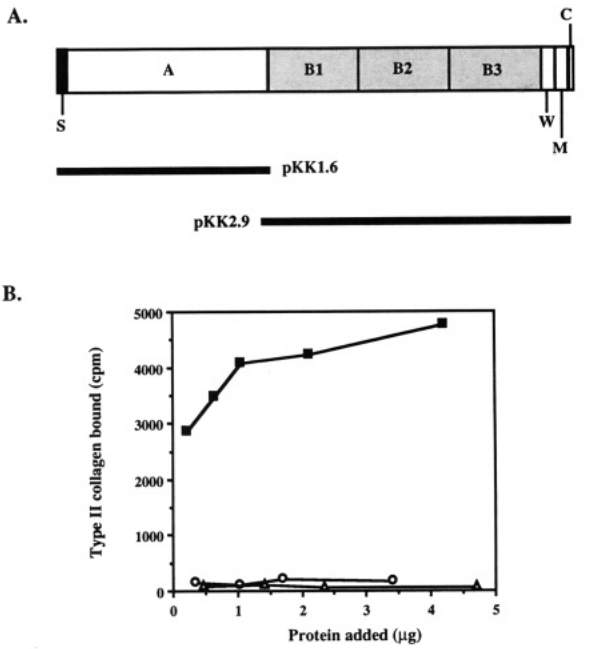

FIGURE 1: Identification of the collagen binding region. (A) Schematic model of the collagen adhesin from *S. aureus* FDA 574 as proposed previously (Patti et al., 1992). The location of the expressed collagen adhesin region is noted. (B) Microtiter wells were coated with the indicated amounts of lysate from pKK1.6(JM105) (■), pKK2.9(JM105) (○), or pKK223-3(JM105) (△), containing vector without the *cna* gene. The wells were incubated with type II [^{125}I]collagen, and the amount of bound ligand was determined by counting individual wells in a γ counter. For further details, see the Experimental Procedures section.

an ATG start codon was synthesized. Two partially complementary oligonucleotides, JP1 (5′-AATTACCATGGAATTCCTGCA-3′) and JP2 (5′-TGGTACCTTAAGG-3′), were heated to 70 °C and slowly cooled to allow annealing. Once annealed, the double-stranded linker was phosphorylated by the addition of ATP and T4 polynucleotide kinase. The DNA linker contained *Eco*RI and *Pst*I restriction sites at the 5′- and 3′-termini, respectively. These sites were used to insert the linker into pKK223-3. A 2.9-kb *Eco*RI/*Pst*I DNA fragment, originally isolated from λGT11 clone pCOL11 (Patti et al., 1992), was ligated to vector pKK223-3 to create plasmid pKK2.9. The collagen adhesin fragment encoded by pKK2.9 contains three repeated domains (B1, B2, and B3), the carboxyl terminus, and downstream sequences (Figure 1A).

In the second expression system, the glutathione-*S*-transferase (GST) gene fusion plasmid pGEX-2T (Pharmacia LKB Biotechnology) was used to generate plasmids which overproduce GST–collagen adhesin fusion proteins. A series of GST–CNA derivatives spanning the A domain of the collagen adhesin were constructed. PCR-amplified *cna* gene fragments were generated using the oligonucleotides listed in Table I. The amplified gene fragments that contain *Bam*HI and *Eco*RI restriction sites incorporated at the 5′- and 3′-ends, respectively, were digested with the appropriate restriction enzymes and ligated to the 3′-end of the previously digested GST gene using standard techniques. Expression plasmid pGEX-1.1 (Figure 2), containing a *cna* gene fragment with an internal *Hin*dIII site, was digested with both *Hin*dIII and *Eco*RI restriction enzymes. Double digestion of the pGEX-1.1 DNA, followed by a fill-in reaction to create blunt ends and subsequent religation of the plasmid, was used to create pGEX-800. The pGEX expression constructs generated are listed in Figure 2. A thrombin cleavage site present between the carrier protein and the collagen adhesin fragment may allow the removal of the GST from the fusion protein. However, efficient proteolysis at the cleavage site was often not obtained.

Table I: Synthetic Oligonucleotides Used To Amplify *cna* Gene Fragments from Genomic DNA Isolated from *S. aureus* Strain FDA 574[a]

	expression constructs	amino acids
CNA 20 5' ATATGAATTCGAGTATAAGGAAGGGGTT 3'	pKK1.6	1-530
CNA 21 5' TTTGGATCCGTTTTTCAGTATTAGTAACCA 3'		
CNA 22 5' CCGGGATCCCAAGCTTGGTATCAAGAG 3'	pGEX 324	296-403
CNA 23 5' GCGGAATTCTTACTTATCAAATGTATACGG 3'		
CNA 24 5' CGCGGATCCATGTGCTACCAAGAAGATACG 3'	pGEX 417	180-318
CNA 25 5' CCGAATTCTTAATTGTGCACAGTATGATT 3'	pQE 417	
CNA 26 5' CGCGGATCCCCAGGTTCTAAAATAACTGTT 3'	pGEX 468	248-403
CNA 23 5' GCGGAATTCTTACTTATCAAATGTATACGG 3'		
CNA 27 5' GCGGATCCATAACATCTGGGAATAAA 3'	pGEX 504	151-318
CNA 25 5' CCGAATTCTTAATTGTGCACAGTATGATT 3'	pQE 504	
CNA 28 5' GGGGATCCGAAAAATTAAGTGATGTTTCG 3'	pGEX 597	120-318
CNA 25 5' CCGAATTCTTAATTGTGCACAGTATGATT 3'		
CNA 24 5' CGCGGATCCATGTGCTACCAAGAAGATACG 3'	pGEX 672	180-403
CNA 23 5' GCGGAATTCTTACTTATCAAATGTATACGG 3'		
CNA 28 5' GGGGATCCGAAAAATTAAGTGATGTTTCG 3'	pGEX 673	120-343
CNA 29 5' CGCGAATTCCTTGGTATCTTTATCCTGTTT 3'		
CNA 30 5' TCCGGATCCGATGATGACGATAAAAATGGAAAA 3'	pGEX 849	61-343
CNA 29 5' CGCGAATTCCTTGGTATCTTTATCCTGTTT 3'	pQE 849	
CNA 30 5' TCCGGATCCGATGATGACGATAAAAATGGAAAA 3'	pGEX 1.0	61-403
CNA 23 5' GCGGAATTCTTACTTATCAAATGTATACGG 3'		
CNA 31 5' GCGGATCCGCACGAGATATTTCA 3'	pGEX 1.1	30-403
CNA 23 5' GCGGAATTCTTACTTATCAAATGTATACGG 3'		

