

Novel fold and assembly of the repetitive B region of the *Staphylococcus aureus* collagen-binding surface protein

Champion CS Deivanayagam¹, Rebecca L Rich², Mike Carson¹, Rick T Owens², Sita Danthuluri¹, Todd Bice¹, Magnus Höök² and Sthanam VL Narayana^{1*}

Background: The *Staphylococcus aureus* collagen-binding protein Cna mediates bacterial adherence to collagen. The primary sequence of Cna has a non-repetitive collagen-binding A region, followed by the repetitive B region. The B region has one to four 23 kDa repeat units (B₁–B₄), depending on the strain of origin. The affinity of the A region for collagen is independent of the B region. However, the B repeat units have been suggested to serve as a 'stalk' that projects the A region from the bacterial surface and thus facilitate bacterial adherence to collagen. To understand the biological role of these B-region repeats we determined their three-dimensional structure.

Results: B₁ has two domains (D₁ and D₂) placed side-by-side. D₁ and D₂ have similar secondary structure and exhibit a unique fold that resembles but is the inverse of the immunoglobulin-like (IgG-like) domains. Comparison with similar immunoglobulin superfamily (IgSF) structures shows novel packing arrangements between the D₁ and D₂ domains. In the B₁B₂ crystal structure, an omission of a single glycine residue in the D₂–D₃ linker loop, compared to the D₁–D₂ and D₃–D₄ linker loops, resulted in projection of the D₃ and D₄ in a spatially new orientation. We also present a model for B₁B₂B₃B₄.

Conclusions: The B region of the Cna collagen adhesin has a novel fold that is reminiscent of but is inverse in nature to the IgG fold. This B region assembly could effectively provide the needed flexibility and stability for presenting the ligand binding A region away from the bacterial cell surface.

Introduction

Bacterial binding to extracellular matrix proteins such as fibrinogen [1], fibronectin [2], vitronectin [3] and collagen [4] is an important mechanism of host-tissue adherence and is considered to be the first step in the pathogenic process of many infections [5]. Many MSCRAMMs (microbial surface components recognizing adhesive matrix molecules) have been identified in the past decade (for reviews see [5–7]). Sequence analyses of the MSCRAMMs revealed a common structural theme composed of unique and repeated sequences. A multidomain organization was proposed for these proteins, with a specific ligand-binding activity being ascribed to some subdomains. The *Staphylococcus aureus* collagen-binding MSCRAMM, Cna, binds to a variety of collagens with moderate affinity [8]. Molecular characterization of the Cna gene revealed the presence of a 55 kDa nonrepetitive A region (previously denoted as the A domain [9]) and the repetitive B region that consists of multiple 23 kDa B-repeat units (previously denoted as the B domain [9]; Figure 1).

Cna mediates bacterial adherence to collagenous tissues such as cartilage, a process that is important in the

Addresses: ¹Center for Macromolecular Crystallography, School of Optometry, University of Alabama at Birmingham, 250 BHSB, 1918 University Boulevard, Birmingham, AL 35294-0005, USA.

²Center for Extracellular Matrix Biology, Albert B. Alkek Institute of Biosciences and Technology, Texas A&M University, 2121 West Holcombe Boulevard, Houston, TX 77030-3303, USA.

*Corresponding author.

E-mail: narayana@pearl.cmc.uab.edu

Key words: collagen, Cna, IgSF, MSCRAMM, *Staphylococcus aureus*

Received: 26 July 1999

Revisions requested: 31 August 1999

Revisions received: 30 September 1999

Accepted: 21 October 1999

Published: 22 December 1999

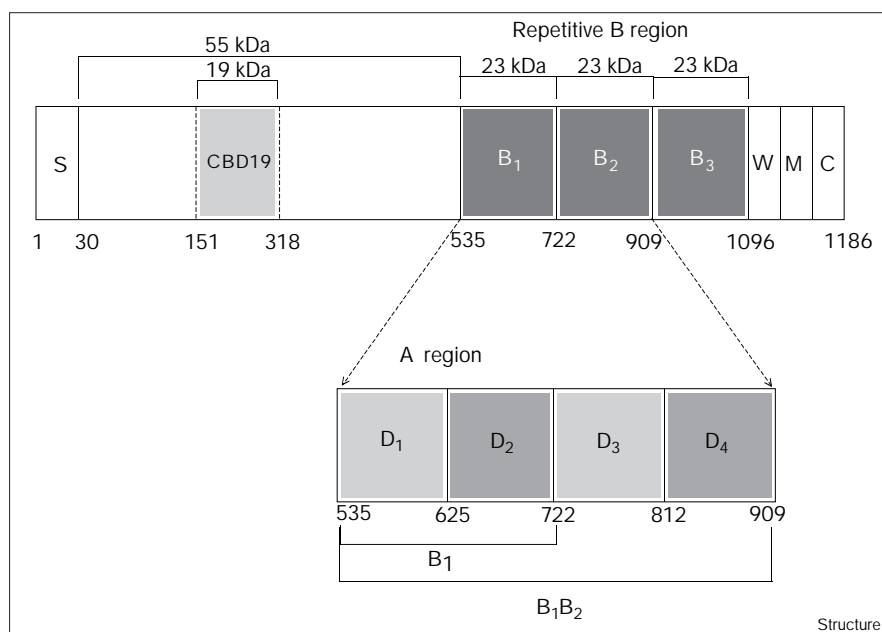
Structure 2000, 8:67–78

0969-2126/00/\$ – see front matter

© 2000 Elsevier Science Ltd. All rights reserved.

pathogenesis of septic arthritis caused by *Staphylococci*. An isogenic mutant in which the Cna gene was inactivated was considerably less virulent than the wild-type Cna-expressing bacteria in a mouse model of septic arthritis [10]. In addition, an *S. aureus* strain that normally lacked the Cna gene became more virulent in this model when the Cna gene was introduced. The collagen-binding activity was attributed to the Cna A region. A 19 kDa portion of the A region, Cna151–318 (previously denoted as CBD19 [collagen-binding domain 19] [11]), with measurable collagen-binding activity, was used in both biochemical and X-ray crystallographic studies to illustrate the protein's mode of binding to collagen. The crystal structure of Cna151–318 was determined in our laboratory, and consisted of two parallel β sheets connected by two short α helices [11]. One of the sheets had a noticeable trench traversing its surface. Molecular modeling studies were helpful in docking a collagen triple-helical probe [(Gly-Hyp-Pro)₄]₃ (where Hyp-Pro is hydroxy-proline) into this trench. Analyses of single point mutations performed in and around the trench confirmed it to be the binding site for collagen. Recently, a similar binding scheme has also been suggested for other collagen-binding proteins [12–16].

Figure 1



Schematic representation of the *Staphylococcus aureus* strain FDA574 collagen adhesin (Cna). The signal peptide (S) is followed by the nonrepetitive A region. CBD19 represents the smallest fragment retaining measurable collagen-binding activity for which the crystal structure is known [11]. This is followed by the repetitive B region, cell wall (W), membrane-spanning (M) domain and charged C terminus (C). D₁, D₂, D₃ and D₄ are the four domains observed in the crystal structures of B₁ and B₁B₂.

The B region of Cna in *S. aureus* strain FDA574 is a triplet of 23 kDa repeat units B₁, B₂ and B₃. Gillaspay *et al.* reported that the B region is composed of one to four nearly identical 23 kDa repeats, depending on the *S. aureus* strain from which the Cna was isolated [17]. Earlier studies indicated that the B region neither binds collagen nor influences the A region's collagen-binding activity [18]. These B region repeat units have been proposed to serve as a 'stalk' that projects the A region away from the bacterial cell surface and positions it for binding to collagen [18].

In our quest to understand the significance of the repetitive B region in the architecture of the *S. aureus* collagen adhesin, in this paper we describe the crystal structures of B₁ and the covalently linked B₁B₂ dimer, contrast their unique fold to the immunoglobulin (IgG) domains, and compare their structures with similar multidomain crystal structures. In addition, we present modeling studies of multiple B region repeat units based on the B₁ and B₁B₂ crystal structures, and highlight some of the important structural features.

