WestVirginiaUniversity
THE RESEARCH REPOSITORY @ WVU

2016

Streptococcal collagen-like protein 1, Scl1, modulates group a Streptococcus adhesion, biofilm formation and virulence

Beth Alexandra Bachert

Follow this and additional works at: https://researchrepository.wvu.edu/etd

STREPTOCOCCAL COLLAGEN-LIKE PROTEIN 1, SCL1, MODULATES GROUP A *STREPTOCOCCUS* ADHESION, BIOFILM FORMATION AND VIRULENCE

Beth Alexandra Bachert

Dissertation submitted to the School of Medicine
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
In
Immunology and Microbial Pathogenesis

Slawomir Lukomski, Ph.D., Committee Chairperson
Thomas Elliott, Ph.D.
Christopher F. Cuff, Ph.D.
Karen H. Martin, Ph.D.
P. Rocco LaSala, M.D.

Department of Microbiology, Immunology, and Cell Biology

Morgantown, West Virginia
2016

Keywords: group A *Streptococcus*, collagen-like proteins, Scl1, Scl2, biofilm, host colonization, M3 type

ABSTRACT

Streptococcal collagen-like protein 1, Scl1, modulates group A *Streptococcus* adhesion, biofilm formation and virulence

Beth A. Bachert

Background: The collagens comprise a large family of versatile proteins found in all three domains of life. The streptococcal collagen-like protein 1, Scl1, of group A *Streptococcus* (GAS) binds extracellular matrix components (ECM), cellular fibronectin and laminin, via the surface-exposed globular domain. GAS strains express Scl1 and form biofilm *in vitro*, except for M3-type strains that are particularly invasive to humans. Hypothesis: Lack of Scl1 adhesin in M3 GAS results in decreased adherence and biofilm formation, and increased virulence. Results and Discussion: First crystal structure of the globular domain revealed a unique six-helical bundle fold, consisting of three pairs of alpha helices connected by variable loops. ECM binding by Scl1 promotes the formation of stable tissue microcolonies, which was demonstrated *in vitro* during infection of wounded human skin equivalents. A conserved nonsense mutation was identified in the *scl1* allele of the M3-type strains (*scl1.3*) that truncates the coding sequence, presumably resulting in a secreted Scl1 variant. Absence of Scl1 on the surface of M3-type GAS was demonstrated experimentally, as well as diminished expression of the *scl1* transcript in M3 strains relative to other M-types. Therefore, M3-type strains have reduced biofilm capacity on ECM coatings relative to other M-types. Constructed full-length recombinant Scl1.3 protein displayed binding capacity to cellular fibronectin and laminin, and M3 strains complemented with functional Scl1.3 adhesin displayed increased biofilm formation. The isoallelic M3 strain, carrying a rare "carrier" allele encoding cell-associated Scl1.3 variant, showed decreased pathology in mice, compared to the invasive M3 strain. Similarly, *scl1* inactivation in biofilm-capable M28- and M41-type GAS led to increased lesion size during subcutaneous infection. Conclusions: The studies presented here demonstrate the importance of surface Scl1 in modulating biofilm formation and virulence of GAS, and provide insight into the structure and function of Scl proteins.

Preface

The Ph.D. Thesis presented here represents a subset of research performed by Beth Bachert that is thematically focused on the importance of the major surface adhesin, the streptococcal collagen-like protein 1 (Scl1), in the pathogenesis of group A *Streptococcus*. The chapters herein correspond to published peer-reviewed research papers and are grouped into two parts that address: Part I) the role of Scl1 in adherence, biofilm formation, and virulence, and Part II) structural analysis of the Scl proteins. Additional published data are presented in Part III that resulted from Author's projects, focused on utilization of collagen-like genes as epidemiological and detection markers, which were carried out under a two-year Graduate Fellowship in Nanotechnology Sensing Advances in Field and Environment (NanoSAFE).

Research included in **Part I** demonstrates that Scl1 significantly contributes to biofilm formation in multiple M-types, except for invasive M3-type strains that were naturally diminished in the capacity to form biofilm (**Chapter 1**). Non-invasive M28- and M41-type strains produced significant biomass and biofilm structure associated with extracellular glycocalyx. Biofilm phenotype was significantly reduced in *scl1* isogenic mutants but was gained following heterologous complementation of *Lactococcus lactis* with Scl1. A unique *scl1.3* allele is reported that is conserved in M3 strains, which contains a null mutation in the collagen-like region, resulting in a truncated secreted protein. Research in **Chapter 2** assesses the effect of this null mutation on adherence and biofilm formation. The absence of Scl1 on the surface of M3-type GAS was demonstrated, as well as diminished *scl1*-transcript level in comparison to other M-types. Complementation of Scl1-deficient M3 GAS with full-length Scl1.3 surface protein conferred biofilm formation by these strains. The M3 wild-type strain lacking surface Scl1 did not form tissue microcolonies in an *in vitro* pseudo-organ skin equivalents, whereas Scl1-expressing M41 strain formed glycocalyx-embedded microcolonies. Inactivation of *scl1* in M28 and M41 GAS resulted in increased virulence in mice. Research in **Chapter 3** focuses on M3 strains harboring a rare carrier allele of *scl1.3*, which restored the open reading frame and the expression of full-length Scl1.3 protein. These strains were attenuated for virulence and had increased host cell adherence compared to the wild-type strain.

Part II focuses on crystallization and structural characterization of Scl proteins. The globular non-collagenous domain of Scl2 from M3-type GAS was crystallized (**Chapter 4**), and revealed a six-helical bundle fold that is rare in bacterial proteins (**Chapter 5**).

Overall, studies presented in **Parts I** and **II** demonstrate the importance of Scl1 in adherence and biofilm formation, while revealing an inverse correlation with invasiveness. Importantly, our studies provide new insights into the structure-function relationship of Scl proteins.

Additional data presented in **Part III**, describe interdisciplinary projects aimed at development of bacterial models for the microfluidic assessment of epidemiology of group A streptococcal strains (**Chapter 6**) and the detection of the select agents, *Burkholderia pseudomallei* and *B. mallei,* by targeting collagen-like genes (**Chapter 7**).

DEDICATION

This work is dedicated to my fiancé Eric Witt, who has been a constant source of love, support, and encouragement over the last 5 years. Your strength, kindness, thoughtfulness, and lighthearted sense of humor has had a tremendous impact on my life and been an inspiration to me. I also dedicate this work to my parents Yvonne and Danny Bachert, who have taught me the value of education, hard work, and perseverance. Without their unwavering faith and support, this work would not have been possible.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF NOMENCLATURE

1. Scl, streptococcal collagen-like protein
2. rScl, recombinant streptococcal collagen-like protein
3. Scl1.3, Scl1 from M3-type strain
4. Scl2.3, Scl2 from M3-type strain
5. GAS, group A *Streptococcus*
6. cFn, cellular fibronectin
7. ECM, extracellular matrix
8. EDA/cFn, EDA-containing cellular fibronectin
9. Lm, laminin
10. Bucl, *Burkholderia* collagen-like protein

GENERAL INTRODUCTION

COLLAGEN-LIKE PROTEINS OF PATHOGENIC STREPTOCOCCI: BIOLOGY, STRUCTURE AND FUNCTION

Summary

The collagen domain, which is defined by the presence of the Gly-X-Y triplet repeats, is amongst the most versatile and widespread known structures found in proteins from organisms representing all three domains of life. The streptococcal collagen-like (Scl) proteins are widely present in pathogenic streptococci, including *Streptococcus pyogenes*, *S. agalactiae*, *S. pneumoniae*, and *S. equi*. Experiments and bioinformatic analyses support that all Scl proteins are homotrimeric and cell wall-anchored. These proteins contain the rod-shaped collagenous domain proximal to cell surface, as well as a variety of outermost non-collagenous domains that generally lack predicted functions but can be grouped into one of six clusters based on sequence similarity. The well-characterized Scl1 proteins of *S. pyogenes* show a dichotomous switch in ligand binding modalities between human tissue and blood environments. Scl1 adhesin specifically recognizes the wound microenvironment, resists killing by neutrophil extracellular traps, promotes adhesion and biofilm formation, and modulates *S. pyogenes* virulence. In blood, binding modalities include components of the complement and coagulation-fibrinolytic systems, as well as plasma lipoproteins. In all, the Scl proteins signify a large family of structurally related surface proteins, which contribute to the ability of streptococci to colonize and cause diseases in humans and animals.

Overview of collagens and collagen-like proteins

A common structure among diverse proteins

Collagens are ubiquitous in nature. The common feature of the collagen module is a triple-helical structure consisting of three polyproline-II-like helices supercoiled in a right-handed direction around a central axis (1). The tight packing of the polypeptide chains requires a glycine every third residue, defining the Gly-X-Y repeat motif, where proline and hydroxyproline often occupy the X and Y positions of human collagen, respectively. Current knowledge on collagen structure is a culmination of fiber diffraction studies,

modeling, and crystallographic studies on collagen mimetic peptides. Since a history of the structural dissection of collagen is out of the scope of this review, we direct the reader to several excellent reviews (2-5). The collagens emerged as essential group of modular proteins of metazoans (6). It is a versatile structure, appearing in human extracellular matrix proteins, host defense proteins, and anchoring fibrils (5). The collagen domain is also present in important proteins in invertebrates, such as the exoskeleton collagens of sponges (7), holdfast structure of the mussel byssus (8), and basement membrane and cuticle collagens of the nematode *Caenorhabditis elegans* (9). There has been a large number of 18,874 collagen-like proteins (CLPs) annotated in bacteria, 695 in viruses, and 157 in archaea (search conducted on 9/24/16 in Uniprot database). The name streptococcal collagen-like proteins, Scl, in *S. pyogenes* was coined (10), which was followed by *Bacillus* proteins Bcl (11,12), pneumococcal protein Pcl (13), Lcl of *Legionella pneumophila* (14), and Bucl proteins of *Burkholderia* spp. (15). However, only a small proportion of predicted bacterial CLPs have been investigated thus far.

Origin of the collagenous domain in bacteria

The origin of the collagenous domain in prokaryotes is still unknown. However, the composition of the collagen domain in human and bacterial collagens differs considerably. In human collagens, proline and hydroxyproline residues are found preferentially in the X and Y positions, respectively, with frequencies of 27% and 38% (3). The abundance of hydroxyprolines in mammalian collagens is a major contributor to structural stability of the triple helix (16,17) but prokaryotes lack the prolyl hydroxylase enzyme to perform this post-translational modification. Therefore, in bacteria more than 30% of proline residues are found in the X position but only 5% in the Y position (18). Despite the lack of hydroxyproline residues, bacterial CLPs have been shown to form stable triple helices, with thermal stabilities similar to human collagens, in the range of 35°-39°C, (19-23). These CLPs rely on other mechanisms of helix stabilization, including hydration-mediated hydrogen bonding networks, electrostatic interactions between side chains, and the presence of specific stabilizing tripeptide repeats (19,24). Proper folding of the triple helix is necessary for collagen functionality.

Horizontal transfer of collagenous sequences from eukaryotes to prokaryotes has been proposed (25). However, recent studies have shown that collagen-like repeats arose independently via separate repeat amplification (26). Low-complexity repeats of the collagen triple helix could emerge by spontaneous mutations and amplify via simple repeat expansion, supporting the hypothesis that eukaryotic and prokaryotic collagens may have emerged by convergent evolution. Here, we present clustering analysis, which identified common domains in Scl proteins from different streptococcal species and subspecies, suggesting that horizontal *scl*-gene transfer is possible between bacteria that share the same human and animal hosts.

Classification of Scl proteins

CLuster ANalysis of Sequences (CLANS)

CLANS analysis was performed to obtain a sequence-based classification of Scl proteins. This analysis groups protein sequences based on all-against-all pairwise sequence similarities without phylogenetic reconstruction (27). CLANS identified six distinct clusters (1-6) of Scl proteins across pathogenic streptococci and two clusters (7-8) with phage proteins that will not be discussed here (Figure 1A). All Scl proteins share a distinct set of conserved features and a similar domain organization (Fig. 1B). They contain a signal peptide (not shown), an N-terminal non-collagenous sequence-variable (V) domain, a central collagen-like (CL) domain, and a cell-wall associated domain containing the LPXTG anchor motif (Gram-positive anchor). The CL domains of different Scl proteins are comprised of varying types of Gly-X-Y triplet repeats, and exhibit significant length variation due to the expansion and contraction of these repeats.

Figure 1. Sequence similarity and structure of Scl proteins in streptococci.
A. Clustering of Scl proteins using CLuster Analysis of Sequences (CLANS). CLANS clusters and visualizes groups of protein sequences based on all-against-all pairwise sequence similarities (*e*-value cutoff for analysis was 10^{-13}) without phylogenetic reconstruction. Each point on the plot represents a Scl amino acid sequence, while the connecting lines represent BLAST high-scoring segment pairs. High confidence clusters are indicated by circles, while a single low confidence cluster is indicated by a dashed circle. A list of Scl proteins identified in each cluster is provided in Table 1. clusters 7 and 8a/8b consist of phage-associated proteins and are not discussed in this review.
B. Cartoon representation of domain organization shared by mature Scl proteins. The Scl proteins consist of an N-terminal noncollagenous variable domain (V), a collagen-like domain (CL) containing repeats of the classical Gly-X-Y motif, and a C-terminally located cell wall-associated Gram-positive anchor with LPXTG motif.
C. Ribbon and surface models of the homotrimeric Scl2-protein globular domain. *Left.* Side view of the six-helix bundle, containing three pairs of antiparallel helices, colored green, orange and magenta, and the point of CL-region attachment. *Right.* Top view, showing the location of exposed hypervariable loop regions between each alpha helix pair and the arrangement of the three external helices wrapped around three internal helices.

Scl proteins of S. pyogenes (clusters 1,2)

The Scl1 and Scl2 proteins (also known as SclA and SclB) of *S. pyogenes* are the first bacterial CLPs to be reported and studied (10,28-31). Both proteins are homotrimeric and contain the globular V domain projected away from the cell surface by the rod-shaped CL domain. The V region sequences differ significantly both between and within Scl1 and Scl2 variants; for each Scl protein, variants are conserved in strains of the same M-type but vary between M-types. The *scl1* and *scl2* genes have been found in all *S. pyogenes* strains tested and are co-expressed in the exponential phase of growth, although their expression is regulated differently. Transcription of the *scl1* gene is positively regulated by the multiple gene regulator of *S. pyogenes*, Mga (28,30,32,33). Scl2 expression depends on phase variation associated with CAAAA repeats located downstream of the start codon (28,29,31).

Scl proteins in streptococci pathogenic to animals (clusters 3, 4 and 6)

The Scl3, 4 and 6 proteins are found in *S. zooepidemicus*, a commensal organism found in domesticated animals rarely transmitting to humans, and *S. equi,* a causative agent of the serious disease strangles in horses (34). CLANS classified several Scl3 proteins, originally denoted SclC for group C *Streptococcus* (35), as well as the related proteins SclD-I and SclZ.1-5, 7 and 12 (36), as belonging to the same cluster 3, whereas SclF (37) formed an independent cluster 4. SclC was shown to be expressed during strangles infection, and immunization with the recombinant SclC protein partially protects against infection in both mouse and horse models (38-40). An effective multi-component vaccine for strangles was developed that include recombinant

Table 1. Scl classification by CLANS analysis

Cluster No.	Proteins identified	Organisms*
1	Scl1 (SclA)	*S. pyogenes* (GAS)
2	Scl2 (SclB)	*S. pyogenes* (GAS)
3	SclC, D, E, G, H, I SclZ.1-5, 7, 12	*S. equi, S. zooepidemicus* (GCS)
4	SclF	*S. equi, S. zooepidemicus* (GCS)
5	PclA	*S. pneumoniae, S. agalactiae* (GBS)
6	SclZ.6, 9, 10 FneC, E, F	*S. equi, S. zooepidemicus* (GCS)
7	Phage minor structural protein	*S. pneumoniae, S. agalactiae* (GBS)
8a/b	Phage-associated hyaluronidase	*S. pyogenes* (GAS), *S. equi* and *S. dysgalactiae* (GCS)

*Abbreviations in parentheses refer to group A, B, and C *Streptococcus*

SclC (40). Several cluster 3 proteins share signature sequences with cluster 2 proteins (35), suggesting some Scl3 variants are related to Scl2 and could emerge via inter-species horizontal gene transfer. A second group of proteins found in both *S. equi* and *S. zooepidemicus*, classified in cluster 6, included the SclZ.6, 9, and 10 proteins (36), as well as FneC, E, and F that were annotated as fibronectin-binding proteins (41).

Scl proteins in S. pneumoniae and S. agalactiae (cluster 5)

The pneumococcal CLP, PclA (13), is a large surface protein (265 kDa) that contains predicted G5 and FIVAR domains, though FIVAR was predicted with low confidence. The G5 domain is associated with binding to N-acetylglucosamine and biofilm formation for a variety of proteins found in streptococcal and staphylococcal species (42), as well as the mycobacterial protein, RpfB, which also contains a triple helix motif (43,44). The FIVAR domain is found in surface-associated proteins SasC of *Staphylococcus aureus* and Embp of *S. epidermidis* that promote fibronectin binding and biofilm formation (45,46). The *S. agalactiae* CLPs also contain the G5 domain, but lack the FIVAR domain. The presence of *pclA* gene was correlated with pneumococcal adherence (13) and also with resistance to penicillin and trimethoprim-sulfamethoxazole (47).

The non-collagenous domain in streptococcal CLPs

The best structurally characterized streptococcal collagens are the two CLPs, Scl1 and Scl2, of *S. pyogenes*. The proteins form stable triple-helical structures when expressed as recombinant (rScl) polypeptides (20,24). The non-collagenous V region constitutes trimerization domain that augments proper collagen assembly to avoid the misfolding of the triple helix due to its repeating structure. However, the V domain of Scls is not necessary for triple helix formation *in vivo*, since the CL region of Scl1 can be expressed without the V region as recombinant protein in a folded triple helical state; still, this rScl-CL construct could not re-fold after thermal denaturation *in vitro* (20,48). These results suggest that V domains of Scls present both structural and ligand-binding functions reviewed below.

Recently, the crystal structure of the V domain of Scl2 from M3-type *S. pyogenes* was reported (49,50). The V domain folds into a six-helical bundle, with three pairs of

antiparallel alpha helices each connected by a variable loop region. Three of the helices are wound in a left-handed super-helix forming the inner core, which is further wrapped by three external alpha helices antiparallel to the internal helices (Fig. 1C). The six-helix bundle forms an elongated cylinder measuring about 30 Å in diameter and 60 Å in height. This fold is consistent with previous secondary structure predictions that deduced a helix-loop-helix motif in the primary amino acid sequence (30,51) and is predicted to be conserved among Scl1 and Scl2 proteins. Indeed, hydrophobic residues located at regular positions of Scl sequences were shown to play a central role in the stabilization of the inner core of the 6-helix bundle fold. In addition, the variable loops adopt a well-defined polyproline II conformation (50).

Scl proteins in pathogenesis: dichotomy of Scl1-ligand binding in human tissue and blood

The Scl1 variants of *S. pyogenes* bind a wide range of host ligands. The nature of Scl1-binding modalities between tissue and blood environments is highly significant, as described below (Fig. 2). In contrast, the Scl2 proteins failed to bind the majority of ligands and their role in pathogenesis is less understood.

Binding modalities in tissue to extracellular matrix and cell receptors

Scl1 selectively binds cellular fibronectin (cFn), but not plasma fibronectin (pFn), and laminin (Lm) (Caswell *et al.*, 2010). Cellular fibronectin is deposited by a variety of cells as an insoluble crosslinked protein in tissues and humans express over 20 cFn isoforms (52,53). Both pFn and cFn are encoded by a single fibronectin gene and contain a conserved structure, consisting of three regions of repeats, type I, II, and III. cFn differs from pFn via the inclusion by alternative splicing of extra domains A (EDA) and B (EDB), as well as varying numbers of the variable V domain (53). Scl1 proteins specifically bind cFn via recognition of the type III repeat, EDA (54). EDA/cFn isoforms are found in low levels in normal adult tissue but are upregulated in wounded tissue (55,56), where the EDA domain interacts with keratinocyte integrin receptors (57) and is important in the wound healing processes (58). *S. pyogenes* strains may express multiple fibronectin-binding proteins, including SfbI/PrtF1, PrtF2, SOF, FbaB, SfbX, and Shr, reviewed in (59)

that bind the type I and type II repeats; thus, Scl1 binds cFn via unique mechanism different from other Fn-binding proteins of *S. pyogenes*

Laminin binding was observed for the same Scl1 variants that bind EDA/cFn isoforms (60). The molecular basis for recognizing both ligands is undetermined due to a lack of sequence similarity between EDA and Lm chains α, β, or γ. At least 15 laminin heterotrimers have been identified in human tissues and binding studies with individual laminins represent a technical challenge. Two Lm-binding proteins, Lbp and Shr, in *S. pyogenes* have been reported in addition to Scl1 (61-63). Given the localization of some laminins to basement membranes, Lm binding by *S. pyogenes* likely represents a relevant pathogenesis trait.

The Scl1.41 variant, expressed by M41-type strains, directly binds human collagen receptors $\alpha_2\beta_1$ and $\alpha_{11}\beta_1$ integrins through the GLPGER motif in the Scl1-collagenous domain (64,65). A similar binding motif GF/LOGER (O represents hydroxyproline), as well as derived sequence motifs GR/AOGER and GASGER, were identified in human collagens as integrin-binding sites (66). Other Scl-integrin binding motifs are found in Scl2-CL regions, such as RGD and KGD sequences. In addition, the GAPGER and GKPGER motifs are found in SclZ/Scl3 proteins (36).

Significance of ligand-binding modalities in tissue
S. pyogenes forms biofilms, or microcolonies, in infected tissue (67,68). *In vitro* assays showed enhanced biofilms formed by strains of multiple M-types on ECM coatings, including Fn and collagens type I and IV (69,70). Scl1 plays an important role in biofilm formation, as isogenic *scl1*-inactivated mutants had significantly reduced overall biofilm biomass and decreased biofilm thickness (71). In addition, enhanced Scl1-mediated biofilms were observed on simple cFn and Lm coatings, as well as on complex extracellular matrix deposited by human dermal fibroblasts (54). Furthermore, inhibition experiments employing EDA-derived peptide and anti-EDA mAb showed significantly reduced adherence of *S. pyogenes* cells to fibroblast-derived matrix, indicating the Scl1-EDA/cFn interaction supports bacterial adherence

The M3-type *S. pyogenes* strains, which are particularly invasive to humans, lack Scl1 protein due to a null mutation within the CL-region coding sequence (10), and form

insignificant biofilms (71). Restoration of full-length Scl1 on the surface of M3-type *S. pyogenes* restored biofilm formation on cFn and Lm coatings (72). Also, biofilm-capable Scl1-expressing M41-type strain produced glycocalyx-embedded tissue microcolonies in an *in vitro* skin equivalent infection model, while biofilm-poor Scl1-lacking M3 strain did not. Rare M3 strains containing the "*scl1.3* carrier allele", with restored open reading frame, were less invasive in a mouse model of necrotizing fasciitis than typical M3 strains lacking Scl1 (73). Scl1-deficient mutants of M28- and M41-type *S. pyogenes* as well as M3-type *S. pyogenes* naturally lacking Scl1, displayed increased invasiveness in mice (72). Altogether, these results support the concept that *S. pyogenes* adherence to ECM and stable biofilm formation that are conferred by surface-attached Scl1, promote localized infection over invasive spread.

Scl1 binding to the $\alpha_2\beta_1$ integrin promoted fibroblast adhesion and spreading, and induced intracellular signaling typical of integrin pathway, thus, mimicking the functional role of human collagen. Direct binding of Scl1 to the $\alpha_2\beta_1$ integrins promotes *S. pyogenes* internalization by epithelial cells, resulting in increased intracellular pools as well as increased re-emergence of bacteria (74). The $\alpha_2\beta_1$ and $\alpha_{11}\beta_1$ integrins are expressed by fibroblasts, endothelial, and epithelial cells (75,76), indicating Scl1 has the potential to adhere to and invade a variety of cell types in the human host. Additionally, the RGD and KGD sequences that are cryptic within the collagen triple helix, become available for binding during tissue remodeling (77).

A recent study reported a novel role for Scl1 of M1 *S. pyogenes* in resistance to neutrophil extracellular traps (NETs). The Scl1-deficient mutant produced smaller skin lesions in mice *in vivo*, compared to the parental strain, which was associated with increased killing by NETs *in vitro*. Neutrophils incubated with Scl1-deficient mutant *in vitro* displayed increased NET formation than with wild-type strain, indicating Scl1 inhibited NET production by neutrophils. Moreover, Scl1-deficient *S. pyogenes* had increased sensitivity to cathelicidin killing *in vitro* (78). Additional studies would confirm if Scl1 variants from divergent M-types also harbor this function.
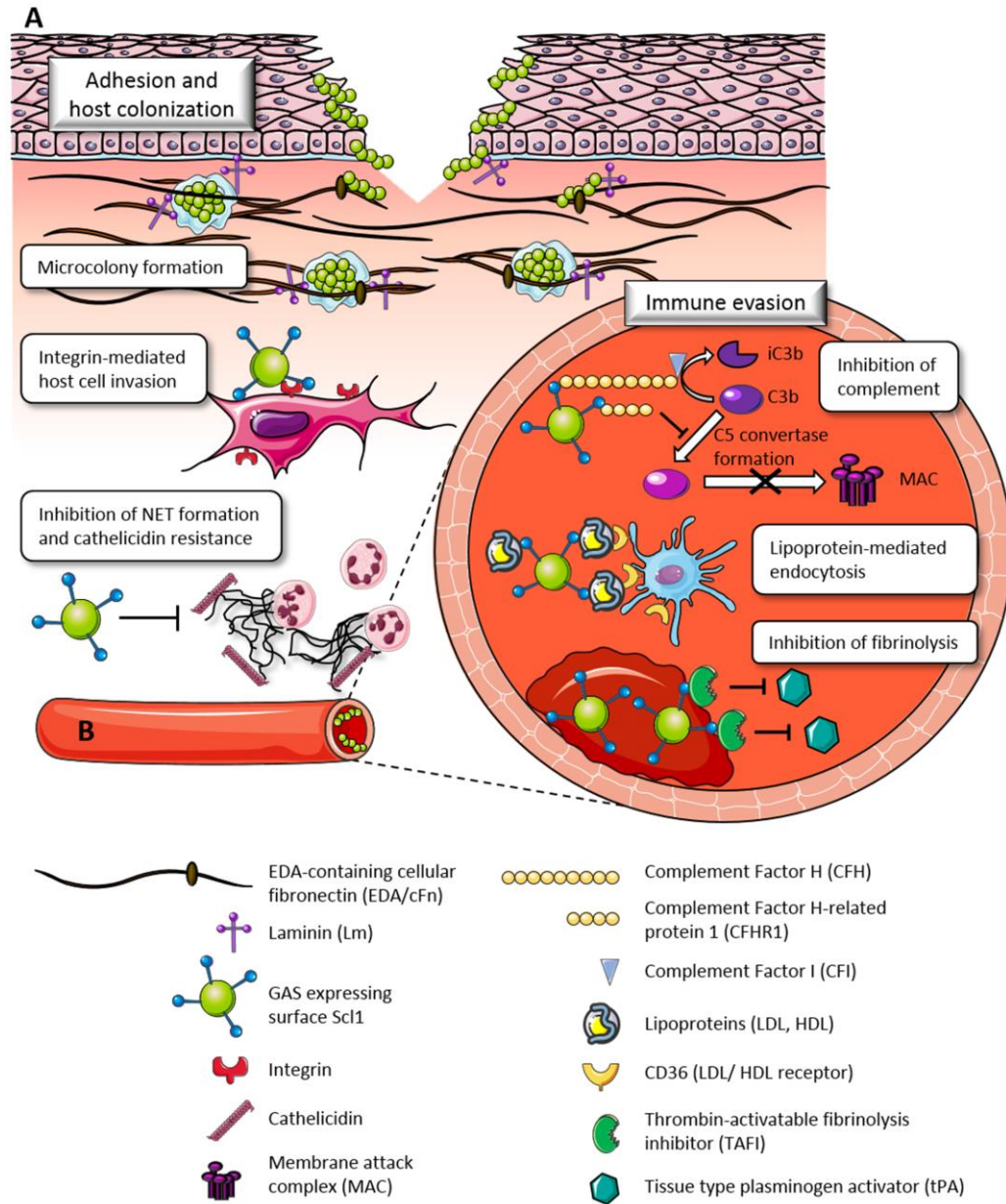
Figure 2. Dichotomous nature of ligand binding by Scl1 in human tissue and blood.
A. Binding modalities in tissue. *S. pyogenes* accesses human tissue via portal of entry (skin, pharyngeal mucosa) represented here as a breach of skin epidermis extending into the underlying dermis. The globular domain of Scl1 surface adhesin selectively recognizes the EDA-containing isoforms of cellular fibronectin (EDA/cFn) and laminin (Lm). Scl1-mediated tissue microcolonies, embedded in glycocalyx, are formed and stabilized by interactions between Scl1 and the surrounding wound microenvironment. The collagenous domain of Scl1 directly binds human collagen receptors, integrins $\alpha_2\beta_1$ and $\alpha_{11}\beta_1$ on the host cell surface, promoting pathogen internalization, survival and reemergence. Scl1 decreases the formation of extracellular traps (NETs) produced by infiltrating neutrophils and killing by NET-associated cathelicidin. **B.** Binding modalities in blood. During dissemination to the blood and deeper tissue, Scl1 contributes immune evasion and survival of *S. pyogenes* via several mechanisms. Binding to complement factor H (CFH) mediates cleavage of C3b by factor I, thus, preventing *S. pyogenes* opsonization, while binding to factor H-related protein 1 (CFHR1) prevents the formation of the C5 convertase and assembly of the membrane attack complex (MAC). Scl1-mediated adsorption of plasma lipoproteins LDL/HDL on the *S. pyogenes* surface may prevent immune recognition and/or promote LDL/HDL receptor-mediated endocytosis, intracellular survival and reemergence. Scl1 binding to and activation of the thrombin-activatable fibrinolysis inhibitor (TAFI) prevents clot breakdown, providing a protective niche for streptococci.

Binding modalities in blood

The Scl1 proteins from M6- and M55-type *S. pyogenes* bind the complement factor H (CFH) and complement factor H-related protein 1 (CFHR1) via Scl1's globular V domain (79). The CFH protein is composed of 20 short consensus repeats (SCRs) that facilitate ligand binding and co-factor function for complement factor I (CFI)-mediated C3b degradation; the C-terminal SCRs 18-20 are involved in binding to cell surfaces, while SCRs 1-4 harbors the co-factor function (80). The CFHR1 protein comprises five SCRs, of which the C-terminal SCRs 3-5 share 99% sequence identity, and therefore surface-binding function, with the SCRs 18-20 of CFH. CFHR1 inhibits the C5 convertase and the formation of the membrane attack complex (81). Using recombinant fragments of the CFH proteins, the Scl1 binding site was mapped in the C-terminal SCRs 19-20 of CFH and SCR 4 of CFHR1 (82). Notably, other known CFH-binding proteins of *S. pyogenes*, such as certain M and M-like proteins and Fba, bind to SCR7 of CFH, as well as complement factor H-like (CFHL) protein (82). *Borrelia burgdorferi* also expresses surface proteins with similar CFH-binding characteristics via the C-terminus (83,84).

Recombinant Scl1 proteins derived from diverse M-types bind the apolipoprotein ApoB100 of low density lipoprotein (LDL) (51). Binding of rScl1 constructs to LDL was facilitated by the Scl1-globular V domain with binding affinities of K_D values in the nanomolar range. Importantly, LDL from human plasma was absorbed by the wild-type cells of *S. pyogenes* but not by the *scl1*-inactivated mutant cells (51). A similar binding with ApoAI apolipoprotein and high density lipoprotein (HDL) was demonstrated for rScl1 construct derived from M41-type *S. pyogenes* (85). HDL binding was inhibited by low concentration of the nonionic detergent Tween20, suggesting hydrophobic interaction to both rScl1 protein and *S. pyogenes* cells.

Some recombinant Scl proteins also bind thrombin-activatable fibrinolysis inhibitor (TAFI) with K_D values in the nanomolar range (86) and the binding site was mapped to residues 205-232 within the TAFI protein (87). TAFI is a zinc-dependent procarboxypeptidase, which acts as an important fibrinolysis regulator and inflammatory mediator upon activation by thrombin, thrombin-thrombomodulin complex, or plasmin.

Significance of ligand-binding modalities in blood

Diverse pathogens express surface proteins that bind complement regulatory proteins, including *Neisseria gonorrhoeae*, *Borrelia burgdorferi*, *Yersinia enterocolitica*, and several streptococcal species (88). CFH binding to the cell surface prevents C3b deposition and phagocytosis, as well as downstream complement-mediated cell lysis. *In vitro*, rScl1-bound CFH retained its co-factor function, mediating the proteolytic breakdown of C3b by CFI (79). Similarly, CFHR1 binding to rScl1 inhibited the formation of the membrane attack complex *in vitro*. Both CFH and CFHR1 bound to M6-*S. pyogenes* cells. However, CFH binding by M5 *S. pyogenes* did not contribute significantly to phagocytosis resistance or virulence (89). Since M6 protein, was the first bacterial CFH-binding protein reported (90), CFH binding by Scl1, in addition to M6 protein, might be necessary for anti-phagocytic phenotype observed for M6 *S. pyogenes*.

Plasma lipoproteins are being increasingly recognized as innate immune components. For example, HDL and LDL neutralize LPS endotoxin (91) and *Staphylococcus aureus* α-toxin (92), and HDL is known to downregulate host adhesion molecules and inflammatory cytokines (93). Elevated lipoprotein levels may be protective against bacterial infections and sepsis in humans (94), and LDL-deficient mice showed increased susceptibility to infections with Gram-negative bacteria and *Candida albicans* (95,96). Interaction between *Yersinia pestis* and ApoB-containing lipoproteins, mediated by the pH6 antigen, prevented binding of the bacteria to macrophages *in vitro* (97). In contrast, LDL and HDL could act as opsonins to increase *S. pyogenes* phagocytosis and killing via a CD36-mediated endocytosis (98,99). Notably, half of *S. pyogenes* strains express serum opacity factor, which binds and disrupts the HDL structure (100), which could protect *S. pyogenes* from lipoprotein-mediated opsonization. Clearly, the Scl1-lipoprotein interactions may have multiple functions during infection and *in vivo* studies are required to determine the effects of these interactions on the host.

Scl-recruited TAFI to the *S. pyogenes* cell surface was cleaved and activated by plasmin and thrombin-thrombomodulin (86). TAFI functions by removing exposed C-terminal lysine residues from fibrin during blood clot formation, thereby preventing recognition and cleavage of these residues by the tissue-type plasminogen activator, ultimately inhibiting fibrinolysis, or the breakdown of clots (101). Additionally, it can

regulate inflammation by cleaving the C-terminal residues of bradykinin, osteopontin, and the chemoattractants C3a, C5a, and chemerin (102). Therefore, binding and prolonged activation of plasmin on the surface of *S. pyogenes* could contribute significantly to TAFI activation. Additionally, activation of TAFI on the *S. pyogenes* cell surface induced inflammation via modulation of the kallikrein/kinin system (103). Since this interaction maintains the formation of clots, it may also represent a mechanism for bacteria to remain associated with the fibrin clot and evade recognition by immune defenses.

Final Remarks

    i)   Modular collagens evolved in higher eukaryotes as members of extracellular matrix to support tissue structure and provide an essential network for cell function. Bacterial collagens likely emerged by convergent evolution via simple Gly-X-Y-repeat amplification; sequence similarity amongst Scl proteins suggests horizontal gene transfer between closely related species.

    ii)   Scl proteins are homotrimeric and surface attached, and have several modalities of sensing host surroundings. The amino-terminal globular "sensing" domain is variable in primary sequence but displays a conserved structure.

    iii)   Scl1 proteins have adapted binding modalities in tissue and blood to mediate *S. pyogenes* colonization and immune evasion.

    iv)   Scl1 down-modulates virulence by fostering the formation of stable tissue microcolonies, thus, promoting a focused nidus of infection.

    v)   The majority of Scl proteins remain uncharacterized and may exert similar functions to those described for Scl proteins of *S. pyogenes* that are important in pathogenesis.

PART I

ROLE OF SCL1 IN ADHERENCE, BIOFILM, AND VIRULENCE

CHAPTER 1

THE STREPTOCOCCAL COLLAGEN-LIKE PROTEIN-1 (SCL1) IS A SIGNIFICANT DETERMINANT FOR BIOFILM FORMATION BY GROUP A *STREPTOCOCCUS*

Heaven A Oliver-Kozup, Meenal Elliott, Beth A Bachert, Karen H Martin, Sean D Reid, Diane E Schwegler-Berry, Brett J Green, and Slawomir Lukomski

ABSTRACT

Background: Group A *Streptococcus* (GAS) is a human-specific pathogen responsible for a number of diseases characterized by a wide range of clinical manifestations. During host colonization GAS-cell aggregates or microcolonies are observed in tissues. GAS biofilm, which is an *in vitro* equivalent of tissue microcolony, has only recently been studied and little is known about the specific surface determinants that aid biofilm formation. In this study, we demonstrate that surface-associated streptococcal collagen-like protein-1 (Scl1) plays an important role in GAS biofilm formation.

Results: Biofilm formation by M1-, M3-, M28-, and M41-type GAS strains, representing an intraspecies breadth, were analyzed spectrophotometrically following crystal violet staining, and characterized using confocal and field emission scanning electron microscopy. The M41-type strain formed the most robust biofilm under static conditions, followed by M28- and M1-type strains, while the M3-type strains analyzed here did not form biofilm under the same experimental conditions. Differences in architecture and cell-surface morphology were observed in biofilms formed by the M1- and M41-wild-type strains, accompanied by varying amounts of deposited extracellular matrix and differences in cell-to-cell junctions within each biofilm. Importantly, all Scl1-negative mutants examined showed significantly decreased ability to form biofilm *in vitro*. Furthermore, the Scl1 protein expressed on the surface of a heterologous host, *Lactococcus lactis*, was sufficient to induce biofilm formation by this organism.

Conclusions: Overall, this work (i) identifies variations in biofilm formation capacity among pathogenically different GAS strains, (ii) identifies GAS surface properties that may aid in biofilm stability and, (iii) establishes that the Scl1 surface protein is an important determinant of GAS biofilm, which is sufficient to enable biofilm formation in the

heterologous host *Lactococcus.* In summary, the GAS surface adhesin Scl1 may have an important role in biofilm-associated pathogenicity.

INTRODUCTION

Microbial biofilm formation is an important virulence mechanism, which allows immune evasion and survival against antibiotic treatments (1,2). Many bacterial nosocomial infections are associated with biofilms formed on contaminated medical devices. Dispersal of biofilm has also been proposed to augment infection spread (3-8). For group A *Streptococcus* (GAS), biofilm research is an emerging field and little is known about the specific surface determinants that aid in biofilm formation. GAS is characteristically associated with significant human morbidity and it is responsible for the clinically common superficial throat and skin infections, such as pharyngitis and impetigo, as well as invasive soft tissue and blood infections like necrotizing fasciitis and toxic shock syndrome (9). Although GAS biofilm has not been associated with implanted medical devices, tissue microcolonies of GAS encased in an extracellular matrix were demonstrated in human clinical specimens (10). Studies reported to date support the involvement of GAS surface components in biofilm formation, including the M and M-like proteins, hyaluronic acid capsule, pili and lipoteichoic acid (11-13). As shown by Cho and Caparon (11), multiple genes are upregulated during biofilm formation and development, including the streptococcal collagen-like protein-1 (Scl1).

The *scl1* gene encoding the Scl1 protein has been found in every GAS strain investigated and its transcription is positively regulated by Mga (14-18), indicating that Scl1 is co-expressed with a number of proven virulence factors. Structurally, the extracellular portion of Scl1 protein extends from the GAS surface as a homotrimeric molecule composed of distinct domains that include the most outward N-terminal variable (V) region and the adjacent collagen-like (CL) region composed of repeating GlyXaaYaa (GXY) sequence. The linker (L) region is close to the cell surface and contains a series of conserved direct repeats. The Scl1 protein can bind selected human extracellular matrix components (19) and cellular integrin receptors (20-22), as well as plasma components (23-27).

16

In this study, we investigated the importance of Scl1 in GAS biofilm using defined isogenic wild-type and *scl1*-inactivated mutant strains of GAS. We report that (i) the pathogenically diverse M41-, M28-, M3- and M1-type GAS wild-type strains have varying capacities to produce biofilm on an abiotic surface; (ii) Scl1 plays an important role during the main stages of biofilm formation with Scl1-negative mutants having an abrogated capacity for adhesion, microcolony formation and biofilm maturation; and (iii) variations in surface morphology as well as in extracellular matrix associated with bacterial cells suggest two distinct but plausible mechanisms that potentially stabilize bacterial microcolonies. We additionally show that expression of Scl1 in *Lactococcus lactis* is sufficient to support a biofilm phenotype. Overall, this work reveals a significant role for the Scl1 protein as a cell-surface component during GAS biofilm formation among pathogenically varying strains.

MATERIALS AND METHODS

GAS strains and growth conditions

The wild-type GAS strains M41- MGAS6183, M1- MGAS5005, and M28-type MGAS6143, as well as their *scl1*-inactivated isogenic mutants and complemented M41Δ*scl1* mutant have been previously described (22,27,28). In addition, a set of the wild-type M3-type GAS strains MGAS158, MGAS274, MGAS315, MGAS335, MGAS1313, and MGAS2079 was also used. GAS cultures were routinely grown on brain–heart infusion agar (BD Biosciences) and in Todd–Hewitt broth (BD Biosciences) supplemented with 0.2% yeast extract (THY medium) at 37°C in an atmosphere of 5% CO_2–20% O_2. Logarithmic phase cultures harvested at the optical density (600 nm) of about 0.5 (OD_{600} ~0.5) were used to prepare GAS inocula for biofilm analysis. Colony counts were verified by plating on tryptic soy agar with 5% sheep's blood (Remel). *Lactococcus lactis* subsp. *cremoris* strain MG1363 (provided by Dr. Anton Steen) were grown using M17 broth or agar media (Oxoid) supplemented with 0.5 M sucrose and 0.5% glucose (SGM17 media) at 30°C in an atmosphere of 5% CO_2–20% O_2.

Heterologous Scl1 expression in Lactococcus lactis

Lactococcus **transformation:** To obtain electrocompetent cells, 500 ml of SGM17 broth supplemented with 2% glycine was inoculated with an overnight culture and grown until OD_{600} ~0.4 was reached. Cells were harvested and washed twice with ice-cold solution A (0.5 M sucrose, 10% glycerol); cells were then re-suspended in solution A (1/1000 of original culture volume) and stored at -80ºC (29). For transformation, cells were thawed on ice and mixed with 1 μl of DNA of the Scl1.41-expressing plasmid pSL230 or pJRS525-vector(22); and transferred to a cold 1-mm electrode-gap cuvette. Cells were pulsed with 2.0 kV at 25 μF and 400 ohms. Immediately following, suspensions were mixed with 1 ml outgrowth medium (SGM17 broth supplemented with 20 mM $MgCl_2$ and 2 mM $CaCl_2$) and incubated for 2.5 h before plating on SGM17 agar supplemented with spectinomycin (30).

Molecular characterization of transformants: The pSL230 was detected in *Lactococcus lactis* MG1363 transformants by PCR amplification directly from bacterial colonies with *scl1.41*-gene specific primers 232up (5'-CTCCACAAAGAGTGATCAGTC) and 232rev (5'-TTAGTTGTTTTCTTTGCGTTT); pSL230 plasmid DNA was used as a positive control. PCR samples were analyzed on 1% agarose gel in Tris-acetate-EDTA buffer and stained with ethidium bromide. Inocula from colonies *of L. lactis* MG1363, as well as colonies harboring either pJRS525 vector or pSL230 construct were used in subsequent experiments.

Western blot analysis: Cell-wall extracts were prepared as previously described (22). Briefly, cells grown to OD_{600} ~0.4 were harvested, washed with TES (10 mM Tris, 1mM EDTA, 25% Sucrose), re-suspended in TES-LMR (TES containing 1mg/ml hen egg lysozyme, 0.1 mg/ml mutanolysin, 0.1 mg/ml RNAseA and 1 mM PMSF) and incubated at 37ºC for 1 h. After centrifugation at 2500g for 10 min, the supernatants were precipitated with ice-cold TCA (16% final) at -20ºC overnight. Precipitates were rinsed thoroughly with ice-cold acetone and dissolved in 1x sample buffer at 250 μl per unit OD_{600}. Samples were subjected to 10% SDS-PAGE, transferred to nitrocellulose, and probed with anti-P176 antiserum followed by goat anti-rabbit-HRP and detected employing chemiluminescent substrate (Pierce).

Flow cytometry: Bacterial cells were grown to mid-log phase (OD_{600} ~0.4), washed once with filtered DPBS containing 1% FBS and re-suspended in the same buffer. Five million cells were incubated with 1:400 dilution of primary reagents, either rabbit pre-bleed (control) or rabbit anti-P176 antiserum for 30 min on ice, washed with DPBS-FBS and then incubated with 1:200 dilution of second reagent donkey anti-rabbit-APC (Jackson ImmunoResearch) for 30 min on ice. After a final wash and re-suspension in DPBS-FBS, flow cytometric data were acquired with FACSCaliber (BD Biosciences) and analyzed employing FCS Express (De Novo Software).

Analysis of biofilm formation

Crystal violet staining assay: Biofilm formation was tested using tissue culture treated polystyrene 24-well plates. 1.5 ml of logarithmic-phase GAS or *Lactococcus* cultures were seeded without dilution into wells and incubated at 37°C for GAS and 30°C for *Lactococcus* in an atmosphere of 5% CO_2–20% O_2 according to indicated time points upon which medium was aspirated. Wells were washed with PBS and 500 µl of 1% crystal violet was added to each well, and incubated at room temperature for 30 min. Dye was then aspirated, wells were washed with PBS, and stain was solubilized with 500 µl of 100% ethanol. Spectrophotometric readings at OD_{600} were recorded for each sample per time point. Samples were analyzed in triplicate in at least three experiments.

Confocal Laser Scanning Microscopy (CLSM): To visualize GAS and *L. lactis* strains by CLSM, bacterial cells were transformed with a GFP-encoding plasmid, pSB027 (31). 15-mm glass cover slips were placed into 24-well tissue culture plate wells. Logarithmic-phase bacterial cultures were inoculated without dilution and grown for 24 h. Cover slips were rinsed with PBS and fixed with 3% paraformaldehyde at room temperature for 30 min. Biofilms present on cover slips were washed with PBS and mounted onto slides using Prolong Gold mounting media (Invitrogen). Confocal images were acquired using a 63x/1.40 Plan-Apochromat objective and a Zeiss LSM 510 laser scanning confocal on an AxioImager Z1 microscope. An orthogonal view of the Z-stacks was used to display and measure biofilm thickness using Zeiss LSM software. Ten representative images

within a single experiment were used to calculate the average vertical thickness measured in micrometers.

To visualize extracellular matrix associated with GAS cells, 24 h biofilm samples were reacted with 100 μg of tetramethyl rhodamine isothiocyanate- (TRITC)-conjugated concanavalin A (TRITC-ConA) (Invitrogen) for 30 min at room temperature in the dark prior to mounting with Prolong Gold medium. An average of ten microscopic views within each sample was reviewed using the 63x/1.40 objective, as described above.

Field emission scanning electron microscopy (FESEM): GAS biofilm samples were grown for 24 h on glass cover slips, washed with PBS, and fixed with 3% paraformaldehyde for 2 h and post-fixed in osmium tetroxide. Samples were next dehydrated in an ethanol gradient, dried using hexamethyldisalizane, mounted onto aluminum stubs and sputter-coated with gold/palladium. The samples were then imaged on a Hitachi S-4800 field emission scanning electron microscope.

Quantitation of hydrophobicity

A modified hexadecane method (12,32,33) was used to determine the cell hydrophobicity. Briefly, 5 ml of the logarithmic-phase GAS or *Lactococcus* cultures (OD$_{600}$ ~0.5) were pelleted, washed and re-suspended in 5 ml of PBS. One ml of hexadecane was added, vortexed for 1 min and incubated for 10 min at 30°C. Mixtures were then vortexed for an additional 1 min and allowed to stand for 2 min for phase separation at room temperature. The absorbance of the lower aqueous phase was read at OD$_{600}$ and compared against the PBS control. Actual hydrophobicity value was calculated using the following equation: Actual Value = [1 − (A/A$_{o}$)] x 100, where A is OD$_{600}$ value after hexadecane treatment and A$_{o}$ is OD$_{600}$ prior to hexadecane treatment.

Statistical analysis

Statistical significance was determined using a two-tailed paired Student's *t*-test. The results were considered statistically significant with $P \le 0.05$ (*) and $P \le 0.001$ (**).

RESULTS

Wild-type GAS strains have heterogeneous capacity for biofilm formation on abiotic surfaces

Biofilm formation was compared between M41-, M28-, M3- and M1-type GAS strains representing distinct epidemiological traits (Figure 1). To assess biofilm formation after 24 h, we used spectrophotometric measurements recorded following crystal violet staining (Figure 1a). Both the M41- and M28-type strains produced more biomass as compared with M1 strain. Furthermore, the M3-type strain produced the lowest absorbance values in a crystal violet assay, indicative of lower cell biomass, as compared with the other wild-type strains. These experiments confirm previous observations (1,34) that GAS strains have varying capacity to form biofilm *in vitro*.

The failure of M3-type strain MGAS315 to produce substantial cellular biomass in the above assay was intriguing because sequence analysis of the *scl1.3* allele found in MGAS315 revealed the presence of a TAA stop codon in the 11th GXY repeat of the Scl1.3-CL region containing a total of 25 GXY triplets (35). This premature stop codon results in a truncated Scl1.3 variant composed of 102 amino acids (~11.4 kDa), which does not contain the cell wall-membrane (WM) associated region, thus, preventing it from anchoring to the bacterial cell surface (Figure 1b). This prompted us to investigate the biofilm formation by five additional M3-type strains, all harboring the same *scl1.3* allele. Five additional M3-type strains, MGAS335, MGAS1313, MGAS2079, MGAS274 and MGAS158, all harboring the same *scl1.3* allele (35) also produced poor biofilm under static conditions, as measured by crystal violet staining. Confocal laser scanning microscopy (CLSM) of three representative strains (MGAS315, MGAS2079, and MGAS158) corroborated results obtained from the crystal violet assay, indicating that these M3-type strains lack the ability to form appreciable biofilm structure. Our data suggest that the lack of capacity for biofilm-formation among M3-type GAS strains examined here might be correlated, at least in part, with lack of surface-attached Scl1.3 protein.

Microscopic evaluation reveals differences in biofilm surface morphology

We next conducted microscopic analysis of the biofilms formed by the wild-type (WT) M41-, M28-, and M1-type GAS strains. First, we examined the overall structural characteristics of biofilms formed after 24 h using CLSM (Figure 4 d-f); Additional file 1: Figure S1 a-f). The average biofilm thickness (see Methods section) differed among all three strains with M1 producing considerably thinner biofilm (mean value of 9 μm) compared to M28 (12 μm) and M41 (15 μm), a result consistent with lower spectrophotometric absorbance values (Figure 1a). In addition to measured differences in biofilm thickness, closer examination of the X-Y orthogonal Z-stack views, representing biofilm cross-sections, revealed architectural differences among the M41, M28, and M1 biofilms. The M1 biofilm, although the thinnest, seems to consist of densely-packed cells that form continuous layers, while the M28 and especially M41 biofilms seem to be less dense but exhibit more elevated supracellular assembly. We therefore used field emission scanning electron microscopy (FESEM) to define more accurately these supracellular differences observed by CLSM between the biofilms produced by the WT M1 and M41 GAS (Figure 2). FESEM exposed notable architectural differences between biofilms formed by these two strains. The M41 (Figure 2, panel a) biofilm was characterized by more diverse surface architecture with the evidence of depressions or crypts, whereas the M1 biofilm (panel b) seems to lack such pronounced surface characteristics. At higher magnification, the M41 cells have a studded cell surface morphology with protrusions linking both sister cells and cells in adjacent chains (panel c). In contrast, the M1 cells had a relatively smoother appearance likely due to the rich bacterial-associated extracellular matrix (BAEM) surrounding these cells and covering their surface (panel d). BAEM material, which was clearly seen at higher resolution between the M1-type cells, was not as evident between cells of the M41-type GAS.

GAS biofilms differ in production of bacterial-associated extracellular matrix

The production of BAEM has been shown to be an integral component in the structural integrity of a biofilm, imparting protection from dehydration, host immune attack, and antibiotic sensitivity (36,37). GAS cells encased in a glycocalyx were first identified by Akiyama *et al.* in skin biopsies obtained from impetigo patients. We therefore compared

the production of BAEM within biofilms employing GFP-expressing GAS strains of the M1 and M41 type (Figure 3). Cells were grown to form biofilms on glass cover slips for 24 h and stained with TRITC-concanavalin A (ConA), a fluorescently-labeled lectin that binds to the extracellular polysaccharides in biofilms (38). Fluorescent microscopy was performed to compare matrix production (red staining) by GAS strains (green). Visual screening of both biofilms indicated that the M41-type strain formed a more dispersed extracellular matrix as compared to the M1 strain, which had a dense, more closely associated matrix. In addition, averages of at least 10 fields of ConA stained matrix by CLSM support our FESEM observations that more BAEM is deposited within the biofilm by the M1 GAS cells as compared to M41 GAS. This is in agreement with the report from Akiyama *et al* that showed a substantial FITC-ConA stained matrix associated with T1-type GAS microcolonies *in vivo* and *in vitro* (10).

Scl1 protein significantly contributes to biofilm formation by GAS

Variations in GAS pathogenicity and capacity to form biofilm are driven by specific proteins and components present on the cell surface or are secreted by the organism. It has been shown that deletion of the M and M-like surface proteins or capsule, as well as increased expression of the secreted SpeB protease decreases biofilm formation dramatically for some strains of GAS (12,39,40). Therefore, we investigated the role of Scl1 in biofilm formation by comparing biofilms formed by GAS WT and *scl1*-inactivated (Δ*scl1)* mutant strains (Figure 4; Additional file 1: Figure S1 a-f). Bacterial biomass was evaluated spectrophotometrically following crystal violet staining at 1, 6, 12, and 24 h time points, representing different stages of biofilm formation, and absorbance values rendered for the WT and Δ*scl1* isogenic mutant strains were compared. The M41Δ*scl1* mutant showed a 29-35% decrease in biofilm formation (the OD_{600} value obtained for the WT strain at each time point was considered 100%), which was sustained throughout all time points. This reduction was statistically significant at initial adherence (1 h), as well as during biofilm development (6-12 h) and at maturation (24 h) (Figure 4a; $P \leq 0.05$ at 1 and 12 h, $P \leq 0.001$ at 6 and 24 h). Complementation of Scl1.41 expression in the M41Δ*scl1* mutant (M41 C) restored its ability to form biofilm to WT levels. Similarly, the M28Δ*scl1* mutant had a significantly decreased capacity for biofilm formation in the range

of 29-44% as compared to WT strain (Figure 4b; $P \le 0.05$ at 1 and 6 h, $P \le 0.001$ at 3, 12 and 24 h). Likewise, there was a statistically significant decrease in M1$\Delta scl1$ biofilm biomass by ~42-75% compared to the WT strain (Figure 4c; $P \le 0.001$ at 1-24 h). CLSM analysis of corresponding 24-h biofilms of these strains confirmed our crystal violet staining results at 24 h. The $\Delta scl1$ mutants had substantially decreased average biofilm thickness by more than 50% (mean values) as compared to the parental WT organisms (Figure 4d-f). While these low average biofilm thickness values measured for the M1$\Delta scl41$ (6 μM) and M28$\Delta scl1$ (5 μM) correspond to residual biofilms made by those mutants (Figure S1a-d), by comparison, the M1$\Delta scl1$ (4 μM) was shown not to produce a continuous biofilm layer under these conditions (Figure S1e-f). Our data support the hypothesis that the Scl1 protein plays an important functional role during GAS biofilm formation and that Scl1 contribution varies among GAS strains with different genetic backgrounds.

Scl1 expression affects surface hydrophobicity

The surface hydrophobicity of GAS has been shown to influence the adherence to abiotic surfaces. The presence of pili (13) , M and M-like proteins, and lipoteichoic acid contributes to cell surface hydrophobic properties (12,41), which in turn may influence biofilm formation by GAS. Here, we have investigated the contribution of Scl1 to surface hydrophobicity of M41-, M28-, and M1-type GAS strains using a modified hexadecane binding assay (12,33,42). As shown in Table 1, the M28-type GAS strain MGAS6143 gave the highest actual hydrophobicity value of 94.3 ± 0.73, followed by the M41-type strain MGAS6183 (92.6 ± 0.86). In contrast, the overall surface hydrophobicity of the M1-type GAS strain MGAS5005 (80.3 ± 0.89) was significantly lower compared to both M28 and M41 strains ($P \le 0.001$ for each comparison). Inactivation of $scl1.41$ in M41-type GAS resulted in a modest, although statistically significant, reduction in the hydrophobicity index (100% for WT vs. 92% for mutant, $P \le 0.001$). *In-trans* complementation of the Scl1.41 expression in M41$\Delta scl1$-C restored the hydrophobic phenotype of the cells to WT level (hydrophobicity index ~105%). In comparison, the contribution of the Scl1.1 and Scl1.28 proteins to surface hydrophobicity is more substantial, as evidenced by a ~21% and ~22% reduction of the hydrophobicity indices of the mutants as compared to the

24

corresponding WT strains, respectively ($P \leq 0.001$ for both). Thus, the Scl1-mediated GAS-cell surface hydrophobicity reported here may contribute to the ability of this organism to form biofilm, as suggested for other cell surface components (12,41).

Scl1 is sufficient to support biofilm formation in *Lactococcus lactis*

To assess whether Scl1 expression is sufficient to confer the ability for biofilm formation, we chose to express this protein in a heterologous *L. lactis* system (43,44). The wild-type *L. lactis* strain MG1363 was transformed with plasmid pSL230 encoding the Scl1.41 protein (22) or with the shuttle vector pJRS525 alone. As shown in Figure 5a, PCR amplification of the *scl1.41* gene employing specific primers yielded no product from the WT *L. lactis* MG1363 (lane 1) and the MG1363::pJRS525 transformant (lane 2). A product of the expected size of 1.4 kb was amplified from the pSL230 plasmid DNA control (lane 4,) as well as was amplified from the MG1363::pSL230 transformant (lane 3). Surface expression of Scl1.41 was confirmed by immunoblot analysis of cell-wall extracts prepared from *L. lactis* WT, and the MG1363::pJRS525 and MG1363::pSL230 transformants, as well as MGAS6163 (WT M41 GAS). As shown in Figure 5b, rabbit antiserum raised against purified recombinant Scl1.41 protein P176 lacking the WM region detected the corresponding immunogen (lane 1), and the homologous full length protein in cell-wall extracts of MGAS6183 (lane 5) as well as MG1363::pSL230 *L. lactis* transformant (lane 4). This band was absent in cell-wall extracts prepared from the WT *L. lactis* MG1363 (lane 2) and MG1363::pJRS525 transformant (lane 3). Expression of Scl1.41 at the cell surface was further established by flow cytometry. Rabbit anti-p176 antibodies stained Scl1.41 MG1363::pSL230 transformant, confirming the expression of Scl1.41 protein at the cell surface in the heterologous host *L. lactis* (Figure 5c, red trace). This protein was absent at the surface of WT MG1363 (black trace) and MG1363::pJRS525 transformant (green trace).

The capacity of *L. lactis* expressing Scl1.41 to form biofilm was evaluated spectrophotometrically following crystal violet staining. As shown in Figure 5d, the MG1363::pSL230 transformant demonstrated a significant increase in biofilm-associated biomass at 24 h, as compared to wild type *L. lactis* or *L. lactis*-containing pJRS525 vector ($P \leq 0.001$). Crystal violet stained wells were photographed for visual representation of

biofilm formation prior to spectrophotometric assay. Biofilm thickness and architecture were evaluated by CLSM (Figure 5e; Additional file 1: Figure S2 a-c). The MG1363::pSL230 transformant produced a substantially thicker biofilm (14 μm) as compared to both MG1363 WT (6 μm) and the vector-only transformant MG1363::pJRS525 (6 μm). The MG1363::pSL230 cells formed highly aggregated structures, thus, acquiring a phenotype consistent with biofilm formation. As shown in Table 2, the MG1363::pSL230 transformant, expressing Scl1.41 surface protein, had significantly enhanced cell surface hydrophobicity (hydrophobicity index of ~137% vs. 100% WT, $P \leq 0.001$) with an actual value of 82.0 ± 2.6, when compared to the MG1363 WT (59.7 ± 7.2) and the vector-only MGAS1363::pJRS525 control (56.6 ± 5.5). These data suggest a direct relationship between Scl1 expression and cell surface hydrophobicity and establish their involvement in the microorganism's ability to form biofilm *in vitro*.

DISCUSSION

Group A *Streptococcus* strains vary because of the vast number of M-protein types, and this variation is associated with varying frequency of isolation and exacerbation of disease (45,46). The M41-, M28-, M3-, and M1-type strains selected for the current study represent a significant intraspecies diversity among clinical isolates of GAS. M41 GAS was a major causative agent of superficial skin infections (47-49), and strain MGAS6183, harboring the Scl1.41 protein, has been studied extensively (19,21,22). M28-type GAS (strain MGAS6143) has historically been associated with puerperal fever and currently is responsible for extensive human infections world-wide (50). M1T1 GAS, represented by strain MGAS5005, is a globally disseminated clone responsible for both pharyngitis and invasive infections (51-53). The M3-type strains of GAS cause a disproportionally large number of invasive GAS infections that are responsible for traumatic morbidity and death (54,55).

Initial studies by Lembke *et al.* that characterized biofilm formation among various M types of GAS typically included several strains of the same M type (1,34). These studies reported a significant strain-to-strain variation in ability to form biofilms within each M type. Studies that followed compared biofilm formation by defined isogenic WT and mutant

strains to assess the contribution of specific GAS surface components responsible for a biofilm phenotype, including M and M-like proteins, hyaluronic acid capsule, lipoteichoic acid, and pili (12,13). In the current study, we have assessed the role and contribution of the surface protein Scl1 in the ability to support biofilm formation by GAS strains of four distinct M types.

Recent advances in molecular mega- and pathogenomics has enabled the characterization of numerous M3-type strains with a single nucleotide resolution (56,57). Interestingly, all five M3-type strains MGAS158, 274, 315, 335, and 1313 that were originally used for *scl1*-gene sequencing (14), plus an additional strain MGAS2079 (not reported) harbor the same *scl1.3* allele containing a null mutation that would result in secretion of a truncated Scl1.3-protein variant. Here, we demonstrate that these GAS strains do not form biofilm on an abiotic surface. Recently, bioinformatic screening of the sequences of ~250 invasive M3-type strains isolated globally, has led to the detection of this single nucleotide polymorphism that results in disruption of Scl1.3 protein (Steve Beres and Jim Musser, personal communication). Lembke *et al.* reported heterogeneous biofilm formation among four M3-type GAS strains examined over a 24, 48, and 72-hour period (34). Biofilm was detected for one strain at a 48 h time point, on a fibrinogen-coated surface; however, it is not known whether this clinical isolate forms biofilm on abiotic surface, whether it expresses the truncated or full-length Scl1.3 protein, and whether it produces an unknown fibrinogen-binding protein, which could augment the attachment and biofilm formation. Therefore, additional studies are necessary to define the contributions of other biofilm-formation determinants in M3-type strains.

Inasmuch as, variation in biofilm formation among GAS isolates of the same M-type has been established, the molecular basis of this phenotypic variation is not known. Several GAS surface-associated and secreted components were shown to contribute to variation in biofilm (12,13,39). In addition, transcription regulators, such as Mga, CovR, and Srv are likely to play substantial roles in GAS biofilm formation (11,39) due to their transcriptional regulation of numerous genes. Therefore, it is logical to assume that the combination of genomic/proteomic make up, allelic polymorphisms, and transcription regulation all contribute to this phenomenon. In addition, discrepancies between *in vitro* data obtained with laboratory-stored strains and microcolony formation *in vivo* likely exist

and add yet another unknown to the complexity of GAS biofilm/microcolony formation and its role in pathogenesis. Despite this complexity, the analyses involving isogenic strains of the same genetic background provide valuable information that allows assessment of the role and contribution of a given GAS component to biofilm formation.

The M1 MGAS5005 strain was shown to form biofilm *in vitro* and in experimental animals (8,39,58), and the present study demonstrates a significant role of Scl1.1 in this process. Likewise, the MGAS6183 strain, representing M41-type isolates often associated with pyoderma, produced a more robust biofilm biomass under the same experimental conditions and Scl1.41-deficient mutant was found to be an important determinant in this process. Similarly, Scl1.28 protein significantly contributes to a robust biofilm made by the M28-type strain MGAS6143. However, a recent study reported that another surface protein, designated AspA, found in M28-type GAS significantly contributed to biofilm formation (59). The Δ*aspA* isogenic mutant showed 60% reduction in biofilm formation. The strain MGAS6180, which they used, expresses the same Scl1.28 variant present in the MGAS6143 strain we used; our Scl1.28 mutant showed ~44% reduction in 24 h biofilm. We propose that several surface proteins contribute to biofilm formation by M28-type strains including proteins AspA and Scl1.28, and potentially, proteins F1/SfbI and F2 that are also present in these strains (22). This redundancy is likely responsible for the observed residual biofilms produced by the AspA- and Scl1.28-deficient mutants.

The observed heterogeneity in biofilm architecture of different GAS strains was previously observed by Lembke *et al.* (34) and was also documented in the current study using FESEM. In addition, here we report the differences in GAS-cell surface morphology and within cell-to-cell junctions in biofilms formed by M1- and M41-type strains. The structural and genetic determination of these differences is not known since M41 genome has not been sequenced, but may be associated with the presence of additional surface proteins, such as the F2 protein (60) encoded by *prtf2* gene found in this strain (22). Even more striking was an observed difference in the amount of the extracellular material associated with each strain, referred to as BAEM (bacteria-associated extracellular matrix). It has been shown that extracellular matrix, also called glycocalyx, is produced by biofilm-forming bacteria. DNA, lipids, proteins (39), polysaccharides and dead cell

debris (61) were identified in this matrix and for gram-positive bacteria, teichoic acids have also been detected (62,63). The exopolysaccharide component of the glycocalyx is detected using carbohydrate-binding lectins, such as concanavalin A (ConA) (10). Both FESEM analysis and ConA staining detected more BAEM associated with M1 biofilm compared to M41, which produced larger biofilm. These observations suggest that GAS biofilm is stabilized differently by different strains and that higher BAEM production does not necessarily pre-determine larger biofilm mass. Consequently, a combination of biofilm features rather than biofilm size alone may be more relevant to pathogenicity of a given GAS strain.

Diminished adherence and biofilm formation could be associated with changes in cell surface hydrophobicity (64) of the *scl1* mutants. Indeed, the lack of Scl1 resulted in both decreased hydrophobicity and the ability to form biofilm, albeit in a somewhat disproportionate manner. A decrease in the hydrophobicity index by only ~8%, as compared to the wild type-strain, was measured for the M41Δ*scl1* mutant and this modest decrease was accompanied by a rather large reduction in biofilm formation capacity after 24 h by 30%. Greater decrease in cell-surface hydrophobicity was measured for the M1Δ*scl1* (~21%) and M28Δ*scl1* (~22%) mutants, which was accompanied by a significant loss in biofilm formation after 24 h by both isogenic strains by ~55% and ~41% ($P ≤ 0.001$ for each comparison), respectively. In addition, heterologous expression of Scl.41 in *L. lactis* increased hydrophobicity index of this organism to ~137% of the WT level, which was accompanied by significant increase in its ability to form biofilm. Similar observations have been reported for the M and M-like protein mutants that typically, but not always, exhibit concurrent loss of both biological features (12). For example, isogenic ΔMrp49 mutant had a non-significant drop in hydrophobicity (~2%) but significantly lower biofilm formation after 48 h by ~30%, whereas ΔEmm1 mutant lost ~78% hydrophobicity and ~44% biofilm formation capacity. In summary: (i) here we report that the Scl1 adhesin is also a hydrophobin with varying contribution to the overall surface hydrophobicity among GAS strains representing different M types and (ii) Scl1-associated surface hydrophobicity is likely to contribute to Scl1-mediated biofilm formation.

To test whether Scl1 alone could support biofilm formation, we used a heterologous *L. lactis* strain, which provides an expression system for membrane-bound

proteins of gram- positive bacteria with LPXTG cell-wall anchoring motifs (30,44,65,66), including the group A streptococcal M6 protein (43,67). In a recent study by Maddocks *et al.* (59) it was shown that heterologous expression of AspA GAS surface protein was able to induce a biofilm phenotype in *L. lactis* MG1363. We were also able to achieve a gain-of-function derivative of the *L. lactis* WT MG1363 strain, (MG1363::pSL230), displaying an altered phenotype associated with biofilm formation, as compared to wild-type parental and vector-only controls. These data support our current model that Scl1 protein is an important determinant of GAS biofilm formation.

As shown by crystal violet staining and CLSM, biofilm formation by the Scl1-negative mutants was compromised during the initial stage of adherence, as well as microcolony stabilization and maturation. Consequently, their capacity for biofilm formation as compared to the respective WT controls was greatly reduced. This comparison identifies for the first time that the Scl1 protein contributes significantly to biofilm assembly and stability. Based on these observations, as well as previous work by us and others, we propose the following model of Scl1 contribution to GAS tissue microcolony formation (Figure 6). First, the Scl1 hydrophobin (current study) initiates bacterial adhesion to animate surfaces within the host (64). Next, the Scl1 adhesin anchors the outside edge of growing microcolony in tissue by direct binding to tissue extracellular matrix components, cellular fibronectin and laminin (19). Microcolony development is stabilized by Scl1-Scl1 scaffolding resulting from Scl1's capacity to form head-to-head dimers (68) between molecules located on adjacent chains. This model will be tested experimentally in future studies.

CONCLUSIONS

In the present work, using pathogenically differing GAS strains, we have demonstrated three concepts. First, we have confirmed previous observations that biofilm formation is an innate property of GAS strains. The M41-type strain used formed a more robust biofilm when compared to M28-type strain as well as M1-type strain. Importantly, the highly invasive M3-type strains devoid of the surface-associated Scl1 also lack the ability to form biofilm. Secondly, the absence of surface-associated Scl1 decreases GAS-cell hydrophobicity suggesting that Scl1 plays a role on the GAS surface as a hydrophobin.

Thirdly, we have established that the Scl1 protein is a significant determinant for GAS biofilm formation. This concept was further tested by the heterologous expression of Scl1 in *Lactococcus,* an organism found in dairy fermentation environments, enabling it to form biofilm. Altogether, these data underscore the importance of Scl1 in biofilm-associated regulation of GAS pathogenicity. Recently published work has shown that the recombinant Scl1 binds to the extracellular matrix components, cellular fibronectin and laminin (19). Our current research provides a foundation warranting additional investigation as to whether direct Scl1-ECM binding may promote GAS biofilm as a bridging mechanism within host tissues.