[a] Restriction endonuclease sites are underlined. The amino acid residues and the corresponding expression constructs are denoted for each pair of primers.

A third expression system that places a histidine hexamer (HIS_6) at the N-terminus of the protein was used to minimize the length of the carrier protein. PCR-amplified *cna* gene fragments were generated using the oligonucleotides listed in Table I. The amplified gene fragments containing *Bam*HI and *Eco*RI restriction sites at the 5'- and 3'-ends, respectively, were ligated to pBS/SK(+) (Stratagene, La Jolla, CA) that had been linearized with the same enzymes. The resulting plasmids were used to transform *E. coli* JM101 cells. Clones containing the proper plasmid constructions were identified by restriction digest analysis. The *cna* gene fragments were removed from pBS/SK(+) by digestion with *Bam*HI and *Sal*I restriction endonucleases and subsequently purified from an agarose gel, as described previously. The gel-purified gene fragments containing *Bam*HI and *Sal*I restriction sites at the 5'- and 3'-ends, respectively, were then ligated to a *Bam*HI and *Sal*I site 3' of the segment encoding the poly(histidine) tag in the expression vector pQE-30 (QIAGEN, Inc.). This procedure was used to create expression plasmids pQE-849, pQE-504, and pQE-417 (Figure 2). As mentioned previously, the *cna* gene has an internal *Hin*dIII restriction site; taking advantage of this fact, genetic construct pQE-441 was obtained by the digestion of plasmid pQE-504 with *Hin*dIII followed by religation of the plasmid.

Preparation of Bacterial Lysates Containing Collagen Adhesin Protein Derivatives. Saturated overnight cultures of *E. coli* JM101 *supE, endA, sbcB*15, *hsd R*4, *rpsL, thi* Δ*(lac-proAB)* (F'*tra*D36 *proAB*$^+$ *lacI*^{q} *Z*ΔM15) or *E. coli* JM105 *supE thi* Δ(*lac-proAB*) (F'*tra*D36 *proAB*$^+$ *lacI*^{q} *Z*ΔM15) harboring expression plasmids were diluted 1:50 in LB supplemented with ampicillin and allowed to grow until the culture reached an OD_{600} of 0.6–0.7. IPTG (final concentration 0.2 mM) was added to the cells and growth continued for another 2.5 h at 37 °C. The bacteria were collected by centrifugation, and the bacterial pellets were resuspended in phosphate-buffered saline (PBS; 10 mM phosphate and 0.14 M NaCl, pH 7.4). The cells were lysed by passage through a French press twice at 20 000 lb/in^2. The bacterial lysate was centrifuged at 102000*g* for 10 min to remove bacterial debris. The supernatant containing the soluble proteins was filtered through a 0.45 μM membrane (Corning, Corning, NY) and retained for further purification.

Purification of Recombinant Collagen Adhesin Proteins. The GST–CNA fusion proteins were purified as described previously (Smith & Johnson, 1988). The HIS_6–CNA proteins were purified by immobilized metal chelate affinity chromatography (Porath et al., 1975, Hochuli et al., 1988). A column containing iminodiacetic acid/Sepharose 6B Fast Flow (Sigma), connected to an FPLC system, was charged with 150 mM Ni^{2+} and equilibrated with buffer A (5 mM imidazole, 0.5 M NaCl, and 20 mM Tris, pH 7.9). After equilibration, the bacterial supernatant was applied to the column, and the column was washed with 10 bed vol of buffer A. Subsequently, the column was eluted with buffer B (200 mM imidazole, 0.5 M NaCl, and 20 mM Tris, pH 7.9). The eluate was monitored for protein by the absorbance at 280 nm, and peak fractions were analyzed by SDS–PAGE. Positive fractions were pooled, dialyzed extensively to 5 mM sodium phosphate (monobasic), pH 7.5, and stored at −20 °C.

Analysis of Collagen Binding Activity by Expressed Collagen Adhesin Proteins. (A) Western Ligand Blot Assay. Proteins were fractionated on a 12% polyacrylamide gel and transferred to a nitrocellulose membrane (Schleicher & Scheull) for 1.5 h at 100 V in 20 mM Tris, 150 mM glycine, and 20% (v/v) methanol (Towbin et al., 1979). Additional protein binding sites were blocked by incubation of the membrane at room temperature for 1 h in 3% (w/v) nonfat dry milk in TBS. Collagen binding proteins were detected by incubating the membranes with chicken type II [^{125}I]collagen in 3% (w/v) nonfat dry milk/TBS (pH 8.0) overnight at 4 °C. The membrane was then washed three times in TBS containing 0.1% Tween 80, dried, and subjected to autoradiography.