Results

Crystal structure of B₁

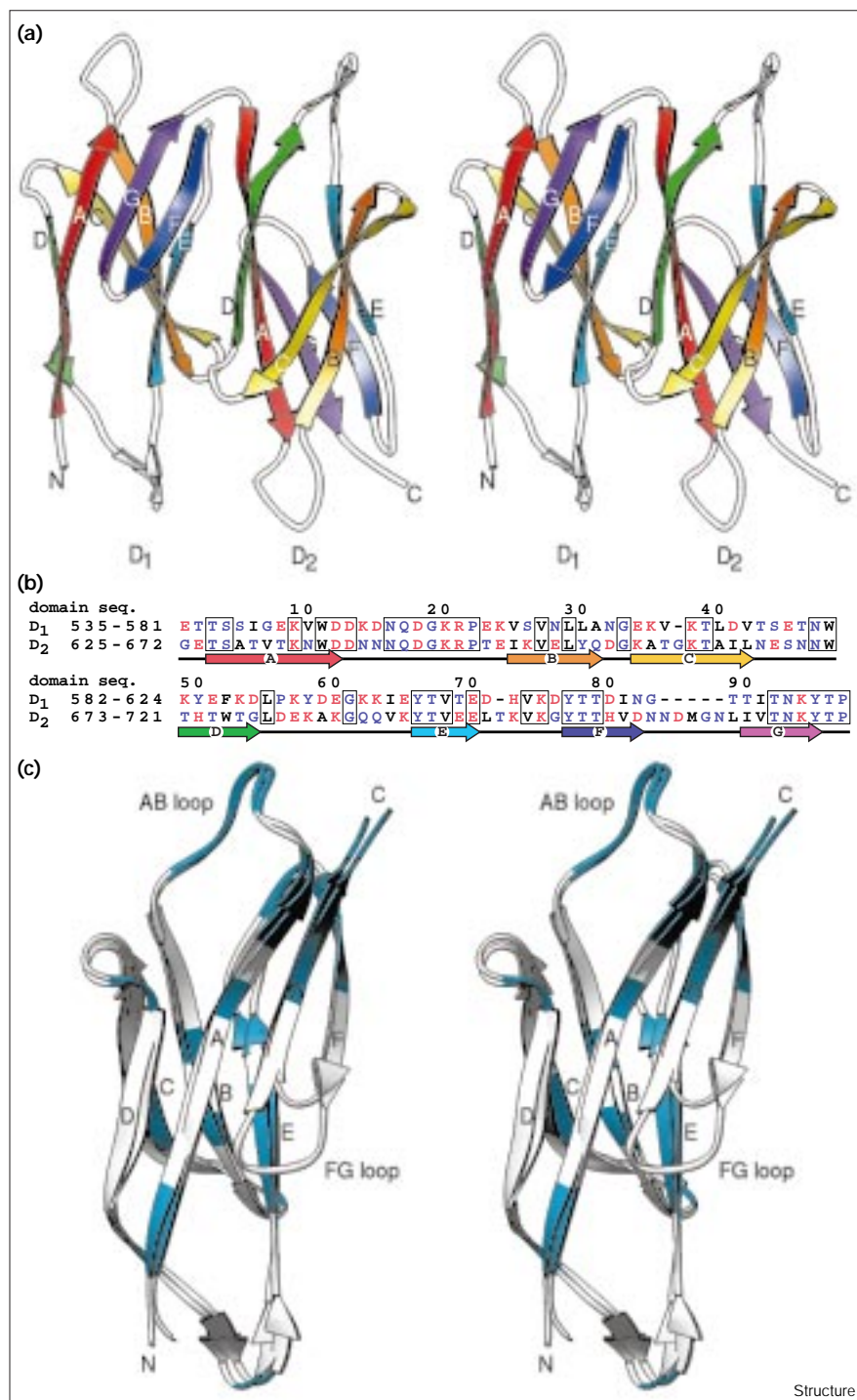
The crystal structure of B₁ displays two similar domains (D₁ and D₂) placed side by side and related by an approximate twofold screw axis (2₁; Figure 2a). Each D domain exhibits a four- and three-strand arrangement similar to that observed in the IgG domains but differs in topology from all the previously described V, C, H and I types (V, variable; C, conserved; H, hybrid; I, intermediate; respectively) in the immunoglobulin superfamily (IgSF)

[19,20]. The order of the β strands and their linking regions are different in comparison with all the described IgSF-like domains. We designate this fold as the inverse IgG (inv-IgG) fold. The interface between D₁ and D₂ is marked by several hydrogen bonds between the two domains. There are only three hydrophobic residues (Leu564, Met710 and Ile714) in the interface and the remaining interactions are via either charged or polar amino acids. A combined surface area of 782 Å² is buried in the interface between the two domains (see the Materials and methods section for a calculation of the buried surface area).

There is 41% sequence identity between the two domains, and the 89-residue D₁ domain and the 97-residue D₂ domain superpose with a root mean square deviation (rmsd; on C _{α} atoms) of 1.28 Å (Figure 2b,c). The existence of two domains and their significant sequence identity in a single B region repeat unit was identified upon solving the crystal structure of B₁. The AB loop regions in these two domains share a striking sequence identity (67%) and have similar structures (Figure 2c). These AB loops in D₁ and D₂ have the sequences DKDNQDGKRPEK (single-letter amino acid code) and DNNNQDGKRPTTE, respectively, where eight out of twelve residues are conserved and resemble the consensus calcium-binding motif of the form DxxNx DxKxxE (where X is any amino acid) [21] that is present in the *S. aureus* surface protein SdrD B domain [22]. The observed coordination geometry of the water molecule present in this loop suggested a possible metal binding site. However, structural analyses of B₁ crystalized in the presence of calcium did not reveal any

Figure 2

The primary and tertiary structure of B₁. (a) Stereoview of a ribbon drawing of the B₁ crystal structure. The two domains D₁ and D₂ are placed side by side and are related by an approximate twofold screw along the axis horizontal to the plane of the paper. The strands of B₁ are assigned rainbow colors to enable comparison with other IgSF domains. (b) Sequence alignment of the D₁ and D₂ domains: a 41% identity is indicated by boxed residues. Red, blue and black represent charged, polar and hydrophobic amino acids and the strands (rainbow colors) indicate the consecutive ABCDEFG strands of the β sheets. (c) Stereoview of ribbon drawings of the superposition of the D₁ and D₂ domains of B₁. Regions in cyan represent identical residues. The AB loop region is highly conserved. The FG loop region in D₂ has five extra residues.

