Surface-exposed loops and an acidic patch in the Scl1 protein of group A Streptococcus enable Scl1 binding to wound-associated fibronectin

Received for publication, February 2, 2018, and in revised form, March 29, 2018 Published, Papers in Press, April 2, 2018, DOI 10.1074/jbc.RA118.002250

Dudley H. McNitt1,‡, Soo Jeon Choi,§, Douglas R. Keene5, Livingston Van De Water4, Flavia Squeglia1, Rita Berisio1, and Slawomir Lukomska1,‡

From the 1Department of Microbiology, Immunology, and Cell Biology, West Virginia University School of Medicine, Morgantown, West Virginia 26506, the 2Micro-imaging Center, Shriners Hospital for Children, Portland, Oregon 97239, the 3Departments of Surgery and Regenerative and Cancer Cell Biology, Albany Medical College, Albany, New York 12208, and the 4Institute of Biostructures and Bioimaging, Italian National Research Council, Via Mezzocannone 16, I-80134 Naples, Italy

Edited by Norma M. Allerwell

Keratinized epidermis constitutes a powerful barrier of the mucosa and skin, effectively preventing bacterial invasion, unless it is wounded and no longer protective. Wound healing involves deposition of distinct extracellular matrix (ECM) proteins enriched in cellular fibronectin (cFn) isoforms containing extra domain A (EDA). The streptococcal collagen-like protein 1 (Scl1) is a surface adhesin of group A Streptococcus (GAS), which contains an N-terminal variable (V) domain and a C-terminally located collagen-like domain. During wound infection, Scl1 selectively binds EDA/cFn isoforms and laminin, as well as low-density lipoprotein (LDL), through its V domain. The trimeric V domain has a six-helical bundle fold composed of three pairs of anti-parallel $\alpha$-helices interconnected by hypervariable loops, but the roles of these structures in EDA/cFn binding are unclear. Here, using recombinant Scl (rScl) constructs to investigate structure–function determinants of the Scl1–EDA/cFn interaction, we found that full-length rScl1, containing both the globular V and the collagen domains, is necessary for EDA/cFn binding. We established that the surface-exposed loops, interconnecting conserved $\alpha$-helices, guide recognition and binding of Scl1-V to EDA and binding to laminin and LDL. Moreover, electrostatic surface potential models of the Scl1-V domains pointed to a conserved, negatively charged pocket, surrounded by positively charged and neutral regions, as a determining factor for the binding. In light of these findings, we propose an updated model of EDA/cFn recognition by the Scl1 adhesin from GAS, representing a significant step in understanding the Scl1–ECM interactions within the wound microenvironment that underlie GAS pathogenesis.

Group A Streptococcus (GAS)3 is a human-adapted pathogen responsible for over 700 million disease-associated infections across the globe each year and an estimated healthcare cost of 224–539 million dollars in the United States alone (1, 2). GAS asymptotically colonizes the general population of adults (5–15%) and children (20–30%), and most of the morbidity is associated with superficial infections of the throat and skin (3). In addition, bacterial spread may result in severe soft tissue and systemic infections, like necrotizing fasciitis and streptococcal toxic shock syndrome (4), or may lead to post-infectious sequelae, such as acute rheumatic fever and rheumatic heart disease, post-streptococcal glomerulonephritis, and pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections (5, 6). For the initial colonization of the human host, GAS requires a portal of entry, which is constituted by a breach in the epithelial cell lining (referred to hereafter as a wound) that allows the pathogen to spread into deeper tissue and blood. GAS expresses numerous surface adhesins that bind to host extracellular matrix (ECM) (7). ECM binding facilitates initial localized colonization, during which GAS forms tissue microcolonies embedded in the glycocalyx (8). The interplay between GAS adhesins and the host’s ECM components, stable tissue microcolony formation, and superficial versus invasive infection are of significant clinical interest.

Among prominent GAS surface proteins are the streptococcal collagen-like (Scl) proteins 1 and 2 (9–13). Scl1 and Scl2 are homotrimeric proteins that have similar domain organization, which is composed of an N-terminal globular sequence-variable (V) domain, a collagen-like (CL) domain, and the C-terminal cell wall–anchoring domain (9, 10). The CL domain of both Scl1 and Scl2 consists of varying numbers of Gly-Xaa-Yaa col-

1Supported by the Integrative Graduate Education and Research Training (IGERT) Traineeship for Research and Education in Nanotoxicology under Award 1144676 and by the Dr. Jennifer Gossling Scholarship in Immunology, and Cell Biology, West Virginia University School of Medicine, Morgantown, West Virginia 26506, the Micro-imaging Center, Shriners Hospital for Children, Portland, Oregon 97239, the Departments of Surgery and Regenerative and Cancer Cell Biology, Albany Medical College, Albany, New York 12208, and the Institute of Biostructures and Bioimaging, Italian National Research Council, Via Mezzocannone 16, I-80134 Naples, Italy.

2To whom correspondence should be addressed: Dept. of Microbiology, Immunology, and Cell Biology, West Virginia University School of Medicine, 2095 Heath Sciences North, PO Box 9177, Morgantown, WV 26506. Tel.: 304-293-6405; Fax: 304-293-7823; E-mail: slukomska@hsc.wvu.edu.

This work was supported in part by National Institutes of Health Grants AI50666 and AI083683, as well as West Virginia University HSC Bridge Grant Funding (to S. L.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

This is an Open Access article under the CC BY license.

3The abbreviations used are: GAS, group A Streptococcus; Scl, streptococcal collagen-like protein; rScl, recombinant Scl; ECM, extracellular matrix; V, variable; CL, collagen-like; CFH, complement factor H; CFHR1, CFH-related protein 1; LDL, low-density lipoprotein; Fn, fibronectin; pFn, plasma fibronectin; cFn, cellular Fn; EDA and EDB, extra domain A and B, respectively; apoB100, apolipoprotein B100; PDB, Protein Data Bank; Lm, laminin; rEDA, recombinant EDA; HRP, horseradish peroxidase; Ab, antibody.
lagent repeats that adopt a right-handed triple-helix structure, similar to mammalian collagen (14–16); the Scl1-CL domain may interact with human collagen-binding integrins, promoting GAS internalization (17–19). The outermost-exposed V domains vary in amino acid sequence between and within Scl1 and Scl2 variants and are typically associated with M-type antigen (9, 11). Despite significant sequence variation, the V domains of Scl1 and Scl2 were predicted to contain two conserved α-helices, spaced by hypervariable regions (11, 20). Recent reports of a first crystal structure for the Scl2-V domain validated these secondary structure predictions (21, 22). The recombinant Scl2-V globular domain forms a trimeric structure, which folds into a six-helix bundle, with three pairs of anti-parallel α-helices interconnected by hypervariable loops (22). Despite both Scl1 and Scl2 having similar structural organization, they play different roles in GAS pathogenesis (23). Whereas the number of Scl2 targets identified remains scanty (24), Scl1 variants from diverse GAS M types bind several host proteins. Ligands include tissue components, cellular fibronectin (cFn), and laminin (Lm), as well as blood components, low- and high-density lipoproteins, and complement Factor H (CFH) and Factor H-related protein 1 (CFHR1) (20, 25–28). It is not clear how the variable sequences of the Scl1-V domain bind the same ligands, but two mutually exclusive binding patterns were observed. Scl1 proteins from several M types bind to ECM proteins, cFn and Lm, and low-density lipoprotein (LDL), but not FH and CFHR1; the converse is also true.

One of the major ECM targets for GAS within wounded tissue is fibronectin (Fn) (7). Fn is a high-molecular weight glycoprotein that is found throughout the human body in two main forms, either as a soluble circulating plasma fibronectin (pFn) secreted by hepatocytes or as an insoluble cellular fibronectin (cFn) secreted by a variety of cells within tissues (29). All forms of Fn are transcribed from a single gene, FNI, and are composed of three modules of fibronectin repeats: types I, II, and III (29, 30). There are about 20 isoforms of human cFn that include combinations of the type III repeats (e.g. extra domain A (EDA/EIIIA) and B (EDB/EIIIB)) and a varied number of V repeats (30–32). These isoforms differ by spatial-temporal distribution following injury (33, 34). Specifically, the EDA-containing isoforms are normally expressed during embryogenesis, although they are found in negligible levels in most adult tissues, except in healing wounds, where EDA/cFn is highly up-regulated (33, 35). The EDA fold is a β-sandwich structure formed by seven antiparallel β-strands (i.e. A, B, C, C’, D, E, and F) connected by loops (36, 37). The C-C’ loop is an important site for cell attachment, via integrin receptors α5β1 and α6β1 (38, 39), and attachment to the C-C’ loop by these integrins has been shown to regulate important tissue-repair and wound-healing functions (40–44).

Here, we established that surface-exposed loops within the V domain structure guide Scl1 to bind EDA/cFn, as well as two other Scl1 ligands, laminin and apolipoprotein B100 (apoB100) of LDL. A second conserved constituent within variable Scl1-V domain is a common topography of the electrostatic surface potential surrounding guiding loops, which includes a negatively charged patch formed by collective Asp and Glu residues contributed by all three monomeric chains. The proposed model accommodates binding of common ligands by phylogenetically diverse Scl1 variants. This work implies the evolutionary response of the Scl1-V domain to selective pressure, resulting in the retention of the conserved structure and function of ligand binding. This underscores the importance of Scl1-binding to EDA-containing fibronectins deposited at the pathogen portal of entry.