(B) Microtiter Plate Assay. Immulon 2 (Dynatech Laboratories, Inc., Chantilly, VA) detachable wells were coated with increasing amounts of bacterial lysate (200 μL total volume) or purified proteins for 2 h at room temperature. A 1% solution of BSA was then added to the wells to block any additional protein binding sites on the well surface. After a 1-h incubation at room temperature, the wells were washed twice with PBS containing 0.01% Tween 80 and 0.1% BSA. Finally, 50 μL of chicken type II [^{125}I]collagen (3×10^4 cpm) was added to the wells, and the mixtures were incubated for 2 h at room temperature. The wells were washed twice and the amounts of radioactivity bound to the protein-coated wells

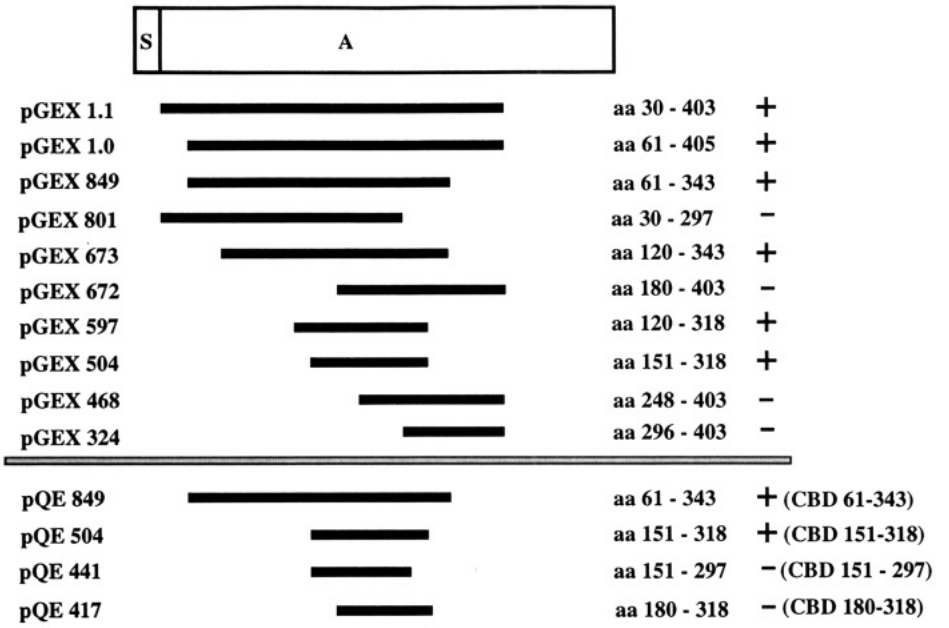

FIGURE 2: Localization of the collagen binding domain within the A region of the *Staphylococcus aureus* collagen adhesin. DNA fragments which span different segments of the A region of the collagen adhesin were generated by PCR. The amplified *cna* gene fragments were digested with the appropriate restriction enzymes, ligated either to the 3′-end of the glutathione-*S*-transferase gene (pGEX) or to the vector (pQE) containing a poly(histidine) tag, and expressed as fusion proteins. The locations of *cna* gene fragments with respect to the A domain and the amino acids represented are indicated for each genetic construct. Recombinant proteins which contain collagen binding activity are denoted by a +, and proteins that do not possess binding activity are indicated by a −.

were determined in a γ-counter. All binding assays were done in duplicate.

Circular Dichroism. The average secondary structures of several recombinant proteins were monitored by circular dichroism (CD) spectroscopy on a Jasco J720 spectropolarimeter calibrated with a 0.1% (w/v) 10-camphorsulfonic acid-*d* solution. CD spectra were measured at 25 °C in a 0.2 mm path length quartz cell, and four scans from 250 to 190 nm (far UV) were generated and averaged. Protein concentrations determined by amino acid analysis were typically 30 μM in 5 mM sodium phosphate (monobasic), pH 7.5. CD spectra were determined between 320 and 250 nm (near UV) utilizing a 1.0 cm path length quartz cell. Protein concentrations of 70 μM were used for the observation of near-UV CD. The molar ellipticity was expressed in deg·cm^2/mol.

Determination of Equilibrium Constants. Real-time biospecific interaction analysis (BIA) was used to assess affinity constants for the interaction of collagen binding domains of the *S. aureus* adhesin with bovine type II collagen. Surface plasmon resonance (SPR) measurements were performed utilizing the BIAcore system (Pharmacia Biosensor AB, La Jolla, CA). Sensor chip CM5 and an Amine Coupling Kit for the immobilization of bovine type II collagen were also obtained from Pharmacia Biosensor AB. Briefly, the detector monitors the interaction of analyte with a biospecific sensor surface (to which biological molecules are covalently bound, i.e., type II collagen), using the optical phenomenon SPR (Liedberg et al., 1983). The binding of an analyte present in solution to the immobilized ligand may be monitored by the change in the SPR signal (reflecting the change in surface concentration) in real time. A plot of maximum (equilibrium) response vs analyte concentration results in the typical saturation profile used to determine affinity constants. SPR response is measured in resonance units (RU). One thousand RU corresponds to a surface concentration of 1 ng/mm^2. Because of the direct correlation between SPR response and the concentration of bound ligand, we were able to use maximum equilibrium response values to determine molar binding ratios of CBD–collagen interactionsusing the following equation:

$$\frac{\text{CBD}}{\text{type II collagen}} = \left(\frac{\text{CBD RU}}{\text{type II collagen RU}}\right)\left(\frac{\text{MW type II collagen}}{\text{MW CBD}}\right)$$

where MW type II collagen = 3.0×10^5 and MW CBD = 1.9×10^4. These data were collected for a series of CBD concentrations representing 10–90% saturation of the CBD binding sites in type II collagen. Data representing 10–50% saturation of binding sites were then utilized to construct a Scatchard plot (Scatchard, 1949).