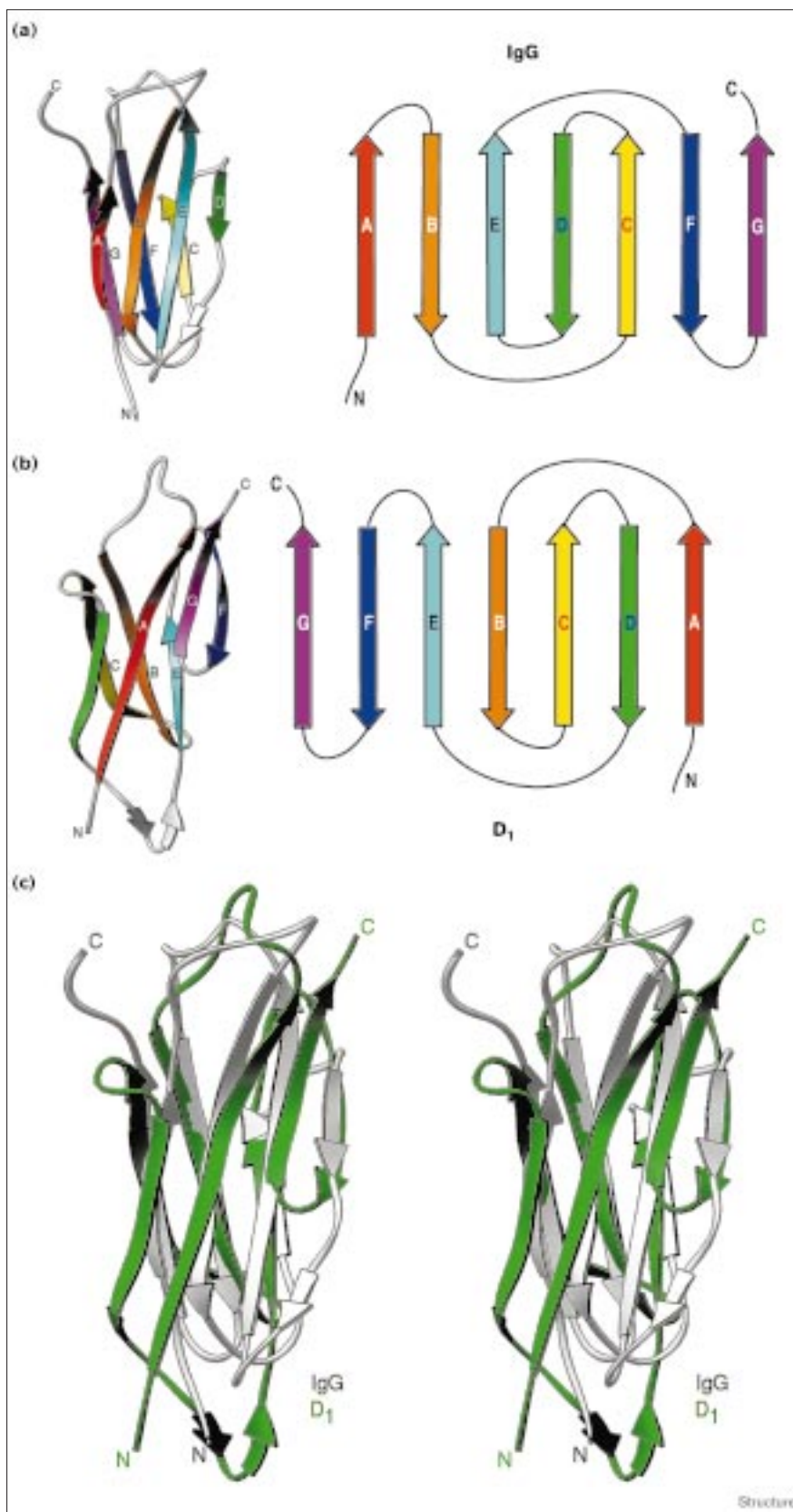


calcium ions in this location (data not shown). Moreover, no change in structure upon addition of Ca²⁺ was observed in the circular dichroism (CD) and fluorescence emission spectra of B₁ or B₁B₂ (data not shown). Further comparison of the D₁ and D₂ structures revealed that the FG loop is the only segment that differs significantly in

length and composition; D₂ has five additional residues including Asp709 and Met710, and it exhibits a resultant negative charge on the molecular surface.

The two independent copies of B₁ in the asymmetric unit have distinct crystal-packing interactions but exhibit

Figure 3



The inverse fold. (a) The IgG constant domain and (b) D₁ as rainbow-colored ribbon and topology diagrams. The D₁ resembles the IgG-like structure with the four- and three-stranded barrel, but exhibits a novel fold. The inverse relation between the two antiparallel β -sandwich structures can be seen as follows: (IgG/D₁) A/F, D/C, C/D, B/E, E/B and F/A. In the rainbow-colored topology diagrams (Figure 3a,b), if any one of the colors on the IgG and D₁ are aligned an adjacent strand of D₁ (to the right or left) is observed to be in the opposite side in IgG. (c) Stereoview superposing the IgG constant (white) and the D₁ (green) domains.

nearly identical structures, superposing well with an rmsd (C_{α} atoms) of 0.68 Å. The presence of an identical second copy of B_1 in the asymmetric unit suggests that the ‘folding pattern’ of D_1 (and D_2), and the ‘domain packing’ between D_1 and D_2 could be intrinsic in nature.

Comparison with the IgG domain

The IgG constant domain generally has a four-stranded β sheet, ABED, on one side of the barrel and a three-stranded β sheet, CFG, on the other side (Figure 3a). The D_1 and D_2 domains also have a four-(DAGF) and three-stranded (EBC) β -sheet arrangement but display a different topology (Figure 3b). When overlaying the four strands of D_1 (DAGF) on IgG (ABED) the directions of the first two strands D and A of D_1 run in opposite directions to the A and B strands of IgG. Whereas the strands A and G of D_1 are parallel, the positional equivalent strands B and E of IgG are antiparallel.

The IgG and D_1 folds seem to be the inverse of one another. Keeping the C-terminal G strand common, the inverse relationship between the two antiparallel β -sandwich structures starting from the N-terminal end of the IgG is as follows: (IgG/ D_1) A/F, B/E, C/D, D/C, E/B, and F/A. It seems that the evolution of the D_1 fold has relied on tail-to-head rather than head-to-tail folding, if directionality could be assigned to the folding. This inverse nature is observed in several other aspects of the fold. For example, in the rainbow-colored topology diagrams (Figure 3a,b), if any one of the strand colors on the IgG and D_1 are aligned an adjacent strand of D_1 (to the right or left) is observed to be on the opposite side in IgG. Also, viewing the four-stranded β sheet as the frontal plane of the molecule, the CD loop region is on the right side for IgG, whereas in D_1 it is on the left side (Figure 3a,b). The N termini of D_1 and IgG originate from nearly identical positions. The C terminus of D_1 is on the right side of the frontal plane (four strands) whereas in IgG it is on the left side of the back plane (three strands) of the module (Figure 3c). Thus, the D_1 and D_2 domains have a four- and three-stranded arrangement similar to that observed in IgG domains, but they exhibit a novel inverse fold.

Three-dimensional alignment with DALI

In a comparison of D_1 with the three-dimensional (3D) structures of other protein structures present in the Protein Data Bank (PDB), the DALI program [23,24] generated the best fit with the intercellular cell adhesion molecule 2 (ICAM-2) fragment followed by coagulation factor fibronectin type III and other cell-adhesion molecules. The crystal structures of human ICAMs, vascular cell adhesion molecule (VCAMs), and mucosal addressin cell adhesion molecule (MAdCAM) consist of two catenated IgSF-like domains, where domain 1 (83–90 residues) is smaller than domain 2 (102–112 residues) [25–28]. B_1 also has a smaller N-terminal domain (D_1 ; 87 residues) and a larger C-terminal

domain (D_2 ; 97 residues). The first domains of VCAMs and MAdCAM carry Leu–Asp–Val (LDV) motifs whereas the ICAMs carry a functionally homologous ETxxxK motifs, which were shown to be essential for binding to their ligands (the integrins). The key integrin-binding residues Asp43 in MAdCAM and Asp40 in VCAM-1 are located in the CD loop. In ICAM-1 and ICAM-2 the critical integrin-binding residues Glu34 and Glu37, respectively, are located at the end of the C strand on a rather flat surface [29]. Notably, the structure of Cna B_1 includes an LDV motif in the first domain (D_1) at the end of the C strand, where the structurally homologous Asp574 is present on a similar flat surface. Also present in D_1 is the ETxxK motif (with one missing x) next to the LDV motif; these motifs are separated by two residues in the CD loop, and they structurally resemble those observed in ICAMs and VCAMs. If the CD loops were to be on the right side (viewing the four strands on the frontal plane; Figure 3b) of D_1 in B_1 , in the present packing arrangement (Figure 2a) these binding motifs would be buried in the surfaces between D_1 and D_2 . The newly observed fold for the D_1 domain has allowed the presentation of the LDV motif at the exposed extremity of the molecule.

Comparison with two-domain IgSF structures

Many two-domain IgSF structures have been reported in the PDB, and they exhibit interesting variations in their packing. Two angles, namely an elbow and a twist angle, have been defined in the past to compare the IgSF domains and their packing. We have derived the new ‘swivel’ angle to compare the 3D packing of IgSF-like domains (see the Materials and methods section). The interferon gamma receptor (IFN- γ R α) [30], human high-affinity IgE receptor (IgE) [31], killer-cell inhibitor (KIR) [32] and human growth hormone (hGHR) [33] are representative proteins having crystal structures with various elbow, twist and swivel angles (Table 1, Figure 4). The elbow angle between the D_1 and D_2 domains of IFN- γ R α , IgE, KIR and hGHR are 110°, 52°, 47° and 72°, respectively (Table 1). The variation in elbow angles observed in these crystal structures might represent the intrinsic characteristics of their domain packing, but they also seem flexible enough to allow minor conformational changes induced by the forces of crystal packing. B_1 has two independent copies related by noncrystallographic symmetry in the asymmetric unit and both copies exhibit a very acute elbow angle of 7° (± 1) between D_1 and D_2 (Table 1) when compared with 47° and 52° for KIR and IgE structures [24,25], respectively (Table 1). The B_1 repeat unit has distinct packing, where all of the three angles have contributed to a tight packing between the D_1 and D_2 domains.