Results

The V domains of Scl1 and Scl2 contain conserved helix-loop-helix structures.

Two streptococcal collagen-like proteins, Scl1 and Scl2, display a similar “lollipop-like” structural organization (Fig. 1A), which is composed of an N-terminal globular V-domain head and a C-terminal CL-domain stalk (14). Multiple-sequence alignments illustrate a significant variations in the V regions, both between and within the Scl1 and Scl2 families (Fig. 1B). Despite sequence variation, several secondary structure prediction algorithms deduced that both Scl1-V and Scl2-V domains contain two conserved α-helices interconnected by hypervariable regions. Crystallography data rendered for the recombinant Scl2-V construct, derived from M3-type GAS, revealed that rScl2.3-V protein forms a homotrimer, with three pairs of conserved α-helices forming a six-helix bundle and hypervariable regions forming surface-exposed loops (PDB code 4NSM). To examine the conservation of the helix-loop-helix motif, homology models of two sequence-diverse Scl1- and Scl2-V domains were generated (Fig. 1C), using the structure of the Scl2.3-V domain as template. The homology model structure of the Scl1.1-V domain from M1-type GAS was based on the consensus sequence alignment (24% sequence identity on 65 amino acid residues), and the model of the Scl2.28-V domain from M28 GAS was obtained with the same approach (51.5% identity with Scl2.3-V sequence on 74 residues). Homology models predict with high confidence that Scl1.1-V (E = 3.5 × 10⁻¹⁶) and Scl2.28-V (E = 1.8 × 10⁻⁴⁰) domains each adopt a six-helix bundle trimeric structure similar to that of Scl2.3-V template. The solved crystal structure and homology models rendered here show that globular domains of Scl1.1 and Scl2.28 variants can trimerize without the presence of the CL domain, an observation that was also found in Scl2 proteins from other M types (45). Furthermore, we hypothesized that trimeric rScl1-V constructs will retain their binding function to ECM ligands.

The conserved structure of Scl1-V domain requires the collagen domain for ligand binding.

Full-length rScl1 proteins, derived from different GAS M types, bind to host cFn via interactions with EDA, which is mediated by Scl1-V domain (25, 46). Here, we examined whether the V domain–alone rScl1 constructs bind to cFn and rEDA. Full-length recombinant Scl1 constructs, derived from M1-type (rScl1.1FL) and M3-type (rScl1.3FL) strains, as well as their corresponding V domain–only proteins (rScl1.1V and rScl1.3V) were immobilized in wells and tested for ECM–ligand binding. Unexpectedly, neither rScl1.1V nor rScl1.3V constructs bound rEDA and cFn, whereas the corresponding full-length rScl1FL control proteins bound both ligands (Fig. 2). Thus, whereas the V domains of Scl1 proteins can fold indepen-
Structural adaptation of Scl1-V domain

Figure 1. Conserved structure of the Scl1 and Scl2 globular domains. A, cartoon model of full-length Scl proteins. The globular N-terminal V domain, central rod-shaped CL domain, and C-terminal cell wall anchor domain are shown. B, multiple sequence alignment of the V regions of Scl1 and Scl2 variants. MSAs were generated using MUSCLE; amino acid residues are colored according to their physico-chemical character. Red, negative (Asp and Glu); blue, positive (Arg, Lys, and His); gray, hydrophobic, aliphatic, or aromatic (Ile, Leu, Val, Ala, Met, Phe, Tyr, and Trp); green, polar (Ser, Thr, Gin, Asn, Gly, and Pro). Predicted α-helices, N-terminal helix 1 (gray) and C-terminal helix 2 (blue), are shown below the alignments (H). C, homology models of the Scl1.1-V and Scl2.28-V domains. Models were generated based on the crystal structure of the Scl2.3-V variant (PDB code 4NSM). Modeled α-helices, N-terminal helix 1 (gray) and C-terminal helix 2 (blue), are shown.

dently of their CL domains, there are currently unknown structural constraints that require the presence of continuous CL domains for EDA/cFn recognition.

The loop region of Scl1-V domain plays a central role in ECM recognition and binding

The binding site for the Scl1-V domain within the loop segment was located between the C- and C'-β-strands of the EDA β-sandwich (46). Here, we hypothesized that EDA binds to surface-exposed loops of the Scl1.1-V domain, which we have since identified in the crystal structure and deduced in the homology models (Fig. 1C). To test this hypothesis, we generated two full-length chimeric rScl1.1 variants: (i) we inserted the 22-amino acid loop sequence of the ECM binding–positive rScl1.1 variant (residues 25–46 of the V domain) in place of the respective loop sequence of the ECM binding–negative protein...
rScl2.28 (residues 20 – 45), resulting in the chimeric construct rScl.chi1FL (Fig. 3A); and (ii) the reverse was also constructed, by inserting the 26-amino acid loop region of rScl2.28-V in place of loop sequence of rScl1.1-V, resulting in rScl.chi2FL. To identify the proper site for inserting these sequences, we considered secondary structure predictions, computed with JPRED; resulting models were then minimized using GROMACS.

To ensure that we did not substantially alter the overall structure of the chimeric proteins, both rScl.chi1FL and rScl.chi2FL were examined by CD spectroscopy and EM of rotary-shadowed samples (Fig. 3B). The wavelength scans of the original rScl1.1FL and rScl2.28FL proteins each exhibit a positive peak of molar ellipticity at 220 nm, \( \theta \) 220, (14), which is a characteristic of the collagen triple helix (47). The CD spectra of both chimeric proteins, rScl.chi1FL and rScl.chi2FL, rendered at 20 °C, show peak ellipticity values at 220 nm (Fig. 3B, solid lines), thus demonstrating that both chimeric constructs adopt triple helix structures. These peaks were not observed at 50 °C, indicating that each triple helix unfolded (Fig. 3B, dashed lines).

The domain organization of both chimeric rScls was further examined via rotary shadowing and subsequent EM. This technique has been used previously by our laboratory and others to investigate the structure of multidomain collagens, containing collagenous and noncollagenous domains (14, 48). Both chimeric rScls displayed the characteristic two-domain lollipop-like organization, with the CL domain, as the stalk, and the globular V-domain head (Fig. 3B). Thus, swapping the loop regions between rScl1.1FL and rScl2.28FL did not alter the overall structural organization and triple helices in the chimeric proteins.

We next examined the binding characteristics of both chimeric proteins to rEDA and cFn ligands, with rScl1.1FL and rScl2.28FL serving as positive and negative controls, respectively. Binding data show that rScl.chi1FL chimera, containing donor-loop sequence of rScl1.1FL, binds both rEDA and cFn, whereas the reverse construct rScl.chi2FL did not bind to either ligand, similar to negative control protein rScl2.28FL (Fig. 3C). These data demonstrate the importance of the loop region in ECM binding for Scl1.1FL and hypothetically for other Scl1 variants.

To assess the importance of the Scl1.1-V loop in rEDA recognition in the context of whole-GAS cells, we complemented in trans the scl2-inactivated mutant (\( \Delta s c l 2 \)) of M1-type strain (9) for the expression of original Scl1.1 (\( \Delta s c l 1::s c l 1.1 \)) or chimeric Scl.chi2 protein (\( \Delta s c l 1::s c l 1.2 \)), using plasmids pSL620 and pSL621 (Table 1). The M1-type WT, and \( \Delta s c l 1 \) mutant strains were used as positive and negative controls, respectively (Fig. 3D, left). Expression of surface Scl1.1 and Scl.chi2 was assessed via Western immunoblotting of the cell wall fractions. Additionally, rScl1.1FL protein was loaded, as a positive control (rScl1.1FL is smaller on the SDS gel because it is devoid of the cell wall region of native Scl1.1 protein).

We next investigated the rEDA binding to whole GAS cells by flow cytometry (Fig. 3D, right). As expected, the WT and \( \Delta s c l 1::s c l 1.1 \)-complemented cells bind rEDA, although the complemented strain binds 4 times more ligand due to overexpression of the surface Scl1.1 protein. In contrast, neither \( \Delta s c l 2 \) mutant nor complemented GAS cells, overexpressing the Scl.chi2 protein, bind rEDA. Thus, GAS cells that express native Scl1 adhere to the cell surface but lose this capacity once cells express a mutated variant with loop substitution from Scl2 protein, which does not bind EDA/cFn.