Preparation of Sensor Chip Surfaces. Pepsin-treated bovine type II collagen was covalently coupled to sensor chip CM5 via primary amine groups using the following conditions: Carboxymethylated dextran (CM5 surface) was activated by derivatization with 100 mM N-hydroxysuccinimide (NHS) mediated by treatment with 400 mM *N*-ethyl-*N*′-[(dimethylamino)propyl]carbodiimide (EDC) for 6.0 min. Generally, 40% of the carboxyl groups on the dextran surface were converted to reactive *N*-hydroxylsuccinimide esters during activation. The NHS ester readily reacts with uncharged primary amino groups on collagen. After activation, a 50 μg/mL solution of bovine type II collagen in 10 mM maleate buffer (pH 6.0) was passed over the surface, which resulted in 5500 RU bound (approx. 5 ng/mm^2). In a second experiment using a 2.0-min activation period, the amount of collagen immobilized corresponded to 2100 RU (2.1 ng/mm^2). Noncovalently associated collagen was washed off with elution buffer (HBS; 150 mM NaCl, 50 mM HEPES, and 10% P-20 surfactant, pH 7.4), and unreacted sites on the dextran surface were blocked by treatment with 1 M ethanolamine hydrochloride for 6 min.

RESULTS

A collagen binding adhesin has previously been isolated from *S. aureus* strain Cowan (Switalski et al., 1989). Cloning and sequencing of the corresponding gene *cna* (Patti et al.,

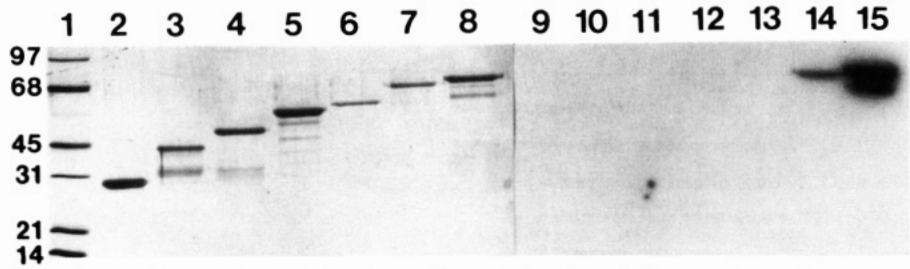

FIGURE 3: SDS–PAGE and autoradiographic analysis of recombinant GST–CNA fusion proteins. Affinity-purified fusion proteins were fractionated by electrophoresis in a 10% polyacrylamide gel, and proteins were stained with Coomassie Brilliant Blue. Migration distances and molecular massess of the standard proteins (lane 1) are indicated by the numbers on the left: lane 2, pGEX-2T; lane 3, pGEX-324; lane 4, pGEX-468; lane 5, pGEX-672; lane 6, pGEX-801; lane 7, pGEX-1.0; lane 8, pGEX-1.1. The proteins were electroblotted to a nitrocellulose membrane and probed with type II [^{125}I]collagen: lane 9, pGEX-2T; lane 10, pGEX-324; lane 11, pGEX-468; lane 12, pGEX-672; lane 13, pGEX-801; lane 14, pGEX-1.0; lane 15, pGEX-1.1.

1992) revealed a 133-kDa polypeptide, a size that correlates closely with the 135 kDa reported for the native protein. A model of the *S. aureus* collagen adhesin as previously proposed (Patti et al., 1992) is shown in Figure 1A. The protein is proposed to consist of a signal sequence (S) that is followed by a large nonrepetitive region (A). Immediately following the A region are three consecutive repeats of a 167 amino acid long unit (B1, B2, and B3). A cell wall (W) region consisting of a 64 amino acid proline- and lysine-rich domain is followed by a stretch of hydrophobic amino acids (M), presumably constituting the cell membrane spanning region. Finally, the extreme carboxyl terminus (C) is made up of a few positively charged amino acids. This model was used as the starting point in an attempt to identify the collagen binding domain.

The Collagen Binding Domain is Contained within the A Region. Initial experiments were designed to identify which region of the collagen adhesin contained the collagen binding site. Plasmid pKK1.6 encodes the amino-terminal portion of the collagen adhesin, including the A domain and plasmid pKK2.9 which encodes the C-terminal portion of the adhesin from the three repeated domains (B1, B2, and B3) to the carboxyl terminus, and downstream sequences were initially constructed (Figure 1A). Bacterial lysates were prepared from IPTG-induced *E. coli* cells harboring the expression plasmids pKK1.6(JM105) and pKK2.9(JM105). Detachable polystyrene microtiter wells were coated with increasing amounts of bacterial lysate, and type II [^{125}I]collagen was added to the immobilized proteins. Collagen binding activity was quantitated as described in Experimental Procedures. The lysate from pKK1.6(JM105) bound type II collagen in a dose-dependent manner, while lysates from pKK2.9(JM105) and cells containing the vector without insert pKK223-3(JM105) did not bind type II collagen above background levels (Figure 1B). Expression of collagen adhesin domains in *E. coli* containing the expression plasmids pKK1.6(JM105) and pKK2.9(JM105) was confirmed by immunoblot analysis using a monospecific antibody against the collagen adhesin (Switalski et al., 1989) (data not shown). These data indicate that the collagen binding site is contained within the A domain of the collagen adhesin.

Localization of the Collagen Binding Domain. To map the collagen binding domain (CBD) more precisely within the collagen adhesin, a series of deletion mutants was constructed. Segments of the *cna* gene were generated by the polymerase chain reaction. Together, the amplified segments span the part of the gene encoding the A domain (Figure 2). The amplified PCR gene products were ligated to the 3′-end of the glutathione-*S*-transferase (GST) gene from *Schistosoma japonicum* (Smith et al., 1986) contained on the prokaryotic plasmid vector pGEX-2T. The constructed pGEX-2T derivatives were used to transform *E. coli* JM101 cells. Expression of the GST fusion proteins was induced by the addition of 0.2 mM IPTG. Subsequently, the bacteria were lysed and the GST fusion proteins were purified by affinity chromatography on glutathione–Sepharose (Smith & Johnson, 1988). SDS–PAGE analysis was used to analyze purified fusion proteins (for examples, see Figure 3, lanes 2–8).