ACKNOWLEDGMENTS

REFERENCES

1. Conley, J., Olson, M. E., Cook, L. S., Ceri, H., Phan, V., and Davies, H. D. (2003) Biofilm formation by group a streptococci: is there a relationship with treatment failure? *J Clin Microbiol* **41**, 4043-4048

2. Ogawa, T., Terao, Y., Okuni, H., Ninomiya, K., Sakata, H., Ikebe, K., Maeda, Y., and Kawabata, S. (2011) Biofilm formation or internalization into epithelial cells enable Streptococcus pyogenes to evade antibiotic eradication in patients with pharyngitis. *Microb Pathog* **51**, 58-68

3. Boles, B. R., Thoendel, M., and Singh, P. K. (2005) Genetic variation in biofilms and the insurance effects of diversity. *Microbiology* **151**, 2816-2818

4. Lauderdale, K. J., Malone, C. L., Boles, B. R., Morcuende, J., and Horswill, A. R. (2010) Biofilm dispersal of community-associated methicillin-resistant *Staphylococcus aureus* on orthopedic implant material. *J Orthop Res* **28**, 55-61

5. Kaplan, J. B., Meyenhofer, M. F., and Fine, D. H. (2003) Biofilm growth and detachment of *Actinobacillus actinomycetemcomitans. J Bacteriol* **185**, 1399-1404

6. Raad, II, Fang, X., Keutgen, X. M., Jiang, Y., Sherertz, R., and Hachem, R. (2008) The role of chelators in preventing biofilm formation and catheter-related bloodstream infections. *Curr Opin Infect Dis* **21**, 385-392

7. Wang, R., Khan, B. A., Cheung, G. Y., Bach, T. H., Jameson-Lee, M., Kong, K. F., Queck, S. Y., and Otto, M. (2011) *Staphylococcus epidermidis* surfactant peptides promote biofilm maturation and dissemination of biofilm-associated infection in mice. *J Clin Invest* **121**, 238-248

8. Connolly, K. L., Roberts, A. L., Holder, R. C., and Reid, S. D. (2011) Dispersal of Group A streptococcal biofilms by the cysteine protease SpeB leads to increased disease severity in a murine model. *PLoS ONE* **6**, e18984

9. Cunningham, M. W. (2000) Pathogenesis of group A streptococcal infections. *Clin Microbiol Rev* **13**, 470-511

10. Akiyama, H., Morizane, S., Yamasaki, O., Oono, T., and Iwatsuki, K. (2003) Assessment of *Streptococcus pyogenes* microcolony formation in infected skin by confocal laser scanning microscopy. *J Dermatol Sci* **32**, 193-199

11. Cho, K., and Caparon, M. (2005) Patterns of virulence gene expression differ between biofilm and tissue communities of *Streptococcus pyogenes. Mol Microbiol* **57**, 1545-1556

12. Courtney, H. S., Ofek, I., Penfound, T., Nizet, V., Pence, M. A., Kreikemeyer, B., Podbielski, A., Hasty, D. L., et al. (2009) Relationship between expression of the family of M proteins and lipoteichoic acid to hydrophobicity and biofilm formation in *Streptococcus pyogenes. PLoS ONE* **4**, e4166

13. Manetti, A., Zingaretti, C., Falugi, F., Capo, S., Bombaci, M., Bagnoli, F., Gambellini, G., Bensi, G., et al. (2007) *Streptococcus pyogenes* pili promote pharyngeal cell adhesion and biofilm formation. *Mol Microbiol* **64**, 968-983

14. Lukomski, 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

15. Lukomski, S., Nakashima, K., Abdi, I., Cipriano, V. J., Shelvin, B. J., Graviss, 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

16. Rasmussen, M., Eden, A., and Bjorck, L. (2000) SclA, a novel collagen-like surface protein of *Streptococcus pyogenes*. *Infect Immun* **68**, 6370-6377

17. Almengor, A. C., and McIver, K. S. (2004) Transcriptional activation of *sclA* by Mga requires a distal binding site in *Streptococcus pyogenes*. *J. Bacteriol.* **186**, 7847-7857

18. Almengor, A. C., Walters, M. S., and McIver, K. S. (2006) Mga is sufficient to activate transcription in vitro of *sof-sfbX* and other Mga-regulated virulence genes in the group A *Streptococcus*. *J. Bacteriol.* **188**, 2038-2047

19. Caswell, C. C., Oliver-Kozup, H., Han, R., Lukomska, E., and Lukomski, S. (2010) Scl1, the multifunctional adhesin of group A *Streptococcus*, selectively binds cellular fibronectin and laminin, and mediates pathogen internalization by human cells. *FEMS Microbiol Lett* **303**, 61-68

20. Humtsoe, J. O., Kim, J. K., Xu, Y., Keene, D. R., Hook, M., Lukomski, S., and Wary, K. K. (2005) A streptococcal collagen-like protein interacts with the a2b1 integrin and induces intracellular signaling. *J. Biol. Chem.* **280**, 13848-13857

21. Caswell, C. C., Barczyk, M., Keene, D. R., Lukomska, E., Gullberg, D. E., and Lukomski, S. (2008) Identification of the first prokaryotic collagen sequence motif that mediates binding to human collagen receptors, integrins a2b1 and a11b1. *J Biol Chem* **283**, 36168-36175

22. Caswell, C. C., Lukomska, E., Seo, N. S., Hook, M., and Lukomski, S. (2007) Scl1-dependent internalization of group A *Streptococcus* via direct interactions with the alpha2beta(1) integrin enhances pathogen survival and re-emergence. *Mol Microbiol* **64**, 1319-1331

23. Gao, Y., Liang, C., Zhao, R., Lukomski, S., and Han, R. (2010) The Scl1 of M41-type group A *Streptococcus* binds the high-density lipoprotein. *FEMS Microbiol Lett* **309**, 55-61

24. Pahlman, L. I., Marx, P. F., Morgelin, M., Lukomski, S., Meijers, J. C. M., and Herwald, H. (2007) Thrombin-activatable Fibrinolysis Inhibitor Binds to Streptococcus pyogenes by Interacting with Collagen-like Proteins A and B. *J. Biol. Chem.* **282**, 24873-24881

25. Caswell, C., Han, R., Hovis, K., Ciborowski, P., Keene, D., Marconi, R., and Lukomski, S. (2008) The Scl1 protein of M6-type group A *Streptococcus* binds the human complement regulatory protein, factor H, and inhibits the alternative pathway of complement. *Mol Microbiol* **67**, 584-596

26. Reuter, M., Caswell, C. C., Lukomski, S., and Zipfel, P. F. (2010) Binding of the human complement regulators CFHR1 and factor H by streptococcal collagen-like protein 1 (Scl1) via their conserved C termini allows control of the complement cascade at multiple levels. *J Biol Chem* **285**, 38473-38485

27. Han, R., Caswell, C. C., Lukomska, E., Keene, D. R., Pawlowski, M., Bujnicki, J. M., Kim, J. K., and Lukomski, S. (2006) Binding of the low-density lipoprotein by streptococcal collagen-like protein Scl1 of *Streptococcus pyogenes*. *Mol Microbiol* **61(2)**, 351-367

28. Lukomski, S., Hoe, N. P., Abdi, I., Rurangirwa, J., Kordari, P., Liu, M., Dou, S. J., Adams, G. G., et al. (2000) Nonpolar inactivation of the hypervariable

streptococcal inhibitor of complement gene (sic) in serotype M1 Streptococcus pyogenes significantly decreases mouse mucosal colonization. *Infect Immun* **68**, 535-542

29.   Holo, H., and Nes, I. F. (1989) High-Frequency Transformation, by Electroporation, of *Lactococcus lactis* subsp. *cremoris* Grown with Glycine in Osmotically Stabilized Media. *Appl Environ Microbiol* **55**, 3119-3123

30.   Que, Y. A., Haefliger, J. A., Francioli, P., and Moreillon, P. (2000) Expression of *Staphylococcus aureus* clumping factor A in *Lactococcus lactis* subsp. *cremoris* using a new shuttle vector. *Infect Immun* **68**, 3516-3522

31.   Cramer, T., Yamanishi, Y., Clausen, B. E., Forster, I., Pawlinski, R., Mackman, N., Haase, V. H., Jaenisch, R., et al. (2003) HIF-1alpha is essential for myeloid cell-mediated inflammation. *Cell* **112**, 645-657

32.   Grivet, M., Morrier, J. J., Benay, G., and Barsotti, O. (2000) Effect of hydrophobicity on in vitro streptococcal adhesion to dental alloys. *J Mater Sci Mater Med* **11**, 637-642

33.   Pan, W. H., Li, P. L., and Liu, Z. (2006) The correlation between surface hydrophobicity and adherence of *Bifidobacterium* strains from centenarians' faeces. *Anaerobe* **12**, 148-152

34.   Lembke, C., Podbielski, A., Hidalgo-Grass, C., Jonas, L., Hanski, E., and Kreikemeyer, B. (2006) Characterization of Biofilm Formation by Clinically Relevant Serotypes of Group A Streptococci. *Appl. Environ. Microbiol.* **72**, 2864-2875

35.   Lukomski, S., Sreevatsan, S., Amberg, C., Reichardt, W., Woischnik, M., Podbielski, A., and Musser, J. M. (1997) Inactivation of *Streptococcus pyogenes* extracellular cysteine protease significantly decreases mouse lethality of serotype M3 and M49 strains. *J. Clin. Invest.* **99**, 2574-2580

36.   Donlan, R. M. (2002) Biofilms: microbial life on surfaces. *Emerg Infect Dis* **8**, 881-890

37.   Kania, R. E., Lamers, G. E., Vonk, M. J., Huy, P. T., Hiemstra, P. S., Bloemberg, G. V., and Grote, J. J. (2007) Demonstration of bacterial cells and glycocalyx in biofilms on human tonsils. *Arch Otolaryngol Head Neck Surg* **133**, 115-121

38.   Maeyama, R., Mizunoe, Y., Anderson, J. M., Tanaka, M., and Matsuda, T. (2004) Confocal imaging of biofilm formation process using fluoroprobed *Escherichia coli* and fluoro-stained exopolysaccharide. *J Biomed Mater Res A* **70**, 274-282

39.   Doern, C. D., Roberts, A. L., Hong, W., Nelson, J., Lukomski, S., Swords, W. E., and Reid, S. D. (2009) Biofilm formation by group A *Streptococcus:* a role for the streptococcal regulator of virulence (Srv) and streptococcal cysteine protease (SpeB). *Microbiology* **155**, 46-52

40.   Roberts, A. L., Connolly, K. L., Doern, C. D., Holder, R. C., and Reid, S. D. (2010) Loss of the group A *Streptococcus* regulator Srv decreases biofilm formation in vivo in an otitis media model of infection. *Infect Immun* **78**, 4800-4808

41.   Wadstrom, T., Schmidt, K. H., Kuhnemund, O., Havlicek, J., and Kohler, W. (1984) Comparative Studies on Surface Hydrophobicity of Streptococcal Strains of Group-a, Group-B, Group-C, Group-D and Group-G. *J. Gen. Microbiol.* **130**, 657-664

42.	Grivetti, L. E., and Ogle, B. M. (2000) Value of traditional foods in meeting macro- and micronutrient needs: the wild plant connection. *Nutr Res Rev* **13**, 31-46

43.	Piard, J. C., Hautefort, I., Fischetti, V. A., Ehrlich, S. D., Fons, M., and Gruss, A. (1997) Cell wall anchoring of the Streptococcus pyogenes M6 protein in various lactic acid bacteria. *J. Bacteriol.* **179**, 3068-3072

44.	Linares, D. M., Kok, J., and Poolman, B. (2010) Genome Sequences of *Lactococcus lactis* MG1363 (Revised) and NZ9000 and Comparative Physiological Studies. *J. Bacteriol.* **192**, 5806-5812

45.	Whatmore, A. M., Kapur, V., Sullivan, D. J., Musser, J. M., and Kehoe, M. A. (1994) Non-congruent relationships between variation in *emm* gene sequences and the population genetic structure of group A streptococci. *Mol. Microbiol.* **14**, 619-631

46.	Bessen, D. E., Sotir, C. M., Readdy, T. L., and Hollingshead, S. K. (1996) Genetic corelates of throat and skin isolates of group A streptococci. *J. Infect. Dis.* **173**, 896-900

47.	Anthony, B. F. (2000) Streptococcal pyoderma. in *Streptococcal infections* (Stevens, D. L., and Kaplan, E. L. eds.), Oxford University Press, New York, N. Y. pp 144-151

48.	Anthony, B. F., Perlman, L. V., and Wannamaker, L. W. (1967) Skin Infections And Acute Nephritis In American Indian Children. *Pediatrics* **39**, 263-279

49.	Top, F. H., Jr., Wannamaker, L. W., Maxted, W. R., and Anthony, B. F. (1967) M antigens among group A streptococci isolated from skin lesions. *J Exp Med* **126**, 667-685

50.	Green, N. M., Beres, S. B., Graviss, E. A., Allison, J. E., McGeer, A. J., Vuopio-Varkila, J., LeFebvre, R. B., and Musser, J. M. (2005) Genetic Diversity among Type emm28 Group A Streptococcus Strains Causing Invasive Infections and Pharyngitis. *J Clin Microbiol* **43**, 4083-4091

51.	Aziz, R., and Kotb, M. (2008) Rise and persistence of global M1T1 clone of *Streptococcus pyogenes. Emerg Infect Dis* **14**, 1511-1517

52.	Aziz, R. K., Edwards, R. A., Taylor, W. W., Low, D. E., McGeer, A., and Kotb, M. (2005) Mosaic Prophages with Horizontally Acquired Genes Account for the Emergence and Diversification of the Globally Disseminated M1T1 Clone of *Streptococcus pyogenes. J. Bacteriol.* **187**, 3311-3318

53.	Musser, J. M., Kapur, V., Szeto, J., Pan, X., Swanson, D. S., and Martin, D. R. (1995) Genetic diversity and relationships among *Streptococcus pyogenes* strains expressing serotype M1 protein: recent intercontinental spread of a subclone causing episodes of invasive disease. *Infect. Immun.* **63**, 994-1003

54.	Kaul, R., McGeer, A., Low, D. E., Green, K., Schwartz, B., Study, O. G. A. S., and Simor, A. E. (1997) Population-based surveillance for group A streptococcal necrotizing fasciitis: clinical features, prognostic indicators, and microbiologic analysis of seventy-seven cases. *Am. J. Med.* **103**, 18-24

55.	Sharkawy, A., Low, D. E., Saginur, R., Gregson, D., Schwartz, B., Jessamine, P., Green, K., and McGeer, A. (2002) Severe group a streptococcal soft-tissue infections in Ontario: 1992-1996. *Clin Infect Dis* **34**, 454-460

56.	Beres, S. B., Sylva, G. L., Barbian, K. D., Lei, B., Hoff, J. S., Mammarella, N. D., Liu, M.-Y., Smoot, J. C., et al. (2002) Genome sequence of a serotype M3 strain

of group A *Streptococcus*: Phage-encoded toxins, the high-virulence phenotype, and clone emergence. *Proc. Natl. Acad. Sci. USA* **99**, 10078-10083

57. Beres, S. B., Sylva, G. L., Sturdevant, D. E., Granville, C. N., Liu, M., Ricklefs, S. M., Whitney, A. R., Parkins, L. D., et al. (2004) Genome-wide molecular dissection of serotype M3 group A *Streptococcus* strains causing two epidemics of invasive infections. *Proc. Natl. Acad. Sci. USA* **101**, 11833-11838

58. Roberts, A. L., Connolly, K. L., Doern, C. D., Holder, R. C., and Reid, S. D. (2010) Loss of the group A *Streptococcus* regulator Srv decreases biofilm formation in vivo in an otitis media model of infection. *Infect Immun* **78**, 4800-4808

59. Maddocks, S. E., Wright, C. J., Nobbs, A. H., Brittan, J. L., Franklin, L., Stromberg, N., Kadioglu, A., Jepson, M. A., et al. (2011) Streptococcus pyogenes antigen I/II-family polypeptide AspA shows differential ligand-binding properties and mediates biofilm formation. *Mol Microbiol* **81**, 1034-1049

60. Jaffe, J., Natanson-Yaron, S., Caparon, M. G., and Hanski, E. (1996) Protein F2, a novel fibronectin-binding protein from *Streptococcus pyogenes*, possesses two domains. *Mol Microbiol* **21**, 373-384

61. Branda, S. S., Gonzalez-Pastor, J. E., Dervyn, E., Ehrlich, S. D., Losick, R., and Kolter, R. (2004) Genes involved in formation of structured multicellular communities by Bacillus subtilis. *J. Bacteriol.* **186**, 3970-3979

62. Gotz, F. (2002) *Staphylococcus* and biofilms. *Mol. Microbiol.* **43**, 1367-1378

63. Nadell, C. D., Xavier, J. B., and Foster, K. R. (2009) The sociobiology of biofilms. *FEMS Microbiol. Rev.* **33**, 206-224

64. Courtney, H. S., Dale, J. B., and Hasty, D. L. (2000) Strategies for preventing group A streptococcal adhesion and infection. in *Handbook of bacterial adhesion: principles, methods, and applications* (An, Y. H., and Friedman, R. J. eds.), Humana Press, Inc., Totowa, N. J. pp 553-579

65. Luo, H. L., Wan, K., and Wang, H. H. (2005) High-frequency conjugation system facilitates biofilm formation and pAM beta 1 transmission by *Lactococcus lactis*. *Appl Environ Microb* **71**, 2970-2978

66. Gerber, S. D., and Solioz, M. (2007) Efficient transformation of *Lactococcus lactis* IL1403 and generation of knock-out mutants by homologous recombination. *J Basic Microb* **47**, 281-286

67. Piard, J. C., JimenezDiaz, R., Fischetti, V. A., Ehrlich, S. D., and Gruss, A. (1997) The M6 protein of Streptococcus pyogenes and its potential as a tool to anchor biologically active molecules at the surface of lactic acid, bacteria. *Streptococci and the Host* **418**, 545-550

68. Xu, Y., Keene, D. R., Bujnicki, J. M., Hook, M., and Lukomski, S. (2002) Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. *J Biol Chem* **277**, 27312-27318
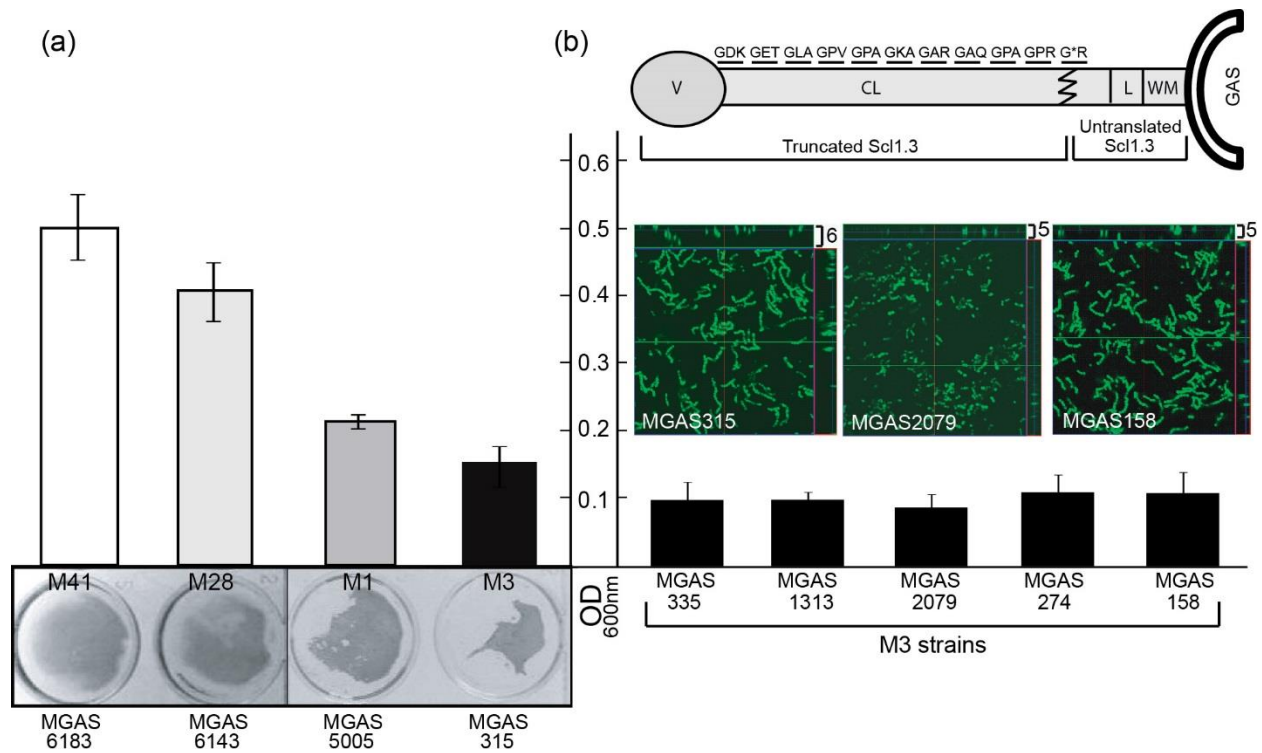
Figure 1

Variation in biofilm formation among GAS strains. (a) Wild type M41-, M28-, M3-, and M1-type GAS strains were grown 24 h under static conditions and analyzed spectrophotometrically following crystal violet staining (top). Visual representation of corresponding wells is shown below. (b) Schematic representation (not to scale) of Scl1.3 protein of M3-type GAS. Translated GXY repeats within the collagen-like (CL) region are shown with an asterisk representing the location of the premature stop codon resulting in a truncated protein. V, variable region; L, linker region; WM, wall-membrane associated region. Below, spectrophotometric measurements of 24-h biofilms following crystal violet staining are graphed for M3-type GAS strains. Absorbance values (OD_{600}) are averages of at least three experiments done in triplicate wells. Corresponding confocal analyses of 24-h biofilms of MGAS315, MGAS2079, and MGAS158 are shown. Images are X-Y orthogonal Z-stack views and average vertical thickness is indicated in micrometers (top right).

37
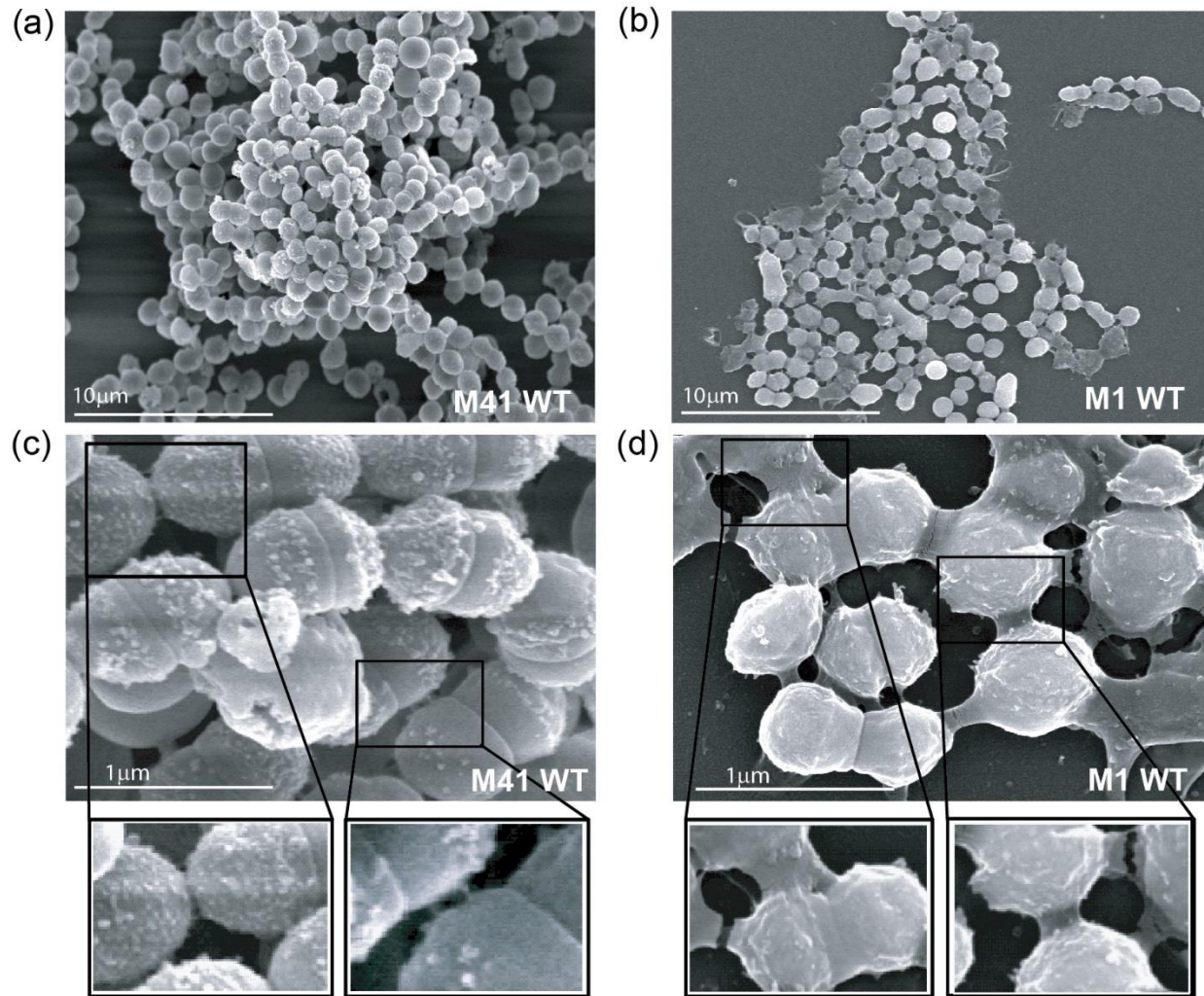
Figure 2

Field emission scanning electron microscopy of GAS biofilms. 24-h biofilms of the M1- and M41-type GAS strains were grown on glass cover slips and analyzed by FESEM. (a-b) Architecture of GAS microcolonies shown at low magnification. (c-d) Cell surface morphology and cell-to-cell junctions observed at higher magnification. Enlargements of cell-to-cell junctions are shown below.
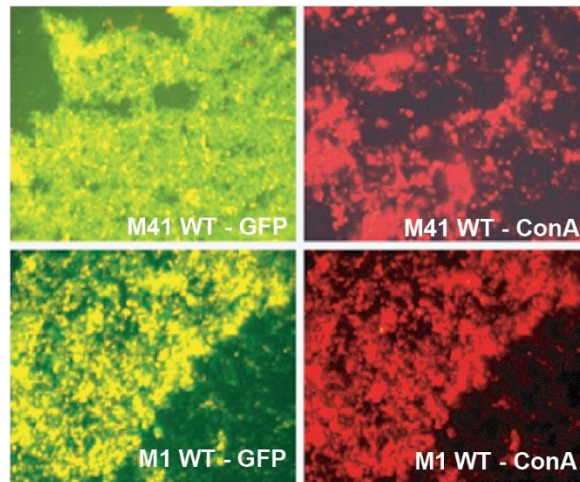
Figure 3

Production of bacterial-associated extracellular matrix. GFP-expressing wild type (WT) M41- and M1-type GAS strains were grown on glass cover slips for 24 h and stained with TRITC-conjugated concanavalin A (ConA). Confocal laser scanning microscopic (CLSM) images were separated to represent green GFP-expressing GAS cells (left images) and red ConA-TRITC staining (right images) for detection of extracellular matrix associated with each strain. Images are from one representative experiment.
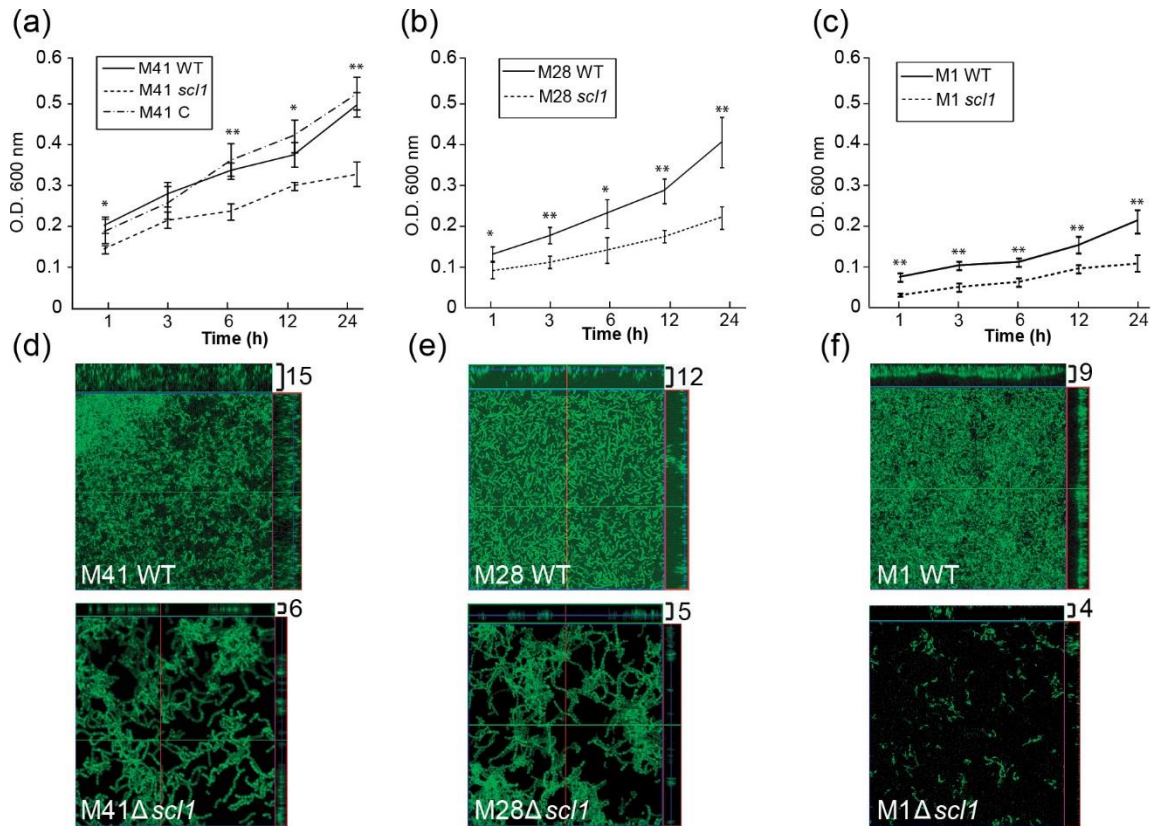
Figure 4

Biofilm formation by wild type and *scl1*-inactivated isogenic mutants. Crystal violet staining and confocal laser scanning microscopy (CLSM) of the GFP-expressing GAS were used to compare biofilm formation by GAS strains. Wild type (WT) M41-, M28-, and M1-type strains, *scl1*-inactivated mutants (*scl1*), and M41 mutant complemented for Scl1.41 expression (M41 C) were used. (a-c) Isogenic GAS strains were grown under static conditions for 24 h and bacterial biomass was detected spectrophotometrically at indicated time points following crystal violet staining. Absorbance values at OD$_{600nm}$ are representative of at least three experiments performed in quadruplicate. Statistical significance is denoted as *$P \leq$ 0.05 and **$P \leq$ 0.001. (d-f) CLSM analysis of corresponding 24 h biofilms from same experiment. Images are X-Y orthogonal Z-stack views of WT (top) and mutant (bottom) GAS strains. Views are representative of ten images within a single experiment. Average vertical biofilm thickness is indicated in micrometers (top right).
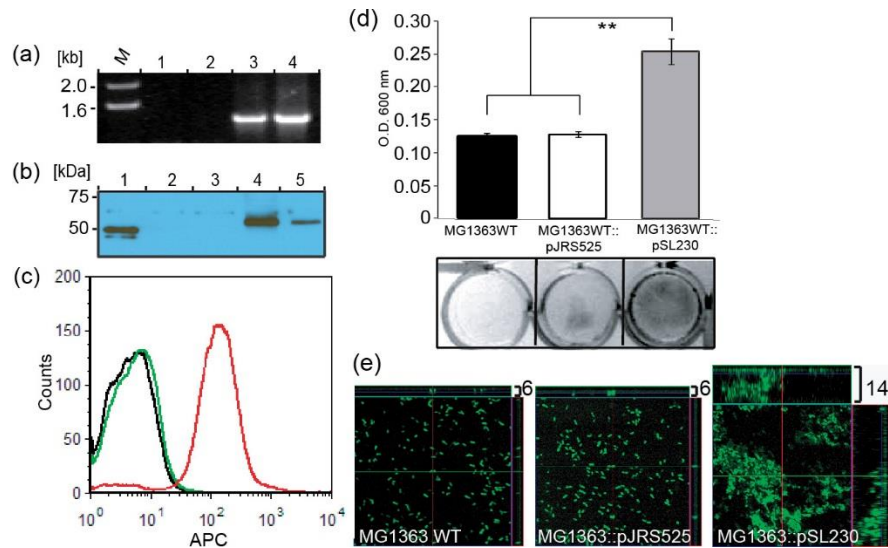
Figure 5

Scl1 expression in *L. lactis* promotes biofilm formation. *L. lactis* was transformed with the plasmid construct pSL230 to express Scl1.41 surface protein or with pJRS525 vector. (a) PCR analysis of *L. lactis* transformants using *scl1.41*-gene-specific primers; lanes: (1) MG1363 wild-type (WT) cells; (2) MG1363::pJRS525 vector-only control; (3) MG1363::pSL230 transformant; (4) control pSL230 plasmid DNA. (b) Scl1.41 expression by western blot analysis of cell-wall extracts prepared from transformed *L. lactis* and control GAS strains using anti- P176 (rScl1.41) antibodies; lanes: (1) purified recombinant P176 protein (truncated Scl1.41); (2) MG1363 WT strain; (3) MG1363::pJRS525 vector; (4) MG1363::pSL230 transformant; (5) MGAS6183 (M41) control. (c) Analysis of Sc1.41 expression by flow cytometry with anti-P176 (rScl1.41) rabbit polyclonal antibodies on the surface of MGAS1363 WT strain (black trace), MGAS1363::pJRS525 vector-only control (green trace) and MG1363:pSL230 transformant (red trace). (d) Crystal violet staining of 24 h biofilms formed by *L. lactis* WT strain*,* MG1363::pJRS525 vector-only control or MG1363::pSL230 transformant (top) with visual representation of the corresponding wells (bottom). Statistical significance is denoted as **P ≤ 0.001. (e) CLSM analysis of 24 h biofilms from same experiment shown in (d). Images are X-Y orthogonal Z-stack views representative of ten images within a single experiment. Average vertical biofilm thickness is indicated in micrometers (top right).
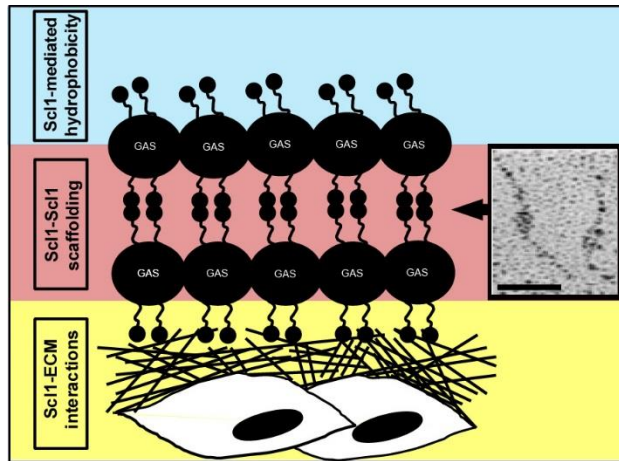
Figure 6

Scl1-mediated model of GAS biofilm (not to scale). Scl1 hydrophobin (current study) initiates bacterial adhesion to animate surfaces (64) within the host (blue field). Scl1 adhesin anchors the growing microcolony by direct binding to tissue extracellular matrix (ECM) components, cellular fibronectin and laminin (19), initiating microcolony formation and anchoring the outside edge of GAS microcolony in tissue (yellow field). Microcolony scaffolding is stabilized by the formation of head-to-head dimers between Scl1 molecules on adjacent chains (pink field). Inset shows Scl1-Scl1 head-to-head dimers formed by rScl1.1 as viewed by electron microscopy after rotary shadowing (68). *Bar*: 50 nm.

Table 1. Cell surface hydrophobicity of GAS strains

GAS Strain	M-Type	Actual Value[†]	Hydrophobicity Index[‡]
MGAS6183 WT	M41	92.6 ± .86	100
MGAS6183 Δ*scl1*	M41	85.2 ± 2.2	**92
MGAS6183 Δ*scl1-C*	M41	98.0 ± .31	105
MGAS5005 WT	M1	80.3 ± .89	100
MGAS5005 Δ*scl1*	M1	63.3 ± 3.2	**79
MGAS6143 WT	M28	94.3 ± .73	100
MGAS6143 Δ*scl1*	M28	72.6 ± .62	**78

[†] Actual hydrophobicity values were calculated based on hexadecane binding as described in Methods. Values are representative of three separate experiments with ten replicates ±SD.

[‡] Hydrophobicity Index represents the ration of actual hydrophobicity value for each strain to that of the isogenic wild-type (WT) strain multiplied by 100.

** Asterisks denote a statistically significant difference of Δ*scl1* mutants versus WTs at P ≤ 0.001.

Table 2. Cell surface hydrophobicity of *Lactococcus* strains

Lactococcus Strain	Actual Value[†]	Hydrophobicity Index[‡]
L.lactis 1363 WT	59.7 ± 7.2	100
L. lactis 1363::pJRS525	56.6 ± 5.5	98
L. lactis 1363::pSL230	82.0 ± 2.6	**137

[†] Actual hydrophobicity values were calculated based on hexadecane binding as described in Methods. Values are representative of three separate experiments with ten replicates ±SD.

[‡] Hydrophobicity Index represents the ration of actual hydrophobicity value for each strain to that of the isogenic wild-type (WT) strain multiplied by 100.

** Asterisks denote a statistically significant difference of Δ*scl1* mutants versus WTs at P ≤ 0.001.

CHAPTER 2

UNIQUE FOOTPRINT IN THE *SCL1.3* LOCUS AFFECTS ADHESION AND BIOFILM FORMATION OF THE INVASIVE M3-TYPE GROUP A *STREPTOCOCCUS*

Beth A. Bachert, Soo Jeon Choi, Paul R. LaSala, Tiffany Harper, Dudley H. McNitt, Dylan T. Boehm, Clayton C. Caswell, Pawel Ciborowski, Douglas R. Keene, Anthony R. Flores, James M. Musser, Flavia Squeglia, Daniela Marasco, Rita Berisio, and Slawomir Lukomski

ABSTRACT

The streptococcal collagen-like proteins 1 and 2 (Scl1 and Scl2) are major surface adhesins that are ubiquitous among group A *Streptococcus* (GAS). Invasive M3-type strains, however, have evolved two unique conserved features in the *scl1* locus: (i) an IS1548 element insertion in the *scl1* promoter region and (ii) a nonsense mutation within the *scl1* coding sequence. The *scl1* transcript is drastically reduced in M3-type GAS, contrasting with a high transcription level of *scl1* allele in invasive M1-type GAS. This leads to a lack of Scl1 expression in M3 strains. In contrast, while *scl2* transcription and Scl2 production are elevated in M3 strains, M1 GAS lack Scl2 surface expression. M3-type strains were shown to have reduced biofilm formation on inanimate surfaces coated with cellular fibronectin and laminin, and in human skin equivalents. Repair of the nonsense mutation and restoration of Scl1 expression on M3-GAS cells, restores biofilm formation on cellular fibronectin and laminin coatings. Inactivation of *scl1* in biofilm-capable M28 and M41 strains results in larger skin lesions in a mouse model, indicating that lack of Scl1 adhesin promotes bacterial spread over localized infection. These studies suggest the uniquely evolved *scl1* locus in the M3-type strains, which prevents surface expression of the major Scl1 adhesin, contributed to the emergence of the invasive M3-type strains. Furthermore these studies provide insight into the molecular mechanisms mediating colonization, biofilm formation, and pathogenesis of group A streptococci.

INTRODUCTION

Group A *Streptococcus* (GAS) or *Streptococcus pyogenes* is a human-specific Gram-positive pathogen responsible for significant morbidity and mortality worldwide (1,2). The clinical outcomes resulting from GAS infection range from superficial pharyngitis and impetigo to severe life-threatening diseases, such as streptococcal toxic shock syndrome and necrotizing fasciitis, as well as post-infectious sequelae, including rheumatic fever, rheumatic heart disease, and post-streptococcal glomerulonephritis (3). Historically, GAS has been a significant cause of puerperal sepsis, scarlet fever, erysipelas, and pharyngitis (4). GAS strains are epidemiologically subtyped based on nucleotide sequence variation at the 5'-end of the *emm* gene, reflecting differences in the hypervariable N-terminal region of the anti-phagocytic surface protein M. Over 240 M-types have been identified (http://www.cdc.gov/abcs/index.html), and certain M-types have been shown to have nonrandom associations with specific disease outcomes (3,5). Since the 1980's there has been resurgence in invasive GAS diseases in the U.S. and other parts of the world. Numerous epidemiology studies conducted in the U.S. (6-11), Canada (12-15), and Europe (16-19) have found associations between infections with M1- and M3-type strains and invasive diseases. Specifically, M3-type strains have been associated with severe invasive disease (6,19) and fatal outcomes (14,16,17).

These epidemiological observations have spurred significant whole genome sequencing efforts aimed at identification of the underlying genetic basis for virulence in M3-type GAS. The complete annotated genomes of invasive M3 strains MGAS315 and SSI-1 have been reported (20,21). These studies have revealed that differing phage elements, insertion sequences, and the large-scale chromosomal inversion identified in SSI-1, contributed to much of the genetic variation between M3 and other M-types. Acquisition of prophages and SNP's drive the expansion of different M3 subclones during epidemic waves of infection (22,23). Additionally, strains causing pharyngitis and those causing invasive infections are derived from the same pool of M3 strains (24). In addition, genetic variation in virulence regulators, including RopB and CovRS, which are "hot-spots" for the accumulation of mutations, also contributes to the hypervirulence of M3 strains (23-28). The cause of the hyper-invasive phenotype of M3 strains is multifactorial

and involves multiple virulence factors controlled by complex regulatory networks that continually undergo remodeling.

The streptococcal collagen-like proteins 1 and 2, or Scl1 and Scl2 (also known as SclA and SclB), are major GAS surface adhesins known to contribute to pathogenesis. Both Scl proteins contain an N-terminal variable region, followed by a collagen-like region containing Gly-X-Y repeats and a cell-wall-anchored C-terminal region (29-33). Transcription of *scl1* is positively regulated by the multiple gene regulator Mga (30,32,34). Scl1 binds cellular fibronectin and laminin (35), and contributes to the formation of biofilm by strains of multiple M-types (36,37). GAS adherence and biofilm formation is enhanced on extracellular matrix (ECM) coatings and on fibroblast-deposited ECM network (37,38), supporting a role for Scl1 in tissue-microcolony formation described during GAS skin infection (39). While the *scl1* gene has been found in every GAS strain tested (29,32), the *scl1.3*-allele in M3-type strains harbors a null mutation within the coding sequence (29). A rare natural reversion of this polymorphism was identified in a small subset of M3 carrier strains (40).

Scl2-protein expression is regulated during translation by varying numbers of short CAAAA repeats downstream of the start codon that determine whether the protein-coding sequence is in-frame or translation will be prematurely terminated (30,31); genome sequencing indicates majority of the M3-type strain contain in-frame *scl2.3* alleles (18,22). Scl2 has been shown in some strains to bind the human thrombin-activatable fibrinolysis inhibitor (41) and contribute to adherence to fibroblasts (31), although its role in GAS pathogenesis is less understood. Recently, the crystal structure of the Scl2.3 globular domain, which is structurally conserved between Scl1 and Scl2, has been reported (42,43) providing insight into the potential binding interactions with host ligands.

In this study, we show that invasive M3-type GAS harbor two unique conserved features of the *scl1* locus including the IS1548 insertion in the promoter region and the null mutation in the coding sequence, which results in a secreted instead of cell-attached Scl1 protein. We demonstrate significantly different expression patterns of *scl1* and *scl2* in M3-type GAS compared to representative strains of M1-, M28- and M41-types. We demonstrate that the expression of the Scl1 adhesin is deficient in serotype M3 strains, as opposed to M1-, M28-, and M41-type strains. However, the Scl2 protein is upregulated

in M3 strains compared to M1, M28, and M41 GAS. The M3 strains lacked significant biofilms on cellular fibronectin and laminin coatings, compared to M41-type GAS, and did not form tissue microcolonies in a wounded pseudo-organ skin equivalent model of infection. Recombinant Scl1.3 specifically bound cellular fibronectin and laminin, and restoration of surface expression of Scl1.3 conferred significant biofilm formation by M3 strains. Inactivation of Scl1 expression in biofilm-capable M28- and M41-type GAS resulted in larger skin lesions produced by the mutants in a mouse model of subcutaneous infection, supporting a role for Scl1 in maintaining a localized infection. Our model advocates that the lack of surface-associated Scl1 adhesin in M3-type strains causes decreased tissue adherence and decreased capacity for stable microcolony formation, thus, promoting bacterial spread over localized nidus of infection.

MATERIALS AND METHODS

1. Bacterial strains and growth

MGAS315 and MGAS10870 are fully sequenced invasive M3-type strains (20,23). MGAS315 was isolated from a patient with streptococcal toxic syndrome in the 1980's (6) and MGAS10870 was isolated from a patient with soft tissue infection in Ontario in 2002 (23). Additional strains from epidemiologically diverse M-types were used for comparison: MGAS6183 (M41), MGAS5005 (M1), and MGAS6143 (M28). The MGAS6143Δ*scl1*-, MGAS6183Δ*scl1*-, and MGAS10870Δ*scl1*-inactivated mutant strains have been previously described (40,44,45). Group A *Streptococcus* cultures were grown in Todd-Hewitt broth (Becton Dickinson and Co.) supplemented with 0.2% yeast extract (THY medium) and on Brain Heart Infusion agar (Becton Dickinson and Co.) at 37°C in an atmosphere with 5% CO_2. For antibiotic selection, erythromycin (5 µg mL^{-1}), chloramphenicol (10 µg mL^{-1}), kanamycin (200 µg mL^{-1}), and spectinomycin (100 µg mL^{-1}) were added to the medium.

Cloning experiments were performed in XL-1 Blue and TB1 *E. coli* cells, while protein expression experiments were performed in BL21 and TB1 *E. coli* cells grown in Luria-Bertani media (Difco Laboratories) at 37°C. For antibiotic selection, chloramphenicol (10 µg mL^{-1}), kanamycin (50 µg mL^{-1}), spectinomycin (100 µg mL^{-1}), and ampicillin (100 µg mL^{-1}) were added to the medium.

2. PCR assays

2.1. Analytical PCR

2.1.1. Detection of IS element upstream of *scl1* in GAS strains

The presence of IS1548 upstream of *scl1* was determined by PCR with primers IS1548F and Scl1R (Figure 1) using Qiagen Taq DNA polymerase (Qiagen, Germantown, MD) as follows: 95°C, 5 min-[95°C 1 min, 62°C 1 min, 72°C 1 min] x30 cycles- 72°C, 10 min. Sequences of primers used in all PCR assays are listed in Table S1.

2.1.2. PCR amplification of *scl1.3* and *scl2.3* alleles

PCR was performed on genomic DNA isolated from M3 strains with primer pairs 232 Up/ 232 Rev for *scl1.3* amplification and sequencing, and Scl2.3 F/R and SclUp/ SclRev for length polymorphism analysis and sequencing of *scl2.3*, respectively. Amplification was performed using Qiagen Taq DNA polymerase as follows: *scl1.3*: 95°C, 5 min-[95°C 1 min, 55°C 1 min, 72°C 1 min] x30 cycles- 72°C, 10 min; *scl2.3*: 95°C, 5 min-[95°C 1 min, 51°C 1 min, 72°C 1 min 45 sec] x30 cycles- 72°C, 10 min. All PCR products were analyzed in a 1% agarose gel with 1 kb Plus DNA ladder as a molecular size marker (Life Technologies, Grand Island, NY).

2.2. Quantitative reverse transcription PCR (qRT-PCR)

2.2.1. RNA isolation

Total RNA was isolated from the logarithmic phase (OD_{600} 0.5) GAS cultures using RNeasy Protect Bacteria Mini kit (Qiagen), employing an additional pretreatment step with 250 µg lysozyme, 100 µg proteinase K, and 12.5 U mutanolysin per sample to augment cell wall disruption. Trace genomic DNA was removed by incubation with TurboDNase enzyme (Ambion). RNA quality was assessed in 1% agarose gel and spectrophotometrically; A260/280 and A260/230 ratios of >1.8 were considered acceptable. RNA was used immediately in cDNA synthesis or stored at -80°C for no more than 1 week.

2.2.2. Determination of transcription by qRT-PCR

Synthesis of cDNA was performed using iScript Select cDNA synthesis kit (Bio-Rad) with 1 µg of RNA per reaction and random primer mix. For each sample, a no reverse transcriptase control was performed, containing only RNA, reaction buffer, and random primer mix, to ensure the absence of genomic DNA. Incubations were carried out on a Bio-Rad C1000 Touch Thermal cycler: 25°C 5 min, 42°C 30 min, 85°C 5 min, and cDNA was stored at -20°C until used in qRT-PCR. qRT-PCR reactions were performed with 250 nM primers using SsoAdvanced SYBR Green Universal Supermix (Bio-Rad). For each qRT-PCR reaction, template cDNA, diluted 1:20, control genomic DNA, or no reverse transcriptase control were used. PCR was performed using the following conditions on a Bio-Rad CFX96 Thermal cycler: 95°C 2 min- [95°C 5 sec, 60°C 30 sec]x30- [65°C- 95°C, 5 sec, 0.5°C/ step]. Standard curves were generated for each primer set using cDNA from MGAS315 to determine the linear range and estimate reaction efficiency. Gene expression of *scl1*, *scl2, emm,* and *mga* was normalized against the expression of *tufA* gene, which has previously been identified and validated as an appropriate endogenous control (46). The ΔΔCt method was utilized to compare gene expression between MGAS315 and other M3 strains, as well as M1, M28, and M41-type strains. Data was averaged from three independent experiments, each containing three technical replicates. Statistical significance was determined using an unpaired *t*-test.

3. Recombinant Scl (rScl) proteins

3.1. Cloning, expression, and purification of M3-derived rScl proteins

The rScl proteins were generated 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 (47). rScl proteins rScl1.3V and rScl1.3FL are derived from MGAS315, whereas rScl2.3 protein is derived from the M3 strain MGAS3375. The recombinant Scl2.3V protein has been described previously (42). Construct containing the recombinant full-length Scl1.3FL (rScl1.3FL) protein was previously described (40). Briefly, the *scl1.3* coding region from MGAS315 was cloned into the *E. coli*/GAS shuttle vector pJRS525 (48), generating plasmid pSL501. PCR mutagenesis was used to convert the internal TAA stop codon into

a GAA glutamate codon, resulting in plasmid pSL502 with continuous full-length *scl1.3FL* allele. This sequence was subsequently cloned into the pASK-IBA2 expression vector for recombinant Scl1.3FL protein production; each clone was verified by sequencing. Protein expression constructs are listed in Table 1.

Protein expression was induced by the addition of anhydrotetracycline at 0.2 µg mL^{-1} for 3 hours. Cells were centrifuged and resuspended 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. Proteins were dialyzed against 25 mM HEPES, pH 8.0, and analyzed by SDS-PAGE stained with RAPID*stain*^{TM} (G-Biosciences). Protein sequence was confirmed using mass spectrometry (University of Nebraska Medical Center) and N-terminal Edman Degradation sequencing (Iowa State University).

3.2. Electron microscopy of rotary shadowed rScl proteins

Electron microscopy visualization of the rotary shadowed preparations of rScl1.3FL and rScl2.3 was used to assess the domain organization of rScl proteins, as employed previously (49). The rScl proteins were dialyzed against 0.1 M ammonium bicarbonate and mixed with glycerol to a final concentration of 70% (vol:vol). Samples were nebulized onto mica chips with an airbrush and rotary-shadowed with carbon/platinum using an electron beam gun. Photomicrograph images were acquired using a transmission electron microscope FEI G2 operated at 80KV.

3.3. Mass spectrometry analysis

Samples from in-gel trypsin digested proteins were cleaned using a Millipore µC18 ZipTip, then resuspended in 0.1% formic acid. Samples were fractionated on a Eksigent cHiPLC column (75 µm x 15 cm ChromXP C18-CL 3 µm 120 Å), and resulting peptides were sequenced using 5600 TripleTOF (typical gradient 2-60% ACN in 60 minutes). Peptides identified were searched against the NCBI protein database with Protein Pilot software employing the following settings: search effort, thorough; taxonomy, none. Positive identification was considered as the identification of two or more unique peptides at high

confidence ≥95%, FDR=0.05, 0.01, or 0.001, which matched the same protein entry in the database searched.

3.4. Binding of rScl proteins to extracellular matrix proteins and synthetic peptide

3.4.1. ELISA binding assays

Binding of rScl proteins to extracellular matrix proteins, cellular fibronectin (cFn) and laminin (Lm), was tested by ELISA (35). rScl proteins were immobilized onto *Strep*-Tactin® coated microplate wells (IBA GmbH) at 0.5 µM at room temperature for 1.5 hours, then blocked with Tris-buffered saline (TBS) containing 1% bovine serum albumin (BSA) overnight at 4°C. The cellular fibronectin from human foreskin fibroblasts (Sigma) and murine laminin (Invitrogen) were added to the wells at 1 µg per well and incubated at room temperature for 1 hour. Bound ECM proteins were detected with rabbit anti-laminin at 1:100 (Sigma) and anti- human fibronectin at 1:4,000 (Sigma) polyclonal antibodies. Secondary antibody goat anti-rabbit IgG (H+L) HRP conjugate (Bio-Rad) was next used with 1-step ABTS substrate (2,2'-Azinobis [3-ethylbenzothiazoline-6-sulfonic acid]-diammonium salt) (ThermoScientific). Absorption was measured using Spectramax 190 at a wavelength of 415 nm. Statistical analysis is based on three independent experiments each containing three technical replicates, using an unpaired *t*-test.

  In antibody inhibition assay, the IST-9 mAb targeting the C-C' loop of EDA domain was utilized (37). cFn was either untreated or pre-treated with increasing concentrations of IST-9 blocking mAb (0.1 µg- 1.0 µg), and added to *Strep*-Tactin®-coated microplate wells immobilized with rScl proteins, then incubated for 1 hour. Bound ECM proteins were detected as above.

3.4.2. Design and synthesis of the C-C' cyclic peptide

The C-C' cyclic peptide was designed based on the crystal structure of the EDA domain of cFn (PDB code 1J8K). In particular, the region, which was reported to be involved in Scl1 binding, connecting to two β-strands C and C' of EDA was elongated to the whole C-C' β-hairpin by allowing the formation the electrostatic interaction between Arg33 and Glu45. The terminal Tyr32 and Pro48 were mutated to Cys to stabilize the β-hairpin by introducing a disulfide bond (**Figure S4**). The obtained sequence,

CRVTYSSPEDGIHELF**C** (molecular weight: 1997.1 Da), endowed with a cyclic structure to mimic the structure of this region in EDA, was acetylated and amidated at the N- and C-terminus, respectively. Synthesis of the designed peptide was performed employing the solid phase method on a 50 µmol scale initially following standard Fmoc strategies (50). Due to aspartimide formation during traditional acylation reactions, peptide synthesis was carried out employing microwave technology (51). Cyclization was achieved by treating the peptide at 0.1 mg/mL (to avoid intermolecular disulphide formation) with buffer carbonate 50 mM, pH=9, overnight. The peptide was purified by RP-HPLC and the identity and purity (> 97%) was assessed by LC-MS (data not shown).

3.4.3. Surface plasmon resonance (SPR) experiments

Real time binding assays were performed at 25 °C on a Biacore 3000 Surface Plasmon Resonance (SPR) instrument (GE Healthcare). For immobilization, rScl1.3FL protein containing the C-terminal *Strep*-tag II was injected at a concentration of 40 µM on streptavidin-coated sensor chip, SA Biacore, until the desired level of immobilization was achieved (averaged value of 100 RU). Binding assays were carried out by injecting the C-C' cyclic peptide at concentrations ranging between 10-500 µM. Experiments were carried out in HBS buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, pH 7.4). The association phase (k_{on}) was followed for 270 s, whereas the dissociation phase (k_{off}) was followed for 300 s. The reference chip sensorgrams were subtracted from sample sensorgrams. Experiments were carried out in duplicates. Kinetic parameters were estimated assuming a 1:1 Langmuir binding model and using version 4.1 Evaluation Software (GE Healthcare).

3.4.4. Fluorescence binding analysis

rScl1.3FL protein, at a concentration of 30 µM, was incubated with increasing concentrations of C-C' cyclic peptide (0 - 300 µM) at 25.0 °C, using an excitation wavelength of 298.0 nm and a fluorescence emission wavelength ranging from 300 to 400 nm. The acquisition parameters were set as follows: excitation and emission slits at 5 nm; 120 nm/min scan rate; 1.00 nm data interval averaging time at 0.500 s, PMT voltage at "high". Fluorescence values were recorded at 333 nm, and subtracted from the

fluorescence intensity of the ligand-free protein, generating -Δfluorescence. The -Δfluorescence values were plotted against the peptide concentration. (52). Experiments were carried out in duplicates. A control assay was carried out employing the buffer as titrant to assess that the dilution effect was under 3%, not affecting the results.

4. Complementation of GAS strains with full-length Scl1.3FL

For *trans*-complementation experiments, plasmid pSL502, encoding full-length cell-associated Scl1.3FL protein, was electroporated into MGAS315 WT and MGAS6183Δ*scl1* electrocompetent cells. The pJRS525 vector was electroporated as a control. Transformants were selected on BHI agar containing 100 µg mL^{-1} spectinomycin, and plasmids were re-sequenced. For MGAS10870Δ*scl1*, which contains a spectinomycin resistance cassette in place of the *scl1.3* allele, the spectinomycin resistance marker in pSL502 was replaced with a kanamycin resistance, generating the plasmid pSL518 and colonies were selected on BHI agar containing 150 µg mL^{-1} kanamycin.

4.1. Determination of Scl1.3- and Scl2.3-protein expression in wild-type and complemented GAS strains

4.1.1. Western blot analysis

Expression of the Scl1.3 and Scl2.3 proteins was determined by western immunoblotting of the bacterial cell wall and culture supernatant protein fractions, as described before (29,30). Briefly, bacterial cultures were grown to an OD$_{600}$ of 0.5 and cells were harvested by centrifugation. Culture supernatant proteins were precipitated with trichloroacetic acid (Sigma) to a final concentration of 10% (vol:vol). The cell wall protein fraction was obtained after cell digestion with lysozyme and mutanolysin in a high sucrose buffer. A total of 10 µg of protein samples were separated by SDS-PAGE and transferred to a nitrocellulose membrane. Detection of Scl1.3, Scl2.3, and M3 proteins was performed using the same sample preparations with rabbit polyclonal antibodies generated against the truncated rScl1.3WT protein (anti-Scl1 1:15,000 dilution, reported in (40)) and the rScl2.3V region (anti-rScl2.3V 1:2,500; generated by Proteintech (43)). Horseradish peroxidase-conjugated goat anti-rabbit IgG (H+L) secondary antibody (Bio-Rad),

combined with Pierce™ ECL western blotting substrate (Thermo Scientific) was used for detection. Images were acquired using a ChemiDoc Touch Imaging System (Bio-Rad).

4.1.2. Flow cytometry analysis

Surface detection of Scl1.3 and Scl2.3 proteins was determined by flow cytometry. Bacteria grown to an OD_{600} of 0.5 were harvested by centrifugation, and washed with flow cytometry buffer (sterile phosphate-buffered saline containing 10% Todd-Hewitt broth supplemented with 0.2% yeast extract). Cells were incubated with polyclonal antibodies against Scl1.3 and Scl2.3 described above at a dilution of 1:100 for 30 minutes on ice, then washed and incubated with Allophycocyanin (APC)-conjugated donkey anti-rabbit IgG (H+L) (Jackson ImmunoResearch) for 30 minutes on ice. Cells were washed and fixed in 0.4% paraformaldehyde, and stored at 4°C until analysis. Before analysis, cells were washed twice and resuspended in flow cytometry buffer. Cells were analyzed using a BD LSRFortessa, and 50,000 events were collected per sample. Data was analyzed using the FCS Express Flow 5 software.

5. Assessment of biofilm formation

5.1. Crystal violet staining assay

Wild-type, mutant, and complemented strains were grown to OD_{600} of 0.5 and seeded into 24-well culture plates coated with ECM at 2 µg per well, then incubated at 37°C with 5% CO_2 for 24 hours. Wells were washed with PBS followed by the addition of 0.5 mL 1% crystal violet solution (Fisher Scientific) diluted in PBS and incubation at room temperature for 30 minutes. Wells were rinsed twice with PBS and stain was solubilized with 0.5 mL of 75% ethanol. Spectrophotometric readings were taken for each sample at OD_{600}. Statistical analysis is shown based on three independent experiments, each containing three technical replicates, using an unpaired *t*-test.

5.2. Confocal laser scanning microscopy (CLSM)

To visualize GAS by CLSM, bacterial cells were transformed with a GFP-encoding plasmid pSB027 (53), as before (36). 15-mm glass cover slips were placed into 24-well tissue culture plate wells and coated with 2 µg of cFn or Lm per well. Bacterial cultures

grown to logarithmic-phase were added to the wells and allowed to form biofilms for 24 hours. Wells were rinsed with PBS and bacterial cells were fixed with 3% paraformaldehyde at room temperature for 30 min. Wells were washed again and coverslips were mounted onto slides using ProLong Gold Antifade Mountant (Thermo Scientific). Confocal images were acquired using a 63×/1.40 Plan-Apochromat objective and a Zeiss LSM 510 laser scanning confocal microscope.

6. *In vitro* and *in vivo* GAS infection models
6.1. GAS infection of *in vitro* cultured human skin equivalents

Wounded full-thickness skin equivalents, EpiDerm-FT (MatTek, Boston, MA) were used. The tissues are discs 8 mm in diameter, which are provided in transwells. A 3-mm wound is generated by performing a punch biopsy to remove the keratinocyte layer. Immediately upon arrival, tissues were equilibrated in antibiotic-free manufacturer's medium overnight at 37°C in atmosphere with 5% CO_2. Wounds were infected with 10 µL of GFP-expressing log-phase group A streptococcal inocula and incubated in a humid environment at 37°C with 5% CO_2 in daily-fresh media; a total of 4 experiments were performed and variables included the inoculum size of 3×10^6 -1.8×10^7 CFU and collection time points between 1-5 days. Tissues designated for histopathological evaluation were fixed in 10% formalin, whereas tissues for two-photon fluorescence (TPF) microscopy imaging were fixed in 4% paraformaldehyde for several hours and then transferred to petri dishes containing PBS. For visualization of glycocalyx produced by GAS strains, tissues were permeabilized with 0.1% Triton X-100 in 1 x PBS and blocked with 0.05% Triton X-100 in 1x PBS with 1% BSA before staining with concanavalin A- tetramethylrhodamine (Molecular Probes). Tissues were then rinsed and stored in PBS at 4°C until TPF imaging was performed. For TPF analysis, an Olympus 60x/ 1.2NA water dipping objective was used. Fixed tissues were imaged by two-photon microscopy with the Ti:sapphire laser (Mira, Coherent) intensity at 60 mW and input wavelength of 850 nm. Laser scanning images were collected at 0.5- 1 µm incremental depths using ScanImage (Janelia Farms, HHMI). Images were saved in a single TIFF file with 16 bit depth. Deconvolution of the images was performed using AutoQuant x3 and 3D models of z-stacks were generated using Imaris software.

6.2. Mouse model of soft tissue infection

Animal experiments were conducted in compliance with the regulations and standards under the Animal Welfare Act, the Public Health Service Policy on Humane Care and Use of Laboratory Animals, and the Guide for the Care and Use of Laboratory Animals. The protocol was approved by the West Virginia University Institutional Animal Care and Use Committee (IACUC).

Subcutaneous infections of mice were carried out as described previously (54). Briefly, 5-week-old male, immunocompetent, hairless mice (strain Crl:SKH1-*hr*BR) were used (Charles River, Wilmington, MA). Groups of 10-15 mice anesthetized with isoflurane were infected subcutaneously at the right flank with ~10^9 GAS CFU of WT or *scl1*-mutant strains, and mice were observed for 14 days. The weight and abscess dimensions (length [L] and width [W]) of each mouse were recorded daily during the first week and every other day thereafter. To analyze differences between mice infected with WT and *scl1*-mutant GAS, the area of each abscess was calculated with the equation for the area (A) of an spherical ellipsoid: A=π(L/2)x(W/2); statistical differences were calculated using the student's *t*-test. At the conclusion of the experiments, mice were anesthetized and sacrificed by cervical dislocation.

RESULTS

M3-type GAS contain unique insertion of IS1548 element and nonsense mutation within the *scl1.3* locus

Since the resurgence of invasive GAS infections in the 1980s and an advent of molecular epidemiology fostered by large-scale sequencing, significant efforts have been made to define the molecular basis for the invasive phenotype of M3-type strains. In parallel to these advances, we identified two unique genomic traits in the *scl1.3* locus, encoding streptococcal collagen-like protein 1: (i) the presence of an insertional sequence element, IS1548, in the promoter region and (ii) the presence of a nonsense mutation within the coding sequence of *scl1.3* allele that was absent in other M-types.

Studies employing *mga*-inactivated mutants have indicated that Scl1 expression was positively regulated by the GAS-global transcriptional regulator, Mga (30,32). Two

putative Mga binding sites were identified upstream of the *scl1* coding sequence in the M1-type strain SF370 and experiments demonstrated that the Mga binding site distal to *scl1* was responsible for transcription activation (34). Identical Mga binding sites I and II were also identified upstream of *scl1.3* in the sequenced M3-type strain MGAS315 with IS1548 element inserted 38 bp upstream of the distal Mga I-binding site (Figure 1A). To determine if the IS1548 insertion was specific to M3 strains, we BLAST-searched this element in 45 completed GAS genomes representing 21 different M-types. We observed the presence of IS1548 in all strains searched with varying locations and occurrences from one to twelve per genome (Figure 1B). However, the IS1548 insertion upstream of the *scl1.3* allele was only found in the sequenced genomes of M3-type strains MGAS315, SSI-1, and M3-b. Interestingly, a complete IS1548 element was not present upstream of *scl1.3* in the recently reported genome of the M3 strain STAB902, which represents a non-invasive isolate (55); instead, a 34-bp remnant of IS1548, including the inverted repeat and additional 14 bp, was found. Based on this bioinformatics data, we examined the presence of the IS1548 element upstream of *scl1* by PCR in a panel of 40 M3-type strains, using primers located in the IS1548 and *scl1.3* sequences (IS1548F and Scl1R, Table S1). All M3-type strains examined were positive for the IS1548-*scl1.3* amplicon, while M1, M41, and M28-type controls were negative (Figure 1C), demonstrating a broad and conserved presence of the IS1548 insertion in this location among M3-type GAS.

The nonsense mutation in the 11th Gly-X-Y repeat in the Scl1.3-CL collagenous region (Figure 1A, red box), presumably results in a truncated secreted variant, which consists of the Scl1-V variable region and 10 Gly-X-Y repeats but lacks the cell wall anchor. This polymorphism was originally identified in five M3-type invasive GAS strains but was absent in 45 other strains analyzed, representing 20 different M-types (29). It was later shown in 98.7% of 479 sequenced invasive M3-type isolates (40). We performed targeted PCR amplification of *scl1.3* from 46 additional M3-type strains and observed identical amplicon sizes in all of them (Figure S1A). Sequencing showed that all tested strains harbored an identical *scl1.3* allele, containing 25 Gly-X-Y repeats in the collagenous domain with the null mutation in the 11th repeat (Data Set S1). A complete lack of genetic variation within *scl1.3* is surprising and differs from the length variation that is commonly observed among *scl* alleles from other M-types (29,32,33). These results demonstrate

that the IS1548 insertion and null mutation we identified in the *scl1.3* locus are unique to and conserved among M3-type GAS. Based on these results, we hypothesized M3 strains produce a truncated, but potentially biologically active Scl1.3 variant, which is secreted instead of being cell-attached.

scl1.3 expression is diminished in M3-type GAS

Original reports showed *scl1* transcripts in northern blots, as well as full-length Scl1 proteins (both cell-associated and cell-free fractions) in western blots for strains of *emm* types 1, 28, 52, and 41 (29,30,32,45). To analyze the expression of the truncated Scl1.3 protein in M3-type GAS, western blot analysis was performed on cell-wall (CW samples) and culture-supernatant (Sup samples) protein fractions of several M3-type strains grown to exponential phase (Figure 2A). The expected truncated Scl1.3 protein was not detected by anti-Scl1 antibodies, whereas the rScl1.3V positive control, corresponding to the V region of Scl1.3 variant, produced the expected immunoreactive band of ~8.3 kDa. In an additional control experiment, the same panel of M3 strains was tested for Scl1.3 on the cell surface using flow cytometry (Figure 2B). No shift in median fluorescence intensity was observed in M3 strains incubated with anti-Scl1 antibody compared to a secondary-only antibody control, indicating a lack of Scl1 on the cell surface among M3 strains.

Given the unique IS1548-*scl1.3* location and the lack of truncated Scl1.3 products in culture supernatants, we investigated *scl1.3* expression by qRT-PCR. Total RNA was isolated from exponential phase cultures of 5 M3-type strains, as well as from previously characterized control strains of *emm* types 1, 28, and 41 (29,30,32,45). Expression of *scl1* from each strain was compared to *scl1.3* transcription in M3 strain MGAS315 (Figure 2C). Each non-M3-type strain tested had significantly higher transcription level of respective *scl1* allele compared to MGAS315. The M1 strain had the most increased expression by 21-fold, while M28 and M41 strains exhibited 6-fold and 8-fold higher *scl1* expression, respectively. Interestingly, the *scl1.3* transcripts examined in four additional M3-type MGAS strains 158, 335, 1313, and 10870 were significantly reduced as compared to MGAS315, with a range of 25-45-fold decrease. Overall, we observed that M3-type GAS harboring the IS element upstream of *scl1.3* have drastically decreased

scl1.3 transcript levels and lack the Scl1 protein product, as assessed by western blot and flow cytometry.

Since *scl1* is regulated by the transcriptional activator Mga, we investigated whether the decreased *scl1.3* transcripts in M3 strains were due to lack of *mga* gene expression in these strains. In parallel, we assessed transcript levels of the *emm* gene encoding the M surface protein, a key virulence factor regulated by Mga. For comparison, we included the M1 strain, which had significantly increased *scl1* expression compared to the M3 strains. With the exception of MGAS335, which had significantly downregulated *mga* and *emm* expression, we found no significant differences in either *mga* or *emm* gene expression between MGAS315 and the other M3 strains or M1 strain (Figure 2D). Furthermore, M3 protein was highly expressed, as it was found in both cell wall and supernatant fractions, except for the MGAS335 strain, consistent with transcription data (Figure 2E). These results demonstrate that the striking downregulation of *scl1.3* in M3-type GAS is not due to decreased *mga* expression or non-functional Mga protein, as *emm* is normally expressed in these strains.

scl2.3 is expressed in M3-type GAS

Scl2 shares a similar structure with Scl1 but its biological function is poorly understood. One study demonstrated in a different M-type background that isogenic mutant devoid of Scl2.55 variant had lower adhesion to human skin fibroblasts (Rasmussen and Björck, 2001); however, the Scl2.3 variant present in M3-type GAS has not been investigated for expression and ECM binding. Therefore, we next assessed *scl2.3/*Scl2.3 expression in M3 strains. PCR amplification and sequencing showed that the majority of M3 strains contained in-frame *scl2.3* allele (Figure S1B, Table S2).

Western blot analysis of cell wall (CW samples) and supernatant (Sup) protein fractions found Scl2.3 protein was expressed by M3 strains MGAS315, 10870, 158, 9517 and 1313, whereas samples obtained from strain MGAS335, which contains an out-of-frame *scl2.3* allele, generated no immunoreactive band; rScl2.3V control produced the expected 10.1-kDa band (Figure 3A). Mass spectrometry confirmed the identity of the presumed immunoreactive Scl2.3-protein band from MGAS315 (Table S3). Consistently, Scl2.3 was detected on the surface of all five M3 strains containing in-frame *scl2.3* alleles

by flow cytometry, with a positive shift in median fluorescence intensity ranging in 67-131-fold change, as compared to the secondary-only antibody control (Figure 3B). We next compared the *scl2*-transcription level in M3 strain MGAS315 with *scl2*-transcription levels in M1-, M28-, and M41-type strains. In striking contrast to the pattern of *scl1.3* transcription, *scl2* transcripts were significantly decreased in the M1-type strain by 13-fold, as well as in the M28 (6-fold) and M41 (3-fold) strains (Figure 3C). Additionally, there was no significant difference between *scl2* expression in MGAS315 and the M3 strains MGAS158, 335, 1313, and 10870 (Table S4). These results confirm that the M3 *scl2.3* allele is transcribed at high levels, resulting in considerable expression of the Scl2.3 protein. These results suggested that Scl2 has an important biological function in M3 strains and warranted subsequent experiments assessing Scl2.3 function.

GAS infection disseminates through human tissue and inhibits wound healing

A wounded human skin equivalent, devoid of an inflammatory component, was utilized as a "mechanistic model" of GAS tissue colonization. The epidermal wound of each skin equivalent was infected with GFP-expressing M3-type invasive strain MGAS315 or M41-type non-invasive strain MGAS6183, and analyzed after 1-5 days by standard histopathology (H&E and Gram's stain) and using two-photon fluorescence (TPF) microscopy. H&E of uninfected tissue controls harvested at day 0 showed the absence of a keratinocyte layer where the punch biopsy was performed (Figure 4A). Complete healing of the wound was observed after 5 days, with a newly-generated intact keratinocyte layer covering the punch biopsy site (Figure 4B). In contrast, tissue infected with either GAS strain exhibited delayed wound closure as late as day 5 post-infection (Figure 4E, H). H&E staining of skin equivalents infected with either M3 or M41 after 24 hours revealed bacterial colonization of the exposed dermal surface, as well as bacterial invasion into puncture-associated defects extending deep into the dermal layer, largely located at the wound edges (Figure 4D, G). Notably, extensive bacterial growth and spread was largely confined to tissue crevices in the dermis, whereas the presence of large bacterial colonies directly below the wound bed was rarely seen on microscopic examination. By day 5 of infection, bacterial invasion via these dermal defects extended to the bottom of the dermis for both strains, presenting both vertical and lateral spread of

bacteria (Figure 4E, H). Gram stain of infected tissues showed the formation of superficial colonies near dermal surfaces, as well as biofilm formation on the surface of exposed dermis (Figure 4J-L). Additionally, epidermal tissue neighboring the wound bed exhibited ~60% decreased thickness of the viable keratinocyte layer in tissues infected with MGAS315 and MGAS6183 by day 5, relative to uninfected tissues (Figure 4C, F, I). This suggests bacterial infection of the wound affects epidermal cells distant from the site of infection in this model.

TPF analysis was then performed on whole infected skin equivalents on day 5 post-infection in order to assess bacterial spread directly below the wound bed. This method allowed us to visualize bacterial structures within tissue that were not apparent in H&E or Gram-stained sections. Tissue microcolonies were observed in samples infected with M41 GAS (Figure 4N), whereas M3 cells had a scattered appearance (Figure 4P). TRITC-concanavalin A (TRITC-conA) was utilized to visualize glycocalyx associated with bacteria. TRITC-conA stain colocalized with bacterial microcolonies formed by the M41 GAS (Figure 4O, Figure S3), indicating microcolonies were encased in a glycocalyx, much like a classic biofilm. In contrast, TRITC-conA stain was associated with scattered M3-GAS chains located at the bottom cell layer in MGAS315-infected skin equivalents (Figure 4Q). These results indicate that M41 GAS, but not the M3 GAS, forms microcolonies in the tissue during a human skin infection that are encased in a glycocalyx, consistent with the observation that M41, but not M3, GAS forms robust biofilm *in vitro* on ECM coatings. In addition, M3 GAS disseminates in a form of scattered chains or single cells through the tissue.

M3-type GAS strains form poor biofilms on extracellular matrix coatings

It was previously reported that M3-type strains have no substantial biofilm formation *in vitro* on an inanimate surface (36). Here, we tested biofilm formation on cellular fibronectin (cFn) and laminin (Lm) coatings by a panel of representative M3 strains isolated from invasive cases of GAS disease, as compared to the non-invasive biofilm-capable M41-type model strain MGAS6183. As expected, wells coated with either cFn or Lm supported robust biofilm formation by the M41 strain, whereas significantly less bacterial biomass was measured for all M3-type GAS on both ECM coatings (Figure 5A). There was also

no correlation between Scl2 expression and biofilm formation. We hypothesized that M3-type GAS, devoid of Scl1 adhesin, have decreased binding to cFn and Lm ECM components, thus, preventing the formation of tissue microcolonies, and that restoration of full-length Scl1.3 (Scl1.3FL) on the GAS cell surface will confer binding to host ECM, as well as biofilm capacity *in vivo* (Figure 5B).

Full-length recombinant Scl1.3 binds cellular fibronectin and laminin

To test this hypothesis, we: (i) constructed Scl1.3- and Scl2.3-derived recombinant proteins, (ii) characterized their structural organization, and (iii) assessed their ECM-binding capacities.

First, rScl1.3FL and rScl2.3 proteins were assessed for purity and integrity by SDS-PAGE (Figure 6A). The expected 20.6-kDa rScl1.3FL migrated at ~34 kDa, which is consistent with previous reports of aberrant migration of rScl proteins (29,30), whereas the rScl2.3 protein migrated according to the expected molecular mass of 16.2 kDa; both proteins were verified by mass spectrometry (Table S3). Rotary shadowed rScl1.3FL and rScl2.3 constructs, exhibited the characteristic lollipop-like structural organization (Figure S2), as observed for previously characterized rScl proteins (47,49). Interestingly, rScl1.3FL formed aggregates that were mediated by the intermolecular interactions between the globular domains; such interactions, however, were not observed between rScl2.3 molecules. The appearance of Scl1-Scl1 aggregates implies an attractive hypothesis that V-to-V-region interactions between the Scl1 molecules, but not between the Scl2 molecules, on the surface of neighboring GAS chains may support biofilm structure, as proposed in our model (Figure 5B).

It has been established that Scl1 proteins selectively bind cellular, but not plasma, fibronectin and laminin (35), and that selective cFn binding is achieved by a unique binding mechanism involving the C-C' loop of the extra domain A in cFn (EDA-cFn) (37). Here, we pre-incubated cFn with increasing concentrations of EDA-blocking IST-9 mAb (0.1, 1.0 µg), then added to wells containing immobilized rScl proteins and allowed for binding. We observed significant dose-dependent inhibition of cFn binding to rScl1.3FL, with a 36% inhibition at 0.1 µg and a 60% inhibition at 1.0 µg of IST-9 (Figure 6B). In addition, the rScl2.3 protein did not bind cFn, which is consistent with our previous data

employing several different rScl2 proteins (35). Surface plasmon resonance (SPR) measurements of binding affinity between rScl1.3FL and EDA-derived C-C' cyclic peptide provided a dissociation constant of K_D= 63.3 µM (Figure 6C). To corroborate the rScl1.3FL-EDA binding affinity, an in-solution fluorescence binding assay was performed, in which the variation in tryptophan fluorescence of rScl1.3FL was recorded as a measure of positive binding to the C-C' cyclic peptide. Tryptophan fluorescence emission at 333 nm showed a dose-response quenching upon addition of the C-C' cyclic peptide and -Δ fluorescence intensity was plotted against peptide concentration (Figure 6D). Data were fitted with a 1:1 model of interaction, providing a K_D = 44.54 ±9 µM, in agreement with SPR data (56).

We next assessed binding of rScl1.3FL and rScl2.3 proteins to laminin (Lm) by ELISA. We determined rScl1.3FL had significant Lm binding, whereas rScl2.3 had not (Figure 6E), consistent with previous findings that Scl1-derived recombinant proteins, but not the Scl2-derived, bind ECM proteins (35). Collectively, these results demonstrate specific binding of Scl1.3FL to the EDA domain of cFn and to Lm and its capacity of being surface adhesin.

Homologous complementation of M3 strains with full-length surface-exposed Scl1.3 adhesin confers biofilm formation on ECM

To assess the effect of cell-surface Scl1.3FL expression on the capacity to form biofilm, an *in-trans* complementation of two representative invasive M3 strains was performed, MGAS315 wild-type (WT) strain, naturally lacking Scl1.3 expression, and a previously generated *scl1*-inactivated mutant of MGAS10870 (10870Δ*scl1*), with plasmids pSL502 (Sp^R) and pSL518 (Km^R), respectively, both encoding the full-length Scl1.3FL protein. As a control, MGAS315 was complemented with a shuttle vector pJRS525. The cell wall-associated expression of Scl1.3FL in complemented M3-type GAS was first tested by western blot analysis of the cell wall protein fractions and on the GAS-cell surface by flow cytometry. An expected ~35-kDa immunoreactive band in complemented strains was observed, which was absent in the parent strains (Figure 7A). Mass spectrometry analysis of the corresponding bands extracted from the gel confirmed they represented the Scl1.3FL protein, with five unique peptides identified with 17% sequence coverage for

complemented MGAS315 and four unique peptides identified with 11% sequence coverage for complemented MGAS10870 (Table S3). A 6.8-fold increase in median fluorescence intensity of Scl1.3FL-complemented MGAS315 cells was measured by flow cytometry, as compared to the vector-complemented and WT control strains (Figure 7B). The Scl1.3FL-complemented 10870Δ$scl1$ exhibited a 2.2 -fold increase in median fluorescence intensity, as compared to the mutant control (Figure 7B). These results indicate the Scl1.3FL is indeed expressed and surface-exposed in the complemented M3-type strains.

Biofilm formation by the complemented and parent strains was then assessed after 24 hours following crystal violet staining and using confocal laser scanning microscopy (CLSM). Scl1.3FL-expressing MGAS315 showed significantly increased biomass on both cFn and Lm coatings compared to the WT parent organism, as well as vector-complemented control; 3.6 or 2.5- and 2.5 or 2.1-fold OD_{600} increases on cFn and Lm, respectively, were measured compared to MGAS315 WT or to vector-complemented MGAS315 (Figure 7C). Significantly thicker biofilm formed by Scl1.3FL-complemented MGAS315 was imaged by CLSM. We observed on average a 2.8-fold (P= 0.0002) and 2.6-fold (P= 0.0002) increased biofilm thickness on cFn and Lm, respectively, compared to MGAS315 WT, and a 2.0-fold (P= 0.0036) and 4.8-fold (P= 2.1×10^{-5}) increased biofilm thickness on cFn and Lm, respectively, compared to vector-complemented MGAS315 (Figure 7D and E, representative images). Similar results were obtained for the complemented 10870Δ$scl1$ mutant, which had 1.7-fold increase in biomass staining on both cFn and Lm compared to the MGAS10870 WT strain, and 2.2- and 2.6-fold increase in biomass staining on cFn and Lm, respectively, compared to the 10870Δ$scl1$ mutant strain (Figure 7C). CLSM data, however, could not be rendered for MGAS10870 strains due to poor GFP expression for unknown reasons. Altogether, it was demonstrated that null mutation in the $scl1$ gene, which ablates surface Scl1.3 protein and is unique to M3-type GAS, is responsible for the decreased biofilm capacity since restoration of the full-length surface-exposed Scl1.3 adhesin significantly fosters stable biofilm formation.

Heterologous complementation of M41 *Δscl1* mutant strain with full-length surface-exposed Scl1.3 restores biofilm formation on ECM

In a previous study we showed that *scl1.41*-inactivation in a non-invasive biofilm-rich M41 strain MGAS6183 resulted in significantly decreased biofilm capacity, which was restored to wild-type level by complementation with surface Scl1.41 (37). Notably, this M41 strain expresses at least one additional major Fn-binding protein, protein F2 (45), which binds both plasma and cellular fibronectin by a mechanism different from Scl1 (57). Here, we hypothesized that expression of rScl1.3FL in the heterologous M41 GAS will confer biofilm formation. Western immunoblotting of the cell wall protein fractions detected the ~35-kDa immunoreactive band, corresponding to full-length Scl1.3, associated with complemented cells, while the M41*Δscl1* mutant and WT cells were signal-negative (Figure 8A). Mass spectrometry of the corresponding band extracted from the gel confirmed Scl1.3FL expression, with 3 unique peptides identified, covering 13% of the amino acid sequence (Table S3). In addition, we showed the expression of the larger Scl1.41 variant in the cell wall of M41 WT, but not in the *Δscl1* mutant, by re-probing a portion of the blot with anti-rScl1.41 antibody, using rScl1.41 protein as a positive control. Similarly to complementation with homologous Scl1.41 protein, the heterologous complementation of the M41*Δscl1* mutant with Scl1.3FL also restored biofilm to M41 WT levels on cFn and Lm (Figure 8B-D). Crystal violet staining showed increased bacterial biomass (Figure 8B) and confocal microscopy revealed significantly thicker biofilms, on average 3-fold increased, as compared to the parental M41*Δscl1* mutant strain on both cFn (Figure 8C) and Lm (Figure 8D) coatings (cFn, $P= 0.0105$; Lm, $P= 0.0011$).

Expression of Scl1 adhesin attenuates GAS during subcutaneous infection

We have previously shown that an M3 strain harboring a carrier *scl1.3* allele, producing a shorter cell-attached Scl1.3 variant, had an attenuated phenotype in a murine model of necrotizing fasciitis (40). Here, we tested our hypothesis that Scl1 adhesin in biofilm-rich M28 and M41 background promotes stable colonization and localized infection, using a murine skin infection model. Hairless, immunocompetent SKH1 mice were subcutaneously infected with ~10^9 GAS CFU of the M28 and M41 wild-type (WT) or their

isogenic *scl1*-inactivated mutant (*scl1*) strains (44,45), and mice were assessed for changes in gross pathology of the skin.

Skin lesions caused by both the WT and *scl1* strains were observed as early as 48 hours post-infection, and the lesions began to regress after day 7 with complete resolution of the lesion by the completion of the experiment. The areas of the skin lesions calculated for mice infected with the *scl1* mutant strains were significantly larger than those of WT-infected mice (Figure 9A; 96 hour time-point is shown); images of lesions of representative mice demonstrate differences in lesion severity between WT- and *scl1*-GAS infected mice for both the M28- and M41-infected groups (Figure 9B). Thus, the *in vivo* data, using GAS strains expressing surface Scl1 proteins and producing rich biofilms, support our hypothesis that decreased adhesion and biofilm formation, due to the absence of Scl1.3 on the surface of M3 strains, bears an inverse correlation to the invasive potential of the infecting GAS strain.

DISCUSSION

Since the resurgence of invasive GAS disease in the 1980's, and emergence of invasive M3-type isolates, significant efforts have been made to determine the molecular basis for the invasive phenotype of M3-type strains. Numerous whole-genome sequencing projects have identified specific genomic features of M3 strains that were correlated with their potential to cause invasive infections. In this study, we show that Scl1-negative M3-type GAS have reduced adhesion and biofilm formation within host tissue, and therefore are predisposed to invasive spread over superficial infection (Figure 10).

Unique topography and expression patterns of *scl* loci

First, the insertion of IS1548 element in the promoter region of *scl1* exclusively found in the M3-type strains is an intriguing observation, given the abundance of IS1548 insertions across GAS genomes representing a variety of M-types. IS elements are known to cause genomic rearrangements and affect the expression of important genes that contribute to bacterial virulence (58). For example, the insertion of IS1548 element (59) into the *scpB-lmb* intergenic region in group B *Streptococcus* has been shown to upregulate *lmb*-gene transcription and Lmb-adhesin surface expression, resulting in enhanced laminin binding

(60). The IS1548 insertion in the *scl1.3* promoter was conserved in the genomes of sequenced M3 strains MGAS315, SSI-1, and M3-b, as well as in 40 additional M3 strains tested by PCR. However, the recently sequenced STAB902 M3 strain (55), which represents a non-invasive isolate, contained a 34-bp IS1548 remnant, suggesting that M3 circulating strains exist with this polymorphism. We hypothesize that the proximity of the IS1548 insertion to the Mga I binding site in M3 strains affects *scl1.3* transcription. Comparative qRT-PCR analysis showed significantly higher levels of *scl1* transcription, particularly in M1, as well as in M28 and M41 strains, relative to M3-type strains analyzed.