The C terminus of the loop region of the Scl1-V domain plays a central role in ECM recognition and binding

We next designed additional chimeric rScl constructs to identify the specific region of the loop involved in rEDA and cFn binding. Instead of replacing the entire 22 amino acids, the Scl1.1-V loop was substituted into Scl2.28-V loop in three segments (Fig. 4A). For the first construct, only the N-terminal 11 amino acids of the Scl1.1-V loop (residues 25–35) were substituted in place of the first 11 amino acids of the corresponding Scl2.28-V sequence (residues 20–30), resulting in construct rScl.chiNFL. Similarly, the C-terminal 11 amino acids of Scl1.1-V (residues 36–46) were substituted in place of the...
Structural adaptation of Scl1-V domain

A

\begin{align*}
&\text{rScl1.1} \\
&\quad \text{(Helix 1)} \\
&\quad \text{Loop} \\
&\quad \text{Helix 2}
\end{align*}

\begin{align*}
&\text{rScl2.28} \\
&\quad \text{(Helix 1)} \\
&\quad \text{Loop} \\
&\quad \text{Helix 2}
\end{align*}

\begin{align*}
&\text{rScl,chi1} \\
&\quad \text{(Helix 1)} \\
&\quad \text{Loop} \\
&\quad \text{Helix 2}
\end{align*}

\begin{align*}
&\text{rScl,chi2} \\
&\quad \text{(Helix 1)} \\
&\quad \text{Loop} \\
&\quad \text{Helix 2}
\end{align*}

**Figure 3. Surface exposed loop of the Scl1.1-V domain is central in ECM recognition and binding.** Full-length chimeric rScl proteins were generated by exchanging V-region loop sequences between ECM binding–positive and –negative rScl proteins; the resulting chimeric rScl,chiFL constructs were characterized structurally by CD spectroscopy and EM, as well as functionally for ECM binding. A, amino acid sequences of the V domains of native (rScl1.1FL and rScl2.28FL) and chimeric (rScl,chi1FL and rScl,chi2FL) proteins. rScl,chi1FL resulted from the insertion of 22 amino acids from the ECM binding–negative variant rScl2.28FL, whereas rScl,chi2FL resulted from the reverse replacement; inserted sequences are shown in italic type, and numbers indicate residue number within the V domain. B, structural characterization of the chimeric rScl,chiFL constructs. Left, CD wavelength scans of folded (20 °C, solid lines) or denatured (50 °C, dashed lines) rScl,chiFL proteins are shown. Right, structural organization viewed by EM of the rotary-shadowed rScl,chi1FL and rScl,chi2FL shows characteristic two-domain organization. Bars, 50 nm. C, ECM binding to full-length original and chimeric rScls, containing partial-loop substitutions, were examined by CD spectroscopy and EM to confirm the essential structural characteristics (Fig. 4B). The CD spectra of all three chimeric proteins showed peak ellipticity values at 220 nm at 20 °C, whereas these peaks were not observed at

C-terminal 15 amino acids of Scl2.28 (residues 31–45), resulting in construct rScl,chiCFL. The third construct, rScl,chiFL, contains the internal 11 amino acids of Scl1.1-V (residues 30–40) in place of the central (residues 25–36) 11 amino acids of Scl2.28-V.

All three chimeric rScls, containing partial-loop substitutions, were examined by CD spectroscopy and EM to confirm the essential structural characteristics (Fig. 4B). The CD spectra of all three chimeric proteins showed peak ellipticity values at 220 nm at 20 °C, whereas these peaks were not observed at
Structural adaptation of Scl1-V domain

Table 1

| Plasmid | Description | Source |
|---------|-------------|--------|
| pASK-IBA2 | E. coli expression vector, C-terminal Streptag II | IBA (Göttingen, Germany) |
| pSL503 | pASK-IBA2 encoding rScl1.3FL | Ref. 74 |
| pSL514 | pASK-IBA2 encoding rScl1.3V | Ref. 49 |
| pSL144 | pASK-IBA2 encoding rScl1.1FL | Ref. 14 |
| pSL158 | pASK-IBA2 encoding rScl1.1V | Ref. 14 |
| pSL161 | pASK-IBA2 encoding rScl1.2FL | Ref. 15 |
| pSL163 | pASK-IBA2 encoding rScl1.2FL | Ref. 14 |
| pSL600 | pASK-IBA2 encoding rScl.chi1FL | This study |
| pSL601 | pASK-IBA2 encoding rScl.chi1FL | This study |
| pSL602 | pASK-IBA2 encoding rScl.chi1FL | This study |
| pSL603 | pASK-IBA2 encoding rScl.chi1FL | This study |
| pSL613 | pASK-IBA2 encoding rScl.chi2FL | This study |
| pSL614 | pASK-IBA2 encoding rScl.chi3FL | This study |
| pQE-30 | N-terminal His tag E. coli expression vector | Qiagen |
| pQE-EDA | pQE-30 encoding rEDA | Ref. 70 |
| pSB027 | E. coli-GAS shuttle plasmid encoding GFP | Ref. 73 |
| pSL620 | pSB027 encoding slcl.1 allele from MGAS 5005 | This study |
| pSL621 | pSL620 encoding Scl.chi2FL | This study |

50 °C. Rotary shadowing images show that each chimeric rScl adopts a lollipop-like structural organization, with both CL-domain stalks and globular V domains. Thus, like whole-loop substitutions, partial-loop substitutions did not impact the structures of the chimeric rScls.

Binding data to cFn and rEDA indicated that only one of the three chimeras with partial-loop substitutions, rScl.chiCFL (C-terminal substitution), binds to cFn and rEDA, similar to the whole-loop-substitution construct rScl.chi1FL and the loop donor protein rScl1.1FL (Fig. 4C); the other two chimeric proteins, as well as binding-negative control protein rScl2.28FL, do not bind to cFn and rEDA. We observed substantially lower OD values recorded during rEDA binding to rScl.chi1FL and rScl.chiCFL, compared with rScl1.1FL. This may reflect less than optimal loop conformation in chimeric proteins associated with rScl2 backbone. Additional binding assays were carried out with increasing concentrations (0.039 – 5.0 μM) of rEDA ligand and rScl proteins to assess binding specificity (Fig. 4D). Concentration-dependent binding of rEDA to rScl.chi1FL, rScl.chiCFL, and control protein rScl1.1FL plateaued between 2.5 and 5.0 μM concentrations. In contrast, neither rScl.chiNFL, rScl.chiFL, rScl.chi2FL, nor control rScl2.28FL showed significant binding to rEDA. Therefore, these data provide evidence that the C-terminal 11 amino acids of rScl1.1-V loop are sufficient for conferring EDA/cFn binding, when inserted into binding-negative rScl2.28 protein.

Distribution of the charged residues on the surface of V domain hints at a binding mechanism for EDA/cFn

Our initial examination of the multiple-sequence alignment of Scl1- and Scl2-V domains failed to identify any conserved pattern of residues that could be responsible for EDA/cFn binding (Fig. 1A). Our subsequent experiments, employing the chimeric loop sequences, highlighted a pattern of charged amino acids (e.g. Lys48(Scl1.1), Glu46,47(Scl1.1), and Asp48(Scl1.1)) that were present within the uppermost portion of the Scl1-V loops and transitioning to inner α-helices but were absent in the respective regions of the ECM binding–negative Scl2-V loops (Fig. 1B). Therefore, we sought to investigate the surface potentials of the Scl1-, Scl2- and chimeric rScl.chi-V domains.

Electrostatic surface maps were computed from the homology models of the ECM binding–positive and –negative proteins. Electrostatic maps computed for Scl1.1- and Scl2.28-V domains reveal striking differences in the distribution of charges across their surfaces. A uniform, negatively charged electrostatic potential characterizes the ECM binding–negative Scl2.28-V on the frontal apical region of the molecule, opposite to the triple-helix attachment (Fig. 5, black ovals). These charges are due to acidic residues located on the C-terminal part of the first outer α-helix, the loop region, and N-terminal part of the second inner α-helix. In this same area, the electrostatic potential map of the binding-positive Scl1.1-V domain presents a central negatively charged pocket or cleft (Fig. 5, black arrowhead), surrounded by a neutral and basic crown surface. Although less pronounced, the same features of electrostatic potential, with a central negatively charged region surrounded by a neutral/basic surface, are present in maps of the binding-positive chimeric rScl.chi1 and rScl.chiC-V domains, but not binding-negative chimeric rScl proteins. Therefore, the specific charge pattern of the V-domain head strongly differentiates between binding-positive and binding-negative V domains. These considerations suggest that interactions with partnering molecules occur through electrostatic recognition of the observed charge pattern.

*Scl1-V loop is responsible for the ECM binding in phylogenetically distant Scl1 variants*

Phylogenetic analysis of the amino acid sequences of the Scl1-V and Scl2-V domains show that both Scl1 and Scl2 form two distinct clusters (Fig. 6A). Within Scl1, there are two observed ligand-binding patterns. The first pattern is represented by Scl1 variants (Fig. 6A, black boxes) that bind host ECM proteins, cFn/EDA and Lm (25) and apoB100-LDL (20). The second pattern is represented by two Scl1 variants (Fig. 6A, black dots) that bind host complement components, CFH and CFHR1 (26, 27), suggesting the potential for a unique adaptation of Scl1 proteins in M6 and M55 GAS. In contrast, the majority of rScl1 constructs investigated show the former binding pattern, and these Scl1 variants are distributed all over many branches of the Scl1 cluster, highlighting the importance of ECM and LDL binding among GAS of various M types (20, 25, 46, 49, 50).