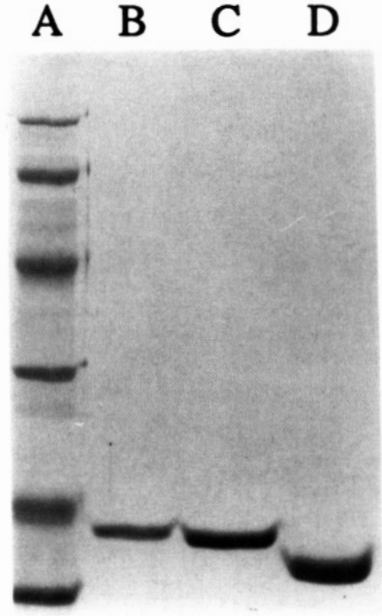

FIGURE 4: SDS–PAGE of recombinant HIS_6–CNA fusion proteins. Affinity-purified fusion proteins were fractionated by electrophoresis on a 15% polyacrylamide gel, and the proteins were stained with Coomassie Brilliant Blue: lane A, molecular weight standards; lane B, CBD(151–318); lane C, CBD(151–297); and lane D, CBD(180–318).

To determine collagen binding activity, the expressed fusion proteins were fractionated by SDS–PAGE , transferred to a nitrocellulose membrane, and probed with type II [^{125}I]collagen. Collagen binding proteins were detected by autoradiography. Constructs pGEX1.0(JM101), which encodes amino acids 61–403 (Figure 3, lane 14), and pGEX1.1(JM101), which encodes amino acids 30–403 of the collagen adhesin (Figure 3, lane 15), expressed a fusion protein which retained collagen binding activity. Fusion proteins constructed in this series and shown in Figure 3 that contained shorter segments of the collagen adhesin (Figure 3, lanes 9–13) failed to bind the type II [^{125}I]collagen. These data indicate that segments of the collagen adhesin fused to the GST carrier protein may retain collagen binding activity. Subsequently, additional constructs were made in the pGEX-2T vector with the goal of identifying the smallest segment of the collagen adhesin that possessed collagen binding activity. The different collagen adhesin segments expressed and their collagen binding activities are depicted schematically in Figure 2. The fusion proteins encoded by pGEX-504 contained the shortest segment

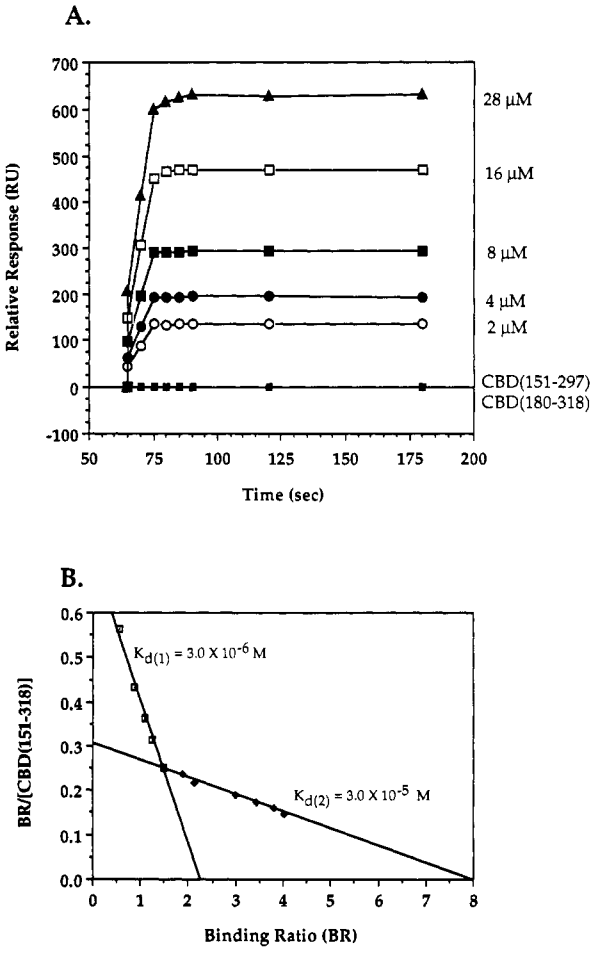

FIGURE 5: BiaCore analysis of the recombinant protein fragments of the collagen adhesin. (A) Sensorgrams of the recombinant collagen binding domain interaction with immobilized type II collagen. Various amounts of CBD(151–318), indicated by the numbers on the right, were passed over immobilized collagen. No response was detected with CBD(151–297) or CBD(180–318) at 30 and 6 μM, respectively. (B) Analysis of the binding of CBD(151–318) to type II collagen by a Scatchard plot. The plot represents 10–60% saturation of binding sites of the data collected from equilibrium response curves. The data presented represent the results of three independent determinations and four independent collagen immobilizations.