Variation in the Mga coding sequence and autoregulated *mga* promoter has been reported to significantly affect Mga-dependent gene expression and virulence properties in GAS (26,27,61). For example, a 12-bp deletion of a single repeat in the VNTR region of the *mga* promoter is present in M3 carrier strains and absent in the invasive M3 strains, causing a downregulation of Mga expression and Mga-regulated genes in the carrier strains (27). We identified an analogous 12-bp deletion in the VNTR of the *mga* promoter of the M1-type MGAS5005 and in M3 strains, MGAS335 and MGAS1313, but not in MGAS315, MGAS158 and MGAS10870. It is likely that the variation observed in the VNTR region impacts Mga expression but could not, alone, explain differences in *scl1* transcription patterns observed in this work. Consequently, qRT-PCR analyses showed *mga* and *emm* transcripts (*emm* is second downstream target of Mga within the Mga regulon) were expressed at similar levels in MGAS5005 and M3 strains, except for the MGAS335, indicating a mechanism of *scl1* downregulation in M3 strains, which is independent from the level of Mga transcription. The presence of *emm* transcripts and M3-protein products indicate that Mga is present and functional in the majority of M3 strains tested, and therefore the polymorphism we observed in the *mga* promoter did not affect Mga expression or activity of Mga-controlled *scl1.3* promoter. Altogether, our data hint at the insertion of the IS1548 in the downregulation of *scl1.3* expression, specifically in M3-type GAS, although additional studies are required to firmly establish whether the IS element insertion is responsible.

It is intriguing that *scl1* and *scl2* expression patterns were drastically different between M1 and M3-type GAS, both associated with invasive infections. The *scl1.3* transcript was decreased by 21-fold in MGAS315, and even further in the remaining M3

strains studied (additional 20-45 fold), as compared to the M1-type strain MGAS5005. The upregulation of *scl1* in M1-type GAS has previously been shown in invasive M1 isolates, including MGAS5005, as compared to pharyngeal M1 isolates (62). Studies comparing wild-type and isogenic *scl1.1* mutants in M1-type GAS reported that Scl1.1 contributed to immune evasion by inhibiting neutrophil extracellular trap production and by protecting bacteria from the cathelicidin LL-37 (63). We conclude the differences we observed in *scl1* expression are related to different pathogenicity requirements for Scl1 protein in M1- versus M3-type strains. In contrast, Scl2 expression is highly upregulated in M3-type strains, as compared to invasive M1-type GAS. While the majority of M3-type strains contain in-frame *scl2.3* alleles, all 21 M1 strains we analyzed contained out-of-frame *scl2.1* alleles (unpublished data). Although the Scl2.3 human ligands are not known, it may have an unidentified biological function, which is important for pathogenesis of M3, but not M1, strains. Hence, we are reporting striking differences in the expression and features of the Scl1 and Scl2 proteins that evolved in the invasive M1 and M3 strains.

Scl1-mediated adhesion and biofilm formation *in vitro*

Scl1 mediates binding to human extracellular matrix components, cellular fibronectin and laminin, as well as biofilm formation (35-37). Previous work identified that Scl1 binds specifically to the C-C' loop of the type III-repeat EDA domain of cellular fibronectin (37). This represents a novel mechanism of fibronectin binding, which is distinct from the mechanism employed by other GAS fibronectin-binding proteins that bind to the N-terminal type I repeats of fibronectin (64). Here, we determined that rScl1.3FL binds to cellular fibronectin via the same C-C' loop-dependent mechanism. It is important for our overall model to acknowledge that the EDA-cFn isoform is specifically produced during embryogenesis and during wound healing in adult tissue (65-67), which indicates Scl1 evolved with a unique function for targeting wounded tissue, a presumed pathogen portal of entry. We also demonstrate significant binding of rScl1.3FL to laminin, a major basement membrane protein at the epidermal-dermal junction; none of those ECM ligands were bound by rScl2.3, consistent with previous knowledge that Scl1, but not Scl2, variants exhibit binding to cFn and Lm (35). These studies show that full-length

Scl1.3, if expressed on the M3-GAS cell surface, would bind ECM, contributing to tissue colonization.

The diminished Scl1.3 expression and poor biofilm formation, led us to hypothesize that lack of surface-expressed Scl1 in M3-type GAS reduces host ECM binding and stable microcolony formation in tissue, thus, shifting the balance towards invasive spread, augmented by other virulence factors expressed by these strains. Recombinant rScl1.3FL formed aggregates that were mediated by the V-to-V region interactions, which could represent a mechanism of biofilm and microcolony stabilization by Scl1 molecules on adjacent GAS cells. Two representative invasive M3 isolates, MGAS315 and MGAS10870, acquired biofilm formation on cFn and Lm when homologous complementation was performed with surface-attached Scl1.3FL. MGAS315 has been shown to contain a missense mutation in the *covS* gene, causing upregulation of CovRS-regulated virulence genes and enhanced virulence during subcutaneous infection of mice, compared to an isogenic strain containing the WT *covS* allele (68). Additionally, MGAS315 contains a mutation in the regulator of protein B allele, *ropB*, which produces a nonfunctional RopB variant, while MGAS10870 contains a wild-type *covR/S* and *ropB* alleles (28), allowing us to demonstrate the effect of Scl1.3FL function in the presence of differing regulatory networks. Similarly, heterologous complementation with surface Scl1.3FL in a *scl1.41*-mutant of the non-invasive biofilm-capable strain M41-type MGAS6183, restored its biofilm capacity to the wild-type level. These results indicate that M3-derived Scl1.3FL variant has the full capacity to support biofilm formation to a similar degree as Scl1 from a divergent M-type. The robust biofilm observed *in vitro* on ECM coatings validates the concept that biofilm capacity combined with adherence to the surrounding host ECM would reinforce the formation of stable tissue microcolonies *in vivo*.

In vitro skin equivalent model of wound colonization and microcolony formation

We observed inhibition of wound re-epithelization by GAS infection of wounded skin equivalents, as well as the thinning of the viable epidermal layer at sites distant from the infected wound. In addition to our study, others have reported changes in skin histopathology and wound healing, resulting from bacterial infections. An *in vivo* study

has reported epidermal defects as a result of GAS infection in a humanized mouse model with human skin graft (69). Previous study of M3-type GAS infection using a skin equivalent model showed that hyaluronic acid capsule interactions with CD44 receptor on keratinocytes induced intracellular signaling, resulting in cytoskeletal rearrangement and monolayer disruption (70). Infection of an *in vitro* skin model containing a burn wound with *Pseudomonas aeruginosa* caused a loss of the keratinocyte layer and basement membrane, while intact epidermis was observed in burned but uninfected tissue (71). Impairment of wound healing has also been demonstrated by staphylococcal infections. Infection of dermal wounds in rabbit ears with *Staphylococcus aureus* showed the formation of biofilm, production of a persistent, low grade inflammatory response, and significantly delayed wound healing (72). Similarly, delayed wound healing by both *S. aureus* and *S. epidermidis* biofilms was observed in a mouse model of cutaneous wounds (73). The inhibition of wound healing we describe here is by and large consistent with reports by other laboratories generated using *in vivo* animal and *in vitro* human skin infection models.

Microcolonies have been identified within human streptococcal impetigo lesions (39) and in tonsils from patients with recurrent GAS pharyngeal tonsillitis (74), and likely represent a superficial or persistent state of GAS colonization. However, streptococcal infections can result in invasive disease due to biofilm disruption and bacterial dissemination (75,76). We observed large rounded microcolonies formed in tissue during infection with M41 strain, while M3 GAS remained scattered throughout the tissue as single cells and chains. Microcolony formation has been previously observed with *S. aureus* infection in organotypic skin model (77). Moreover, wound infection in rabbit ears with *S. aureus* produced mature biofilms encased in exopolysaccharide, as revealed by concanavalin A staining (72). Similarly, we also demonstrated that microcolonies in M41-infected tissue were encapsulated in bacterial-associated glycocalyx. However, a lack of glycocalyx-encapsulated microcolonies was associated with infection by M3-type GAS.

These results support our hypothesis that biofilm-poor M3 strains are abolished in stable microcolony formation *in vivo,* in part due to a lack of the ECM-binding Scl1 protein and an overall lack of surface adhesins, although they likely express the FbaB protein,

identified in M3 GAS to be involved in the adherence and invasion into epithelial and endothelial cells (78-80).

In vivo mouse model of skin infection

Recent study reported that a small proportion of non-invasive M3-type strains (~1.3%) were found to harbor the _scl1.3_ "carrier allele", which resulted from an in-frame deletion in the collagenous region, encompassing the null mutation, producing a shorter cell-attached Scl1.3 variant. This MGAS10870 strain containing the _scl1.3_ carrier allele was attenuated following intramuscular infection (40). In this study, we utilized the M28- and M41-type strains, representing biofilm-rich producers, for subcutaneous inoculation. We observed that _scl1.28-_ and _scl1.41_-inactivated isogenic mutants produced significantly larger skin lesions as compared to the wild-type parent strains. These results, again, support the hypothesis that lack of Scl1 surface adhesin destabilizes focused nidus of infection, resulting in a shift towards increased tissue spread. However, previous studies performed in a M1 GAS background, utilizing _scl1.1_-mutants for subcutaneous infection, reported smaller skin lesions in the mutant groups, which likely reflects a differing predominant function of Scl1.1 in M1-type GAS (29,63). Investigations using intranasal and intraperitoneal mouse infection models of _Streptococcus pneumoniae_ have shown that culture-grown bacteria disseminated to the ear and lungs, while biofilm-grown bacteria stably colonized the nasopharynx (81,82). A similar study on _Streptococcus pyogenes_ showed that bacteria grown in biofilms have downregulated virulence genes and tend to colonize the nasal associated lymphoid tissue of mice, while culture-grown bacteria had significantly increased dissemination and were more virulent in a septicemia model (83). Previous studies reported that inactivation of some GAS genes resulted in increased skin pathology produced by the mutants compared to their wild-type organisms, and these genes often encoded surface proteins, including SpyCEP (84), Mrp (85), protein F1 (86), and Spy0128, encoding a major pilus subunit (87). Similarly, the _covS_ mutant of group A streptococcal M1T1 strain with upregulated SpeB-protease activity was hypervirulent and had reduced capacity to bind human epithelial cells and fibronectin, and also to form biofilm due to increased cleavage of surface proteins (88). Altogether, the concept that expression of a surface adhesin, such as Scl1, involved in biofilm formation

and host tissue attachment, is inversely related to strain invasiveness has gained support from several studies, including this study.

We show the invasive M1- and M3-type GAS evolved *scl1* and *scl2* alleles with opposite expression patterns, with *scl1* downregulated and *scl2* upregulated in M3 compared to M1 GAS. We show M3-type GAS, devoid of surface-expressed Scl1.3, lacked biofilm formation on ECM coatings and microcolony formation during infection of *in vitro* wounded skin equivalent. Complementation with surface Scl1.3FL restored biofilm capacity of M3-type GAS on ECM coatings. Mouse infection with the isogenic *scl1* mutants of biofilm-rich M28- and M41-type GAS produced larger lesions, supporting the role of Scl1 in a localized tissue infection. Lastly we developed a model for Scl1-mediated microcolony formation (Figure 10), whereby Scl1 expressed on the GAS surface strengthens host colonization by attachment to cFn and Lm expressed within wounded tissue, as well as biofilm formation via Scl1-Scl1 interactions, resulting in a local, stabilized microcolony. Conversely, when Scl1 is absent on the GAS cell surface, as is the case for M3-type GAS, bacteria lack stable anchoring in the surrounding host ECM, as well as structural strength within microcolony, promoting cells to disperse more freely.

ACKNOWLEDGMENTS

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
2. Sims Sanyahumbi, A., Colquhoun, S., Wyber, R., and Carapetis, J. R. (2016) Global Disease Burden of Group A *Streptococcus*. in *Streptococcus pyogenes: Basic Biology to Clinical Manifestations* (Ferretti, J. J., Stevens, D. L., and Fischetti, V. A. eds.), The University of Oklahoma Health Sciences Center, Oklahoma City, OK. pp
3. Cunningham, M. W. (2000) Pathogenesis of group A streptococcal infections. *Clin Microbiol Rev* **13**, 470-511
4. Stevens, D. L., and Kaplan, E. L. (eds). (2000) *Streptococcal infections: clinical aspects, microbiology, and molecular pathogenesis*, Oxford University Press, New York, N.Y.
5. Shulman, S. T., Tanz, R. R., Kabat, W., Kabat, K., Cederlund, E., Patel, D., Li, Z., Sakota, V., Dale, J. B., and Beall, B. (2004) Group A streptococcal pharyngitis serotype surveillance in North America, 2000-2002. *Clin Infect Dis* **39**, 325-332
6. Musser, J. M., Hauser, A. R., Kim, M. H., Schlievert, P. M., Nelson, K., and Selander, R. K. (1991) *Streptococcus pyogenes* causing toxic-shock-like syndrome and other invasive diseases: clonal diversity and pyrogenic exotoxin expression. *Proc Natl Acad Sci U S A* **88**, 2668-2672
7. Cleary, P. P., Kaplan, E. L., Handley, J. P., Wlazlo, A., Kim, M. H., Hauser, A. R., and Schlievert, P. M. (1992) Clonal basis for resurgence of serious *Streptococcus pyogenes* disease in the 1980s. *Lancet* **339**, 518-521
8. Cockerill, F. R., 3rd, MacDonald, K. L., Thompson, R. L., Roberson, F., Kohner, P. C., Besser-Wiek, J., Manahan, J. M., Musser, J. M., Schlievert, P. M., Talbot, J., Frankfort, B., Steckelberg, J. M., Wilson, W. R., and Osterholm, M. T. (1997) An outbreak of invasive group A streptococcal disease associated with high carriage rates of the invasive clone among school-aged children. *JAMA* **277**, 38-43
9. Stevens, D. L., Tanner, M. H., Winship, J., Swarts, R., Ries, K. M., Schlievert, P. M., and Kaplan, E. (1989) Severe group A streptococcal infections associated with a toxic shock-like syndrome and scarlet fever toxin A. *N Engl J Med* **321**, 1-7
10. DiPersio, J. R., File, T. M., Jr., Stevens, D. L., Gardner, W. G., Petropoulos, G., and Dinsa, K. (1996) Spread of serious disease-producing M3 clones of group A

Streptococcus among family members and health care workers. *Clin Infect Dis* **22**, 490-495

11. Johnson, D. R., Stevens, D. L., and Kaplan, E. L. (1992) Epidemiologic analysis of group A streptococcal serotypes associated with severe systemic infections, rheumatic fever, or uncomplicated pharyngitis. *J Infect Dis* **166**, 374-382

12. Davies, H. D., McGeer, A., Schwartz, B., Green, K., Cann, D., Simor, A. E., Low, D. E., and Group, O. G. A. S. S. (1996) Invasive group A streptococcal infections in Ontario, Canada. *N Engl J Med* **335**, 547-554

13. Kaul, R., McGeer, A., Low, D. E., Green, K., and Schwartz, B. (1997) Population-based surveillance for group A streptococcal necrotizing fasciitis: Clinical features, prognostic indicators, and microbiologic analysis of seventy-seven cases. Ontario Group A Streptococcal Study. *Am J Med* **103**, 18-24

14. Sharkawy, A., Low, D. E., Saginur, R., Gregson, D., Schwartz, B., Jessamine, P., Green, K., and McGeer, A. (2002) Severe group a streptococcal soft-tissue infections in Ontario: 1992-1996. *Clin Infect Dis* **34**, 454-460

15. Hollm-Delgado, M. G., Allard, R., and Pilon, P. A. (2005) Invasive group A streptococcal infections, clinical manifestations and their predictors, Montreal, 1995-2001. *Emerg Infect Dis* **11**, 77-82

16. Gaworzewska, E., and Colman, G. (1988) Changes in the pattern of infection caused by *Streptococcus pyogenes*. *Epidemiol Infect* **100**, 257-269

17. Colman, G., Tanna, A., Efstratiou, A., and Gaworzewska, E. T. (1993) The serotypes of *Streptococcus pyogenes* present in Britain during 1980-1990 and their association with disease. *J Med Microbiol* **39**, 165-178

18. Meisal, R., Hoiby, E. A., Caugant, D. A., and Musser, J. M. (2010) Molecular characteristics of pharyngeal and invasive *emm3 Streptococcus pyogenes* strains from Norway, 1988-2003. *Eur J Clin Microbiol Infect Dis* **29**, 31-43

19. Lamagni, T. L., Neal, S., Keshishian, C., Alhaddad, N., George, R., Duckworth, G., Vuopio-Varkila, J., and Efstratiou, A. (2008) Severe *Streptococcus pyogenes* infections, United Kingdom, 2003-2004. *Emerg Infect Dis* **14**, 202-209

20. Beres, S. B., Sylva, G. L., Barbian, K. D., Lei, B., Hoff, J. S., Mammarella, N. D., Liu, M. Y., Smoot, J. C., Porcella, S. F., Parkins, L. D., Campbell, D. S., Smith, T. M., McCormick, J. K., Leung, D. Y., Schlievert, P. M., and Musser, J. M. (2002) Genome sequence of a serotype M3 strain of group A *Streptococcus*: phage-encoded toxins, the high-virulence phenotype, and clone emergence. *Proc Natl Acad Sci U S A* **99**, 10078-10083

21. Nakagawa, I., Kurokawa, K., Yamashita, A., Nakata, M., Tomiyasu, Y., Okahashi, N., Kawabata, S., Yamazaki, K., Shiba, T., Yasunaga, T., Hayashi, H., Hattori, M., and Hamada, S. (2003) Genome sequence of an M3 strain of *Streptococcus pyogenes* reveals a large-scale genomic rearrangement in invasive strains and new insights into phage evolution. *Genome Res* **13**, 1042-1055

22. Beres, S. B., Sylva, G. L., Sturdevant, D. E., Granville, C. N., Liu, M., Ricklefs, S. M., Whitney, A. R., Parkins, L. D., Hoe, N. P., Adams, G. J., Low, D. E., DeLeo, F. R., McGeer, A., and Musser, J. M. (2004) Genome-wide molecular dissection of serotype M3 group A *Streptococcus* strains causing two epidemics of invasive infections. *Proc Natl Acad Sci U S A* **101**, 11833-11838

23. Beres, S. B., Carroll, R. K., Shea, P. R., Sitkiewicz, I., Martinez-Gutierrez, J. C., Low, D. E., McGeer, A., Willey, B. M., Green, K., Tyrrell, G. J., Goldman, T. D., Feldgarden, M., Birren, B. W., Fofanov, Y., Boos, J., Wheaton, W. D., Honisch, C., and Musser, J. M. (2010) Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. *Proc Natl Acad Sci U S A* **107**, 4371-4376

24. Shea, P. R., Beres, S. B., Flores, A. R., Ewbank, A. L., Gonzalez-Lugo, J. H., Martagon-Rosado, A. J., Martinez-Gutierrez, J. C., Rehman, H. A., Serrano-Gonzalez, M., Fittipaldi, N., Ayers, S. D., Webb, P., Willey, B. M., Low, D. E., and Musser, J. M. (2011) Distinct signatures of diversifying selection revealed by genome analysis of respiratory tract and invasive bacterial populations. *Proc Natl Acad Sci U S A* **108**, 5039-5044

25. Olsen, R. J., Laucirica, D. R., Watkins, M. E., Feske, M. L., Garcia-Bustillos, J. R., Vu, C., Cantu, C., Shelburne, S. A., 3rd, Fittipaldi, N., Kumaraswami, M., Shea, P. R., Flores, A. R., Beres, S. B., Lovgren, M., Tyrrell, G. J., Efstratiou, A., Low, D. E., Van Beneden, C. A., and Musser, J. M. (2012) Polymorphisms in regulator of protease B (RopB) alter disease phenotype and strain virulence of serotype M3 group A *Streptococcus. J Infect Dis* **205**, 1719-1729

26. Cao, T. N., Liu, Z., Cao, T. H., Pflughoeft, K. J., Trevino, J., Danger, J. L., Beres, S. B., Musser, J. M., and Sumby, P. (2014) Natural disruption of two regulatory networks in serotype M3 group A *Streptococcus* isolates contributes to the virulence factor profile of this hypervirulent serotype. *Infect Immun* **82**, 1744-1754

27. Flores, A. R., Olsen, R. J., Wunsche, A., Kumaraswami, M., Shelburne, S. A., 3rd, Carroll, R. K., and Musser, J. M. (2013) Natural variation in the promoter of the gene encoding the Mga regulator alters host-pathogen interactions in group a *Streptococcus* carrier strains. *Infect Immun* **81**, 4128-4138

28. Carroll, R. K., Shelburne, S. A., 3rd, Olsen, R. J., Suber, B., Sahasrabhojane, P., Kumaraswami, M., Beres, S. B., Shea, P. R., Flores, A. R., and Musser, J. M. (2011) Naturally occurring single amino acid replacements in a regulatory protein alter streptococcal gene expression and virulence in mice. *J Clin Invest* **121**, 1956-1968

29. Lukomski, 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

30. Lukomski, S., Nakashima, K., Abdi, I., Cipriano, V. J., Shelvin, B. J., Graviss, 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

31. Rasmussen, M., and Bjorck, L. (2001) Unique regulation of SclB - a novel collagen-like surface protein of *Streptococcus pyogenes*. *Mol Microbiol* **40**, 1427-1438

32. Rasmussen, M., Eden, A., and Bjorck, L. (2000) SclA, a novel collagen-like surface protein of *Streptococcus pyogenes*. *Infect Immun* **68**, 6370-6377

33. Whatmore, A. M. (2001) *Streptococcus pyogenes sclB* encodes a putative hypervariable surface protein with a collagen-like repetitive structure. *Microbiology* **147**, 419-429

34. Almengor, A. C., and McIver, K. S. (2004) Transcriptional activation of *sclA* by Mga requires a distal binding site in *Streptococcus pyogenes*. *J Bacteriol* **186**, 7847-7857

35. Caswell, C. C., Oliver-Kozup, H., Han, R., Lukomska, E., and Lukomski, S. (2010) Scl1, the multifunctional adhesin of group A *Streptococcus*, selectively binds cellular fibronectin and laminin, and mediates pathogen internalization by human cells. *FEMS Microbiol Lett* **303**, 61-68

36. Oliver-Kozup, H. A., Elliott, M., Bachert, B. A., Martin, K. H., Reid, S. D., Schwegler-Berry, D. E., Green, B. J., and Lukomski, S. (2011) The streptococcal collagen-like protein-1 (Scl1) is a significant determinant for biofilm formation by group A *Streptococcus. BMC Microbiol* **11**, 262

37. Oliver-Kozup, H., Martin, K. H., Schwegler-Berry, D., Green, B. J., Betts, C., Shinde, A. V., Van De Water, L., and Lukomski, S. (2013) The group A streptococcal collagen-like protein-1, Scl1, mediates biofilm formation by targeting the extra domain A-containing variant of cellular fibronectin expressed in wounded tissue. *Mol Microbiol* **87**, 672-689

38. Lembke, C., Podbielski, A., Hidalgo-Grass, C., Jonas, L., Hanski, E., and Kreikemeyer, B. (2006) Characterization of biofilm formation by clinically relevant serotypes of group A streptococci. *Appl Environ Microbiol* **72**, 2864-2875

39. Akiyama, H., Morizane, S., Yamasaki, O., Oono, T., and Iwatsuki, K. (2003) Assessment of *Streptococcus pyogenes* microcolony formation in infected skin by confocal laser scanning microscopy. *J Dermatol Sci* **32**, 193-199

40. Flores, 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

41. Pahlman, L. I., Marx, P. F., Morgelin, M., Lukomski, S., Meijers, J. C., and Herwald, H. (2007) Thrombin-activatable fibrinolysis inhibitor binds to *Streptococcus pyogenes* by interacting with collagen-like proteins A and B. *J Biol Chem* **282**, 24873-24881

42. Squeglia, F., Bachert, B., Romano, M., Lukomski, S., and Berisio, R. (2013) Crystallization and preliminary X-ray crystallographic analysis of the variable domain of Scl2.3, a streptococcal collagen-like protein from invasive M3-type *Streptococcus pyogenes*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **69**, 1023-1025

43. Squeglia, F., Bachert, B., De Simone, A., Lukomski, S., and Berisio, R. (2014) The crystal structure of the streptococcal collagen-like protein 2 globular domain from invasive M3-type group A *Streptococcus* shows significant similarity to immunomodulatory HIV protein gp41. *J Biol Chem* **289**, 5122-5133

44. Han, R., Caswell, C. C., Lukomska, E., Keene, D. R., Pawlowski, M., Bujnicki, J. M., Kim, J. K., and Lukomski, S. (2006) Binding of the low-density lipoprotein by streptococcal collagen-like protein Scl1 of *Streptococcus pyogenes*. *Mol Microbiol* **61**, 351-367

45. Caswell, C. C., Lukomska, E., Seo, N. S., Hook, M., and Lukomski, S. (2007) Scl1-dependent internalization of group A *Streptococcus* via direct interactions with the

alpha2beta(1) integrin enhances pathogen survival and re-emergence. *Mol Microbiol* **64**, 1319-1331

46.	Virtaneva, K., Porcella, S. F., Graham, M. R., Ireland, R. M., Johnson, C. A., Ricklefs, S. M., Babar, I., Parkins, L. D., Romero, R. A., Corn, G. J., Gardner, D. J., Bailey, J. R., Parnell, M. J., and Musser, J. M. (2005) Longitudinal analysis of the group A *Streptococcus* transcriptome in experimental pharyngitis in cynomolgus macaques. *Proc Natl Acad Sci U S A* **102**, 9014-9019

47.	Han, R., Zwiefka, A., Caswell, C. C., Xu, Y., Keene, D. R., Lukomska, E., Zhao, Z., Hook, 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

48.	McIver, K. S., and Scott, J. R. (1997) Role of *mga* in growth phase regulation of virulence genes of the group A *Streptococcus*. *J Bacteriol* **179**, 5178-5187

49.	Xu, Y., Keene, D. R., Bujnicki, J. M., Hook, M., and Lukomski, S. (2002) Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. *J Biol Chem* **277**, 27312-27318

50.	Fields, G. B., and Noble, R. L. (1990) Solid phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids. *Int J Pept Protein Res* **35**, 161-214

51.	Vanier, G. S. (2013) Microwave-assisted solid-phase peptide synthesis based on the Fmoc protecting group strategy (CEM). *Methods Mol Biol* **1047**, 235-249

52.	Williamson, M. P. (2013) Using chemical shift perturbation to characterise ligand binding. *Progress in nuclear magnetic resonance spectroscopy* **73**, 1-16

53.	Cramer, T., Yamanishi, Y., Clausen, B. E., Forster, I., Pawlinski, R., Mackman, N., Haase, V. H., Jaenisch, R., Corr, M., Nizet, V., Firestein, G. S., Gerber, H. P., Ferrara, N., and Johnson, R. S. (2003) HIF-1alpha is essential for myeloid cell-mediated inflammation. *Cell* **112**, 645-657

54.	Lukomski, S., Montgomery, C. A., Rurangirwa, J., Geske, R. S., Barrish, J. P., Adams, G. J., and Musser, J. M. (1999) Extracellular cysteine protease produced by *Streptococcus pyogenes* participates in the pathogenesis of invasive skin infection and dissemination in mice. *Infect Immun* **67**, 1779-1788

55.	Soriano, N., Vincent, P., Moullec, S., Meygret, A., Lagente, V., Kayal, S., and Faili, A. (2014) Closed genome sequence of noninvasive *Streptococcus pyogenes* M/*emm3* strain STAB902. *Genome Announc* **2**

56.	Russo, A., Scognamiglio, P. L., Hong Enriquez, R. P., Santambrogio, C., Grandori, R., Marasco, D., Giordano, A., Scoles, G., and Fortuna, S. (2015) In Silico Generation of Peptides by Replica Exchange Monte Carlo: Docking-Based Optimization of Maltose-Binding-Protein Ligands. *PLoS ONE* **10**, e0133571

57.	Sela, S., Aviv, A., Tovi, A., Burstein, I., Caparon, M. G., and Hanski, E. (1993) Protein F: an adhesin of *Streptococcus pyogenes* binds fibronectin via two distinct domains. *Mol Microbiol* **10**, 1049-1055

58.	Mahillon, J., Leonard, C., and Chandler, M. (1999) IS elements as constituents of bacterial genomes. *Res Microbiol* **150**, 675-687

59.	Granlund, M., Oberg, L., Sellin, M., and Norgren, M. (1998) Identification of a novel insertion element, IS1548, in group B streptococci, predominantly in strains causing endocarditis. *J Infect Dis* **177**, 967-976

60. Al Safadi, R., Amor, S., Hery-Arnaud, G., Spellerberg, B., Lanotte, P., Mereghetti, L., Gannier, F., Quentin, R., and Rosenau, A. (2010) Enhanced expression of *lmb* gene encoding laminin-binding protein in *Streptococcus agalactiae* strains harboring IS1548 in *scpB-lmb* intergenic region. *PLoS ONE* **5**, e10794

61. Sanson, M., O'Neill, B. E., Kachroo, P., Anderson, J. R., Flores, A. R., Valson, C., Cantu, C. C., Makthal, N., Karmonik, C., Fittipaldi, N., Kumaraswami, M., Musser, J. M., and Olsen, R. J. (2015) A naturally occurring single amino acid replacement in multiple gene regulator of group A *Streptococcus* significantly increases virulence. *Am J Pathol* **185**, 462-471

62. Sumby, P., Whitney, A. R., Graviss, E. A., DeLeo, F. R., and Musser, J. M. (2006) Genome-wide analysis of group A streptococci reveals a mutation that modulates global phenotype and disease specificity. *PLoS Pathog* **2**, e5

63. Dohrmann, S., Anik, S., Olson, J., Anderson, E. L., Etesami, N., No, H., Snipper, J., Nizet, V., and Okumura, C. Y. (2014) Role for streptococcal collagen-like protein 1 in M1T1 group A *Streptococcus* resistance to neutrophil extracellular traps. *Infect Immun* **82**, 4011-4020

64. Yamaguchi, M., Terao, Y., and Kawabata, S. (2013) Pleiotropic virulence factor - *Streptococcus pyogenes* fibronectin-binding proteins. *Cell Microbiol* **15**, 503-511

65. 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 alpha9beta1 and one of its ligands, the EIIIA segment of fibronectin, in cutaneous wound healing. *J Invest Dermatol* **123**, 1176-1181

66. 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

67. Jarnagin, W. R., Rockey, D. C., Koteliansky, V. E., Wang, S. S., and Bissell, D. M. (1994) Expression of variant fibronectins in wound healing: cellular source and biological activity of the EIIIA segment in rat hepatic fibrogenesis. *J Cell Biol* **127**, 2037-2048

68. Stetzner, Z. W., Li, D., Feng, W., Liu, M., Liu, G., Wiley, J., and Lei, B. (2015) Serotype M3 and M28 group A streptococci have distinct capacities to evade neutrophil and TNF-alpha responses and to invade soft tissues. *PLoS ONE* **10**, e0129417

69. Scaramuzzino, D. A., McNiff, J. M., and Bessen, D. E. (2000) Humanized *in vivo* model for streptococcal impetigo. *Infect Immun* **68**, 2880-2887

70. Cywes, C., and Wessels, M. R. (2001) Group A *Streptococcus* tissue invasion by CD44-mediated cell signalling. *Nature* **414**, 648-652

71. Shepherd, J., Douglas, I., Rimmer, S., Swanson, L., and MacNeil, S. (2009) Development of three-dimensional tissue-engineered models of bacterial infected human skin wounds. *Tissue Eng Part C Methods* **15**, 475-484

72. Gurjala, A. N., Geringer, M. R., Seth, A. K., Hong, S. J., Smeltzer, M. S., Galiano, R. D., Leung, K. P., and Mustoe, T. A. (2011) Development of a novel, highly quantitative *in vivo* model for the study of biofilm-impaired cutaneous wound healing. *Wound Repair Regen* **19**, 400-410

73. Schierle, C. F., De la Garza, M., Mustoe, T. A., and Galiano, R. D. (2009) Staphylococcal biofilms impair wound healing by delaying reepithelialization in a murine cutaneous wound model. *Wound Repair Regen* **17**, 354-359

74. 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

75. Connolly, K. L., Roberts, A. L., Holder, R. C., and Reid, S. D. (2011) Dispersal of Group A streptococcal biofilms by the cysteine protease SpeB leads to increased disease severity in a murine model. *PLoS ONE* **6**, e18984

76. Connolly, K. L., Braden, A. K., Holder, R. C., and Reid, S. D. (2011) Srv mediated dispersal of streptococcal biofilms through SpeB is observed in CovRS+ strains. *PLoS ONE* **6**, e28640

77. Popov, L., Kovalski, J., Grandi, G., Bagnoli, F., and Amieva, M. R. (2014) Three-Dimensional Human Skin Models to Understand *Staphylococcus aureus* Skin Colonization and Infection. *Frontiers in immunology* **5**, 41

78. Terao, Y., Kawabata, S., Nakata, M., Nakagawa, I., and Hamada, S. (2002) Molecular characterization of a novel fibronectin-binding protein of *Streptococcus pyogenes* strains isolated from toxic shock-like syndrome patients. *J Biol Chem* **277**, 47428-47435

79. Terao, Y., Kawabata, S., Kunitomo, E., Murakami, J., Nakagawa, I., and Hamada, S. (2001) Fba, a novel fibronectin-binding protein from *Streptococcus pyogenes*, promotes bacterial entry into epithelial cells, and the *fba* gene is positively transcribed under the Mga regulator. *Mol Microbiol* **42**, 75-86

80. Amelung, S., Nerlich, A., Rohde, M., Spellerberg, B., Cole, J. N., Nizet, V., Chhatwal, G. S., and Talay, S. R. (2011) The FbaB-type fibronectin-binding protein of *Streptococcus pyogenes* promotes specific invasion into endothelial cells. *Cell Microbiol* **13**, 1200-1211

81. Marks, L. R., Davidson, B. A., Knight, P. R., and Hakansson, A. P. (2013) Interkingdom signaling induces *Streptococcus pneumoniae* biofilm dispersion and transition from asymptomatic colonization to disease. *MBio* **4**

82. Blanchette-Cain, K., Hinojosa, C. A., Akula Suresh Babu, R., Lizcano, A., Gonzalez-Juarbe, N., Munoz-Almagro, C., Sanchez, C. J., Bergman, M. A., and Orihuela, C. J. (2013) *Streptococcus pneumoniae* biofilm formation is strain dependent, multifactorial, and associated with reduced invasiveness and immunoreactivity during colonization. *MBio* **4**, e00745-00713

83. Marks, L. R., Mashburn-Warren, L., Federle, M. J., and Hakansson, A. P. (2014) *Streptococcus pyogenes* biofilm growth *in vitro* and *in vivo* and its role in colonization, virulence, and genetic exchange. *J Infect Dis*

84. Sumby, P., Zhang, S., Whitney, A. R., Falugi, F., Grandi, G., Graviss, E. A., Deleo, F. R., and Musser, J. M. (2008) A chemokine-degrading extracellular protease made by group A *Streptococcus* alters pathogenesis by enhancing evasion of the innate immune response. *Infect Immun* **76**, 978-985

85. Boyle, M. D., Raeder, R., Flosdorff, A., and Podbielski, A. (1998) Role of *emm* and *mrp* genes in the virulence of group A streptococcal isolate 64/14 in a mouse model of skin infection. *J Infect Dis* **177**, 991-997

86.   Nyberg, P., Sakai, T., Cho, K. H., Caparon, M. G., Fassler, R., and Bjorck, L. (2004) Interactions with fibronectin attenuate the virulence of *Streptococcus pyogenes*. *EMBO J.* **23**, 2166-2174

87.   Crotty Alexander, L. E., Maisey, H. C., Timmer, A. M., Rooijakkers, S. H., Gallo, R. L., von Kockritz-Blickwede, M., and Nizet, V. (2010) M1T1 group A streptococcal pili promote epithelial colonization but diminish systemic virulence through neutrophil extracellular entrapment. *J Mol Med (Berl)* **88**, 371-381

88.   Hollands, A., Pence, M. A., Timmer, A. M., Osvath, S. R., Turnbull, L., Whitchurch, C. B., Walker, M. J., and Nizet, V. (2010) Genetic switch to hypervirulence reduces colonization phenotypes of the globally disseminated group A *Streptococcus* M1T1 clone. *J Infect Dis* **202**, 11-19
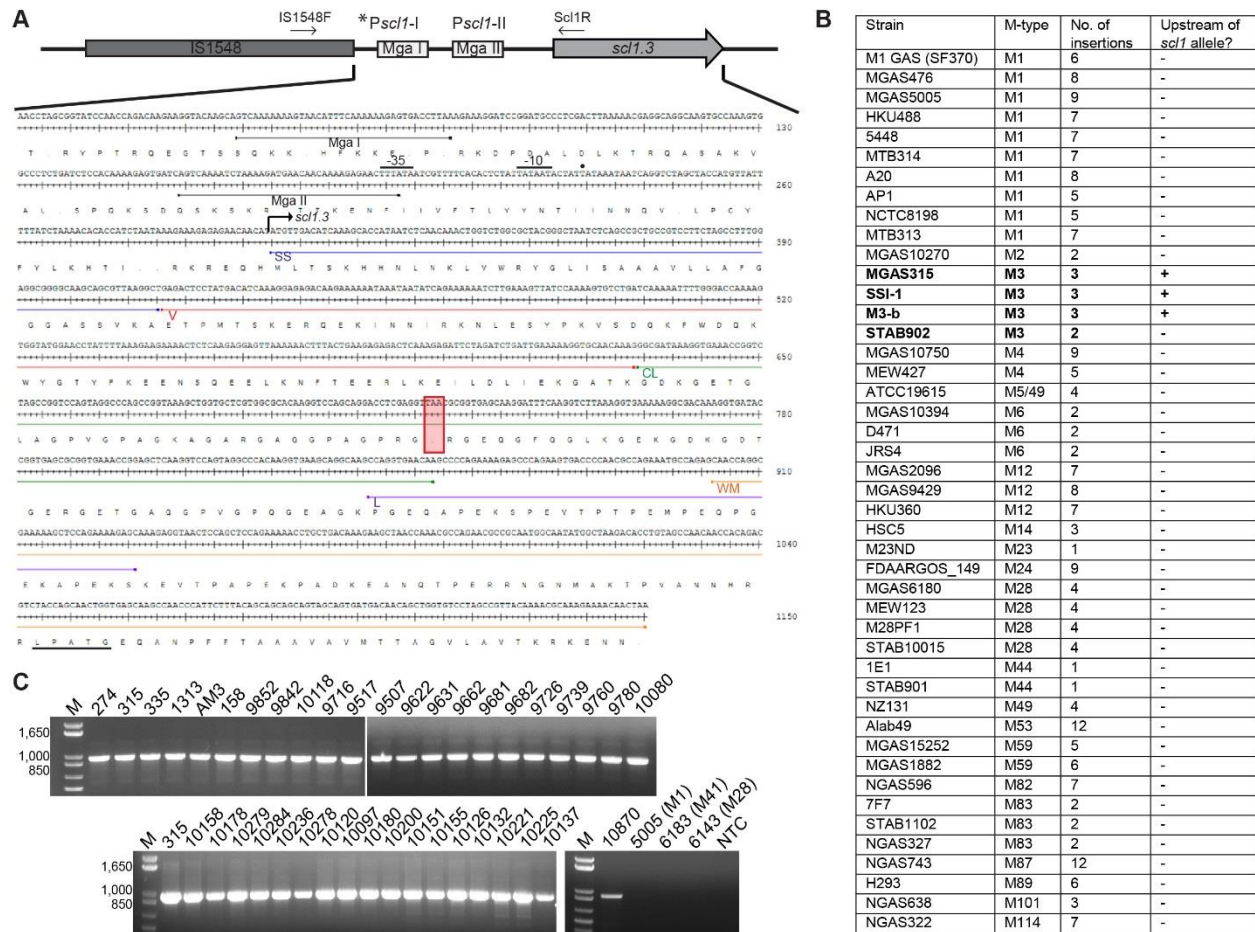
Figure 1

M3-type GAS strains harbor unique polymorphisms in the *scl1.3* locus. (A) Schematic representation and nucleotide sequence are based on the *scl1.3* locus in the M3-type strain MGAS315 genome. The *scl1.3* promoter region contains two putative Mga binding sites, P*scl1*-I with Mga I and P*scl1*-II with Mga II. IS1548 is inserted 38 bp upstream of the preferred *P*scl1*-I/Mga I promoter, which was shown to be necessary for *scl1* transcription. Transcriptional start site (solid dot), -10 and -35 boxes are shown upstream of *scl1.3* coding sequence. Scl1 regions are designated as follows: SS, signal sequence; V, variable region; CL, collagen-like region; L, linker region; WM, wall-membrane region; LPATG, cell-wall anchor. The null mutation in the 11th Gly-X-Y repeat of the CL region is depicted by the red box, presumably resulting in a truncated secreted Scl1.3 protein. Relative location of primers, IS1548F and Scl1R, used to generate amplicons in (C) are shown. **(B)** IS1548 insertion upstream of *scl1.3* is unique to M3 genomes. BLAST search in the NCBI nucleotide (nr/nt) database using IS1548 (1,317 bp) sequence as query identified insertions in 45 GAS genomes representing 21 different M-types. Only genomes of M3-type strains harbored IS1548 element upstream of the *scl1* allele (bold text). **(C)** IS1548 insertion upstream of *scl1.3* is conserved among M3 strains. Genomic DNA was isolated from a collection of 40 M3-type strains and analyzed by PCR for the presence of IS1548 upstream of *scl1* using primers IS1548F and Scl1R (located in conserved *scl1* signal sequence). Additional M1-, M41-, and M28-type control strains, and a no template control (NTC) are included. Expected amplicon size, 963 bp; M, 1 kb Plus DNA Ladder. MGAS designation applies to all strain numbers shown above gel wells, with the exception of strain AM3.

82
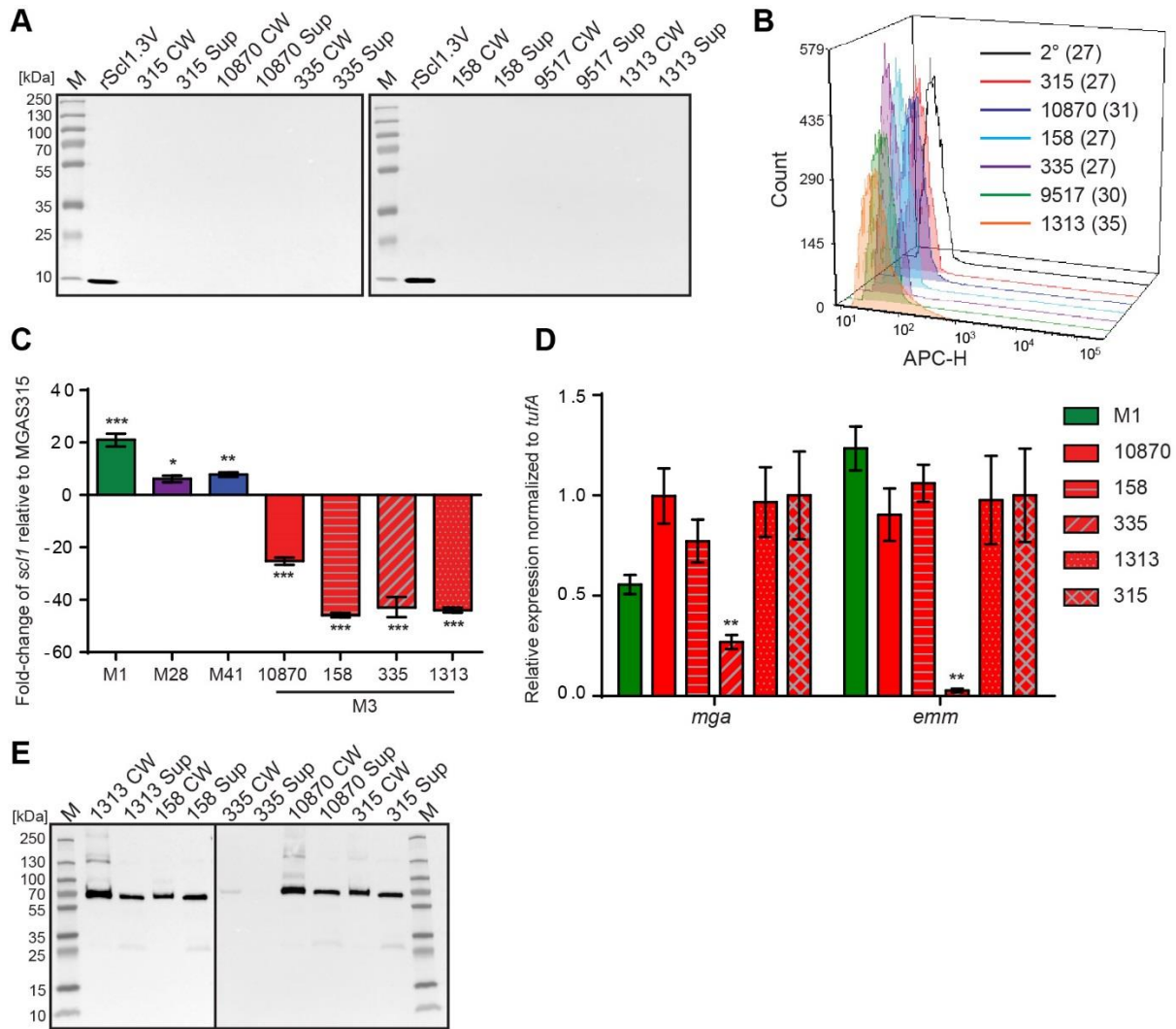
Figure 2

Assessment of Scl1.3 expression. (A) Assessment of Scl1.3 production by M3-type GAS. Cell wall (CW) and culture supernatant (Sup) protein fractions prepared from exponential phase cultures of several M3-type strains were analyzed by western immunoblotting, using anti-Scl1 rabbit polyclonal antibody. Recombinant protein rScl1.3V, corresponding to the variable region of Scl1.3, was used as a positive control. Expected molecular masses: Scl1.3, 11.4 kDa; rScl1.3V, 8.3 kDa. M, PageRuler™ Plus Prestained Protein Ladder. **(B)** Detection of Scl1.3 on the surface of M3-type GAS. Flow cytometry analysis of several M3-type strains is shown using anti-Scl1 antibody described in part (A) (color-shaded histograms) or a secondary only control (2° sample, black outlined histogram). Median fluorescence intensities (MFI) are shown in parentheses for each strain. **(C)** Assessment of *scl1* transcription. Fold-change of *scl1* transcript levels are shown compared to *scl1.3* transcription in M3-type strain MGAS315. qRT-PCR was performed on RNA obtained from exponential phase cultures. Results are shown from three independent experiments, each performed in triplicate wells. Standard errors and statistical analysis were computed from averaged ΔCt values for each biological replicate prior to normalization against the endogenous reference gene *tufA*; *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$ (student's *t*-test). **(D)**

83

Assessment of *mga* and *emm* transcription. Relative expression levels of *mga* and *emm* genes were compared between MGAS315 and four additional M3 strains or the M1 strain MGAS5005. Results are shown from three independent experiments, each performed in triplicate wells. Standard errors and statistical analysis were computed from averaged ΔCt values for each biological replicate prior to normalization against the endogenous reference gene *tufA*; **$P \leq 0.01$. **(E)** Assessment of M3-protein production by M3-type GAS. The same cell wall (CW) and culture supernatant (Sup) protein samples prepared from exponential phase cultures of M3-type strains (used in panel A) were analyzed by western immunoblotting, using anti-M3 protein rabbit polyclonal antibody. Expected molecular mass: 65 kDa. M, PageRuler™ Plus Prestained Protein Ladder.
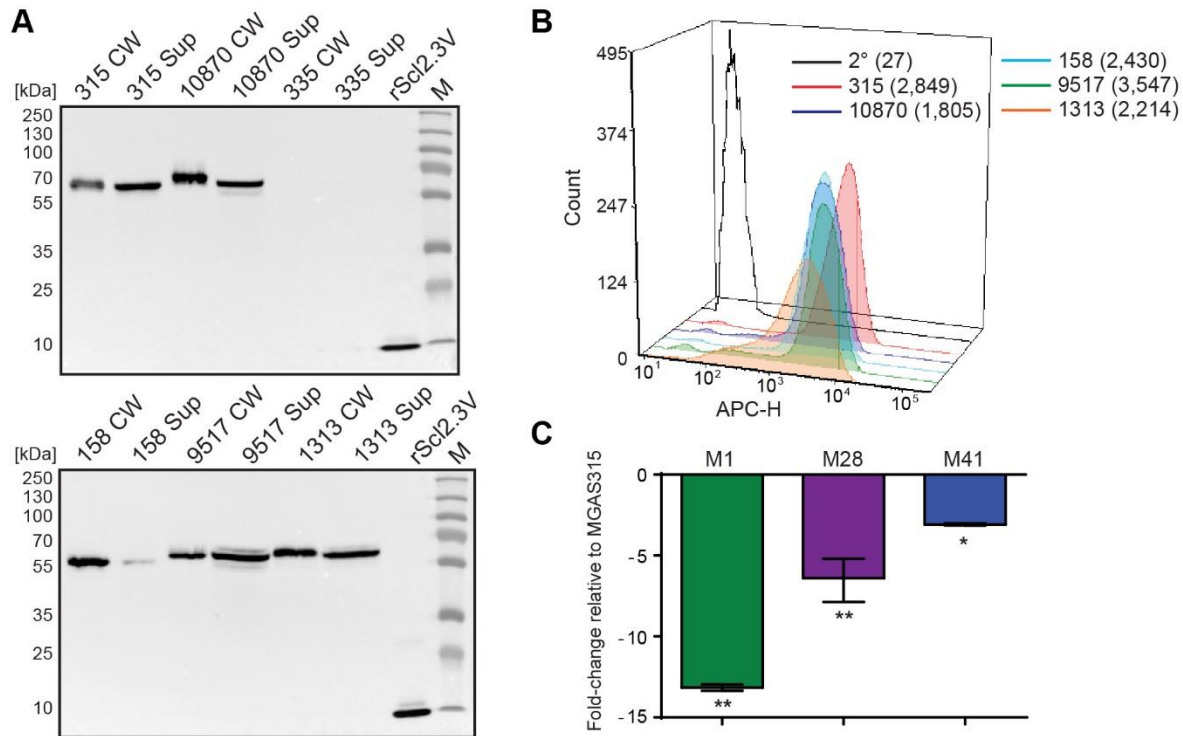
Figure 3

Characterization of the *scl2.3* locus in M3-type GAS. (A) Assessment of Scl2.3 production by M3-type GAS. The same cell wall (CW) and culture supernatant (Sup) protein samples prepared from exponential phase cultures of several M3-type strains (used in Fig. 2A, E) were analyzed by western immunoblotting, using anti-rScl2.3V rabbit polyclonal antibody. Recombinant protein rScl2.3V, corresponding to the variable region of Scl2.3 protein, was used as a positive control. Expected molecular masses based on MGAS315: Scl2.3, 52.5 kDa; rScl2.3V, 10.1 kDa. Aberrant migration of detected Scl2.3 variants is characteristic of Scl proteins. M, PageRuler™ Plus Prestained Protein Ladder. **(B)** Detection of Scl2.3 on the surface of M3-type GAS. Flow cytometry analysis of several M3-type strains is shown using anti-rScl2.3V rabbit polyclonal antibody (color-shaded histograms) or a secondary-only control (2° sample, black outlined histogram). Median fluorescence intensities (MFI) are shown in parentheses for each strain. **(C)** Assessment of *scl2* transcription. Fold-change of *scl2* transcription levels are shown compared to *scl2.3* transcription in M3-type MGAS315. qRT-PCR was performed on reverse-transcribed RNA obtained from exponential phase cultures. Results are shown from three independent experiments, each performed in triplicate wells. Standard errors and statistical analysis were computed from averaged ΔCt values for each biological replicate prior to normalization against the endogenous reference gene *tufA*; $*P \le 0.05$, $**P \le 0.01$ (student's *t*-test).
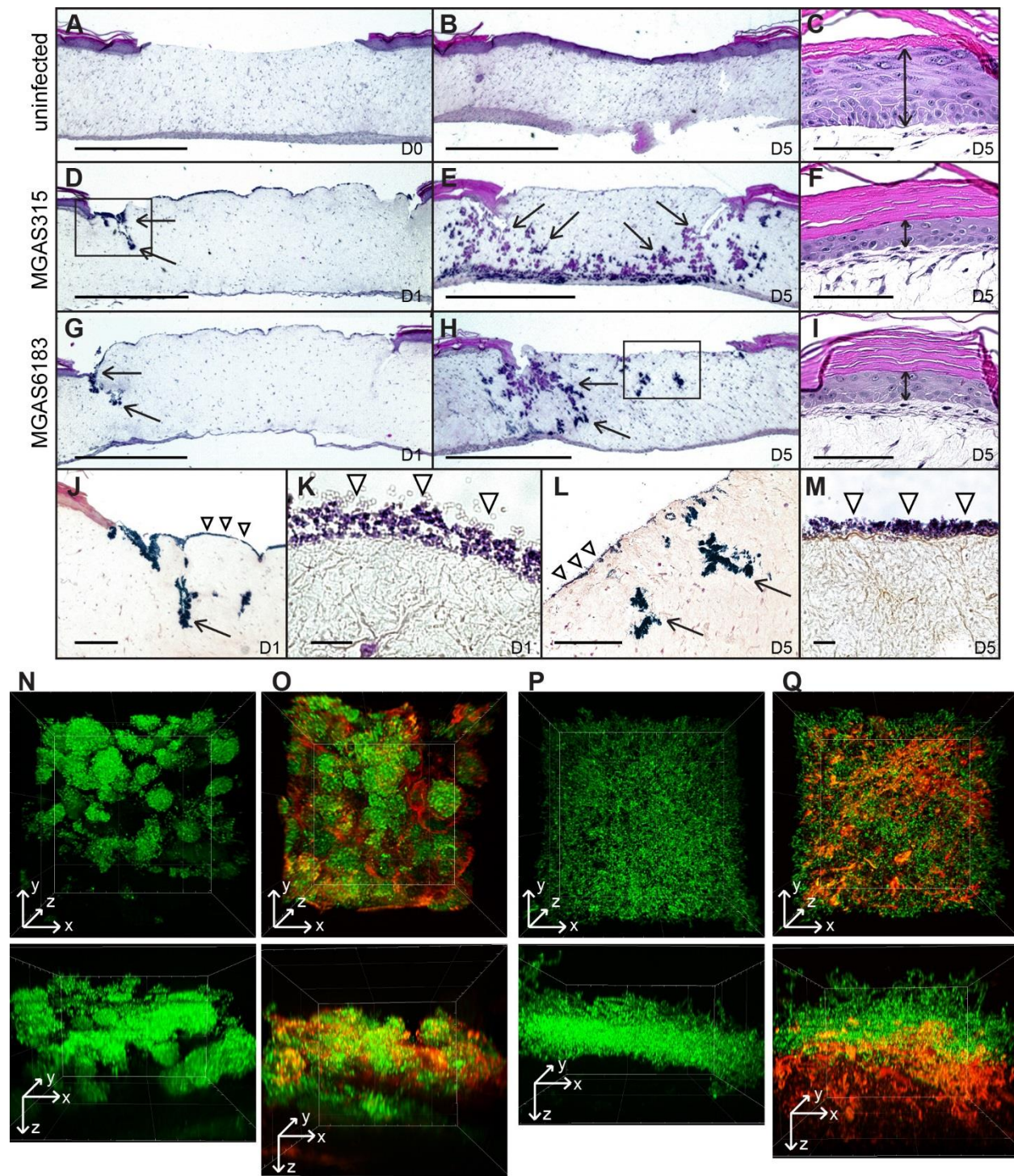
Figure 4

In vitro skin equivalent model of GAS infection. (A-I) H&E stained sections of uninfected (A-C) or infected (D-I) wounded skin equivalents at 40x magnification; scale bar: 1000 µm. Uninfected wound at day 0 (A) shows a lack of the epidermal layer where biopsy punch was

performed, which healed by day 5 (B). At day 1, infection of wounded skin equivalents with M3 and M41 GAS revealed superficial colonization of the wound bed, as well as invasion into the defects formed at the wound edge (D, G, arrows). By day 5, bacteria had disseminated throughout tissue laterally and vertically (arrows), reaching the bottom of the dermis layer (E, H). **(C, F, I)** H&E stained sections of the keratinocyte layer of intact skin surrounding the biopsy punch; scale bar: 100 µm. The thickness of the viable keratinocyte layer outside the wound (C, double arrow) was significantly reduced in tissues infected with M3 (F) and M41 (I) by day 5. **(J, L)** Gram stained sections of wounded tissue, corresponding to boxed areas in panels D and H, show tissue microcolonies (arrows) and superficial bacterial colonization (arrowheads). (J, scale bar: 400 µm; L, scale bar: 200 µm). **(K, M)** 1000x magnification micrographs of surface biofilms shown in J and L (scale bar: 10 µm). **(N-Q)** Two-photon fluorescence microscopy analysis of infected skin equivalents. Vertical dissemination through the wound bed and glycocalyx formation by the GFP-expressing M41- (N-O) and M3-type (P-Q) GAS at day 5. 3D projections of z-stacks are shown from the top view of the z-stack (top panels) or the side view (bottom panels). Multi-channel images show GFP fluorescence of GAS cells (N, P) and TRITC-conA stain of glycocalyx (O, Q). All images were acquired at 600x magnification.
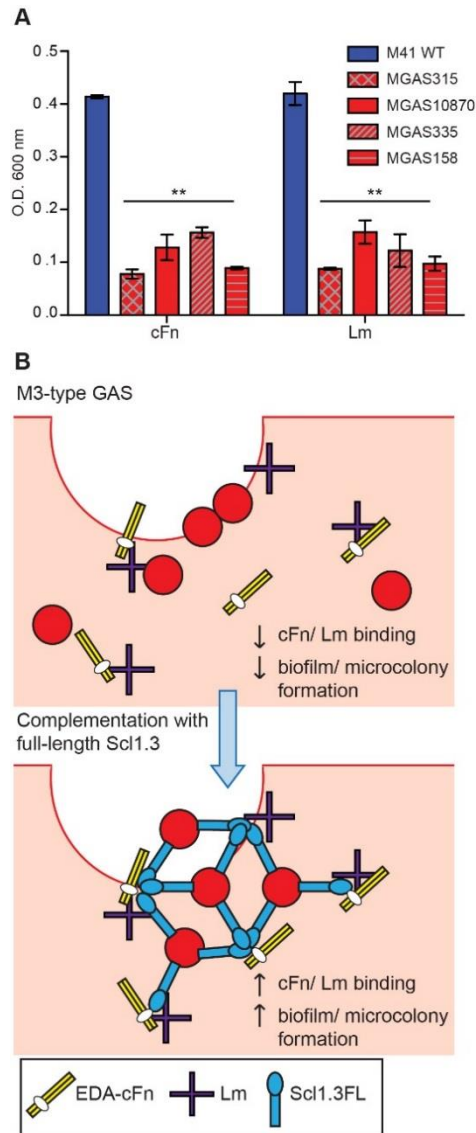
Figure 5

***In vitro* biofilm formation and hypothetical model of microcolony formation by invasive M3-type GAS. (A)** Limited *in vitro* biofilm formation by invasive M3 strains on coatings with cellular fibronectin and laminin. A panel of invasive M3-type strains was compared to a non-invasive biofilm-capable M41-type strain MGAS 6183. Crystal violet staining was used to assess biomass formed after 24 hours of growth in wells coated with cellular fibronectin (cFn) and laminin (Lm); results represent averaged values from at least 3 independent experiments performed in triplicate wells. **P≤0.01 (student's *t*-test). **(B)** Hypothesis model. Top: Infection of wounded skin with wild-type M3-type GAS. Inherent lack of surface-expressed Scl1.3 causes decreased binding to cFn and Lm expressed in wounded tissue, and reduces biofilm and tissue microcolony formation by M3-type bacteria (red circles). Bottom: *In-trans* complementation of M3-type GAS with full-length cell-associated Scl1.3, Scl1.3FL, restores binding to cFn and Lm in tissue, which confers biofilm and tissue microcolony formation during infection.
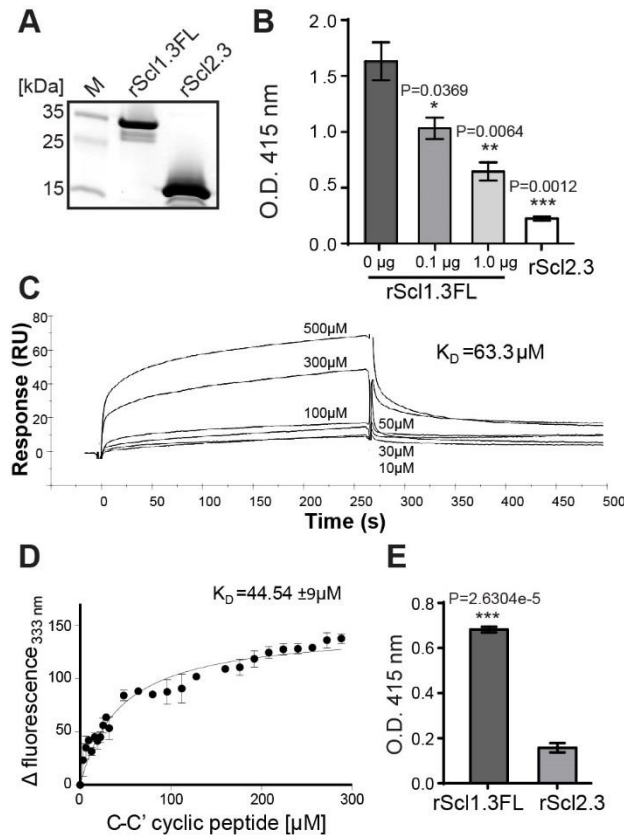
Figure 6

Construction and binding characterization of recombinant full-length Scl1.3FL and Scl2.3 proteins. (A) Affinity purified rScl1.3FL and rScl2.3 proteins were analyzed by 4-20% gradient SDS-PAGE; expected molecular masses: rScl1.3FL, 20.6 kDa and rScl2.3, 16.2 kDa. **(B-D-)** Binding of rScl1.3FL to extra domain A (EDA) of cellular fibronectin. **(B)** IST-9 antibody inhibition identifies rScl1.3FL binding to the C-C' loop of EDA-cFN. Inhibition of rScl1.3FL binding to EDA-cFn was tested by ELISA following pre-incubation of cFn with blocking IST-9 mAb specific to the C-C' loop. Significance of inhibition by 0.1 and 1.0 μg of IST-9 mAb was determined by student's *t*-test as compared to the untreated cFn. A lack of significant cFn binding by rScl2.3 was evident, as compared to rScl1.3FL-cFn binding; student's *t*-test. **(C)** rScl1.3FL binding to a cyclic peptide mimicking the C-C' loop of EDA using surface plasmon resonance. Overlay of sensorgrams for the interaction between immobilized rScl1.3FL and EDA-derived C-C' cyclic peptide is shown. The experimental curves corresponding to different concentrations of peptide (10-500 μM) were fitted according to a single binding model with 1:1 stoichiometry. **(D)** rScl1.3FL binding to a C-C' cyclic peptide using tryptophan fluorescence assay. Tryptophan fluorescence quenching analysis shows the dose-response curve of the fluorescence values of rScl1.3FL at 333 nm plotted against the concentration values of C-C' cyclic peptide. **(E)** Laminin binding to rScl1.3FL and rScl2.3 by ELISA. Recombinant rScl proteins were immobilized onto Strep-Tactin-coated wells and incubated with laminin. Bound laminin was detected with specific primary pAbs and HRP-conjugated secondary Abs. Laminin binding was compared between rScl1.3FL and rSc2.3 and evaluated statistically using student's *t*-test. Results for panels C and F represent averaged values from at least 3 independent experiments performed in triplicate wells. *P≤ 0.05, **P≤ 0.01, ***P≤ 0.001.

89
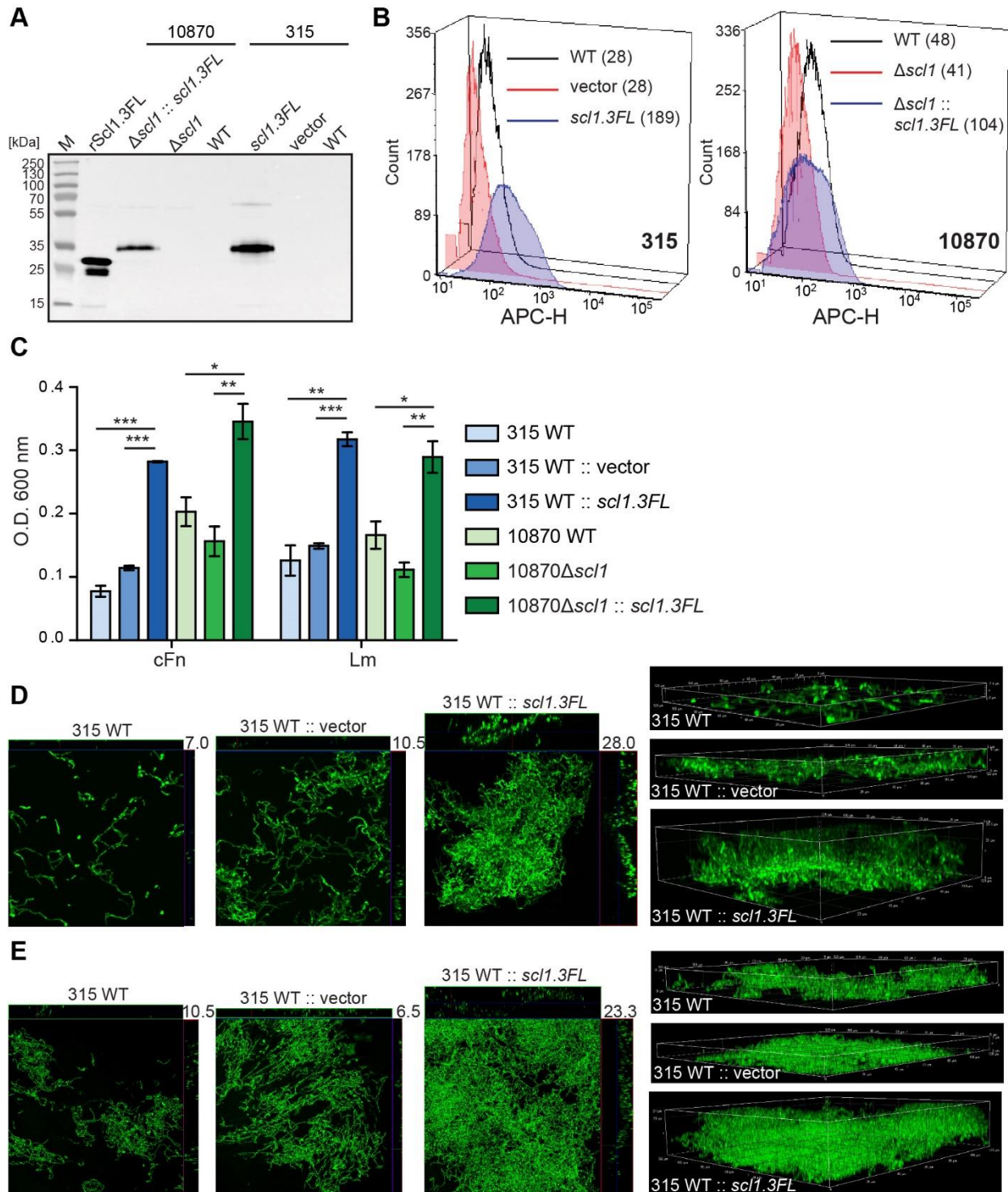
Figure 7

Homologous complementation of M3-type GAS with full-length surface-expressed Scl1.3 protein confers biofilm formation. (A) Western blot detection of full-length Scl1.3 in cell wall fractions of MGAS315 and MGAS10870Δ*scl1* complemented *in-trans* with the *scl1.3FL* allele.

Parent strains were included as negative controls and rScl1.3FL was used as a positive control for detection. **(B)** Flow cytometry detection of Scl1.3FL on the GAS cell surface. Left: Fluorescence intensity of Scl1.3FL-complemented MGAS315 WT strain was compared to vector-complemented or WT parent strain. Right: Fluorescence intensity of Scl1.3FL-complemented MGAS10870Δ*scl1* was compared to WT and Δ*scl1* parent strains. Median fluorescence intensities are shown in parentheses for each strain. **(C)** Crystal violet assessment of bacterial biomass after 24 hours of growth in cFn- or Lm-coated wells. MGAS315 WT complemented *in-trans* with pJRS525 (Spc^R vector) or pSL502 (**Table 1**) harboring the *scl1.3FL* allele was compared to WT parental strain. Spectinomycin resistant MGAS10870Δ*scl1* mutant was complemented *in-trans* with Kan^R pSL518 (**Table 1**), harboring the *scl1.3FL* allele. Biofilm biomass of the complemented strain was compared to MGAS10870 WT and *scl1*-inactivated parental strains. Results represent averaged values from at least 3 independent experiments performed in triplicate wells. *P≤ 0.05, **P≤ 0.01, ***P≤ 0.001; students *t*-test. **(D, E)** Confocal laser scanning microscopy analysis of biofilm formation by GFP-expressing MGAS315 vector and *scl1.3FL*- complemented strains. Biofilms were grown for 24 hours on cFn-coated (D) or Lm-coated (E) coverslips. Maximum intensity projections of GAS biofilms with cross-sectional views (left panels) are representative of z-stacks from ten fields within a single experiment. Average vertical thickness is indicated in micrometers. 3D projections of z-stacks (right panels) are shown from the side view.
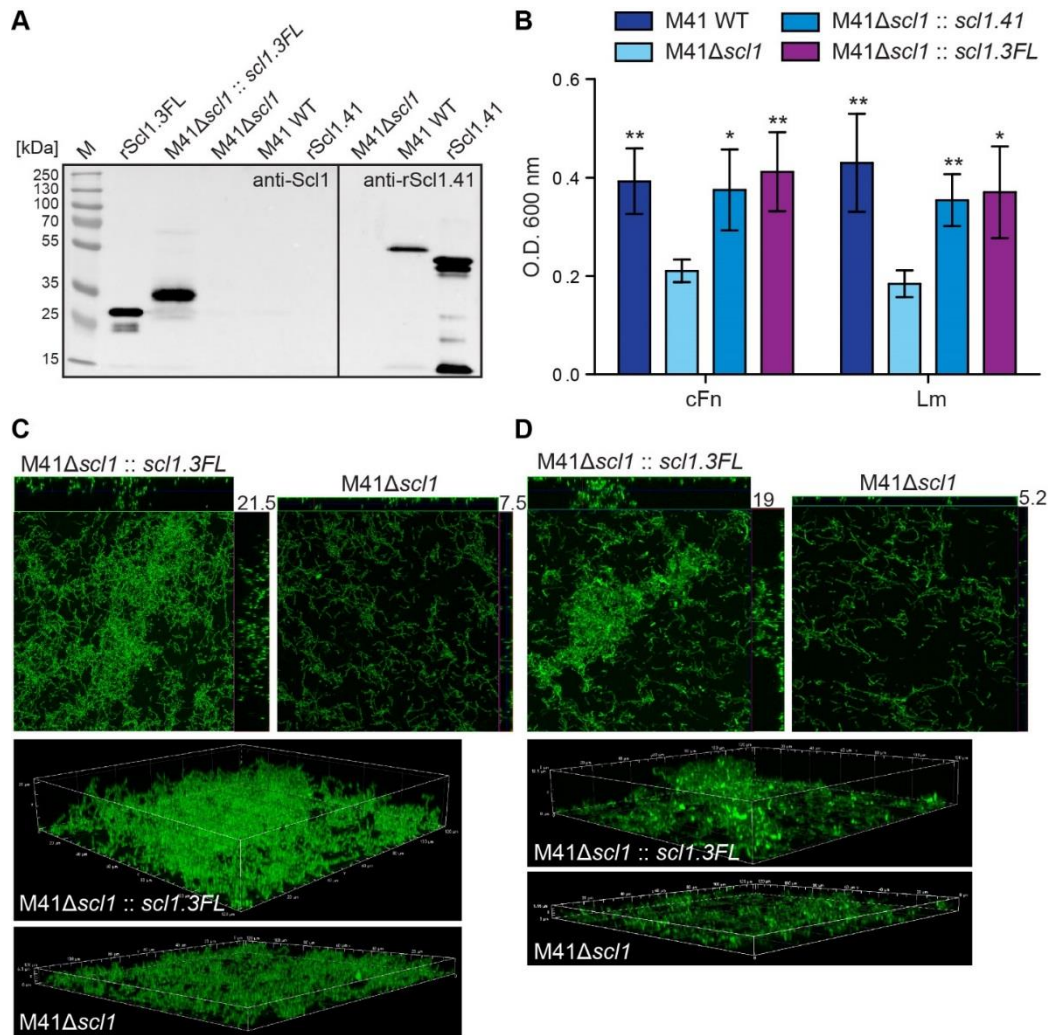
Figure 8

Heterologous complementation of *scl1*-deficient mutant of M41-type GAS with full-length Scl1.3 restores biofilm formation. (A) Left blot: western blot detection of full-length Scl1.3 protein in cell wall fraction of M41Δ*scl1*::*scl1.3FL* using anti-Scl1 polyclonal antibody. The M41Δ*scl1* and M41 WT strains were included as negative controls, and rScl1.3FL was included as a positive control. Right blot: western blot detection of Scl1.41 in M41Δ*scl1* and M41 WT strains using anti-rScl1.41 antibody and rScl1.41 as a positive control. **(B)** Crystal violet assessment of biofilm biomass after 24 hours of growth in cFn- or Lm-coated wells. M41 *scl1*-inactivated (M41Δ*scl1*) mutant was complemented *in-trans* with either homologous *scl1.41* allele harbored on pSL230 or with heterologous *scl1.3FL* allele harbored on pSL502 (**Table 1**). M41 WT, Δ*scl1* mutant and two complemented mutant strains were assessed for biofilm formation and statistical significance was assessed by students *t*-test as compared to M41Δ*scl1*; *P≤ 0.05, **P≤ 0.01. **(C, D)** Confocal laser scanning microscopy analysis of biofilm formation by GFP-expressing M41Δ*scl1* or M41Δ*scl1*:: *scl1.3FL*. Biofilms were grown for 24 hours on cFn-coated (C) or Lm-coated (D) coverslips. Maximum intensity projections of GAS biofilms with cross-sectional views (top panels) are representative of z-stacks from ten fields within a single experiment. Average vertical thickness is indicated in micrometers. 3D projections of z-stacks (bottom panels) are shown from the side view.
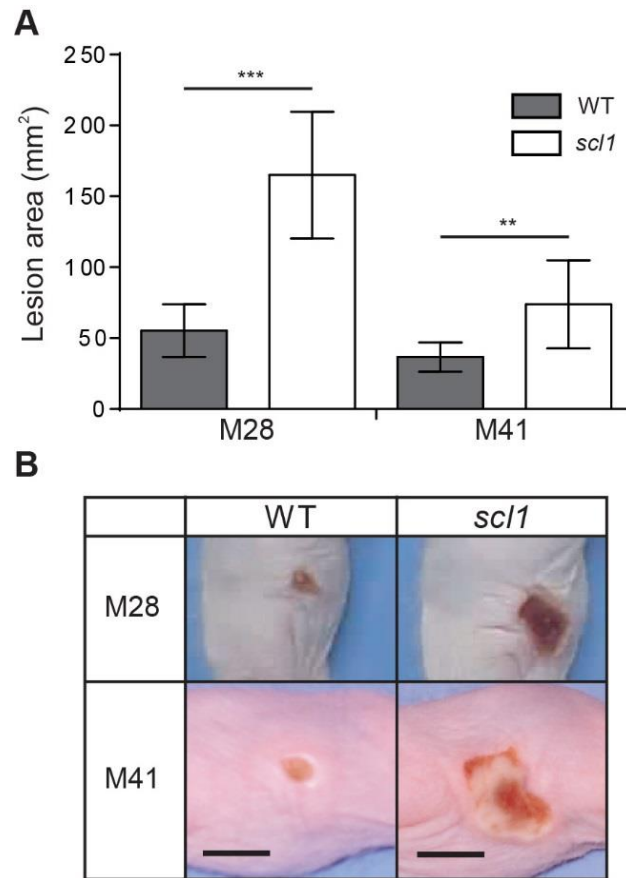
Figure 9

Skin pathology of SKH1 hairless mice infected subcutaneously with wild-type and *scl1*-inactivated mutants of M28- and M41-type GAS. (A) Mean area of skin lesions in mice infected with M28- or M41-type isogenic GAS strains. The severity of skin lesions was compared following infection with wild-type (WT) or *scl1*-mutant (*scl1*) GAS; 96-hour time-point is shown. The mean lesion area and standard deviation were calculated and compared between experimental groups. Data represents mean lesion size from at least 10 mice per experimental group. Statistical differences were calculated using the student's *t*-test; **$P \le 0.01$, ***$P \le 0.001$. **(B)** Gross pathology of the skin lesions. Digital images of the skin lesions of representative animals infected with either WT or *scl1*-mutant GAS. The images show skin lesions developed after 96 hours post-infection, and scale bar represents 10 mm.

93
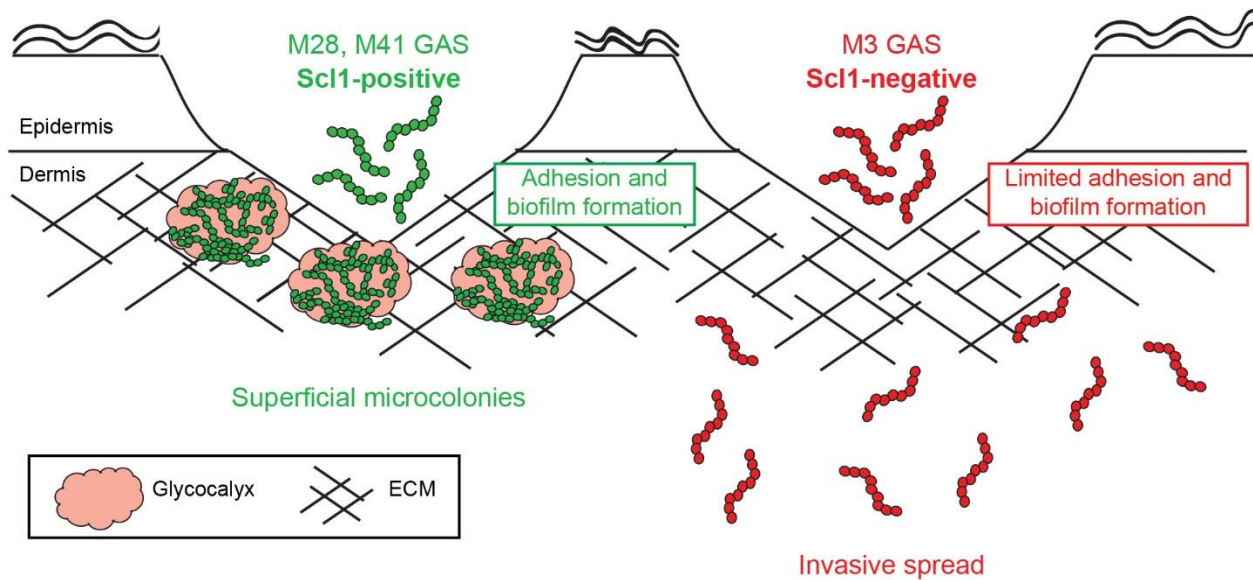
Figure 10

Model of Scl1-mediated GAS adhesion, biofilm formation, and host colonization. M28- and M41-type GAS express full-length cell-attached Scl1, which mediates adherence to cFn and Lm. ECM binding promotes biofilm formation in wounded tissue and allows the formation of superficial tissue microcolonies resulting in non-invasive colonization. On the contrary, in M3-type GAS, naturally lacking full-length cell-attached Scl1.3 adhesin, bacterial cells have reduced adherence to cFn and Lm and reduced capacity for biofilm formation, thereby circumventing the formation of localized tissue microcolonies; infection balance is shifted towards dissemination through tissue, promoting invasive spread.