Crystal structure of B_1B_2

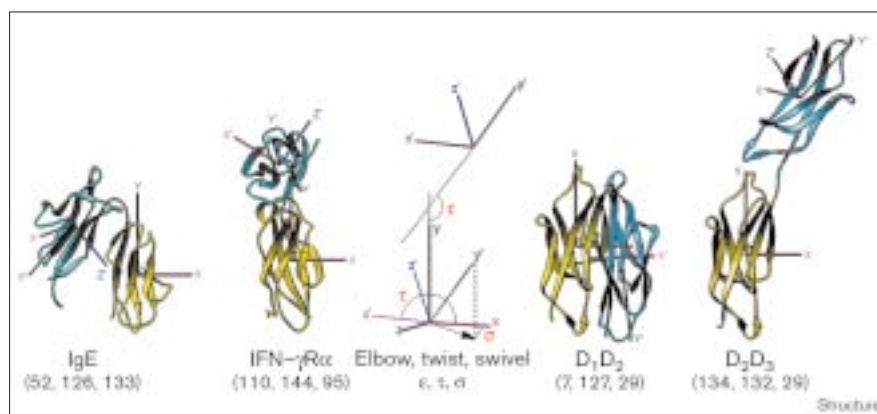
In the crystal structure, the B_1 and B_2 repeat units maintain their individual structural identity as first proposed by Rich *et al.* [18]; each has two domains arranged side by

Table 1

The calculated angles between IgSF domains.				
Protein	Elbow angle ϵ ($^{\circ}$)	Twist angle τ ($^{\circ}$)	Swivel angle σ ($^{\circ}$)	Reference
B ₁ : D ₁ -D ₂	7	127	29	This paper
B ₁ B ₂ : D ₂ -D ₃	134	132	-29	This paper
B ₁ B ₂ : D ₃ -D ₄	6	130	43	This paper
IFN- γ R α receptor	110	144	95	[30]
IgE	52	126	-133	[31]
Killer-cell inhibitor	47	86	164	[32]
Human growth hormone	72	144	128	[33]
Fibronectin: F ₇ -F ₈	121	108	125	[34]
Fibronectin: F ₈ -F ₉	126	160	95	[34]
Fibronectin: F ₉ -F ₁₀	166	44	-114	[34]
Hemolin: H ₁ -H ₂	127	127	170	[35]
Hemolin: H ₂ -H ₃	29	69	-26	[35]
Hemolin: H ₃ -H ₄	165	131	108	[35]

side, similar to their arrangement in the crystal structure of the single B₁ repeat unit (Figures 1,5a). As expected, the B₁ and B₂ repeat units have very similar structures and upon superposition have an rmsd (on C $_{\alpha}$ atoms) of 0.47 Å. The domains of B₁B₂ are designated as D₁, D₂, D₃ and D₄ (Figure 1). Several significant interactions are observed between the D₂ and D₃ domains. The AB loop of D₂ has several hydrogen bonds with the DE loops of D₃, increasing the stability of the packing between the B₁ and B₂ repeat units. B₁ and B₂ have 782 Å² buried surface area between them, and the replacement of residue Asn516 in B₁ by Asp753 in B₂ has no significant impact on the local structure. The crystal-structure determination of B₁B₂ is unique in that the covalently linked B₁ and B₂ repeat units are almost identical (99.5%) in sequence. To our knowledge there have been no similar crystal structures reported of two covalently linked repeat units that exhibit such a high degree of sequence homology.

Figure 4



Representative two-domain IgSF structures of IgE, IFN- γ R α with D₁D₂ and D₂D₃. The first and second domains are colored yellow and cyan respectively. The axes are indicated in magenta (x, x'), black (y, y') and blue (z, z'). The angles elbow (ϵ), twist (τ) and swivel (σ) are schematically represented and the calculated values are given within brackets (see the Materials and methods section for definitions of these angles).

Single amino acid omission results in a different orientation

In B₁B₂ the loops connecting D₁ to D₂ and D₃ to D₄ have the sequence KYTPGET (Figure 5b) and form a β turn similar to the one observed in B₁. The loop region connecting the two repeat units B₁ and B₂ (D₂ to D₃) has a missing glycine residue. This KYTPET polypeptide segment connecting D₂ to D₃ does not adopt a β -turn arrangement (Figure 5c). The presence of the glycine residue in the D₁-D₂ and D₃-D₄ junctions seems to have imparted a conformational flexibility to this polypeptide segment, allowing the loop to form a restrictive β turn. However, the D₂-D₃ connecting loop without the glycine residue has an apparent reduced flexibility, restricting the conformations the linker region can adopt. As a result, the B₂ repeat unit is projected in a spatially different orientation.

Comparison with four-domain IgSF structures

The crystal structures of segments of human fibronectin and insect hemolin with four IgG-like domains have been solved (Figure 6) [34,35]. In the crystal structure of a four-domain segment of human fibronectin, the IgSF domains (F₁, F₂, F₃ and F₄) stack in an elongated fashion roughly along an axis [34]. The average elbow angle between these domains is 137° (Table 1). In the structure of hemolin, an insect IgSF member induced during bacterial infection, the four IgSF domains (H₁, H₂, H₃ and H₄) adopt a horseshoe shape [35]. This structure has variable elbow, twist and swivel angles between the domains (Table 1). In the B₁B₂ structure the domains D₁-D₂ and D₃-D₄ have acute elbow angles of 7° (± 1), partially resembling the H₂-H₃ domains of hemolin. Interestingly, however, the D₂-D₃ domains pack like the elongated fibronectin domains, with an elbow angle of 134°. The calculated buried surface areas between D₁ and D₂ (784 Å²) and D₃ and D₄ (827 Å²) are similar to those observed between H₂ and H₃ (1005 Å²) and H₁ and H₄ (739 Å²) of hemolin. The buried surface area between the elongated domains D₂ and D₃ of B₁B₂ (222 Å²) is smaller than the buried surface area between

Figure 5

B₁B₂ crystal structure. (a) Crystal structure of B₁B₂ in stereo. The two individual repeat units B₁ and B₂ retain the packing arrangements observed in the B₁ structure. (b) Electron density around the KYTPGET polypeptide in the D₁–D₂ and D₃–D₄ linker region. (c) Electron density around the KYTPET polypeptide in the D₂–D₃ linker region.

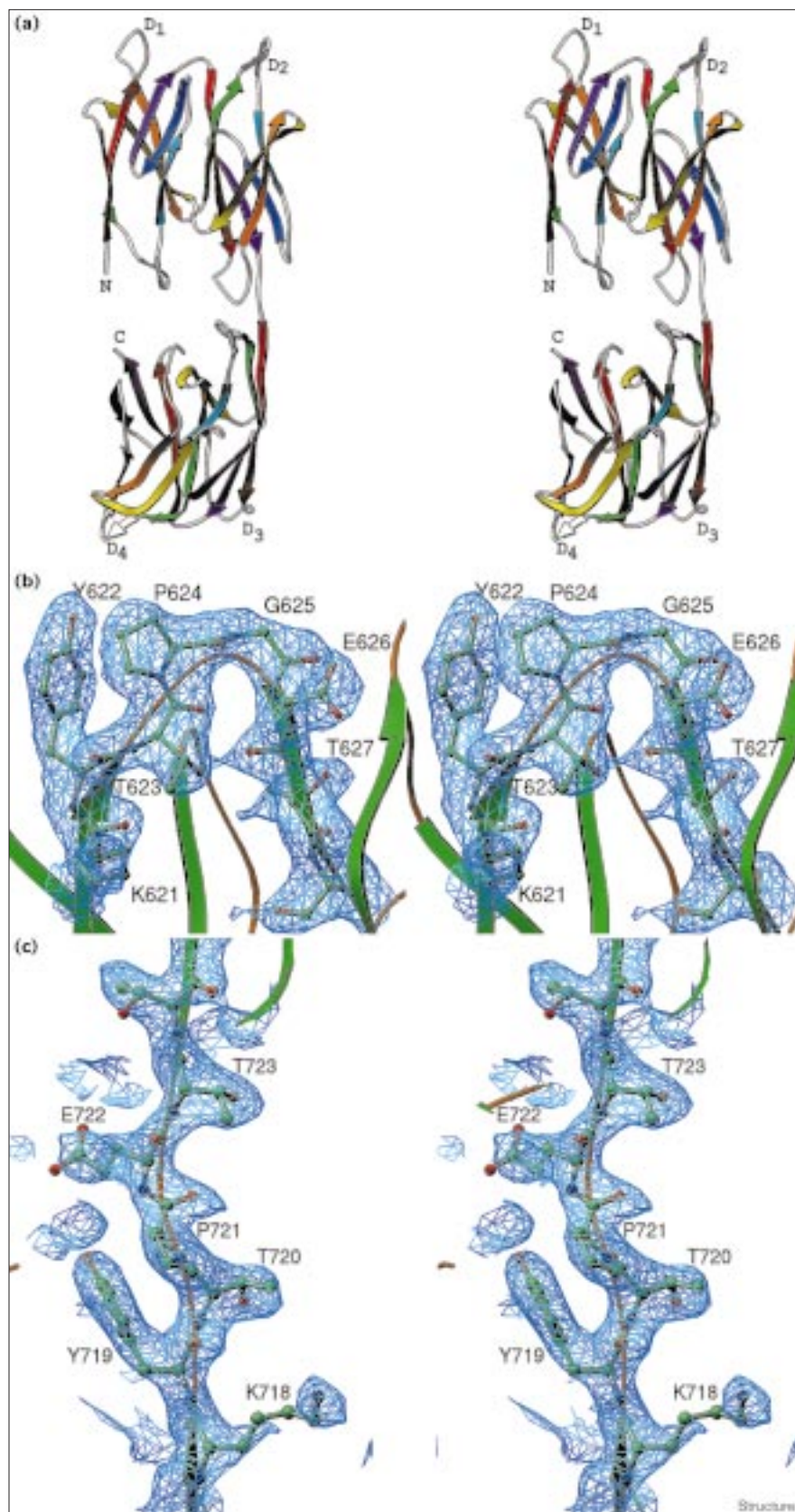
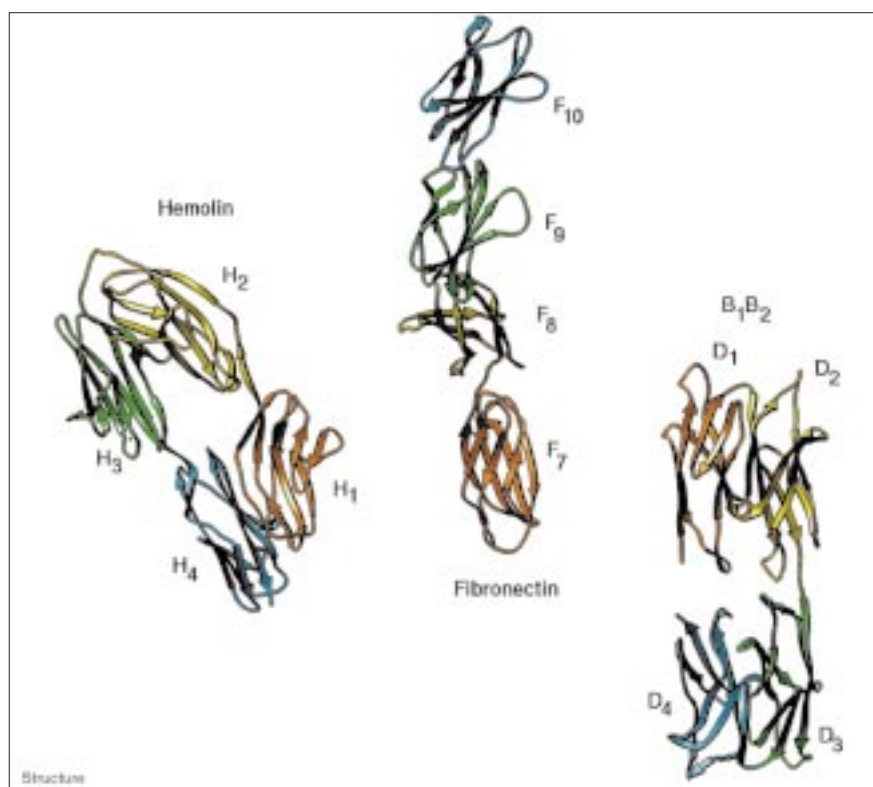