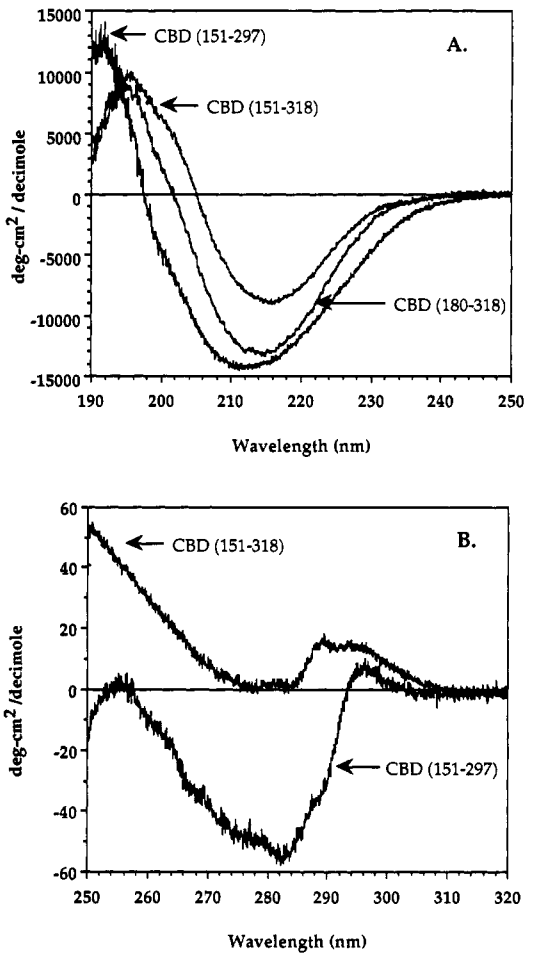

FIGURE 6: Circular dichroism of recombinant protein fragments of the collagen adhesin. (A) Far-UV spectra illustrating the overall similarity in secondary structure between the recombinant constructs. (B) Near-UV spectra illustrating a dramatic change in CD as a result of deleting 21 amino acids from the C-terminus of CBD(151–318) to generate CBD(151–297).

of the collagen adhesin, corresponding to amino acids 151–318, that still retained collagen binding activity. Fusion proteins containing segments of the collagen adhesin beginning at amino acid 180 (pGEX-672) or ending at amino acid residue 297 (pGEX-801) did not bind type II [^{125}I]collagen under the experimental conditions employed.

We were concerned that the large (27 kDa) GST carrier protein could affect the collagen binding ability of the expressed fusion proteins. A thrombin cleavage site has been engineered at the 3′-end of the carrier protein. However, attempts to cleave the generated fusion proteins with thrombin were largely unsuccessful. An alternative expression system (pQE) was therefore employed for the production of some selected constructs. This expression system produces proteins with a histidine hexamer at the N-terminus, allowing the rapid purification of the recombinant proteins by metal chelate chromatography. Four constructs encoding the following collagen binding domain segments of the collagen adhesin were designed using this system: 61–343, 151–318, 151–297, and 180–318. The recombinant proteins were purified as described in the Experimental Procedures. SDS–PAGE analysis of the purified proteins is shown in Figure 4.

Equilibrium Constant Determinations. The ability of the different CBD constructs to bind collagen was further investigated using the BIAcore system. Bovine type II collagen was immobilized on a sensor chip, and the purified proteins from the different CBD constructs were passed over the derivatized chip. CBD(61–343) (data not shown) and CBD(151–318) bound to the immobilized collagen, whereas CBD(151–297) and CBD(180–318) did not bind (Figure 5A). Analysis of CBD(151–318) at varying concentrations resulted in the SPR response curves shown in Figure 5A. When the generated binding data were analyzed by the method of Scatchard, a biphasic curve was obtained (Figure 5B), indicating the presence of two classes of binding sites in the immobilized collagen. If we assume that the immobilized bovine type II collagen is in a triple-helical conformation and has an M_r of 300 000, calculations indicate the presence of two binding sites of the high affinity type with an estimated K_d of 3×10^{-6} M and six binding sites of low affinity with an estimated K_d of 3×10^{-5} M. Hill plots constructed from the binding data indicated no cooperativity in CBD(151–318) binding type II collagen, further supporting the presence of two independent classes of binding sites (data not shown).

Circular Dichroism Analysis. The observation that short truncations of CBD(151–318) at either the N- or C-terminus resulted in the loss of collagen binding activity may be the consequence of conformational changes in the expressed protein rather than a deletion of the collagen binding sequence. The three recombinant proteins CBD(151–318), CBD(151–297), and CBD(180–318) were therefore analyzed by CD. The secondary structure of all three constructs appears to be