Table 1. Constructs used in this study

Plasmid	Description	Source
pJRS525	*E. coli* / *S. pyogenes* shuttle vector	McIver and Scott. 1997
pSL230	pJRS525 with *scl1.41* allele	Caswell *et al.* 2007
pSL501	pJRS525 with *scl1.3WT* allele	Flores *et al.* 2015
pSL502	pSL501 with repaired allele *scl1.3FL*	Flores *et al.* 2015
pSL518	pSL502 with Kan^R in place of Spc^R	This study
pASK-IBA2	*E. coli* expression vector	IBA, Göttingen
pSL154	pASK-IBA2 encoding rScl1.3V	This study
pSL503	pASK-IBA2 encoding rScl1.3FL	Flores *et al.* 2015
pSL514	pASK-IBA2 with rScl2.3	This study
pSB027	GFP-encoding plasmid	Cramer *et al.* 2003

CHAPTER 3

NATURAL VARIANT OF COLLAGEN-LIKE PROTEIN A IN SEROTYPE M3 GROUP A *STREPTOCOCCUS* INCREASES ADHERENCE AND DECREASES INVASIVE POTENTIAL

Anthony R. Flores, Brittany E. Jewell, Erika M. Versalovic, Randall J. Olsen, Beth A. Bachert, Slawomir Lukomski and James M. Musser

ABSTRACT

Group A *Streptococcus* (GAS) predominantly exists as a colonizer of the human oropharynx that occasionally breaches epithelial barriers to cause invasive diseases. Despite the frequency of GAS carriage, few investigations into the contributory molecular mechanisms exist. To this end, we have identified a naturally occurring polymorphism in the gene encoding the streptococcal collagen-like protein A (SclA) in GAS carrier strains. All previously sequenced invasive serotype M3 GAS possess a premature stop codon in the *sclA* gene truncating the protein. The carrier polymorphism is predicted to restore SclA function and was infrequently identified by targeted DNA sequencing in invasive strains of the same serotype. We demonstrate that a strain with the carrier *sclA* allele expressed a full-length SclA protein while the strain with the invasive *sclA* allele expressed a truncated variant. An isoallelic mutant invasive strain with the carrier *sclA* allele exhibited decreased virulence in a mouse model of invasive disease and decreased multiplication in human blood. Further, the isoallelic invasive strain with the carrier *sclA* allele persisted in the mouse nasopharynx and had increased adherence to cultured epithelial cells. Repair of the premature stop codon in the invasive *sclA* allele restored the ability to bind the extracellular matrix proteins laminin and cellular fibronectin. These data demonstrate that a mutation in GAS carrier strains increases adherence and decreases virulence and suggest selection against increased adherence in GAS invasive isolates.

INTRODUCTION

The bacterial pathogen *Streptococcus pyogenes* (group A *Streptococcus*, GAS) causes a wide range of disease in humans. GAS causes severe invasive infections such as toxic shock syndrome and necrotizing fasciitis but is also the cause of milder, more benign infections (e.g. pharyngitis). In addition to causing disease, GAS colonizes the throats of humans in the absence of symptoms. Asymptomatic colonization rates range between 5-15% in healthy children (1), a rate that far exceeds that of GAS invasive disease (2). However, despite the high prevalence of GAS carriage, little is known about the molecular factors that contribute to asymptomatic colonization.

Colonization of the host epithelium is a key first step to establishing GAS carriage or disease. GAS elaborates several key surface proteins that contribute to this process including the M protein, fibronectin binding proteins, the GAS pilus, and streptococcal collagen-like (Scl) proteins. The collagen-like protein SclA (also known as Scl1) is found in all GAS serotypes examined to date (3) and is positively regulated by the well-characterized regulator Mga (4). The SclA protein extends from the GAS cell surface in a homotrimeric, 'lollipop-like' fashion (5). The outermost region consists of a globular head that varies considerably between GAS serotypes followed by a repeating Gly-X-Y (GXY) collagen-like sequence that is in turn linked to the cell wall through a proline-rich linker region (Figure 1A). SclA contributes to GAS adherence and colonization through binding cellular fibronectin and laminin via the variable globular head (6) and integrins through the collagen-like region (7,8). Variability in the degree of binding between GAS serotypes to these and other host molecules is predicted based on the differences in the V and CL regions of SclA.

While several studies have begun to unravel the role of SclA as an adhesin, its contribution to phenotypic differences based on source of GAS isolation (e.g. invasive versus carrier) is unknown. Expression of GAS surface proteins may enhance adherence to host surfaces but may also reduce virulence. Expression of pili in serotype M1 GAS promotes adherence but enhances GAS killing through neutrophil extracellular traps (NETs) reducing systemic virulence (9). Similarly, the fibronectin-binding protein PrtF1 increased adherence to human cells but decreased dissemination in a mouse model of GAS infection (10). Similar paradoxical observations exist for surface proteins in

Staphylococcus aureus (11). In contrast to other GAS serotypes, it has been shown that the *sclA* gene in serotype M3 GAS harbors an internal stop codon truncating the predicted protein (Figure 1A) (12). In fact, the truncated *sclA* gene was identical in recent whole genome sequencing of >200 invasive serotype M3 GAS (13), suggesting selection against a full-length, surface-attached protein. The impact of the truncated *sclA* allele on the pathogenesis of invasive serotype M3 GAS remains unknown.

Here, we report that human carrier strains of serotype M3 GAS contain a natural variant of *sclA* that restores the full open-reading frame. Expression of the full-length SclA in an invasive strain correlates with increased adherence to epithelial cells *in vitro* and decreases virulence in invasive disease in mice. The data presented also suggest that the truncated *sclA* allele contributes to pathogenesis in invasive serotype M3 GAS. Our findings suggest that mutations negatively affecting virulence and enhancing colonization of mucosal surfaces may contribute to asymptomatic carriage in GAS.

MATERIALS AND METHODS

Bacterial strains used and culture conditions

GAS strains are listed in Table S1 and were grown on trypticase soy agar containing 5% sheep blood agar (SBA) (Becton Dickinson), in Todd-Hewitt broth containing 0.2% (wt/vol) yeast extract (THY) (Difco Laboratories), or on THY agar. When needed, GAS media was supplemented with chloramphenicol (Sigma-Aldrich) at 10 μg/mL or spectinomycin (Sigma-Aldrich) at 150 μg/ml. *E. coli* DH5α or TOP10 (Invitrogen) grown in Luria-Bertani (LB) broth or on LB agar (Difco Laboratories) were used for cloning and supplemented with ampicillin (Sigma-Aldrich) at 100 μg/mL or chloramphenicol (Sigma-Aldrich) at 20 μg/mL when appropriate.

PCR amplification and sequencing of *sclA*

Primers for PCR and sequencing of *sclA* in GAS strains are listed in Table S2. Growth of strains, genomic DNA extraction, PCR amplification and Sanger sequencing were performed as previously described (14).

Generation of isoallelic and isogenic mutants in MGAS10870

Plasmids and primers used in this study are listed in Tables S1 and S2. The procedure used in generating the isoallelic mutant MGAS10870*sclA*^{Carrier} was as described (15). Briefly, the *sclA*^{Carrier} allele was amplified using the primers 2902F and 2487R from the carrier strain MGAS23440 and subsequently ligated into the temperature sensitive *E. coli*-Gram positive shuttle vector pJL1055 using *Bam*HI and *Xho*I (16) to generate pJSF41. Electrocompetent cells of MGAS10870 were transformed with pJSF41 and allelic replacement carried out as previously described (17). The isoallelic mutant was confirmed using Sanger sequencing (Applied Biosystems).

Insertional inactivation of *sclA* in MGAS10870 was performed as previously described (18). Briefly, a three-step process was used to generate a PCR fragment in which 810 bases of the *sclA* open reading frame (ORF) were removed. Primer pair 0488R and MSP184 were used to generate a 824-bp fragment including the first 18-bp of the *sclA* ORF. Likewise, primer pair 2901F and MSP183 were used to generate a 791-bp fragment containing the last 18-bp of the *sclA* ORF. Primer pair MSP181 and MSP182 were used to amplify *aad9* with ends homologous to the first and last 18-bp of the *sclA* ORF. The three PCR fragments were then used to generate a 2,581-bp fragment using primer pair 2901F and 0488R in a ligation PCR in which all but 36-bp (12 amino acids) of the *sclA* gene were replaced, in-frame, by the spectinomycin cassette *aad9*. All isoallelic and isogenic mutants were confirmed by Sanger sequencing (Applied Biosystems).

RNA isolation and quantitative real-time PCR analysis

Transcript levels of *sclA* were assayed as previously described (15). Briefly, RNA was isolated and purified using an RNeasy kit according to the manufacturer's instructions (Qiagen). Reverse transcription of RNA to produce cDNA was done with a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). TaqMan quantitative real-time PCR (qQRT-PCR) was performed with an ABI 7500 Fast Real-Time and *tufA* used as the endogenous control gene. TaqMan primers and probes used in analyses are listed in Table S2. All reactions were performed in quadruplicate using RNA purified from at least three biologic replicates.

Generation of anti-SclA rabbit antibodies and western immunoblot analysis of SclA

The *sclA*[Invasive] gene was PCR amplified using primers MSP263 and MSP264 (Table S2) and cloned into the expression vector pET15b (Novagen) to generate pJSF66. The His-tagged SclA[Invasive] protein was overexpressed in *E. coli* (BL21) and purified using nickel affinity chromatography. Affinity-purified rabbit polyclonal anti-SclA antibody was generated by LifeTein (Hillsborough, NJ).

Complementation (*in trans*) of the isogenic deletion mutant MGAS10870Δ*sclA* utilized the Gram-positive expression vector pDC123 (19). The primer pair MSP275 and MSP277 were used to amplify the *sclA* gene from either MGAS10870 (*sclA*[Invasive]) or MGAS23440 (*sclA*[Carrier]) and ligated into the *Eco*RV/*Hind*III sites of pDC123. The resulting plasmids (Table S1) were electroporated into MGAS10870Δ*sclA* to generate the individually transcomplemented strains. Isolation of bacterial proteins and western immunoblot analysis was carried out as described (15).

Cloning, expression, and purification of full-length SclA (SclA[M3-FL])

All plasmids and primers are listed in Tables S1 and S2. The *sclA* coding region of the serotype M3 strain MGAS315 was amplified by PCR using the forward primer 232 Up and the reverse primer ME7, digested with *Eco*RI and *Hpa*I enzymes, and, cloned into the *E. coli*/GAS shuttle vector pJRS525. The resulting plasmid, pSL501 served as a template for PCR mutagenesis to obtain the full-length *sclA* allele (pSL502). The forward PCR primer pJRS525F1 was used with the reverse primer ME6, which contains a single base pair substitution encoding a glutamate residue (GAA) instead of the stop codon (TAA) naturally present in the 11th GXY repeat of *sclA* wild-type allele.

The recombinant full-length SclA protein (rSclA[M3-FL]) was generated employing the *Strep*-tag® II expression and purification system (IBA-GmbH, Goettingen, Germany), as described previously (5,20). PCR amplification was performed using pSL502 DNA template, with the primers Scl1.3 M3VF and Scl1.28WMR. PCR product was digested with *Hind*III and *Eae*I enzymes and cloned into the pASK-IBA2 *E. coli* vector designed for periplasmic expression, resulting in pSL503 encoding rSclA[M3-FL]. Each clone, pSL501-503, was confirmed by sequencing. Protein rSclA[M3-FL] was affinity purified on *Strep*-Tactin® sepharose, dialyzed against 25 mM HEPES, pH 8.0, and analyzed on SDS-

PAGE stained with RAPID*stain*^{TM} (G-Biosciences). Protein identity was confirmed by N-terminal Edman degradation (Iowa State University, Protein Facility).

Mouse virulence experiments

Female CD1 mice (Harlan Laboratories) were used for virulence studies as described previously (21). Mice were inoculated in the right hind limb with 1 x 10^7 CFU of GAS in 100 µl PBS for intramuscular infection or intranasally with 1 x 10^7 CFU in 50 µl PBS for nasopharyngeal colonization. Near-mortality was determined by observation using pre-defined criteria for mice infected intramuscularly (21). Mouse nasopharyngeal colonization was performed as previously described (22). All mouse experiments were approved by the Institutional Animal Care and Use Committee of Houston Methodist Research Institute.

Growth in human blood

Experiments assessing the ability to grow in human blood were conducted under a Houston Methodist Research Institute Institutional Review Board human subjects protocol and carried out as described (22,23). A minimum of two healthy, non-immune, adult donors were used for each experiment.

Human epithelial cell adherence assays

GAS adherence to cultured epithelial cells was carried out as previously described (22). Assays were performed in quadruplicate. Percentage adherence was calculated by dividing the recovered CFU by the original inoculum.

ECM binding assay

To assess ECM binding by ELISA, recombinant Scl (rScl) proteins (0.5 µM) were immobilized onto *Strep*-Tactin® coated microplate wells (IBA-GmbH), as described before (6,24). The cellular fibronectin (cFn) from human foreskin fibroblasts (F2518, Sigma), and laminin (Lm) (23017-015, Invitrogen) were added to wells containing rScls at 1 µg per well. Rabbit anti-human fibronectin (F3648, Sigma) and anti-laminin (L9393, Sigma) primary antibodies were used at concentrations of 1:1000 and 1:250, respectively.

Secondary antibody goat anti-rabbit IgG (H+L) HRP conjugate (BioRad), was used at a concentration of 1:2000, followed by addition of the HRP substrate 2,20-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) (37615, 1-step ABTS, ThermoScientific); absorption was measured at a wavelength of 415 nm. OD values from BSA-coated control wells were subtracted from the OD values of each test well. Shown are combined results from three independent experiments performed in triplicate.

Statistics

A two-tailed *t*-test (unequal variance) was used to compare multiplication factors between strains grown in human blood, adherence of strains to epithelial cells, and ECM binding of recombinant Scl proteins. Kruskal-Wallis analysis of variance was used to compare rates of nasopharyngeal colonization between strains. A *P*-value less than 0.05 was considered significant for all statistical tests.

RESULTS

The variant *sclA* allele occurs infrequently in serotype M3 GAS invasive strains and results in expression of a full-length protein

In an effort to identify key molecular genetic differences contributing to asymptomatic carriage of GAS, we performed whole genome sequencing of 37 serotype M3 GAS isolates cultured serially from 9 individuals over time (22). We discovered that 4 strains recovered from one subject contained a variant *sclA* allele compared to the reference serotype M3 genome. The variant carrier allele (*sclA*Carrier) differed from all previously sequenced serotype M3 *sclA* (*sclA*Invasive) by a 153-bp deletion removing the internal stop codon within the collagen-like region and restoring the complete open-reading frame (Figure 1A). The predicted SclACarrier protein has highest homology (>71% identity) to the SclA proteins from serotypes M6 and M5 GAS.

 All invasive serotype M3 GAS strains studied to date possess the *sclA* allele with an internal stop codon (*sclA*Invasive) (13). The predominance of the truncated *sclA* allele in invasive serotype M3 GAS suggests host selective pressure against expression of the full-length SclA protein in invasive disease. However, the possibility exists that variant *sclA* alleles restoring the open-reading frame and potentially expressing a complete SclA

protein are found in more geographically diverse invasive serotype M3 GAS. We hypothesized that the frequency of variation in *sclA* would be substantially less in a population of invasive serotype M3 GAS than that found in the carrier population examined. To test our hypothesis we performed targeted DNA sequencing of *sclA* using strains obtained from the Centers for Disease Control Active Bacterial Core surveillance spanning the years 1998 to 2008 (http://www.cdc.gov/abcs) (14). Very few (6/479, 1.3%) of the geographically and temporally diverse invasive serotype M3 GAS isolates harbored the same 153-bp deletion in *sclA* as identified in the GAS carrier strains. The *sclA* allele in the remaining invasive strains contained the internal stop codon as in the reference serotype M3 genome (MGAS315).

Given that the *sclA*[Invasive] allele contains an internal stop codon, it is possible invasive strains do not produce *sclA*[Invasive] transcript or that it is unstable and quickly degraded. We previously showed that peak *sclA* transcript coincides with peak *mga* transcript levels in serotype M3 GAS (15). Thus, we hypothesized that *sclA* transcript would be produced in a strain with either *sclA*[Invasive] or *sclA*[Carrier]. To test this hypothesis, we generated an isoallelic mutant strain that differed from the invasive parental strain MGAS10870 only by the presence of the *sclA*[Carrier] gene (MGAS10870*sclA*[Carrier]). We subsequently assayed *sclA* transcript *in vitro* after growth in rich medium (THY). As predicted, we detected *sclA* transcript and observed no significant differences between the isoallelic mutant and parental strains (Figure 1B).

We next hypothesized that the full-length SclA[Carrier] protein is expressed in GAS. To test this hypothesis, we complemented *in trans* the invasive strain MGAS10870 lacking completely the *sclA* gene (MGAS10870Δ*sclA*) with a plasmid expressing either *sclA*[Carrier] or *sclA*[Invasive] under the native *sclA* promoter. Total protein of the knockout and *trans*-complemented strains were probed using a polyclonal anti-SclA antibody generated using purified SclA[Invasive] protein (Figure 1A). Consistent with our hypothesis, we identified the SclA[Carrier] protein at the predicted size (~ 21 kDa) in MGAS10870pDC*sclA*[Carrier] (Figure 1C). In addition, we also identified an approximately 12-kDa SclA[Invasive] protein, consistent with the predicted protein size based on the premature stop codon, in MGAS10870pDC*sclA*[Invasive] (Figure 1A). Thus, the variant *sclA* allele identified in the

serotype M3 carrier strains restores expression of a full-length protein, however, a shorter, truncated version of SclA is also expressed in invasive serotype M3 GAS.

A serotype M3 invasive strain with *sclA*^{Carrier} has reduced virulence.

Inasmuch as expression of cell surface adhesion proteins has been shown to reduce virulence in GAS (9,10), we next tested the hypothesis that invasive serotype M3 GAS with *sclA*^{Carrier} will have decreased virulence compared to invasive strains with *sclA*^{Invasive}. To test this hypothesis, we used a mouse model of necrotizing fasciitis to compare the parental MGAS10870 to the isoallelic mutant MGAS10870*sclA*^{Carrer}. Compared to the parental invasive strain the isoallelic mutant MGAS10870*sclA*^{Carrier} formed a localized, abscess-like lesion on visual inspection of the infected limb at 48h (Figure 2A). Microscopic examination revealed extensive myonecrosis in the parental invasive strain while the isoallelic mutant with *sclA*^{Carrier} showed an abscess-like lesion with surrounding healthy, viable tissue (Figure 2A). This observation is consistent with decreased dissemination of the strain with *sclA*^{Carrier} within the infected tissue. In contrast to the isoallelic carrier mutant, MGAS10870 lacking *sclA* (MGAS10870Δ*sclA*) showed no significant difference following visual and microscopic examination compared to the parental strain.

Given that human antibodies against Scl proteins cross-react with Scl of differing serotypes (25), it is possible that surface expression of SclA enhances phagocytosis of GAS. To further test our hypothesis that the strain harboring *sclA*^{Carrier} reduces GAS virulence we next assessed the ability of the mutants to grow *ex vivo* in human blood. Consistent with differences in mouse tissue destruction, we observed a significantly decreased ability of MGAS10870*sclA*^{Carrier} to grow in human blood but no such difference in MGAS10870Δ*sclA* compared to MGAS10870 (Figure 2B), suggesting enhanced phagocytosis in the mutant with the carrier allele. Similar to previous studies of GAS surface adhesins reducing virulence (9,10), both *in vivo* mouse and *ex vivo* human data confirm a decreased virulence phenotype in an invasive strain expressing SclA^{Carrier}. In addition, deletion of the *sclA*^{Invasive} allele in an invasive strain had no significant effect on virulence in the models tested.

Invasive strain harboring *sclA*[Carrier] persists in mouse nasopharyngeal colonization and demonstrates increased adherence to human epithelial cells.

SclA has previously been shown to contribute to adherence to epithelial cells (8). Furthermore, others have hypothesized that GAS carrier strains are characterized by increased ability to adhere to epithelial surfaces (26). Thus, if our hypothesis is correct and SclA[Carrier] increases adherence, the invasive strain with *sclA*[Carrier] should have increased ability to colonize epithelial surfaces. We first tested the invasive strain MGAS10870 and the isoallelic mutant with *sclA*[Carrier] in a mouse nasopharyngeal model of colonization. We observed a significant decrease in the proportion of mice colonized with the isoallelic mutant MGAS10870*sclA*[Carrier] compared to MGAS10870 (Figure 3A). However, mice were stably and persistently colonized by the isoallelic mutant MGAS10870*sclA*[Carrier] at a level similar to the parental invasive strain at day 14 (Figure 3A) suggesting an ability to persist on the mouse nasopharyngeal mucosa. We also observed a significant decrease in mice colonized by the deletion mutant MGAS10870Δ*sclA* (Figure 3A) compared to the parental strain. In addition to decreased colonization, both the carrier (MGAS10870*sclA*[Carrier]) and deletion (MGAS10870Δ*sclA*) mutants had significantly decreased bacterial burden compared to the parental invasive strain (Figure 3B). The finding of decreased bacterial burden in the mouse nasopharyngeal model is strikingly similar to that observed in GAS carrier strains deficient in capsule production (22).

To further characterize adherence in MGAS10870*sclA*[Carrier] we performed *in vitro* adherence assays to cultured epithelial cells. Given the known role of SclA in adherence to epithelial surfaces (6), we hypothesized that compared to the parental invasive strain, the strain with *sclA*[Carrier] would better adhere to cultured epithelial cells. We assayed the *in vitro* adherence to both a pharyngeal epithelial cell line (HEp-2) and a skin epithelial cell line (HaCaT). Compared to the invasive strain, we observed significantly greater adherence to both HEp-2 and HaCaT cells in the isoallelic strain MGAS10870*sclA*[Carrier] and the deletion mutant MGAS10870Δ*sclA* (Figure 3C-D). Our data are consistent with the model that presence of the naturally-occurring variant *sclA*[Carrier] in serotype M3 GAS results in decreased virulence, stable and persistent colonization of the mouse nasopharynx, and increased ability to adhere to cultured human epithelial cells. Further,

while deletion of *sclA*[Invasive] did not affect virulence in the mouse intramuscular or *ex vivo* human models of GAS disease, the observed phenotypic differences in colonization and adherence suggest biologic activity of the truncated and secreted SclA[Invasive] protein.

Full-length serotype M3 SclA protein (SclA[M3-FL]) binds extracellular matrix components, cellular fibronectin and laminin

Several recombinant SclA (rSclA) constructs derived from GAS of different M protein serotypes have been shown to bind human extracellular matrix (ECM) components such as cellular fibronectin (cFn) and laminin (Lm) (6,24). Since binding is mediated by the SclA variable (V) region, we hypothesized that the observed increase in adherence in MGAS10870*sclA*[Carrier] was at least in part due to the expression of a cell-attached, full-length SclA protein. To test our hypothesis, we cloned the *sclA* coding region from invasive serotype M3 GAS and repaired the nonsense mutation, generating a GAA codon encoding a glutamate residue in place of the TAA premature stop codon (Fig. 1A). The resulting full-length *sclA* allele (*sclA*[M3-FL]) was cloned into an expression vector to generate full-length recombinant SclA protein (rSclA[M3-FL]), purified (Fig. 4A), and subsequently tested for binding to cFn and Lm by ELISA. SclA[M3-FL] demonstrated significantly greater binding to both cFn and Lm compared to SclB[M28] ($P < 0.05$), an ECM-binding negative control derived from SclB protein of serotype M28 GAS (Fig. 4B). Binding of SclA[M3-FL] was similar to recombinant SclA[M1] and SclA[M41] proteins, derived from serotype M1 and M41 GAS, respectively, and included as ECM-binding positive controls. Our data indicate that the full-length cell-attached serotype M3 SclA protein and thus the carrier SclA protein function as adhesins able to bind cFn and Lm in human tissue. Together, these data suggest that serotype M3 GAS acquired the unique null mutation resulting in secretion of a truncated SclA variant rather than cell-attached protein. Consistent with the mouse intramuscular infection model, lack of a cell-attached SclA adhesin decreases tissue adherence and thereby promotes dissemination in serotype M3 GAS.

DISCUSSION

In contrast to our sophisticated knowledge of bacterial mechanisms contributing to virulence in invasive disease, comparatively little is known regarding the molecular genetic factors contributing to asymptomatic carriage of bacterial pathogens. Several studies exist describing mutations, in both virulence genes and virulence regulators, that contribute to hypervirulence and increased invasive disease caused by GAS. However, despite the much higher prevalence of GAS carriage, few studies have shown that specific mutations contribute to a GAS carrier phenotype. Our data not only suggest that small genetic changes may contribute to GAS carriage but that study of GAS carrier strains may provide greater insight into GAS pathogenesis.

Our data demonstrating that the presence of $sclA^{Carrier}$ is associated with decreased virulence and increased adherence are consistent with previous observations in GAS carriage. It has long been suggested that key phenotypic features of GAS carrier strains include decreased virulence and increased ability to adhere to and persist on mucosal surfaces. Further, such a phenotype may contribute to eradication failures in cases of GAS pharyngotonisillitis (27). Sela et al. (26) showed that GAS strains from subjects with eradication failure following treatment for pharyngotonsillitis had higher rates of adherence and internalization compared to strains from subjects with successful eradication. Moreover, previous studies have shown an inverse relationship between expression of bacterial surface adhesins and virulence (9-11). Thus, the mutation identified in *sclA* is one of the first specific mutations in a surface protein potentially linking increased adherence and decreased virulence in GAS carrier strains.

In addition to shedding new light on GAS carriage, our data suggest positive selection to maintain the truncated SclA^{Invasive} protein in the invasive serotype M3 GAS population. Invasive strains of serotype M3 have surprisingly few mutations in *sclA*. Our analysis of geographically and temporally diverse serotype M3 invasive GAS strains demonstrated that ~1% have the carrier mutation. The remaining invasive strains harbored the same *sclA* allele – the allele containing the premature stop codon – and virtually no polymorphisms (467 of 473; 98.7%). The observation that invasive serotype M3 GAS express the truncated protein *in vitro* (Figure 1B) and the isogenic deletion

mutant is affected in models of adherence and colonization is intriguing. Recently, it was shown that the homologous protein from serotype M1 GAS enhances survival in neutrophil extracellular traps (NETs) and interferes with myeloperoxidase release (28). Thus, it is possible that expression of a truncated SclA in serotype M3 GAS acts in a similar manner to enhance survival *in vivo* and promote invasive disease but, as it is not surface attached, does not directly act as an adhesin and instead facilitates dissemination.

The biologic activity of SclA^Invasive appears to be complex and potentially niche-specific based on the differing phenotypes of the deletion mutant in the models tested. The decreased bacterial burden of the deletion mutant in the mouse nasopharynx suggests that, at least in this model, SclA^Invasive contributes to the ability of invasive serotype M3 GAS to multiply at the mucosal surface. However, as indicated by the increased ability of the same mutant to better adhere to cultured epithelial cells (similar to the SclA^Carrier mutant), the truncated SclA may interfere with bacterial cell binding of host components. The observation that a full-length recombinant SclA protein binds the ECM components and that the truncated SclA is conserved among invasive serotype M3 GAS and has the identical variable region responsible for ECM binding further supports this model. Further experimentation is needed to test these hypotheses.

The mutation identified in *sclA* in GAS carrier strains is likely one of many mutations that may contribute to the carrier phenotype. For instance, it has been demonstrated that a mutation in the promoter of the gene for the virulence regulator Mga contributes to decreased virulence in GAS carrier strains (15). More recently, mutations that eliminate GAS capsule have been shown to arise *in vivo* during human carriage and result in decreased virulence and increased adherence to epithelial cell surfaces (22). The current study not only suggests that mutations contribute to GAS carriage but enhances understanding of serotype M3 GAS pathogenesis, further emphasizing the importance of carriage research. One interpretation of the data presented is that of a model in which distinct genetic differences between GAS carrier and invasive strains contribute to a carrier phenotype defined by decreased virulence and increased ability to adhere to epithelial surfaces (Figure 5). Further research is needed to fully define genotype-phenotype relationships in GAS carrier strains.

ACKNOWLEDGMENTS

REFERENCES

1. Shaikh, N., Leonard, E., and Martin, J. M. (2010) Prevalence of streptococcal pharyngitis and streptococcal carriage in children: a meta-analysis. *Pediatrics* **126**, e557-564

2. O'Loughlin, R. E., Roberson, A., Cieslak, P. R., Lynfield, R., Gershman, K., Craig, A., Albanese, B. A., Farley, M. M., Barrett, N. L., Spina, N. L., Beall, B., Harrison, L. H., Reingold, A., and Van Beneden, C. (2007) The epidemiology of invasive group A streptococcal infection and potential vaccine implications: United States, 2000-2004. *Clin Infect Dis* **45**, 853-862

3. Lukomski, 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

4. Almengor, A. C., and McIver, K. S. (2004) Transcriptional activation of *sclA* by Mga requires a distal binding site in *Streptococcus pyogenes*. *J Bacteriol* **186**, 7847-7857

5. Xu, Y., Keene, D. R., Bujnicki, J. M., Hook, M., and Lukomski, S. (2002) Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. *J Biol Chem* **277**, 27312-27318

6. Caswell, C. C., Oliver-Kozup, H., Han, R., Lukomska, E., and Lukomski, S. (2010) Scl1, the multifunctional adhesin of group A *Streptococcus*, selectively binds cellular fibronectin and laminin, and mediates pathogen internalization by human cells. *FEMS Microbiol Lett* **303**, 61-68

7. Caswell, C. C., Barczyk, M., Keene, D. R., Lukomska, E., Gullberg, D. E., and Lukomski, S. (2008) Identification of the first prokaryotic collagen-sequence motif that mediates binding to human collagen receptors, integrins alpha 2beta 1 and alpha 11beta 1. *J Biol Chem*

8. Caswell, C. C., Lukomska, E., Seo, N. S., Hook, M., and Lukomski, S. (2007) Scl1-dependent internalization of group A *Streptococcus* via direct interactions with the alpha2beta(1) integrin enhances pathogen survival and re-emergence. *Mol Microbiol* **64**, 1319-1331

9. Crotty Alexander, L. E., Maisey, H. C., Timmer, A. M., Rooijakkers, S. H., Gallo, R. L., von Kockritz-Blickwede, M., and Nizet, V. (2009) M1T1 group A streptococcal pili promote epithelial colonization but diminish systemic virulence through neutrophil extracellular entrapment. *J Mol Med*

10. Nyberg, P., Sakai, T., Cho, K. H., Caparon, M. G., Fassler, R., and Bjorck, L. (2004) Interactions with fibronectin attenuate the virulence of *Streptococcus pyogenes*. *EMBO J.* **23**, 2166-2174

11. McElroy, M. C., Cain, D. J., Tyrrell, C., Foster, T. J., and Haslett, C. (2002) Increased virulence of a fibronectin-binding protein mutant of Staphylococcus aureus in a rat model of pneumonia. *Infect Immun* **70**, 3865-3873

12. Oliver-Kozup, H. A., Elliott, M., Bachert, B. A., Martin, K. H., Reid, S. D., Schwegler-Berry, D. E., Green, B. J., and Lukomski, S. (2011) The streptococcal collagen-like protein-1 (Scl1) is a significant determinant for biofilm formation by group A *Streptococcus. BMC Microbiol* **11**, 262

13. Shea, P. R., Beres, S. B., Flores, A. R., Ewbank, A. L., Gonzalez-Lugo, J. H., Martagon-Rosado, A. J., Martinez-Gutierrez, J. C., Rehman, H. A., Serrano-Gonzalez, M., Fittipaldi, N., Ayers, S. D., Webb, P., Willey, B. M., Low, D. E., and Musser, J. M. (2011) Distinct signatures of diversifying selection revealed by genome analysis of respiratory tract and invasive bacterial populations. *Proc Natl Acad Sci U S A* **108**, 5039-5044

14. Olsen, R. J., Laucirica, D. R., Watkins, M. E., Feske, M. L., Garcia-Bustillos, J. R., Vu, C., Cantu, C., Shelburne, S. A., 3rd, Fittipaldi, N., Kumaraswami, M., Shea, P. R., Flores, A. R., Beres, S. B., Lovgren, M., Tyrrell, G. J., Efstratiou, A., Low, D. E., Van Beneden, C. A., and Musser, J. M. (2012) Polymorphisms in regulator of protease B (RopB) alter disease phenotype and strain virulence of serotype M3 group A *Streptococcus. J Infect Dis* **205**, 1719-1729

15. Flores, A. R., Olsen, R. J., Wunsche, A., Kumaraswami, M., Shelburne, S. A., 3rd, Carroll, R. K., and Musser, J. M. (2013) Natural variation in the promoter of the gene encoding the Mga regulator alters host-pathogen interactions in group a *Streptococcus* carrier strains. *Infect Immun* **81**, 4128-4138

16. Gryllos, I., Grifantini, R., Colaprico, A., Cary, M. E., Hakansson, A., Carey, D. W., Suarez-Chavez, M., Kalish, L. A., Mitchell, P. D., White, G. L., and Wessels, M. R. (2008) PerR confers phagocytic killing resistance and allows pharyngeal colonization by group A Streptococcus. *PLoS Pathog* **4**, e1000145

17. Carroll, R. K., Shelburne, S. A., 3rd, Olsen, R. J., Suber, B., Sahasrabhojane, P., Kumaraswami, M., Beres, S. B., Shea, P. R., Flores, A. R., and Musser, J. M. (2011) Naturally occurring single amino acid replacements in a regulatory protein alter streptococcal gene expression and virulence in mice. *J Clin Invest* **121**, 1956-1968

18. Lukomski, S., Hoe, N. P., Abdi, I., Rurangirwa, J., Kordari, P., Liu, M., Dou, S. J., Adams, G. G., and Musser, J. M. (2000) Nonpolar inactivation of the hypervariable streptococcal inhibitor of complement gene (sic) in serotype M1 Streptococcus pyogenes significantly decreases mouse mucosal colonization. *Infect Immun* **68**, 535-542

19. Chaffin, D. O., and Rubens, C. E. (1998) Blue/white screening of recombinant plasmids in Gram-positive bacteria by interruption of alkaline phosphatase gene (phoZ) expression. *Gene* **219**, 91-99

20. Han, R., Zwiefka, A., Caswell, C. C., Xu, Y., Keene, D. R., Lukomska, E., Zhao, Z., Hook, 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

21. Olsen, R. J., Sitkiewicz, I., Ayeras, A. A., Gonulal, V. E., Cantu, C., Beres, S. B., Green, N. M., Lei, B., Humbird, T., Greaver, J., Chang, E., Ragasa, W. P., Montgomery, C. A., Cartwright, J., Jr., McGeer, A., Low, D. E., Whitney, A. R., Cagle, P. T., Blasdel, T. L., DeLeo, F. R., and Musser, J. M. (2010) Decreased necrotizing fasciitis capacity caused by a single nucleotide mutation that alters a multiple gene virulence axis. *Proc Natl Acad Sci U S A* **107**, 888-893

22. Flores, A. R., Jewell, B. E., Olsen, R. J., Shelburne, S. A., 3rd, Fittipaldi, N., Beres, S. B., and Musser, J. M. (2014) Asymptomatic carriage of group a streptococcus is associated with elimination of capsule production. *Infect Immun* **82**, 3958-3967

23. Lancefield, R. C. (1959) Persistence of type-specific antibodies in man following infection with group A streptococci. *J Exp Med* **110**, 271-292

24. Oliver-Kozup, H., Martin, K. H., Schwegler-Berry, D., Green, B. J., Betts, C., Shinde, A. V., Van De Water, L., and Lukomski, S. (2013) The group A streptococcal collagen-like protein-1, Scl1, mediates biofilm formation by targeting the extra domain A-containing variant of cellular fibronectin expressed in wounded tissue. *Mol Microbiol* **87**, 672-689

25. Hoe, N. P., Lukomska, E., Musser, J. M., and Lukomski, S. (2007) Characterization of the immune response to collagen-like proteins Scl1 and Scl2 of serotype M1 and M28 group A Streptococcus. *FEMS Microbiol Lett* **277**, 142-149

26. Sela, S., Neeman, R., Keller, N., and Barzilai, A. (2000) Relationship between asymptomatic carriage of Streptococcus pyogenes and the ability of the strains to adhere to and be internalised by cultured epithelial cells. *J Med Microbiol* **49**, 499-502

27. Kaplan, E. L., Gastanaduy, A. S., and Huwe, B. B. (1981) The role of the carrier in treatment failures after antibiotic for group A streptococci in the upper respiratory tract. *J Lab Clin Med* **98**, 326-335

28. Dohrmann, S., Anik, S., Olson, J., Anderson, E. L., Etesami, N., No, H., Snipper, J., Nizet, V., and Okumura, C. Y. (2014) Role for streptococcal collagen-like protein 1 in M1T1 group A *Streptococcus* resistance to neutrophil extracellular traps. *Infect Immun* **82**, 4011-4020
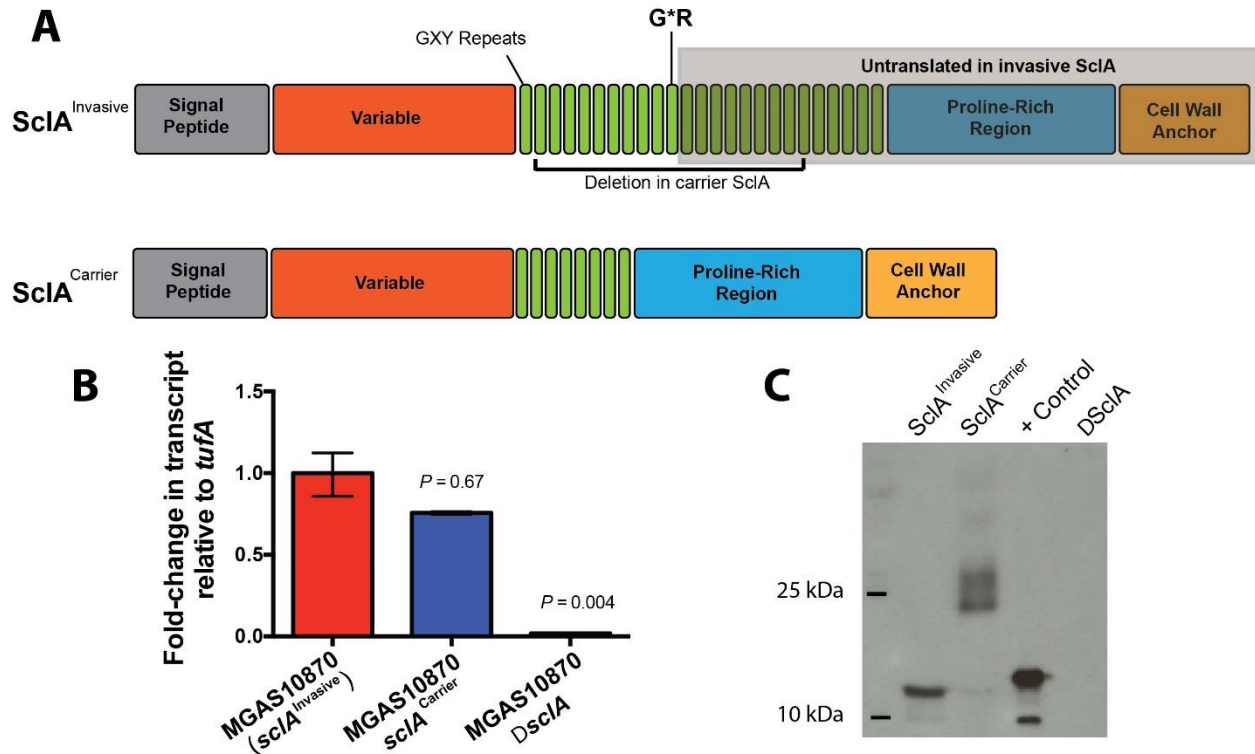
Figure 1

SclA differs in GAS serotype M3 invasive and carrier strains and is expressed. (**A**) Invasive strains of serotype M3 GAS are characterized by an *sclA* allele harboring an internal stop codon truncating the predicted protein after the 11th GXY repeat (G*Y). The shaded region containing the cell wall anchor is untranslated in invasive SclA, indicating the protein is not cell-attached. The naturally occurring carrier *sclA* allele is characterized by a deletion removing 17 GXY repeats in the mature protein including the G*R (bold) present in the invasive protein. To generate the full-length SclA^{Invasive} protein, the G*R (Gly-Stop-Arg) was repaired to GER (Gly-Glu-Arg). (**B**) TaqMan transcript analysis of invasive (MGAS10870), isoallelic (MGAS10870*sclA*^{Carrier}), and isogenic (MGAS10870Δ*sclA*) mutants. *P*-values determined by *t*-test (unequal variance). (**C**) Western immunoblot using anti-SclA^{Invasive} antibody and showing SclA expression in MGAS10870Δ*sclA* transcomplemented with either *sclA*^{Carrier} or *sclA*^{Invasive}. The positive control (+ Control) consisted of purified SclA^{Invasive} protein used for generation of rabbit anti-SclA antibody.

112
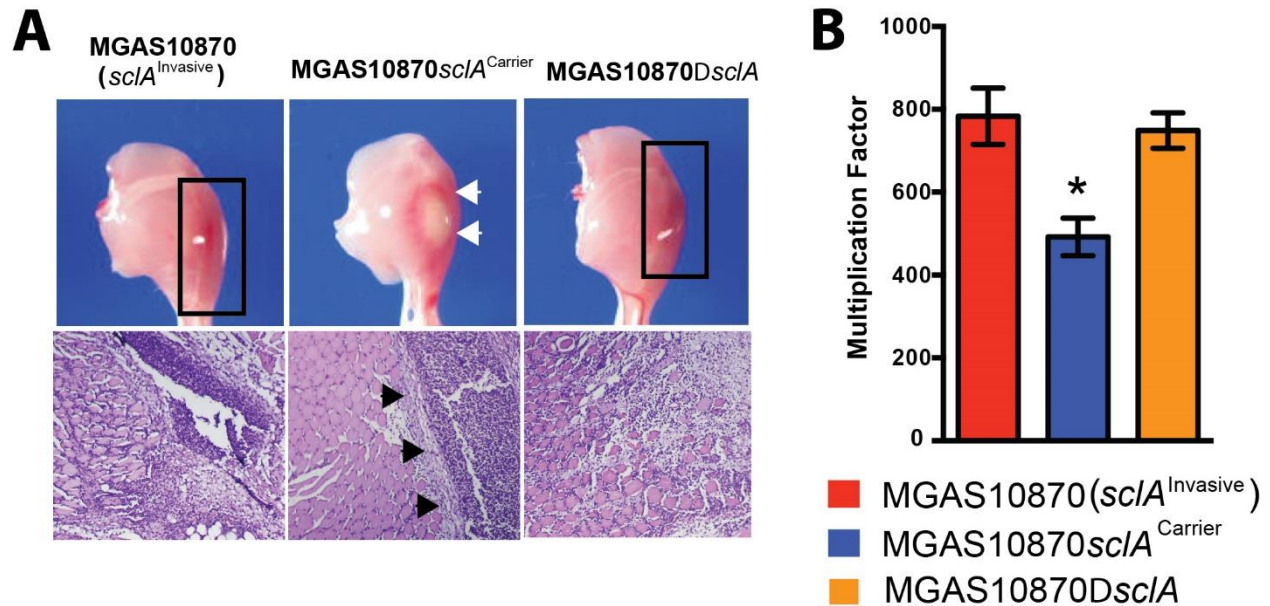
Figure 2

Strain with *sclA*[Carrier] has reduced virulence compared to the parental invasive strain. (**A**) Visual (top panels, original magnification 4X) and microscopic (bottom panels, original magnification 20X) examination of mouse hind limb lesions at 48 hours following intramuscular injection of GAS strains. Whereas mice infected with the isoallelic mutant (*sclA*[Carrier]) had a small, circumscribed, abscess-like lesion (white arrowheads), the parental invasive strain and isogenic deletion mutant caused a comparatively larger, grossly necrotic lesion centered at the inoculation site (boxed). Similarly, compared to the extensive spreading myonecrosis observed in the invasive and isogenic deletion strains, the isoallelic mutant showed an abscess-like lesion (black arrowheads) on microscopic examination. (**B**) Multiplication after growth in human blood. Shown is multiplication after growth in a single donor performed in quadruplicate. Similar results were observed after growth using additional donors. Error bars represent standard deviation and asterisk (*) indicates *P*-value < 0.05 as determined by *t*-test (unequal variance).
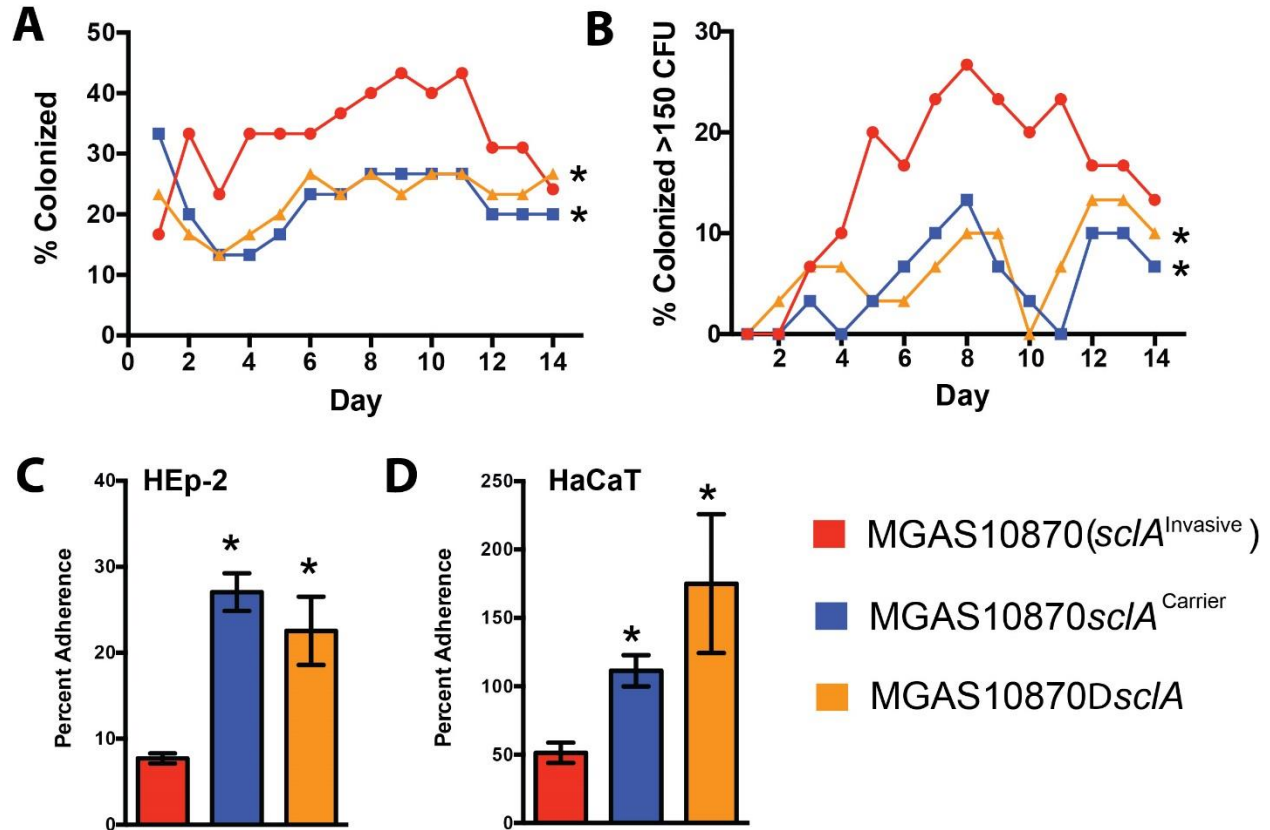
Figure 3

SclA^{Carrier} affects mouse nasopharyngeal colonization and adherence to cultured epithelial cells. (**A**) Mouse nasopharyngeal colonization following inoculation of invasive or isoallelic GAS strains. Mice (*n*=20) were inoculated and swabbed daily for 14 days. (**B**) Bacterial burden of mice colonized after nasopharyngeal infection. Percent of colonized mice with >150 CFU after daily swabbing. Asterisk (*) indicates $P < 0.05$ compared to the parental invasive strain as determined by Kruskal-Wallis. (**C-D**) Adherence of GAS strains to cultured epithelial cells. Error bars represent standard deviation and asterisks (*) indicate $P < 0.05$ compared to the parental invasive strain as determined by *t*-test (unequal variance).
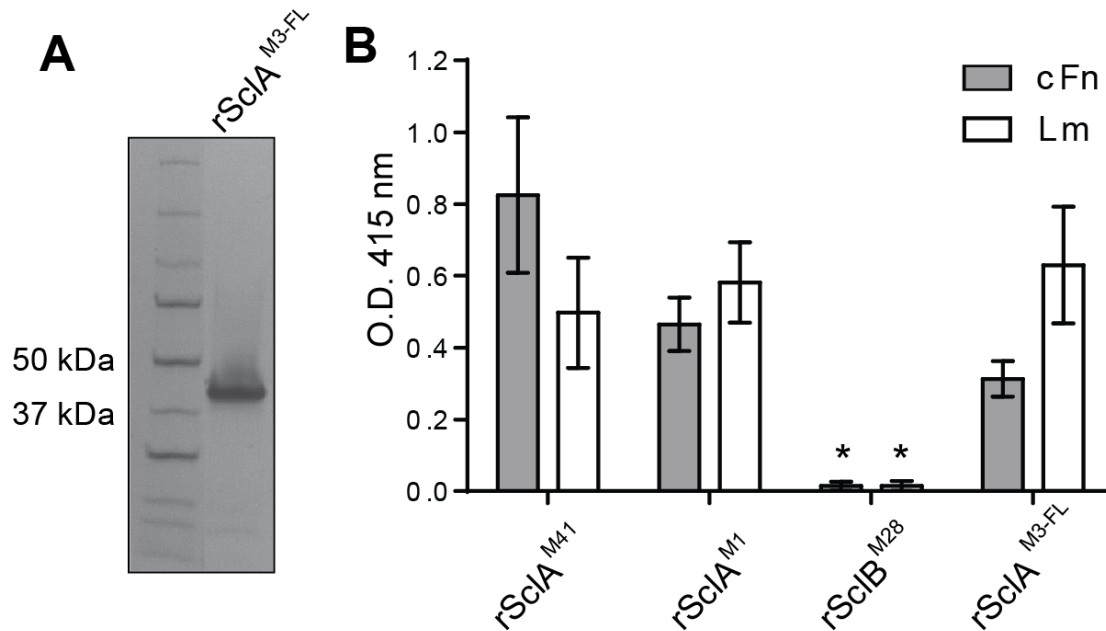
Figure 4

Recombinant full-length SclA (rSclA$^{M3\text{-}FL}$) binds the human extracellular matrix proteins cellular fibronectin (cFn) and laminin (Lm). (**A**) SDS-PAGE gel of purified rSclA$^{M3\text{-}FL}$. Expected size of protein is 20.6 kDa and rSclA$^{M3\text{-}FL}$ migrates at approximately 39 kDa; aberrant migration is characteristic of recombinant Scl proteins. (**B**) ECM binding to rSclA$^{M3\text{-}FL}$ protein. Test protein rSclA$^{M3\text{-}FL}$, as well as control ECM-binding positive (rSclA^{M41} and rSclA^{M1}) and binding negative (rSclB^{M28}) control proteins, were immobilized onto *Strep*-Tactin coated wells, and incubated with cFn and Lm. Detection of rScl-bound ECM was carried out using specific primary antibodies and HRP-conjugated secondary antibodies. Bars represent the mean O.D.$_{415nm}$ values normalized for BSA controls. Asterisks represent *P*-values < 0.05 for rSclA$^{M3\text{-}FL}$, rSclA^{M1}, and rSclA^{M41} in comparison to the negative control rSclB^{M28} determined from three independent experiments each performed in triplicate wells, using one-sample *t*-test.
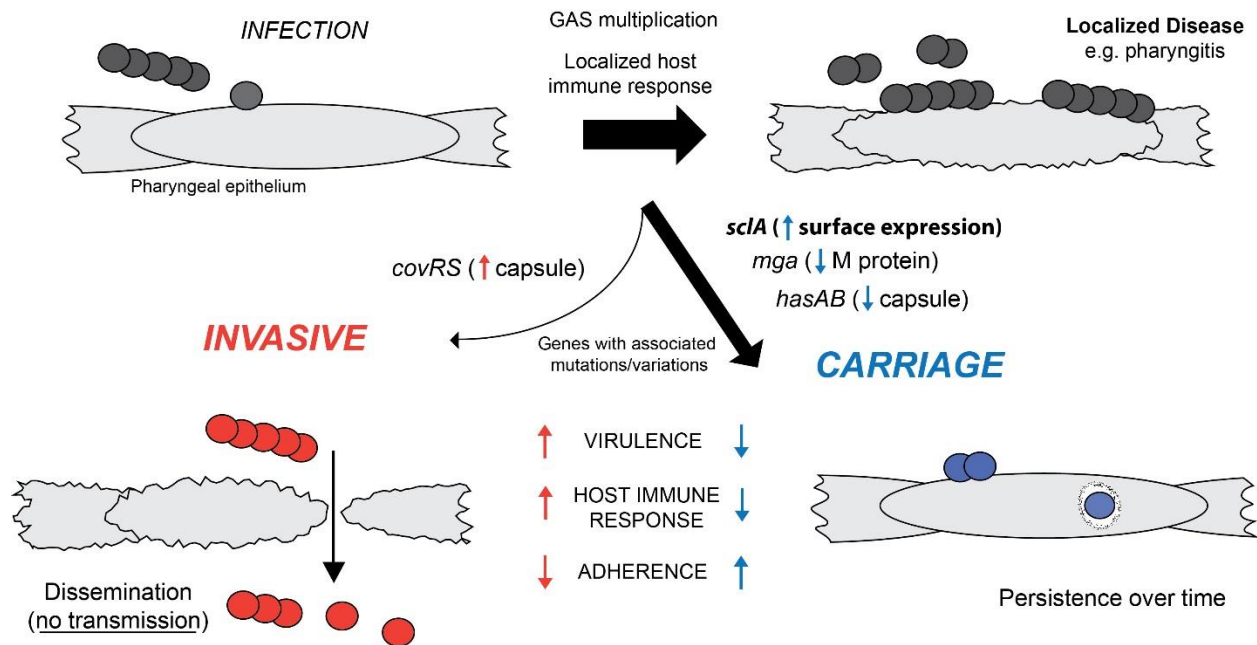
Figure 5

Model summarizing the effect of invasive (*sclA*^{Invasive}) or carrier (*sclA*^{Carrier}) *sclA* alleles in serotype M3 GAS. Following establishment of initial infection on an epithelial surface, GAS may progress to invasive disease (indicated by red circles) or carriage (indicated by blue circles). Invasive strains are characterized by the presence of a truncated and secreted SclA protein, decreased adherence to epithelial surfaces, and increased tissue destruction. It is postulated that the truncated SclA^{Invasive} protein at least partially contributes to decreased adherence to epithelial surfaces and host extracellular matrix, and increased ability to disseminate observed in invasive serotype M3 GAS. Conversely, carrier strains harboring the *sclA*^{Carrier} allele produce a full-length, surface-attached SclA protein, have decreased tissue destruction, and have increased adherence to host epithelial surfaces. Likewise, it is proposed that the decreased virulence and increased adherence in some carrier strains can be attributed to the naturally occurring SclA^{Carrier} protein.

116

PART II
STRUCTURAL ANALYSIS OF THE SCL PROTEINS

CHAPTER 4

CRYSTALLIZATION AND PRELIMINARY X-RAY CRYSTALLOGRAPHIC ANALYSIS OF THE VARIABLE DOMAIN OF SCL2.3, A STREPTOCOCCAL COLLAGEN-LIKE PROTEIN FROM INVASIVE M3-TYPE *STREPTOCOCCUS PYOGENES*

Flavia Squeglia, Beth Bachert, Maria Romano, Slawomir Lukomski*, and Rita Berisio

ABSTRACT

Streptococcal collagen-like proteins, Scls, are widely expressed by the well-recognized human pathogen *Streptococcus pyogenes*. These surface proteins contains a signature central collagen-like region, and an amino-terminal globular domain, termed the variable domain, which is protruded away from the cell surface by the collagen-like domain. Despite their recognized importance for bacterial pathogenicity, no structural information is hitherto available on proteins of the Scl class. The variable domain of Scl2 from invasive M3-type S. pyogenes has successfully been crystallized using vapor-diffusion methods. The crystals diffracted to 1.5 Å resolution and belonged to space group *H*32, with unit-cell parameters a =44.23, b =44.23, c =227.83 Å. The crystal structure was solved by single wavelength anomalous dispersion using anomalous signal from a europium chloride derivative.

INTRODUCTION

Collagen-like proteins that form stable triple-helices have been shown to be present in many bacterial species (1) and to play a role in pathogenicity. Best characterized prokaryotic collagens are the two collagen-like proteins, Scl1 and Scl2, which have been demonstrated to be simultaneously expressed on the cell surface of *S. pyogenes* and to promote bacterial adhesion to the host (2,3). Both Scl1 and Scl2 proteins contain a signal sequence, an N-terminal variable globular domain (V), a highly charged collagen-like triple-helix domain (CL) consisting of (Gly-Xaa-Yaa)$_n$ triplet repeats and a C-terminal Gram-positive cell wall attachment domain. The Scl1 and Scl2 proteins form stable triple-helical structures when expressed as recombinant proteins (4,5), and their N-terminal globular V domain adjacent to the triple-helix domain appears to be important for efficient triple-helix assembly (5-7).

 Bacterial adherence to host tissues, an early critical step in the infection process, often involves surface proteins (8,9). Among these, Scl proteins of *S. pyogenes* are crucial to host-pathogen recognition (10). It has been demonstrated that Scl1 can bind selected human extracellular matrix components (11), cellular integrin receptors (2,12,13), and plasma components (14-17). Importantly, human collagen receptors, such as integrin $\alpha_2\beta_1$, recognize the triple helix CL domain of Scl1, and this event results in cell signaling, indicating that collagen-like bacterial proteins display not only structural but also functional similarities to human collagens (2,12,13). Despite the key role of Scl proteins in bacterial pathogenicity, their three-dimensional structure has not been determined. In addition, whereas the triple helical structure of CL domain of Scls can be predicted based on sequence identity to collagen (5), no structural clues regarding the V domain can be obtain from Scl sequences, owing to poor sequence identity with known structures. Here, we report the crystallization and preliminary crystallographic investigations of the V domain of Scl2 from invasive M3-type *S. pyogenes* (Scl2.3-V).

119

MATERIALS AND METHODS

Cloning, expression and purification of recombinant rScl2.3-V protein

Recombinant rScl2.3-V protein was produced in the *Escherichia coli* periplasm using the *Strep*-tag II expression and purification system (IBA GmbH, Goettingen, Germany), as reported previously (4). Briefly, the 5'-portion of the *scl2.3* gene from strain MGAS315, encoding the amino-terminal Scl2.3-V-region, was PCR-amplified using the forward primer scl2-M3VF (5'-GAGATGGCCGATGGTGAAGATGCCCAAAAAAG) and the reverse primer scl2-M3VR (5'-CAGCGTCTCAGCGCTATCAAGGACATGATCTTGTATGCC) and was cloned into pASK-IBA2 vector, resulting in plasmid pSL155. *Escherichia coli* strain DH5α was used for cloning and *E. coli* BL21 was used for protein expression. *E. coli* harboring plasmid pSL155, which encodes the rScl2.3-V protein, was grown in Luria-Bertani (LB) liquid medium (BD Biosciences) supplemented with ampicillin (100 μg mL^{-1}). Plasmid construct pSL155 was confirmed by DNA sequencing and the identity of purified recombinant protein rScl2.3-V was confirmed by N-terminal Edman degradation.

Crystallization experiments

Crystallization trials were performed at 293 K using the hanging-drop vapor-diffusion method. Preliminary crystallization conditions were set up using a robotic station for high throughput crystallization screening (Hamilton STARlet NanoJet 8+1) and commercially available sparse-matrix kits (Crystal Screen, Crystal Screen 2 and Index, Hampton Research). Optimization of the crystallization conditions was performed manually by fine-tuning the protein and precipitant concentrations.

Data collection and processing

Diffraction data were collected to 1.52 Å in-house from a native crystal at 100K using a Rigaku Micromax 007 HF generator producing Cu *Kα* radiation and equipped with a Saturn944 CCD detector. Cryoprotection of the crystals was achieved without addition of further cryoprotectants, given the composition of the crystallization mother liquor, which contained the cryoprotectant pentaerythritol ethoxylate (15/4 EO/OH) at 30% (v/v). For phasing purposes, native crystals were soaked in solutions containing between 2 and 8 mM EuCl$_3$ for different soaking times. Data were collected from several crystals to identify the best single-wavelength

Anomalous Diffraction (SAD) signal. The data sets were scaled and merged using HKL2000 program package (18). Statistics of data collection are reported in Table 1.

Structure determination and refinement

Phasing was achieved using in-house SAD data. A preliminary evaluation of the anomalous signal was performed for all tested crystals using the SCALEPACK software, implemented in HKL2000 (18). SHELXD was used to identify europium ion sites (19). Phases were then improved by solvent-flattening density modification and phase extension by RESOLVE (20). The obtained model was further improved using ARP/*w*ARP (21).

RESULTS AND DISCUSSION

The recombinant rScl2.3-V protein corresponds to the N-terminal part (residues 1-77) of Scl2.3 variant from the invasive *S. pyogenes* M3-type strain MGAS315 (22). This protein construct has successfully been purified and crystallized using vapor-diffusion methods. The purified rScl2.3-V showed a single band of approximately 10 kDa on SDS–PAGE, which is in good agreement with the predicted molecular mass of 10,105 Da. The initial automated crystallization screening using commercially available solutions provided the first hints of crystallization conditions. Small crystals that were not appropriate for diffractometric experiments were obtained in the presence of pentaerythritol ethoxylate. The quality of these crystals was improved by fine-tuning the concentration of the protein and of the precipitants. Crystals suitable for X-ray diffraction experiments (Figure 1) were obtained in 3 d using a protein concentration of 5 mg mL^{-1} and 0.05 M ammonium sulfate, 0.05M bis-tris pH 6.5, 30% (v/v) pentaerythritol ethoxylate (15/4 EO/OH). These crystals, which diffracted to 1.5 Å resolution (Figure 2), showed three-fold symmetry and belonged to the space group *H*32, with unit-cell parameters a=44.23, b=44.23, c=227.83 Å (Table 1). Matthews coefficient calculations (23) suggested the presence of one molecule per asymmetric unit (VM =2.14 Å^3/Da, with 42.6 % solvent content).

Lanthanides can yield high-phasing-power derivatives using in house copper sources (24-26). Europium chloride derivative crystals were prepared by soaking the native crystals in stabilizing solutions containing between 2 and 8 mM EuCl$_3$ for increasing soaking times. SAD data were collected for all crystals at 100K using a Rigaku Micromax 007 HF generator

producing Cu *Kα* radiation. The best SAD data were obtained upon crystal soaking in a solution consisting of 8 mM $EuCl_3$, 0.05M ammonium sulfate, 0.05M bis-tris, 30% (v/v) pentaerythritol ethoxylate (15/4 EO/OH) at pH 6.5 for 3 hours. The data sets were scaled and merged using HKL2000 program package (18) (Table 1). (19), we could identify four Europium sites in the asymmetric unit of the protein. With this substructure, a correlation coefficient of 31.4% was calculated (CC-all, calculated with all data). The obtained phases were improved by phase extension and density modification, using RESOLVE (27) and ARP/*w*ARP (21). Using this approach, about 80% of the residues present in the asymmetric unit could be automatically modeled (27). Manual model-building sessions (28) aimed at defining the complete Scl2.3-V structure are in progress.

ACKNOWLEDGMENTS

REFERENCES

1.   Rasmussen, M., Jacobsson, M., and Bjorck, L. (2003) Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J. Biol. Chem.* **278**, 32313-32316

2.   Humtsoe, J. O., Kim, J. K., Xu, Y., Keene, D. R., Hook, M., Lukomski, S., and Wary, K. K. (2005) A streptococcal collagen-like protein interacts with the alpha2beta1 integrin and induces intracellular signaling. *The Journal of biological chemistry* **280**, 13848-13857

3.   Rasmussen, M., and Bjorck, L. (2001) Unique regulation of SclB - a novel collagen-like surface protein of Streptococcus pyogenes. *Mol Microbiol* **40**, 1427-1438

4.   Han, R., Zwiefka, A., Caswell, C. C., Xu, Y., Keene, D. R., Lukomska, E., Zhao, Z., Hook, M., et al. (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

5.   Xu, Y., Keene, D. R., Bujnicki, J. M., Hook, M., and Lukomski, S. (2002) Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. *The Journal of biological chemistry* **277**, 27312-27318

6.   Lukomski, 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. *Infection and immunity* **68**, 6542-6553

7. Mohs, A., Silva, T., Yoshida, T., Amin, R., Lukomski, S., Inouye, M., and Brodsky, B. (2007) Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline. *The Journal of biological chemistry* **282**, 29757-29765

8. Esposito, C., Pethoukov, M. V., Svergun, D. I., Ruggiero, A., Pedone, C., Pedone, E., and Berisio, R. (2008) Evidence for an elongated dimeric structure of heparin-binding hemagglutinin from Mycobacterium tuberculosis. *Journal of bacteriology* **190**, 4749-4753

9. Chagnot, C., Listrat, A., Astruc, T., and Desvaux, M. (2012) Bacterial adhesion to animal tissues: protein determinants for recognition of extracellular matrix components. *Cellular microbiology* **14**, 1687-1696

10. Berisio, R., and Vitagliano, L. (2012) Polyproline and triple helix motifs in host-pathogen recognition. *Curr Protein Pept Sci* **13**, 855-865

11. Caswell, C. C., Oliver-Kozup, H., Han, R., Lukomska, E., and Lukomski, S. (2010) Scl1, the multifunctional adhesin of group A Streptococcus, selectively binds cellular fibronectin and laminin, and mediates pathogen internalization by human cells. *FEMS microbiology letters* **303**, 61-68

12. Caswell, C. C., Barczyk, M., Keene, D. R., Lukomska, E., Gullberg, D. E., and Lukomski, S. (2008) Identification of the first prokaryotic collagen sequence motif that mediates binding to human collagen receptors, integrins alpha2beta1 and alpha11beta1. *The Journal of biological chemistry* **283**, 36168-36175

13. Caswell, C. C., Lukomska, E., Seo, N. S., Hook, M., and Lukomski, S. (2007) Scl1-dependent internalization of group A Streptococcus via direct interactions with the alpha2beta(1) integrin enhances pathogen survival and re-emergence. *Molecular microbiology* **64**, 1319-1331

14. Gao, Y., Liang, C., Zhao, R., Lukomski, S., and Han, R. (2010) The Scl1 of M41-type group A Streptococcus binds the high-density lipoprotein. *FEMS microbiology letters* **309**, 55-61

15. Caswell, C. C., Han, R., Hovis, K. M., Ciborowski, P., Keene, D. R., Marconi, R. T., and Lukomski, S. (2008) The Scl1 protein of M6-type group A Streptococcus binds the human complement regulatory protein, factor H, and inhibits the alternative pathway of complement. *Molecular microbiology* **67**, 584-596

16. Reuter, M., Caswell, C. C., Lukomski, S., and Zipfel, P. F. (2010) Binding of the human complement regulators CFHR1 and factor H by streptococcal collagen-like protein 1 (Scl1) via their conserved C termini allows control of the complement cascade at multiple levels. *The Journal of biological chemistry* **285**, 38473-38485

17. Pahlman, L. I., Marx, P. F., Morgelin, M., Lukomski, S., Meijers, J. C., and Herwald, H. (2007) Thrombin-activatable fibrinolysis inhibitor binds to Streptococcus pyogenes by interacting with collagen-like proteins A and B. *The Journal of biological chemistry* **282**, 24873-24881

18. Otwinowski, Z., and Minor, W. (1997). *Methods Enzymol.* **276**, 307–326

19. Sheldrick, G. M. (2008) A short history of SHELX. *Acta Crystallogr A* **64**, 112-122

20. Terwilliger, T. (2004) SOLVE and RESOLVE: automated structure solution, density modification and model building. *J Synchrotron Radiat* **11**, 49-52

21. Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **3**, 1171-1179

22. Beres, S. B., Sylva, G. L., Barbian, K. D., Lei, B., Hoff, J. S., Mammarella, N. D., Liu, M. Y., Smoot, J. C., et al. (2002) Genome sequence of a serotype M3 strain of group A Streptococcus: phage-encoded toxins, the high-virulence phenotype, and clone emergence. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 10078-10083
23. Matthews, B. W. (1968) Solvent content of protein crystals. *J Mol Biol* **33**, 491-497
24. Ruggiero, A., Squeglia, F., Marasco, D., Marchetti, R., Molinaro, A., and Berisio, R. (2011) X-ray structural studies of the entire extracellular region of the serine/threonine kinase PrkC from Staphylococcus aureus. *Biochem J* **435**, 33-41
25. Perez-Dorado, I., Gonzalez, A., Morales, M., Sanles, R., Striker, W., Vollmer, W., Mobashery, S., Garcia, J. L., et al. (2010) Insights into pneumococcal fratricide from the crystal structures of the modular killing factor LytC. *Nature structural & molecular biology* **17**, 576-581
26. Perez-Dorado, I., Sanles, R., Gonzalez, A., Garcia, P., Garcia, J. L., Martinez-Ripoll, M., and Hermoso, J. A. (2010) Crystallization of the pneumococcal autolysin LytC: in-house phasing using novel lanthanide complexes. *Acta crystallographica. Section F, Structural biology and crystallization communications* **66**, 448-451
27. Terwilliger, T. C. (2003) SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol* **374**, 22-37
28. Jones, T. A. (2004) Interactive electron-density map interpretation: from INTER to O. *Acta Crystallogr D Biol Crystallogr* **60**, 2115-2125

Figure 1

Image of typical rScl2.3-V crystals grown using 5 mg mL^{-1} protein solution and 0.05M ammonium sulfate, 0.05M bis-tris, 30% (v/v) pentaerythritol ethoxylate (15/4 EO/OH), pH 6.5.
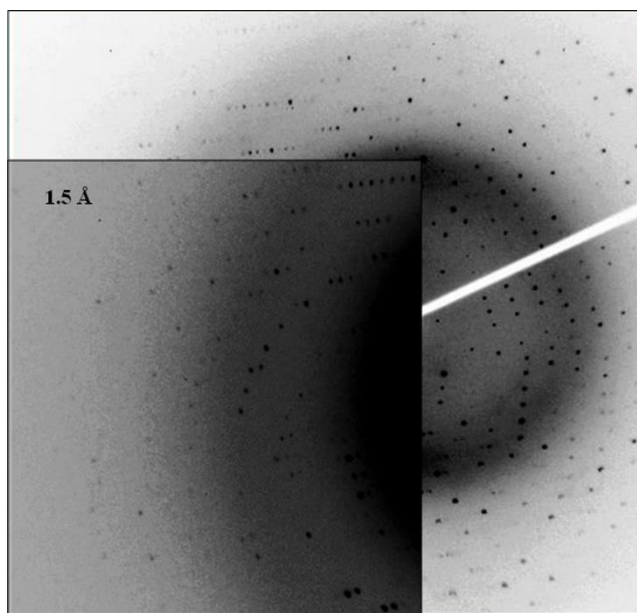
Figure 2

Diffraction pattern of a rScl2.3-V native crystal. Diffraction data are detectable to 1.5 Å resolution.

Table 1. Data collection statistics. Values in parentheses are for highest resolution shells

	EuCl$_3$ derivative	Native
Space group	*H*32	*H*32
Unit cell parameters	$a = b = 44.26$, $c = 228.01$, $\gamma = 120$	$a = b = 44.23$, $c = 227.83$, $\gamma = 120$
Resolution (Å)	1.87	1.52
Average redundancy	9.5 (7.5)	5.3 (2.6)
Unique reflections	7545	13802
Completeness (%)	100 (99.9)	99.2 (86.7)
R$_{merge}$[†] (%)	0.043 (0.419)	0.061 (0.345)
Average I/σ(I)	47.7 (5.1)	49.2 (3.10)

[†] $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$

CHAPTER 5

THE CRYSTAL STRUCTURE OF THE STREPTOCOCCAL COLLAGEN-LIKE PROTEIN 2 GLOBULAR DOMAIN FROM INVASIVE M3-TYPE GROUP A *STREPTOCOCCUS* SHOWS SIGNIFICANT SIMILARITY TO IMMUNOMODULATORY HIV PROTEIN GP41

Flavia Squeglia, Beth Bachert, Alfonso De Simone, Slawomir Lukomski and Rita Berisio

ABSTRACT

The arsenal of virulence factors deployed by streptococci include streptococcal collagen-like (Scl) proteins. These proteins, which are characterized by a globular and a collagen-like domain, play key roles in host-adhesion, host immune defense evasion and biofilm formation. In this work, we demonstrate that the Scl2.3 protein is expressed on the surface of invasive M3-type strain MGAS315 of *Streptococcus pyogenes*. We report the crystal structure of Scl2.3 globular domain, the first of any Scl. This structure shows a novel fold among collagen trimerization domains, of either bacterial or human origin. Despite there being low sequence identity, we observe that Scl2.3 globular domain structurally resembles the gp41 subunit of the envelope glycoprotein from human immunodeficiency virus type 1, an essential subunit for viral fusion to human T cells. We combined crystallographic data with modeling and molecular dynamics techniques to gather information on the entire lollipop-like Scl2.3 structure. Molecular dynamics data evidence a high flexibility of Scl2.3, with remarkable inter-domain motions which are likely instrumental to the protein biological function in mediating adhesive or immune-modulatory functions in host-pathogen interactions. Altogether, our results provide molecular tools for the understanding of Scl-mediated streptococcal pathogenesis and important structural insights for the future design of small molecular inhibitors of streptococcal invasion.

INTRODUCTION

Streptococcus pyogenes or group A *Streptococcus* (GAS) is a human adapted pathogen, causing over 700 million cases worldwide annually (1). GAS infections produce a wide range of clinical outcomes, from superficial throat and skin infections to life-threatening invasive diseases such as streptococcal toxic shock syndrome (STSS) and necrotizing fasciitis (2,3). The mortality resulting from the acquisition of invasive infections is high with 163000 deaths globally each year (4). It is known that M3-type GAS strains are associated with severe infections. In a survey of 108 isolates from the US, 50% of invasive diseases were caused by M1 and M3 type strains, and M3 type strains contributed to the majority of STSS cases (5). Over the past decade, molecular pathogenomics has facilitated our understanding of the molecular basis for the more severe invasive diseases caused by M3-type strains (6-10). GAS produces cell-associated virulence factors that contribute to host colonization and immune evasion, and include the streptococcal collagen-like proteins Scl1 and Scl2, also known as SclA and SclB (11-15).

The Scl1 and Scl2 proteins share a similar structural organization, including an N-terminal variable globular domain (V), a highly charged collagen-like triple-helix domain (CL) consisting of (Gly-Xaa-Yaa)$_n$ triplet repeats and a C-terminal Gram-positive cell wall attachment domain (Figure 1A). Like collagen, an important structural protein in the extracellular matrix of animals, Scl1 and Scl2 form stable triple-helical structures (11,16-19). The collagen triple helix is composed of three left-handed polyproline helices twisted into a right-handed supercoiled structure. In mammals, a strong contribution to triple helix stability is given by a high content of hydroxyproline (Hyp) residues at the Y position of the X-Y-Gly triplets, whereas bacteria lack the prolyl hydroxylase needed for post-translational modification of proline residues (17,20,21). To explore the basis of bacterial collagen triple-helix stability in the absence of Hyp, biophysical studies were carried out on recombinant Scl2 protein and a set of peptides modeling the Scl2 highly charged repetitive (Gly-X-Y)n sequences (17). These studies showed that bacteria have developed alternative strategies to stabilize the triple helix, involving electrostatic interactions, inter-chain hydrogen bonds, and a hydration-mediated hydrogen bonding network (17,22).

129

Similar to that observed for human collagen, the V-domain was hypothesized to be needed for proper folding of the triple helical regions, since their high symmetry constitutes an obstacle for optimal folding. However, the observation that the recombinant CL domain of Scl1 is expressed as a stable triple helix (16,23) contrasts this hypothesis, at least *in vivo*. Scls have characteristic "lollipop-shaped" domain organization, which seems apt for ligand binding. Indeed, antibody mapping and electron microscopy imaging analyses confirmed that the stalk-forming CL region projects the globular V region away from the bacterial surface (16), a feature which may facilitate interactions of V regions with their potential targets. Several biologically-relevant V-region ligands have been identified using experimental approaches. Thus, different Scl variants bind human extracellular matrix proteins, cellular fibronectin and laminin (24), as well as plasma components including the low-density lipoprotein, thrombin-activatable fibrinolysis inhibitor, and complement regulatory proteins factor H and factor H-related protein-1 (18,25-28). In addition, the CL domain of Scl can bind directly to host cells through cellular receptors, integrins $\alpha_2\beta_1$ and $\alpha_{11}\beta_1$ (29-31). Hence, the two main Scl structural domains bind human ligands and are essential for GAS adhesion, host-cell entry, and immunomodulation of host defenses. Because of the importance of invasive M3-type strains in human morbidity and mortality, the presumed expression of Scl2.3 (Scl2 from M3 strain) was previously used as an epidemiological marker of *S. pyogenes* (7), although its actual expression has not been shown. Here, we demonstrate that Scl2.3 protein is expressed on the cell surface of an invasive M3-type group A *Streptococcus*. Since no structural clues on Scl2 are available, we have combined x-ray crystallography with molecular modeling and dynamics to obtain information on the structure of the entire molecule. This structure delivers the first atomic description of a Scl protein and opens the field for the understanding of structure-function relationship of key proteins that mediate essential adhesive and immunomodulatory functions of group A *Streptococcus*.

MATERIALS AND METHODS

Bacterial strains and growth

The group A *Streptococcus* (GAS) M3-type strain MGAS315 (6) used here was isolated from an invasive case of a streptococcal toxic-like shock syndrome in Texas (5). GAS was routinely grown in Todd-Hewitt broth supplemented with 0.2% yeast extract (THY medium) or on tryptose agar with 5% sheep blood (BD Biosciences) at 37°C in 5% CO_2-20% O_2 atmosphere.

The *Escherichia coli* strain DH5α was used in cloning experiments and *E. coli* BL21 was used for protein expression. *E. coli* strains were grown in Luria-Bertani media (BD Biosciences) supplemented with ampicillin (100 µg/ml).

Protein methods

Recombinant Scl2.3-V region polypeptide, designated rScl2.3-V, was generated using the *Strep*-tag II expression and purification system (IBA-GmbH), as described previously (16). Briefly, the 5'-portion of the *scl2.3* allele from strain MGAS315, encoding the amino-terminal variable (V) region of the presumed mature Scl2.3 protein, was PCR-amplified using scl2-M3VF (5'-GAGATGGCCGATGGTGAAGATGCCCAAAAAG) forward primer and scl2-M3VR (5'-CAGCGTCTCAGCGCTATCAAGGACATGATCTTGTATGCC) reverse primer, and subsequently cloned into an *E. coli* expression vector pASK-IBA2, generating plasmid pSL155.

The rScl2.3-V polypeptide is fused at the N-terminus to the OmpA signal peptide mediating periplasmic expression of recombinant protein. The OmpA is selectively cleaved off during protein export by an endogenous signal peptidase, thus, releasing the rScl polypeptide; the N-terminal sequence of purified rScl2.3-V was confirmed by Edman Degradation. The rScl2.3-V polypeptide also has a short affinity tag, the *Strep*-tag II (WSHPQFEK), at the C-terminus, which allowed for affinity chromatography purification on *Strep*-Tactin Sepharose. Purified rScl2.3-V protein was dialyzed against 25 mM HEPES, pH 8.0 and stored -20°C. Recombinant protein rScl2.3-V was tested for purity and integrity on TGX 4-20% gradient gel (Bio-Rad) and stained with RAPIDstain™ (G-Biosciences).