We next investigated whether the loop region of the phylogenetically distant Scl1.28-V variant, found in M28-type GAS, which shares 46% sequence identity with Scl1.1-V sequence (Fig. 6A, arrows), will support EDA/cFn binding. To do this, we substituted the loop region of Scl1.28-V (residues 25–46) in place of the loop region of Scl2.28-V (residues 20–45), resulting in the chimeric construct rScl.chi3FL (Fig. 6B). The CD spectra of rScl.chi3FL showed peak ellipticity at 220 nm at 20 °C, but not at 50 °C, and rotary-shadowed images show characteristic lollipop-like domain organization (Fig. 6C). Binding data to cFn and rEDA show that rScl.chi3FL binds both ligands (Fig. 6D). Thus, as in Scl1.1-V, the loop region of the divergent Scl1.28-V variant confers binding to both rEDA and cFn. Despite significant efforts, we could not generate a homology
Figure 4. The C-terminal 11 amino acids of the loop segment of the Scl1.1-V domain are central in ECM recognition and binding. Chimeric rScl proteins were generated by exchanging portions of the loop from ECM binding–positive rScl1.1-V with respective loop sequences of the binding-negative protein rScl2.28FL; the resulting chimeric rScl.chi constructs were characterized structurally by CD spectroscopy and EM, as well as functionally for ECM binding by ELISA. A, amino acid sequences of the V domains of native (rScl1.1FL and rScl2.28FL) and chimeric (rScl.chiNFL, rScl.chiCFL, and rScl.chiIFL) proteins. rScl.chiNFL resulted from the insertion of the N-terminal 11 amino acids of the rScl1.1FL-loop sequence in place of the corresponding N-terminal loop region of the rScl2.28FL, whereas rScl.chiCFL resulted from the replacement of the C-terminal 11 amino acids and rScl.chiIFL from the replacement of the internal 11 amino acids; inserted sequences are shown in italic type, and numbers indicate residue number within the V domain. B, structural characterization of the chimeric rScl constructs. Left, CD wavelength scans of folded (20 °C, solid lines) or denatured (50 °C, dashed lines) rScl.chi proteins are shown. Right, structural organization viewed by EM of the rotary-shadowed rScl.chiNFL, rScl.chiCFL, and rScl.chiIFL shows characteristic two-domain organization. Bars, 50 nm. C, ECM binding to chimeric rScl constructs by ELISA. rScl proteins were immobilized onto Strep-Tactin–coated microplate wells and incubated with ECM ligands cFn and rEDA. Primary anti-ECM Abs and HRP-conjugated secondary Abs were used for ligand detection. Graphic bars indicate the mean A415 nm normalized against BSA controls. Statistical analysis was calculated using Student’s two-tailed t test, from three independent experiments, each performed in triplicate wells (n = 3 ± S.D.); *, p ≤ 0.05; **, p ≤ 0.01; ***, p ≤ 0.001. Solid (rEDA) and dashed (cFn) lines indicate threshold A415 nm/2 S.D. values (error bars) recorded for experimentally validated binding-negative rScl2.28FL control protein. D, concentration-dependent binding of rEDA to rScl constructs. rScl proteins were immobilized onto Strep-Tactin–coated microplate wells and incubated with increasing concentrations of rEDA. Wells were further developed, as described above. Statistical analysis was calculated using two-way analysis of variance, from two independent experiments, each performed in triplicate wells (n = 2 ± S.D.); *, p ≤ 0.05; **, p ≤ 0.01; ***, p ≤ 0.001. Statistical significance evaluates difference between bindings of each rScl1-chimeric protein to binding-negative control rScl2.28FL.
model of Scl1.28-V domain with a satisfactory global model quality estimation, due to the low sequence identity with the template structure of Scl2.28-V domain, close to 12%.

*The same structural constraints determine recognition and binding of the Scl1-V domain to laminin and apolipoprotein B100-LDL*

Here, we sought to understand whether the same structural constraints apply during Scl1 binding to Lm and apoB100-LDL (Fig. 7). Both rScl1.1FL and rScl1.3FL bound to Lm and apoB100, whereas their respective V-domain-only counterparts, rScl1.1V and rScl1.3V, demonstrated negligible or no binding to both ligands (Fig. 7A). Chimeric constructs rScl.chi1FL, rScl.chiCFL, and rScl.chi3FL also bind to Lm and apoB100-LDL, whereas binding-negative control proteins rScl2.28FL and rScl.chi2FL (reverse loop replacement to that in rScl.chi1FL) did not (Fig. 7, B and C); in addition, both rScl.chiN and rScl.chI did not bind to Lm and apoB100-LDL (data not shown). Collectively, our data show that (i) full-length Scl1 protein is required for binding to all ligands, and (ii) the C-terminal 11 amino acids of the Scl1-V loop region are sufficient to mediate binding of Scl1 to its sequence- and structure-diverse ligands.

**Discussion**

Adhesion to Fn is important for the pathogenesis of group A *Streptococcus* (7) by promoting the development of glycocalyx-embedded tissue microcolonies (*in vivo* biofilms) that are seen in clinical specimens collected from cases of pharyngitis and skin infections (8, 51). We previously reported that the streptococcal collagen-like protein 1, or Scl1, uniquely adheres to cFn isoforms expressed in wounded tissue and that interactions involve the V globular domain of Scl1 and EDA of cFn (25, 46). Here, we report on the structural determinants involved in Scl1 binding to cFn, in the context of diversity among fibronectin-binding proteins of GAS and other bacterial pathogens. We also report that the same determinant of the Scl1 adhesin is responsible for binding to Lm and apoB100-LDL complex.

Several mammalian collagens, such as collagen type I, IV, and X, contain an N-terminal and/or C-terminal noncollagenous domain, which serves as a nucleation or trimerization domain, thereby facilitating proper folding of the three adjacent α-chains into a mature collagen triple helix (52). Similarly, bacterial collagen-like proteins contain noncollagenous trimerization domains, often of unknown functions (14, 45). The BclA of *Bacillus anthracis* endospores possesses a C-terminal noncollagenous trimerization domain, BclA-CTD, for CL-domain folding and increased thermal stability (53). Two different recombinant BclA-CTD constructs had the same crystal structure, similar to a noncollagenous globular domain of C1q, independent of whether these constructs included the continuous CL domain (54, 55). Likewise, the N-terminal V domains of Scl1 and Scl2 assisted in the trimerization of homologous and heterologous CL domains (45, 56), and the recombinant Scl2-V-only construct formed a trimeric six-helix bundle structure (22). However, here we found that the rScl1-V-only constructs, derived from the corresponding full-length rScl1FL proteins, did not bind rEDA, cFn, Lm, and apoB100-LDL. We propose that the triple helix imposes a restraint on the V domain in a way to destabilize it and make it more prone to conformational changes. We expect a slight effect on the tertiary structure but not on the secondary structure. A reason for the restraint may be the one-residue stagger of triple helices,
whereas the V domains are perfectly aligned; these structural features probably distort the V domain slightly, to make it more flexible.

Whereas Scl1 is unique among GAS Fn-binding proteins in recognizing the EDA type III repeat of cFn, several other bacterial adhesins may in fact recognize the EDA, due to their selective binding to cFn but not pFn, such as FimH of pathogenic Escherichia coli, OppA, and Msp of Treponema denticola and Tp0155 of Treponema pallidum (57). In addition, numerous bacterial surface proteins bind Fn type III repeats other than EDA that are present in both pFn and cFn, including proteins ShdA, ComE1, Pap31, and Ali, that are associated with the outer membrane of Gram-negative bacteria (58–61) and fibrillar surface proteins Embp and PavA-B of Gram-positive bacteria (62, 63). Two pneumococcal Fn-binding proteins, PavA and PavB, bind to multiple FnIII repeats via common epitopes

---

**Figure 6. Conservation of ECM recognition and binding by phylogenetically distant Scl1 variant.**

A. Phylogenetic tree of Scl1-V and Scl2-V regions. Branch lengths are scaled to genetic distance, and bootstrap values are shown for all nodes. Scl1 variants with known ECM-binding capacities are boxed, whereas variants known to bind complement components, Factor H and Factor H-related protein 1, are marked with dots. ECM-binding–positive variants, Scl1.28 and Scl1.1, and binding-negative variant Scl2.28, used for loop switching, are marked by arrows.

B. Amino acid sequences of the V domains of native (rScl1.28FL and rScl2.28FL) and chimeric (rScl.chi3FL) proteins. rScl.chi3FL resulted from the insertion of the 22 amino acids from the ECM-binding–positive variant rScl1.28FL in place of the corresponding loop region of the binding-negative variant rScl2.28FL; the inserted sequence is shown in italic type, and numbers indicate residues within the V domain.

C. Structural characterization of rScl.chi3FL. Left, CD wavelength scans of folded (20 °C, solid line) or denatured (50 °C, dashed line) rScl.chi3FL protein are shown. Right, structural organization viewed by EM of the rotary-shadowed rScl.chi3FL shows the characteristic two-domain organization.