Figure 6

Comparison of B_1B_2 and other four-domain IgSF-like domains, hemolin (PDB code 1B1H) and fibronectin (PDB code 1FNF). The four domains are colored consecutively in the order orange, yellow, green and blue. The orientations have been chosen, as described earlier, by aligning the first domain of each of these structures.



H_1 and H_2 (325 \AA^2) and H_3 and H_4 (310 \AA^2) in hemolin and comparable to those between the fibronectin domains F_7 and F_8 (285 \AA^2), F_8 and F_9 (267 \AA^2) and F_9 and F_{10} (163 \AA^2). Overall, the structure of B_1B_2 has less buried surface area than fibronectin and hemolin. It follows a clockwise direction in the arrangement of domains when compared with the counter-clockwise path followed by hemolin (Figure 6), a reflection of the inverse nature of this novel fold. Thus, the crystal structure of B_1B_2 presents a mixture of the packing arrangements observed in fibronectin (D_2 – D_3) and hemolin (D_1 – D_2 , D_3 – D_4).

Modeling multiple repeats

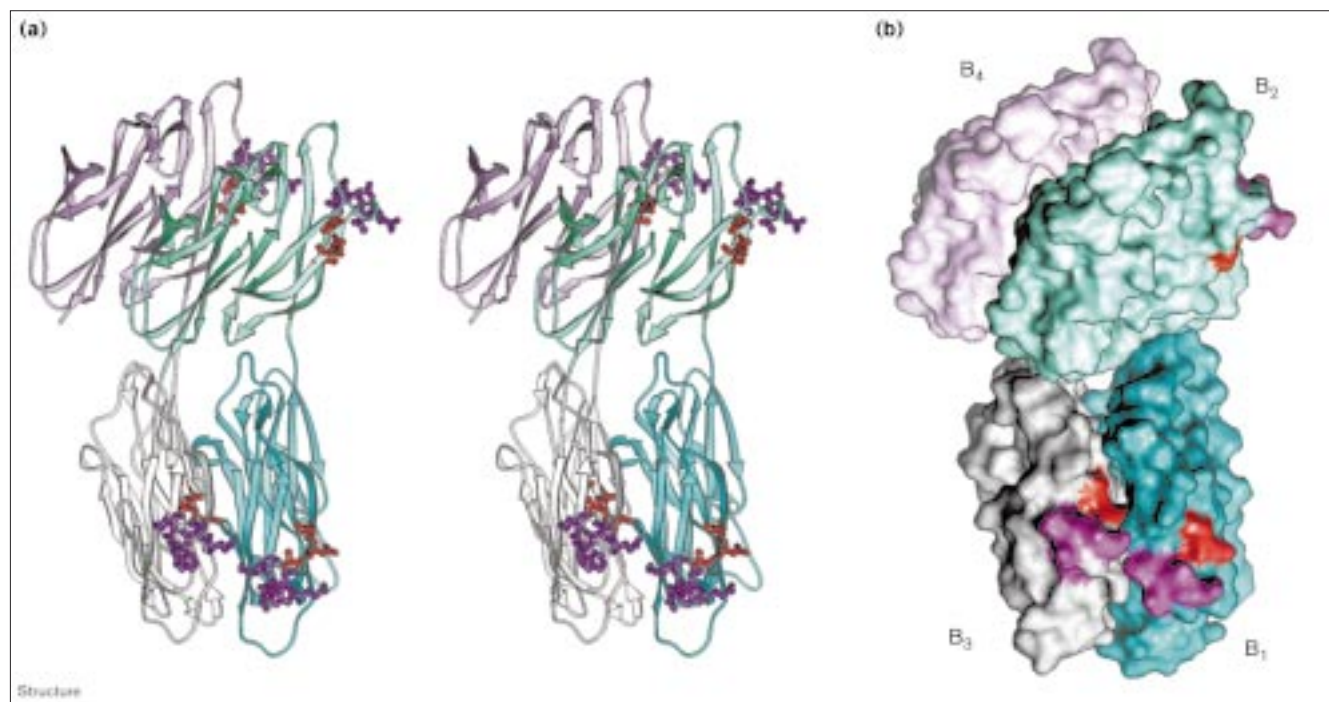
The presence of multiple B region repeat units in Cna derived from different strains of *S. aureus* has been reported [17]. This raises the question of how $B_1B_2B_3B_4$ would pack. In the absence of a crystal structure we initiated modeling studies of these repeat units. A model of $B_1B_2B_3B_4$ was generated assuming that the B_2 – B_3 linkage is similar to that observed in B_1 – B_2 . This assumed constraint places B_3 face to face with B_1 . By analogy, B_4 will lie face to face with B_2 . Rigid-body transformations based on the C_α atoms were used to construct the initial model. In this $B_1B_2B_3B_4$ model there were some close contacts and some interpenetration of subunits. A brute-force docking procedure [11,36] implemented in SoftDock (MC, unpublished program) to dock B_1B_2 and B_3B_4 was initiated; the

B_1B_2 model was held stationary and an additional B_1B_2 (B_3B_4) was allowed to pivot about its N terminus within a 5 \AA radius of the C terminus of B_2 . A set of favorable solutions placed B_3B_4 in approximately the same orientation as the $B_1B_2B_3B_4$ model generated initially via rigid-body transformations, but the bad contacts were relieved. Figure 7 shows the best solution, in which the only buried charged residues form an ion pair. These repeat units are arranged in a zig-zag fashion that resembles an accordion.