The presence of the cell-wall associated Scl2.3 protein was studied using the method described before (11,15). MGAS315 was grown in THY medium until mid-logarithmic phase (OD_{600} ~0.5) before GAS cells were harvested by centrifugation. The cell wall-associated protein fraction was obtained by resuspending the cell pellet in a high-sucrose buffer (10mM Tris pH 8.0, 20% sucrose), containing 25 U of mutanolysin and 1 mg/mL of lysozyme, and incubating at 37°C for 1 hour. rScl2.3-V protein and cell-wall-associated fraction of MGAS315 were analyzed by SDS-PAGE and Western immunoblotting, using rabbit polyclonal antibodies raised against rScl2.3-V (Proteintech Group, Inc.). Alkaline phosphatase-conjugated anti-rabbit IgG H&L goat polyclonal antibodies (Rockland) were used as the secondary antibody, and detection was performed using 1-Step™ NBT/BCIP substrate (Thermo Scientific). PageRuler Plus Prestained Protein Ladder (Thermo Scientific) was used as a molecular weight marker.

CD spectroscopy

To analyze the conformational state of rScl2.3-V, far-UV CD spectra were registered at 20°C. All CD spectra were recorded with a Jasco J-810 spectropolarimeter equipped with a Peltier temperature control system (Model PTC-423-S). Molar ellipticity per mean residue, $[\theta]$ in deg $cm^2 \cdot dmol^{-1}$, was calculated from the equation: $[\theta] = [\theta]obs \cdot mrw \cdot (10 \cdot l \cdot C)^{-1}$, where $[\theta]obs$ is the ellipticity measured in degrees, mrw is the mean residue molecular mass (116.1 Da), C is the protein concentration in $g \cdot L^{-1}$ and l is the optical path length of the cell in cm. Far-UV measurements (190-260 nm) were carried out at 20 °C using a 0.1 cm optical path length cell and a protein concentration of 0.2 $mg \cdot mL^{-1}$. Thermal denaturation studies were conducted at 222 nm with increasing temperature from 20 to 70°C. Proteins were equilibrated at each temperature point for 2 min, and the temperature was increased with an average rate of 0.5 °C/min. Tm was obtained by taking the peak of the first derivative of the melting curve.

Multiple light scattering

Purified rScl2.3-V was analysed by size-exclusion chromatography (SEC) coupled to a DAWN MALS instrument (Wyatt Technology) and an OptilabTM rEX (Wyatt Technology). 1 mg of sample was loaded a S75 10/30 column, equilibrated in 25 mM HEPES, 100 mM

NaCl, pH 7.4. A constant flow rate of 0.5 ml/min was applied. The on-line measurement of the intensity of the Rayleigh scattering as a function of the angle as well as the differential refractive index of the eluting peak in SEC was used to determine the weight average molar mass (Mw) of eluted protein, using the Astra 5.3.4.14 software (Wyatt Technologies).

Crystallization, data collection and processing

Crystallization trials were performed at 293 K using the hanging-drop vapor-diffusion method. Preliminary crystallization conditions were set up using a robot station for high throughput crystallization screening (Hamilton STARlet NanoJet 8+1) and commercially available sparse-matrix kits (Crystal Screen kits I and II, Hampton Research, Index). Optimization of the crystallization conditions was performed manually by tuning protein and precipitant concentrations. Best crystals were grown in 0.05M Ammonium sulfate, 0.05M Bis-Tris pH 6.5, 30% v/v Pentaerythritol ethoxylate (15/4 EO/OH) (32). For structure solution, europium chloride derivative crystals were prepared by soaking a native crystal in a solution containing 8 mM $EuCl_3$, 0.05M ammonium sulphate, 0.05M Bis-Tris, 30% v/v Pentaerythritol ethoxylate (15/4 EO/OH) for 3 hours at pH 6.5. A single-wavelength anomalous diffraction experiment (SAD) was recorded in-house at 100K using a Rigaku Micromax 007 HF generator producing Cu K_α radiation and equipped with a Saturn944 CCD detector. The data sets were scaled and merged using HKL2000 program package (33) (Table 1).

Structure determination and refinement

Phasing was achieved using in-house single anomalous dispersion data, using a previously adopted protocol (34). Using these data, both SHELXD (35) and SOLVE (36) identified five europium ions. Phases, improved by phase extension and density modification by RESOLVE (36) and wARP (37), allowed us to trace nearly the entire molecule structure. Crystallographic refinement was carried out against 95% of the measured data using the CCP4 program suite (38). The remaining 5% of the observed data, which was randomly selected, was used in Rfree calculations to monitor the

progress of refinement. The structures was validated using the program PROCHECK (39) and deposited with the PDB (accession code 4nsm).

Sequence and structure alignments

Alignments of all available Scl2 sequences were performed using the ClustalW program. This sequence alignment was used to obtain phylogenetic relationships. Structure alignments were carried out using the DALI server.

Modeling of the full Scl2.3

Molecular modeling sessions were carried out to model the collagen like domain and obtain the entire Scl2.3 structure. The collagen-like domain was modeled using the structure of the collagen-like peptide (Pro-Pro-Gly)$_{10}$ as a template (40). The full sequence of Scl2.3 was adjusted on the domain structure using *ad hoc* made routines. The full model was energy minimized using the GROMACS package.

Molecular dynamics simulations

Molecular dynamics (MD) simulations were performed using the region 7-123 of the Scl2.3 as a starting model, including the crystallographic V-domain (residues 7-77) and part of the modelled region (residues 78-123). MD simulations were carried out with the GROMACS package by using the all-atom AMBER99sb ILDN force field (41) in combination with the TIP4P-ew explicit water model (42). To avoid any bias on the hydration status of the protein derived from the MD analyses, crystallographic water molecules were removed from the starting model. The simulations were carried out in the NPT ensemble with periodic boundary conditions at a constant temperature of 300 K by using a weak coupling with external bath (V-rescale method) (43) and a constant pressure of 1 atm (berendesen pressure coupling)(44). A rectangular box was used to accommodate the protein, water molecules, and ions. The system included 28827 water molecules and a total of 120438 atoms.

Bending angles between CL and Scl2.3-V, or between regions of the collagen triple helix were defined between centre of masses of three group of atoms. For the definition of the global inter-domain angle, these atoms are the Cα atoms of residues 57 and 60 of

each chain (from Scl2.3-V), residues 75-77 (hinge region) and residues 82-84 (from CL). For the bending angle between three zones of the CL domain, we selected the Cα atoms of residues 77-78 of each chain (bottom region), residues 94 and 95 (hinge region) and residues 112 and 113 (top region).

<div align="center">

RESULTS

</div>

Expression of Scl2.3 protein by M3-type GAS

Expression of the Scl2 proteins is regulated at the level of translation and depends on a number of pentanucleotide repeats CAAAA found downstream to a GTG start codon (12,14,15). Based on the number of these repeats, the *scl2*-coding sequence may be in-frame, resulting in expression of the full-length protein, or out-of-frame, leading to early translation termination. We assessed the cell-surface expression of Scl2.3 protein by MGAS315, a strain representative of global invasive M3 organisms.

To generate tools for the detection of Scl2.3 protein, we cloned, expressed, and purified the recombinant Scl2.3-V protein (rScl2.3-V), corresponding to the V region of Scl2.3 from MGAS315. SDS-PAGE analysis of purified rScl2.3-V shows a single protein band of the expected size of about 10.1 kDa (Figure 1B), as further confirmed by sequencing. Rabbits were immunized with rScl2.3-V to generate specific anti-Scl2.3 antibodies, which we used to test the presence of the Scl2.3 in the cell wall-associated protein fraction of MGAS315 by western immunoblotting (Figure 1C). In addition to positive control (rScl2.3-V lane), we detected a prominent immunoreactive band of approximately 65 kDa in the cell-wall fraction (Scl2.3 lane) using post-immune rabbit serum, while probing with control pre-immune serum was negative for the rScl2.3 and Scl2.3 bands. Based on sequence analysis, the predicted molecular mass of the mature Scl2.3 protein is ~52.5 kDa. However, an aberrant migration of Scl proteins has been well documented (11,13). Altogether, our data show that the Scl2.3 protein is expressed on cell surface of invasive M3-type strain MGAS315 of *S. pyogenes*.

Structural studies in solution

Structural features of rScl2.3-V in solution were checked using circular dichroism (CD) and light scattering studies. As previously shown (22), Scl2-V has a typical α-helical CD

spectrum (Figure 2A). Thermal stability curves, determined by monitoring the CD signal at 222 nm, evidences a cooperative unfolding with melting transition at T_m=50°C. Consistent with previous data, denaturation of rScl2.3-V is fully reversible (Figure 2B). Analytical size-exclusion chromatography (SEC), coupled with multiangle light scattering (MALS) was carried out to investigate the oligomerization state of rScl2.3-V in solution. The on-line measurement of the intensity of the Rayleigh scattering as a function of the angle as well as the differential refractive index of the eluting peak in SEC was used to determine Mw. This analysis produced an Mw value of 26600 ±107 Da, which corresponds to a trimeric organization of the molecule (Figure 2C).

Overall structure of rScl2.3-V

rScl2.3-V was crystallized in the space group H32. The structure was solved by single-wavelength anomalous dispersion (SAD) analysis of europium-derivatized crystal and refined to a resolution of 1.6 Å (Table 1). Analysis of crystal packing using the software PISA confirms that the biologic unit of rScl2.3-V is a trimer. Consistently, a large surface area is buried (32% of the total surface, 5470 $Å^2$) upon trimer formation, with a strong gain of free energy of solvation (Δ^iG=-42.1 kcal/mol). rScl2.3-V molecules are organized about three-fold crystallographic axes to form a six helical bundle structure (Figure 3A). The inner core of this bundle consists of a parallel, trimeric structure in which helices are wrapped in a gradual left-handed super-helix. Three further helices wrap antiparallel to the internal helices in a left-handed direction around the exterior of the central trimer. The six-helix bundle forms an elongated cylinder measuring about 30 Å in diameter and 60 Å in height. Interestingly, external helices are shifted with respect to internal ones, as a 12-residue-long loop, embedding residues from Lys31 to Asp42, connects internal and external helices in each monomer. This region, which contains Pro34 and Pro36, adopts a well-defined polyproline II conformation (Figure 3B).

The V-domain of Scl2 was proposed to be stabilized by coiled coil interactions (23), although prediction servers do not provide a clear answer. We searched rScl2.3-V crystal structure for the typical structural features of coiled coils, named knobs-into-holes, using the software SOCKET (45). In typical coiled coils, hydrophobic side chains at 'a' and 'd' positions on one helix act as knobs and dock into holes formed by diamonds of four

residues on the partnering helix. This analysis shows that rScl2.3-V does not contain coiled coils.

Interactions between inner helices of the rScl2.3-V six-helix bundle involve different types of contacts along the bundle. Hydrophobic interactions exist at the two poles of the molecule, whereas an intricate pattern of salt bridges is formed in the central part (Figure 4). In this pattern, Arg56 bridges Glu60 of two adjacent protomers and interacts with Asp61 of an adjacent protomer. Further salt bridges exist between the central Glu60 and Arg64 and between Asp61 and more peripheral Lys57 (Figure 4). As a result, as many as 16 salt bridges stabilize the central region of the bundle.

Three outer N-terminal helices (residues 7-38) pack obliquely against the outside of the inner trimer in an antiparallel orientation. As such, they interact through hydrophobic interactions and salt bridges with residues in three grooves on the surface of the central helical trimer, whereas interactions mediated by the PPII strand are mostly hydrophobic (Figure 5A). The analysis of the electrostatic potential surface reveals an uneven distribution of charged patches, with a concentration of negative charges in the region opposite to the origin of the collagen triple helix (Figure 5C, D). The negatively charged patch generated by Asp42 and Asp43 of each chain surrounds a solvent exposed hydrophobic region, generated by Leu41 and Met46. Of these residues, the position of Met46 is occupied by hydrophobic residues in all members of a subgroup of Scl2 sequences, identified by phylogenetic analysis (branch C in Figure 6). In the same subgroup, negatively charged residues are often occurring in a region embedding Asp42 and Asp43. Different features characterize the other two subgroups, but all sequences present both charged and hydrophobic residues in loop regions (either experimentally determined or predicted), indicating that these features may be functionally important.

Sequence alignments

Several sequences of both Scl2 and Scl1, deriving from different *S. pyogenes* strains, have been identified. Multiple sequence alignment shows that a hallmark of all Scl2 sequences is the occurrence of hydrophobic residues at regular positions, most of which are conserved in all analyzed sequences. An analysis of rScl2.3-V structure shows that these residues constitute the inner core of the 6-helix bundle fold (Figure 4). This finding

suggests that all Scl2 proteins share the same 6-helix bundle fold we observe in rScl2.3-V structure. The same considerations apply to Scl1 sequences, since most conserved hydrophobic residues are also conserved in Scl1 (data not shown). Phylogenetic analysis shows that Scl2 sequences can be subdivided into three main branching groups (A, B and C in Figure 6). In each branch, specific characteristics are conserved. For example, a striking difference between branches A-B and branch C, which contains Scl2.3, is the presence of a fully conserved Pro residue in A-B in a position corresponding to Scl2.3 Ser26, which belongs to the α-helix α2 in Scl2.3 structure. Another almost conserved Pro residue characterizes branch B, in place of Scl2.3 Ser48, which is embedded in α-helix α2. These considerations suggest that the structures of proteins in each branch differ in the boundaries of α-helices constituting their six-helix bundle fold. Compared to Scl2.3, secondary structure predictions suggest that branch A and B are characterized by shorter α1 and α2 helices, connected by a longer loop in branch A, and by a loop-helix-loop motif in branch B (Figure 6).

Scl2.3-V structurally resembles gp41

A search for similar folds in structural data bases revealed a strong structural relationship between rScl2.3-V and subunit gp41 of the envelope glycoprotein from human immunodeficiency virus 1 (HIV-1) (PDB code 3o40, Z=12.1, r.m.s.d. 2.8 Å), with sequence identity between the two proteins of 9%, after alignment of 165 residues. In addition, the 3-carboxy-cis,cis-muconate lactonizing enzyme from *Agrobacterium radiobacter* (PDB code 2fen, Z=10.0, r.m.s.d. 3.3 Å) and the adenylosuccinate lyase from *E. coli* (PDB code 3gzh, Z=10.0, r.m.s.d. 3.4 Å) also show structural relationship with Scl2.3-V fold, with sequence identities in the range 9-10%. In the case of the two bacterial proteins identified by DALI (PDB codes 2fen and 3gzh), only five of the six helices of rScl2.3-V six-helix bundle are conserved. On the contrary, the superposition of rScl2.3-V structure on that of gp41 evidences a striking similarity in the helical arrangement of the two six-helix bundles (Figure 7). However, the three inner α-helices of gp41 are packed together in the 'knobs-into-hole' arrangement typical of coiled coils whereas coiled coil interactions were not found in rScl2.3-V structure. This feature is likely responsible for a more compact arrangement of inner helices in gp41, compared to

138

rScl2.3-V (Figure 7). Also, whereas gp41 fold is a highly regular six-helix bundle, rScl2.3-V presents a polyproline region at the N-terminal side of external helices. Although the position of these proline residues is not conserved among Scl2 sequences (Figure 6), their presence in the region connecting the two main helices forming protomers of the six-helix bundle is a distinctive feature of all Scl2 sequences.

Modeling of triple helical regions and MD simulations

Good quality electron density maps allowed us to define the conformation of rScl2.3-V C-terminal ends, up to Leu76. In Scl2, this is the site of attachment of the collagen-like triple helix. Notably, whereas all three Leu76 residues from rScl2.3-V are in a plane, the triplets of the collagen triple helix are typically staggered by one residue. This poses a question whether the asymmetry of the triple helix is accommodated by the rScl2.3-V structure or if a kink of the two domains is necessary, as previously observed for the engineered foldon-collagen (46). To tackle this question, we modeled the triple helical part of Scl2.3, thus producing the first structural description of a Scl (Figure 8A), and carried out Molecular Dynamics (MD) simulations. To assess the evolution of the structure in the simulation timescale (100 ns), a number of stereo-chemical parameters (gyration radius, secondary structure and RMSD) were monitored along the trajectories. The evaluation of root-mean-square deviations (RMSD) (calculated on the Cα atoms) from the starting structure evidences that large motions characterize the simulated system (Figure 8B). The RMSD values are smaller when they are separately computed for Scl2.3-V and Scl2.3-CL regions. Of these, RMSD values for the Scl2.3-V region are on average smaller than those of the Scl2.3-CL region (Figure 8B). Likely, the difference in the RMSD behavior for the two regions is in its structural characteristics. Indeed, local fluctuations on an elongated structure (i.e. Scl2.3-CL) propagate into larger effects on the RMSD than in globular structures (i.e. Scl2.3-V). The RMSD data are consistent with the presence of a principal motion involving a global inter-domain bending motion between Scl2.3-CL and Scl2.3-V (Figure 8C). Analysis of the MD trajectory structures evidences a continuous evolution from linear conformations, in which the Scl2.3-V and Scl2.3-CL domains are coaxial, to a more 'bent' state. In particular, the bending angle

between the axis of the Scl2.3-V and Scl2.3-CL domains (See Methods for definition) ranges from 143° to 180° in an elastic fashion (Figure 8C).

A direct indicator of the stability of the triple helix motif is the number of conserved main-chain hydrogen (H)-bonds along the trajectory (Figure 9A). These intermolecular H-bonds are distinctive of the triple helix motif and are established between the amide group of the Gly residues and carbonyl groups from complementary peptides that form the triple helix. The analysis confirms that the force field and simulation setup used were able to maintain the initial H-bonding pattern of the structure: on average, 87.5% of the native main-chain H-bonds were maintained. In addition to the observed inter-domain rearrangements along the trajectory, MD data also evidence a high flexibility of the Scl2.3-CL domain, with a bending angle around the center of the Scl2.3-CL region ranging between 152 and 180° (Figure 8D). The increased flexibility of amino-acid-rich triple helices, compared to imino-acid rich ones, is in line with previous MD analyses of other collagen-like polypeptides (47-49).

DISCUSSION

The arsenal of virulence factors deployed by streptococci include streptococcal collagen-like (Scl) proteins, which arm the cell wall of the bacterium and establish multiple functions, like host-adhesion (29-31), evasion of host immune defenses (27,28), or biofilm formation (50,51). There are nearly 300 collagen-like proteins annotated to streptococci (52), including several pathogenic organisms like *S. pyogenes* (11-15), *S. pneumoniae* (53), and *S. equi* (54,55). In addition to a signature collagen-like domain, Scl proteins contain a globular domain (V domain) and both Gram-positive signal peptide (YSIRK) and cell-wall anchor (LPXTG) domains, predicting that they are all cell-surface proteins. Despite their established importance in bacterial pathogenesis, no three-dimensional structural information is available so far for any of the Scl proteins. In this work, we formally demonstrate that the Scl2 protein is expressed by invasive M3-type strain MGAS315 and is found on the bacterial cell surface (Figure 1). By combining x-ray crystallography with computational techniques, we provide a structural description of the entire Scl2.3 molecule.

Scl2 is known to be regulated at the level of translation by the varying number of CAAAA pentanucleotide repeats directly downstream of the start codon, which may result in frameshift of the *scl2*-gene reading frame and early translation termination (12,14,15). Analysis of *scl2* within 50 GAS strains representing 21 different M types showed that the *scl2* allele is present in virtually all strains tested, though the number of the repeats, as well as resulting Scl2 expression, varied among strains. For example, none of the M1-type strains, whereas about half of the M28- and M12-type strains were predicted to express the full-length Scl2 variants. Interestingly, all of the M3-type strains initially tested (15) and 84% of 255 M3 global isolates (7) were found to contain in-frame *scl2.3* alleles. This suggests that there may be a selective advantage in M3 strains to express Scl2, and it may have an important role in the pathogenesis of M3-type GAS.

Previous binding studies have delineated roles for Scl1 and Scl2 in both host colonization and immune evasion. Thus, some Scl1 variants may aid host colonization by binding to cellular fibronectin and laminin, which are major components of human extracellular matrix, and integrins $\alpha_2\beta_1$ and $\alpha_{11}\beta_1$, which are present on the host cell surface (24,31). Scl1 has also been shown to bind the plasma lipoproteins and complement regulators of the immune system (18,27,28). Furthermore, both Scl1 and Scl2 proteins have been shown to bind thrombin-activatable fibrinolysis inhibitor, interfering with the normal fibrinolytic breakdown of blood clots (26), which may resemble a role of staphylococcal coagulases that produce clots as protective barrier against the immune response (56). These observations suggest Scl2 is involved in evasion or modulation of the immune response, rather than in host colonization. Though the role of Scl2 during infection is currently unclear, the structural data gained from this study provides very interesting clues into its possible function. The crystal structure of the V domain of Scl2.3 (rScl2.3-V) unveils a compact trimeric six-helix bundle fold. Consistently, light scattering experiments evidence that rScl2.3-V exhibits a trimeric arrangement also in solution (Figure 2). Trimeric six-helix bundle folds are, to date, not observed in bacteria, but typically characterize several glycoproteins involved in viral fusion, including the gp41 subunit of the envelope glycoprotein of human immunodeficiency virus type 1 (HIV-1), the glycoprotein B of Herpes Simplex virus (57) and the GP2 domain of the envelope glycoprotein GP from the Marburg and Ebola viruses (58-61). However, different than

gp41 and from previously reported data on Scl2 (23), the structure of rScl2.3-V does not contain coiled coils, but it is stabilized by both hydrophobic and ion pair interactions (Figure 4). Alignment of proteins with known structure shows that, even with low sequence identity (9%), rScl2.3-V structurally resembles the gp41 subunit of HIV-1, a subunit responsible for membrane fusion of the HIV virus (62,63). Since gp41 functions as a viral entry protein into CD4+ T lymphocytes, this might suggest a novel potential role for Scl2 in interacting with T cells and causing hyper-activation of the immune response, which is a hallmark of the STSS infections that are often associated with M3-type strains.

We modeled the triple helical region of Scl2 and performed MD analysis with the aim of investigating structural and dynamic features of Scl2.3. Scl2.3-V is located at the tip of an extremely elongated triple helical structure (about 1030 Å, Figure 8A) and exposes highly hydrophobic residues, like Leu41 and Met46 (Figure 5), a feature which may play a role in Scl2.3-mediated interaction of *S. pyogenes* with the hydrophobic *milieu* in the host. MD data evidenced an extremely flexible nature of Scl2, with a dynamic kink of the inter-domain organization (Figure 8). A kinked structure was previously observed for the an engineered foldon-collagen (46) and reflects the need of the structure to fit the three-fold symmetry of the V domain (which brings the site of attachment of the three collagen-like chains in a plane) with the one-residue stagger of the collagen-like chains. Our data show that the Scl2.3 structure can undertake both kinked and linear conformations, in a rapid equilibrium between them (Figure 8). The high structural flexibility we observed in Scl2.3 is likely instrumental to its biological function in mediating adhesive or immune-modulatory functions in host-pathogen interactions.

The V domain of Scls has been proposed as a trimerization domain which helps collagen folding. Indeed, different than for globular proteins, misfolding of the collagen triple helix is a likely event because of its repeating structure, whose stability is relatively insensitive to lateral shifts by one or more Gly-X-Y repeats. Consistently, trimerization domains have been found in many different proteins containing collagen triple helices. However, it has been shown that the V domain of Scls is not needed *in vivo*, since the CL region of Scl1 can be expressed in a folded triple helical state (16,23). This observation highlights a different folding mechanism of Scl proteins, compared to human collagen, for which trimerization domains are crucial to the correct triple helical

arrangement (64). Likely, V-domains of Scls display dichotomous functions by acting as triple helix stabilization domains, since they exhibit higher folding temperature than the CL regions (23), and by mediating host-pathogen interaction (29-31).

To date, there are five known atomic structures of trimerization domains of collagen, the NC1 domain of collagen IV (65), the homologous NC1 domain of collagens VIII and X (66,67) and the trimerization domains of collagen XV and XVIII (68,69). A trimerization domain was also characterized for BclA, the major component of the exosporium of the *B. anthracis* spore (70). All of them have a high content of β-structure but share no structural homology. The structure of rScl2.3-V presents novel features as it is mainly composed of α-helices. Multiple sequence alignment suggests that the 6-helix bundle fold exhibited by Scl2.3 is conserved in all Scl2 sequences, albeit with different α-helix boundaries and length of the loop connecting alpha helices α1 and α2 (Figure 6).

Phylogenetic analyses of variation in Scl2 V region among different M types revealed several interesting observations. Scl2 sequences from different M types formed three separate clades, referred to as A, B, and C (Figure 6). The invasive M type 3, found in branch C, clustered with M types 1 and 28, which are also associated with invasive infections, including streptococcal toxic shock syndrome and necrotizing fasciitis (5,71,72). Our analysis additionally evidenced that cluster C contains M types associated with rheumatic fever, including types 1, 3, 6, and 18 (71,73), thus suggesting a possible role of Scl2.3 in this disease. Altogether, our work delivers the first atomic description of a Scl protein. This structural information, which can be extended to other members of the Scl family, is precious to the understanding the structural basis of Scl-mediated streptococcal infection.

REFERENCES

1. Bisno, A. L., Rubin, F. A., Cleary, P. P., and Dale, J. B. (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

2. Kaul, R., McGeer, A., Low, D. E., Green, K., Schwartz, B., Study, O. G. A. S., and Simor, A. E. (1997) Population-based surveillance for group A streptococcal necrotizing fasciitis: clinical features, prognostic indicators, and microbiologic analysis of seventy-seven cases. *Am. J. Med.* **103**, 18-24

3. Sharkawy, A., Low, Donald E., Saginur, R., Gregson, D., Schwartz, B., Jessamine, P., Green, K., and McGeer, A. (2002) Severe group A streptococcal soft tissue infections in Ontario: 1992-1996. *Clin. Infect. Dis.* **34**, 454-460

4. Carapetis, J., Steer, A., Mulholland, E., and Weber, M. (2005) The global burden of group A streptococcal diseases. *Lancet Infect Dis* **5**, 685-694

5. Musser, J. M., Hauser, A. R., Kim, M. H., Schlievert, P. M., Nelson, K., and Selander, R. K. (1991) Streptococcus pyogenes causing toxic-shock-like syndrome and other invasive diseases: clonal diversity and pyrogenic exotoxin expression. *Proc. Natl. Acad. Sci. USA* **88**, 2668-2672

6. Beres, S. B., Sylva, G. L., Barbian, K. D., Lei, B., Hoff, J. S., Mammarella, N. D., Liu, M.-Y., Smoot, J. C., et al. (2002) Genome sequence of a serotype M3 strain of group A *Streptococcus*: Phage-encoded toxins, the high-virulence phenotype, and clone emergence. *Proc. Natl. Acad. Sci. USA* **99**, 10078-10083

7. Beres, S. B., Sylva, G. L., Sturdevant, D. E., Granville, C. N., Liu, M., Ricklefs, S. M., Whitney, A. R., Parkins, L. D., et al. (2004) Genome-wide molecular dissection of serotype M3 group A *Streptococcus* strains causing two epidemics of invasive infections. *Proc. Natl. Acad. Sci. USA* **101**, 11833-11838

8. Beres, S. B., Carroll, R. K., Shea, P. R., Sitkiewicz, I., Martinez-Gutierrez, J. C., Low, D. E., McGeer, A., Willey, B. M., et al. (2010) Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. 10.1073/pnas.0911295107 *Proceedings of the National Academy of Sciences* **107** 4371-4376

9. Olsen, R. J., and Musser, J. M. (2010) Molecular pathogenesis of necrotizing fasciitis. *Annual Review of Pathology: Mechanisms of Disease* **5**, 1-31

10. Nakagawa, I., Kurokawa, K., Yamashita, A., Nakata, M., Tomiyasu, Y., Okahashi, N., Kawabata, S., Yamazaki, K., et al. (2003) Genome Sequence of an M3 Strain of Streptococcus pyogenes Reveals a Large-Scale Genomic Rearrangement in

Invasive Strains and New Insights into Phage Evolution. *Genome Res.* **13**, 1042-1055

11. Lukomski, 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

12. Rasmussen, M., and Bjorck, L. (2001) Unique regulation of SclB-a novel collagen-like surface protein of *Streptococcus pyogenes*. *Infect. Immun.* **40**, 1427-1438

13. Rasmussen, M., Eden, A., and Bjorck, L. (2000) SclA, a novel collagen-like surface protein of *Streptococcus pyogenes*. *Infect. Immun.* **68**, 6370-6377

14. Whatmore, A. M. (2001) *Streptococcus pyogenes sclB* encodes a putative hypervariable surface protein with a collagen-like repetitive structure. *Microbiology* **147**, 419-429

15. Lukomski, S., Nakashima, K., Abdi, I., Cipriano, V. J., Shelvin, B. J., Graviss, 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

16. Xu, Y., Keene, D. R., Bujnicki, J. M., Hook, M., and Lukomski, S. (2002) Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. *J. Biol. Chem.* **277**, 27312-27318

17. Mohs, A., Silva, T., Yoshida, T., Amin, R., Lukomski, S., Inouye, M., and Brodsky, B. (2007) Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline. *The Journal of biological chemistry* **282**, 29757-29765

18. Han, R., Caswell, C. C., Lukomska, E., Keene, D. R., Pawlowski, M., Bujnicki, J. M., Kim, J. K., and Lukomski, S. (2006) Binding of the low-density lipoprotein by streptococcal collagen-like protein Scl1 of Streptococcus pyogenes. *Molecular microbiology* **61**, 351-367

19. Berisio, R., and Vitagliano, L. (2012) Polyproline and triple helix motifs in host-pathogen recognition. *Curr Protein Pept Sci* **13**, 855-865

20. Berisio, R., De Simone, A., Ruggiero, A., Improta, R., and Vitagliano, L. (2009) Role of side chains in collagen triple helix stabilization and partner recognition. *Journal of peptide science : an official publication of the European Peptide Society* **15**, 131-140

21. Improta, R., Berisio, R., and Vitagliano, L. (2008) Contribution of dipole-dipole interactions to the stability of the collagen triple helix. *Protein science : a publication of the Protein Society* **17**, 955-961

22. Yu, Z., Brodsky, B., and Inouye, M. (2011) Dissecting a bacterial collagen domain from Streptococcus pyogenes: sequence and length-dependent variations in triple helix stability and folding. *The Journal of biological chemistry* **286**, 18960-18968

23. 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 science : a publication of the Protein Society* **19**, 775-785

24. Caswell, C. C., Oliver-Kozup, H., Han, R., Lukomska, E., and Lukomski, S. (2010) Scl1, the multifunctional adhesin of group A *Streptococcus*, selectively binds

cellular fibronectin and laminin, and mediates pathogen internalization by human cells. *FEMS Microbiol Lett* **303**, 61-68

25. Gao, Y., Liang, C., Zhao, R., Lukomski, S., and Han, R. (2010) The Scl1 of M41-type group A *Streptococcus* binds the high-density lipoprotein. *FEMS Microbiol. Let.* **309**, 55-61

26. Pahlman, L. I., Marx, P. F., Morgelin, M., Lukomski, S., Meijers, J. C. M., and Herwald, H. (2007) Thrombin-activatable fibrinolysis inhibitor binds to *Streptococcus pyogenes* by interacting with collagen-like proteins A and B. *J. Biol. Chem.* **282**, 24873-24881

27. Caswell, C., Han, R., Hovis, K., Ciborowski, P., Keene, D., Marconi, R., and Lukomski, S. (2008) The Scl1 protein of M6-type group A *Streptococcus* binds the human complement regulatory protein, factor H, and inhibits the alternative pathway of complement. *Mol. Microbiol.* **67**, 584-596

28. Reuter, M., Caswell, C. C., Lukomski, S., and Zipfel, P. F. (2010) Binding of the human complement regulators CFHR1 and factor H by streptococcal collagen-like protein 1 (Scl1) via their conserved C termini allows control of the complement cascade at multiple levels. *J. Biol. Chem.* **285**, 38473-38485

29. Humtsoe, J. O., Kim, J. K., Xu, Y., Keene, D. R., Hook, M., Lukomski, S., and Wary, K. K. (2005) A streptococcal collagen-like protein interacts with the a2b1 integrin and induces intracellular signaling. *J. Biol. Chem.* **280**, 13848-13857

30. Caswell, C. C., Lukomska, E., Seo, N. S., Hook, M., and Lukomski, S. (2007) Scl1-dependent internalization of group A *Streptococcus* via direct interactions with the alpha2beta(1) integrin enhances pathogen survival and re-emergence. *Mol Microbiol* **64**, 1319-1331

31. Caswell, C. C., Barczyk, M., Keene, D. R., Lukomska, E., Gullberg, D. E., and Lukomski, S. (2008) Identification of the first prokaryotic collagen sequence motif that mediates binding to human collagen receptors, integrins a2b1 and a11b1. *J Biol Chem* **283**, 36168-36175

32. Squeglia, F., Bachert, B., Romano, M., Lukomski, S., and Berisio, R. (2013) Crystallization and preliminary X-ray crystallographic analysis of the variable domain of Scl2.3, a streptococcal collagen-like protein from invasive M3-type Streptococcus pyogenes. *Acta crystallographica. Section F, Structural biology and crystallization communications* **69**, 1023-1025

33. Otwinowski, Z., and Minor, W. (1997). *Methods Enzymol.* **276**, 307–326

34. Ruggiero, A., Squeglia, F., Marasco, D., Marchetti, R., Molinaro, A., and Berisio, R. (2011) X-ray structural studies of the entire extracellular region of the serine/threonine kinase PrkC from Staphylococcus aureus. *Biochem J* **435**, 33-41

35. Sheldrick, G. M. (2008) A short history of SHELX. *Acta Crystallogr A* **64**, 112-122

36. Terwilliger, T. (2004) SOLVE and RESOLVE: automated structure solution, density modification and model building. *J Synchrotron Radiat* **11**, 49-52

37. Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **3**, 1171-1179

38. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr* **59**, 1131-1137

39. Cheng, T. J., Sung, M. T., Liao, H. Y., Chang, Y. F., Chen, C. W., Huang, C. Y., Chou, L. Y., Wu, Y. D., et al. (2008) Domain requirement of moenomycin binding to bifunctional transglycosylases and development of high-throughput discovery of antibiotics. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 431-436

40. Berisio, R., Vitagliano, L., Mazzarella, L., and Zagari, A. (2002) Crystal structure of the collagen triple helix model [(Pro-Pro-Gly)(10)](3). *Protein science : a publication of the Protein Society* **11**, 262-270

41. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950-1958

42. Horn, H. W., Swope, W. C., Pitera, J. W., Madura, J. D., Dick, T. J., Hura, G. L., and Head-Gordon, T. (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys* **120**, 9665-9678

43. Bussi, G., Donadio, D., and Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J Chem Phys* **126**, 014101

44. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684-3690

45. Walshaw, J., and Woolfson, D. N. (2001) Socket: a program for identifying and analysing coiled-coil motifs within protein structures. *J Mol Biol* **307**, 1427-1450

46. Stetefeld, J., Frank, S., Jenny, M., Schulthess, T., Kammerer, R. A., Boudko, S., Landwehr, R., Okuyama, K., et al. (2003) Collagen stabilization at atomic level: crystal structure of designed (GlyProPro)10foldon. *Structure* **11**, 339-346

47. De Simone, A., Vitagliano, L., and Berisio, R. (2008) Role of hydration in collagen triple helix stabilization. *Biochem Biophys Res Commun* **372**, 121-125

48. Raman, S. S., Parthasarathi, R., Subramanian, V., and Ramasami, T. (2006) Role of aspartic acid in collagen structure and stability: A molecular dynamics investigation. *J Phys Chem B* **110**, 20678-20685

49. Vitagliano, L., Berisio, R., and De Simone, A. (2011) Role of hydration in collagen recognition by bacterial adhesins. *Biophys J* **100**, 2253-2261

50. Oliver-Kozup, H., Martin, K. H., Schwegler-Berry, D., Green, B. J., Betts, C., Shinde, A. V., Van De Water, L., and Lukomski, S. (2013) The group A streptococcal collagen-like protein-1, Scl1, mediates biofilm formation by targeting the extra domain A-containing variant of cellular fibronectin expressed in wounded tissue. *Molecular microbiology* **87**, 672-689

51. Oliver-Kozup, H. A., Elliott, M., Bachert, B. A., Martin, K. H., Reid, S. D., Schwegler-Berry, D. E., Green, B. J., and Lukomski, S. (2011) The streptococcal collagen-like protein-1 (Scl1) is a significant determinant for biofilm formation by group A Streptococcus. *BMC Microbiol* **11**, 262

52. Pawlowski, M., Bujnicki, J. M., and Lukomski, S. (2013) Collagen-like proteins of pathogenic streptococci: biology, structure and function. in *7th International Conference on Gram-Positive Microorganisms*, Montecatini Terme, Italy

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

54. Beres, S., Sesso, R., Pinto, S., Hoe, N., Porcella, S., Deleo, F., and Musser, J. (2008) Genome sequence of a lancefield group C *Streptococcus zooepidemicus* strain causing epidemic nephritis: new information about an old disease. *PLoS ONE* **3**, e3026

55. Karlstrom, A., Jacobsson, K., Flock, M., Flock, J., and Guss, B. (2004) Identification of a novel collagen-like protein, SclC, in *Streptococcus equi* using signal sequence phage display. *Vet Microbiol* **104**, 179-188

56. Cheng, A. G., McAdow, M., Kim, H. K., Bae, T., Missiakas, D. M., and Schneewind, O. (2010) Contribution of coagulases towards Staphylococcus aureus disease and protective immunity. *PLoS Pathog* **6**, e1001036

57. Cantisani, M., Falanga, A., Incoronato, N., Russo, L., De Simone, A., Morelli, G., Berisio, R., Galdiero, M., et al. (2013) Conformational Modifications of gB from Herpes Simplex Virus Type 1 Analyzed by Synthetic Peptides. *J Med Chem*

58. Koellhoffer, J. F., Malashkevich, V. N., Harrison, J. S., Toro, R., Bhosle, R. C., Chandran, K., Almo, S. C., and Lai, J. R. (2012) Crystal structure of the Marburg virus GP2 core domain in its postfusion conformation. *Biochemistry* **51**, 7665-7675

59. Lee, J. E., Fusco, M. L., Hessell, A. J., Oswald, W. B., Burton, D. R., and Saphire, E. O. (2008) Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. *Nature* **454**, 177-182

60. Buzon, V., Natrajan, G., Schibli, D., Campelo, F., Kozlov, M. M., and Weissenhorn, W. (2010) Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions. *PLOS Pathogens* **6**, e1000880

61. Heldwein, E. E., Lou, H., Bender, F. C., Cohen, G. H., Eisenberg, R. J., and Harrison, S. C. (2006) Crystal structure of glycoprotein B from herpes simplex virus 1. *Science* **313**, 217-220

62. Tran, E. E., Borgnia, M. J., Kuybeda, O., Schauder, D. M., Bartesaghi, A., Frank, G. A., Sapiro, G., Milne, J. L., et al. (2012) Structural mechanism of trimeric HIV-1 envelope glycoprotein activation. *PLOS Pathogens* **8**, e1002797

63. Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (1997) Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387**, 426-430

64. Boudko, S. P., Engel, J., and Bachinger, H. P. (2012) The crucial role of trimerization domains in collagen folding. *Int J Biochem Cell Biol* **44**, 21-32

65. Than, M. E., Henrich, S., Huber, R., Ries, A., Mann, K., Kuhn, K., Timpl, R., Bourenkov, G. P., et al. (2002) The 1.9-A crystal structure of the noncollagenous (NC1) domain of human placenta collagen IV shows stabilization via a novel type of covalent Met-Lys cross-link. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 6607-6612

66. Kvansakul, M., Bogin, O., Hohenester, E., and Yayon, A. (2003) Crystal structure of the collagen alpha1(VIII) NC1 trimer. *Matrix Biol* **22**, 145-152

67. Bogin, O., Kvansakul, M., Rom, E., Singer, J., Yayon, A., and Hohenester, E. (2002) Insight into Schmid metaphyseal chondrodysplasia from the crystal structure of the collagen X NC1 domain trimer. *Structure* **10**, 165-173

68. Wirz, J. A., Boudko, S. P., Lerch, T. F., Chapman, M. S., and Bachinger, H. P. (2011) Crystal structure of the human collagen XV trimerization domain: a potent trimerizing unit common to multiplexin collagens. *Matrix Biol* **30**, 9-15

69. Boudko, S. P., Sasaki, T., Engel, J., Lerch, T. F., Nix, J., Chapman, M. S., and Bachinger, H. P. (2009) Crystal structure of human collagen XVIII trimerization domain: A novel collagen trimerization Fold. *Journal of molecular biology* **392**, 787-802

70. Rety, S., Salamitou, S., Garcia-Verdugo, I., Hulmes, D. J., Le Hegarat, F., Chaby, R., and Lewit-Bentley, A. (2005) The crystal structure of the Bacillus anthracis spore surface protein BclA shows remarkable similarity to mammalian proteins. *The Journal of biological chemistry* **280**, 43073-43078

71. Cunningham, M. W. (2000) Pathogenesis of group A streptococcal infections. *Clin. Microbiol. Rev.* **13**, 470-511

72. O'Loughlin, R., Roberson, A., Cieslak, P., Lynfield, R., Gershman, K., Craig, A., Albanese, B., Farley, M., et al. (2007) The epidemiology of invasive group A streptococcal infection and potential vaccine implications: United States, 2000-2004. *Clin Infect Dis* **45**, 853-862

73. Shulman, S. T., Stollerman, G., Beall, B., Dale, J. B., and Tanz, R. R. (2006) Temporal Changes in Streptococcal M Protein Types and the Near-Disappearance of Acute Rheumatic Fever in the United States. *Clin. Infect. Dis.* **42**, 441-447
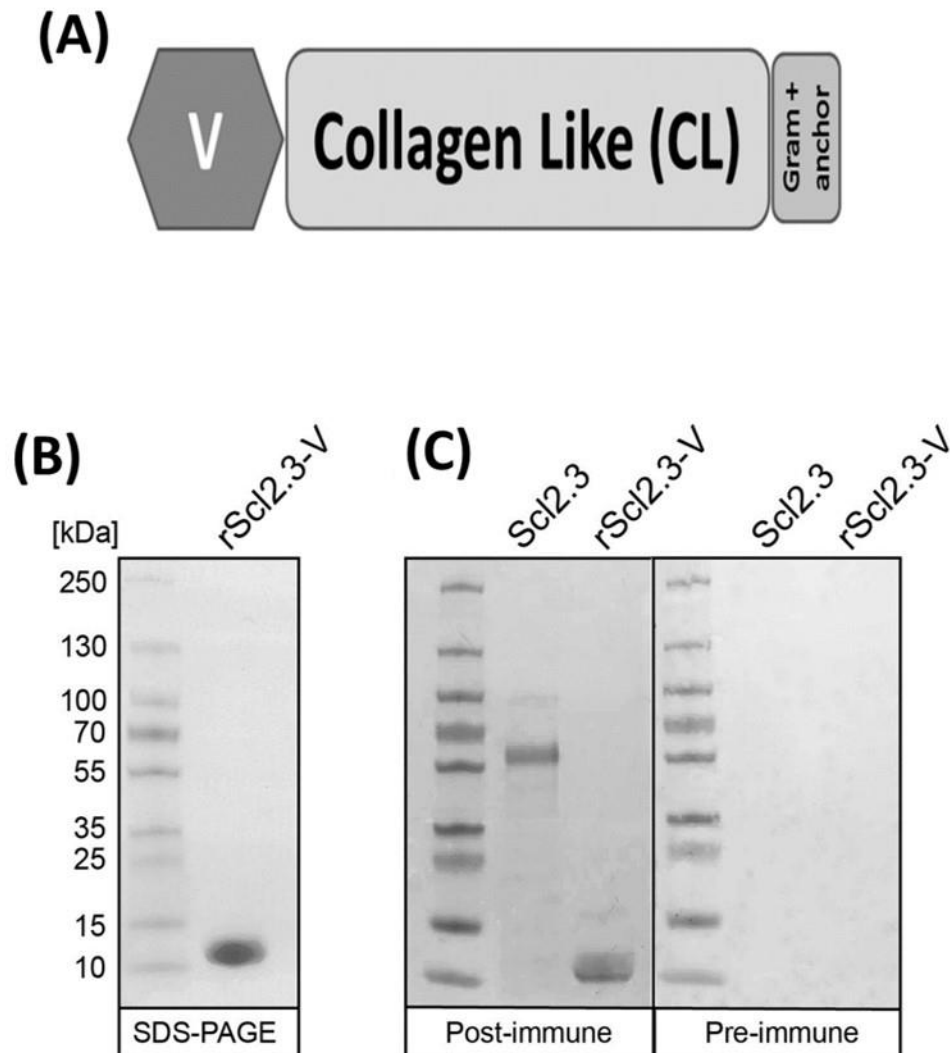
Figure 1

Expression of Scl2.3 surface protein by the invasive M3-type strain MGAS 315. (A) Domain organization of Scl2.3, according to the PFAM database. (B) Characterization of recombinant Scl2.3-V protein (rScl2.3-V). rScl2.3-V polypeptide expressed in *E. coli* and affinity purified was resolved on 4-20% SDS-PAGE gradient gel and stained. Purified rScl2.3-V was used to immunize rabbits and the resulting post-immune sera were used to detect native Scl2.3 protein expressed by MGAS315. (C) Surface expression of Scl2.3 protein by GAS. Cell wall-associated protein fraction of MGAS315 was analyzed by Western immunoblotting to detect native Scl2.3 protein using anti-rScl2.3-V post-immune rabbit sera. The same blot developed with pre-immune sera from the same rabbit is shown as a negative control. PageRuler™ Plus Prestained Protein Ladder was used as a molecular weight marker.
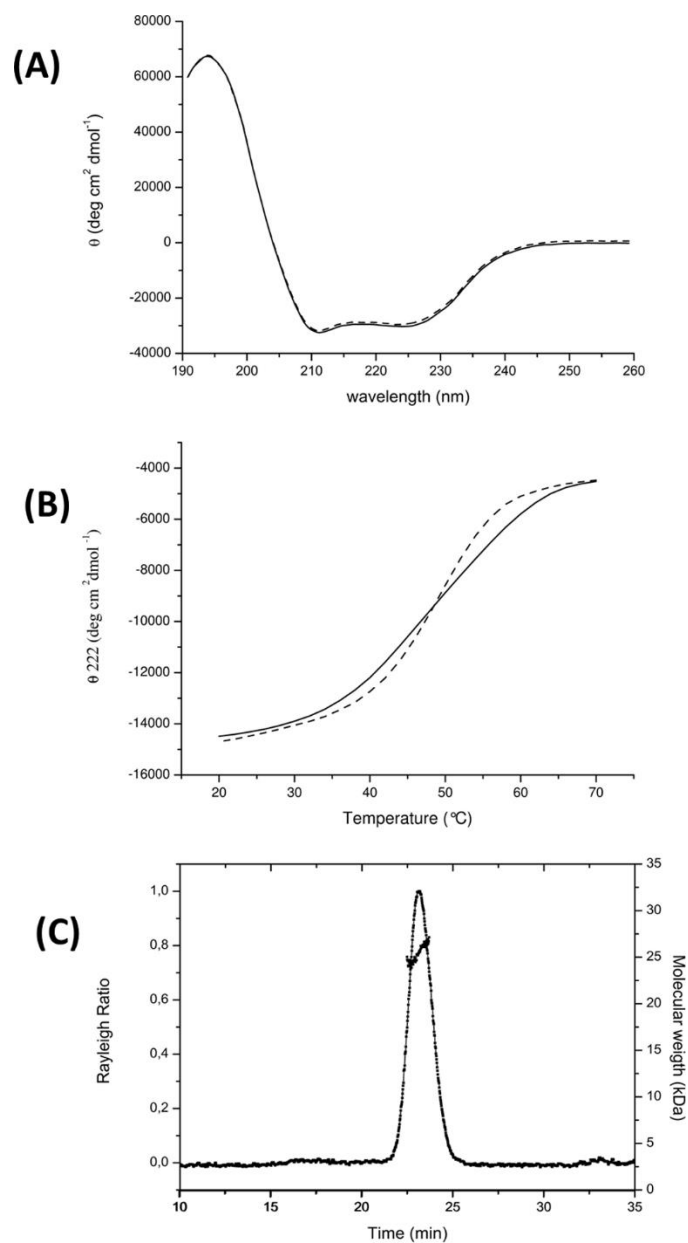
Figure 2

Structural characterization of Scl2.3-V in solution. (A) CD spectra of Scl2.3-V at 20°C after refolding are shown as solid and dashed lines, respectively. (B) Denaturation (solid line) and refolding (dashed line) of Scl2.3-V followed at λ=222 nm. (C) Analytical SEC-MALS; The black curve represents the Rayleigh ratio (left scale) against the retention time. Molecular mass (right scale) values correspond to a trimeric state of Scl2.3-V.
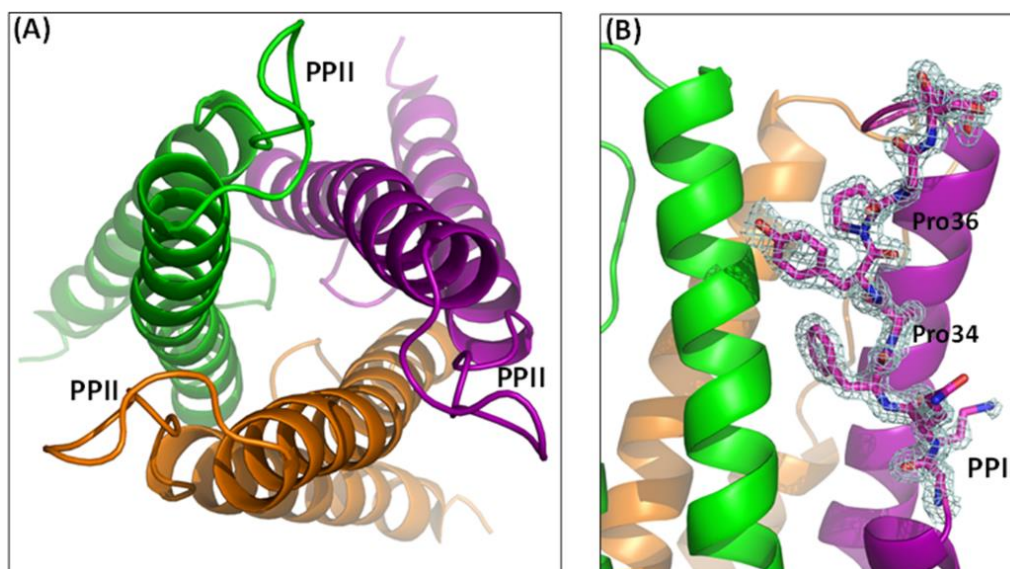
Figure 3

Cartoon representation of the crystal structure of Scl2.3-V trimer. (A) Top view of Scl2.3-V six-helix bundle. (B) A detail of (2Fo-Fc) Electron density on the polyproline II (PPII) region.
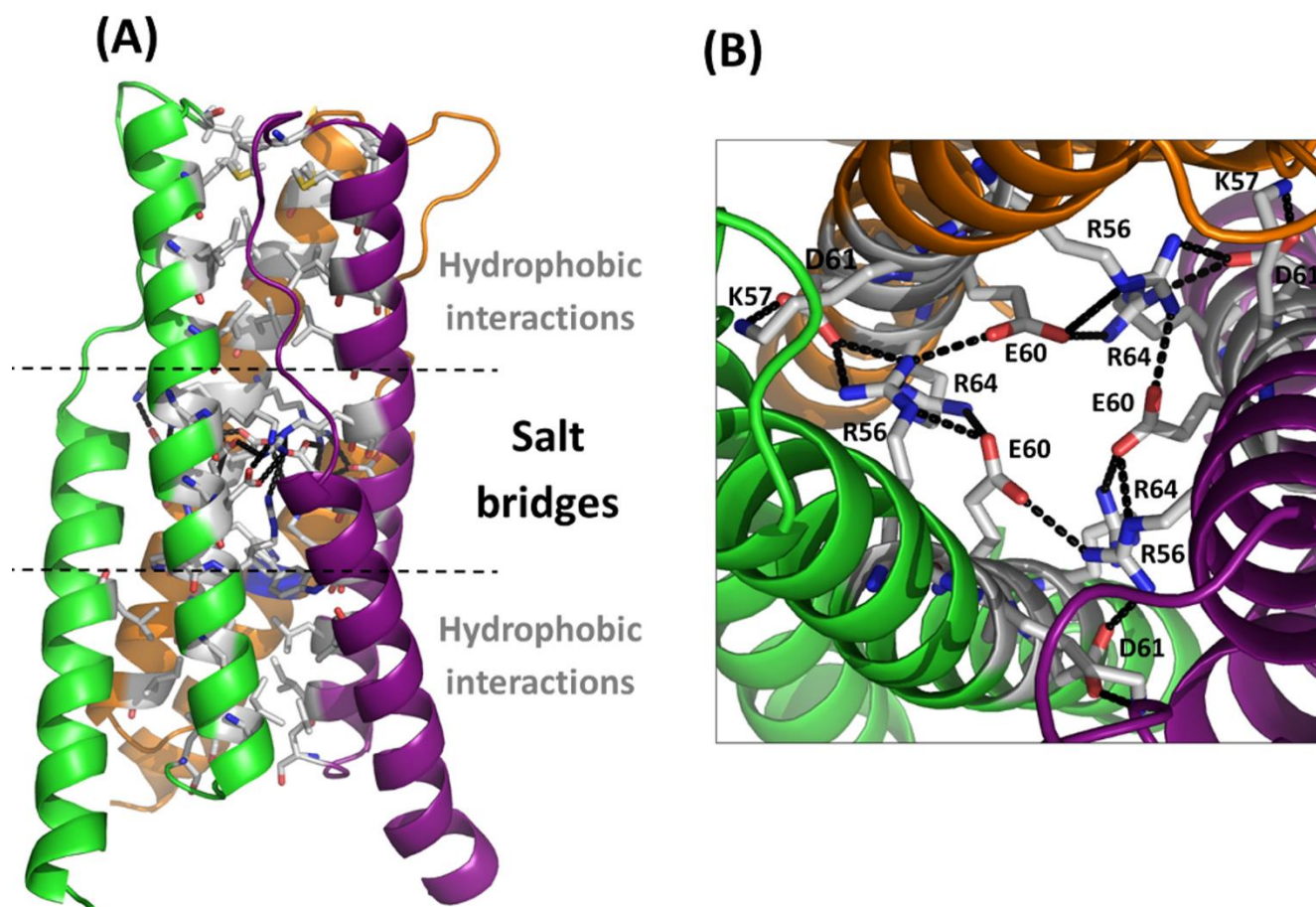
Figure 4

Main interactions stabilizing the six-helix bundle of Scl2.3-V. (A) Three regions are identifiable, with the central one stabilized by salt bridges and the external ones stabilized by hydrophobic interactions. (B) A detail of salt bridge interactions in the central part of the six-helix bundle.
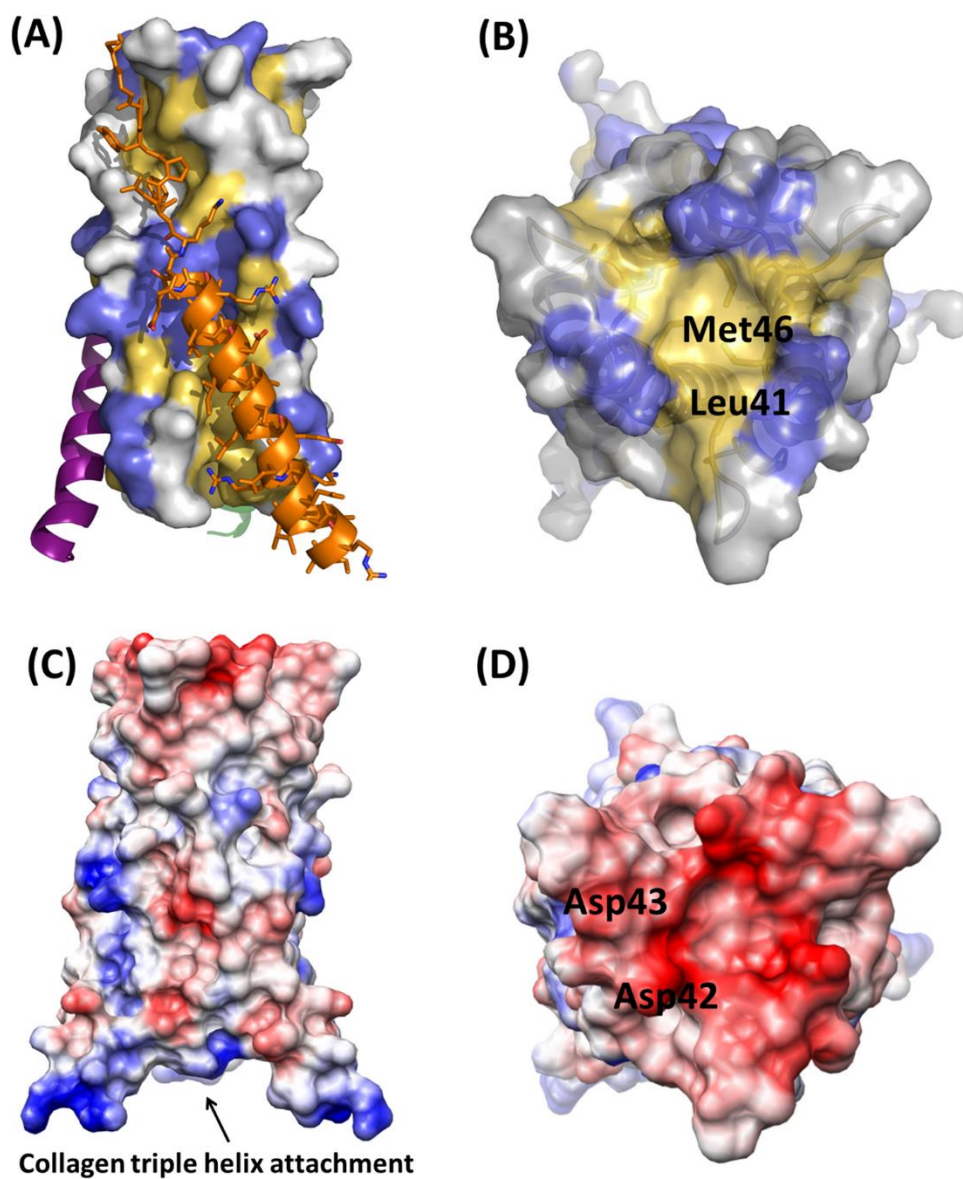
Figure 5

Surface features of Scl2.3-V. (A) Side and (B) top view of Scl2.3-V surface. Hydrophobic and hydrophilic residues are colored yellow and gray, respectively. Charged residues are colored blue. (C, D) Side and top views of Scl2.3-V electrostatic potential surface. The positive and negative charges are colored blue and red, respectively.
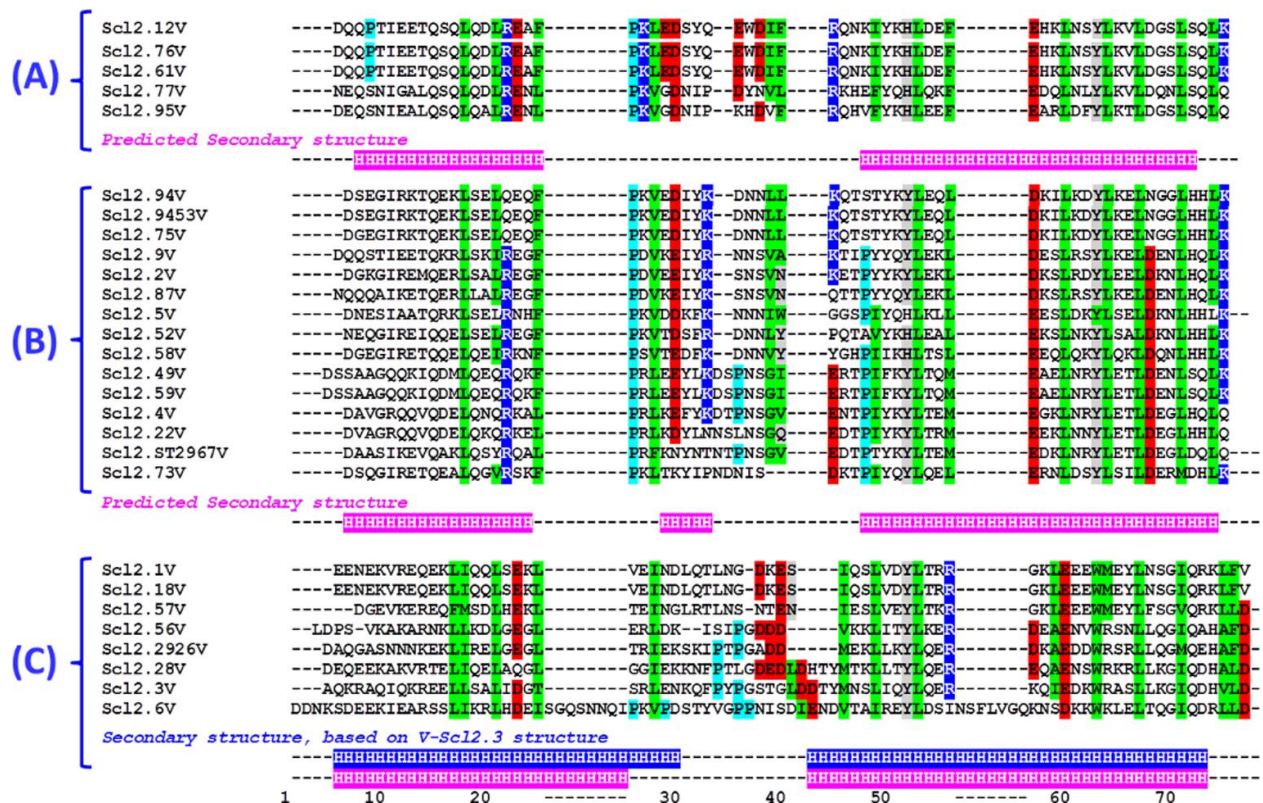
Figure 6

Multiple sequence alignment analysis of Scl2-V region variants. (A) Sequence alignment was performed using CLUSTALW. Conserved residues are shown in green (hydrophobic), blue (positively charged), red (negatively charged), and grey (polar); Pro residues are shown in cyan. A, B and C refer to the three respective branches calculated by phylogenetic analysis using CLUSTALW. Secondary structure prediction, according to JPRED, is reported in magenta whereas the secondary structure based on Scl2.3-V crystal structure is reported in blue.
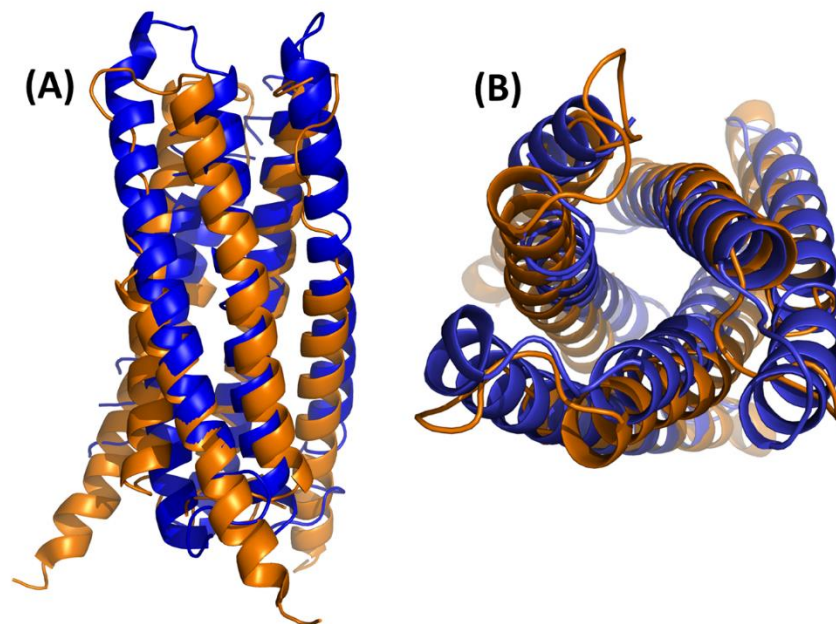
Figure 7

Scl2.3-V structural alignment with the HIV protein gp41. Side (A) and top (B) views of the superposition between Scl2.3-V (orange) with gp41 (blue, PDB code 3o40).

(A)

(B)

(C)

(D)

Global

Scl2.3-CL
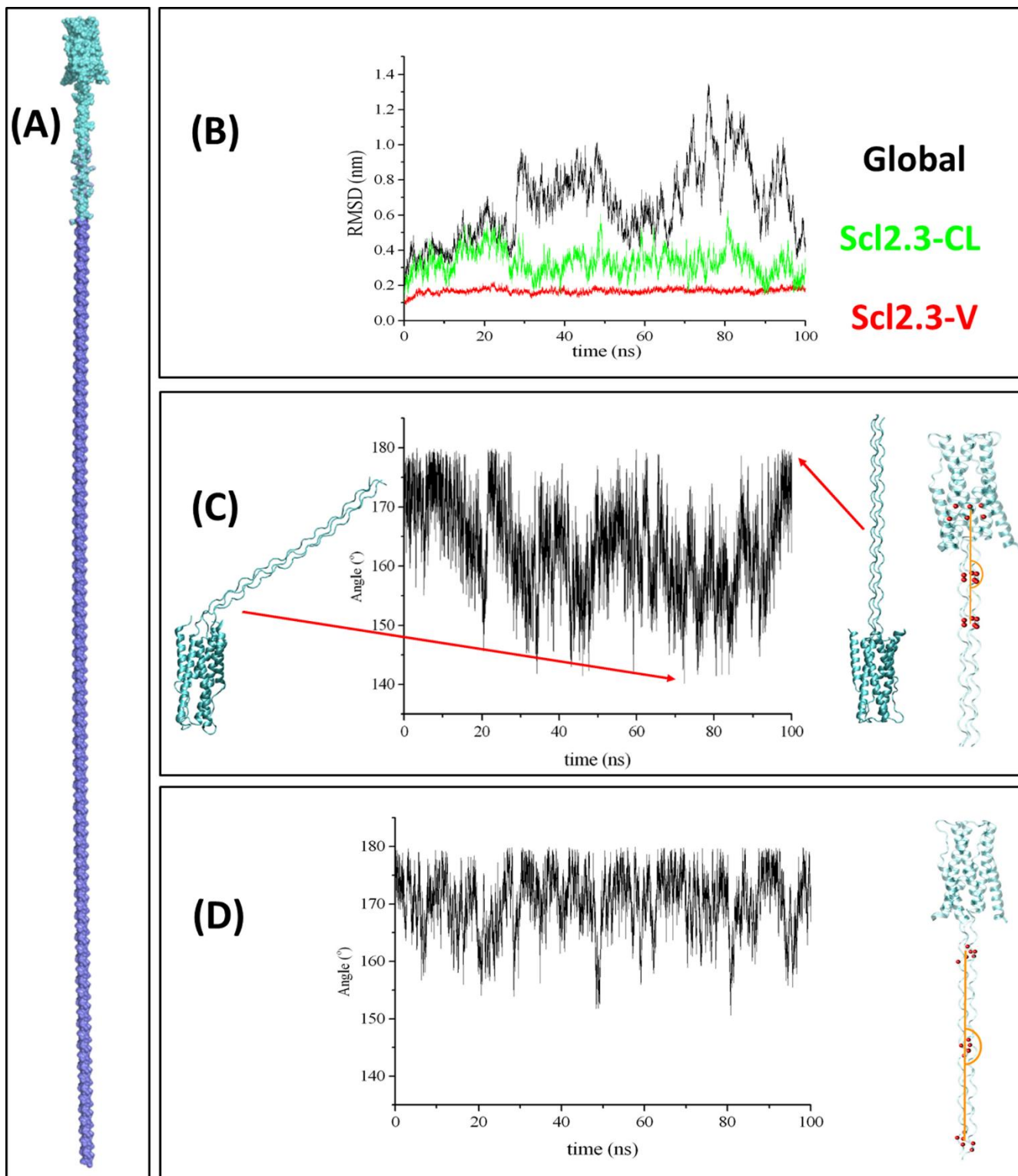
Scl2.3-V

Modeling and molecular dynamics analyses. (A) Modeling of Scl2.3 structure: a view of the entire Scl2.3 structure; The cyan region (residues 7-123 of each chain of the Scl2.3 trimer) was used for MD studies. (B) Evolution of the RMSD calculated on Cα atoms, throughout the MD trajectory for the entire model (black), the Scl2.3-V (red) and Scl2.3-CL (green). (C) Evolution of the bending angle between the Scl2.3-V and Scl2.3-CL domains throughout the MD trajectory; (D) Evolution of the bending angle between three zones of the Scl2.3-CL domain throughout the MD trajectory; The insets of panels C and D show atoms used for the computation of the bending angle as red balls.

Table 1. Data collection and refinement statistics

A. *Data collection*[a]

	Europium chloride derivative	Native
Space group	H32	H32
a, b, c (Å), γ (°)	44.3, 44.3, 228.0, 120.0	44.4, 44.4, 227.8, 120.0
Resolution (Å)	1.87	1.52
Average redundancy	9.5 (7.5)	5.3 (2.6)
Completeness (%)	100 (99.9)	99.2 (86.7)
Rmerge (%)	0.043 (0.419)	0.061 (0.345)
Average I/σ(I)	47.7 (5.1)	49.2 (3.1)

B. Refinement

Resolution range (Å)	15.00-1.52
Rwork, Rfree (%)	18.5, 23.3
No. atoms (protein, ligands, water)	623, 13, 100
r.m.s. deviations	
Bond lengths (Å)	0.02
Bond angles (°)	1.6

[a]Values in parentheses are for highest resolution shells: 1.90-1.87 Å for $EuCl_3$ derivative and 1.55-1.52 Å for the native crystal.

PART III

ADDITIONAL PUBLISHED DATA

CHAPTER 6

REVERSIBLE PHOSPHOLIPID NANOGELS FOR DEOXYRIBONUCLEIC ACID FRAGMENT SIZE DETERMINATIONS UP TO 1,500 BASE PAIRS AND INTEGRATED SAMPLE STACKING

Brandon C. Durney, Beth A. Bachert, Hillary S. Sloane, Slawomir Lukomski, James P. Landers, and Lisa A. Holland

ABSTRACT

Phospholipid additives are a cost-effective medium to separate deoxyribonucleic acid (DNA) fragments and possess a thermally-responsive viscosity. This provides a mechanism to easily create and replace a highly viscous nanogel in a narrow bore capillary with only a 10 °C change in temperature. Preparations composed of dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) and 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine (DHPC) self-assemble, forming structures such as nanodisks and wormlike micelles. Factors that influence the morphology of a particular DMPC-DHPC preparation include the concentration of lipid in solution, the temperature, and the ratio of DMPC and DHPC. It has previously been established that an aqueous solution containing 10% phospholipid with a ratio of [DMPC]/[DHPC] = 2.5 separates DNA fragments with nearly single base resolution for DNA fragments up to 500 base pairs in length, but beyond this size the resolution decreases dramatically. A new DMPC-DHPC medium is developed to effectively separate and size DNA fragments up to 1,500 base pairs by decreasing the total lipid concentration to 2.5%. A 2.5% phospholipid nanogel generates a resolution of 1% of the DNA fragment size up to 1,500 base pairs. This increase in the upper size limit is accomplished using commercially available phospholipids at an even lower material cost than is achieved with the 10% preparation. The separation additive is used to evaluate size markers ranging between 200 and 1,500 base pairs in order to distinguish invasive strains of *Streptococcus pyogenes* and *Aspergillus* species by harnessing differences in gene sequences of collagen-like proteins in these organisms. For the first time, a reversible stacking gel is integrated in a capillary sieving separation by utilizing the thermally-responsive viscosity of these self-assembled phospholipid preparations. A

discontinuous matrix is created that is composed of a cartridge of highly viscous phospholipid assimilated into a separation matrix of low viscosity. DNA sample stacking is facilitated with longer injection times without sacrificing separation efficiency.

INTRODUCTION

Estimating the size of DNA is critical in genetic analysis associated with human identification (1), species identification (2,3), detecting personal biomarkers (4), analyzing food (5), and categorizing pathogenicity (6). Moreover, following sample preparation, accurate determination of the size of DNA fragments is a mandatory analytical step in sophisticated high-throughput sequencing techniques. It is also a critical endpoint for methods based on PCR amplification, where the length of the product is used to determine the presence of a specific target sequence in the DNA template. Confirmation of size is achieved by comparing the fragment migration time during electrophoretic sieving using various sieving agents. Slab gel electrophoresis, the workhorse of modern analytical science, is a commonly used tool for DNA sizing due to the simplicity and low cost of the method. Sizing is achieved by comparing the migration time of the targeted DNA with a DNA ladder that contains a mixture of fragments of known size. The DNA ladder standard is run in a separate lane. A serious disadvantage of slab gels is the poor separation efficiency and the reduced throughput even when completed in parallel lanes in the gel (7,8). Although more expensive than slab gel methods, capillary gel electrophoresis provides a size discrimination that is improved substantially, and higher sample throughput. Sequencing of the human genome provides direct proof of this (9).

Capillary gel electrophoresis separations of DNA fragments are accomplished using polymer solutions that provide a size-based separation of DNA fragments by creating networks or matrices that form dynamic pores for sieving (10-12), or separation through DNA entanglement (13). DNA fragments ranging from 50-1,500 bases are often present in an electropherogram, but the resolution obtained for the larger DNA fragments is poor and accurate sizing is limited to DNA fragments that are shorter than 500 bases in length. Ultra-high resolution separations were achieved with 0.75% hydroxyethyl cellulose used to size a 256 base pair (bp) DNA fragment (14). Although larger fragments could not be sized, 1,078 bp and 1,353 bp DNA were baseline resolved (14). A

162

commercial system utilized hydroxyethyl cellulose to separate 50 bp to 10 kbp DNA and to accurately size 275 to 815 bp DNA (15). In another case, a matrix of 2% linear polyacrylamide was used to sequence a DNA fragment 1,300 bases long (16). Several polymer properties dictate the utility of the material and the useful size range of the separation (17). Increasing the upper fragment size limit of the separation can be accomplished by increasing the length of the polymer and maintaining the concentration of the gel at or above the entanglement threshold (10). This increase in concentration is accompanied by an increase in viscosity (11), which provides challenges for the introduction to or replacement of, the gel in the separation capillary.

Temporary gels, which rupture and reform during the separation, are an exciting alternative to permanent gels because the upper size limit increases when the rate of gel rupture is optimized (18). Wormlike micelles form large self-assembled structures that dynamically rupture and reform. Aqueous solutions of the long-chain phospholipid dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) and the short-chain phospholipid 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine (DHPC) are known to form wormlike micelles (19,20). A dynamic sieving matrix of 10% phospholipid with [DMPC]/[DHPC] = 2.5 was optimized for single-base pair resolution that was adequate for detection of short tandem repeats (STR) in selected loci utilized for human identification using a 100 cm long capillary within a 30 minute run time (21). This is comparable to separations on commercial capillary electrophoresis instruments for forensic STR analyses (22). What makes this particularly attractive, is that the phospholipid nanogel is substantially more cost-effective at roughly one-fourth the cost of commercial gels utilized for human identification (23-25). A significant advantage of the dynamic sieving matrix is the thermally-responsive viscosity. The sieving matrix has low viscosity below the gel-phase transition temperature of the phospholipids and becomes gel-like at higher temperatures. As a result, the sieving agent is easily introduced into or expelled from the narrow bore separation capillary at a temperature of 19 °C. Prior to the separation, the matrix is transformed into a viscous gel in-capillary at a separation temperature of 30 °C.

While a phospholipid additive has been previously described for the separation of DNA fragments below 500 bp (21), we show that the DNA sizing range of the separation can be extended by reducing the concentration of wormlike micelles in solution. The size,

shape, and stability of the phospholipid structures depend upon the phospholipid concentration, the ratio of DMPC and DHPC, and the temperature. The effect of these parameters on the size limit of DNA separations was evaluated, and the linear range for size separation was extended to 1,500 bp by tuning the concentration of the phospholipid nanogel and the ratio of [DMPC]/[DHPC]. This preparation was used to compare the size of the DNA fragments to internal standards. The effectiveness of the sieving matrix was demonstrated with the analysis of PCR-amplified DNA sequences specific to select pathogens. The difference between the true size of a DNA fragment and the size calculated from internal standards is less than 2.3% for fragment sizes between 200-500 bp. For fragments that range from 600-1,500 bp this difference between the true size and measured size is less than or equal to 4%. Different properties of the high and low concentration phospholipids were exploited by integrating both dynamic sieving matrices into a single capillary to create a discontinuous gel in-capillary used to concentrate DNA samples through gel stacking.

MATERIALS AND METHODS

Chemicals and reagents

The fluorescently labeled DNA ladder (catalog # MM-1000-FAM, 50-1,000 bp) was purchased from BioVentures (Murfreesboro, TN). The fluorescent intercalating dye SYBR Green 1 nucleic acid gel stain (catalog # S-7567), 1 Kb Plus (catalog # 10787-018, 100-12,000 bp), and 100 bp (catalog # 15628-019, 100-2,072 bp) DNA ladders were purchased from Life Technologies (Grand Island, NY). The methanol was purchased through Calbiochem/EMD (Gibbstown, NJ). The DMPC and DHPC phospholipids were purchased through Avanti Polar Lipids (Alabaster, AL), and the 3-(*N*-morpholino)-propanesulfonic acid (MOPS) was purchased from Alfa Aesar (Ward Hill, MA). Sodium hydroxide was obtained from Sigma-Aldrich (St. Louis, MO). All deionized water used was obtained from an ELGA PURELAB ultra water filtration system (Lowell, MA).

Preparation of phospholipid mixture

The phospholipid preparations were made as described previously (26) to obtain molar ratios of [DMPC]/[DHPC] = 0.5 and [DMPC]/[DHPC] = 2.5. The phospholipids were stored

at -20 °C and thawed before being weighed out. Once they were weighed, an aqueous buffer of 100 mM MOPS (pH 7.0) was added to obtain the appropriate concentration of phospholipid per volume of aqueous diluent (% w/v). The solution was then mixed using a vortex mixer for several minutes to dissolve the phospholipids. The phospholipid medium was then exposed to three rapid freeze thaw cycles using liquid nitrogen followed by centrifugation (10,000 rpm at 4 °C) for 10 minutes. The matrix was then portioned into working volume aliquots of 200 µL [DMPC]/[DHPC] = 0.5 (5% w/v) and 100 µL [DMPC]/[DHPC] = 2.5 (10% w/v). The [DMPC]/[DHPC] = 2.5 lipids were then diluted with the same 100 mM MOPS (pH 7.0) buffer to obtain the 7.5, 5.0, and 2.5% phospholipid used for electrophoresis. No freeze thaws were done after the dilution from the 10% phospholipid. The additional phospholipid matrices with [DMPC]/[DHPC] = 3.0, 4.0, and 5.0 used in the study were prepared the same manner. The MOPS buffer was also used for capillary rinses between runs and served as the solution in the anodic and cathodic reservoirs during electrophoresis separations.