D. ECM binding to rScl.chi3FL by ELISA. rScl proteins were immobilized onto Strep-Tactin–coated microplate wells and incubated with ECM ligands cFn and rEDA. Primary anti-ECM Abs and HRP-conjugated secondary Abs were used for ligand detection. Graphic bars indicate the mean A415 nm normalized against BSA controls. Statistical analysis was calculated using Student’s two-tailed t test, from three independent experiments, each performed in triplicate wells (n = 3 ± S.D.). **, p ≤ 0.01; ***, p ≤ 0.001. Solid (rEDA) and dashed (cFn) lines indicate threshold A415 nm ± 2 S.D. values (error bars) recorded for experimentally validated ECM-binding–negative rScl2.28FL control protein.
located on β-strands B, C, and F, as well as the AB and BC loops (63). This is, however, different from the Scl1-binding site on EDA, mapped to the C–C loop (46). Interestingly, PavA/B did not bind the RGD sequence on FnIII10, which is recognized by the integrin α5β1, or the synergistic sites on repeats III8 and III9 (63), whereas the Scl1 binding site on the C–C loop maps in the vicinity of the binding site for integrins α5β1 and α4β1 (39, 46).

PavA and PavB binding to the FnIII repeats was modeled to be largely due to electrostatic interactions; however, basic amino acids were primarily responsible for binding, rather than negatively charged residues, as modeled here for Scl1-V. Thus, Scl1 remains a unique adhesin that targets EDA/cFn isoforms that are expressed at the pathogen portal of entry.

Here, we identified the loop region of the Scl1-V domain as the structural determinant that is central to EDA/cFn recognition and binding. When the 22-amino acid loop sequence of the EDA binding–positive Scl1-V variant, derived from the global M1-type strain, was transplanted into the corresponding region of the binding-negative Scl2 protein, the chimeric construct gained EDA-binding activity. When the reverse transplantation was performed, ligand binding was lost both in recombinant protein and when expressed on the surface of Scl1 mutant GAS. A similar gain of binding function by the chimeric protein was achieved when the loop region from a phylogenetically distant Scl1 variant from M28-type GAS was substituted into the Scl2-V loop. A novel “catch-clamp” Fn-binding mechanism was deciphered for the multifunctional filamentous adhesin of Streptococcus gordonii, CshA (64). The CshA protein, which also lacks the classical Fn-binding repeats, binds both pFn and cFn via its nonrepetitive domains 1 and 2. The NR2 globular domain has a β-sandwich structure with a lectin-like fold and probably binds glycosylated sites on Fn.
NR2 domain is capped by a flexible ligand-gating loop, which, upon ligand recognition, uncovers a presumed binding cleft. The Scl1-V domain is different, because it forms a six-helix bundle structure, not lectin-like fold, and binds nonglycosylated recombinant EDA protein; in addition, the EDA repeat in human cFn is not glycosylated (31). The lectin-like domain, containing a ligand-gating loop capping the ligand-binding cleft, is represented in numerous fibrillar streptococcal adhesins, including Streptococcus mutans SpaA (65), S. gordonii SspB (66), Streptococcus agalactiae BspA (67), and M28-type Streptococcus pyogenes AspA (68), all correlated with the extended antigen I/II protein family. Whereas the detailed Scl1 structure and structural organization are different from those of antigen I/II affilates, there are several intangible similarities, where the elongated CL domain projects the globular ligand-binding domain away from the cell surface, with the ligand-engaging loop placed at the apex of the Scl1-V structure.

Homology modeling of the Scl1-V domain revealed that electrostatic surface potential and distribution of charged residues probably participate in ligand recognition and binding. A negatively charged region is formed in the center of the Scl-V trimer of ligand binding–positive models modeled (Scl1.1, rScl.chi1, and rScl.chiC) that are surrounded by neutral and positively charged residues. In contrast, the electrostatic models of the binding-negative Scl-V domains (Scl2.28, rScl.chi2, and rScl.chiN) display a uniform negative surface charge throughout the crown of the Scl-V domain. The arrangement of the surface potential observed in the Scl1-V domains was associated with a pattern of conserved negatively charged amino acids, primarily Glu 46-47(Scl1.1) and Asp 48(Scl1.1), located at the C-terminus of the loop sequence at the junction with the α-helices forming the inner core of Scl.1-V bundle. This charged distribution is reminiscent of the ligand-gating loops, capping negatively charged binding clefts, proposed in CshA and BspA (64, 67). This conserved pattern is observed in Scl1 variants that are evolutionarily distant on the phylogenetic tree; for example, Glu 48-50 of Scl1.28 from M28 GAS. Therefore, phylogenetically distant Scl1 proteins, which have significant amino acid sequence variation, exhibit conservation in the arrangement of charged amino acids and similar ligand-binding partners.

We previously reported that the Scl1-V domain bound EDA/cFn, Lm, and apoB100-LDL, thus supporting a dichotomous nature of Scl1-ligand binding in tissue and blood (23). Here, we established that the same regions of the globular Scl1 domain are critical in binding with all three ligands. It is not unusual that bacterial adhesins are multifunctional and bind several distinct ligands. Such proteins typically contain clearly defined functional domains; for example, the ECM-binding protein of S. epidermidis Embp, discussed earlier, contains both the Fn-binding FIVAR domains and the G-related albumin–binding GA domains (62). The collagen domain in streptococcal surface proteins is also found in association with additional domains; for instance, the pneumococcal protein PcaA (69) harbors both FIVAR and G5 (which binds GlicNAc) domains, in addition to the collagen domain (23). The Scl1-V domain is different in that the short sequence-variable loop region and a conserved arrangement of the surface potential at the apical plane of the Scl1-V structure are involved in binding of assorted ligands that do not share substantial sequence or structural similarities. These data, combined with previous molecular evolutionary analysis (9, 11), as well as structural and phylogenetic analyses (20, 22), indicate that retaining the conserved structure of the Scl1-V domain is the selective constraint that influenced sequence variation.

Our work demonstrates several critical structure–function constraints that determine ligand-binding capacity by the globular domain of a major group A streptococcal adhesin, Scl1. Current prospects for a global GAS vaccine, based on M protein, are problematic because more than 240 M-protein types have been reported, especially in regions outside North America and Europe, where vaccines are acutely needed. The EDA binding to diverse Scl1 variants could be developed into a nonconventional treatment option for GAS infections. The EDA-derived cyclic peptide bound Scl1 with an affinity in the micromolar range (49). The ongoing research explores the possibility of designing an improved EDA-derived peptide for targeting diverse GAS strains. This work provides an important structural understanding of Scl1 adhesin, which is essential during early stages of GAS infection and lays the foundation for the development of nonvaccine inhibitors of Scl1–ligand interactions for the prevention of infection progression.

**Experimental procedures**

**Production of rScls**

Gene cloning and rScl protein production were performed in E. coli TB1 and BL-21, respectively, grown in Luria–Bertani medium (Fisher) with ampicillin (100 µg ml⁻¹) at 37 °C. rScl-encoding clones, which were derived from the original scl alleles, were generated by PCR amplification from GAS genomic DNA; primers contained flanking sequences compatible with E. coli expression vector pASK-IBA2 digested with BsaI. Clones encoding the chimeric rScl proteins were generated using synthetic dsDNA fragments (gBlocks; Integrated DNA Technologies) flanked by PflMI and BsaI sites. All plasmids were verified by DNA sequencing. Plasmid constructs used in this study are described in Table 1, and gBlocks are shown in Table S1.

All rScl proteins were produced using the Strep-tag II cloning, expression, and purification system (IBA-GmbH, Gottingen, Germany). Proteins were expressed with a C-terminal affinity tag and purified on Strep-Tactin Sepharose, as described (14, 15). The full-length (rSclFL) and V-region-only (rSclIV) constructs were made, as designated. rScl1.1FL and rScl1.1V are derived from Scl1 protein in M1-type strain MGA6708 (14); rScl1.28FL and rScl2.28FL originate from M28 strains MGA6274 (14); and rScl1.3FL and rScl1.3V are derived from M3 strain MGAS315 (49). Both naturally derived and chimeric rScl proteins were expressed in E. coli BL21 periplasm following the induction with anhydrotetracycline at 0.2 µg ml⁻¹ for 3 h. Cells were centrifuged and suspended in high-sucrose buffer (100 mM Tris-HCl, 1 mM EDTA, pH 8.0, 500 mM sucrose) or Cell Lytic B Buffer (Sigma) for separation of the periplasmic fraction and subsequent affinity purification. Purified proteins were analyzed by SDS-PAGE and stained with RAPIDstain™(G-Biosciences); proteins were dialyzed against...
25 mM HEPES, pH 8.0, and stored at −20 °C. Protein concentrations were determined using Qubit fluorometric quantitation (ThermoFisher).

**rEDA production**

Recombinant EDA (Table 1) was produced using the pQE-30 His tag cloning, expression, and purification system (Qiagen) in the *E. coli* strain JM-109 (Promega), as described (70). *E. coli* containing rEDA-encoding plasmid was grown in Luria–Bertani broth supplemented with ampicillin (100 μg/ml−1). Protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside for 3 h. Cells were harvested by centrifugation, and pellets were frozen at −20 °C for 2 h or overnight. Cells were next suspended in a lysis buffer (50 mM Tris/HCl, pH 8.0, 50 mM NaCl, 2 mM MgCl2, 2% Triton X-100, 10 mM β-mercaptoethanol, 0.2 mg/ml lysozyme, 1 EDTA-free protease inhibitor mixture tablet (per 10 ml) (Pierce), 1 mM phenylmethylsulfonyl fluoride, 10 units/ml DNaseI) and incubated on ice for 20 min. Cell lysate was centrifuged at high speed (16,000 g for 20 min), and supernatant was collected. Supernatant was mixed with nickel-nitritriacetic acid–agarose resin (Qiagen) by inversion for 1.5 h and then poured into a column (Bio-Rad). Sample was washed with a 50 mM ammonium bicarbonate. rScl samples containing rEDA-encoding sequences were normalized by subtracting the BSA controls in each experimental set-up.