Discussion

The IgSF motif is evolutionarily ancient, as exemplified by both intracellular and extracellular IgSF domains in *Caenorhabditis elegans* [37]. On the human leukocyte surface about 40% of the proteins are predicted to contain one or more IgSF domains [38]. The crystal structure of the bacterial periplasmic chaperone (PapD) has a topology identical to that of an IgG fold [39]. Here we report the structure of the B region of the bacterial surface protein Cna, which has a β sandwich of novel architecture that resembles the IgG fold. The novel inverse nature of this fold compared with that of IgG is observed in: the reversal of directions in the placement of adjacent strands in the topology diagrams compared with IgG (Figure 3a,b); the presence of the CD strand on the opposite side of the molecule compared with the IgG domain (Figure 3c); and the clockwise pattern of packing of domains in the B_1B_2

Figure 7



Model of $B_1B_2B_3B_4$ repeat units. (a) Stereoview of the ribbon model. The LDV and ETxxK motifs are in red and purple ball-and-stick representation. The model displays a zig-zag arrangement like an accordion. (b) Molecular surface model of $B_1B_2B_3B_4$ repeat units.

structure compared with the counter-clockwise packing pattern observed in hemolin (Figure 6). The secondary structure predictions based on the sequence of D_1 proposed an IgSF fold; however, the observed inverse fold could not be predicted. The preservation of structural homology between the D domains is observed not only in the β -sheet regions but also in the loop regions. The conservation of sequence and tertiary structure by lower-level organisms like bacteria indicates that this might be an essential structural element for a specific, but as yet unidentified, biological function.

It is interesting to observe both the integrin-binding LDV and ETxxK motifs in the CD loop region of the first domain (D_1) of B_1 ; they closely resemble those of the ICAMs and VCAMs. The LDV motifs in B_1B_2 are well exposed at the extremities of the molecule and are separated by a distance of 60 Å. It is very tempting to speculate and suggest that each B repeat unit binds to one integrin molecule. However, cell binding experiments with $\alpha_4\beta_1$ integrin, which recognizes the LDV motif [40] did not show any positive binding to the recombinant construct of B_1B_2 (data not shown).

Domain duplication by genetic evolution has led to interesting combinations of events in both sequence and structure, and divergent evolution could lead to a difference in the

sequence of the duplicated domain [41]. Domain arrangements are also crucial during the initial assembly of proteins to optimally adapt to their function. There are several examples of tandem-duplicated crystal structures (e.g. chymotrypsin, zinc-finger motifs found in transcription factor IIIA, and fibronectin type III repeats found in the muscle protein titin; for a review see [41]). However, these duplicated domains vary in amino acid composition and in some cases in their secondary structure. The B_1 and B_2 repeat units have nearly identical sequence and secondary structure. Further, the B_1B_2 structure has less buried surface area between the domains when compared with the other similar four-domain structures of fibronectin and hemolin. This could mean that the interactions between B_1 and B_2 are more susceptible to changes by the external forces in comparison with the other four-domain IgSF structures.

The observation of the omission of a residue resulting in a different orientation of the next B repeat unit is similar to that suggested by Eisenberg and coworkers [42,43]. They proposed that a hinge-loop mutation/deletion lead to domain swapping, a phenomena observed in the packing of multiple copies of the protein molecule. In the case of B_1B_2 we do not observe any domain-swapping characteristics, however, we do observe an arrangement that is equivalent to domain swiveling. The structure of B_1B_2 resembles the linear (open-ended) 3D domain-swapped oligomer, which

was proposed for multiple aggregation of monomers. Interestingly, Pokkuluri *et al.* [44] reported domain flipping resulting from the substitution of one amino acid (Glu38) in the crystal structure of modified human immunoglobulin light-chain variable domains (V_L). However, in this structure the 'conventional dimer' has flipped by 180° around an axis perpendicular to the original twofold axis. It is important to note that a single glycine residue produces a radically different conformation in a sequence (KYTPET, KYTPGET) of high homology. Homology modeling and threading analyses tend to conclude that the presence of a glycine in a sequence of this kind would not produce major conformational change; in the case of B_1B_2 this conclusion would obviously be incorrect.

Cna was proposed to be a mosaic protein, where one region/domain functions independently of another [18]. In our modeling studies the accordion model of the B-repeat units could project the A region from the cell surface and aid in binding to collagen. If these repeat units were not serving as a stalk, but instead have an unknown function, then flexibility of the $B_1B_2B_3B_4$ model would be needed for these repeat units to move and function independently. The $B_1B_2B_3B_4$ accordion model seems to possess the flexibility and stability needed for holding out the larger A region (55 kDa) for collagen binding. The necessary trigger or force to activate them to move in conjunction (stretching/retracting) could come from any of the many known and unknown cell events.

Biological implications

The collagen-binding surface protein of *Staphylococcus aureus* is composed of a non-repetitive A region and a repetitive B region. In our previous study of this protein, the A region was shown to bind to collagen. The crystal structure was solved for a portion of the A region, Cna151–318, that contains the collagen-binding activity. Subsequently, docking studies implicated a trench observed in the structure as the collagen-binding site, and mutations in and around this trench supported this hypothesis. The presence of one to four B-repeat units has been observed in different strains of *S. aureus*. The importance of the presence of the B-repeat region and the functions of these repeats are not yet known. However, it has been suggested that the B-repeat units might serve as a 'stalk' that projects the A region away from the bacterial surface, thereby facilitating bacterial adherence to collagen. Analyses of the crystal structure of the single B-repeat unit (B_1), revealed two immunoglobulin-like (IgG-like) domains with a novel fold. This new fold has a four- and three-strand arrangement, and is the inverse of the fold found in the IgG-like domains. Notably, the B-repeat units include the integrin-binding LDV and ETxxK motifs in the CD loop regions of their first domains, although the significance of these domains remains unclear. The crystal structure of two B-repeat units

(B_1B_2) has a packing arrangement where a single missing glycine residue in the linker region between B_1 and B_2 has resulted in a new orientation of the domains. Our modeling studies on the $B_1B_2B_3B_4$ repeats shows that these domains pack in a zig-zag fashion, like an accordion; they might stretch and contract from the bacterial cell wall and thus aid in the projection of the A region away from the cell surface. Perhaps in the event of proteolytic loss of a Cna A region resulting from a specific cleavage by the host's extracellular defensive apparatus, the B-repeat units could be pressed into an adhesive process or any other function essential for bacterial survival. The repeating number of these domains might reflect the added stability the bacteria would achieve by multiple anchoring.

Materials and methods

Crystallization and data collection

The expression, purification and crystallization of the recombinant protein B_1 (22.6 kDa) were reported previously [45]. Crystals were grown by the hanging-drop vapor-diffusion method. A droplet containing 2 μ l of protein, 1.6 μ l of well solution (1 ml of 4M $(NH_4)_2SO_4$, 50 mM succinic acid, pH 4.6) and 0.4 μ l of 10% (w/v) octyl β -D-glucopyranoside was equilibrated against the reservoir well solution. Crystals grew at room temperature in 4–5 days to a size of $0.5 \times 0.5 \times 0.2$ mm. Data were collected on the Raxis IIc image plate detector mounted on a Rigaku RU-200 rotating anode operating at 100 mA and 40 kV. A complete data set to 2.5 Å resolution was collected at room temperature. The crystallographic data statistics are presented in Table 2.

In the case of crystallization of the B_1B_2 (43.8 kDa) recombinant protein, a droplet containing 1 μ l of protein, 1 μ l of 5 mM $CaCl_2$ and 7 μ l of well solution (1 ml of 16% (w/v) polyethylene glycol 6000 (PEG6000), 100 mM sodium cacodylic acid, pH 7.5) was equilibrated against the reservoir well solution [46]. Diffraction-quality crystals grew at room temperature in 5–6 days to a size of $0.4 \times 0.4 \times 0.15$ mm. A complete data set was collected on a Raxis IV image plate mounted on a Rigaku RU-H3R rotating anode operating at 100 mA and 40 kV to 2.3 Å resolution at room temperature. Crystallographic data statistics are presented in Table 2.