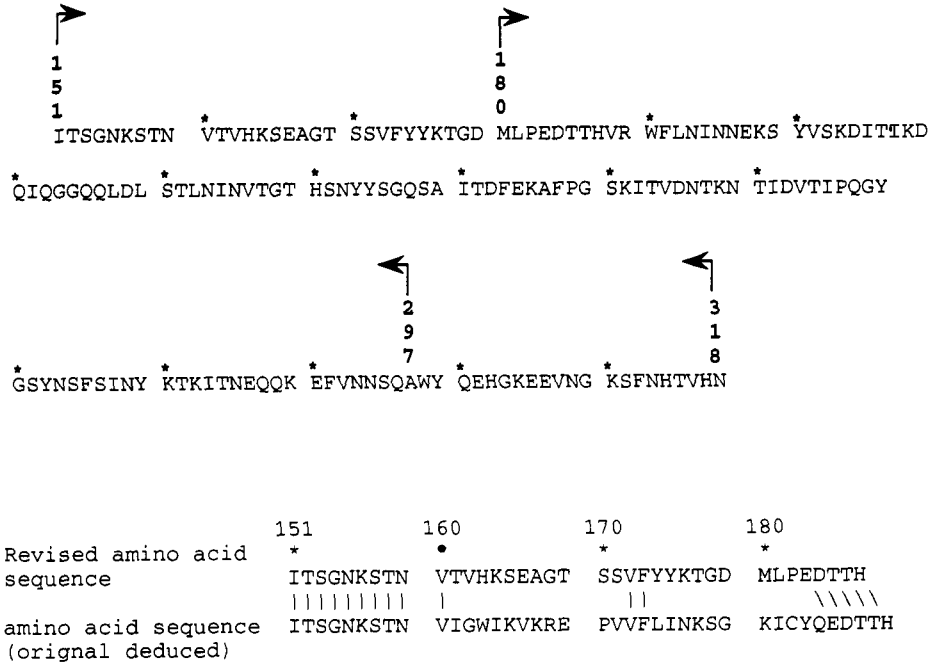

FIGURE 7: Figure 7. Amino acid sequence of the ligand binding domain of the *S. aureus* collagen adhesin. Deduced amino acid sequences of CBD(151–318), CBD(151–297), and CBD(180–318) encoded by the *cna* gene from *S. aureus* FDA 574. It should be noted that, due to an error in the initial DNA sequencing, amino acid residues 161–171 and 174–184 are incorrect as originally reported (Patti et al., 1992). The original deduced amino acid sequence in comparison with the corrected sequence is also depicted. The corrected deduced sequence shown here was confirmed by direct amino acid sequencing of residues 151–180 from CBD(151–318).

predominantly β-sheet, as indicated by the far-UV spectra (Figure 6A). The near-UV CD spectra indicated that CBD(151–318) has a much higher degree of intermolecular packing (tertiary structure) than CBD(151–297) (Figure 6B). CBD(151–318) exhibits maxima at both 250–260 and 289.5 nm, which is considered characteristic for tyrosine or phenylalanine side chains and tryptophan side chains, respectively. Conversely, CBD(151–297) exhibits significantly reduced intensity at 250–260 and 290 nm, suggesting that the environments of the aromatic residues are markedly different in the two recombinants, possibly reflecting differences in tertiary conformations. CBD(180–318) had low solubility, and a near-UV spectrum could not be obtained.

DISCUSSION

Recent investigations have focused on the identification of matrix binding sites within putative staphylococcal adhesins. *S. aureus* coagulase is an $M_r \approx 87\,000$ extracellular protein that may occur in a cell-associated form (Bodén et al., 1989) and induces the polymerization of fibrinogen to form clots in mammalian plasma (Jeljasewicz et al., 1983). Coagulase can bind both prothrombin and fibrinogen independently of each other. Using *lacZ-coa* fusions, McDevitt and co-workers recently localized the fibrinogen binding domain to the 177 C-terminal amino acids (McDevitt et al., 1992). The prothrombin binding domain of coagulase was previously localized to the N-terminal region (Kawabata et al., 1986).

The ligand binding domain of the *S. aureus* fibronectin receptor has also been determined. Flock and colleagues expressed a fusion protein (ZZ–Fr) consisting of the promoter, signal sequence, and immunoglobulin binding regions from protein A ligated to a 600-nucleotide fragment of the *fnbp* gene (Flock et al., 1986). The recombinant fusion protein expressed in *E. coli* possessed fibronectin binding activity. The fibronectin binding domain was further localized to a 38 amino acid unit that is repeated three times within the carboxyl-terminal portion of FNBP (Signäs et al., 1989). Synthetic peptides mimicking each of the repeated units (called D) inhibited the binding of the [^{125}I]fibronectin to *S. aureus*. Furthermore, chemical modification studies of the repeated D3 unit within the fibronectin receptor have indicated that acidic residues within amino acids 21–33 are essential for fibronectin binding (McGavin et al., 1991).

In the present study, the ligand binding domain of the *S. aureus* collagen adhesin was initially located to the A region. Molecular characterization of the *cna* gene has shown the 1.6-kb segment encoding the A domain to be conserved among clinical *S. aureus* isolates that bind collagen type II, whereas the carboxyl half of the collagen adhesin can contain either two or three B repeats of 167 amino acids (Switalski et al., 1993).

To further localize the collagen binding site, a series of *cna* gene fragments spanning segments of the A domain was generated by PCR. The amplified *cna* gene fragments were expressed in *E. coli* as either glutathione-*S*-transferase fusion proteins or polypeptides that contained a histidine hexamer tag. The shortest segment expressed as a fusion protein that retained collagen binding activity contained amino acid residues 151–318. The amino acid sequence of this segment is shown in Figure 7. The binding of these recombinant domains was analyzed by three different methods. In the BIAcore system, the soluble recombinant domains were assayed for their ability to bind immobilized collagen. In a Western blot type assay, soluble [^{125}I]collagen was analyzed for the ability to bind the immobilized recombinant domains. In a third assay (data not shown), recombinant domains immobilized on microtiter wells bound soluble [^{125}I]collagen in a dose-dependent manner. The results obtained with these three assays are consistent with one exception. When CBD(151–318) was expressed as a GST fusion, it bound [^{125}I]collagen in the Western blot assay; however, when expressed as a poly(histidine) fusion protein, it did not bind the soluble [^{125}I]collagen. These results suggest that immobilization of CBD(151–318) on a supporting substrate after being subjected to denaturing conditions may result in

improper refolding of the collagen binding domain when the additional protein sequence is not present. Short truncations of either the N-terminal [CBD(180–318)] or the C-terminal [CBD(151–297)] portion of CBD(151–318) resulted in the loss of collagen binding activity. One possible explanation for this finding is that the collagen binding domain is conformationally specific, in which a segment of CBD(151–318) is directly involved in binding to collagen and flanking protein sequences are necessary to maintain proper conformation.

CD analyses of CBD(151–318), CBD(151–297), and CBD(180–318) indicated differences in secondary structure and dramatic differences in tertiary structure, although the proteins were essentially identical (100% in sequence, 90% in overall homogeneity). The near-UV CD spectrum of CBD(151–198) was indicative of a protein containing a low degree of intermolecular packing, as compared to the very high intramolecular packing observed for CBD(151–318). Additionally, CBD(151–318) demonstrated greater solubility (50 mg/mL in 5 mM sodium phosphate, pH 7.5) than both CBD(151–297) and CBD(180–318), whose observed solubilities were determined to be 2 and 0.5 mg/mL, respectively. Taken together, these observations support the idea that the shorter constructs are subjected to an impaired folding that is incompatible with collagen binding. Thus, an active collagen binding domain is conformationally specific, and any perturbations of the conformation of this domain may result in the loss of collagen binding activity.