In regular binding assays, ECM ligands were added to the rScl-immobilized wells at 1 μg/well and incubated at room temperature for 1 h. To assess concentration-dependent binding, rEDA was added to the wells at concentrations ranging between 0.039 and 5 μM. Bound ECM ligands were detected with anti-cFn polyclonal antibody (Sigma/F3648; 1:4000), anti-His tag mAb (Proteintech/66005; 1:100), anti-Ln polyclonal antibody (Sigma/L939; 1:500), and anti-apoB100 (Chemicon/AB742; 1:500), followed by goat anti-rabbit (Bio-Rad/1721019; 1:2000), anti-mouse (Bio-Rad/1721011; 1:1000), or rabbit anti-goat (Bio-Rad/1721034; 1:1000) secondary antibodies conjugated to horseradish peroxidase (HRP) (Bio-Rad). The HRP substrate, 2,2’-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid) (Fisher) was used, and colorimetric reactions were recorded at A415 nm.

**Rotary shadowing and EM**

rScl proteins were rotary-shadowed and analyzed by EM, as described previously (14). Briefly, recombinant proteins were dialyzed against 25 mM HEPES buffer, pH 8.0, and protein concentration was measured with Qubit fluorometric quantitation; protein integrity and purity were assessed by 15% SDS-PAGE. Protein was stored at −20 °C until future use.

**CD spectroscopy**

CD spectroscopy of rScl proteins was performed as described previously (14). Briefly, protein samples were dialyzed against 0.1 mM ammonium bicarbonate. rScl samples were mixed with glycerol to a final glycerol concentration of 70% (v/v), and 100 μl of each sample was nebulized with an airflow onto freshly cleaved mica. Samples were then dried in a vacuum and rotary-shadowed with carbon/platinum using an electron beam gun tilted at an angle of 6°, relative to the mica surface, in a Balzers BAE 250 evaporator. The replicas were then dried with carbon at 90° and then were floated in distilled water and picked up onto bare 600-mesh copper grids. Photomicrographs were taken using an FEI G2 electron microscope operated at 120 kV with a 30-μm objective aperture.

**Protein-binding assays**

rScl proteins (0.5 μM solutions) were immobilized onto Strep-Tactin-coated microplate wells for 1.5 h at room temperature and blocked with TBS supplemented with 1% BSA overnight at 4 °C, followed by incubation with ECM ligands: cFn (Sigma), rEDA, murine Lm-111 (Sigma), and apoB100 and LDL (Sigma). The no-rScl controls were performed in BSA-coated wells for each ligand and each antibody used. Final ODs were normalized by subtracting the BSA controls in each experimental set-up.

In regular binding assays, ECM ligands were added to the rScl-immobilized wells at 1 μg/well and incubated at room temperature for 1 h. To assess concentration-dependent binding, rEDA was added to the wells at concentrations ranging between 0.039 and 5 μM. Bound ECM ligands were detected with anti-cFn polyclonal antibody (Sigma/F3648; 1:4000), anti-His tag mAb (Proteintech/66005; 1:100), anti-Ln polyclonal antibody (Sigma/L939; 1:500), and anti-apoB100 (Chemicon/AB742; 1:500), followed by goat anti-rabbit (Bio-Rad/1721019; 1:2000), anti-mouse (Bio-Rad/1721011; 1:1000), or rabbit anti-goat (Bio-Rad/1721034; 1:1000) secondary antibodies conjugated to horseradish peroxidase (HRP) (Bio-Rad). The HRP substrate, 2,2’-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid) (Fisher) was used, and colorimetric reactions were recorded at A415 nm.

**Protein modeling**

Multiple-sequence alignments of protein sequences were generated using MUSCLE and refined manually based on the results of secondary structure predictions. Predicted α-helices represent the consensus between PSIPRED, PHD, and SSPro. The phylogenetic tree of Scl1-V and Scl2-V regions was inferred based on multiple-sequence alignments, using the neighbor-joining method (adapted from Ref. 20). Branch lengths were scaled to the distance computed using the Gonnet250 matrix, and bootstrap values (statistical support for given bifurcation) were calculated for all nodes.

The homology models of Scl1.1-V and Scl2.28-V domains were obtained after consensus-based sequence alignment using HHMer, against the amino acid sequence of Scl2.3-V domain with solved crystal structure (22) (PDB code 4NSM). The homology models were built with MODELLER (71), and stereochemical quality of the models was improved by energy minimization with GROMACS (72).

**Complementation of M1 group A Streptococcus**

The *E. coli* shuttle plasmid pSB027 (73) (Table 1) was used for in trans complementation. The DNA fragment encompassing the *scl1.1* coding sequence with upstream promoter was PCR-amplified using genomic DNA from the M1-type strain MGAS5005 (WT) with primers 232Reg and 232Rev_BgIII (Table S1). The amplified DNA was cloned into pSB027 between BgIII and HindIII sites, resulting in pSL620 (Table 1). The *scl.chi2* sequence was cloned as a synthetic dsDNA fragment (gBlocks (Integrated DNA Technologies); SL621 gBlock, Table S1) between SspI and PpuMI sites of pSL620, resulting in pSL621 (Table 1). Both pSL620 and pSL621 were electroporated into MGAS5005 Δscl1 mutant GAS (9), and transformants were selected on BHI agar containing 10 μg/ml−1 chloramphenicol; plasmids were verified by DNA sequencing.
GAS cultures were grown in Todd Hewitt broth (BD Biosciences) supplemented with 0.2% yeast extract (THY medium) and on brain heart infusion agar (BD Biosciences) at 37 °C in an atmosphere with 5% CO₂. For antibiotic selection, chloramphenicol (5 μg ml⁻¹) was added to the medium.

**Western blot analysis**

Expression of Scl1.1 and Scl.chi2 proteins was determined by Western immunoblotting of the bacterial cell wall protein fraction, as described previously (9, 10). Briefly, bacterial cultures were grown to an A₆₀₀ of 0.5, and cells were harvested by centrifugation. GAS cells were next digested with 100 μg of lysozyme and 5 units of mutanolysin and phenylmethylsulfonyl fluoride in a 20% sucrose buffer. 5–25 μl of the sample supernatants were separated by SDS-PAGE and transferred to a nitrocellulose membrane. Detection of Scl1.1 and Scl.chi2 proteins was performed using rabbit antibody (1:2000) specific to the Scl1.1-V domain (9). Horseradish peroxidase–conjugated goat anti-rabbit IgG secondary antibody (Bio-Rad; 1:2000) was added to blots, followed by the addition of Clarity ECL substrate (Bio-Rad) for detection. Images were acquired using a ChemiDoc touch imaging system (Bio-Rad).

**Flow cytometry analysis**

Determination of Scl surface expression and rEDA binding by GAS cells was measured by flow cytometry. Bacteria grown to an A₆₀₀ of 0.5 were harvested by centrifugation and washed by flow cytometry buffer (PBS containing 10% Todd Hewitt broth with 0.2% yeast extract). For Scl surface detection, the anti-Scl1.1-V antibody (9) was preabsorbed with MGAS5005 Δscl cells, and then Ab (1:10 dilution) was incubated with GAS cells tested for 30 min on ice. Cells were next washed and incubated with Allophycocyanin-conjugated donkey anti-rabbit antibody (Jackson ImmunoResearch; 1:150). For rEDA binding, GAS cells tested were incubated with rEDA for 30 min at room temperature, and bound rEDA was detected with anti-His tag mAb (1:10 dilution) (Proteintech) preabsorbed with MGAS5005 WT cells. Samples were then incubated with goat anti-mouse polyclonal antibody conjugated with Alexa Fluor® 568 (ThermoFisher; 1:150). Cells were washed and fixed in 0.4% paraformaldehyde and stored at 4 °C until analysis. 50,000 events were collected per sample using a BD LSRFortessa (BD Biosciences), and data were analyzed with FCS Express Flow version 6 software.

**Statistical analyses**

Statistics were performed using the two-tailed paired Student’s t test or two-way analysis of variance. Significance was denoted at the following levels: *, p ≤ 0.05; **, p ≤ 0.01; ***, p ≤ 0.001. Error bars represent S.D. with analyses based on three independent experimental repeats (n = 3), each performed in triplicate technical replicates.

**Author contributions**—D. H. M., R. B., and S. L. conceptualization; D. H. M., S. J. C., and S. L. data curation; D. H. M., R. B., and S. L. formal analysis; D. H. M., S. J. C., D. R. K., L. V. D. W., F. S., and S. L. investigation; D. H. M. visualization; D. H. M., S. J. C., R. B., and S. L. methodology; D. H. M., F. S., and R. B. writing-original draft; L. V. D. W., R. B., and S. L. writing-review and editing; R. B. and S. L. supervision; S. L. funding acquisition; S. L. validation; S. L. project administration.