PCR amplification

Primers were based on *scl2* genes with the following gene bank accession numbers and genome locations; M3: AE014074, location 791329-793050; M12: CP000259, location 862738-863880; M28: CP000056, location 781169-782437. Forward primers specific for each *scl2* allele, are as follows: M3-Scl2.3_CLflank_2F, 5'-AGGCATACAAGATCATGTCCTTGA-3'; M12-Scl2.12_CLfl_1F, 5'-AAGAGTGGGATATCTTTAGGC-3'; M28-Scl2.28_CLfl_1F, 5'-CTACAGGAACGAGAACAAGC-3'. A single reverse primer Scl2.3_CLflank_2R, 5'-TTTGGTGTATGTGCTGCGGT-3', targeting a conserved 3' region of *scl2* gene was paired with each forward primer. Genomic DNA was purified and isolated as described previously (27). PCR reactions were carried out with 1.5 units of *Taq* DNA polymerase (Qiagen, Germantown, MD) in a final volume of 50 µL per reaction and DNA template in the nanogram range using a C1000 Touch Thermal Cycler (BioRad, Hercules, CA). The manufacturer's PCR buffer was supplemented with 0.2 µM concentration of each primer and 0.2 mM deoxynucleotide triphosphates (Bioline, Taunton, MA). Amplification conditions consisted of an initial denaturation step at 94 °C for 3 min, followed by 30

cycles of 1 min at 94 °C, 1 min at 51 °C, and 1 min 45 s at 72 °C, followed by a final extension step at 72 °C for 7 min. The amplicons were analyzed by agarose gel electrophoresis and the true size of each *scl2*-PCR product was determined by Sanger DNA sequencing (Eurofins Genomics, Huntsville, AL). The DNA sequences were assembled and analyzed with the Lasergene 11 Genomics Suite (DNASTAR, Inc., Madison, WI) software, as previously described (28). Isolation of genomic DNA and purification from fungal cultures, PCR primers (Eurofins Genomics, Huntsville, AL), amplification conditions, and verification of conserved amplicon sizes for *acl*-based detection of *Aspergillus* species were previously established (29).

Gel extraction of DNA ladder standards

Fragments from the two unlabeled DNA ladders were separated using a VWR horizontal gel box Midi Plus 15x10cm connected to a Hoefer hsi (San Francisco, CA) power supply using a 1% agarose (CalbioChem/EMD Chemicals) in 1X TBE buffer (G Biosciences, St. Louis, MO) gel. Slab gel electrophoresis was accomplished using an applied voltage of 110 V for 3 hours, and alternate gel lanes were stained using ethidium bromide (part # 15585011, Life Technologies) for UV visualization. Extraction and purification of DNA bands in unstained lanes was done using the illustra™ GFX™ PCR DNA and Gel Band Purification Kit manufactured by GE Healthcare (Buckinghamshire, UK) prior to analysis on the P/ACE MDQ Capillary electrophoresis instrument (Beckman Coulter, Fullerton, CA). DNA that was isolated and recovered from the agarose gel was then utilized as an internal standard to size PCR amplicons. The 600, 1,000, 1,500, and 1,650 bp fragments were used as internal standards to size *S. pyogenes* and *B. anthracis* amplicons, whereas the 150, 450, and 500 bp were used to size *Aspergillus* spp. amplicons.

Pre-treating the separation capillary

Separations were performed using fused silica capillary with an outer diameter of 360 μm and an inner diameter of 25 μm (Polymicro Technologies, Phoenix, AZ). The separation capillary had a total length of 40 cm, and an effective length of 30.2 cm. New capillary was conditioned with a 4-step rinse procedure performed at 140 kPa with (1) 1N NaOH for 30 min, (2) deionized water for 15 min, (3) methanol for 15 min, (4) deionized water

for 15 min. The conditioned capillary was then passivated with a semi-permanent phospholipid coating to suppress electroosmotic flow as previously described (30). This was accomplished using a 2-step rinse at 140 kPa with (1) 5% phospholipid of [DMPC]/[DHPC] = 0.5 containing 1.25 mM Ca^{2+} for 20 min and (2) a 2 min flush with 100 mM MOPS buffered to pH 7. This passivation rinse is repeated once (i.e., 4-steps). The capillary was then filled with phospholipid nanogel at an ambient temperature of 19 °C. Once filled with the phospholipid, the capillary temperature was increased to the desired separation temperature. A reverse polarity voltage of 20 kV was applied for 5 min prior to the first run of the day.

Capillary electrophoresis sample introduction and separation

Sample was injected by first introducing a pre-plug of run buffer at 6.9 kPa for 7 sec, then applying a reverse polarity electrokinetic injection of DNA for the specified time, and finally by inserting a post-plug of run buffer at 3.4 kPa for 5 sec. The purpose of these pre- and post-plugs was to facilitate stacking and improve the theoretical plate count (31). Separations were performed at 4 kV under reverse polarity conditions. Between each separation the capillary was flushed at 140 kPa for 3 min with 5% [DMPC]/[DHPC] = 0.5, 2 min MOPS (pH 7), and 3 min with the phospholipid preparation used to separate DNA.

Fluorescence detection and analysis of DNA

The Beckman Coulter P/ACE MDQ was equipped with a laser induced fluorescence detection module and an air cooled argon ion laser (λ_{ex}=488 nm and λ_{em}=520 nm). The optimization of concentration and composition of phospholipid nanogel was accomplished using a fluorescein end-labeled DNA ladder. The intercalating dye SYBR Green 1 was utilized for separations of PCR amplicons. SYBR dye was diluted to a concentration of 20 μM SYBR in phospholipid preparations of [DMPC]/[DHPC] = 2.5 at a concentration equal to that used for the separation. The concentration of the plug was calculated based on a literature estimate of 20 mM stock concentration from the manufacturer (32). Unless otherwise stated, a 7 second 34.5 kPa plug of phospholipid modified with dye was pushed toward the detection window in the separation capillary prior to sample injection. Under an applied field, the positively-charged dye and negatively-charged DNA migrated toward

each other and the two bound effectively upon contact as the K_d is 3.1 nM (33). Any variation in the ratio of dye and DNA in a separation will change the migration time of the DNA. Reproducible run times are achieved by utilizing a dye concentration compatible with the detector output, in this case an output of approximately 1-5 relative fluorescent units was achievable with a plug of 20 μM SYBR Green 1. As summarized in Table S-1, the reproducibility in migration time is less than or equal to 5% for separations with a 7 second dye plug. The use of internal size standards in each separation further mitigates any bias in migration time due to variation in the ratio of dye to DNA.

The size of the PCR amplicons was calculated using internal standards purified from the DNA ladders listed in the materials section. Sizing of fragments was accomplished using two DNA standards that bracketed the size of the PCR amplicon. All DNA fragments were co-injected and a linear fit was used to relate the migration time to the known size in bp of the DNA standards. Using this fit, the size of each PCR amplicon was calculated based on the migration time. Resolution between fragments was calculated using the expression $Rs = 0.589*(\Delta t/w_{1/2av})$, where Δt is the difference in migration time between the two peaks of interest, and $w_{1/2av}$ is the average width at half height for the same two peaks. Resolution in base pairs was calculated by dividing the difference in the number of base pairs for the two fragments (Δbp) by the calculated resolution. Therefore, base pair resolution can be expressed as $Rs_{bp}=(\Delta bp)/(0.589*(\Delta t/w_{1/2av}))$. Data collection and analysis were performed using 32 Karat Software version 7.0 (Beckman Coulter), and peak width at 50% height as well as theoretical plate count were calculated using the "USP" criterion.

RESULTS AND DISCUSSION

Effect of gel concentration on the separation

The total lipid concentration (% w/v) determined the number of 'ribbons' or 'disks' in solution and the degree of long-range interaction, including the extent of lipid entanglement of the nanogel. Using a fluorescently-labeled DNA ladder, the separations were examined in the 500-1,000 bp range to quantify the effect of phospholipid concentration on separation performance. Our previous literature reported the effect of concentrations of phospholipid ranging from 5-15% for separations based on subtle difference in hydrodynamic volume (26) and from 6-12% for separations based on sieving (21). For both studies (21,26) a 10% phospholipid concentration generated the most effective separation efficiency and peak resolution. In the current study concentrations of 2.5, 5.0, 7.5, and 10% were used to evaluate the effect of phospholipid concentration on the useful upper size limit for DNA separations. Preparations below 2.5% phospholipid did not generate reproducible DNA separations, presumably because wormlike micelles are not stable under these conditions, and consequently, these were not assessed. The peak resolution summarized in Table 1 was measured for a set of DNA fragments at 450 and 475 bp as well as a set at 900 and 950 bp. There was no significant improvement in resolution of DNA fragments with a length of 450 and 475 bp for separations accomplished at 23 °C. When the separation was performed at the optimum temperature for each lipid preparation, which was 25 °C for 5% nanogel, and 30 °C for both 7.5% and 10% nanogel, the resolution of DNA fragments smaller than 500 bp was similar when comparing a 2.5% to a 10% phospholipid nanogel (see Table S-2 in the supporting information). The resolution between the DNA fragments with a length of 900 and 950 bp improved substantially (2.3-fold) when comparing a nanogel of 2.5% phospholipid to one of 10% phospholipid. When the separation was performed at the optimum temperature for each lipid preparation (i.e., 25 °C for 5% nanogel, and 30 °C for both 7.5% and 10% nanogel), the resolution between fragments of 900 and 950 bp improved 1.8-fold (see Table S-2 in the supporting information). These results indicated that the lipid preparations should be maintained at a concentration of 2.5% lipid.

The effect of temperature and composition, factors that impact nanogel morphology, on separation

The ratio of long-to-short-chain lipid determines the shape of the self-assembled aggregate formed in solution. When the relative amount of long chain lipid increases, the aggregates are transformed from a bilayer disk conformation with a small diameter, to disks with larger diameters and 'ribbons' (19,20). The temperature also impacts the transformation from nanodisks to nanoribbons and a substantial change in apparent viscosity is observed over a temperature range specific to the preparation (34-36). The morphologies adopted by the matrix were dependent on the ratio of long-to-short chain lipid as well as the temperature of the preparation. At a total lipid concentration of 2.5%, preparations were examined with a ratio of long-to-short chain lipid of 2.5, 3.0, and 4.0. The results, summarized in Table 2, revealed that the lipid composition had no significant effect on the resolution of DNA fragments under these conditions. When the ratio of long-to-short chain lipid was 5.0, the separation performance decreased. The resolution improved marginally for nanogels with different lipid composition when the optimum performance was observed at a temperature different from 23 °C (see Table S-3 in the supporting information). Although, the performance of the separation matrix was tolerant of changes in lipid composition, the preparation made with a 2.5 ratio of long-to-short chain lipid was used for analytical separations and the effect of temperature on separation performance was tested. The results for a 2.5% w/v preparation composed of [DMPC]/[DHPC] equal to 2.5 are summarized in Table S-4 in the supporting information. The best resolution for both the 450 and 475 bp DNA and the 900 and 950 bp DNA was obtained at 23 °C. At this temperature 450 and 475 bp DNA fragments were separated with 5 bp resolution, and 900 and 950 bp DNA fragments were separated with 8 bp resolution. For both size domains the resolution obtained using a 2.5% nanogel is 1% when normalized by dividing the resolution (in bp) by the size of the separated DNA fragments. This is a substantial increase over the 10% resolution obtained with a commercial system that utilizes a hydroxypropylmethylcellulose matrix for DNA shorter than 1,000 bp (15). This improved sizing obtained with the phospholipid nanogel is an extraordinary advantage for DNA analyses with sieving that require accurate size determinations.

Expanded size range of sieving using a wormlike nanogel

These optimization studies confirmed that decreasing the concentration of the lipid preparation separated a wider range of DNA fragments than previously reported. This is demonstrated in Figure 1, which depicts the separation of the 1kilobase DNA ladder. The separation performance obtained with the DNA size ladder improves by successively decreasing the viscosity of the nanogel used for sieving DNA. Figure 1 trace A is of 10% lipid with [DMPC]/[DHPC] equal to 2.5 at 30 °C, which is the higher viscosity formulation and provides accurate sizing for DNA fragments smaller than 500 bp in length. When the temperature of the 10% nanogel matrix is decreased the viscosity is decreased as well. The trace in Figure 1 B obtained using a 10% nanogel matrix at a separation temperature of 23°C demonstrates some improvement in the resolution of larger fragments ranging from 1,000 to 12,000 bp in length. The trace in Figure 1C is obtained using a 2.5% nanogel matrix at a separation temperature of 30°C. This material has a lower viscosity than the 10% nanogel matrix and relative to the 10% nanogel the resolution is improved for the fragments ranging from 1,000 to 12,000 base pairs. Figure 1 trace D is of the optimized preparation of 2.5% lipid with [DMPC]/[DHPC] equal to 2.5 at 23 °C. This nanogel has the lowest viscosity and generates the best resolution for DNA fragments larger than 500 bp. For DNA fragments shorter than 500 bp, the performance of the optimized preparation was similar. The resolution of 450 and 475 bp fragments decreased from 6.2, using the 10% phospholipid preparation with [DMPC]/[DHPC] = 2.5 at 30 °C (see Table S-2), to 5.0, using the optimized preparation of 2.5% phospholipid with [DMPC]/[DHPC] = 2.5 at 23 °C. This slight decrease in resolution performance to 78% of the previously reported value of 6.2, did not limit the application of the material to analyze and accurately calculate the size of these PCR amplicons.

Application to assess length polymorphism of PCR amplicons derived from pathogenic bacteria

PCR amplicons were identified using bioinformatics to predict unique DNA biomarkers of different strains of *Streptococcus pyogenes* ranging from 634 to 1,332 bp in length. Over 700 million infections annually are attributed to this pathogenic bacterium associated with common throat and skin diseases with more than 0.5 million deaths resulting from

invasive and autoimmune complications (37). Strains are typically characterized epidemiologically by serological typing of the M protein or, more recently, by sequencing of the *emm* gene (38-40). However, epidemiological surveillance during outbreaks of infection requires additional tools to distinguish between strains of the same M-type. The *scl2* gene, encoding a common *S. pyogenes* surface protein, exhibits significant length polymorphism between strains of the same M-type, due to the presence of repetitive Gly-X-Y motifs in the collagenous domain (28,41). Strain typing by length variation in collagen-like genes has been successfully demonstrated for the *bcl* genes of *Bacillus anthracis* (42) and is a viable strategy for typing *S. pyogenes*. The *scl2* alleles from strains of M-types 3, 12, and 28 were selected here as epidemiological markers since these are among the most prevalent M-types causing infection in the United States and globally (43,44). Importantly, M3-type strains are associated with invasive diseases like necrotizing fasciitis and streptococcal toxic shock syndrome (43,45-47).

Primers were designed to generate PCR amplicons of *scl2* alleles associated with each M-type. The 2.5% phospholipid preparation was utilized to evaluate the applicability of the method for the analysis of PCR amplicons. Three amplicons were separated, as shown in Figure 2a, and sized with internal standards of 600 and 1,000 bp. The 634 bp and 715 bp amplicons, which were sequenced in the genome of three different M12-type strains, and the 842 bp amplicon, which was sequenced in the M28-type strain, were detected and sized using the phospholipid nanogel. The remaining amplicons (1,071, 1,152, and 1,332 bp) sequenced in the M3-type strains were detected and sized (see Fig. 2b) using internal standards of 1,000 and 1,500 bp. The extended sizing range of the phospholipid nanogel enabled detection of larger *scl2*-derived amplicons to generate unique DNA markers that distinguish strains of *S. pyogenes*. To further demonstrate the effectiveness of the phospholipid nanogel for size determination, a previously reported amplicon of *bcl* gene that is 1,446 bp in length (42) was analyzed using internal standards of 1,000 and 1,650 bp as shown in Figure 2c.

Characterization of the difference between true and measured size (bias)

To demonstrate that the phospholipid preparation optimized for the separation of large DNA fragments remained effective for the analysis of DNA fragments below 500 bp in

size, unique biomarkers of several *Aspergillus* species were analyzed (see Figure S-1 in the supporting information). This fungus leads to increased mortality among immunocompromised patients suffering from invasive aspergillosis (48-51) and is also associated with crop contamination (52). Detection of *Aspergillus* through unique DNA sequences is significant because PCR methods are faster and more conclusive than the traditional means of identifying *Aspergillus* species based on morphology observed in microbiology laboratory cell cultures. Unique sequences in *Aspergillus* collagen-like genes (*acl*) found in five pathogenic species were targeted by PCR resulting in amplicons of distinct lengths: *A. nidulans* 227 bp; *A. terreus* 262 bp; *A. niger* 297 bp; *A. flavus* 402 bp; and *A. fumigatus* 338 and 489 bp. Using internal standards of 150 and 450 bp to calculate the size, these six biomarkers were resolved in a capillary with 30.2 cm effective length and a separation additive containing 2.5% phospholipid preparation of [DMPC]/[DHPC] equal to 2.5 at 23 °C. This separation additive optimized for larger DNA fragments estimated the size with a bias ranging from 0-6 bp (see Table 3), which represented a size accuracy within 2.5% of the true fragment size. Previously, these biomarkers were separated using a 40 cm effective length and a 10% phospholipid preparation of [DMPC]/[DHPC] equal to 2.5 at 30 °C (29). In that study, the amplicon size was estimated using three size standards of 150, 250, and 350 bp with a bias ranging from 1-7 bp, which represented a size accuracy within 2% of the true fragment size.

The analysis of DNA biomarkers for strains of *S. pyogenes* revealed that the phospholipid nanogel generated a relative bias in size accuracy less than 1% (2-4 bp bias) of the total fragment size for 634, 715 and 842 bp, and increased to 2%, 3%, and 4% (19-48 bp bias) of the total fragment size for 1,071, 1,152, and 1,332 bp, respectively (Table 4). The analysis of the 1,446 bp PCR amplicon for the *bcl* C gene of *Bacillus anthracis* shown in Figure 2c had an estimated size of 1,449 ± 8 bp (*n* = 5), generating a size bias of less than 1%. Each single Gly-X-Y repeating unit in *scl2* or *bcl* is encoded by 9 bp. Thus, the phospholipid nanogel will separate two distinct *scl2* or *bcl* alleles differing only by one Gly-X-Y repeat at the ~500-850-bp range and by 1-5 Gly-X-Y repeats at the ~1,000-1,400-bp range. The phospholipid nanogel performance is similar to the values reported for a commercial system that utilizes hydroxyethyl cellulose, which generated a

size bias ranging from 0-7% for 119 to 1270 bp DNA fragments (15). Clearly, the material can be used for DNA fragments approaching 1,500 bp.

This is a valuable assay to be used alone or in combination with other epidemiological markers to trace the emergence of new strains with a higher discriminatory power. Accurate size evaluation for any extended repeating sequence will also be an important tool assisting in long-range genomic assemblies produced from short (50-150 bp) DNA sequences obtained using next generation sequencers (53).

Advanced material properties enable in-capillary sample stacking

Continuous nanogels were used to generate the data in Figures 1 and 2 as well as Tables 1-4; however, the separations in Figure 1 confirmed that DNA had a faster mobility in the 2.5% phospholipid preparation than in the 10% preparation, which also has a substantially higher apparent viscosity. The different viscosity of the 10% vs the 2.5% phospholipid nanogels was integrated to create a discontinuous nanogel to stack DNA within a small plug of the highly viscous 10% phospholipid nanogel and then separate the DNA in the low viscosity 2.5% nanogel that yielded higher separation performance. These two nanogel preparations were introduced serially into the separation capillary. DNA was stacked using a 10 kV, 30 second injection in the discontinuous nanogel as shown in Figure 3A. At a temperature of 19 °C, the capillary was first filled with the 2.5% nanogel for 3 minutes at 140 kPa (20 psi). This was followed by a small plug of 10% nanogel introduced for 10 seconds at 69 kPa (10 psi), and finally a pre-plug of aqueous buffer was introduced for 15 seconds at 34 kPa (5 psi). The temperature was raised to 30 °C to increase the viscosity, which caused the 10% phospholipid to form a highly viscous nanogel. The DNA was electrokinetically-injected at 30 °C and stacked. The temperature was then dropped to 23 °C and the separation proceeded. The enhancement in peak area achieved with stacking is apparent by comparing the electropherograms shown in Figure 3A with 3B and 3C. The DNA in the electropherogram in Figure 3B, obtained with a 10 kV, 2 second injection in a discontinuous nanogel, was difficult to detect. The peaks were broad when the large 10 kV, 30 second injection was performed in the absence of the stacking gel as shown in Figure 3C. The plate count of the 600 bp peak in Figure 3C (40,000 theoretical plates) was 12-fold lower than that obtained with stacking in Figure

3A (500,000 theoretical plates). Integrating both materials offered a means to harness advantages of each medium for unique processes of injection and separation of large DNA fragments.

CONCLUSIONS

A practical alternative to slab gels was developed to separate and size DNA fragments as large as 1,500 bp, with a size accuracy ranging from 1-4%. The material was compatible with the higher concentration medium reported previously (21) and was suitable for thermally controlled gel-stacking and separations. By decreasing the concentration of the phospholipid preparation from 10% to 2.5%, the net cost of the dynamic sieving gel was reduced. The ease with which the dynamic sieving nanogel was introduced into capillary channels was ideal for microscale separations because the low flow resistance required only nominal pressures for loading. This holds potential to benefit microfluidic separations as well because when the separation device can be used repeatedly, the lifetime of the device is increased, which in turn decreases the overall cost of the analytical measurement.

ACKNOWLEDGMENTS

REFERENCES

1. Chakraborty, R., Stivers, D. N., Su, B., Zhong, Y., and Budowle, B. (1999) The utility of short tandem repeat loci beyond human identification: Implications for development of new DNA typing systems. *Electrophoresis* **20**, 1682-1696
2. Turenne, C. Y., Sanche, S. E., Hoban, D. J., Karlowsky, J. A., and Kabani, A. M. (1999) Rapid identification of fungi by using the ITS2 genetic region and an automated fluorescent capillary electrophoresis system. *J. Clin. Microbiol.* **37**, 1846-1851
3. Chen, Y.-C., Eisner, J. D., Kattar, M. M., Rassoulian-Barrett, S. L., Lafe, K., Bui, U., Limaye, A. P., and Cookson, B. T. (2001) Polymorphic internal transcribed spacer region 1 DNA sequences identify medically important yeasts. *J. Clin. Microbiol.* **39**, 4042-4051
4. Klepárník, K., and Boček, P. (2007) DNA diagnostics by capillary electrophoresis. *Chem. Rev. (Washington, DC, U. S.)* **107**, 5279-5317
5. Castro-Puyana, M., García-Cañas, V., Simó, C., and Cifuentes, A. (2012) Recent advances in the application of capillary electromigration methods for food analysis and Foodomics. *Electrophoresis* **33**, 147-167
6. Lagally, E. T., Scherer, J. R., Blazej, R. G., Toriello, N. M., Diep, B. A., Ramchandani, M., Sensabaugh, G. F., Riley, L. W., et al. (2004) Integrated portable genetic analysis microsystem for pathogen/infectious disease detection. *Anal. Chem.* **76**, 3162-3170
7. Swerdlow, H., and Gesteland, R. (1990) Capillary gel electrophoresis for rapid, high resolution DNA sequencing. *Nucleic Acids Res.* **18**, 1415-1419
8. Drossman, H., Luckey, J. A., Kostichka, A. J., D'Cunha, J., and Smith, L. M. (1990) High-speed separations of DNA sequencing reactions by capillary electrophoresis. *Anal. Chem.* **62**, 900-903
9. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., et al. (2001) The sequence of the human genome. *Science* **291**, 1304-1351
10. Grossman, P. D., and Soane, D. S. (1991) Experimental and theoretical studies of DNA separations by capillary electrophoresis in entangled polymer solutions. *Biopolymers* **31**, 1221-1228
11. Viovy, J.-L., and Duke, T. (1993) DNA electrophoresis in polymer solutions: Ogston sieving, reptation and constraint release. *Electrophoresis* **14**, 322-329
12. Andras, G. (1993) Capillary column containing a dynamically cross-linked composition and method of use. (Patent, U. S. ed.
13. Barron, A. E., Blanch, H. W., and Soane, D. S. (1994) A transient entanglement coupling mechanism for DNA separation by capillary electrophoresis in ultradilute polymer solutions. *Electrophoresis* **15**, 597-615
14. Woolley, A. T., and Mathies, R. A. (1994) Ultra-high-speed DNA fragment separations using microfabricated capillary array electrophoresis chips. *Proc. Natl. Acad. Sci. U S A* **91**, 11348-11352
15. Mueller, O., Hahnenberger, K., Dittmann, M., Yee, H., Dubrow, R., Nagle, R., and Ilsley, D. (2000) A microfluidic system for high-speed reproducible DNA sizing and quantitation. *Electrophoresis* **21**, 128-134

16. Zhou, H., Miller, A. W., Sosic, Z., Buchholz, B., Barron, A. E., Kotler, L., and Karger, B. L. (2000) DNA sequencing up to 1300 Bases in two hours by capillary electrophoresis with mixed replaceable linear polyacrylamide solutions. *Anal. Chem.* **72**, 1045-1052

17. Barron, A. E., Sunada, W. M., and Blanch, H. W. (1996) The effects of polymer properties on DNA separations by capillary electrophoresis in uncross-linked polymer solutions. *Electrophoresis* **17**, 744-757

18. Duke, T., and Viovy, J. L. (1994) Theory of DNA electrophoresis in physical gels and entangled polymer solutions. *Physical Review E* **49**, 2408-2416

19. Harroun, T. A., Koslowsky, M., Nieh, M.-P., de_Lannoy, C.-F., Raghunathan, V. A., and Katsaras, J. (2005) Comprehensive examination of mesophases formed by DMPC and DHPC mixtures. *Langmuir* **21**, 5356-5361

20. Nieh, M. P., Raghunathan, V. A., Glinka, C. J., Harroun, T. A., Pabst, G., and Katsaras, J. (2004) Magnetically alignable phase of phospholipid "bicelle" mixtures is a chiral nematic made up of wormlike micelles. *Langmuir* **20**, 7893-7897

21. Durney, B. C., Lounsbury, J. A., Poe, B. L., Landers, J. P., and Holland, L. (2013) A Thermo-responsive Phospholipid Pseudo-gel: Tunable DNA Sieving with Capillary Electrophoresis *Anal. Chem.* **85**, 6617-6625

22. Lounsbury, J. A., Bienvenue, J. M., and Landers, J. P. (2012) Sample-to-result STR genotyping systems: Potential and status. *Forensic Sci Rev* **24**, 123-142

23. http://avantilipids.com/index.php?option=com_content&view=article&id=214&Itemid=206&catnumber=850345. (accessed September 24, 2014) 1,2-ditetradecanoyl-*sn*-glycero-3-phosphocholine (DMPC) PC(14:0/14:0), CAS Number 18194-24-6, catalog number 850345P. Avanti Polar Lipids, Inc

24. http://avantilipids.com/index.php?option=com_content&view=article&id=206&Itemid=206&catnumber=850305. (accessed September 24, 2014) 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine 06:0 PC (DHPC), CAS number 34506-67-7, catalog number 850305P. Avanti Polar Lipids, Inc

25. http://www.lifetechnologies.com/search/global/searchAction.action?query=pop-4+polymer&resultPage=1&resultsPerPage=15&autocomplete=true. (accessed September 24, 2014) POP-4® Polymer for 3130/3130xl Genetic Analyzers, 7000 µl (Applied Biosystems®), catalog number 4352755. 2012 Life Technologies Corporation

26. Luo, R., Archer-Hartmann, S. A., and Holland, L. A. (2010) Transformable capillary electrophoresis for oligosaccharide separations using phospholipid additives. *Anal. Chem.* **82**, 1228-1233

27. Reda, K. B., Kapur, V., Goela, D., Lamphear, J. G., Musser, J. M., and Rich, R. R. (1996) Phylogenetic distribution of streptococcal superantigen SSA allelic variants provides evidence for horizontal transfer of ssa within *Streptococcus pyogenes. Infect. Immun.* **64**, 1161-1165

28. Lukomski, S., Nakashima, K., Abdi, I., Cipriano, V. J., Shelvin, B. J., Graviss, 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

29. Tuntevski, K., Durney, B. C., Snyder, A. K., LaSala, P. R., Nayak, A. P., Green, B. J., Beezhold, D. H., Rio, R. V. M., et al. (2013) *Aspergillus* collagen-like (*acl*) genes: identification, sequence polymorphism and assessment for PCR-based pathogen detection. *Appl. Environ. Microbiol.* **79**, 7882-7895

30. White, C. M., Luo, R., Archer-Hartmann, S. A., and Holland, L. A. (2007) Electrophoretic screening of ligands under suppressed EOF with an inert phospholipid coating. *Electrophoresis* **28**, 3049-3055

31. Archer-Hartmann, S. A., Sargent, L. M., Lowry, D. T., and Holland, L. A. (2011) Microscale exoglycosidase processing and lectin capture of glycans with phospholipid assisted capillary electrophoresis separations. *Anal. Chem.* **83**, 2740-2747

32. Zipper, H., Brunner, H., Bernhagen, J. r., and Vitzthum, F. (2004) Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. *Nucleic Acids Res.* **32**, e103

33. Dragan, A. I., Pavlovic, R., McGivney, J. B., Casas-Finet, J. R., Bishop, E. S., Strouse, R. J., Schenerman, M. A., and Geddes, C. D. (2012) SYBR Green I: fluorescence properties and interaction with DNA. *Journal of Fluorescence* **22**, 1189-1199

34. Flynn, A., Ducey, M., Yethiraj, A., and Morrow, M. R. (2012) Dynamic properties of bicellar lipid mixtures observed by rheometry and quadrupole echo decay. *Langmuir* **28**, 2782-2790

35. Wu, X., Langan, T. J., Durney, B. C., and Holland, L. A. (2012) Thermally responsive phospholipid preparations for fluid steering and separation in microfluidics. *Electrophoresis* **33**, 2674-2681

36. Pappas, T., and Holland, L. (2008) Fluid steering in a microfluidic chip by means of thermally responsive phospholipids. *Sens. Actuators, B, Chem.* **128**, 427-434

37. 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

38. Beall, B., Facklam, R., Hoenes, T., and Schwartz, B. (1997) Survey of emm gene sequences and T-antigen types from systemic *Streptococcus pyogenes* infection isolates collected in San Francisco, California; Atlanta, Georgia; and Connecticut in 1994 and 1995. *J. Clin. Microbiol.* **35**, 1231-1235

39. Courtney, H. S., Hasty, D. L., Li, Y., Chiang, H. C., Thacker, J. L., and Dale, J. B. (1999) Serum opacity factor is a major fibronectin-binding protein and a virulence determinant of M type 2 Streptococcus pyogenes. in *Mol. Microbiol.*, Wiley-Blackwell

40. Johnson, D. R., and Kaplan, E. L. (1993) A review of the correlation of T-agglutination patterns and M-protein typing and opacity factor production in the identification of group A streptococci. *J. Med. Microbiol.* **38**, 311-315

41. Rasmussen, M., and Björck, L. (2001) Unique regulation of SclB – a novel collagen-like surface protein of *Streptococcus pyogenes*. *Mol. Microbiol.* **40**, 1427-1438

42. Leski, T. A., Caswell, C. C., Pawlowski, M., Klinke, D. J., Bujnicki, J. M., Hart, S. J., and Lukomski, S. (2009) Identification and classification of *bcl* genes and

proteins of *bacillus cereus* group organisms and their application in *bacillus anthracis* detection and fingerprinting. *Appl. Environ. Microbiol.* **75**, 7163-7172

43. O'Loughlin, R. E., Roberson, A., Cieslak, P. R., Lynfield, R., Gershman, K., Craig, A., Albanese, B. A., Farley, M. M., et al. (2007) The epidemiology of invasive group A streptococcal infection and potential vaccine implications: United States, 2000-2004. *Clin. Infect. Dis.* **45**, 853-862

44. Steer, A. C., Law, I., Matatolu, L., Beall, B. W., and Carapetis, J. R. (2009) Global emm type distribution of group A streptococci: systematic review and implications for vaccine development. *The Lancet Infectious Diseases* **9**, 611-616

45. O'Brien, K. L., Beall, B., Barrett, N. L., Cieslak, P. R., Reingold, A., Farley, M. M., Danila, R., Zell, E. R., et al. (2002) Epidemiology of invasive group A *Streptococcus* disease in the United States, 1995-1999. *Clin. Infect. Dis.* **35**, 268-276

46. Sharkawy, A., Low, D. E., Saginur, R., Gregson, D., Schwartz, B., Jessamine, P., Green, K., McGeer, A., et al. (2002) Severe group A Streptococcal soft-tissue infections in Ontario: 1992-1996. *Clin. Infect. Dis.* **34**, 454-460

47. Musser, J. M., Hauser, A. R., Kim, M. H., Schlievert, P. M., Nelson, K., and Selander, R. K. (1991) Streptococcus pyogenes causing toxic-shock-like syndrome and other invasive diseases: clonal diversity and pyrogenic exotoxin expression. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 2668-2672

48. Dimopoulos, G., Frantzeskaki, F., Poulakou, G., and Armaganidis, A. (2012) Invasive aspergillosis in the intensive care unit. *Ann. N. Y. Acad. Sci.* **1272**, 31-39

49. Groll, A. H., Kurz, M., Schneider, W., Witt, V., Schmidt, H., Schneider, M., and Schwabe, D. (1999) Five-year-survey of invasive aspergillosis in a paediatric cancer centre. Epidemiology, management and long-term survival. *Mycoses* **42**, 431-442

50. Latge, J. P. (1999) *Aspergillus fumigatus* and aspergillosis. *Clin Microbiol Rev* **12**, 310-350

51. Zmeili, O. S., and Soubani, A. O. (2007) Pulmonary aspergillosis: a clinical update. *QJM* **100**, 317-334

52. Yu, J., Chang, P. K., Cary, J. W., Wright, M., Bhatnagar, D., Cleveland, T. E., Payne, G. A., and Linz, J. E. (1995) Comparative mapping of aflatoxin pathway gene clusters in *Aspergillus parasiticus* and *Aspergillus flavus*. *Appl. Environ. Microbiol.* **61**, 2365-2371

53. Salzberg, S. L., Phillippy, A. M., Zimin, A., Puiu, D., Magoc, T., Koren, S., Treangen, T. J., Schatz, M. C., et al. (2012) GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Res.* **22**, 557-567
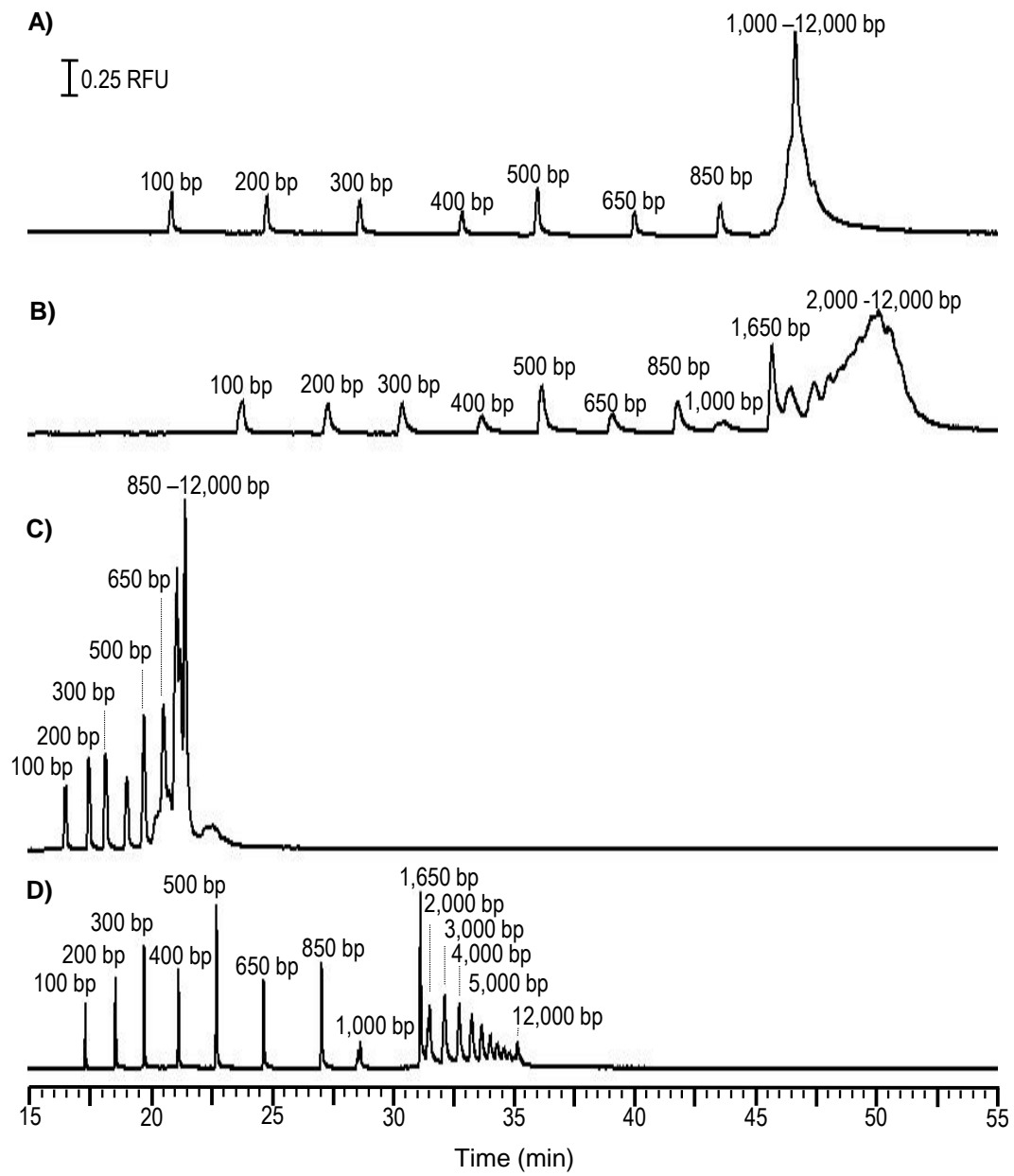
Figure 1

Separations obtained at different phospholipid concentration and run temperature.
Electropherograms of a 1 kilobase DNA ladder are detected using SYBR Green 1 fluorescent dye using a phospholipid nanogel with [DMPC]/[DHPC] = 2.5: (A) at a concentration of 10% and temperature of 30 °C, (B) at a concentration of 10% and temperature of 23 °C, (C) at a concentration of 2.5% and temperature of 30 °C, and (d) at a concentration of 2.5% and temperature of 23 °C. The DNA fragments in the ladder (Life Technologies, #10787-018) range from 100-12,000 bp and the size in bp specified by the manufacturer is indicated above each corresponding peak.
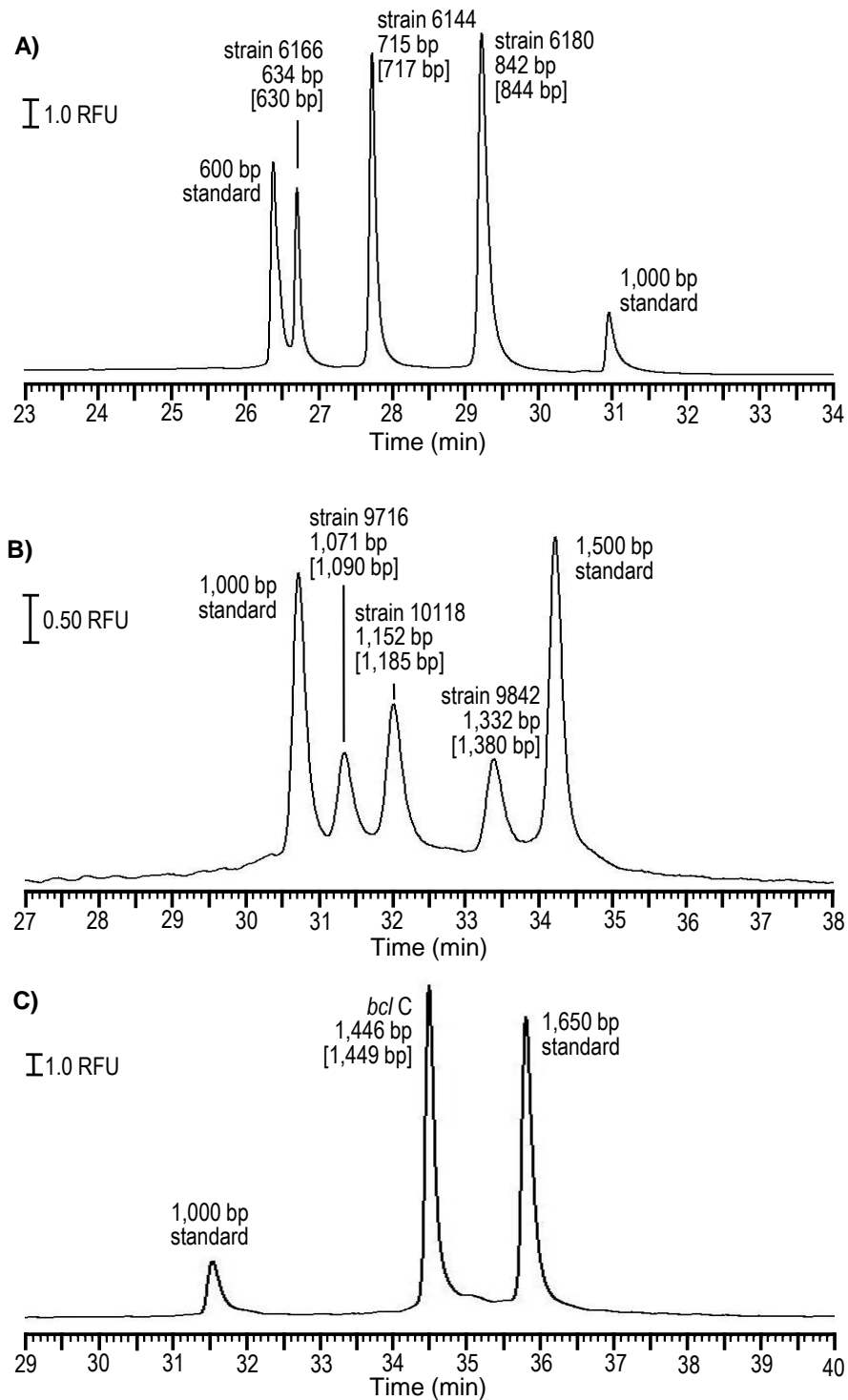
Figure 2

Separation of PCR amplicon obtained with the phospholipid nanogel optimized for DNA up to 1,500 bp in length. Electropherograms detected using SYBR Green 1 fluorescent dye and a phospholipid nanogel with [DMPC]/[DHPC] = 2.5 at a concentration

of 2.5% and temperature of 23 °C. The true sizes are specified above each peak while the measured sizes are denoted in the brackets. The peaks for each marker are labeled to indicate: (A) the three *scl2* strains designated 6166, 6140, 6180 (B) the three *scl2* strains designated 9716, 10118, 9842, and (C) *bcl* C. The separation is accomplished using a capillary that is 40 cm long, 25 μm id, with an effective length of 30.2 cm and E_{app} = 100 V/cm.
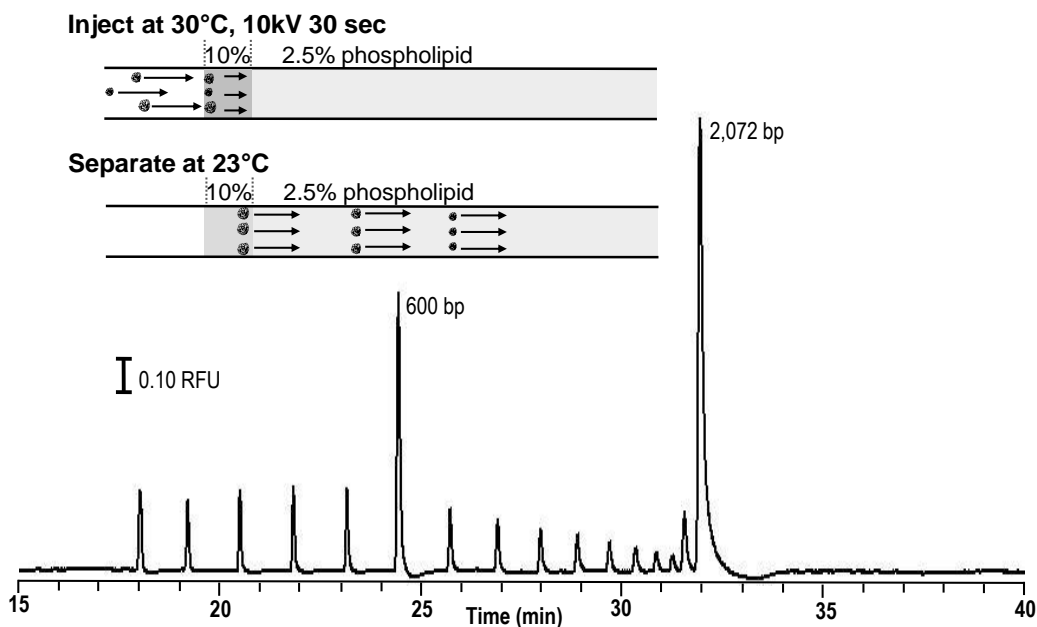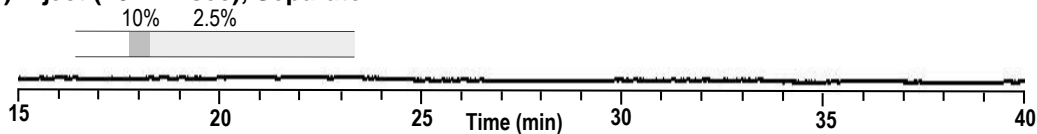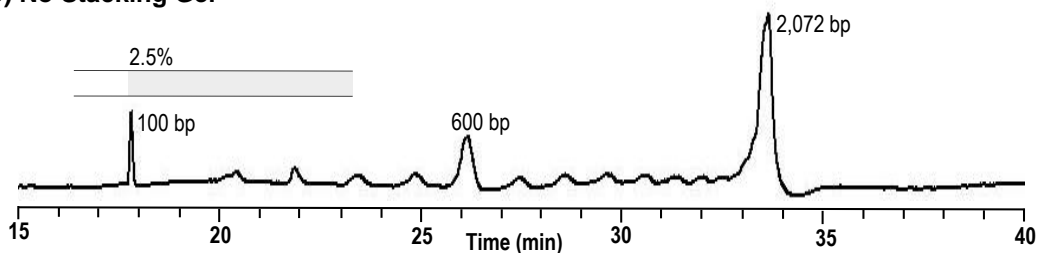
Figure 3

DNA stacking with phospholipid additives. Depicts the process of (A) discontinuous phospholipid nanogels in which the DNA is stacked in the 10% phospholipid and separated in the 2.5% phospholipid, (B) discontinuous phospholipid nanogels in which the 10 kV 2 sec injection is too short to stack DNA, and (C) a continuous nanogel using a 10 kV 30 sec injection without a stacking plug. The separation is accomplished using a capillary that is 40 cm long, 25 μm id, with an effective length of 30.2 cm and E_{app} = 100 V/cm.

184

Table 1. Effect of concentration on resolution[a]

Phospholipid (%)	Resolution 450/475 bp DNA		Resolution 900/950 bp DNA	
	Resolution (CV)[b]	in bp[c]	Resolution (CV)[b]	in bp[c]
2.5	5.0 (6)	5	6.4 (7)	8
5.0	4.6 (10)	5	4.3 (10)	12
7.5	4.5 (9)	6	3.4 (10)	15
10.0	4.4 (10)	6	2.8 (9)	18

[a]Separations (n = 5) of DNA base ladder MM-1000-FAM are achieved at 23 °C with [DMPC]/[DHPC] = 2.5 in a 40 cm long, 25 μm id capillary. The effective length is 30.2 cm and E_{app} = 100 V/cm.

[b]Resolution is calculated as 0.589($\Delta t/w_{1/2av}$), where Δt is the difference in peak migration times and $w_{1/2av}$ is the average peak width at half height

[c]Resolution in base pairs is calculated as the difference in base pairs/resolution.

Table 2. Effect of [DMPC]/[DHPC] ratio on resolution at 23°C[a]

[DMPC] [DHPC]	Resolution 450/475 bp DNA		Resolution 900/950 bp DNA	
	Resolution (CV)	In bp	Resolution (CV)	In bp
2.5	5.0 (6)	5	6.4 (7)	8
3.0	4.4 (10)	6	5.2 (10)	10
4.0	4.8 (10)	5	5.5 (10)	9
5.0	1.1 (20)	23	1.2 (20)	42

[a]Separation conditions, calculation of resolution, and calculation of resolution in bp are identical to that reported in Table 1.

Table 3. Performance of 2.5% phospholipid for *Aspergillus* biomarkers

Species	True size in bp[a]	Measured size in bp[b]	(CV)	Bias in bp[c]
A. nidulans	227	227	(0.3)	0
A. terreus	262	256	(0.4)	6
A. niger	297	301	(0.3)	4
A. fumigatus	338	333	(0.5)	5
A. flavus	402	402	(0.1)	0
A. fumigatus	489	495	(0.4)	6

[a]True size established as described in [26]. *A. fumigatus* derived amplicons of gene *aclF1*-5' and -3' were sequenced and represent true sizes of 338, 489 bp. Amplicons derived from *A. terreus* (gene *aclT2*-5'), *A. nidulans* (gene *aclNl*-3'), and *A. niger* (gene *aclNi*-5') were based on bioinformatics and were also partially sequenced with forward, but not with reverse, primers that were used to generate these amplicons. The amplicon from *A. flavus* (gene *aclFL*-3') was sequenced in full and the size of this amplicon is 402 bp.

[b]Separation conditions are identical to those reported in Table 1. Size standards of 150 and 450 bp are used.

[c]Bias is calculated as the difference between the true fragment size and the measured size.

Table 4. Performance of 2.5% phospholipid for *Streptococcus* biomarkers

Strain, type	True size in bp[a]	Measured size in bp[b]	(CV)	Bias in bp[c]
6166, M12	634	630	(0.2)	4
6144, M12	715	717	(0.4)	2
6180, M28	842	844	(0.5)	2
9716, M3	1071	1090	(0.4)	19
10118, M3	1152	1185	(0.3)	33
9842, M3	1332	1380	(0.2)	48

[a]True size established using DNA sequencing as reported in reference [27].
[b]Separations conditions are identical to those reported in Table 1. For strains 6166, 6144, and 6180 the size is determined using 600 and 1,000 bp standards. For strains 9716, 10118, and 9842, the size is determined using 1,000 and 1,500 bp standards
[c]Bias is calculated as the difference between the true fragment size and the measured size.

CHAPTER 7

A UNIQUE SET OF THE *BURKHOLDERIA* COLLAGEN-LIKE PROTEINS PROVIDES INSIGHT INTO PATHOGENESIS, GENOME EVOLUTION AND NICHE ADAPTATION, AND INFECTION DETECTION

Beth A. Bachert**,** Soo J. Choi, Anna K. Snyder, Rita V.M. Rio, Brandon C. Durney, Lisa A. Holland, Kei Amemiya, Susan L. Welkos, Joel A. Bozue, Christopher K. Cote, Rita Berisio, and Slawomir Lukomski

ABSTRACT

Burkholderia pseudomallei and *Burkholderia mallei*, classified as category B priority pathogens, are significant human and animal pathogens that are highly infectious and broad-spectrum antibiotic resistant. Currently, the pathogenicity mechanisms utilized by *Burkholderia* are not fully understood, and correct diagnosis of *B. pseudomallei* and *B. mallei* infection remains a challenge due to limited detection methods. Here, we provide a comprehensive analysis of a set of 13 novel *Burkholderia* collagen-like proteins (Bucl) that were identified among *B. pseudomallei* and *B. mallei* select agents. We infer that several Bucl proteins participate in pathogenesis based on their noncollagenous domains that are associated with the components of a type III secretion apparatus and membrane transport systems. Homology modeling of the outer membrane efflux domain of Bucl8 points to a role in multi-drug resistance. We determined that *bucl* genes are widespread in *B. pseudomallei* and *B. mallei*; Fischer's exact test and Cramer's V^2 values indicate that the majority of *bucl* genes are highly associated with these pathogenic species versus nonpathogenic *B. thailandensis*. We designed a *bucl*-based quantitative PCR assay which was able to detect *B. pseudomallei* infection in a mouse with a detection limit of 50 CFU. Finally, chromosomal mapping and phylogenetic analysis of *bucl* loci revealed considerable genomic plasticity and adaptation of *Burkholderia* spp. to host and environmental niches. In this study, we identified a large set of phylogenetically unrelated *bucl* genes commonly found in *Burkholderia* select agents, encoding predicted pathogenicity factors, detection targets, and vaccine candidates.

INTRODUCTION

Collagen structure is formed by three polypeptide chains of continuous repetitive Gly-Xaa-Yaa (GXY) sequence, each adopting left handed polyproline II type helices that combined form a right-handed superhelix (1). It is a universal structure that is broadly found among members of all three domains of life. It is the most abundant protein in mammals where it harbors important structural functions in the extracellular matrix and in support of cell adhesion, differentiation and growth (2,3). The prokaryotic collagen was identified and studied more recently, and has similar GXY sequence and triple helical structure (4-8). In mammalian collagens, proline (Pro) in the Y position is hydroxylated post-translationally and resulting Hyp (hydroxyproline) residues confer the maximum stability to the triple helix. As bacteria lack the prolyl hydroxylase required for these residues, bacterial collagens must be stabilized by other mechanisms, including increased proline content and electrostatic interactions between amino acid side chains (9-12). Several bacterial collagen-like proteins have been shown to form stable triple helices, including streptococcal collagen-like proteins 1 and 2 of *Streptococcus pyogenes* (4,13), rCL$_{Cp}$ from *Clostridium perfringens* (14), and BclA of *Bacillus anthracis* (15,16). Bacterial collagen-like proteins are found in species that are pathogenic to humans and animals (5-8,16-22). They are often surface-exposed and participate in important pathogenesis processes, including adherence and biofilm formation, host colonization and immune evasion (6,7,18,19,23-30). Several collagen-like genes have been evaluated as biomarkers for pathogen detection by targeting their conserved non-collagenous regions (31,32) and for strain fingerprinting by targeting highly polymorphic repetitive collagen-like sequences (32-35).

The *Burkholderia* species are ubiquitous in the environment but also include animal and plant pathogens. A group of 17 closely related species, designated *B. cepacia* complex organisms, cause pulmonary infections primarily in patients with cystic fibrosis (36). Two other species, *Burkholderia pseudomallei* and *Burkholderia mallei,* are significant human and animal pathogens in endemic regions and also represent biowarfare threats. These bacteria have been classified as category B priority pathogens, in part due to their high infectivity, an intrinsic broad-spectrum antibiotic resistance, and previous use as biological weapons during wartime (37). *B. pseudomallei* is a soil

saprophyte endemic to southeastern Asia and northern Australia, which causes melioidosis in humans. Melioidosis has a variety of clinical outcomes, from localized skin infection to pneumonia and acute septicemia, as well as chronic illness with abscess formation in major organs (38). As 50% of patients with septicemic melioidosis die within 48 hours, rapid diagnosis is crucial to patient survival (39). *B. pseudomallei* has a large genome of about 7.2 Mb, which undergoes frequent horizontal gene transfer as evidenced by multiple genomic islands that differ between strains (40). *B. mallei* is a closely related bacterium with a smaller genome, ~5.8 Mb (41). It is the causative agent of glanders in horses and other animals that can be transmitted to humans. It has been demonstrated by multi-locus sequence typing analysis that *B. mallei* is a clonal derivative of *B. pseudomallei* (42), which has undergone significant genomic reduction and rearrangement during host-adaptation (41). Consequently, *B. mallei* is unable to survive outside the host. *B. mallei* was one of the first microbes to be weaponized during World War I to infect livestock and humans (37). A third closely related organism, *B. thailandensis*, is considered non-pathogenic for humans (43). *B. thailandensis* is also a soil saprophyte with a large genome of ~6.7 Mb, which is endemic to geographical regions coinciding with *B. pseudomallei* (43,44); therefore, it is necessary to differentiate between the two species.

In this study, we identified and characterized an unexpectedly large set of 13 distinct *Burkholderia* collagen-like (*bucl*/Bucl) genes and proteins that are conserved in pathogenic *B. pseudomallei* and *B. mallei* species. We report the widespread presence of *bucl* genes in *B. pseudomallei* and *B. mallei* assessed by bioinformatics and analytical PCR, explore their phylogenetic relationships, infer important pathogenicity traits and antibiotic resistance mechanisms associated with Bucl proteins, and demonstrate the use of *bucl* genes as detection markers for these select agents in an animal model of infection.

RESULTS

Identification of *Burkholderia* collagen-like (*bucl*) genes

An increasing number of collagen-like proteins have recently been identified and studied in a variety of bacterial species, including Gram-positive pathogenic group A (5-8,17), B (SL, unpublished data), C (20,45) streptococci and pneumococci (18), bacilli and clostridia (16,21,32), as well as Gram-negative respiratory pathogen *Legionella pneumophila ((19);* SL, unpublished data). Here, we assessed the presence and distribution of the collagen-like proteins among *Burkholderia* species in the Pfam collagen family database (PF01391). We identified a total of 85 sequences among the members of the *Burkholderiaceae* family, with 77 of these sequences designated *Burkholderia* collagen-like (Bucl) proteins, among various species of the *Burkholderia* genus. We next focused on 59 protein sequences found in three closely related species of *Burkholderia*, *B. pseudomallei* (Bp), *B. mallei* (Bm), and *B. thailandensis* (Bt) that we initially categorized into 16 (Bucl1-16) protein types, based on domain organization and GXY-repeat types in their collagen-like (CL) regions; subsequent refinement eliminated three Bucl types, resulting in 13 Bucl proteins 1, 2, 3, 4, 5, 6, 7, 8, 10, 13, 14, 15, and 16. To assess their distribution, nucleotide sequences of these 13 *bucl* genes were used as independent queries to BLASTn-search the NCBI nonredundant database. Though we observed collagen-like sequences in other *Burkholderia* species, this set of 13 *bucl* genes and proteins were unique to Bp, Bm, and Bt species.

Identification of *bucl* genes in Bp K96243, proof of principle

A BLAST search of *bucl* alleles from various strains against the genome sequence of the reference strain Bp K96243 revealed that all 13 *bucl* genes were present and were distributed around both chromosomes (Fig. 1A). Six *bucl* genes were localized on chromosome one and seven *bucl* genes on chromosome two, and were found on both plus and minus strands (Fig. 1A, B). The presence of each *bucl* gene in Bp K96243 genome was confirmed by PCR with primers targeting the noncollagenous regions (Fig. 1C). Mapping of *bucl* genes in additional seven Bp and four Bm fully sequenced genomes revealed significant intra- and inter-species genomic rearrangements involving *bucl* loci (Fig. 2). For example, the region encoding *bucl* genes 6, 8, and 10 in Bp 668 was inverted

compared to Bp K96243 genome (Fig. 2A). Additionally, we observed both rearrangements (Fig. 2B, C) and deletions of *bucl* loci (Fig. 2C) in Bm genomes, compared to Bp, which is consistent with Bm-genomic plasticity as well as the evolution of Bm from Bp through genome reduction (41,46). To further characterize the genomic organization of these strains, organizational patterns (OP) of *bucl* biomarkers were assigned according to their position and orientation on each chromosome (Table 1). In aggregate, chromosomal rearrangements occur more frequently on chromosome one (six distinct organizational patterns were observed for both Bp and Bm strains analyzed) compared to chromosome two (three organizational patterns observed). While only one major organizational pattern on chromosome 1, Ch1 OPII, was found exclusively among Bp strains, major organizational pattern on chromosome 2, Ch2 OPI, was found in both Bp and Bm genomes. All observed rearrangements were intrachromosomal in both species, indicating no exchange of genetic material involving *bucl* markers occurred between the chromosomes. In summary, consistent with bioinformatic data, we here confirmed by PCR the presence of all 13 *bucl* genes in Bp K96243. We also captured significant genomic plasticity of the Bp and Bm species by employing *bucl* markers.

Characterization of Bucl proteins

Overall characteristics of Bucl proteins were examined in a set of geographically diverse *Burkholderia* strains sequenced, including 13 Bp, 11 Bm, and 9 Bt strains (Table 2, Table 3). All 13 Bucl proteins identified contained a collagen-like region (CL) flanked by noncollagenous N- and C-terminal regions. The noncollagenous regions were conserved among all three species within each Bucl with sporadic length variations observed (Table 2). As expected, the CL regions of the same Bucl varied significantly in length between strains due to differing numbers of GXY repeats. For example, Bucl3 varied from 38 repeats to 63 repeats in different strains of Bp, Bm, and Bt (Table 2). The triplet usage was unique to each Bucl across species and usually one or two GXY-repeat types dominated each CL region. For example, Bucl1 and Bucl8 contained exclusively GAN and GAS repeats, respectively, while Bucl3 contained predominantly GTS repeats and Bucl10 had predominantly GIH triplets.

In order to assess whether the Bucl proteins will form collagen-like triple helices, stability predictions were performed on representative Bucl-CL amino acid sequences. GXY repeat number in Bucl proteins varies from 2 in Bucl14 to 63 in Bucl3 (Table 2). Stability of the predicted collagenous regions of each Bucl was computed using an approach derived from host-guest peptide studies (47). Examination of the stability profiles shows highest stabilities for Bucl2, Bucl5, Bucl13, and Bucl15, with predicted melting temperatures ranging between 35-38°C, while all other Bucl proteins had melting temperatures between 20-35°C (Fig. 3). Transmembrane regions were predicted in CL domains of Bucl proteins 4, 6, 7, 8, 14, 15, and 16, whose stability ranks low (Fig. 3). Hydrophobic interactions occurring in a membrane environment likely stabilize these triple helices.

Structural Predictions

In addition to the CL region, four Bucl proteins were predicted to contain putative domains proven to participate in pathogenesis in other bacterial species (Table 2, Fig. 4A). Bucl3 contains a putative Talin-1 domain; Talin-1 is a cytoskeletal protein that binds and activates integrins in mammals and talin-1-integrin interaction links the cytoskeleton with the extracellular matrix, allowing cell adhesion and migration (48-50). Bucl4 contained a Bac_export_1 domain (Bacterial export proteins, family 1; PF01311) found in members of type III secretion protein family, including the SpaR of *Shigella* and *Salmonella*, and the YscT of *Yersinia*. These proteins form the inner-membrane part of the needle complex, which transports bacterial effector proteins to afflict host cells. Bucl8 contained the OEP domain; the members of outer membrane efflux protein family (PF02321) form channels that allow export of various compounds, including anti-microbial agents, in Gram-negative bacteria across the outer membrane (51). Bucl13 contained a SBP_bac_3 domain (Bacterial extracellular solute-binding proteins, family 3; PF00497), which is found in periplasmic proteins that bind specific solutes within the periplasmic space and are often associated with ABC-type transporters (52).

Signal sequences were predicted in Bucl proteins 3, 4, 8, and 15, additionally supporting extracellular location for Bucl4 and Bucl8 (Table 2). Most Bucl proteins had transmembrane regions, interestingly, often associated with the CL regions (Table 2).

Modeling of the OEP domains in Bucl8

The OEP domains found in Bucl8 are inferred in the formation of an efflux pump, thus, contributing to multi-drug resistance of Bp and Bm species (53). Two tandem OEP domains were predicted with high confidence (E-values $7x10^{-22}$ and $4.5x10^{-18}$). The Bucl8 is also predicted to be a lipoprotein with an amino-terminal lipid-binding cysteine residue and a transmembrane region predicted with TMpred (54).

HMM search in the PDB database using Bucl8-OEP region as a query identified closest similarity (E-value=$6.6x10^{-53}$) to the drug discharge outer-membrane lipoprotein OprM of *P. aeruginosa* (55,56). Using OprM structure as a template (pdb code 3d5k, sequence identity 27%), the model of Bucl8 was generated with MODELLER 9 v.9 (57).

The OEP domains of Bucl8 form a trimeric structure containing the characteristic α-barrel, which spans the periplasmic space, and the β-barrel, which spans the outer membrane (Fig. 4B). In OprM, the β-barrel is known to anchor the protein to the outer membrane, and also contains a series of surface exposed loops that are involved in constriction of the β-barrel pore, thereby preventing influx of xenobiotics at the resting state (56,58). The α-barrel contains an arrangement of twelve short helices and six long helices that form a bundle which is constricted at both ends but contains a bulge in the middle that can accommodate antibiotics. Twisting of the helices to loosen the pores forms a funnel-channel structure allowing for the active transport of antibiotics across the outer membrane outside of the bacterial cell (56).

The *bucl8* gene was found in all Bp and Bm strains tested by PCR and bioinformatics (Table 4), signifying the potential importance of Bucl8-efflux pump in the survival and pathogenesis of these species. Interestingly, all Bt strains analyzed contained DNA sequence homologous to the OEP-domain of Bucl8 in Bp and Bm but lacked the sequence corresponding to the Bucl8-collagenous domain; thus, it could not be recognized as a true Bucl. Additionally, a single nucleotide insertion at position 52, directly preceding the OEP-encoding region, was found, causing a frameshift mutation, which resulted in an altered amino acid downstream sequence.

Phylogenetic analyses of *bucl* genes

To better understand the relationship of *bucl* genes among *Burkholderia* spp., parsimony and model-based phylogenetic analyses were performed. All 13 *bucl* sequences, originally identified in collagen Pfam database, were BLASTn-searched against completed genomes of Bp, Bm, and Bt, and each *bucl* sequence was downloaded. The 13 *bucl* genes demonstrate no sequence similarity, indicating these are non-homologous genes, whereas alleles encoding the same *bucl* gene were orthologous across species. Nucleotide sequence alignments were generated for each *bucl* gene present in 13 Bp and 11 Bm strains; analysis of *bucl3* and *bucl4* also included 9 Bt strains (S1 data set). Pairwise alignments of each *bucl* among the different strains revealed that percent identities ranged from 42%-100%, with the average percent identity for each *bucl* ranging from 76.5-94.9% (S2 data set). In general, the non-collagenous regions of *bucl* genes were conserved, while the CL regions showed significant length polymorphisms. Consequently, the *bucl1* phylogeny based on non-CL region sequence produced a star pattern, while the *bucl1* phylogeny generated based on the entire *bucl1* sequence showed more extensive branching patterns, most of which were supported by Bayesian Posterior Probability values and several of which were also supported by maximum parsimony bootstrap values (S1 Fig). The CL region of *bucl1* encodes a single GAN-repeat type, therefore, the only difference between *bucl1* alleles from different strains represented in this tree arises from different GAN-repeat numbers. Since this is a common feature of all *bucl*s, and incorporation of these regions would likely lead to long branch attraction, only the non-CL regions were used in further analyses. Multiple sequence alignments of *bucl* genes 2, 5, 6, 7, 10, 13, 14, 15 and 16 showed highly conserved nucleotide sequence, similar to *bucl1*, therefore phylogenetic analysis was not performed.

Phylogenetic trees were generated for individual and concatenated *bucl3*, *bucl4*, and *bucl8*, as these genes were present in all three species and contained the most informative characters. We included the OEP-encoding sequence of *bucl8* from Bt strains in this analysis, despite the lack of CL-encoding sequence and conserved frameshift mutation, because of significant sequence similarity to *bucl8*-OEP sequences shared with Bp and Bm. The phylogeny generated from concatenated sequences showed similar associations as phylogenies for the individual genes, although usually with higher

statistical support. All analyses showed Bp and Bm strains were more closely related to each other than to Bt strains, which formed a main separate branch (Fig. 5, 6, S2). This observation is consistent with the hypothesis that the pathogenic Bp and Bm strains diversified from Bt (42,59-61). On the concatenated tree, Bm strains formed a single clade without further resolution that was strongly supported by both Bayesian posterior probability (PP, 100) and maximum parsimony (MP, 100) bootstrap values (Fig. 5). This observation indicates either inadequate time for the diversification of Bm strains or purifying selection for the retention of nucleotide identity due to importance in adapting to its host pathogen niche (41). In contrast, Bp strains exhibited higher diversification as shown by the presence of multiple clades. Four supported clusters were observed, two of which, Cluster 1 and Cluster 4, showed geographical associations as these strains were all Australian isolates. Cluster 1 (PP 98, MP 100) contained Bp strains 20B16, MSHR146, MSHR511, and NCTC 13178, all isolates obtained from Australia. Cluster 2 (PP 58, MP 100) contained Bp strains NCTC13179 and 1026b, isolated from human infections in Australia and Thailand, respectively. Cluster 3 (PP 100, MP 100) contained Bp strains 1106a and BPC006, obtained from northeast Thailand and China, respectively. Finally, Cluster 4 (PP 100, MP 100) contained Bp strains MSHR305 and MSHR520, which are both human infection isolates from Australia. Clusters 1 and 4 were also supported by trees based on individual *bucl3, bucl4,* and *bucl8* genes, although strain NCTC 13178 as part of Cluster 1 was only supported by the tree based on *bucl4* (Fig. 6, S2). Similar to Bp, Bt strains showed significant diversification as evidenced by the formation of three supported clusters in the concatenated tree. These clusters were numbered consecutively Cluster 5, Cluster 6, and Cluster 7 (Fig. 5). Clusters 5 and 7 were supported by individual *bucl3* and *bucl4* phylogenies (Fig. 6), while only Cluster 5 was supported by *bucl8* phylogeny (S2 Fig). Analysis performed using amino acid sequences of Bucl proteins generated phylogenetic trees with similar patterns, though the support values were lower (S3 Fig), indicating many of the nucleotide changes were synonymous.

Overall, most *bucl* genes were highly conserved among Bp and Bm with most of the variation occurring in the CL region due to differing numbers of GXY repeats. Variation in non-CL regions of *bucl3*, *bucl4*, and *bucl8* revealed divergence between Bt and select agents Bp and Bm, as well as diversification among Bt strains. Bp and Bm appear more

closely related, but only Bp strains showed diversification across the *bucl* loci by the formation of multiple distinct clades with strong statistical support.

Assessment of *bucl* distribution across *Burkholderia* spp.

In order to assess the distribution of *bucl* genes across *Burkholderia*, nucleotide BLAST searches were performed using *bucl*-gene sequences from the reference strain Bp K96243, as queries against completed genomes of 13 Bp, 11 Bm, and 9 Bt strains. All 13 *bucl* genes were present in all Bp genomes, while the majority of *bucl* genes were maintained within Bm genomes (Table 4). Up to three *bucl* genes were missing in 8 Bm genomes, which is consistent with the reduced genetic material in this species (41,46). In contrast, only complete open reading frames of *bucl3* and *bucl4* were present in Bt genomes, presumably encoding a lipoprotein with a putative Talin-1 domain and a type III secretion inner membrane protein (Table 2, Fig. 4A), respectively.

In addition to bioinformatic data, we tested distribution of *bucl* genes by standard PCR in a collection of genomic DNA from 25 Bp and 20 Bm strains, as well as the DNA from non-select agent controls 4 Bt, 3 *B. cepacia* (Bc), 5 *B. cenocepacia* (Bce), and 6 *B. multivorans* (Bmv) strains (Table 5, Table 6). Consistent with bioinformatic data, virtually all 25 Bp strains were found to contain all 13 *bucl* genes, with the exception of strain China 3 (BpCh3) which was missing *bucl1* and *bucl4* (Table 4, Fig. 7, S4 Fig). Almost all Bm strains tested (15 out of 20) were lacking up to three *bucl* genes, in agreement with bioinformatic results. We calculated *bucl* frequencies as the proportion of Bp and Bm strains positive for each *bucl,* as tested by both PCR and bioinformatics. High frequencies were observed for *bucl3*, *bucl4*, *bucl7*, and *bucl15* (0.90-0.98), while lower frequencies were observed for *bucl2* (0.85) and *bucl10* (0.82). The *bucl2* and *bucl10* genes were most frequently absent from Bm strains, missing in about one-third of strains analyzed, indicating these genes are nonessential for Bm survival in mammalian host. Finally, all Bt strains contained only *bucl3* and *bucl4*, while no amplification of these two *bucl* genes was obtained for other control *Burkholderia* spp. (Table 4, Fig. 7, S4 Fig).

We next evaluated the association of *bucl* presence with pathogenicity among Bp, Bm, and Bt strains. The Fisher Exact Probability Test and Cramer's V analysis were performed on the number of *bucl* genes present and absent among two groups: 1)

pathogenic Bp and Bm strains and 2) nonpathogenic Bt strains. The Fisher test provides a measure of the statistical significance between two groups, and Cramer's V squared (V^2) is a value, which measures the degree of association between two variables on a scale of zero (no association) to one (perfect association). The Fisher test showed significant differences between group 1 and 2 for all *bucl* genes, except for *bucl3* and *bucl4*, indicating the presence of collagen-like genes is significantly associated with pathogenic *B. pseudomallei* and *B. mallei* species as compared with non-pathogenic *B. thailandensis* (Table 4). Further calculation of Cramer's V^2 showed perfect association (V^2=1) for *bucl* genes *5, 6, 8, 13, 14,* and *16* that were present in all Bp and Bm strains, while absent in all Bt strains. High V^2 values were calculated for *bucl1* (V^2=0.829), *bucl7* (V^2=0.829), and *bucl15* (V^2=0.908), indicating positive association with these *bucl* genes with pathogenic Bp and Bm, as compared with nonpathogenic Bt lacking them. The remaining *bucl* genes, 2, 3, 4, and 10, had little or no association with pathogenic Bp and Bm compared to Bt (V^2<0.5). Hence, our statistical analyses strongly infer association between the presence of the majority of Bucl proteins and pathogenicity.

Detection of *Burkholderia* select agents by analytical PCR

Four conserved amplicons generated from *bucl* genes that were uniformly found in all Bp and Bm strains, but were absent in Bt, Bc, Bce, and Bmv strains, were assessed for select agent detection by standard agarose gel electrophoresis and capillary gel electrophoresis: *bucl5* (216 bp), *bucl13* (212 bp), *bucl14* (178 bp), and *bucl16* (123 bp) (Fig. 7C). Size-identification of *bucl*-based amplicons by capillary gel electrophoresis was performed in a 10% phospholipid nanogel, allowing near single base pair resolution (62), including *bucl5* and *bucl13* amplicons that differ by 4 bp. Sizing of the target DNA fragments was accomplished by linear regression analysis for DNA size (in bp) versus migration time. The *bucl* gene amplicon sizes were calculated using the linear fit obtained for the migration times of internal standards with lengths of 100 bp and 250 bp, and the standard deviation calculated from 5 replicate measurements. The bias is calculated as the difference between the true fragment size and the measured size. Sizing results are reported as follows for n=5 separations: [gene name (true size): calculated size ± standard deviation, percent relative size bias defined as bias divided by the true size];

bucl5 (216 bp): 218 ± 2 bp, 0.9%; *bucl13* (212 bp): 215 ± 1 bp, 1%; *bucl14* (178 bp): 181 ± 1 bp, 2%; *bucl16* (123 bp): 120 ± 1 bp, 2%.

Detection of *Burkholderia* select agents by quantitative PCR

Identification of molecular targets for *Burkholderia* select agents is challenging due to the high genomic plasticity reported in these organisms that include significant genomic rearrangements and deletions. PCR assays developed for *Burkholderia* detection include BurkDiff, a dual-probe assay able to detect and differentiate Bp and Bm (63,64), and the TTS1 assay targeting *orf2* of type three secretion system I, detecting Bp only (64-66). Here, we developed a qPCR assay for the detection of Bp and Bm based on *bucl16* target. A locked nucleic acid hydrolysis probe specific for *bucl16* gave robust amplification using DNA of Bp K96243 (Cq= 21.85±1.37). This probe was then tested against the genomic DNA collection, providing amplification of all Bp and all Bm strains, with no amplification from non-select agent controls including Bt, Bce, and Bmv, as well as a no DNA template control (Fig. 8A). 30 ng of DNA was used for each strain and Cq values ranged from 23.42- 29.05.

We next tested the *bucl16*-based qPCR assay towards detection of an infection with *Burkholderia* select agents by employing samples spiked with human plasma, and with samples obtained from experimental animals. PCR reactions performed with 30 ng Bp K96243 DNA and spiked with 5% human plasma produced positive amplification with average Cq=24.28±2.14 (Fig. 8B), whereas reactions spiked with 10% and 20% human plasma produced averaged Cq= 25.89±1.76 and Cq=27.96±1.82, respectively.

Next, Bp strain HBPUB10134a was used for the detection of *Burkholderia* infection *in vivo*. Our recent studies have shown that Bp HBPUB10134a was the most virulent in the intraperitoneal infection model among a panel of 11 Bp strains, with an LD_{50} of 10 CFU at day 21 post-infection (67). Following the injection, mice presented common clinical manifestations, including abscess and pyrogranuloma formation in the spleen and liver, and in some cases lesions and inflammation in the eyes and tail. A common pathological observation was the loss of rear limb function occurring between 6 and 30 days post-infection, associated with the pyrogranulomatous inflammation in the skin, skeletal muscle, bone, and peripheral nerves in the hind limbs. Here, mice that were

injected intraperitoneally, were euthanized and sampled after 3, 7, or 14 days postinfection. Homogenized spleen samples were plated on blood agar to assess bacterial loads and 1 μL samples of irradiated sterile spleen extracts were used directly in qPCR reactions. Four samples, with original bacterial loads of 5×10^7, 5×10^6, 2×10^5, and 10^3 CFU/ mL, thus, presumably corresponding to 5×10^4, 5×10^3, 2×10^2, and 10 CFU per 1 μL added to each qPCR reaction, respectively, were tested using our *bucl16*-based assay. When crude spleen extracts were used in qPCR, positive detection was obtained for 5×10^3 CFU and 5×10^4 CFU samples with averaged Cq values of 29.49±1.67 and 26.39±1.71, respectively (Fig. 8B). Importantly, we observed that 1:10 dilution of the sample containing 5×10^4 CFU/ μL, resulted in improved amplification, as evidenced by lower Cq value (23.32±0.42), while 1:100 dilution resulted in similar amplification as undiluted crude sample (Cq= 27.23±1.10) (Fig. 8B, red curves). Further 1:1000 dilution of spleen extract provided detection level as low as 50 CFU per reaction with a Cq value of 32.63 ±1.57. On the other hand, 1:10 dilution of the sample originally containing 5×10^3 CFU/ μL resulted in poorer amplification (Cq= 32.66±2.46) than crude undiluted sample (gray curves). We think that crude spleen extracts contained varying levels of inhibitors that differentially affected amplifications in these two samples. Finally, in addition to *bucl16*, *bucl* genes *5, 6, 8, 13,* and *14* that were found in all Bp and Bm strains are similarly good candidate markers for the development of diagnostic qPCR assays.

DISCUSSION

Traditionally, collagen has been associated with multicellular animals, although, the number of collagen-like proteins identified in bacterial genomes has recently increased with 2554 sequences currently (search on 04/12/15) deposited in the Pfam collagen data base. The distribution of these collagen-like proteins is not uniform, however; they are absent in some bacteria and are overrepresented in other species. Here, we identified and characterized a group of 13 discrete collagen-like proteins in *Burkholderia,* referred to as Bucl, which are largely found in the pathogenic Bp and Bm species. Furthermore, we found that *bucl* genes provided important clues on the genomic plasticity and evolution of *Burkholderia* select agents. We observed Bucl proteins contained domains that are known to be involved in pathogenesis and antibiotic resistance, including an outer

membrane efflux protein which we modelled. Finally, we utilized *bucl* genes as detection targets and successfully detected Bp infection in a mouse model.

Characterization of Bucl-CL Region

Collagen-like sequences, embedding the typical repetition of triplets of the type Gly-X-Y (2,68-70) have been identified in all Bucl sequences. We observed that for each Bucl, one or two GXY types predominated the CL region. This limited variation in GXY content resembles that seen in Bcl proteins of *Bacillus anthracis* (32) but is in contrast to Scl proteins of *Streptococcus pyogenes*, whose GXY sequence varies significantly within the CL region (13). Typical of prokaryotic collagens, these sequences do not contain triple-helix-stabilizing hydroxyprolines, since bacteria lack the prolyl-hydroxylase enzyme necessary for post-translational modification of Pro to Hyp. The highest triple helix stabilities were predicted for Bucl2, Bucl5, Bucl13 and Bucl15, within the range of 35-38°, which is similar to that of previously studied bacterial collagens as well as human collagen (11,13,71,72). Similar to the CL regions of other prokaryotic proteins, like Scls from *S. pyogenes* (7,8), the CL regions of these proteins share the common characteristics of possessing charged residues GEX, GLE and GXR triplets, respectively (Fig. 3, Table 2). Indeed, ion pairs play a major role in stabilizing the triple helix, with an enthalpic stabilization, which likely involves interactions of polar groups with an ordered hydration network (9,10,12). Additionally, specific GXY triplets were found to have favorable enthalpy values, corresponding to increased hydrogen bonding potential, including GPE (71), which is a common GXY triplet in the Bucl5 CL region. These regions are likely to be of biological importance in establishing interactions with charged counterparts. Interestingly, bacterial collagens have been shown to have relatively high proline content, 20% in *S. pyogenes* and up to 40% in *B. anthracis* (11), especially in the X position (73), whereas Bucl proteins lack Pro residues; only Bucl5 contains GPE repeats, likely contributing to its predicted high stability. Other Bucl proteins with lower thermal stability may rely on the hydrophobic membrane environment for triple helix stabilization, as those were predicted to have transmembrane regions, especially within the CL regions. Stability predictions shown here were computed using long CL sequences, whereas some Bucl variants had short CL regions, which may not form triple helices. This is substantiated by

the fact that few triplets may also exist in other folds e.g., G5 domain, whose structure presents a pseudo-triple helix (74). In summary, while overall characteristics of the Bucl proteins we identified were similar to previously described bacterial collagen-like proteins, i.e., presence of collagenous and non-collagenous domains and length variation in collagen region, the GXY content observed in Bucls was unique and likely impacts the structural stability of the Bucl-CL triple helix.