Scatchard analysis revealed the presence of eight binding sites for CBD(151–318) on a type II collagen triple-helical monomer. Of these eight sites, two appear to bind CBD(151–318) with 10-fold higher affinity than the remaining six sites. These data would indicate the presence of a distinct number of binding sites on type II collagen and contradict earlier results indicating that generic collagen structures are recognized by the adhesin (Speziale et al., 1986). The collagen receptor, as originally purified from heat treated *S. aureus* strain Cowan, did not bind collagen with an appreciable affinity, nor could it inhibit the binding of type II [^{125}I]collagen to bacterial cells (Switalski et al., 1989). The conformation of the native receptor was possibly altered or destroyed in the heat treatment of the bacteria or during the isolation and purification procedure. On the other hand, the recombinant collagen adhesin expressed in *E. coli* and purified under nondenaturing conditions can be purified by collagen–Sepharose chromatography and binds soluble collagen directly. Furthermore, it inhibits the binding of *S. aureus* to cartilage (Switalski et al., 1993). It is therefore possible that, in prior studies with heat-killed *S. aureus* cells, the collagen adhesin was partly denatured and did not exhibit the restricted binding specificity indicated in the current studies. We are currently undertaking an in-depth analysis of different forms of the collagen adhesin with various types of collagen to better understand the nature of this interaction.

ACKNOWLEDGMENT

We thank N. Garza and A. Höök for their technical assistance, Dr. T. Baldwin and Dr. A. Chaffotte (Department of Biochemistry, Texas A&M University) for their helpful-discussions concerning the CD analysis, and Dr. P. Neame (Shriners Hospital, Tampa, FL) for protein sequencing.

REFERENCES

Ausubel, F. A., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., & Struhl, K. (Eds.) (1991) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York.

Bodén, M. K., & Flock, J.-I. (1989) *Infect. Immunol. 57*, 2358–2363.

Brosius, J., & Holy, A. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6929–6933.

Crowe, J. S., Cooper, H. J., Smith, M. A., Sims, M. J., Parker, D., & Gewert, D. (1991) *Nucleic Acids Res. 19*, 184.

de Boer, H. A., Comstock, L. J., & Vasser, M. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 22–25.

Flock, J.-I., Fröman, G., Jönsson, K., Guss, B., Signäs, C., Nilsson, B., Raucci, G., Höök, M., Wadström, T., & Lindberg, M. (1987) *EMBO J. 6*, 2351–2357.

Gullberg, D., Turner, D. C., Terracio, L., Borg, T. K., & Rubin, K. (1990) *Exp. Cell Res. 190*, 254–264.

Hay, E. D. (1991) *Cell Biology of Extracellular Matrix*, 2nd ed., Plenum Press, New York.

Heinegård, D., & Oldberg, Å. (1989) *FASEB J. 3*, 2042–2051.

Hochuli, E., Döbeli, H., & Schacher, A. (1988) *J. Chromatogr. 411*, 177–184.

Horton, W., Miyashita, T., Kohno, K., Hassell, J. R., & Yamada, Y. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8864–8868.

Hunter, W. M. (1978) *Handbook of Experimental Immunology* (Weir, D. M., Ed.) pp 14.1–14.40, Blackwell Scientific Publications Ltd., Oxford, UK.

Kawabata, S., Miyata, T., Morita, T., Miyata, T., Inagawa, S., & Igarashi, H. (1986) *J. Biol. Chem. 261*, 527–531.

Liedberg, B., Nylander, C., & Lundström, I. (1983) *Sens. Actuators 4*, 299–304.

Marmur, J. (1961) *J. Mol. Biol. 3*, 208–218.

Mayne, R., & Burgeson, R. E. (1987) *Structure and Function of Collagen Types*, Academic Press, Orlando, FL.

McDevitt, D., Vaudaux, P., & Foster, T. J. (1992) *Infect. Immunol. 60*, 1514–1523.

McGavin, M. J., Raucci, G., Gurusiddappa, S., & Höök, M. (1991) *J. Biol. Chem. 266*, 8343–8347.

Miller, E. J., & Gay, S. (1987) *Methods Enzymol. 144*, 3–41.

Patti, J. M., Jonsson, H., Guss, G., Switalski, L. M., Wiberg, K., Lindberg, M., & Höök, M. (1992) *J. Biol. Chem. 267*, 4766–4772.

Porath, J., Carlsson, J., Olsson, I., & Belfrage, G. (1975) *Nature 258*, 598–599.

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

Scatchard, G. (1949) *Ann. N.Y. Acad. Sci. 51*, 660–672.

Signäs, C., Raucci, G., Jönsson, K., Lindgren, P.-E., Anantharamaiah, G. M., Höök, M., & Lindberg, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 699–703.

Smith, D. B., & Johnson, K. S. (1988) *Gene 67*, 31–40.

Smith, D. B., Davern, K. M., Board, P. G., Tiu, W. U., Garcia, E. G., & Mitchell, G. F. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8703–8707.

Speziale, P., Raucci, G., Viasi, L., Switalski, L. M., Timpl, R., & Höök, M. (1986) *J. Bacteriol. 167*, 77–81.

Switalski, L. M., Speziale, P., & Höök, M. (1989) *J. Biol. Chem. 264*, 21080–21086.

Switalski, L. M., Patti, J. M., Butcher, W., Gristina, A. G., Speziale, P., & Höök, M. (1993) *Mol. Microbiol. 7*, 99–107.

Towbin, H., Staehlin, T., & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4350-4354.