Structure determination and refinement

The structure of the B_1 protein was solved by the MIR (multiple isomorphous replacement) method. Three derivatives (trimethyl lead acetate [TMLA], potassium platinum tetrachloride and samarium nitrate) were identified by difference Patterson analysis suitable for MIR phasing (using the program XTALVIEW [46]). The details of the heavy-atom derivative statistics are presented in Table 3. Each derivative had one major site and one minor site. Refinement of heavy-atom derivatives and subsequent phase calculations were carried out using the program PHASES [47]. An initial 3.0 Å electron-density map revealed clearly defined solvent boundaries and most of the sidechain densities were visible. The initial C_α trace was made using the program CHAIN [48]. Sidechains were incorporated into the model and further refinements were performed with X-PLOR [49]. All reflections ($\geq 2\sigma$) from 8.0 to 2.5 Å were used in the refinement, with 10% partitioned in a test set for monitoring the refinement process [50]. The initial R_{free}/R factor of the fully built model with sidechains was 40.1%/35.7%. The molecular model building was performed with the graphics program O [51], in conjunction with OOPS [52]. Solvent molecules were added based on geometry and electron density. Model convergence was achieved after several cycles of model building and refinement. Maximum-likelihood refinement and bulk-solvent correction were carried out with CNS_0.4 [53] in the last few cycles, resulting in 84 solvent molecules in the asymmetric unit with the final R_{free}/R factor being 24.5%/19.1%. The average B factor for B_1 was 38.8 Å². Although the B factors are on the

Table 2

B ₁ and B ₁ B ₂ crystallographic data.		
Parameter	B ₁	B ₁ B ₂
a (Å)	96.97	42.39
b (Å)	101.32	79.42
c (Å)	120.79	130.40
α=β=γ (°)	90.00	90.00
Space group	I222	P2 ₁ 2 ₁ 2 ₁
Resolution (Å)	2.5	2.3
I/σI	11.7	22.3
R _{sym} * (%)	7.0	6.9
V _m (Å ³ Da ⁻¹)	3.2	2.4
Unique reflections	20,953	19,750
Reflections > 3σ (%)	61.5	48.7
Completeness (%)	99.0	97.0
No. of molecules in asymmetric unit	2	1
R factor	18.85	18.29
R _{free}	24.21	24.44
Rmsd (bonds)	0.005	0.009
Rmsd (angles)	1.180	1.360
Average B factor of protein (Å ²)	36.58	41.28
Average B factor mainchain (Å ²)	34.58	39.58
Average B factor sidechain (Å ²)	38.86	42.97
No. of solvent molecules	84	72

*R_{sym} = $\sum |I_h - \langle I_h \rangle| / \sum I_h$, where $\langle I_h \rangle$ is the average intensity over symmetry equivalents.

high side, they are still comparable with the calculated B value from the Wilson plot of 47.6 Å². The average B factors for IgG-like domains vary between 13.0 Å² (for fibronectin [34]) and 50.4 Å² (for ICAM-2 [25]). The model has 99.1% of the residues in the allowed regions of the Ramachandran plot, two residues in the generously allowed regions, and one residue in the disallowed region. These latter three residues appear in loop region where the electron density is not clearly defined.

The crystal structure of B₁B₂ was solved by the molecular-replacement method. The crystal structure of B₁ was used as the initial model. A good rotation solution was obtained using the program AMORE [54]. Using the first rotation solution we obtained two good peaks for the translation search. Fixing the properly rotated and translated molecule we could obtain the rotation and translation values for the second molecule. Following rigid-body refinement in AMORE, the R factor was 31.1% and correlation coefficient of 0.51 was obtained for each individual solution. The model was further refined by X-PLOR using simulated annealing. The R_{free}/R factor dropped from 42.8%/35.6% to 37.7%/27.7%. Further refitting was carried out with the program O in conjunction with OOPS. Maximum-likelihood refinement and bulk-solvent correction were performed with CNS_0.4 in the last few cycles. With 72 solvent molecules in the asymmetric unit the model had a final R_{free}/R factor of 24.44%/18.29%. This model had an average B factor of 41.2 Å², which was comparable with that of the B₁ model and had all the residues within the allowed regions of the Ramachandran plot.

Orientation and superposition of IgSF-like domains

The comparison of the different IgSF domains requires a common coordinate system and here we describe a new method for comparing them. The first and last C_α atoms of the seven core strands of the barrel each defined a strand vector and their average, a strand center. The average of the sheet centers was taken as the origin. The seven normalized vectors (negated if necessary to point in the same direction as the first strand) were averaged and normalized to define a y axis. The average direction of the vectors between the first three sheet centers on each side of the barrel defined an intermediate vector, x'. Then the z axis was defined as the cross product of x' and y. The x axis was then defined as the cross product of y and z. The y axis gives the average direction of

Table 3

Heavy-atom derivative statistics for B ₁ .					
Heavy-atom compound	Resolution (Å)	R _{sym} * (%)	Unique reflections	R _{cullis} [†]	Phasing power [‡]
(CH ₃) ₃ PbOOCCH ₃	3.0	6.4	16,335	0.526	2.11
K ₂ PtCl ₄	3.5	9.2	13,031	0.543	1.87
Sm(NO ₃) ₃	4.0	8.1	10,177	0.625	1.64

*R_{sym} = $\sum |I_h - \langle I_h \rangle| / \sum I_h$, where $\langle I_h \rangle$ is the average intensity over symmetry equivalents. [†]R_{cullis} = $\sum |F_{PH} \pm F_P - F_{Hcalc}| / \sum |F_{PH} - F_P|$. [‡]Phasing power = $\sum |F_H| / \sum |F_{PHobs} - |F_{PHcalc}|$.

sheets, with the strand centers roughly parallel to the x axis. This orientation is used for all figures and superpositions presented in this paper. All the figures in this paper were made using the program RIBBONS [55].

Elbow, twist and swivel angles

The relationship between the two domains is based on the coordinate system described earlier to align domains. We define an elbow angle as the complement of the angle between the y axes of the two domains, a twist angle as the angle between the x axes of the two domains, and a swivel angle as the projection of the second domain's y axis on the xz plane of the first domain (Figure 4). The original method used in calculating the elbow angles for antibodies was based on the intersection of the dyads [56]. In the structure of hemolin it was defined as the angle between the long axes of adjacent domains, approximated by the ellipsoids [35], and in fibronectin by connecting the C_α positions of residues at the midpoints of the B–C and E–F loops [34]. The rotation angles required to superimpose adjacent domains described for fibronectin closely resemble our calculated values for the twist angles. The swivel angle is a new definition and to our knowledge has not been described yet for comparing the IgSF-like domains. (The program used to calculate these angles is available on request from MC).

Buried surface area

The analytical molecular surface area was calculated using the MSP program [57]. The coordinates of the two domains, D₁ and D₂, were combined to form D₁D₂ and the surface area was determined for each set. Buried surface was approximated with the areas: D₁ + D₂ – D₁D₂.

Accession numbers

The coordinates of B₁ and B₁B₂ have been deposited with the Protein Data Bank with accession codes 1D2O and 1D2P respectively.

Acknowledgements

Authors thank Mark Walter and Theodore Jardetzky for providing the coordinates of IFN-γRα and IgE, respectively. CCSD thanks Dwight Moore and Laurent Chantalat for discussion on crystallization. SVLN acknowledges the support for this project from NIH and NASA.

References

- Hawiger, J., Timmons, S., Strong, D.D., Cottrell, B.A., Riley, M. & Doolittle, R.F. (1982). Identification of A domain of human fibrinogen interacting with *Staphylococcal* clumping factor. *Biochemistry* **21**, 1407-1413.
- Kuusela, P. (1978). Fibronectin binds to *Staphylococcus aureus*. *Nature* **276**, 718-720.
- Chhatawal, G.S., Preissner, K.T., Muller-Berghaus, G. & Blobel, H. (1987). Specific binding of the human S protein (vitronectin) to *Streptococci*, *Staphylococcus aureus*, and *Escherichia coli*. *Infect. Immunol.* **55**, 1878-1883.
- Speziale, P., Raucci, G., Visai, L., Switalski, L.M., Timpl, R. & Höök, M. (1986). Binding of collagen to *Staphylococcus aureus* Cowan I. *J. Bacteriol.* **167**, 77-81.
- Foster, T.J. & Höök, M. (1998). Surface protein adhesins of *Staphylococcus aureus*. *Trends Microbiol.* **6**, 484-488.