Characterization of Bucl non-collagenous domains and their inferred roles in *Burkholderia* pathogenesis

It has been observed that collagen-like proteins are often surface associated. Indeed, among 53 bacterial and viral collagen-like proteins analyzed in an initial genome-based study, 16 were annotated as cell-wall attached or membrane associated (73). Additionally, surface expression of collagen-like proteins including Scls of *S. pyogenes*, PclA of *S. pneuomonia*, and Lcl of *L. pneumophila*, has been demonstrated experimentally (7,8,18,19). Structural predictions performed for Bucl proteins revealed that their majority, 10 out of 13, have transmembrane regions, supporting the location of Bucl proteins in the inner or outer membrane of *Burkholderia* spp. Moreover, four of these proteins were predicted to contain both signal sequences and transmembrane domains, further supporting surface association. Further non-collagenous features include well-conserved domains (in Bucl3, Bucl4, Bucl8 and Bucl13), which are inferred in pathogenesis.

Bucl3 was predicted to have a Talin-1 domain. Talin-1 in eukaryotes is known to bind and activate integrins as well as link the cell cytoskeleton to the extracellular matrix (50). Cell-to-cell invasion by *Burkholderia* is largely achieved by the disruption of the host cytoskeletal network, as well as the fusion of host cells resulting in the formation of multinucleated giant cells, mediated mainly by type III and type VI secretion systems (75,76). The intra- and inter-cellular spread is facilitated by the formation of actin tails which propel the bacterial cells. Thus far, several type III secretion effector proteins are known to be involved with host actin polymerization allowing cell invasion, including BimA, BopE, and BipD (43). The putative Talin-1 domain found in Bucl3 may also be involved

in interactions with host actin that allow for cell invasion or the formation of actin tails during infection.

The Bucl4 protein is putative inner membrane protein part of the type III secretion T3SS-2 system (77). There are three known type III secretion gene clusters (T3SS-1, T3SS-2, and T3SS-3) distributed among Bp, Bm, and Bt species. T3SS-1 is specific to Bp while T3SS-2 and 3 are found ubiquitously in all three species (78). The T3SS-3 is known to be important for virulence in Bp (44,75,79), as mutants deficient in the T3SS-3 have reduced replication in host cells, and are unable to escape endocytic vacuoles, and to form membrane protrusions and actin tails (80). The other two secretion systems are less well characterized, and the role of the T3SS-2 secretion system in pathogenesis is not known. The unique association of a collagenous domain with Bac_export_1 domain in this inner membrane protein of T3SS-2 has not been previously acknowledged.

Bucl13 contains the SBP_bac_3 domain, and is predicted to be a bacterial periplasmic solute binding protein. Binding of the solute causes a conformational change, which allows interaction of the solute with inner membrane proteins and subsequent transport of the solute into the cell. Family 3 solute-binding proteins are known to bind polar amino acids and opines (52), therefore Bucl13 is likely associated with amino acid transport; interestingly, Bucl13 is present in all Bp and Bm strains tested, while it is absent in non-pathogenic Bt. Bucl13 was also predicted to have a collagenous domain with a relatively high thermal stability, possibly contributing to its function.

Of particular interest is the Bucl8 protein, which was found to contain two tandem outer membrane efflux protein (OEP) domains that are known to contribute to the multidrug resistant phenotype of Bp and Bm species. These organisms are intrinsically resistant to multiple antibiotics including aminoglycosides, macrolides, and β-lactams (53,81). The outer membrane protein is an integral component of a tripartite Resistance-Nodulation-Division (RND) efflux pump that also requires an accessory protein in the periplasm and an inner membrane transport protein. It is known that there are 10 RND efflux pumps annotated in the Bp K96243 strain, many of which have not been explored (40). Currently, only three of these systems, BpeAB-OprB, AmrAB-OprA, and BpeEF-OprC, have been investigated for their roles in multidrug resistance (82-84). Interestingly, Bucl8 was found to be present in all Bp and Bm strains and absent in the non-pathogenic

Bt, suggesting selective pressure for the Bucl8-OEP in human or animal infection. We homology-modeled the Bucl8-OEP region based on OprM protein of *Pseudomonas aeruginosa* and observed a trimeric arrangement forming an outer membrane-spanning β-barrel and periplasmic α-barrel. The presence of the CL domain in Bucl8 is an unexpected observation, as collagenous regions have not been reported as part of efflux pump systems. On the other hand, the trimeric arrangement of Bucl8 is consistent with the formation of a collagen triple helix. TMPred predicted a transmembrane region, albeit with lower score, for a part of the collagen-like region (amino acids 539-557) indicating the CL region folds back across the membrane. The CL region is then predicted to extend into the extracellular space, projecting the carboxyl-terminal region from the cell surface. The triple-helical CL region may have a number of functions: i) to project the C-terminal region, which may serve as a surface adhesin, ii) to stabilize the trimeric arrangement of the OEP, and iii) may assist in blocking the β-barrel pore at the resting state, thus, preventing entry of xenobiotics into the cell. Ongoing studies will determine the potential role of Bucl8-OEP in drug resistance, Bp and Bm pathogenesis, as well as a potential as vaccine candidate.

Bucl phylogeny

The presence of 13 collagen-like genes in Bp and Bm genomes poses the question how have these unique sequences been acquired in *Burkholderia*? The GXY repeats found in bacterial collagens may have arisen through mechanisms including *de novo* spontaneous mutation and subsequent triplet repeat expansion independent within each gene, or by horizontal gene transfer. It has been initially suggested that collagen sequences are acquired by horizontal transfer from eukaryotes to prokaryotes based on the lack of collagen sequences in ancestral archaeal genomes and relatively few sequences identified in bacterial genomes (73). However, current collagen Pfam contains 2,554 bacterial collagen sequences, as well as 14 archaeal. A recent study, focused on bacterial molecular mimics of host proteins, proposed that collagen-like sequences found in pathogens evolved independently to mimic human host proteins (85). The uniformity of GXY content within each Bucl indicated they are likely to have evolved from the accumulation of repeats within each gene, resulting in diverse Bucl proteins that share

the GXY motif but with different GXY composition. Additionally, gene-enrichment analysis showed that collagen-like proteins were related to extracellular matrix mimicry and cell adhesion, supporting the evolution of repetitive sequences in virulence factors. Our phylogenetic analyses show that 13 collagen-like genes observed in numerous Bp and Bm genomes are unrelated to each other, which supports their independent acquisition, as well as selective adaptation of their collagen-like sequences in the host environment. This is further supported by the lack of collagen-like proteins (11 out of 13) in the closely related environmental species of Bt, indicating these sequences were acquired after divergence of Bp and Bm from Bt. Since Bucl proteins are unrelated and encoded in various locations in the genome, within-gene expansion of GXY-repeat motifs may point to convergent evolution of collagenous sequences to fulfill a similar function.

Phylogenetic trees based on three *bucl* loci showed Bt strains formed a distinct separate branch from Bp and Bm strains. This is consistent with previous studies based on phylogenetic analyses of seven MLST loci (42) and over 11,000 SNPs (60) which showed Bp and Bt isolates were resolved into two groups that were supported in 100% of bootstrap replicates. We also observed Bm strains share high sequence similarity, while Bp strains exhibited more intraspecies diversity, forming more extensive clusters that often corresponded to geographical associations. Previous phylogeographic reconstruction of *Burkholderia* strains based on over 14,000 SNPs showed that Bp and Bm strains formed separate clusters. The same study also showed Bp strains were significantly divided between those originating from Australia and Asia (60), in agreement with our observation that Australian isolates formed distinct clusters in *bucl*-based phylogenetic trees.

Bucl distribution

Both bioinformatic and PCR analyses showed that the majority of *bucl* genes are unique to Bp and Bm strains, with the exception of *bucl3* and *bucl4.* This observation may indicate these two genes are selected for in the environment of Bp and Bt, as several genomes of host-adapted Bm, lack these genes. The absence of most *bucl*s from Bt is a surprising observation since its genome is overall similar to that of Bp (59)*,* which may suggest either the acquisition of *bucl*s in Bp or the loss of *bucl*s in Bt after divergence of the two species.

Both Bp and Bt species have large genomes of approximately 7.2 and 6.7 Mb, respectively, divided between two chromosomes. Comparative genomics showed that Bp and Bt genomes share a large number of conserved genes involved in both core and accessory functions, while genes associated with virulence in Bp have increased diversity (59). Interestingly, *bucl3* and *bucl4*, encoding proteins potentially involved in pathogenesis, are found in the avirulent Bt. It has been shown that 71% of virulence-related genes in Bp are conserved in Bt with similarities of over 80%, including type III secretion gene clusters (59). Amino acid differences in virulence proteins present in both species may confer functional differences impacting virulence in Bp vs. Bt. Lastly, the presence of *bucl3* and *bucl4* alone in Bt was not sufficient to cause pathogenesis, a new biological trait acquired by Bp and Bm after the acquisition of additional virulence factors, including additional Bucls. A prominent feature of *Burkholderia* genomes is the presence of multiple horizontally acquired genomic islands that differ between Bp and Bt (40,59). These genomic islands are associated with survival in the soil environment and are absent in Bm genomes, possibly explaining why Bm cannot persist in the environment (46). The presence of most *bucl* biomarkers in both Bp and Bm genomes indicates they are not located within genomic islands but are rather a part of the core genome.

Previously, it has been reported that Bm is a clonal derivative of Bp, which has evolved to adapt to the host environment. Multilocus sequence typing analyses show that, in contrast to Bp, Bm strains are genetically homogenous, while relatively few new genes are being identified, as additional genomes are sequenced (46). However, the variable portion of the genome, though not acquiring new genetic information, is continuing to alter via expansion of IS elements and chromosomal rearrangements. Our phylogenetic analysis of *bucl* genes within Bm supports the genetic homogeneity among Bm strains, and mapping of *bucl* markers showed considerable chromosomal rearrangements occurring between Bm strains.

Different collagen-like proteins, unrelated to 13 Bucls characterized here, were also present in other *Burkholderia* species. We noticed these collagen-like proteins found in *B. cepacia*, *B. cenocepacia*, *B. multivorans*, *B. ambifaria*, *B. glumae*, *B. gladioli*, and *B. xenovorans*, contained GTS repeats within the CL region, similar to Bucl3. However, outside of the CL region, sequence identity was very low, therefore these proteins were

207

not included in the Bucl3 group. Given the importance of *Burkholderia* species as human pathogens and part of the *Burkholderia cepacia* complex (*B. cepacia*, *B. cenocepacia*, *B. multivorans*, and *B. ambifaria*), plant pathogens (*B. glumae* and *B. gladioli*), and plant symbionts (*B. xenovorans*), investigation of these collagen-like proteins is an interesting area for further study.

bucl-based infection detection

Bp and Bm are reported to have fatality rates up to 80% and 95%, respectively (86,87), making early diagnosis and treatment critical for patient survival. Currently, culture-based identification of *Burkholderia* select agents remains the gold standard for diagnosis (86,88). Highly variable genomes present a challenge in finding reliable genetic targets that are not subjected to chromosomal deletion, especially for *B. mallei*. Although a few laboratory-developed qPCR tests have been reported, there are no FDA-approved assays for the detection of *Burkholderia* select agents. The TTS1 assay (65,66) detects specifically Bp, and the BurkDiff assay (63) detects both organisms and differentiates them based on a SNP-associated shift of approximately 1 ΔCt (64). Here, we assessed *bucl* markers as detection targets. Standard PCR performed on a large collection of gDNA yielded specific amplicons for *bucl5*, *13, 14*, and *16* from Bp and Bm but not from the non-select agent controls. This PCR test with 4 *bucl* targets detected Bp infection in laboratory animals using spleen extracts as a specimen. Separation of these conserved amplicons by capillary electrophoresis in a phospholipid nanogel matrix allowed for size-based identification, similarly to previously described identification of *Aspergillus* spp. (31) and *Streptococcus pyogenes* (35). Precise microfluidic separation could be used for strain fingerprinting based on multiplexed amplicons generated with primers flanking the repetitive CL region, as previously tested with *B. anthracis* strains (32). Our achieved resolution <9 bp along a wide range of amplicon sizes will allow to differentiate two strains that differ by a single GXY repeat. We further developed a qPCR assay for *bucl16,* which detects both Bp and Bm; it was tested with purified genomic DNA templates, gDNA spiked with 5% human plasma, and spleen extracts from infected mice. The *bucl16* assay was able to detect as low as 50 CFU per reaction in diluted spleen samples; however, it should be noted that sample-to-sample variation was observed. The specimen type will also

affect detection outcome. For example, sputum and pus typically contain high bacterial loads (10^2-10^9 CFU/ mL) (89), whereas blood of 45% of patients with septicemic melioidosis had less than 1 CFU/ mL bacteria in the blood (90), which presents a sensitivity challenge, even for highly performing qPCR assays. In as much as current work was focused on a single assay, which would simultaneously detect both select agents similarly to BurkDiff assay, the ongoing research is focused on the development of a probe-based qPCR assay targeting nucleotide polymorphisms identified in *bucl3* and *bucl4* genes. In summary, selected *bucl* genes represent promising detection targets as they are both specific to and ubiquitously found in Bp and Bm strains.

MATERIALS AND METHODS

Ethics statement

Animal Studies: Animal research at the United States Army Medical Research Institute of Infectious Diseases (USAMRIID) was conducted under an animal use protocol approved by the USAMRIID Institutional Animal Care and Use Committee (approved by the USAMRIID-IACUC) in compliance with the Animal Welfare Act, PHS Policy, and other Federal statutes and regulations relating to animals and experiments involving animals. The facility where this research was conducted is accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International (AAALAC) and adheres to principles stated in the Guide for the Care and Use of Laboratory Animals, National Research Council, 2011. Tissue samples used in this study were generated in a previously published work (67). Briefly, challenged mice were observed at least daily for 14 days for clinical signs of illness. Humane endpoints were used during all studies, and mice were humanely euthanized when moribund, according to an endpoint score sheet. Animals were scored on a scale of 0–11: 0–2 = no significant clinical signs; 3–7 = significant clinical symptoms; such as subdued behavior, hunched appearance, absence of grooming, and impacted hind limb function and hind limb paralysis (increased monitoring was warranted and mice were checked at least twice per day); 8–11 = distress. Those animals receiving a score of 8–11 were humanely euthanized by CO_2 exposure using compressed CO_2 gas followed by cervical dislocation. The mice that were serially sampled were deeply anesthetized and then euthanized by exsanguination followed by

cervical dislocation. However, even with multiple observations per day, some animals died as a direct result of the infection in between observation periods.

Human plasma collection: Anonymized human plasma samples were utilized in quantitative PCR experiments. Plasma samples were obtained from an already-existing collection, which was established by the corresponding author (SL; IRB Protocol Number: 1308076685). Collection of human blood of healthy adults was performed in accordance with the Human Research Protections Policy at West Virginia University. This study was approved by the Institutional Review Board at West Virginia University (IORG0000194) and written informed consent was obtained from all participants.

Bioinformatic analyses

Burkholderia collagen-like proteins, designated Bucl, were identified by searching the Sanger Institute Pfam collagen database (PF01391). Bucl proteins found in *B. pseudomallei*, *B. mallei*, and *B. thailandensis* were categorized into 13 Bucl-protein types based on similar domain organization and primary sequence similarity. Next, nucleotide BLASTn search was performed using each of 13 *bucl*-gene sequence as a query against the NCBI Nucleotide collection (nr/nt) database, as well as whole genome shotgun contigs (wgs) database, to determine *bucl* distribution in completed *Burkholderia* spp. genomes. DNA analyses were performed using the Lasergene Core Suite v. 12 (DNASTAR, Inc., Madison, WI).

Protein structure prediction and modeling

Domain organization of Bucl proteins was adapted from the Pfam collagen database (91) and verified independently using the Fugue 2.0 Server (92), which additionally identified the putative Talin-1 domain within Bucl3. Presence of a signal peptide was predicted with the hidden Markov model component of the SignalP 3.0 Server (http://www.cbs.dtu.dk/services/SignalP-3.0/) (93-95). The presence of transmembrane domains was predicted with TMpred (96).

When possible, as in the case of Bucl8, a 3D model was generated by homology modeling. Best template was identified by employing profile hidden Markov models

(profile HMMs) and the program HMMer (97). Once the best template was identified (pdb code 3d5k, sequence identity 27%, residues 51-516), the model of Bucl8 outer membrane efflux protein (OEP) domains was generated using MODELLER 9V9 (57). Stereo-chemical quality of the model was improved by energy minimization using GROMACS (98).

Thermal stability along the predicted triple helices of Bucl-collagen domains was assessed with an algorithm developed by Persikov *et al.* 2005 (47). With this approach, a stability coefficient is assigned for every GXY triplet and averaged over a window of 5 tripeptide units. The averaged relative stability values are plotted against the tripeptide number in the collagen sequence.

Phylogenetic analyses

Both individual (with and without the collagen-like domains) and concatenated nucleotide sequences were aligned with ClustalV in the Megalign module in DNASTAR® Lasergene software, and verified manually. Maximum parsimony analyses were performed with 1000 bootstrap replicates using MEGA 6.06 (99), with the Tree-Bisection Reconnection heuristic search and 200 max trees saved. The evolutionary models used for each data-set were determined by MrModelTest 2.3 (100) with the Akaike Information Criterion (AIC). Bayesian analyses were performed within MrBayes 3.1.2 (101) implementing six Markov chains, 1000000 generations, with trees sampled every 100 iterations. Posterior probabilities were calculated using the last 20% of saved trees (burnin=8000). Cutoff values for significance were assigned 95 for Bayesian analysis and 70 for maximum parsimony analysis. All phylogenetic trees were constructed using the majority rule consensus. Trees were viewed in FIGTREE v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/). Phylogenies were constructed based on single *bucl* genes as well as concatenated *bucl* genes.

bucl distribution among *Burkholderia* species

bucl distribution was assessed in a broad collection of *Burkholderia* strains using genomic DNA (Table 2) obtained from: (i) NIH Biodefense and Emerging Infections Research Resources Repository, NIAID, NIH, (ii) Dr. Christopher Cote (The United States Army

Medical Research Institute of Infectious Disease), and (iii) Dr. Joanna Goldberg (Emory University). The total collection consisted of DNA from 25 *B. pseudomallei* and 20 *B. mallei* strains, and non-select agent control DNA from 4 *B. thailandensis*, 3 *B. cepacia*, 5 *B. cenocepacia*, and 6 *B. multivorans* strains. Analytical PCR was performed with primers targeting conserved non-collagenous regions of *bucl* alleles present in the reference strain *B. pseudomallei* K96243. PCR buffer (10 mM Tris-HCl, 1.5 mM $MgCl_2$, 50 mM KCl, pH 8.3) included 0.2 µM primers, 0.2 mM dNTP's, and 1.5 M betaine (Sigma-Aldrich, St. Louis, MO) to ameliorate amplification problems associated with high GC content (~68%) of *Burkholderia* genomes (41). A temperature gradient of 50-65°C was tested for each primer pair and gDNA of *B. pseudomallei* K96243 harboring all 13 *bucl* genes as a template; uniform amplification conditions were established for all *bucl* genes at an annealing temperature of 64°C. Amplification was performed with an in-house Taq polymerase as follows: 95°C, 5 min-[95°C 30 sec, 64°C 30 sec, 72°C 45 sec] x30 cycles-72°C, 10 min. 40 ng of template DNA was used for screening genomic DNA collection and reactions were carried out on a Bio-Rad S1000 thermal cycler. Resultant PCR products were analyzed on a 2% agarose gel with a 50-bp ladder DNA standard (New England Biolabs Inc., Boston, MA). Gels were imaged using the Eagle Eye II (Stratagene, La Jolla, CA), and FOTO/ Analyst Investigator/ Eclipse gel documentation workstation (Fotodyne, Harland, WI).

qPCR amplification of *bucl* targets

Testing of selected *bucl* amplicons by real-time PCR with SYBR green intercalating dye was performed to assess potential candidates for probe-based detection of *B. pseudomallei* and *B. mallei* species. Reactions were carried out with SsoAdvanced™ SYBR® Green Supermix (Bio-Rad, Hercules, CA), 0.5 µM concentration of each primer and 25 ng of gDNA from strain *B. pseudomallei* K96243 as a template in a total volume of 20 µL. Amplification curves were obtained with the following program: 95°C, 3 min-[95°C 5 sec, 64°C 10 sec]x35 cycles. qPCR was performed using a Bio-Rad CFX96 instrument and data analyzed with the CFX Manager™ software Version 3.0. PrimeTime® qPCR probe was developed for the *bucl16* gene, which yielded robust amplification in 5' nuclease qPCR assays, to detect *B. pseudomallei* and *B. mallei*

species. Locked nucleic acid (LNA) *bucl16*-based probe (Table 4) contained a 5'-FAM fluorophore and a 3'-Iowa Black fluorescent quencher. Reactions were carried out using SsoAdvanced™ Universal Probes supermix (Bio-Rad), 0.5 µM primers, 0.2 µM concentration of probe and 25 ng of gDNA template in a total volume of 20 µL. Amplification curves were obtained with the following program: 95°C 3 min-[95°C 5 sec, 64°C 10 sec]x35 cycles.

Capillary gel electrophoresis

Reagents for separation of DNA by capillary gel electrophoresis included the nanogel matrix composed of the phospholipids dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) and 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine (DHPC) (Avanti Polar Lipids, Alabaster, AL), 3-(N-morpholino)-propanesulfonic acid (MOPS) (Alfa Aesar, Ward Hill, MA) buffer, and SYBR green 1 (Life Technologies, Grand Island, NY). The phospholipid pseudogel was prepared at a molar ratio of [DMPC]/[DHPC] =2.5 at 10% wt/vol in an aqueous solution of 100 mM MOPS buffer (pH 7) in order to generate the nanogel separation matrix. Intercalating dye was incorporated into the nanogel at 1x concentration to enable fluorescent DNA detection. The 50-bp DNA ladder (New England BioLabs, Ipswich, MA) was used as a molecular size marker.

Separations were performed on a Beckman Coulter P/ACE MDQ system equipped with a laser-induced fluorescence detection module and a 3 mW air-cooled argon ion laser (λ_{ex} = 488 nm and λ_{em} = 520 nm). The fused silica capillary was conditioned prior to electrophoresis separation of DNA using previously described rinsing (62) and coating (102) procedures. The capillary was filled with liquid nanogel solution a temperature below 24°C (19-21°C); then the temperature was increased to 30°C in order to form the sieving gel for accurate sizing separations of PCR amplicons. DNA samples were electrokinetically injected under reverse polarity as previously described (103). Data collection and analysis were performed with 32 Karat Software version 5.0 (Beckman Coulter). Sizing was accomplished by co-injecting the *bucl5*, *13*, *14*, and *16* amplicons with two internal standards of known length that bracketed the size of the DNA targets. Internal standards of 100 bp and 250 bp were used to create a linear fit for DNA size (in bp) versus migration time. The resulting slope and intercept were then used to calculate

the size (length in bp) of the *bucl* gene targets based on their migration times. The reported values for the calculated DNA size and standard deviation (in bp) are an average for *n*=5 consecutive separations.

Detection of *B. pseudomallei* gDNA in infected mice and human plasma using *bucl* markers

BALB/c mice (female 7-10 weeks of age at time of challenge-National Cancer Institute, NCI-Frederick, MD) were injected by the intraperitoneal (i.p.) route. Mice were infected with a dose equivalent to approximately 6 times the LD_{50} of *B. pseudomallei* HBPUB10134a (LD_{50} is 10 CFU) (67). At various time points after infection mice were euthanized by exsanguination under deep anesthesia and spleens were harvested. Spleens were weighed and homogenized in RPMI 1640 medium (Life Technology, Grand Island, NY). Bacterial load in the freshly prepared spleen extracts was determined by plating serial dilutions on sheep blood agar (ThermoScientific Remel Products, KS). Plates were incubated at 37°C for two days before determining CFU counts. The spleen extracts were irradiated and confirmed sterile before use in PCR assays, and were stored at -70°C. Standard PCR for *bucl* genes *5*, *13*, *14*, and *16*, and probe-based qPCR for *bucl16* were performed as described above using 1 µL of DNA-containing spleen specimen. Additionally, qPCR reactions were performed with 30 ng *B. pseudomallei* K96243 gDNA spiked with 5% human plasma collected in EDTA tubes to test the feasibility of the assay on clinical samples containing plasma. qPCR experiments were performed in triplicate and Cq values were averaged.

Statistical analyses

Statistical significance of *bucl* presence in pathogenic *B. pseudomallei* and *B. mallei* strains vs. nonpathogenic *B. thailandensis* was performed using the Fisher Exact Probability Test, followed by calculation of Cramer's V squared.

ACKNOWLEDGMENTS

REFERENCES

1. Brodsky, B., and Ramshaw, J. A. (1997) The collagen triple-helix structure. *Matrix Biol* **15**, 545-554
2. Brodsky, B., and Persikov, A. V. (2005) Molecular structure of the collagen triple helix. *Adv Protein Chem* **70**, 301-339
3. Ricard-Blum, S. (2011) The collagen family. *Cold Spring Harbor perspectives in biology* **3**, a004978
4. Xu, Y., Keene, D. R., Bujnicki, J. M., Höök, M., and Lukomski, S. (2002) Streptococcal Scl1 and Scl2 Proteins Form Collagen-like Triple Helices. *J. Biol. Chem.* **277**, 27312-27318
5. Rasmussen, M., Eden, A., and Bjorck, L. (2000) SclA, a novel collagen-like surface protein of *Streptococcus pyogenes*. *Infect Immun* **68**, 6370-6377
6. Rasmussen, M., and Bjorck, L. (2001) Unique regulation of SclB-a novel collagen-like surface protein of *Streptococcus pyogenes*. *Infect Immun* **40**, 1427-1438
7. Lukomski, 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
8. Lukomski, S., Nakashima, K., Abdi, I., Cipriano, V. J., Shelvin, B. J., Graviss, 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
9. De Simone, A., Vitagliano, L., and Berisio, R. (2008) Role of hydration in collagen triple helix stabilization. *Biochem Biophys Res Commun* **372**, 121-125
10. Berisio, R., De Simone, A., Ruggiero, A., Improta, R., and Vitagliano, L. (2009) Role of side chains in collagen triple helix stabilization and partner recognition. *J Pept Sci* **15**, 131-140
11. Yu, Z., An, B., Ramshaw, J. A., and Brodsky, B. (2014) Bacterial collagen-like proteins that form triple-helical structures. *J Struct Biol* **186**, 451-461
12. Mohs, A., Silva, T., Yoshida, T., Amin, R., Lukomski, S., Inouye, M., and Brodsky, B. (2007) Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline. *J. Biol. Chem.* **282**, 29757-29765

13. Han, R., Zwiefka, A., Caswell, C. C., Xu, Y., Keene, D. R., Lukomska, E., Zhao, Z., Hook, M., et al. (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

14. 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**, 348-356

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

16. Sylvestre, P., Couture-Tosi, E., and Mock, M. (2002) A collagen-like surface glycoprotein is a structural component of the *Bacillus anthracis* exosporium. *Mol Microbiol* **45**, 169-178

17. Whatmore, A. M. (2001) *Streptococcus pyogenes sclB* encodes a putative hypervariable surface protein with a collagen-like repetitive structure. *Microbiol* **147**, 419-429

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

19. Vandersmissen, L., De Buck, E., Saels, V., Coil, D. A., and Anne, J. (2010) A *Legionella pneumophila* collagen-like protein encoded by a gene with a variable number of tandem repeats is involved in the adherence and invasion of host cells. *FEMS Microbiol Lett* **306**, 168-176

20. Karlstrom, A., Jacobsson, K., Flock, M., Flock, J., and Guss, B. (2004) Identification of a novel collagen-like protein, SclC, in *Streptococcus equi* using signal sequence phage display. *Vet Microbiol* **104**, 179-188

21. Pizarro-Guajardo, M., Olguin-Araneda, V., Barra-Carrasco, J., Brito-Silva, C., Sarker, M. R., and Paredes-Sabja, D. (2014) Characterization of the collagen-like exosporium protein, BclA1, of *Clostridium difficile* spores. *Anaerobe* **25**, 18-30

22. Beres, S., Sesso, R., Pinto, S., Hoe, N., Porcella, S., Deleo, F., and Musser, J. (2008) Genome sequence of a lancefield group C *Streptococcus zooepidemicus* strain causing epidemic nephritis: new information about an old disease. *PLoS ONE* **3**, e3026

23. Han, R., Caswell, C. C., Lukomska, E., Keene, D. R., Pawlowski, M., Bujnicki, J. M., Kim, J. K., and Lukomski, S. (2006) Binding of the low-density lipoprotein by streptococcal collagen-like protein Scl1 of *Streptococcus pyogenes*. *Mol Microbiol* **61**, 351-367

24. Caswell, C., Lukomska, E., Seo, N., Hook, M., and Lukomski, S. (2007) Scl1-dependent internalization of group A *Streptococcus* via direct interactions with the a2b1 integrin enhances pathogen survival and re-emergence. *Mol Microbiol* **64**, 1319-1331

25. Caswell, C. C., Barczyk, M., Keene, D. R., Lukomska, E., Gullberg, D. E., and Lukomski, S. (2008) Identification of the first prokaryotic collagen sequence motif that mediates binding to human collagen receptors, integrins a2b1 and a11b1. *J Biol Chem* **283**, 36168-36175

26. Caswell, C. C., Oliver-Kozup, H., Han, R., Lukomska, E., and Lukomski, S. (2010) Scl1, the multifunctional adhesin of group A *Streptococcus*, selectively binds cellular fibronectin and laminin, and mediates pathogen internalization by human cells. *FEMS Microbiol Lett* **303**, 61-68

27. Reuter, M., Caswell, C. C., Lukomski, S., and Zipfel, P. F. (2010) Binding of the human complement regulators CFHR1 and factor H by streptococcal collagen-like protein 1 (Scl1) via their conserved C termini allows control of the complement cascade at multiple levels. *J Biol Chem* **285**, 38473-38485

28. Oliver-Kozup, H., Martin, K. H., Schwegler-Berry, D., Green, B. J., Betts, C., Shinde, A. V., Van De Water, L., and Lukomski, S. (2013) The group A streptococcal collagen-like protein-1, Scl1, mediates biofilm formation by targeting the extra domain A-containing variant of cellular fibronectin expressed in wounded tissue. *Mol Microbiol* **87**, 672-689

29. Pahlman, L. I., Marx, P. F., Morgelin, M., Lukomski, S., Meijers, J. C. M., and Herwald, H. (2007) Thrombin-activatable fibrinolysis inhibitor binds to *Streptococcus pyogenes* by interacting with collagen-like proteins A and B. *J. Biol. Chem.* **282**, 24873-24881

30. Bozue, J., Moody, K., Cote, C., Stiles, B., Friedlander, A., Welkos, S., and Hale, M. (2007) *Bacillus anthracis* spores of the *bclA* mutant exhibit increased adherence to epithelial, fibroblast, and endothelial cells but not macrophages. *Infect Immun* **75**, 4498-4505

31. Tuntevski, K., Durney, B. C., Snyder, A. K., Lasala, P. R., Nayak, A. P., Green, B. J., Beezhold, D. H., Rio, R. V., et al. (2013) *Aspergillus* collagen-like (*acl*) genes: identification, sequence polymorphism and assessment for PCR-based pathogen detection. *Appl Environ Microbiol* **79**, 7882-7895

32. Leski, T. A., Caswell, C. C., Pawlowski, M., Klinke, D. J., Bujnicki, J. M., Hart, S. J., and Lukomski, S. (2009) Identification and classification of *bcl* genes and proteins of *Bacillus cereus* group organisms and their application in *Bacillus anthracis* detection and fingerprinting. *Appl Environ Microbiol* **75**, 7163-7172

33. Sylvestre, P., Couture-Tosi, E., and Mock, M. (2003) Polymorphism in the collagen-like region of the *Bacillus anthracis* BclA protein leads to variation in exosporium filament length. *J Bacteriol* **185**, 1555 - 1563

34. Castanha, E. R., Swiger, R. R., Senior, B., Fox, A., Waller, L. N., and Fox, K. F. (2006) Strain discrimination among *B. anthracis* and related organisms by characterization of *bclA* polymorphisms using PCR coupled with agarose gel or microchannel fluidics electrophoresis. *J Microbiol Methods* **64**, 27-45

35. Durney, B. C., Bachert, B. A., Sloane, B. S., Lukomski, S., Landers, J. P., and Holland, L. A. (2015) Reversible phospholipid nanogels for deoxyribonucleic acid fragment size determinations up to 1,500 base pairs and integrated sample stacking. *Anal Chim Acta*

36. Vandamme, P., and Dawyndt, P. (2011) Classification and identification of the *Burkholderia cepacia* complex: Past, present and future. *Systematic and applied microbiology* **34**, 87-95

37. Wheelis, M. (1998) First shots fired in biological warfare. *Nature* **395**, 213

38. White, N. J. Melioidosis. *The Lancet* **361**, 1715-1722

39. Anuntagool, N., Naigowit, P., Petkanchanapong, V., Aramsri, P., Panichakul, T., and Sirisinha, S. (2000) Monoclonal antibody-based rapid identification of *Burkholderia pseudomallei* in blood culture fluid from patients with community-acquired septicaemia. *J Med Microbiol* **49**, 1075-1078

40. Holden, M. T., Titball, R. W., Peacock, S. J., Cerdeno-Tarraga, A. M., Atkins, T., Crossman, L. C., Pitt, T., Churcher, C., et al. (2004) Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proc Natl Acad Sci U S A* **101**, 14240-14245

41. Nierman, W. C., DeShazer, D., Kim, H. S., Tettelin, H., Nelson, K. E., Feldblyum, T., Ulrich, R. L., Ronning, C. M., et al. (2004) Structural flexibility in the *Burkholderia mallei* genome. *Proc. Natl. Acad. Sci. USA* **101**, 14246-14251

42. Godoy, D., Randle, G., Simpson, A. J., Aanensen, D. M., Pitt, T. L., Kinoshita, R., and Spratt, B. G. (2003) Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, *Burkholderia pseudomallei* and *Burkholderia mallei*. *J Clin Microbiol* **41**, 2068-2079

43. Lazar Adler, N. R., Govan, B., Cullinane, M., Harper, M., Adler, B., and Boyce, J. D. (2009) The molecular and cellular basis of pathogenesis in melioidosis: how does *Burkholderia pseudomallei* cause disease? *FEMS Microbiol Rev* **33**, 1079-1099

44. Galyov, E. E., Brett, P. J., and DeShazer, D. (2010) Molecular insights into *Burkholderia pseudomallei* and *Burkholderia mallei* pathogenesis. *Annu. Rev. Microbiol.* **64**, 495-517

45. Karlstrom, A., Jacobsson, K., and Guss, B. (2005) SclC is a member of a novel family of collagen-like proteins in *Streptococcus equi* subspecies *equi* that are recognised by antibodies against SclC. *Vet Microbiol*

46. Losada, L., Ronning, C. M., DeShazer, D., Woods, D., Fedorova, N., Kim, H. S., Shabalina, S. A., Pearson, T. R., et al. (2010) Continuing evolution of *Burkholderia mallei* through genome reduction and large-scale rearrangements. *Genome biology and evolution* **2**, 102-116

47. Persikov, A. V., Ramshaw, J. A. M., and Brodsky, B. (2005) Prediction of collagen stability from amino acid sequence. *J. Biol. Chem.* **280**, 19343-19349

48. Brown, N. H., Gregory, S. L., Rickoll, W. L., Fessler, L. I., Prout, M., White, R. A., and Fristrom, J. W. (2002) Talin is essential for integrin function in Drosophila. *Dev Cell* **3**, 569-579

49. Tadokoro, S., Shattil, S. J., Eto, K., Tai, V., Liddington, R. C., de Pereda, J. M., Ginsberg, M. H., and Calderwood, D. A. (2003) Talin binding to integrin beta tails: a final common step in integrin activation. *Science* **302**, 103-106

50. Wegener, K. L., Partridge, A. W., Han, J., Pickford, A. R., Liddington, R. C., Ginsberg, M. H., and Campbell, I. D. (2007) Structural basis of integrin activation by talin. *Cell* **128**, 171-182

51. Sun, J., Deng, Z., and Yan, A. (2014) Bacterial multidrug efflux pumps: mechanisms, physiology and pharmacological exploitations. *Biochem Biophys Res Commun* **453**, 254-267

52. Tam, R., and Saier, M. H., Jr. (1993) Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiological reviews* **57**, 320-346

53. Thibault, F. M., Hernandez, E., Vidal, D. R., Girardet, M., and Cavallo, J.-D. (2004) Antibiotic susceptibility of 65 isolates of *Burkholderia pseudomallei* and *Burkholderia mallei* to 35 antimicrobial agents. *J Antimicrob Chemother* **54**, 1134-1138

54. Ikeda, M., Arai, M., Lao, D. M., and Shimizu, T. (2002) Transmembrane topology prediction methods: a re-assessment and improvement by a consensus method using a dataset of experimentally-characterized transmembrane topologies. *In Silico Biol* **2**, 19-33

55. Johnson, J. M., and Church, G. M. (1999) Alignment and structure prediction of divergent protein families: periplasmic and outer membrane proteins of bacterial efflux pumps. *J Mol Biol* **287**, 695-715

56. Phan, G., Benabdelhak, H., Lascombe, M. B., Benas, P., Rety, S., Picard, M., Ducruix, A., Etchebest, C., et al. (2010) Structural and dynamical insights into the opening mechanism of *P. aeruginosa* OprM channel. *Structure* **18**, 507-517

57. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M.-y., Pieper, U., and Sali, A. (2002) Comparative Protein Structure Modeling Using Modeller. in *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]*, John Wiley & Sons, Inc. pp

58. Akama, H., Matsuura, T., Kashiwagi, S., Yoneyama, H., Narita, S., Tsukihara, T., Nakagawa, A., and Nakae, T. (2004) Crystal structure of the membrane fusion protein, MexA, of the multidrug transporter in *Pseudomonas aeruginosa*. *J Biol Chem* **279**, 25939-25942

59. Yu, Y., Kim, H. S., Chua, H. H., Lin, C. H., Sim, S. H., Lin, D., Derr, A., Engels, R., et al. (2006) Genomic patterns of pathogen evolution revealed by comparison of *Burkholderia pseudomallei*, the causative agent of melioidosis, to avirulent *Burkholderia thailandensis*. *BMC Microbiol* **6**, 46

60. Pearson, T., Giffard, P., Beckstrom-Sternberg, S., Auerbach, R., Hornstra, H., Tuanyok, A., Price, E. P., Glass, M. B., et al. (2009) Phylogeographic reconstruction of a bacterial species with high levels of lateral gene transfer. *BMC Biol* **7**, 78

61. Brett, P. J., DeShazer, D., and Woods, D. E. (1998) *Burkholderia thailandensis* sp. nov., a *Burkholderia pseudomallei*-like species. *International journal of systematic bacteriology* **48 Pt 1**, 317-320

62. Durney, B. C., Lounsbury, J. A., Poe, B. L., Landers, J. P., and Holland, L. A. (2013) A thermally responsive phospholipid pseudogel: tunable DNA sieving with capillary electrophoresis. *Anal Chem* **85**, 6617-6625

63. Bowers, J. R., Engelthaler, D. M., Ginther, J. L., Pearson, T., Peacock, S. J., Tuanyok, A., Wagner, D. M., Currie, B. J., et al. (2010) BurkDiff: a real-time PCR allelic discrimination assay for *Burkholderia pseudomallei* and *B. mallei*. *PLoS ONE* **5**, e15413

64. Price, E. P., Dale, J. L., Cook, J. M., Sarovich, D. S., Seymour, M. L., Ginther, J. L., Kaufman, E. L., Beckstrom-Sternberg, S. M., et al. (2012) Development and

validation of *Burkholderia pseudomallei*-specific real-time PCR assays for clinical, environmental or forensic detection applications. *PLoS ONE* **7**, e37723

65.   Novak, R. T., Glass, M. B., Gee, J. E., Gal, D., Mayo, M. J., Currie, B. J., and Wilkins, P. P. (2006) Development and evaluation of a real-time PCR assay targeting the type III secretion system of *Burkholderia pseudomallei*. *J Clin Microbiol* **44**, 85-90

66.   Meumann, E. M., Novak, R. T., Gal, D., Kaestli, M. E., Mayo, M., Hanson, J. P., Spencer, E., Glass, M. B., et al. (2006) Clinical evaluation of a type III secretion system real-time PCR assay for diagnosing melioidosis. *J Clin Microbiol* **44**, 3028-3030

67.   Welkos, S. L., Klimko, C. P., Kern, S., Bearss, J., Bozue, J. A., Bernhards, R. C., Trevino, S., Waag, D., et al. (2015) Characterization of *Burkholderia pseudomallei* strains using a murine intraperitoneal infection model and in vitro macrophage assays. *PLoS ONE*

68.   Berisio, R., Vitagliano, L., Mazzarella, L., and Zagari, A. (2002) Recent progress on collagen triple helix structure, stability and assembly. *Protein Pept Lett* **9**, 107-116

69.   Okuyama, K. (2008) Revisiting the molecular structure of collagen. *Connect Tissue Res* **49**, 299-310

70.   Shoulders, M. D., and Raines, R. T. (2009) Collagen structure and stability. *Annu Rev Biochem* **78**, 929-958

71.   Chan, V. C., Ramshaw, J. A., Kirkpatrick, A., Beck, K., and Brodsky, B. (1997) Positional preferences of ionizable residues in Gly-X-Y triplets of the collagen triple-helix. *J Biol Chem* **272**, 31441-31446

72.   Leikina, E., Mertts, M. V., Kuznetsova, N., and Leikin, S. (2002) Type I collagen is thermally unstable at body temperature. *Proc Natl Acad Sci U S A* **99**, 1314-1318

73.   Rasmussen, M., Jacobsson, M., and Bjorck, L. (2003) Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J. Biol. Chem.* **278**, 32313-32316

74.   Ruggiero, A., Tizzano, B., Pedone, E., Pedone, C., Wilmanns, M., and Berisio, R. (2009) Crystal structure of the resuscitation-promoting factor (DeltaDUF)RpfB from *M. tuberculosis*. *J Mol Biol* **385**, 153-162

75.   Warawa, J., and Woods, D. E. (2005) Type III secretion system cluster 3 is required for maximal virulence of *Burkholderia pseudomallei* in a hamster infection model. *FEMS Microbiol Lett* **242**, 101-108

76.   Schwarz, S., Singh, P., Robertson, J. D., LeRoux, M., Skerrett, S. J., Goodlett, D. R., West, T. E., and Mougous, J. D. (2014) VgrG-5 is a *Burkholderia* type VI secretion system-exported protein required for multinucleated giant cell formation and virulence. *Infect Immun* **82**, 1445-1452

77.   Angus, A. A., Agapakis, C. M., Fong, S., Yerrapragada, S., Estrada-de los Santos, P., Yang, P., Song, N., Kano, S., et al. (2014) Plant-associated symbiotic *Burkholderia* species lack hallmark strategies required in mammalian pathogenesis. *PLoS ONE* **9**, e83779

78. Rainbow, L., Hart, C. A., and Winstanley, C. (2002) Distribution of type III secretion gene clusters in *Burkholderia pseudomallei*, *B. thailandensis* and *B. mallei*. *J Med Microbiol* **51**, 374-384

79. French, C. T., Toesca, I. J., Wu, T. H., Teslaa, T., Beaty, S. M., Wong, W., Liu, M., Schroder, I., et al. (2011) Dissection of the *Burkholderia* intracellular life cycle using a photothermal nanoblade. *Proc Natl Acad Sci U S A* **108**, 12095-12100

80. Stevens, M. P., Wood, M. W., Taylor, L. A., Monaghan, P., Hawes, P., Jones, P. W., Wallis, T. S., and Galyov, E. E. (2002) An Inv/Mxi-Spa-like type III protein secretion system in *Burkholderia pseudomallei* modulates intracellular behaviour of the pathogen. *Mol Microbiol* **46**, 649-659

81. Choh, L. C., Ong, G. H., Vellasamy, K. M., Kalaiselvam, K., Kang, W. T., Al-Maleki, A. R., Mariappan, V., and Vadivelu, J. (2013) *Burkholderia* vaccines: are we moving forward? *Frontiers in cellular and infection microbiology* **3**, 5

82. Moore, R. A., DeShazer, D., Reckseidler, S., Weissman, A., and Woods, D. E. (1999) Efflux-mediated aminoglycoside and macrolide resistance in *Burkholderia pseudomallei*. *Antimicrob Agents Chemother* **43**, 465-470

83. Chan, Y. Y., Tan, T. M., Ong, Y. M., and Chua, K. L. (2004) BpeAB-OprB, a multidrug efflux pump in *Burkholderia pseudomallei*. *Antimicrob Agents Chemother* **48**, 1128-1135

84. Podnecky, N. L., Wuthiekanun, V., Peacock, S. J., and Schweizer, H. P. (2013) The BpeEF-OprC efflux pump is responsible for widespread trimethoprim resistance in clinical and environmental *Burkholderia pseudomallei* isolates. *Antimicrob Agents Chemother* **57**, 4381-4386

85. Doxey, A. C., and McConkey, B. J. (2013) Prediction of molecular mimicry candidates in human pathogenic bacteria. *Virulence* **4**, 453-466

86. Hoffmaster, A. R., AuCoin, D., Baccam, P., Baggett, H. C., Baird, R., Bhengsri, S., Blaney, D. D., Brett, P. J., et al. (2015) Melioidosis diagnostic workshop, 2013. *Emerg Infect Dis* **21**

87. Van Zandt, K. E., Greer, M. T., and Gelhaus, H. C. (2013) Glanders: an overview of infection in humans. *Orphanet journal of rare diseases* **8**, 131

88. Currie, B. J. (2015) Melioidosis: evolving concepts in epidemiology, pathogenesis, and treatment. *Semin Respir Crit Care Med* **36**, 111-125

89. Wongsuvan, G., Limmathurotsakul, D., Wannapasni, S., Chierakul, W., Teerawattanasook, N., and Wuthiekanun, V. (2009) Lack of correlation of *Burkholderia pseudomallei* quantities in blood, urine, sputum and pus. *The Southeast Asian journal of tropical medicine and public health* **40**, 781-784

90. Walsh, A. L., Smith, M. D., Wuthiekanun, V., Suputtamongkol, Y., Chaowagul, W., Dance, D. A., Angus, B., and White, N. J. (1995) Prognostic significance of quantitative bacteremia in septicemic melioidosis. *Clin Infect Dis* **21**, 1498-1500

91. Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., et al. (2012) The Pfam protein families database. *Nucleic Acids Res* **40**, D290-301

92. Shi, J., Blundell, T. L., and Mizuguchi, K. (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.* **310**, 243-257

93. Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**, 783-795

94. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10**, 1-6

95. Nielsen, H., and Krogh, A. (1998) Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc Int Conf Intell Syst Mol Biol* **6**, 122-130

96. Hofmann, K., and Stoffel, W. (1993) TMBASE - A database of membrane spanning protein segments. *Biological Chemistry Hoppe-Seyler* **374**

97. Eddy, S. R. (2011) Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**, e1002195

98. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005) GROMACS: fast, flexible, and free. *J Comput Chem* **26**, 1701-1718

99. Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**, 2725-2729

100. Nylander, J. A. A. (2004) MrModeltest v2. Evolutionary Biology Centre, Uppsala University

101. Ronquist, F., and Huelsenbeck, J. P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574

102. White, C. M., Luo, R., Archer-Hartmann, S. A., and Holland, L. A. (2007) Electrophoretic screening of ligands under suppressed EOF with an inert phospholipid coating. *Electrophoresis* **28**, 3049-3055

103. Luo, R., Archer-Hartmann, S. A., and Holland, L. A. (2010) Transformable capillary electrophoresis for oligosaccharide separations using phospholipid additives. *Anal Chem* **82**, 1228-1233

A

bucl15
bucl14

B. pseudomallei
K96243_Ch1

4074.542 kb

bucl16

bucl6
bucl8 *bucl10*

bucl3
bucl5
bucl2

B. pseudomallei
K96243_Ch2

3173.005 kb

bucl13

bucl4
bucl1
bucl7

B

Bucl No.	Ch/ strand	Location	Length (bp)	Molecular weight (kDa)
1	2/-	2119909-2120421	513	17.21
2	2/+	2647246-2647895	650	24.14
3	2/+	2667859-2669577	1719	51.03
4	2/+	2202787-2203818	1032	34.85
5	2/+	326839-327465	627	21.42
6	1/+	2398535-2398690	156	4.80
7	2/+	1102922-1103488	567	20.51
8	1/+	2346940-2348973	2034	68.30
10	1/-	2171241-2171663	423	14.97
13	2/-	2293634-2294872	1239	44.13
14	1/-	952706-952977	272	10.52
15	1/-	3424165-3424360	196	7.08
16	1/+	1191473-1192156	684	22.83

C

M bucl1 bucl2 bucl3 bucl4 bucl5 bucl6 bucl7 bucl8 bucl10 bucl13 bucl14 bucl15 bucl16

[bp] 300 250 200 150 100 50

Figure 1

Identification and characterization of *bucl* genes in *B. pseudomallei* reference strain K96243. (A) Schematic representation of *bucl* distribution. Relative position and orientation of each *bucl* gene is shown; six *bucl* genes are present on chromosome one and seven on chromosome two. (B) Summary table of *bucl* distribution. *bucl* location, orientation, and length are mapped to the genome of Bp K96243. Molecular weight of each Bucl protein encoded by each *bucl* allele is shown. (C) PCR amplification of 13 *bucl* genes from Bp K96243. Primers were designed targeting the non-collagenous conserved regions, and PCR conditions were established for all *bucl* amplicons at a uniform annealing temperature of 64°C. Amplicon sizes; *bucl1*, 123 bp; *bucl2* 133 bp; *bucl3*, 166 bp; *bucl4*, 176 bp; *bucl5*, 216 bp; *bucl6*, 115 bp; *bucl7*, 264 bp; *bucl8*, 96 bp; *bucl10*, 109 bp; *bucl13*, 212 bp; *bucl14*, 178 bp; *bucl15*, 95 bp; and *bucl16*, 123 bp ; M, 50-bp DNA size marker.
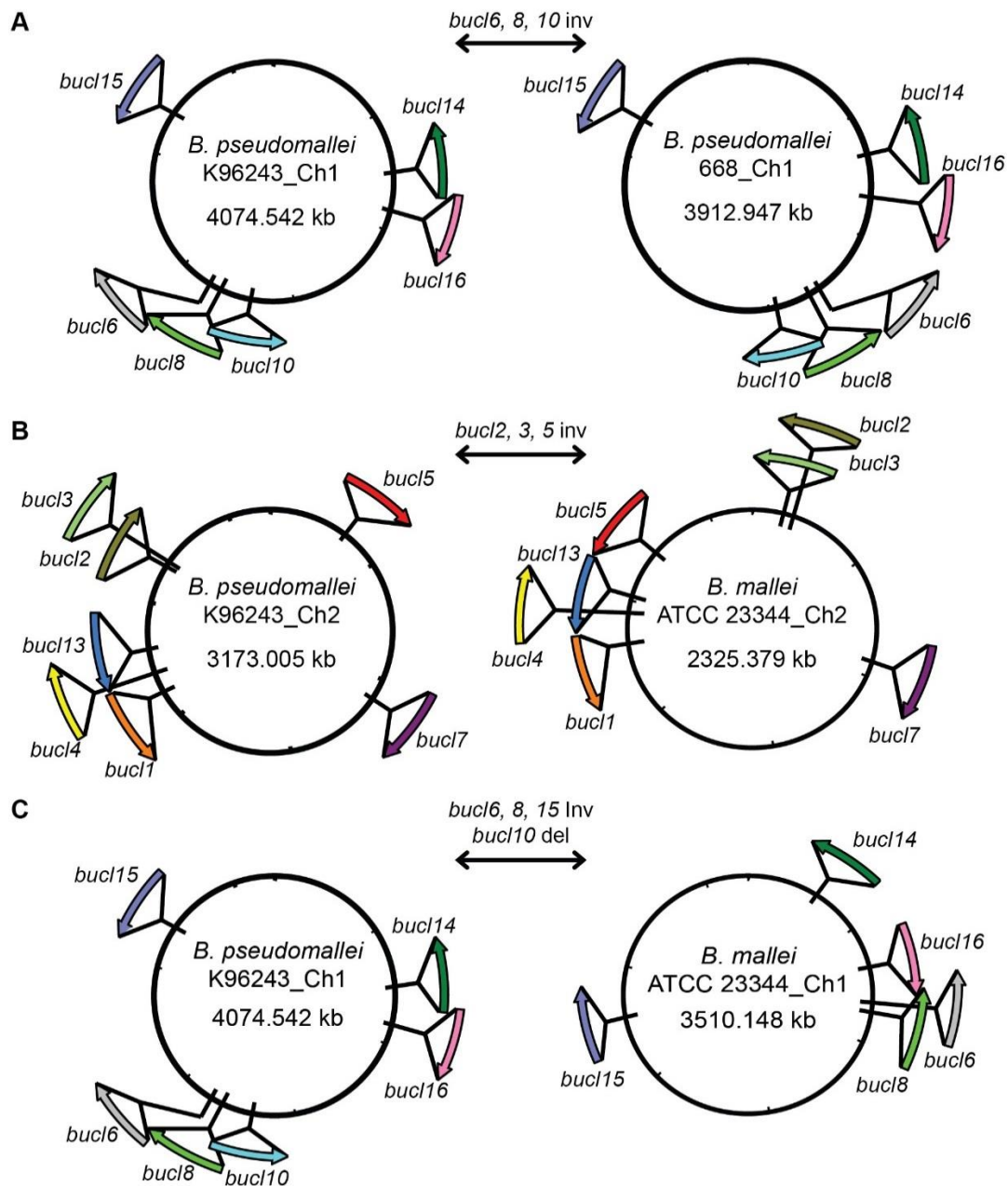
Figure 2

Chromosomal rearrangements and deletions involving *bucl* loci. Relative positions and orientations of each *bucl* gene was rendered from the NCBI database, and used for chromosomal mapping. (A) Intraspecies chromosomal inversion (inv) between *B. pseudomallei* strains K96243 and 668 involving the region encoding *bucl* genes 6, 8, and 10. (B) Interspecies chromosomal inversion between Bp K96243 and Bm ATCC 23344 involving the region encoding *bucl* genes 2, 3, and 5 on chromosome 2. (C) Interspecies chromosomal inversion involving *bucl* genes 6, 8, and 15, and deletion of *bucl*10 between Bp K96243 and Bm ATCC 23344 on chromosome 1. Ch, chromosome.

A

Bucl No.	CL region sequence	Average stability
1	$(GAN)_{16}$ -$(GTN)_4$-(GAA)-(GAD)	25.6°C
2	$(GEV)_{14}$-(GEA)-AEA-$(GEA)_2$-AEA-(GEV)-(GEA)	35.3°C
3	$(GTS)_{49}$-(GTA)	24.0°C
4	$(GVS)_{10}$-$(GAS)_{20}$-(GTS)	30.7°C
5	$(GLE)_{20}$-$(GPE)_2$-$(GLD)_2$-(GFD)	37.2°C
6	(GAL)-$(GAS)_2$-$(GAA)_{10}$-(GAS)-$(GAA)_5$-(GAE)	32.8°C
7	(GLS)-$(GSS)_{10}$-$(GAS)_2$-(GVA)	27.3°C
8	$(GAS)_{21}$	33.0°C
10	$(GIH)_{24}$-(GMR)	27.1°C
13	$(GIR)_{12}$-DIR-$(GIR)_2$-$(GVR)_8$-(GIR)	36.0°C
14	$(GWC)_4$-$(GRC)_5$-$(GWC)_2$-$(GRC)_{20}$-$(GRR)_5$-(GRA)	30.6°C
15	$(GAT)_2$-$(GAI)_6$-(GAA)	35.3°C
16	$(GFG)_2$-(GVD)-$(GFD)_{26}$-(GFG)	20.3°C

B

Figure 3

Thermal stability of the Bucl collagen regions. (A) The CL region sequences, representative of all 13 Bucl proteins, plotted in B) are shown with averaged stability values calculated for the entire CL region. (B) Triple helix thermal stability plot. Amino acid sequences for Bucl-CL regions shown in A) were used to model thermal stability with an algorithm developed by Persikov et al. 2005. Relative thermal stability is shown as the melting temperature for each GXY triplet along each Bucl-CL region.
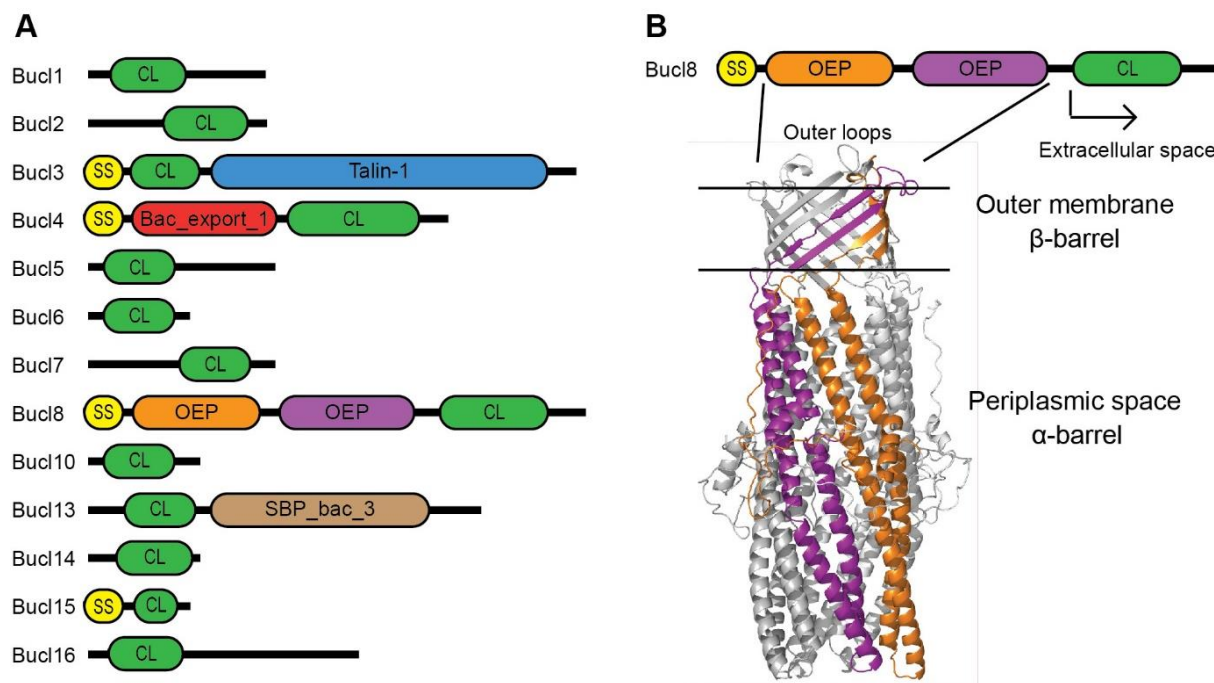
Figure 4

Characterization of *Burkholderia* collagen-like proteins. (A) Architecture of Bucl proteins identified in collagen Pfam data base (not to scale). Proteins were categorized into 13 distinct Bucl types based on sequence similarities and domain organization. Predicted domains in each Bucl are shown: SS, signal sequence; CL, collagen-like domain; Talin-1 domain; Bac_export_1, bacterial export protein family 1; OEP, Outer Membrane Efflux Protein; and SBP_bac_3, bacterial extracellular solute-binding protein family 3. (B) Cellular organization of Bucl8 and homology modelling of the OEP domains. Bucl8 protein schematic is shown above homology model of OEP domains generated with MODELLER. Three monomers, each containing two OEP domains, assemble to form a homotrimer. Shown from top to bottom are the cell-surface exposed loops, the β-barrel spanning the outer membrane and the α-barrel spanning the periplasmic space, corresponding to the predicted OEP domains. The two OEP domains from a single monomer are highlighted in orange and purple, and the remaining monomers are colored gray. Following the OEP domains, the CL region is predicted to be partially extracellular with an additional C-terminal non-collagenous domain.
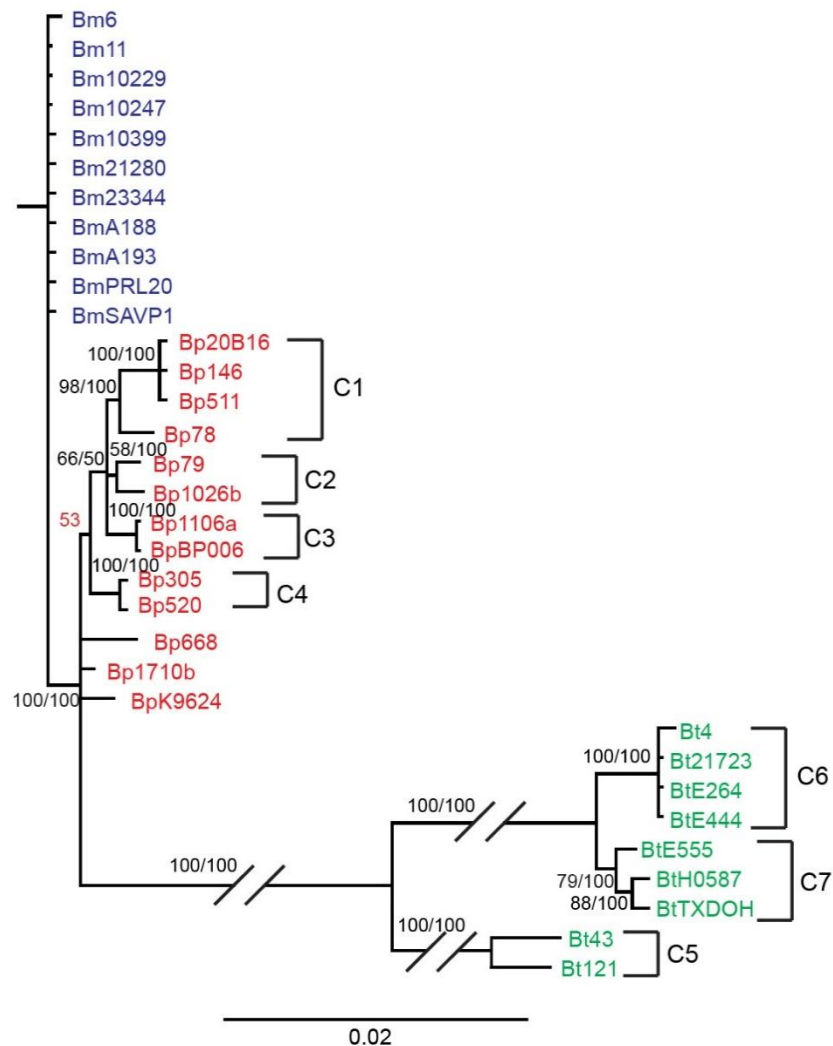
Figure 5

Phylogenetic analysis of *B. pseudomallei*, *B. mallei*, and *B. thailandensis* strains by *bucl*-locus typing. Bayesian analysis was performed on concatenated nucleotide sequences of the non-collagenous regions of *bucl3*, *bucl4,* and *bucl8* present in a set of 13 *B. pseudomallei*, 11 *B. mallei*, and 9 *B. thailandensis* strains (as shown in Table 3). Support values for each branch are shown as posterior probability from Bayesian analysis and bootstrap values from maximum parsimony analysis, respectively (PP/MP). Posterior probability value which was not supported by maximum parsimony analysis is shown in red. Phylogenetic Clusters 1-4 (C1-C4) correlated with geographic location of *B. pseudomallei* strains, whereas Clusters 5-7 (C5-C7) contained *B. thailandensis* strains that made up a separate branch from *B. pseudomallei* and *B. mallei* strains. Scale bar is representative of evolutionary distance in substitutions per nucleotide.

227
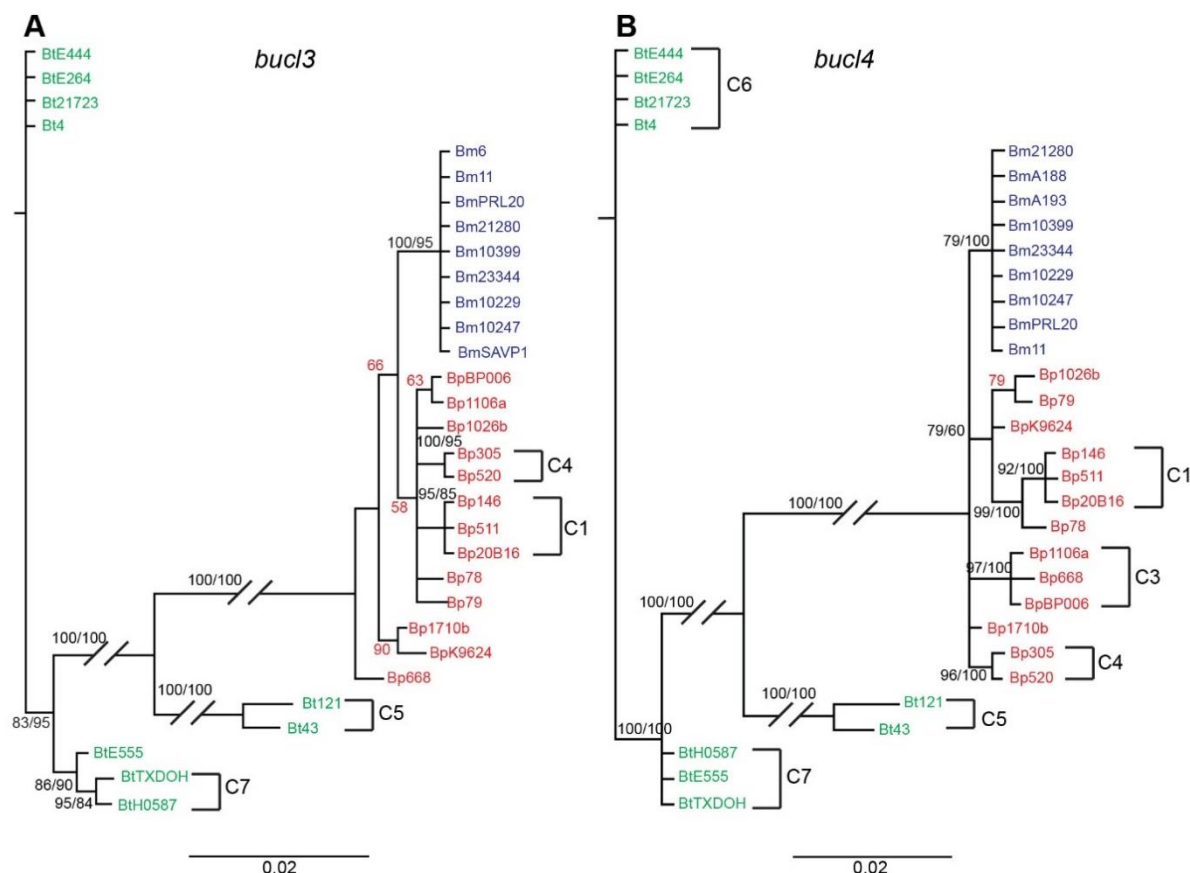
Figure 6

Phylogenetic analysis of *B. pseudomallei*, *B. mallei*, and *B. thailandensis* strains using individual *bucl3* and *bucl4* genes. Bayesian analysis was performed on nucleotide sequences of non-collagenous regions of a set of *Burkholderia* strains described in Table 3. Support values for each branch are shown as posterior probability from Bayesian analysis and bootstrap values from maximum parsimony analysis, respectively (PP/MP). Posterior probability values not supported by parsimony analysis are shown in red. Scale bar is representative of evolutionary distance in substitutions per nucleotide. Several clusters of strains corresponding to those observed in the concatenated analysis, C1-C7 in Fig. 5, were also observed in the individual trees.
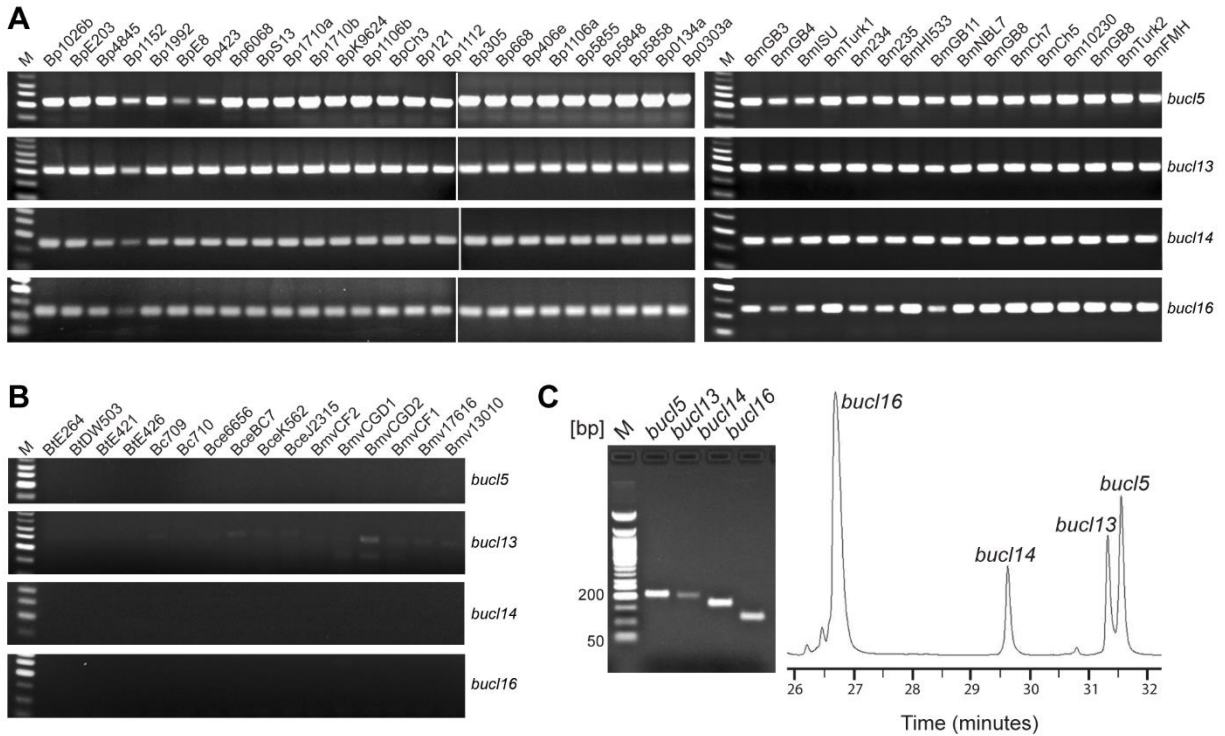
Figure 7

Distribution of *bucl* genes among *Burkholderia* spp. select agents by PCR. Presence of *bucl* genes was assessed by PCR on (A) a collection of genomic DNA from 25 *B. pseudomallei* and 16 *B. mallei* strains, as well as (B) in control strains of *B. thailandensis*, *B. cepacia*, *B. cenocepacia*, and *B. multivorans*; selected *bucl* genes *5, 13, 14,* and *16* are shown. (C) Detection and separation of selected *bucl* amplicons generated from the *B. pseudomallei* reference strain K96243 by traditional 2% agarose gel electrophoresis (left) or by capillary gel electrophoresis (right). Electropherogram generated by capillary gel electrophoresis with phospholipid nanogel matrix shows separation of amplicons over time. Amplicon sizes: *bucl5*, 216 bp; *bucl13*, 214 bp; *bucl14*, 178 bp; and *bucl16*, 123 bp. M, 50-bp DNA ladder. PCR data shown in Panel A for 25 Bp strains come from two merged gel images.

229

Figure 8

Detection of *B. pseudomallei* and *B. mallei* by qPCR. (A) Real-time qPCR detection of *bucl16*-gene target. Genomic DNA of 25 *B. pseudomallei* (red) and 15 *B. mallei* strains (blue), and control DNA from 4 *B. thailandensis*, 4 *B. cenocepacia*, and 6 *B. multivorans* strains (gray). (B) qPCR detection of *bucl16* target in the presence of human plasma and in spleen extracts from infected mice. 25 ng of gDNA from Bp K96243 was used as a positive control (blue line). Amplification of *bucl16* in qPCR reaction spiked with 5% human plasma is shown (green line). Mice were infected with Bp HBPUB10134a and CFU counts used in each qPCR reaction were based on plating spleen extracts on blood agar. Positive amplification is shown for spleen samples with 5×10^4 CFU (red lines: square, undiluted; triangle, 1:10 dilution; circle, 1:100 dilution; diamond, 1:1000 dilution) and 5×10^3 CFU (gray lines: square, undiluted; triangle, 1:10 dilution), while no amplification was obtained for crude spleen samples with original 2×10^2 CFU and 10 CFU per reaction and no template control, NTC (black lines). Inset; amplification of *bucl* markers *5, 13, 14,* and *16* by standard PCR using crude spleen samples containing 5×10^4 CFU per reaction. M, 50-bp DNA ladder.

Table 1. Assessment of genomic plasticity of *B. pseudomallei* and *B. mallei* using biomarkers.

| OP[a] | Strains | *bucl*s on Chromosome 1 | | Ch2 OP | Strains | *bucl*s on Chromosome 2 | |
		plus strand	minus strand			plus strand	minus strand
Ch1 OPI	BpK9624	6, 8, 16	10, 14, 15	Ch2 OPI	BpK9624, Bp668, Bp1026b, Bp1106a, Bp1710b, BpBP006, Bp305, Bp146, Bp511, Bp520, Bp20B16, Bp78, Bm10229, Bm10247, BmSAVP1	2, 3, 4, 5 , 7	1, 13
Ch1 OPII	Bp668, Bp1026b, Bp1106a, Bp1710b, BpBP006, Bp305, Bp79, Bp146, Bp511, Bp520, Bp20B16, Bp78	10, 16	6, 8, 14, 15	Ch2 OPII	Bp79	2, 3, 4, 5	1, 7, 13
Ch1 OPIII	Bm23344	15, 16	6, 8, 14	Ch2 OPIII	Bm23344	4, 7	1, 2, 3, 5, 13
Ch1 OPIV	Bm10229	6, 8, 14, 15, 16	10				
Ch1 OPV	Bm10247	10, 15	6, 8, 14, 16				
Ch1 OPVI	BmSAVP1	15	6, 8, 10, 14, 16				

[a]Organizational patterns (OP) of *bucl* genes were assigned to each chromosome, Ch1 and Ch2, according to position and orientation. OPs were labeled I-VI for chromosome 1, and I-III for chromosome 2. *bucl* position on the plus or minus strand is shown corresponding to each OP.

Table 2. Characterization of Bucl proteins in *Burkholderia*[a].

Bucl No.	No. of amino acids			Collagen-like region (CL)		Structural predictions		
	Total	N-terminus	C-terminus	No. of GXY repeats	GXY type	Putative domains	[b]SS	[c]TM
Bucl1	152-197	42	80-89	7-22	**GAN**	N/A	No	Yes
Bucl2	171-228	141	21	3-19	**GEV.** GEA	N/A	No	No
Bucl3	551-640	44	372-405	38-63	**GTS.** GSS	Talin-1	Yes	Yes
Bucl4	297-379	271	9-18	7-30	**GVS. GAS**	Bac_export_1	Yes	Yes, CL region
Bucl5	168-230	35-41	114	7-25	**GLE.** GPE. GLD. GFD	N/A	No	No
Bucl6	40-88	1	27	4-20	GAL. **GAS. GAA.** GAE	N/A	No	Yes, CL region
Bucl7	188-212	134	36	5-14	**GLS. GSS. GAS.** GVA	N/A	No	Yes, CL region
Bucl8	608-677	522	74	4-24	**GAS**	OEP	Yes	Yes, CL region
Bucl10	92-155	2-8	63	8-25	**GIH.** GMH. GMR	N/A	No	No
Bucl13	385-433	24	318	12-25	**GIR.** GVR. GSG. GGS.	SBP_bac_3	No	Yes
Bucl14	83-191	11	69	2-37	**GWC. GRC. GRR.** GRH	N/A	No	Yes, CL region
Bucl15	56-91	21-69	4	5-14	GVL. GAL. GML. **GAT. GAI.** GAA	N/A	Yes	Yes, CL region
Bucl16	227-307	65-67	148	4-30	GFG. GVD. **GFD. GAF**	N/A	No	Yes, CL region

[a]Characteristics of Bucl proteins are shown based on analysis of completed genomes of 13 Bp, 11 Bm, and 9 Bt strains (see Table 3). The total protein length and length of protein sequences that are amino- and carboxyl-terminal to CL regions in each Bucl protein is shown as amino acid number, whereas the length of each CL region, which varies between strains, is expressed as the number of GXY repeats. Predominant GXY repeats are represented in bold text. Putative domains in the noncollagenous regions of each Bucl are shown: Talin-1 domain; Bac_export_1, bacterial export protein family 1; OEP, outer membrane efflux protein; and SBP_bac_3, bacterial extracellular solute-binding protein family 3.

[b]SS; Signal sequence predictions are based on hidden Markov model predictions in the SignalP 3.0 server.

[c]TM; Transmembrane domain predictions were made using TMPred.

Table 3. *Burkholderia* strains used in this study[a].

Species	Abbreviation	Strain	Isolate information	
B. pseudomallei	BpK9624[b]	K96243	female diabetic patient- Khon Kaen hospital, Northeast Thailand	1996
B. pseudomallei	Bp1710b[b]	1710b	relapse of same patient infected with 1710a, blood culture, Northeast Thailand, Sappasithiprasong hospital	1999
B. pseudomallei	Bp305[b]	MSHR305	brain sample, fatal encephalomyelitis, Australia, Royal Darwin hospital	1994
B. pseudomallei	Bp1026b[b]	1026b	blood culture from 29-year old female rice farmer with diabetes milletus, Northeast Thailand, Sappasithiprasong hospital	1993
B. pseudomallei	BpBP006	BPC006	Blood from patient with Type I diabetes and multiple abscesses, China, Baoting Town, Hainan	2008
B. pseudomallei	Bp1106a[b]	1106a	female rice farmer, Northeast Thailand, Sappasithiprasong hospital	1993
B. pseudomallei	Bp79	NCTC 13179	skin ulcer, Australia	2014
B. pseudomallei	Bp668[b]	668	blood culture from 53-year old male patient with severe melioidosis encephalomyelitis, Darwin Australia	1995
B. pseudomallei	Bp146	MSHR146	goat udder, Australia	1992
B. pseudomallei	Bp511	MSHR511	throat of goat, Australia	1997
B. pseudomallei	Bp520	MSHR520	human blood culture, Australia	1998
B. pseudomallei	Bp20B16	NAU20B-16	soil, Australia	2006
B. pseudomallei	Bp78	NCTC 13178	human post-mortem brain, Australia	N/A
B. mallei	BmSAVP1	SAVP1	pathogenic strain which became avirulent after passage through 6 equids, originally caused disease in a mule in India	
B. mallei	Bm10229	NCTC 10229	Europe	
B. mallei	Bm10247	NCTC 10247	Europe	
B. mallei	Bm23344[b]	ATCC 23344	human post-mortem knee fluid, skin pustules and blood, Burma	1944
B. mallei	Bm21280[b]	2002721280	Iran	1952
B. mallei	BmA188	A188	>8 passages	
B. mallei	BmA193	A193	Pasteur Institute, France	1964
B. mallei	Bm10399[b]	ATCC 10399	horse lung, Southern China	1949
B. mallei	BmPRL20	PRL-20	blood of a gelding from the Lahore Polo Club, Lahore, Pakistan	2005
B. mallei	Bm11	strain_11	human, Turkey	1949
B. mallei	Bm6	strain_6	human	1950
B. thailandensis	BtE264	E264	rice field soil sample, Central Thailand	
B. thailandensis	Bt121	MSMB121	soil, Australia	2007
B. thailandensis	BtH0587	H0587	human pleural wound, LA, United States	1997
B. thailandensis	BtE444	E444	soil, Thailand	2002
B. thailandensis	Bt43	MSMB43	bore water source in Darwin, Australia; first isolate of *B. thailandensis* in Australia	

B. thailandensis	BtTXDOH	TXDOH	United States		
B. thailandensis	Bt21723	2002721723	Human, CDC		2010
B. thailandensis	Bt4	4			
B. thailandensis	BtE555	E555			

[a]13 Bp, 11 Bm, and 9 Bt strains listed in this table were used for analysis of Bucl characteristics (Table 2), in part for distribution assessment (Table 4), and for phylogenetic analyses. Strain abbreviations listed in this table are used in all figures.
[b]Strains which were also tested by PCR for *bucl* distribution.