**Acknowledgments**—We thank Paolo Fagone for assistance with rEDA purification, Kathleen Brandage for flow cytometry assistance, and Janusz Bujnicki for initial bioinformatic analyses. We thank Nyles Charron for critical reading of the manuscript. We especially thank the Shriners Hospital for Children for support of the Micro-Imaging Center in the Portland Research Unit. Flow cytometry experiments were performed in the West Virginia University Flow Cytometry and Single Cell Core Facility, which is supported by National Institutes of Health (NIH) Equipment Grant S10OD016165 and NIGMS, NIH, Grant P50GM103488 (CoBRE).

**References**

1. Carapetis, J. R., Steer, A. C., Mulholland, E. K., and Weber, M. (2005) The global burden of group A streptococcal diseases. Lancet Infect. Dis. 5, 685–694 CrossRef Medline
2. Pföh, E., Wessels, M. R., Goldmann, D., and Lee, G. M. (2008) Burden and economic cost of group A streptococcal pharyngitis. Pediatrics 121, 229–234 CrossRef Medline
3. Elfratiou, A., and Lamagni, T. (2016) Epidemiology of Streptococcus pyogenes. In Streptococcus pyogenes: Basic Biology to Clinical Manifestations [Internet] (Ferretti, J. J., Stevens, D. L., and Fischetti, V. A., eds) University of Oklahoma Health Sciences Center, Oklahoma City, OK
4. Bisno, A. L., Rubin, F. A., Cleary, P. P., Dale, J. B., and National Institute of Allergy and Infectious Diseases (2005) Prospects for a group A streptococcal vaccine: rationale, feasibility, and obstacles—report of a National Institute of Allergy and Infectious Diseases workshop. Clin. Infect. Dis. 41, 1150–1156 CrossRef Medline
5. Cunningham, M. W. (2000) Pathogenesis of group A streptococcal infections. Clin. Microbiol. Rev. 13, 470–511 CrossRef Medline
6. Pavone, P., Parano, E., Rizzo, R., and Trifiletti, R. R. (2006) Autoimmune neuropsychiatric disorders associated with streptococcal infection: Sydenham chorea, PANDAS, and PANDAS variants. J. Child Neurol. 21, 727–736 CrossRef Medline
7. Yamaguchi, M., Terao, Y., and Kawabata, S. (2013) Pleiotropic virulence factor: Streptococcus pyogenes fibronectin-binding proteins. Cell. Microbiol. 15, 503–511 CrossRef Medline
8. Akiyama, H., Morizane, S., Yamasaki, O., Oono, T., and Iwashuki, K. (2003) Assessment of Streptococcus pyogenes microcolony formation in infected skin by confocal laser scanning microscopy. J. Dermatol. Sci. 21, 193–199 CrossRef Medline
9. Lukoms, S., Nakashima, K., Abdi, I., Cipriano, V. J., Ireland, R. M., Reid, S. D., Adams, G. G., and Musser, J. M. (2000) Identification and characterization of the scl gene encoding a group A Streptococcus extracellular protein virulence factor with similarity to human collagen. Infect. Immun. 68, 6542–6553 CrossRef Medline
10. Lukoms, S., Nakashima, K., Abdi, I., Cipriano, V. J., Shelvin, B. J., Gravis, E. A., and Musser, J. M. (2001) Identification and characterization of a second extracellular collagen-like protein made by group A Streptococcus: control of production at the level of translation. Infect. Immun. 69, 1729–1738 CrossRef Medline
11. Rasmussen, M., Edén, A., and Björck, L. (2000) SclA, a novel collagen-like surface protein of Streptococcus pyogenes. Infect. Immun. 68, 6370–6377 CrossRef Medline
12. Whatmore, A. M. (2001) Streptococcus pyogenes scbA encodes a putative hypervariable surface protein with a collagen-like repetitive structure. Microbiology 147, 419–429 CrossRef Medline
13. Rasmussen, M., and Björck, L. (2001) Unique regulation of ScbB: a novel collagen-like surface protein of Streptococcus pyogenes. Mol. Microbiol. 40, 1427–1438 CrossRef Medline
14. Xu, Y., Keene, D. R., Bujnicki, J. M., Höök, M., and Lukoms, S. (2002) Streptococcal Sc11 and Sc12 proteins form collagen-like triple helices. J. Biol. Chem. 277, 27312–27318 CrossRef Medline
28. Gao, Y., Liang, C., Zhao, R., Lukomska, E., Zhao, Z., Höök, M., and Lukomski, S. (2006) Assessment of prokaryotic collagen-like sequences derived from streptococcal Scl1 and Scl2 proteins as a source of recombinant GXY polymers. Appl. Microbiol. Biotechnol. 72, 109–115 CrossRef Medline

29. Pankov, R., and Yamada, K. M. (2002) Fibronectin at a glance. J. Biol. Chem. 278, 29757–29765 CrossRef Medline

30. ffrench-Constant, C. (1995) Alternative splicing of fibronectin—many different proteins but few different functions. Exp. Cell Res. 221, 261–271 CrossRef Medline

31. Hynes, R. O. (1990) Fibronectins, Springer-Verlag, New York, NY

32. To, W. S., and Midwood, K. S. (2011) Plasma and cellular fibronectin: distinct and independent functions during tissue repair. Fibrogenesis Tissue Repair 4, 21 CrossRef Medline

33. ffrench-Constant, C., and Hynes, R. O. (1989) Alternative splicing of fibronectin is temporally and spatially regulated in the chicken embryo. Development 106, 375–388 Medline

34. Coito, A. J., Brown, L. F., Peters, J. H., Kupiec-Weglinski, J. W., and van de Water, L. (1997) Expression of fibronectin splicing variants in organ transplantation: a differential pattern between rat cardiac allografts and isografts. Am. J. Pathol. 150, 1757–1772 Medline

35. ffrench-Constant, C., Van de Water, L., Dvorak, H. F., and Hynes, R. O. (1989) Reappearance of an embryonic pattern of fibronectin splicing during wound healing in the adult rat. J. Cell Biol. 109, 903–914 CrossRef Medline

36. Leahy, D. J., Aukhil, L., and Erickson, H. P. (1996) A crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. Cell 84, 155–164 CrossRef Medline

37. Niimi, T., Osawa, M., Yamaji, N., Yasunaga, K., Sakashita, H., Mase, T., Tanaka, A., and Fujita, S. (2001) NMR structure of human fibronectin EDA. J. Biochem. NMR 21, 281–294 CrossRef Medline

38. Liao, Y. F., Gotwals, P. J., Kotelsiansky, V. E., Sheppard, D., and Van De Water, L. (2002) The EIIIA segment of fibronectin is a ligand for integrins α9β1 and α6β1 providing a novel mechanism for regulating cell adhesion by alternative splicing. J. Biol. Chem. 277, 14467–14474 CrossRef Medline

39. Shinde, A. V., Bystroff, C., Wang, C., Vogeleyzang, M. G., Vincent, P. A., Hynes, R. O., and Van De Water, L. (2008) Identification of the peptide sequences within the EIIIA (EDA) segment of fibronectin that mediate integrin α9β1-dependent cellular activities. J. Biol. Chem. 283, 2858–2870 CrossRef Medline

40. Palmer, E. L., Ruegg, C., Ferrando, R., Pytela, R., and Sheppard, D. (1993) Sequence and tissue distribution of the integrin α9 subunit, a novel partner of β1 that is widely distributed in epithelia and muscle. J. Cell Biol. 123, 1289–1297 CrossRef Medline

41. Singh, P., Reimer, C. L., Peters, J. H., Stepp, M. A., Hynes, R. O., and Van De Water, L. (2004) The spatial and temporal expression patterns of integrin α9β1 and one of its ligands, the EIIIA segment of fibronectin, in cutaneous wound healing. J. Invest. Dermatol. 123, 1176–1181 CrossRef Medline

42. Singh, P., Chen, C., Pal-Ghosh, S., Stepp, M. A., Sheppard, D., and Van De Water, L. (2009) Loss of integrin α9β1 results in defects in proliferation, causing poor re-epithelialization during cutaneous wound healing. J. Invest. Dermatol. 129, 217–228 CrossRef Medline

43. Shinde, A. V., Kelsch, R., Peters, J. H., Sekiguchi, K., Van De Water, L., and McKeown-Longo, P. J. (2015) The α8β1 integrin and the EDA domain of fibronectin regulate a profibrotic phenotype in dermal fibroblasts. Matrix Biol. 41, 26–35 CrossRef Medline

44. Longmate, W. M., Lyons, S. P., Chittur, S. V., Pumiglia, K. M., Van De Water, L., and DiPersio, C. M. (2017) Suppression of integrin α3β1 by α9β1 in the epidermis controls the paracrine resolution of wound angiogenesis. J. Cell Biol. 216, 1473–1488 CrossRef Medline

45. Yu, Z., Mirochnitchenko, O., Xu, C., Yoshizumi, A., Brodsky, B., and Inouye, M. (2010) Noncollagenous region of the streptococcal collagen-like protein is a trimerization domain that supports refolding of adjacent homologous and heterologous collagenous domains. Protein Sci. 19, 775–785 CrossRef Medline