6. Patti, J.M., Allen, B.L., McGavin, M.J., Gurusiddappa, S. & Höök, M. (1994). MSCRAMM-mediated adherence of microorganisms to host tissues. *Annu. Rev. Microbiol.* **48**, 585-617.
7. Patti, J.M. & Höök, M. (1994). Microbial adhesins recognizing extracellular matrix macromolecules. *Curr. Opin. Cell. Biol.* **6**, 752-758.
8. House-Pompeo, K., Boles, J.O. & Höök, M. (1994). METHODS: A companion to *Methods Enzymol.* **6**, 134-142.
9. Patti, J.M., Boles, J.O. & Höök, M. (1993). Identification and biochemical characterization of the ligand binding domain of the collagen adhesin from *Staphylococcus aureus*. *Biochemistry* **32**, 11428-11435.
10. Patti, J.M., *et al.*, & Höök, M. (1994). The *Staphylococcus aureus* collagen adhesin is a virulence determinant in experimental septic arthritis. *Infect. Immunol.* **62**, 152-161.
11. Symersky, J., *et al.*, & Narayana, S.V.L. (1997). Structure of the collagen-binding domain from a *Staphylococcus aureus* adhesin. *Nat. Struct. Biol.* **4**, 833-838.
12. Emsley, J., King, S.L., Bergelson, J.M. & Liddington, R.C. (1997). Crystal structure of the I domain from integrin $\alpha_2\beta_1$. *J. Biol. Chem.* **272**, 28512-28517.
13. Perona, J.J., Tsu, C.A., Craik, C.S. & Fletterick, R.J. (1997). Crystal structure of an ecotin-collagenase complex suggests a model for recognition and cleavage of the collagen triple helix. *Biochemistry* **36**, 5381-5392.
14. Rich, R.L., *et al.*, & Höök, M. (1999). Trench-shaped binding sites promote multiple classes of interactions between collagen and the adherence receptors, $\alpha_1\beta_1$ integrin and *Staphylococcus aureus* Cna MSCRAMM. *J. Biol. Chem.* **274**, 24906-24913.
15. Lovejoy, B., *et al.*, & Jordan, S.R. (1994). Structure of the catalytic domain of fibroblast collagenase complexed with an inhibitor. *Science* **263**, 375-377.
16. Rich, R.L., Kriekemeyer, B., Owens, R.T., LaBrenz, S., Narayana, S.V.L., Weinstock, G.M., Murray, B.E. and Höök, M. (1999). Ace: A collagen-binding MSCRAMM from *Enterococcus faecalis*. *J. Biol. Chem.* **272**, 26939-26945.
17. Gillaspay, A.F., Patti, J.M., Pratt, F.L., Jr, landolo, J.J. & Smeltzer, M.S. (1997). The *Staphylococcus aureus* collagen adhesin-encoding gene (*Cna*) is within a discrete genetic element. *Gene* **196**, 239-48.
18. Rich, R.L., *et al.*, & Höök, M. (1998). Domain structure of the *Staphylococcus aureus* collagen adhesin. *Biochemistry* **37**, 15423-15433.
19. Harpaz, Y. & Chothia, C. (1994). Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J. Mol. Biol.* **238**, 528-539.
20. Bork, P., Holm, L. & Sander, C. (1994). The immunoglobulin fold: structural classification, sequence patterns and common core. *J. Mol. Biol.* **242**, 309-320.
21. Krestinger, R.H. (1987). Calcium coordination and the calmodulin fold: divergent versus convergent evolution. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 499-510.
22. Josefsson, E., O'Connell, D., Foster, T.J., Durussel, I. & Cox, J.A. (1998). The binding of calcium to the B-repeat segment of SdrD, a cell surface protein of *Staphylococcus aureus*. *J. Biol. Chem.* **273**, 31145-31152.
23. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.
24. Holm, L. & Sander, C. (1994). Searching protein structure databases has come of age. *Proteins* **19**, 165-173.
25. Casasnovas, J.M., Springer T.A., Liu J.H., Harrison, S.C. & Wang, J.-H. (1997). The crystal structure of ICAM-2 reveals a distinctive integrin recognition surface. *Nature* **387**, 312-315.
26. Jones, E.Y., *et al.*, & Stuart, D.I. (1995). Crystal structure of an integrin-binding fragment of vascular cell adhesion molecule-1 at 1.8 Å resolution. *Nature* **373**, 539-544.
27. Wang, J.H., *et al.*, & Osborn, L. (1995). The crystal structure of an N-terminal two-domain fragment of vascular cell adhesion molecule 1 (VCAM-1): a cyclic peptide based on the domain 1 C-D loop can inhibit VCAM-1-alpha 4 integrin interaction. *Proc. Natl Acad. Sci. USA* **92**, 5714-5718.
28. Tan, K., Casasnovas, J.M., Liu, J.H., Briskin, M.J., Springer, T.A. & Wang, J.H. (1998). The structure of immunoglobulin superfamily domains 1 and 2 of MAdCAM-1 reveals novel features important for integrin recognition. *Structure* **6**, 793-801.
29. Wang, J.-H. & Springer, T. (1998). Structural specializations of immunoglobulin superfamily members for adhesion to integrins and viruses. *Immunol. Rev.* **163**, 197-215.
30. Walter, M.R., *et al.*, & Narula, S.K. (1995). Crystal structure of a complex between interferon- γ and its soluble high-affinity receptor. *Nature* **376**, 230-235.
31. Fan, Q.R., Mosyak, L., Winter, C.C., Wagtmann, N., Long, E.O. & Wiley, D.C. (1997). Structure of the inhibitory receptor for human natural killer cells resembles haematopoietic receptors. *Nature* **389**, 96-100.
32. Garman G.C., Kinet, J.-P. & Jadretzky, S. (1998). Crystal structure of the human high-affinity IgE receptor. *Cell* **95**, 961-965.
33. De Vos, A.M., Ultsch, M. & Kossiakoff, A.A. (1992). Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* **255**, 306-312.
34. Leahy, D.J., Aukhil, I. & Erickson, H.P. (1996). 2.0 Å Crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell* **84**, 155-164.
35. Su, Xu-D., Gastinel, L.N., Vaughn, D.E., Faye, I., Poon, P. & Bjorkman, P.J. (1998). Crystal structure of hemolin: A horseshoe shape with implications for homophilic adhesion. *Science* **281**, 991-995.
36. Jiang, F. & Kim, S.H. (1991). "Soft Docking": matching of molecular surface cubes. *J. Mol. Biol.* **219**, 79-102.
37. Chothia, C. & Jones, E.Y. (1997). The molecular structure of cell adhesion molecules. *Annu. Rev. Biochem.* **66**, 823-862.
38. Barclay, A.N. (1993). *The leukocyte antigen facts book*. Academic Press, London.
39. Holmgren, A., Juehn, M.J., Brändén, C.-I. & Hultgren, S.J. (1992). Conserved immunoglobulin-like features in a family of periplasmic pilus chaperones in bacteria. *EMBO J.* **11**, 1617-1622.
40. Komoriya, A., Green, L.J., Mervic, M., Yamada, S.S., Yamada, K.M. & Humphries M.J. (1991). The minimal essential sequence for a major cell type-specific adhesion site (CS1) within the alternatively spliced type III connecting segment domain of fibronectin is leucine-aspartic acid-valine. *J. Biol. Chem.* **266**, 15075-15079.
41. Heringa, J. & Taylor, W.R. (1997). Three-dimensional domain duplication, swapping and stealing. *Curr. Opin. Struct. Biol.* **7**, 416-421.
42. Bennett, M.J., Schlunegger, M.P. & Eisenberg, D. (1995). 3D domain swapping: A mechanism for oligomer assembly. *Protein Sci.* **4**, 2455-2468.
43. Schlunegger, M.P., Bennett, M.J. & Eisenberg, D. (1997). Oligomer formation by 3D domain swapping: A model for protein assembly and misassembly. *Adv. Protein Chem.* **50**, 61-122.
44. Pokkuluri, P.R., *et al.*, & Schiffer, M. (1997). A region flip as a result of a single amino acid substitution. *Structure* **6**, 1067-1073.
45. Deivanayagam, C.C.S., *et al.*, & Narayana, S.V.L. (1998). Crystallization and preliminary X-ray analysis of B domain fragments of a *Staphylococcus aureus* collagen-binding protein. *Acta Crystallogr. D* **55**, 525-527.
46. McRee, D.E. (1993). *Practical Protein Crystallography*. Academic Press, California.
47. PHASES. (1995). A program package for processing and analyzing diffraction data for macromolecules. *Methods Enzymol.* **277**, 590-620.
48. CHAIN: A crystallographic model building program: Version 7.0, Baylor College of Medicine, Houston, Texas.
49. Brünger, A.T. (1996). *X-PLOR (version 3.85): A system for X-ray crystallography and NMR*, Yale University Press, New Haven, CT.
50. Brünger, A.T. (1992). Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475.
51. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. D* **47**, 110-119.
52. Kleywegt, G.J. & Jones, T.A. (1997). Model rebuilding and refinement practice. *Methods Enzymol.* **276**, 208-330.
53. Adams, P.D., Pannu, N.S., Read, R.J. & Brünger, A.T. (1997). Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc. Natl Acad. Sci. USA* **94**, 5018-5023.
54. Navaza, J. (1994). AmoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157-163.
55. Carson, M. (1997). RIBBONS. *Methods Enzymol.* **277**, 493-505.
56. Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L. & Wiley, D.C. (1987). The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* **329**, 506-512.
57. Connolly, M.L. (1993). The molecular surface package. *J. Mol. Graphics.* **11**, 139-141.