46. Oliver-Kozup, H., Martin, K. H., Schwegler-Berry, D., Green, B. J., Betts, C., Shinde, A. V., Van De Water, L., and Lukomska, S. (2013) The group A streptococcal collagen-like protein Scl1 mediates biofilm formation by targeting the extra domain A-containing variant of cellular fibronectin expressed in wounded tissue. Mol. Microbiol. 87, 672–689 CrossRef Medline

47. Brodsky-Doyle, B., Leonard, K. R., and Reid, K. B. (1976) Circular dichroism and electron-microscopy studies of human subcomponent C1q before and after limited proteolysis by pepstatin. Biochem. J. 159, 279–286 CrossRef Medline

48. Engel, J., and Furrthmayr, H. (1987) Electron microscopy and other physical methods for the characterization of extracellular matrix components: laminin, fibronectin, collagen IV, collagen VI, and proteoglycans. Methods Enzymol. 145, 3–78 CrossRef Medline
Structural adaptation of Scl1-V domain

49. Bachert, B. A., Choi, S. J., LaSala, P. R., Harper, T. J., McNitt, D. H., Boehm, D. T., Caswell, C. C., Ciborowski, P., Keene, D. R., Flores, A. R., Musser, J. M., Squegla, F., Marasco, D., Berisio, R., and Lukomski, S. (2016) Unique footprint in the scl1.3 locus affects adhesion and biofilm formation of the invasive M3-type group A Streptococcus. Front. Cell Infect. Microbiol. 6, 90

50. Chen, S. M., Tsai, Y. S., Wu, C. M., Liao, S. K., Wu, L. C., Chang, C. S., Liu, Y. H., and Tsai, P. J. (2010) Streptococcal collagen-like surface protein 1 promotes adhesion to the respiratory epithelial cell. BMC Microbiol. 10, 320

51. Roberts, A. L., Connolly, K. L., Kirse, D. J., Evans, A. K., Poehling, K. A., Peters, T. R., and Reid, S. D. (2012) Detection of group A Streptococcus in tonsils from pediatric patients reveals high rate of asymptomatic streptococcal carriage. BMC Pediatr. 12, 3 Medicline

52. Bourhis, J.-M., Mariano, N., Zhao, Y., Harlos, K., Exposito, J.-Y., Jones, E. Y., Mosli, C., Aghajani, N., and Hulmes, D. J. S. (2012) Structural basis of fibrillar collagen trimerization and related genetic disorders. Nat. Struct. Mol. Biol. 19, 1031–1036

53. Boydston, J. A., Chen, P., Steichen, C. T. Jr. and Turnbough, C. L. Jr. (2005) Orientation within the exosporium and structural stability of the collagen-like glycoprotein BcA of Bacillus anthracis. J. Bacteriol. 187, 5310–5317

54. Réty, S., Salamitou, S., Garcia-Verdugo, I., Aghajani, N., and Hulmes, D. J. S. (2016) Unique structural basis of fibrillar collagen trimerization and related genetic disorders. Nat. Struct. Mol. Biol. 19, 1031–1036

55. Xu, C., Yu, Z., Inouye, M., Brodsky, B., and Mirochnitchenko, O. (2010) Expanding the family of collagen proteins: recombinant bacterial collagens of varying composition form triple-helices of similar stability. Biomacromolecules 11, 948–956

56. Henderson, B., Nair, S., Pallas, J., and Williams, M. A. (2011) Fibronectin: a multidomain host adhesion targeted by bacterial fibronectin-binding proteins. FEMS Microbiol. Rev. 35, 147–200

57. Kingdley, R. A., Keestra, A. M., de Zoete, M. R., and Baumler, A. J. (2004) The Shda adhesin binds to the cationic cradle of the fibronectin 13FnIII repeat module: evidence for molecular mimicry of heparin binding. Mol. Microbiol. 52, 345–355

58. Mullen, L. M., Nair, S. P., Ward, J. M., Rycroft, A. N., Williams, R. J., Robertson, G., Mordan, N. J., and Henderson, B. (2008) Novel adhesin from Pasteurella multocida that binds to the integrin-binding fibronectin FnI9–10 repeats. Infect. Immun. 76, 1093–1104

59. Dabo, S. M., Confer, A. W., Anderson, B. E., and Gupta, S. (2006) Bartonella henselae Pap3L, an extracellular matrix adhesin, binds the fibronectin repeat module III13 module. Infect. Immun. 74, 2513–2521

60. Tsang, T. M., Annis, D. S., Kronshage, M., Fenno, J. T., Usselman, L. D., Mosher, D. F., and Krukonis, E. S. (2012) All protein binds ninth type III fibronectin repeat (9FnIII) within central 120-kDa region of fibronectin to facilitate cell binding by Yersinia pestis. J. Bacteriol. 287, 16759–16767

61. Christner, M., Franke, G. C., Schommer, N. N., Wendt, U., Wegert, K., Pehle, P., Kroll, G., Schulze, C., Buck, F., Mack, D., Aepfelbacher, M., and Rohde, H. (2010) The giant extracellular matrix-binding protein of Staphylococcus epidermidis mediates biofilm accumulation and attachment to fibronectin. Mol. Microbiol. 75, 187–207

62. Bachert, B. A., Choi, S. J., LaSala, P. R., Harper, T. J., McNitt, D. H., Boehm, D. T., Caswell, C. C., Ciborowski, P., Keene, D. R., Flores, A. R., Musser, J. M., Squegla, F., Marasco, D., Berisio, R., and Lukomski, S. (2016) Unique footprint in the scl1.3 locus affects adhesion and biofilm formation of the invasive M3-type group A Streptococcus. Front. Cell Infect. Microbiol. 6, 90

63. Kanwal, S., Jensch, I., Palm, G. J., Brönstrup, M., Rohde, M., Kohler, T. P., Somplatzki, D., Tege, W., Jenkinson, H. F., and Hammerschmidt, S. (2017) Mapping the recognition domains of pneumococcal fibronectin-binding proteins PavA and PavB demonstrates a common pattern of molecular interactions with fibronectin type III repeats. Mol. Microbiol. 105, 839–859

64. Back, C. R., Sztukowska, M. N., Till, M., Lamont, R. J., Jenkinson, H. F., Nobbs, A. H., and Race, P. R. (2017) The Streptococcus gordonii adhesin CshA protein binds host fibronectin via a catch-clamp mechanism. J. Biol. Chem. 292, 1538–1549

65. Troller-Charlier, N., Ogier, J., Moras, D., and Cavarelli, J. (2002) Crystal structure of the V-region of Streptococcus mutans antigen I/II at 2.4 Å resolution suggests a sugar preformed binding site. J. Mol. Biol. 318, 179–188

66. Forsgren, N., Lamont, R. J., and Persson, K. (2009) Crystal structure of the variable domain of the Streptococcus gordonii surface protein SspB. Protein Sci. 18, 1896–1905

67. Rego, S., Heal, T. J., Pinwill, G. R., Till, M., Robson, A., Lamont, R. J., Sessions, R. B., Jenkinson, H. F., Race, P. R., and Nobbs, A. H. (2016) Structural and functional analysis of cell wall-anchored polypeptide adhesin BspA in Streptococcus agalactiae. J. Biol. Chem. 291, 15985–16000

68. Green, N. M., Zhang, S., Porcella, S. F., Nagiec, M. J., Barbian, K. D., Beres, S. B., Lefebvre, R. B., and Musser, J. M. (2005) Genome sequence of a serotype M28 strain of group A Streptococcus: potential new insights into uuerperal sepsis and bacterial disease specificity. J. Infect. Dis. 192, 760–770

69. Paterson, G. K., Nieminen, L., Jefferies, J. M. C., and Mitchell, T. J. (2008) PcaA, a pneumococcal collagen-like protein with selected strain distribution, contributes to adherence and invasion of host cells. FEMS Microbiol. Lett. 285, 170–176

70. Webb, B., and Sali, A. (2017) Protein structure modeling with MODELLER. Methods Mol. Biol. 1654, 39–54

71. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., and Lindahl, E. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29, 845–854

72. Cramer, T., Yamanishi, Y., Clausen, B. E., Frister, I., Pawlinski, R., Mack, J. R., and Krukonis, E. S. (2012) Ail protein binds ninth type III repeat module: evidence for molecular mimicry of heparin binding. Mol. Microbiol. 52, 345–355

73. Webb, B., and Sali, A. (2017) Protein structure modeling with MODELLER. Methods Mol. Biol. 1654, 39–54

74. Prins, K., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., and Lindahl, E. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29, 845–854

75. Cramer, T., Yamanishi, Y., Clausen, B. E., Frister, I., Pawlinski, R., Mack, J. R., and Krukonis, E. S. (2012) Ail protein binds ninth type III repeat module: evidence for molecular mimicry of heparin binding. Mol. Microbiol. 52, 345–355

76. Flore, A. R., Jewell, B. E., Versalovic, E. M., Olsen, R. J., Bachert, B. A., Lukomski, S., and Musser, J. M. (2015) Natural variant of collagen-like protein A in serotype M3 Group A Streptococcus increases adherence and decreases invasive potential. Infect. Immun. 83, 1122–1129