Table 4. Distribution of all *bucl* genes in *Burkholderia spp.* as assessed by bioinformatics and PCR amplification[a].

Strain Abbreviatio	*bucl* No.												
	1	2	3	4	5	6	7	8	10	13	14	15	16
BpBP006[b]	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp79[b]	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp146[b]	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp511[b]	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp520[b]	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp20B16[b]	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp78[b]	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp1026b[b]	+	+	+	+	+	+	+	+	+	+	+	+	+
BpE203	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp4845	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp1152	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp1992	+	+	+	+	+	+	+	+	+	+	+	+	+
BpE8	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp423	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp6068	+	+	+	+	+	+	+	+	+	+	+	+	+
BpS13	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp1710a	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp1710b[b]	+	+	+	+	+	+	+	+	+	+	+	+	+
BpK9624[b]	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp1106b	+	+	+	+	+	+	+	+	+	+	+	+	+
BpCh3	-	NT	NT	-	+	+	+	+	NT	+	+	+	+
Bp121	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp1112	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp305[b]	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp668[b]	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp406e	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp1106a[b]	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp5855	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp5848	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp5858	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp0134a	+	+	+	+	+	+	+	+	+	+	+	+	+
Bp0303a	+	+	+	+	+	+	+	+	+	+	+	+	+
Bm10229[b]	+	-	+	+	+	+	+	+	+	+	+	+	+
Bm10247[b]	+	+	+	+	+	+	+	+	+	+	+	+	+
BmA188[b]	+	-	-	+	+	+	+	+	+	+	+	+	+
BmA193[b]	+	-	-	+	+	+	-	+	+	+	+	+	+
Bm10399[b]	+	+	+	+	+	+	+	+	+	+	+	+	+
BmPRL20[b]	+	-	+	+	+	+	+	+	+	+	+	+	+
Bm6[b]	+	-	+	-	+	+	+	+	+	+	+	+	+
Bm11[b]	+	+	+	+	+	+	+	+	+	+	+	+	+
BmSAVP1[b]	-	+	+	-	+	+	-	+	+	+	+	+	+
BmGB3	+	-	-	-	+	+	+	+	+	+	+	+	+
BmGB4	+	+	+	-	+	+	+	+	+	+	+	+	+
BmISU	+	+	+	+	+	+	+	+	-	+	+	+	+
BmTurk1[c]	+	+	+	+	+	+	+	+	+	+	+	+	+
Bm234	+	+	+	+	+	+	+	+	-	+	+	+	+
Bm235	+	+	+	+	+	+	+	+	-	+	+	+	+
BmHI533	+	+	+	+	+	+	+	+	-	+	+	+	+
BmGB11	+	-	+	+	+	+	+	+	+	+	+	+	+
BmNBL7	+	+	+	+	+	+	+	+	-	+	+	+	+
BmGB8	+	+	+	+	+	+	+	+	-	+	+	+	+
Bm23344[b]	+	+	+	+	+	+	+	+	-	+	+	+	+
BmTurk2[c]	+	+	+	-	+	+	+	+	+	+	+	+	+

BmFMH	+	+	+	+	+	+	+	+	-	+	+	+	+
Bm21280[b]	+	+	+	+	+	+	+	+	-	+	+	-	+
Bm85567	+	+	+	+	+	+	+	+	+	+	+	+	+
Bm2700C	+	-	+	+	+	+	+	+	+	+	+	+	+
BmCh7	+	+	+	+	+	+	+	+	-	+	+	+	+
BmCh5	+	+	+	+	+	+	+	+	+	+	+	+	+
Bm10230	+	-	+	+	+	+	+	+	+	+	+	+	+
BmGB8	+	+	+	+	+	+	+	+	-	+	+	+	+
Bt21723[b]	-	-	+	+	-	-	-	-	-	-	-	-	-
BtH0587[b]	-	-	+	+	-	-	-	-	-	-	-	-	-
BtE444[b]	-	-	+	+	-	-	-	-	-	-	-	-	-
Bt121[b]	-	-	+	+	-	-	-	-	-	-	-	-	-
BtE555[b]	-	-	+	+	-	-	-	-	-	-	-	-	-
Bt43[b]	-	-	+	+	-	-	-	-	-	-	-	-	-
Bt4[b]	-	-	+	+	-	-	-	-	-	-	-	-	-
BtTXDOH[b]	-	-	+	+	-	-	-	-	-	-	-	-	-
BtE264[b]	-	-	+	+	-	-	-	-	-	-	-	-	-
BtDW503	-	-	+	+	-	-	-	-	-	-	-	-	-
BtE421	-	-	+	+	-	-	-	-	-	-	-	-	-
BtE426	-	-	+	+	-	-	-	-	-	-	-	-	-
Fisher p-value	<0.0001		0.581	0.353	<0.0001								
Cramer's V²	0.829	0.486	0.009	0.018	1	1	0.829	1	0.426	1	1	0.908	1
Bc706	-	-	-	-	NT	NT	NT	NT	-	NT	NT	NT	NT
Bc709	-	-	-	-	-	-	-	-	-	-	-	-	-
Bc710	-	-	-	-	-	-	-	-	-	-	-	-	-
Bce6656	-	-	-	-	-	-	-	-	-	-	-	-	-
BceBC7	-	-	-	-	-	-	-	-	-	-	-	-	-
BceK562	-	-	-	-	-	-	-	-	-	-	-	-	-
BceJ2315	-	-	-	-	-	-	-	-	-	-	-	-	-
Bce103a2	-	-	-	-	NT	NT	-	NT	-	NT	-	NT	-
BmvCF2	-	-	-	-	-	-	-	-	-	-	-	-	-
BmvCGD1	-	-	-	-	-	-	-	-	-	-	-	-	-
BmvCGD2	-	-	-	-	-	-	-	-	-	-	-	-	-
BmvCF1	-	-	-	-	-	-	-	-	-	-	-	-	-
Bmv17616	-	-	-	-	-	-	-	-	-	-	-	-	-
Bmv13010	-	-	-	-	-	-	-	-	-	-	-	-	-

[a]Presence or absence of *bucl* amplicons are indicated by + or –, respectively. NT, not tested (not sufficient amount of gDNA available). Association of *bucl* presence with pathogenic Bp and Bm species, compared to nonpathogenic Bt strains, was assessed using the Fisher Exact Probability Test and Cramer's V² analysis.
[b]Strains for which *bucl* presence was determined by bioinformatics.
[c]Smaller-sized amplicons observed for *bucl2* amplicon (S4 Fig. 4A).

Table 5. Primers and probe used for *bucl* amplification[a].

bucl#	Primer name	Primer sequence 5'-3'	Amplicon size
bucl1	Bucl1_4F	GTGGCGCTGGCGCATCGTGAACGGC	103 bp
	Bucl1_4R	CTTCGTCGGTTGCGTGTCGTCCGTTGC	
bucl2	Bucl2_1F	CGGCGTGCGACGGAA	133 bp
	Bucl2_1R	GCCCACTTCGCGATTCTTC	
bucl3	Bucl3_2F	CTGCTCGGCGGCCTGTCGGGTTCGG	166 bp
	Bucl3_2R	CGGGCGCGGTCGTCGTCGA	
bucl4	Bucl4_2F_ext	GACGAATTCATCCGCTTCATCGTG	176 bp
	Bucl4_2R_ext2	CCGCTGCGCATCGGGCCTTTCA	
bucl5	Bucl5_2F	AACTCGACGAACTCAACGCGAATCGAC	216 bp
	Bucl5_2R	GCGCGCCGTTCTTTCTAGCGCTGC	
bucl6	Bucl6_CL flank_F	AGGAGCGGCGCTTGCCGGGCG	115 bp[b]
	Bucl6_Clflank_2R	GAACGGCGACGGTCCGACGCAGC	
bucl7	Bucl7_2F	ATGGACACGACCACGCAGGACGGG	264 bp
	Bucl7_2R	CCAATGAACGGCCCGCGTCGCTTTC	
bucl8	Bucl8_2F	GCAGCTCGATTCGTGGAT	243 bp
	Bucl8_2R	AGGTGGTACGACAGGCTCAG	
	Bucl8_3F	CTACGCGCTCCTCGACATCGCGC	96 bp
	Bucl8_3R	TGCGTGCCGATGCCCGCGCGCA	
bucl10	Bucl10_1F	GCATGCGTTGGACACGA	109 bp
	Bucl10_1R	GCAACGTCGTCATCTCGTC	
bucl13	Bucl13_2F	GTTCGATTTCACGACGTACCGGCTCG	212 bp
	Bucl13_2R	CGTCGTCGTCGAAGTACAGCACGTC	
bucl14	Bucl14_1F	TCGGCACATCTGTCGCCGCGAACC	178 bp
	Bucl14_1R	CGTATGGCCGCCGTGTCGATCGG	
bucl15	Bucl15_1F	GATCGCTCGACGCGCCCGRCGTGC	95 bp[b]
	Bucl15_1R	CTAAAACCGCCGGCGYGCCGCGC	
bucl16	Bucl16_2F	CCGGCAGCACCGACTCGAGCGTGCG	123 bp
	Bucl16_2R	CGTCGTTCGMGCTCGCCGATCGCTCG	
	[c]Bucl16_5'FAM_3'IBQ	TCTGCA+CG+G+CG+GTG+AGCCGCTTCA	

[a]Primers were designed to generate conserved amplicons within the non-collagenous region of each *bucl* gene. Primers for *bucl6* and *bucl15* were designed flanking the collagenous region, which varies in size among strains (S4 Fig. 4B). Primers Bucl8_3F/3R were used to generate amplicon from Bp K96243 reference strain, whereas primers Bucl8_2F/2R selectively amplify products from Bp and Bm gDNA and were used in large-scale PCR screening (S4 Fig. 4B).
[b]Amplicon sizes expected for the Bp K96243 reference strain.
[c]LNA probe for *bucl16* detection; + symbols precede LNA bases.

Table 6. Genomic DNA collection.

Species	Abbreviation	Strain	Isolate information			Source of DNA
			Alternative designations	Source of isolate	Year	
B. pseudomallei	Bp1026b	1026B		blood culture from 29-year old female rice farmer with diabetes milletus, Northeast Thailand, Sappasithiprasong hospital	1993	USAMRIID[a]
B. pseudomallei	BpE203	E203		Soil sample from Roi Et, Thailand	1997	USAMRIID
B. pseudomallei	Bp4845	NCTC4845	(S. 397, NRRL B-1112, CCEB 472)	Monkey, Singapore	1935	USAMRIID
B. pseudomallei	Bp1152	STW-115-2		water, Thailand	1965	USAMRIID
B. pseudomallei	Bp1992	STW-199-2		water, Thailand	1965	USAMRIID
B. pseudomallei	BpE8	E8		Soil sample obtained on road to Trakan Phuet Phon District, Ubon Ratchathani Thailand		USAMRIID
B. pseudomallei	Bp423	423		Blood culture, Cambodia	2008	USAMRIID
B. pseudomallei	Bp6068	Pasteur 6068	2002721763	Vietnam		BEI Resources[b]
B. pseudomallei	Bp13	S13		muicodal strain, environmental isolate, Singapore		BEI Resources
B. pseudomallei	Bp1710a	1710a		blood culture of 52-year old male rice farmer with diabetes milletus, Northeast Thailand	1996	BEI Resources
B. pseudomallei	Bp1710b	1710b		relapse of same patient infected with 1710a, blood culture, Northeast Thailand, Sappasithiprasong hospital	1999	BEI Resources
B. pseudomallei	BpK9624	K96243		female diabetic patient- Khon Kaen hospital, Northeast Thailand	1996	BEI Resources
B. pseudomallei	Bp1106b	1106b		relapse of same patient infected with 1106a- female rice farmer, pus aspirated from liver abscess, Northeast Thailand, Sappasithiprasong hospital	1996	BEI Resources
B. pseudomallei	BpCh3	China 3		septicemia of American soldier, Burma		BEI Resources
B. pseudomallei	Bp121	NBL 121	strain 286, MP-S	chronic melioidosis case, infection acquired while living in Far East, Louisiana, United States	1953	BEI Resources
B. pseudomallei	Bp1112	NRRL B-1112	strain S 397, CCEB 472	naturally infected lab monkey, Singapore	1935	BEI Resources
B. pseudomallei	Bp305	MSHR305		brain sample, fatal encephalomyelitis, Australia, Royal Darwin hospital	1994	USAMRIID
B. pseudomallei	Bp668	MSHR668		blood culture from 53-year old male patient with severe melioidosis encephalomyelitis, Darwin Australia	1995	USAMRIID
B. pseudomallei	Bp406e	406e		disseminated melioidosis patient, toe swab, Ubon Ratchathani province, Northeast Thailand	1988	USAMRIID
B. pseudomallei	Bp1106a	1106a		female rice farmer, Northeast Thailand, Sappasithiprasong hospital	1993	USAMRIID
B. pseudomallei	Bp5855	MSHR5855		Australia	2011	USAMRIID

Species	Strain	Alt name	Other designations	Source/Notes	Year	Source
B. pseudomallei	Bp5848	MSHR5848		inhalational melioidosis, Australia	2011	USAMRIID
B. pseudomallei	Bp5858	MSHR5858				USAMRIID
B. pseudomallei	Bp0134a	HBPUB 10134a		sputum, Thailand, Mahidol University	2010	USAMRIID
B. pseudomallei	Bp0303a	HBPUB 10303a		sputum, Thailand, Mahidol University	2011	USAMRIID
B. mallei	BmGB3	GB3	2002734306, 2002734311, strain A, NCTC120	Lister Institute, London	1920	USAMRIID
B. mallei	BmGB4	GB4	M4, 2002734304, strain 6, NCTC10248	human, Ankara, Turkey	1950	USAMRIID
B. mallei	BmISU	ISU		Iowa State University		USAMRIID
B. mallei	BmTurk1	Turkey 1	2000031065, #1 Turkey	Turkey, isolated by Dr. Linda Schlater	2003	USAMRIID
B. mallei	Bm234	KC234	2002721273, 3783	human, Burma- isolated via CA Gleisser Army Medical School	1956	USAMRIID
B. mallei	Bm235	KC235	2002721274	Fort Detrick, Maryland, United States	1956	USAMRIID
B. mallei	BmHI533	HI533	2000031304, 2000031281	human liver abscess drainage, Maryland, United States	2000	USAMRIID
B. mallei	BmGB11	GB11	NCTC 10245, 2002721275, China 5, ATCC10399	horse lung, Southern China	1949	USAMRIID
B. mallei	BmNBL7	NBL 7	China 7	Prep of B mallei China 7 derived from ATCC23344 via passage through several individuals		BEI Resources
B. mallei	BmGB8	GB8 horse 4		derivative of ATCC23344 passaged through horse and isolated from the lung as a single colony, Manitoba, Canada		BEI Resources
B. mallei	Bm23344	ATCC 23344		human post-mortem knee fluid, skin pustules and blood, Burma	1944	BEI Resources
B. mallei	BmTurk2	Turkey 2	T2	Turkey		BEI Resources
B. mallei	BmFMH	FMH		derivative of ATCC23344 passaged through human, laboratory acquired infection- blood	2000	USAMRIID
B. mallei	Bm21280	2002721280	KC1092, 52-236	Iran	1952	BEI Resources
B. mallei	Bm86567	86-567	India86-567-2, 2000031064	mule, East India		BEI Resources
B. mallei	Bm2700C	SR092700C				BEI Resources
B. mallei	BmCh7	China 7	NBL7	preparation produced directly from ATCC 23344	1942	BEI Resources
B. mallei	BmCh5	China 5	MM-A, NBL4	lung and nose of infected horse, Kweiyang, China	1942	BEI Resources
B. mallei	Bm10230	NCTC 10230	strain Ivan	horse with glanders, Hungary	1961	BEI Resources
B. mallei	BmGB8**	GB8 (atcc23344)		Laboratory passage of ATCC 23344 in mouse	1997	USAMRIID
B. thailandensis	BtE264	E264		rice field soil sample, Central Thailand		BEI Resources
B. thailandensis	BtDW503	DW503		Derived from E264; (ΔamrR-oprA) (Kms Gms Sms); rpsL (Smr), Central Thailand		BEI Resources

Species	Strain	Alt. designation	Other designations	Source	Year	Repository
B. thailandensis	BtE421	E421		rice field soil sample from Ubon Ratchathani province, Northeast Thailand	2001	BEI Resources
B. thailandensis	BtE426	E426		rice field soil sample from Ubon Ratchathani province, Northeast Thailand	2001	BEI Resources
B. cepacia	Bc706	DD-706				BEI Resources
B. cepacia	Bc709	DD-709				BEI Resources
B. cepacia	Bc710	DD-710				BEI Resources
B. cenocepacia	Bce6656	LMG16656		sputum of cystic fibrosis patient, Edinburgh, United Kingdom	1989	BEI Resources
B. cenocepacia	BceBC7	BC7		sputum from 15-year old patient with "cepacia syndrome", Canada		Emory U[c]
B. cenocepacia	BceK562	K56-2		less antibiotic resistant derivative of BC7, Canada		Emory U
B. cenocepacia	BceJ2315	J2315		sputum from cystic fibrosis patient, Edinburgh, United Kingdom	1989	Emory U
B. cenocepacia	Bce103a2	DD-707				BEI Resources
B. multivorans	BmvCF2	CF2		sputum from cystic fibrosis patient, NIH Clinical Center		Emory U
B. multivorans	BmvCGD1	CGD1		sputum from chronic granulomatous disease patient, NIH Clinical Center		Emory U
B. multivorans	BmvCGD2	CGD2		blood from chronic granulomatous disease patient, NIH Clinical Center		Emory U
B. multivorans	BmvCF1	CF1		sputum from cystic fibrosis patient, Belgium		Emory U
B. multivorans	Bmv17616	ATCC 17616		soil sample, United States		Emory U
B. multivorans	Bmv13010	LMG13010	CCUG 34080, Lauwers Cepa 002, CIP 105495, DSM 13243, NCTC 13007	sputum of cystic fibrosis patient, Belgium	1992	BEI Resources

[a]USAMRIID; United States Army Medical Research Institute of Infectious Disease.

[b]BEI Resources; NIH Biodefense and Emerging Infections Research Resources Repository, NIAID, NIH.

[c]Emory U; Dr. Joanna Goldberg, Emory University School of Medicine, Atlanta, GA.

CONCLUSIONS

Adhesion and biofilm formation: friend or foe?

The main goal of this work was to investigate the role of the major surface adhesin of GAS, streptococcal collagen-like protein 1 (Scl1), in host tissue attachment, biofilm formation, and virulence. Specifically, our studies focused on the invasive M3-type strains that form poor biofilms *in vitro* and harbor a unique *scl1.3* allele, which encodes a truncated cell-free protein variant. Biofilm formation by bacteria is typically considered disadvantageous to the host because biofilm-embedded bacteria are considerably more resistant to phagocytosis by PMNs and antibiotic treatment (104). However, it has been an emerging concept that adherence and biofilm formation are inversely related to invasiveness and virulence during bacterial infection. Specifically, up to 99.5% of GAS infections are superficial infections of the throat and skin that are associated with localized tissue microcolonies or biofilms, whereas the remaining 0.5% are invasive infections that are not biofilm-associated (105). Our unique observations in biofilm-deficient M3-type strains led us to devise the following hypothesis to explain this phenomenon: **Invasive M3-type GAS, devoid of Scl1 surface adhesin, have reduced adherence to host ECM proteins, including cellular fibronectin (cFn) and laminin (Lm), and reduced biofilm capacity, thus circumventing the formation of stable tissue microcolonies and shifting the balance towards invasive dissemination instead of localized infection.** Our major findings, summarized below in the context of relevant literature, support this model.

Poor biofilm formation and the unique scl1.3 allele of M3-type GAS

Our initial study demonstrated that Scl1 significantly contributes to biofilm formation by GAS on inanimate surfaces (Chapter 1). We showed that deletion of *scl1* in strains of M1, M28 and M41 type resulted in decreased biofilms, with at least 50% reduced average thickness. Heterologous complementation of the biofilm-deficient *Lactococcus lactis* with *scl1* conferred biofilm formation. Although the mechanism of Scl1-mediated biofilm formation was not fully established, Scl1 expression was correlated with increased surface hydrophobicity in these strains. Notably, several invasive M3 wild-type strains had

diminished biofilms in comparison to biofilm-capable M1-, M28-, and M41-type strains. We hypothesized that poor biofilm formation was associated with the predicted lack of surface Scl1, based on the presence of a unique null mutation in the *scl1.3* allele of the M3-strains. The nonsense mutation occurred in the 11th Gly-X-Y repeat of the collagenous region, presumably resulting in a truncated secreted Scl1 variant. This unique *scl1.3* allele was conserved in >98% of invasive M3 isolates (Chapter 3). Our successive studies confirmed the lack of Scl1 expression on the surface of M3 strains by flow cytometry, and in cell wall extracts of M3 strains by western blot. Interestingly, the truncated Scl1.3 variant was not detected in supernatant fraction, likely a result of the diminished expression of the *scl1.3* transcript that was also perceived for these strains in comparison to other M-types (Chapter 2). Altogether, association between the lack of Scl1 expression, accompanied by limited biofilm formation, prompted further investigations into the invasive traits of M3-type strains.

Scl1-mediated biofilm and ECM binding: implications in tissue microcolony formation
Accumulating evidence has supported the importance of Scl1 in mediating the formation of *in vivo* tissue microcolonies, or biofilms, during infection. First, previous data showed that Scl1-mediated biofilms were enhanced on cFn and Lm coatings, as well as on a complex ECM matrix deposited by human fibroblasts (54). Next, Scl1 specifically binds to extra domain A-containing isoforms of cellular fibronectin (EDA/cFn) and basement membrane laminin (Lm), which are both ECM components found in wounded tissue (54,60). Consequently, both binding functions could foster ligand recognition at the pathogen portal of entry, which supports bacterial adherence and formation of stable tissue microcolonies. Here, we used an *in vitro* model of wounded human skin equivalents to compare wound colonization and tissue invasion by GAS. The Scl1-lacking biofilm-poor M3-type GAS invaded wound bed in a form of loose chains and cells, whereas Scl1-expressing biofilm-capable M41-type GAS formed large rounded microcolonies encased in glycocalyx, as imaged by two-photon microscopy (Chapter 2). These results demonstrated striking phenotypic difference in tissue invasion by M3 and M41 GAS.

Full-length surface-attached Scl1.3 supports adherence and biofilm formation

Figure 1. Hypothesis model. Top: Infection of wounded skin with wild-type M3-type GAS. Inherent lack of surface-expressed Scl1.3 causes decreased binding to cFn and Lm expressed in wounded tissue, and reduces biofilm and tissue microcolony formation by M3-type bacteria (red circles). Bottom: *In-trans* complementation of M3-type GAS with full-length cell-associated Scl1.3, Scl1.3FL, restores binding to cFn and Lm in tissue, which confers biofilm and tissue microcolony formation during infection.

To test our hypothesis shown in **Figure 1**, we first had to assess whether full-length Scl1.3 variant binds cFn and Lm, and second, if it restores biofilm formation of M3-type GAS on cFn- and Lm-coated surfaces (Chapter 2). First, recombinant full-length Scl1.3 (rScl1.3FL) protein was generated, following the reversion of the null mutation and restoring the open reading frame, and tested for ECM binding. Positive binding of rScl1.3FL to cFn and Lm was demonstrated by ELISA, surface plasmon resonance, and tryptophan fluorescence binding assays. These results support the concept that Scl1.3FL on the GAS cell surface would behave as an adhesin conferring adherence to EDA/cFn and Lm. Interestingly, rotary shadowing and electron microscopy of rScl1.3 proteins showed aggregates mediated by V-to-V domain interactions that could stabilize biofilm structure.

A second set of experiments was designed to test the effect of Scl1.3FL surface expression on GAS biofilm formation. Homologous complementation of two invasive M3-type strains with Scl1.3FL conferred biofilm formation on cFn- and Lm-coated surfaces. Additionally, heterologous complementation of a Δ*scl1* mutant of M41-type strain with Scl1.3FL restored biofilm formation to the wild-type level. Biofilms formed by the complemented strains were at least two-fold thicker on average, compared to vector-complemented controls. Overall, these results reinforce our hypothesis that M3-type
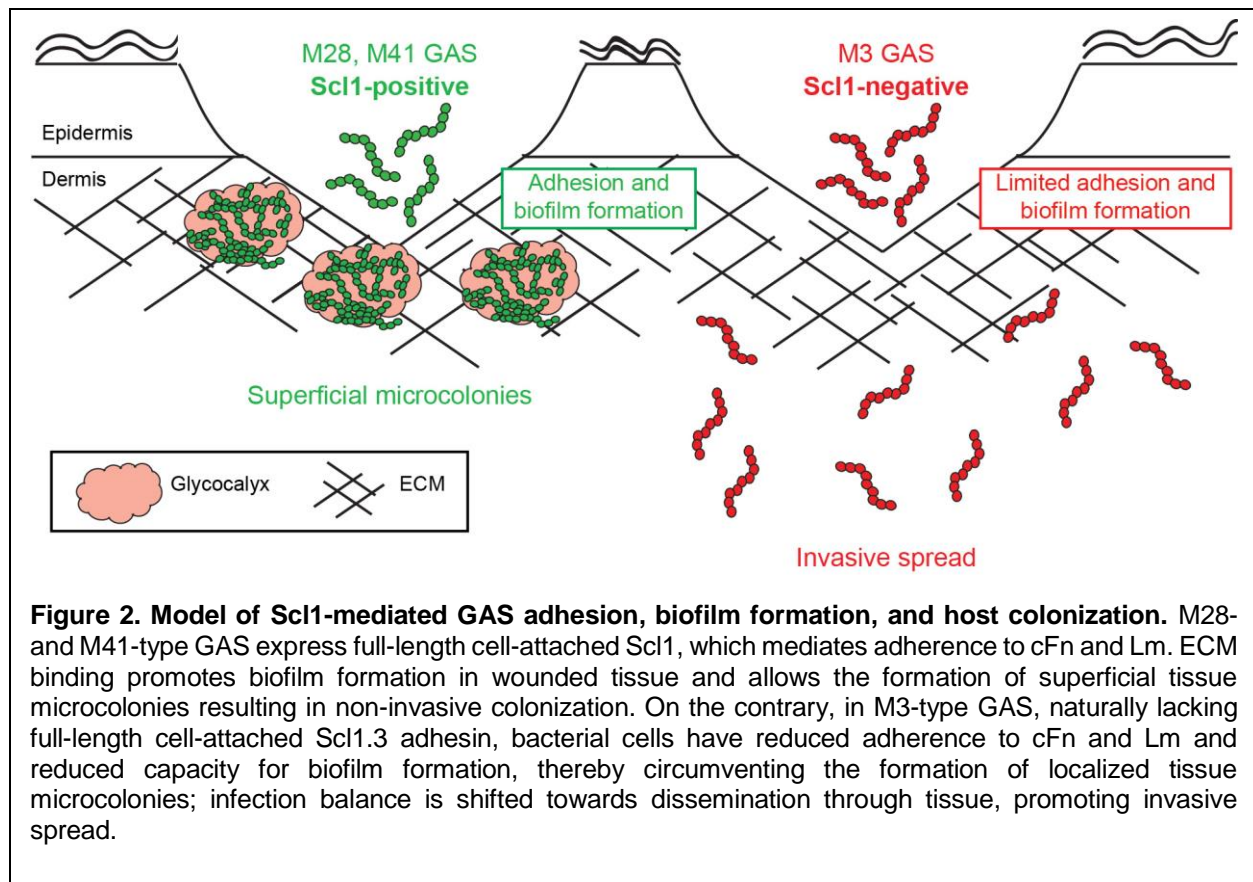
strains have reduced adherence and microcolony formation due to a natural lack of surface Scl1 with intrinsic capacity for ECM binding and biofilm support (Figure 1).

Inactivation of Scl1 in GAS results in hypervirulent phenotype

If our hypothesis is correct, the inactivation of *scl1* in biofilm-capable GAS strains should result in increased virulence during infection. We validated this concept by demonstrating that *scl1*-inactivated isogenic mutants of the M28- and M41-type strains produced larger skin lesions during subcutaneous infection of mice than the wild-type parental strains (Chapter 2). A similar phenotype was displayed by Scl1-deficient M3-type GAS in comparison to M3 strain containing a rare cell-attached Scl1 variant. While the majority of M3-type strains harbor the *scl1.3* allele containing the null mutation, ~1.3% of M3 strains harbor a "*scl1.3* carrier allele" containing an in-frame deletion encompassing the null mutation, and thus, express a cell-attached variant. An isoallelic replacement strain, harboring the carrier allele, formed a limited abscess compared to the invasive parental M3 strain (Chapter 3). The carrier strain also exhibited increased adherence to human epithelial cells.

Adherence and biofilm formation: an inverse relationship with invasiveness

Overall, our results support the hypothesis that M3-type GAS, naturally lacking surface Scl1, are reduced in adherence and microcolony formation, thus resulting in bacterial dissemination and invasive infection. **Figure 2** depicts a model for Scl1-mediated adhesion and microcolony formation during infection, comparing M28- and M41-type GAS expressing Scl1 to M3-type GAS lacking Scl1.

Figure 2. Model of Scl1-mediated GAS adhesion, biofilm formation, and host colonization. M28- and M41-type GAS express full-length cell-attached Scl1, which mediates adherence to cFn and Lm. ECM binding promotes biofilm formation in wounded tissue and allows the formation of superficial tissue microcolonies resulting in non-invasive colonization. On the contrary, in M3-type GAS, naturally lacking full-length cell-attached Scl1.3 adhesin, bacterial cells have reduced adherence to cFn and Lm and reduced capacity for biofilm formation, thereby circumventing the formation of localized tissue microcolonies; infection balance is shifted towards dissemination through tissue, promoting invasive spread.

The main hypothesis presented in my thesis, which postulates an inverse relationship between strain invasiveness and its capacity to form biofilm, has gained recent support from several studies. *Streptococcus pneumoniae* grown in biofilm was less able to disseminate to the blood in mouse models of intranasal and intratracheal infection (106). Biofilm-grown bacteria also had increased adherence to host epithelial cells, as compared to bacteria grown in culture (106). Another study showed that biofilm-deficient mutants of *S. pneumoniae* elicited greater cytokine responses in a mouse model of intranasal infection colonization, and had decreased adherence and increased invasion of epithelial cells *in vitro* (107). Destabilization of *S. pneumoniae* biofilm and subsequent dissemination was triggered by co-infection with influenza A virus *in vivo*, as well as febrile temperatures and nutrient availability in vitro. The dispersed bacteria had altered gene expression profile and enhanced dissemination in mice (108). Similarly, GAS cells that formed biofilm *in vitro* on epithelial cells had downregulated expression of virulence genes, including streptolysins, *emm*, and hyaluronic acid capsule gene *hasA*, and showed enhanced colonization of nasal-associated lymphoid tissue, while dissemination to distant

245

organs was decreased (109). Likewise, the cleavage of bacterial surface adhesins by the potent GAS protease SpeB, resulted in dispersal of biofilm, which was associated with increased lesion sizes in mice (110).

Scl variants of GAS- a shared structure with diverse ligand binding properties

The majority of our work has focused on adherence and biofilm conferred by the Scl1 protein, functions that are lacking for the Scl2 protein; however both Scl proteins are ubiquitous among GAS strains and share a similar overall structure. Early studies based on recombinant Scl proteins of GAS showed they harbor a conserved lollipop-like domain organization with a globular variable domain and stalk-like collagenous domain (20,21). Additionally, multiple sequence alignment and secondary structure analyses of the Scl1- and Scl2-variable regions predicted two conserved α-helices interspaced by the hypervariable sequence (51). A major contribution that came from current work was the report of the crystal structure for the V domain of the Scl2 protein of M3-type GAS, which provided insights into possible ligand binding sites and function of the Scl proteins (Chapters 4 and 5). The Scl-globular (homotrimeric) domain folds into unique six helical bundle structure, which is predicted to be conserved across all Scl1 and Scl2 variants. The two antiparallel α-helices that are conserved in all variants form a structural core for the exposed variable loops that are likely involved in ligand binding. The V domain was also found to be structurally similar to the gp41 subunit of the envelope glycoprotein of HIV-1, a membrane fusion protein involved in viral entry into CD4+ T cells, although the two proteins had low sequence identity. The six-helix bundle structure is largely absent in bacteria, with the exception of the cholix toxin which was reported to contain a bundle of six helices within the translocation domain (111). Membrane fusion and entry or translocation across the membrane seems to be a function offered by this structural element. Since internalization of GAS by host cells has been reported (112), it would be interesting if the V domain of Scl also mediates an interaction with host cell membranes to accomplish host cell entry.

SUPPLEMENTARY MATERIAL FOR CHAPTER 1

THE STREPTOCOCCAL COLLAGEN-LIKE PROTEIN-1 (SCL1) IS A SIGNIFICANT DETERMINANT FOR BIOFILM FORMATION BY GROUP A *STREPTOCOCCUS*

Heaven A Oliver-Kozup, Meenal Elliott, Beth A Bachert, Karen H Martin, Sean D Reid, Diane E Schwegler-Berry, Brett J Green, and Slawomir Lukomski

Figure S1

Figure S1 (a). Confocal laser scanning microscopy (CLSM) of GFP-expressing M41 WT GAS biofilm at 24 h. Panels represent a gallery view of consecutive images taken at 1 micrometer increments. Panel shown in lower right corner represents an X-Y orthogonal Z-stack view (Fig. 4d). Thickness is indicated in micrometers.

Figure S1 (b). Confocal laser scanning microscopy (CLSM) of GFP-expressing M41Δ*scl1* GAS biofilm at 24 h. Panels represent a gallery view of consecutive images taken at 0.75 micrometer increments. Panel shown in lower right corner represents an X-Y orthogonal Z-stack view (Fig. 4d). Thickness is indicated in micrometers.
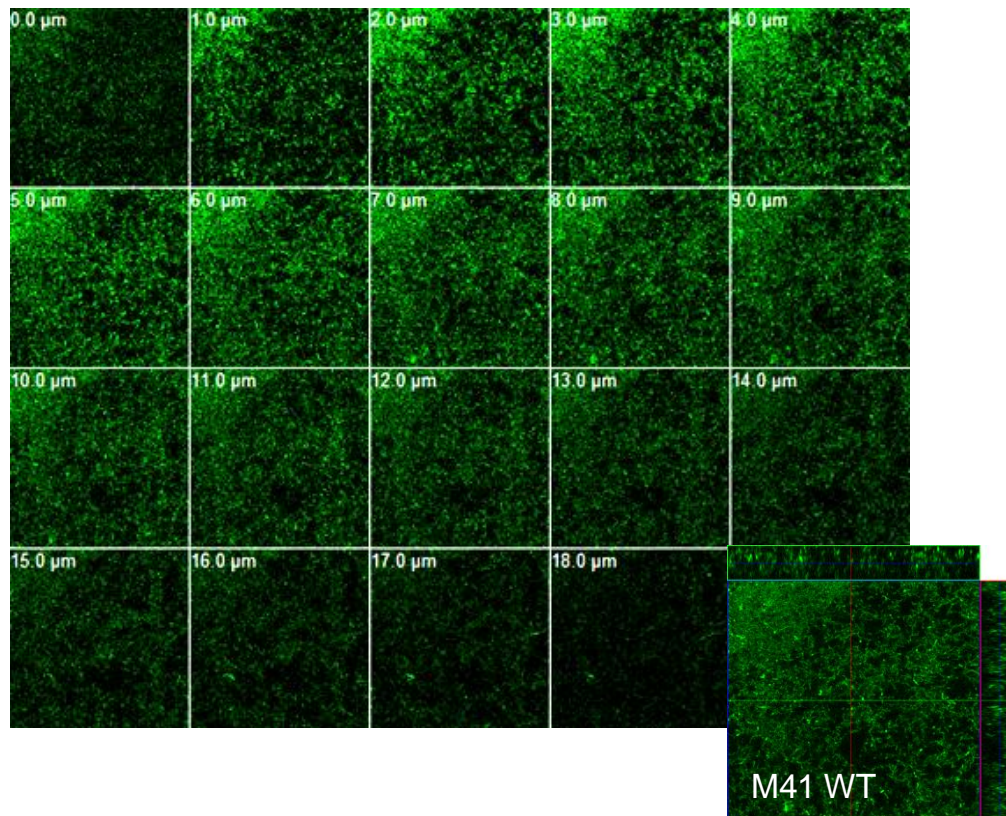
Figure S1 (c). Confocal laser scanning microscopy (CLSM) of GFP-expressing M28 WT GAS biofilm at 24 h. Panels represent a gallery view of consecutive images taken at 1 micrometer increments. Panel shown in lower right corner represents an X-Y orthogonal Z-stack view (Fig. 4e). Thickness is indicated in micrometers.

Figure S1 (d). Confocal laser scanning microscopy (CLSM) of GFP-expressing M28Δ*scl1* GAS biofilm at 24 h. Panels represent a gallery view of consecutive images taken at 1 micrometer increments. Panel shown in lower right corner represents an X-Y orthogonal Z-stack view (Fig. 4e). Thickness is indicated in micrometers.
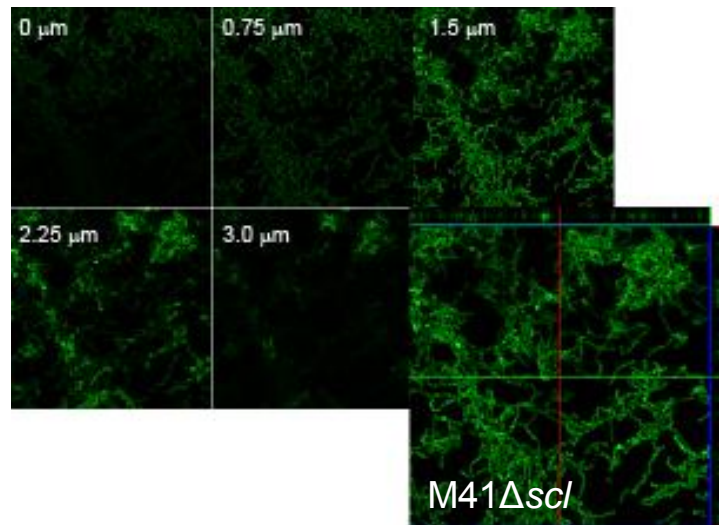
Figure S1 (e). Confocal laser scanning microscopy (CLSM) of GFP-expressing M1 WT GAS biofilm at 24 h. Panels represent a gallery view of consecutive images taken at 1 micrometer increments. Panel shown in lower right corner represents an X-Y orthogonal Z-stack view (Fig. 4f). Thickness is indicated in micrometers.
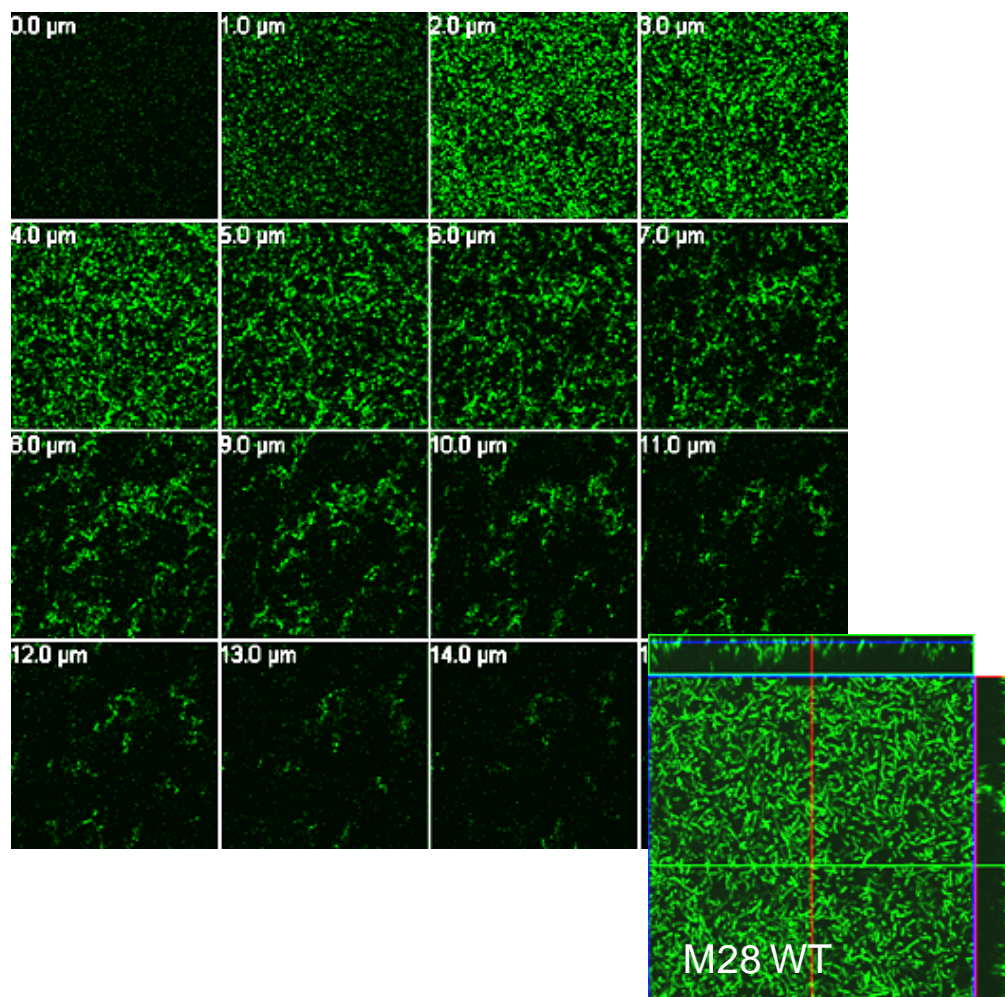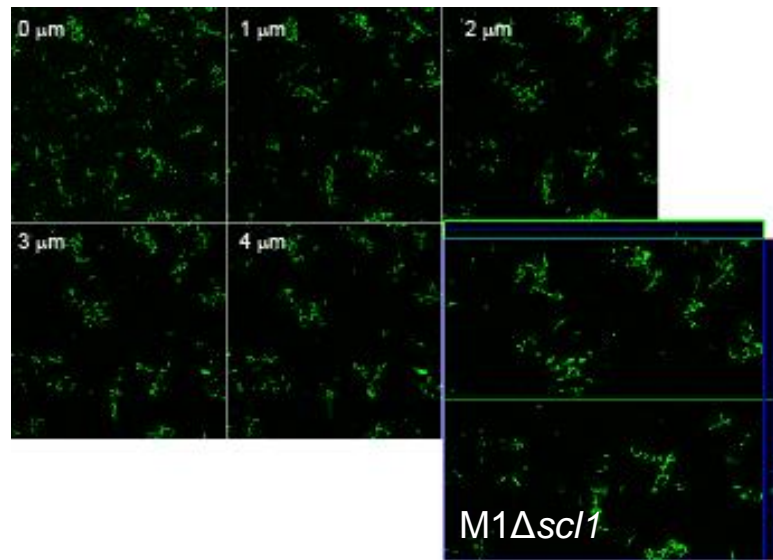
Figure S1 (d). Confocal laser scanning microscopy (CLSM) of GFP-expressing M1Δ*scl1* GAS biofilm at 24 h. Panels represent a gallery view of consecutive images taken at 1 micrometer increments. Panel shown in lower right corner represents an X-Y orthogonal Z-stack view (Fig. 4f). Thickness is indicated in micrometers.
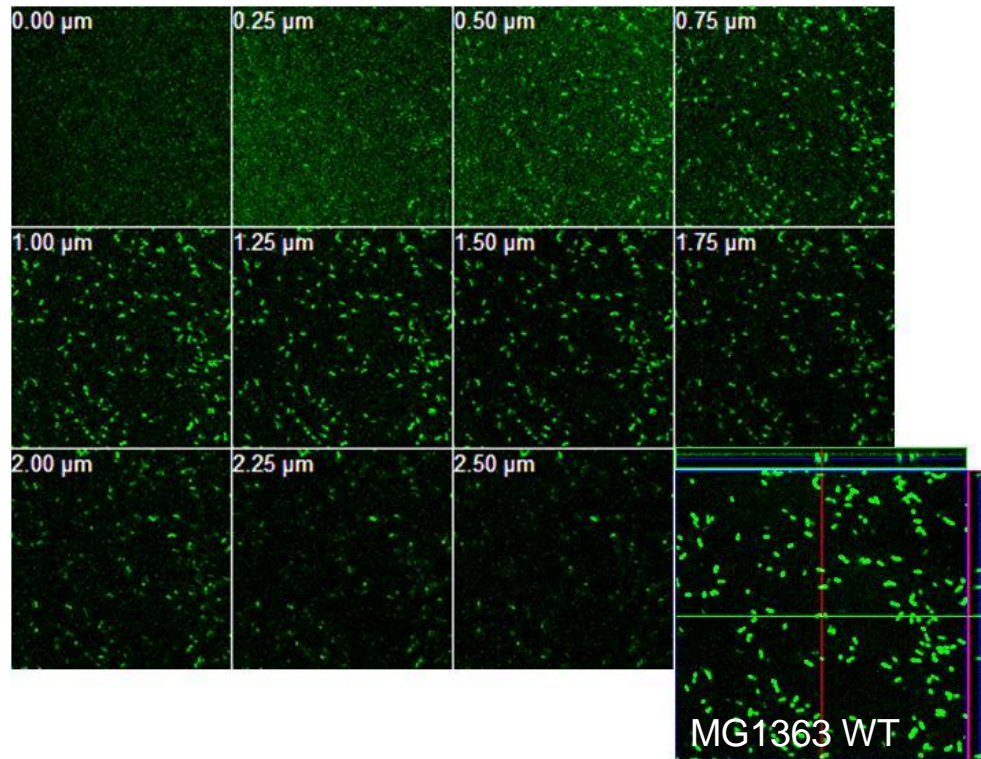
Figure S2

Figure S2 (a). Confocal laser scanning microscopy (CLSM) of GFP-expressing *Lactococcus lactis* MG1363 WT biofilm at 24 h. Panels represent a gallery view of consecutive images taken at 0.25 micrometer increments. Panel shown in lower right corner represents an X-Y orthogonal Z-stack view (Fig. 5e). Thickness is indicated in micrometers.
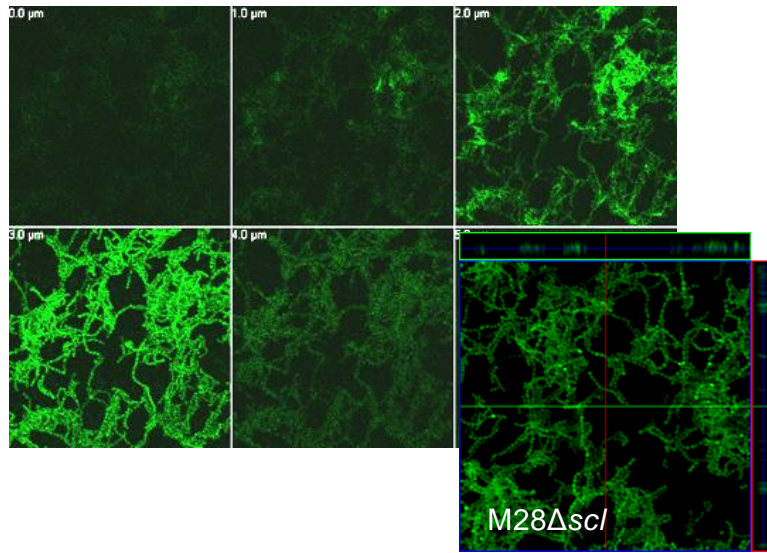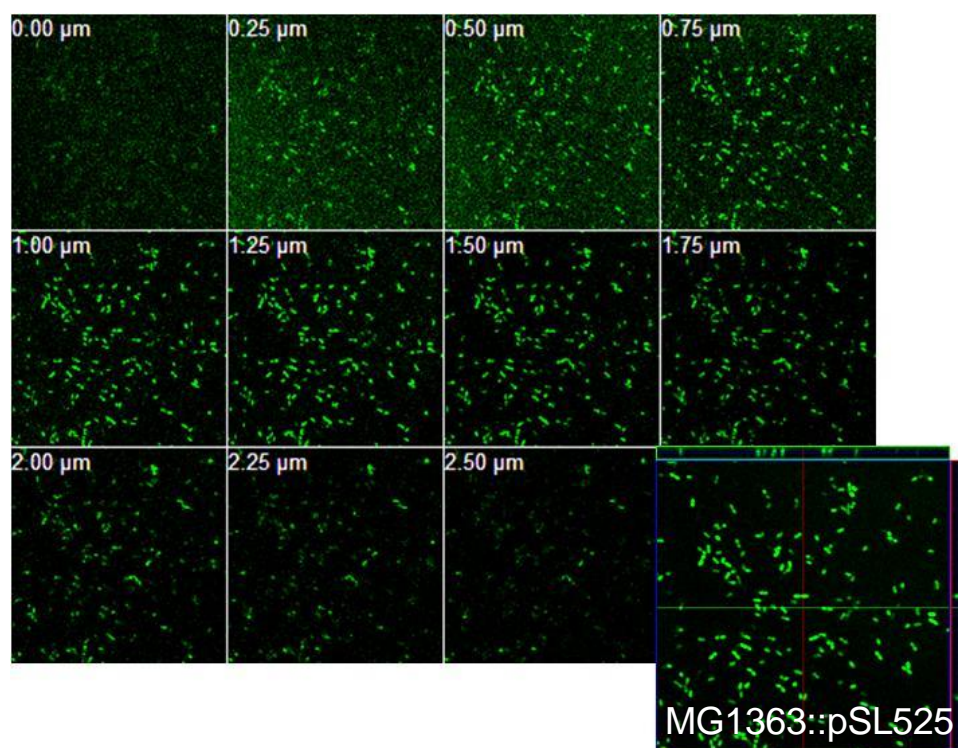
Figure S2 (b). Confocal laser scanning microscopy (CLSM) of GFP-expressing *Lactococcus lactis* biofilm at 24 h. *L. lactis* was transformed with the shuttle vector pJRS525 (MG1363::pJRS525). Panels represent a gallery view of consecutive images taken at 0.25 micrometer increments. Panel shown in lower right corner represents an X-Y orthogonal Z-stack view (Fig. 5e). Thickness is indicated in micrometers.
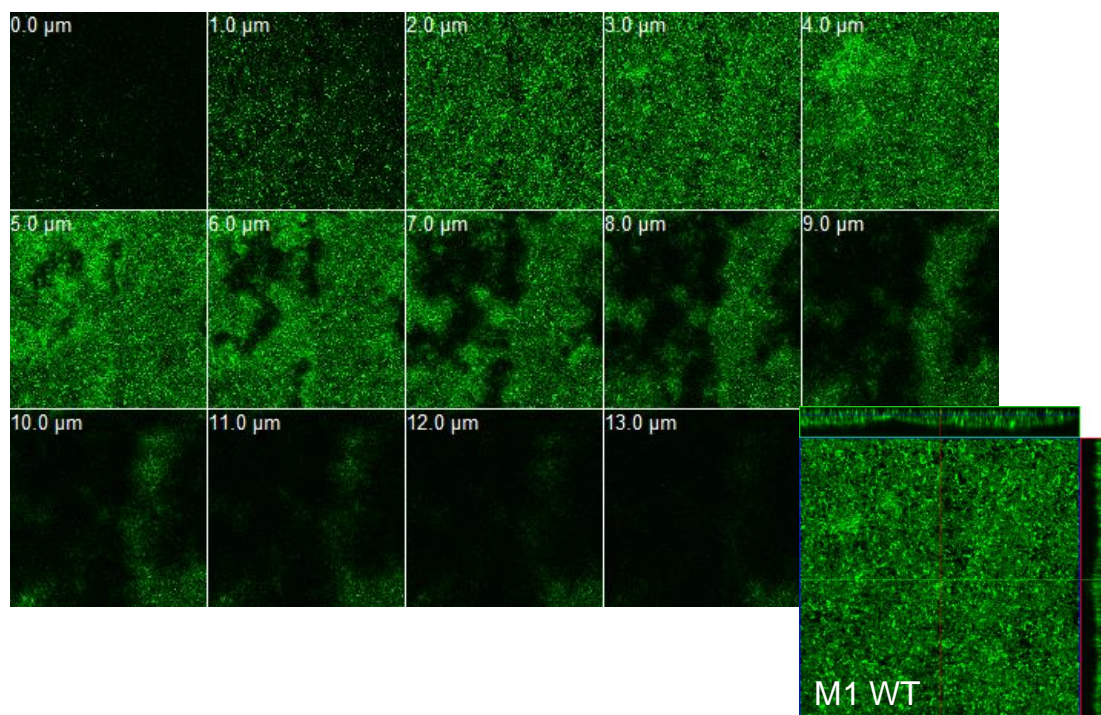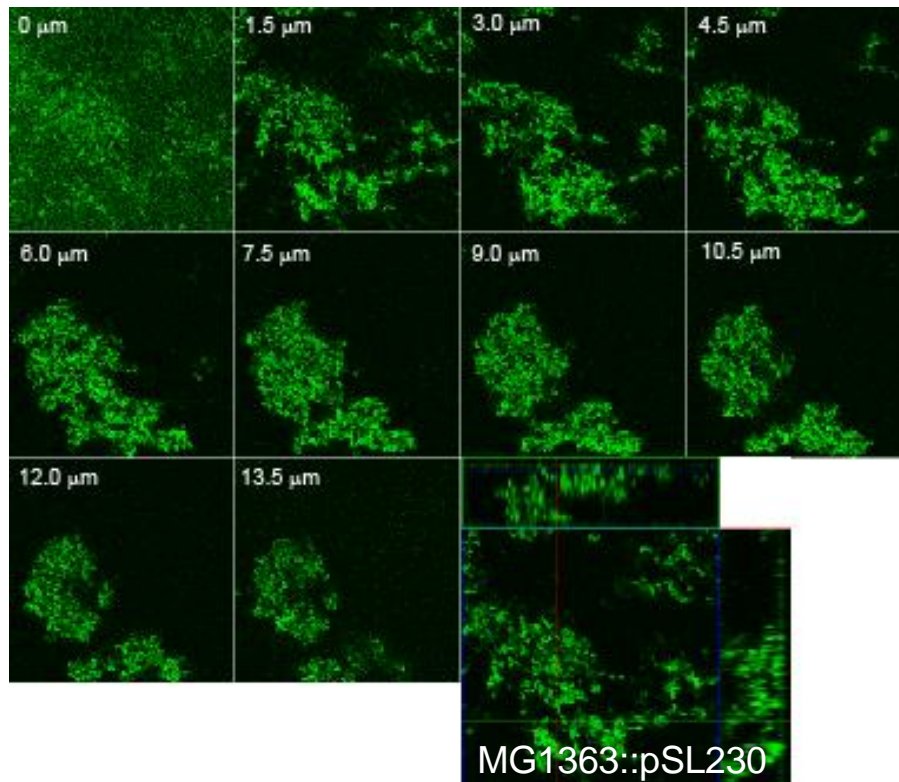
Figure S2 (c). Confocal laser scanning microscopy (CLSM) of GFP-expressing *Lactococcus lactis* expressing Scl1.41 protein (MG1363::pSL230) biofilm at 24 h. *L. lactis* was transformed with the plasmid construct pSL230 encoding Scl1.41 protein (MG1363::pSL230). Panels represent a gallery view of consecutive images taken at 0.5 micrometer increments. Panel shown in lower right corner represents an X-Y orthogonal Z-stack view (Fig. 5e). Thickness is indicated in micrometers.

SUPPLEMENTARY MATERIAL FOR CHAPTER 2

UNIQUE FOOTPRINT IN THE *SCL1.3* LOCUS AFFECTS ADHESION AND BIOFILM FORMATION OF THE INVASIVE M3-TYPE GROUP A *STREPTOCOCCUS*

Beth A. Bachert, Soo Jeon Choi, Paul R. LaSala, Tiffany Harper, Dudley H. McNitt, Dylan T. Boehm, Clayton C. Caswell, Pawel Ciborowski, Douglas R. Keene, Anthony R. Flores, James M. Musser, Flavia Squeglia, Daniela Marasco, Rita Berisio, and Slawomir Lukomski
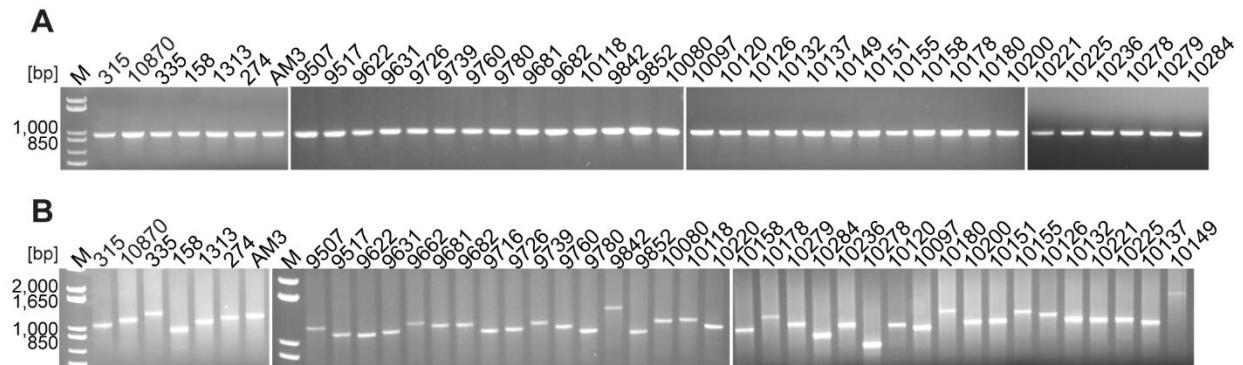
Figure S1. Analytical PCR of *scl1.3* and *scl2.3* in M3-type GAS. (A) PCR of *scl1.3* in 42 M3-type strains was performed with primers 232Up and 232Rev flanking the *scl1* locus (**Table S1**). Expected size based on MGAS315: 1,010 bp. M, 1 kb Plus DNA Ladder. **(B)** PCR of *scl2.3* in 42 M3-type strains was performed with primers Scl2.3 F and Scl2.3 R flanking the *scl2.3* collagen-like region (**Table S1**). Expected size based on MGAS315 genome is 1,125 bp. M, 1 kb Plus DNA Ladder. MGAS designation applies to all strain numbers shown above gel wells, with the exception of strain AM3.
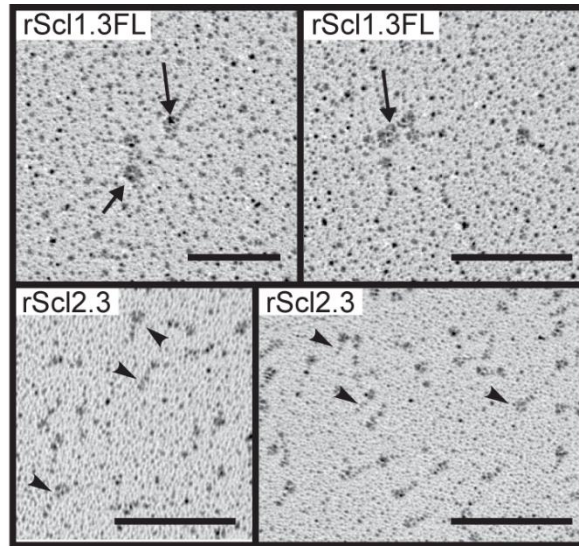
Figure S2. Domain organization of rScl1.3FL and rScl2.3 proteins. Rotary shadowing of rScl proteins demonstrates the conserved lollipop-like domain organization. Aggregates mediated by the interactions between V regions of rScl1.3FL are seen in the top panels (depicted by arrows). No apparent aggregation was observed in the rScl2.3 preparation (single lollipops are depicted by arrowheads). All scale bars reflect 100 nm.
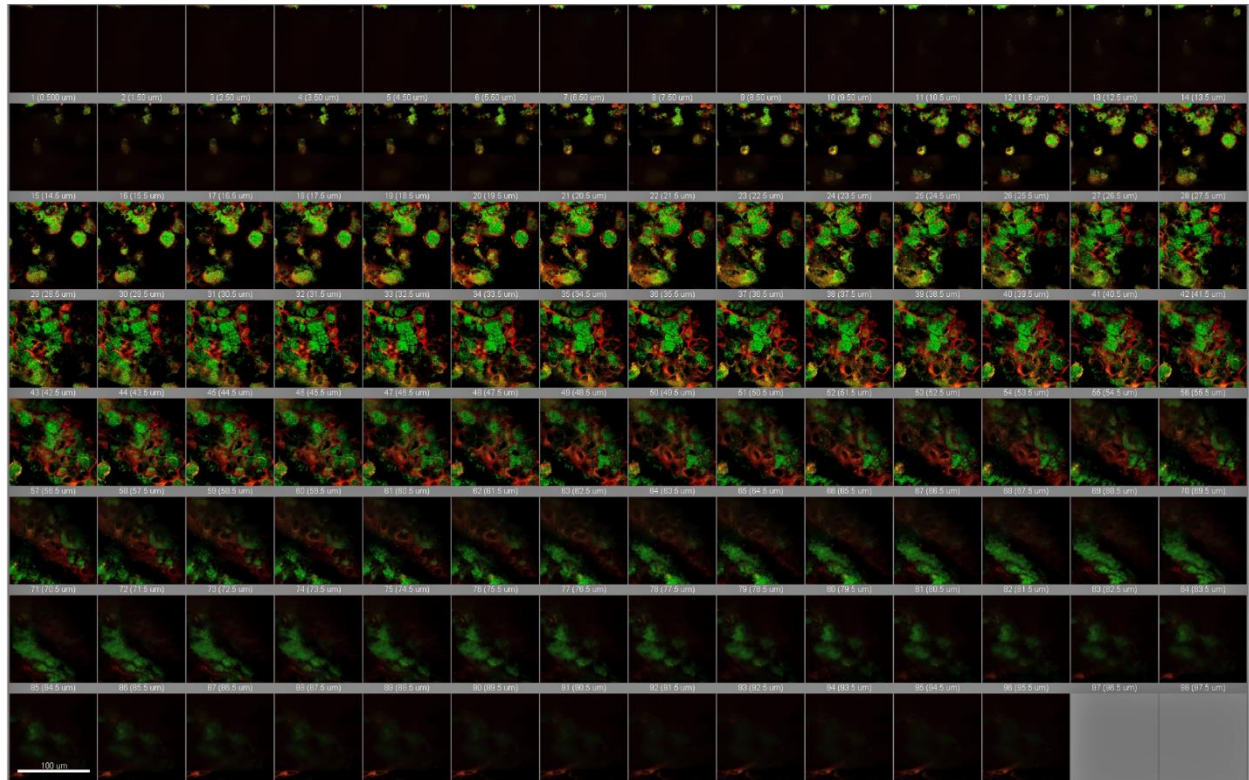
Figure S3. Colocalization of glycocalyx staining with GFP-expressing M41-type GAS during *in vitro* infection of skin equivalent. Representative TPF gallery image of z-stack layers from 3D projection (**Figure 5** of manuscript), showing colocalization of TRITC-concanavalin A staining of glycocalyx (red) with GFP-expressing GAS microcolonies (green). Z-stack step size: 1 μm; 600x magnification.
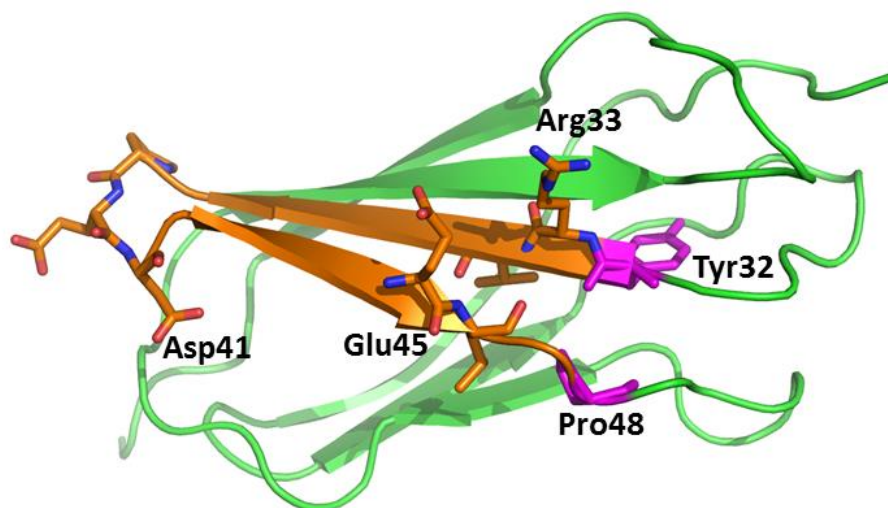
Figure S4. Strategy for C-C' loop peptide design. Cartoon model of the EDA domain of cFn (PDB code 1J8K) used for the C-C' peptide design. The designed peptide includes the C-C' beta-hairpin (orange). The terminal residues of the beta-hairpin (magenta) were mutated to cysteine to induce a disulfide bond and mimic the conformation of the peptide in the protein. The key residue for recognition, Asp41, is reported in ball-and-stick along with the residues Arg33 and Glu45, which form a salt bridge, and the terminal residues mutated to cysteine.

Table S1. Primers used in this study

Primer name	Sequence (5'-3')	Use
Analytical PCR		
IS1548F	GCCGTCTGCGTGCCCATTGCGTCTA	Detection of
Scl1 R	ACTAGATCTGAGATTATGGTGCTTTGATGTC	IS1548
232 Up	CTCCACAAAGAGTGATCAGTC	Amplificatio
232 Rev	TTAGTTGTTTTCTTTGCGTTT	n of *scl1.3*
Scl2.3 F	AGGCATACAAGATCATGTCCTTGA	Amplificatio
Scl2.3 R	TTTGGTGTATGTGGTGCGGT	n of *scl2.3*
Scl Up	CTTTCAATGGATGACGATACC	Amplificatio
Scl Rev	ACTTTCCATCAGTTAGGTAGC	n of *scl2.3*
Cloning		
Scl1.3 M3VF	GAGATGGCCGAGACTCCTATGACATCAAAGG	Cloning of
Scl1.3 M3VR	CAGCGTCTCAGCGCTCTTTGTTGCACCTTTTTCAATCAG	*scl1.3V* region
232 Up	CTCCACAAAGAGTGATCAGTC	Cloning of
ME7	TCAGTGAATTCTCTTTAGAGGATTAG	*scl1.3WT*
pJRS525F1	GGGTTTTCCCAGTCACG	Repair of
ME6ᵃ	TCCAGCAGGACCTCGAGGT**GAA**CGC	*scl1.3* null
Scl1.3 M3VF	GAGATGGCCGAGACTCCTATGACATCAAAGG	Cloning of
Scl1.28WMR	GTCAAGCTTATTATTTTTCGAACTGCGGGTGGCTCCAAGGTTTTTCTGGAGCTGGAGTTACC	*scl1.3FL*
Scl1.3 M3VF ext	GAGATGGCCGAGACTCCTATGACATCAAAGGAGAGACAAG	Cloning of *scl1.3WT*
Scl1.3 truncR2	GGTCTCAGCGCTACCTCGAGGTCCTGCTGGACCTTG	
Scl2 M3VF	GAGATGGCCGATGGTGAAGATGCCCAAAAAAG	Cloning of
Scl2.28_WMR	GTCAAGCTTATTATTTTTCGAACTGCGGGTGGCTCCAT	*scl2.3*
qRT-PCR analysis		
tufA_F	CAACTCGTCACTATGCGCACAT	qRT-PCR of
tufA_R	GAGCGGCACCAGTGATCAT	*tufA*
Scl1_WMR_exp_F	TGCTGACAAAGAAGCTAACCAAAC	qRT-PCR of
Scl1_WMR_exp_R	GTGGTTGTTGGCTACAGGTGTCT	*scl1*
Scl2_WMR_exp_F	TCCTAAAACACCAGAGGTCC	qRT-PCR of
Scl2_WMR_exp_R	TGTGTGTGTCGTGAGCTGC	*scl2*
emm3_exp_F2	AACAAATCTCAGACGCAAGCCGTC	qRT-PCR of
emm3_exp_R2	TTCAAGCTCTTTGTTWAGTTTTTCAAG	*emm*
Mga_TMF	CAAGTCAACAGTGGAGAGAACTAAAATT	qRT-PCR of
Mga_TMR	ATGGAGATGTTGAGAGCTTTGCT	*mga*
Mga_F1	ATCAGACAAAAACATTAAATTGCATG	Sequencing
Mga_R1	TTGCATGTTAGTGAGACAAGTTTGC	of *mga*

ᵃ Bolded GAA in primer ME6 indicates codon for glutamine in place of the original TAA stop codon.

Table S2. Variation in *scl2.3* gene among M3 strains

Strain	No. of GXY repeats	No. of bp in the CL region	Amplicon size (bp)[a]	No. of CAAAA repeats	GTG in frame[b]
MGAS274	~125	~1125	~1206	8	yes
MGAS315	116	1044	1125	5	yes
MGAS335	~135	~1215	~1296	4	no
MGAS1313	~119	~1071	~1152	8	yes
AM3	~129	~1161	~1242	10	no
MGAS3375	17	153	234	5	yes
MGAS9517	113	1017	1098	8	yes
MGAS9622	111	999	1080	11	yes
MGAS9631	113	1017	1098	18	no
MGAS9716	110	990	1071	14	yes
MGAS9726	111	999	1080	11	yes
MGAS9739	119	1071	1152	5	yes
MGAS9760	113	1017	1098	11	yes
MGAS9780	106	954	1035	14	yes
MGAS9842	139	1251	1332	9	no
MGAS9852	103	927	1008	8	yes
MGAS10080	117	1053	1134	5	yes
MGAS10118	119	1071	1152	5	yes
MGAS10220	110	990	1071	12	no

[a] Amplicons were generated using primers Scl2.3 F/ R (**Table S1**)
[b] GTG is *scl2* start codon

Table S3. Mass spectrometry identification of rScl and native Scl proteins[a]

Sample	No. of peptides >95% confident	Peptide	Confidence	Sequence coverage	Confident ID
rScl1.3V	4	EENSQEELK	99%	40%	yes
		EENSQEELKNFTEER	99%		
		LKEILDLIEK	99%		
		WYGTYFKEENSQEELK	99%		
rScl2.3	2	GIQDHVLDGQDGDR	99%	18%	yes
		EELLSALIDGTSR	98.13%		

MGAS315 WT

Sample	No. of peptides >95% confident	Peptide	Confidence	Sequence coverage	Confident ID
Scl2.3 in Sup	12	DVTPAPQNPSN	99%	17%	yes
		DVTPAPQNPSNR	99%		
		EELLSALIDGTSR	99%		
		GEAGPAGPR	99%		
		GIQDHVLDGQDGDR	99%		
		GIQDHVLDGQDGDRGEAGPAGPR	99%		
		GLNKPQTQGGNQL	99%		
		GLNKPQTQGGNQLAK	99%		
		NKPQTQGGNQLAK	99%		
		REELLSALIDGTSR	99%		
		TPEVPQKPDTAPHTPK	99%		
		TPQIPGQSK	99%		
Scl2.3 in CW	7	DVTPAPQNPSNR	99%	15%	yes
		EELLSALIDGTSR	99%		
		GIQDHVLDGQDGDR	99%		
		GIQDHVLDGQDGDRGEAGPAGPR	99%		
		GLNKPQTQGGNQLAK	99%		
		REELLSALIDGTSR	99%		
		TPEVPQKPDTAPHTPK	99%		

Scl1.3FL- complemented GAS

Sample	No. of peptides >95% confident	Peptide	Confidence	Sequence coverage	Confident ID
315WT :: *scl1.3FL*	5	EENSQEELKNFTEER	99%	17%	yes
		EILDLIEK	99%		
		GDKGETGLAGPVGPAGK	99%		
		GETGLAGPVGPAGK	99%		
		LKEILDLIEK	99%		
10870Δ*scl1* :: *scl1.3FL*	4	EILDLIEK	99%	11%	yes
		GDKGETGLAGPVGPAGK	99%		
		GETGLAGPVGPAGK	99%		
		LKEILDLIEK	99%		
M41Δ*scl1* :: *scl1.3FL*	3	EENSQEELKNFTEER	99%	13%	yes
		GDKGETGLAGPVGPAGK	99%		
		GETGLAGPVGPAGK	99%		

[a] Protein identification is confident by the presence of at least two distinct peptides with >95% confidence each. Sup, supernatant protein fractions; CW, cell-wall association protein fractions.

Table S4. Summary table of fold-changes in scl1 and scl2 expression[a]

Strain	Gene	Fold change	+SE	- SE	P value	Significance
M1	*scl1*	21.05195084	-2.33583	-2.62734	0.001184755	**
	scl2	-23.6920597	-1.97551	-1.82347	0.01549836	*
M28	*scl1*	6.237987683	-1.1098	-1.34997	0.01395444	*
	scl2	-7.25032516	0.960352	1.106978	0.034822349	*
M41	*scl1*	7.787441864	-0.82018	-0.91673	0.004897482	**
	scl2	-3.22615419	0.376259	0.425935	0.141911016	N.S.
M3 MGAS:						
10870	*scl1*	-25.2160816	1.366976	1.445328	5.48581E-06	***
	scl2	-1.84753589	-0.01174	-0.01166	0.077646106	N.S.
158	*scl1*	-45.8489636	-0.86051	-0.84465	0.000121468	***
	scl2	1.077129939	0.059307	0.056212	0.753558444	N.S.
335	*scl1*	-42.9309881	-4.03211	-3.68593	5.18378E-05	***
	scl2	-2.44408041	-1.65928	-0.98832	0.283398851	N.S.
1313	*scl1*	-43.929358	0.938527	0.959016	5.45812E-06	***
	scl2	-3.89390614	-2.38312	-1.47835	0.117229672	N.S.

[a]Fold-changes in *scl1* and *scl2* expression relative to M3 MGAS315 are shown for each strain. Data is based on qRT-PCR analysis of RNA isolated during exponential growth phase and expression is normalized to the expression of *tufA* gene. +SE, positive standard error; -SE, negative standard error. Significance is calculated based on three independent experiments, each performed in triplicate wells. *P<0.05, **P<0.01, ***P<0.001; student's *t*-test.

SUPPLEMENTARY MATERIAL FOR CHAPTER 3

NATURAL VARIANT OF COLLAGEN-LIKE PROTEIN A IN SEROTYPE M3 GROUP A
***STREPTOCOCCUS* INCREASES ADHERENCE AND DECREASES INVASIVE**
POTENTIAL

Anthony R. Flores, Brittany E. Jewell, Erika M. Versalovic, Randall J. Olsen, Beth A. Bachert, Slawomir Lukomski and James M. Musser

Table S1. Strains and plasmids used in this study

Strain	Description	Source
MGAS315	Serotype M3 invasive strain	(1)
MGAS10870	Serotype M3 invasive strain	(2)
MGAS23440	Serotype M3 invasive strain	(3)
MGAS10870*sclA*^{Carrier}	Isoallelic mutant with *sclA*^{Carrier}	This study
MGAS10870Δ*sclA*	Isogenic deletion mutant lacking *sclA*	This study
MGAS10870Δ*sclA*::pDC*sclA*^{Invasive}	Isogenic deletion mutant complemented *in trans* with pDC*sclA*^{Invasive}	This study
MGAS10870Δ*sclA*::pDC*sclA*^{Carrier}	Isogenic deletion mutant complemented *in trans* with pDC*sclA*^{Carrier}	This study
Plasmid	**Description**	**Source**
pDC123	*E. coli*/*S. pyogenes* shuttle vector used for trans-complementation and expression	(4)
pJL1055	*E. coli*/*S. pyogenes* shuttle vector used for allelic exchange in GAS	
pCR2.1	*E. coli* cloning vector	Invitrogen
pET15b	*E. coli* expression vector	Novagen
pDC*sclA*^{Invasive}	pDC123 shuttle vector with *sclA*^{Invasive} allele and native promoter from MGAS10870	This study
pDC*sclA*^{Carrier}	pDC123 shuttle vector with *sclA*^{Carrier} allele and native promoter from carrier strain MGAS23440	This study
pJSF41	pJL1055 with *sclA*^{Carrier} from MGAS23431	This study
pJSF66	pET15b containing *sclA*^{Invasive} from MGAS10870 used for overexpression and subsequent antibody generation	This study
pJRS525	*E. coli*/ *S. pyogenes* shuttle vector used for cloning and repair of *sclA*^{Invasive} allele from MGAS315 to generate *sclA*$^{M3-FL}$	(5)
pSL501	pJRS525 with *sclA*^{Invasive} from MGAS315	This study
pSL502	pSL501 with repaired allele *sclA*$^{M3-FL}$ generated by PCR mutagenesis	This study
pASK-IBA2	*E. coli* expression vector used for generation of rSclA^{M3-FL}	IBA, Göttingen
pSL503	pASK-IBA2 with *sclA*$^{M3-FL}$	This study

Table S2. Primers used in this study

Primer	Use	Sequence (5'-3')
2902F	3' end of *sclA* used to amplify *sclA* from invasive and carrier strains	ATTTTTGCCATTTCGTTCCT
2487R	5' end of *sclA* upstream of promoter used to amplify *sclA* from invasive and carrier strains	CGAATTTTCCAAGATTGACGA
2901F	Downstream of *sclA* used to generate in-frame allelic replacement of *sclA* with *aad9*	GAAATGCGCTCTTGTTTGTC
MSP166	Forward primer used for PCR amplification and sequencing of *sclA* in invasive GAS	TCTTTTGGGATCTCTCAGGC
MSP167	Reverse primer used for PCR amplification and sequencing of *sclA* in invasive GAS	TAAATAATCAGGTCTAGCTACC
MSP183	Used with 2901F to generate 3'-flanking region overlapping with *aad9*	CTATTTAAATAACAGATTAAAAAAATTATAACGCAAAGAAAACAACTAATCCTCTAAATTGAG
MSP181	Used with MSP182 to amplify *aad9*; contains 18-bp at end of *sclA* ORF	CTCAATTTAGAGGATTAGTTGTTTTCTTTGCGTTATAATTTTTTTAATCTGTTATTTAAATAG
MSP182	Used with MSP181 to amplify *aad9*; contains 18-bp at beginning of *sclA* ORF	GAAAGAGAGAACAACATATGTTGACATCAAAGCACAATACATGTTATAATAACTATAAC
MSP184	Used with primer 0488R to generate 5'-flanking region overlapping with *aad9* for in-frame deletion of *sclA*	GTTATAGTTATTATAACATGTATTGTGCTTTGATGTCAACATATGTTGTTCTCTCTTTC
0488R	Upstream of *sclA* used to generate in-frame allelic replacement of *sclA* with *aad9*	TATGAACAGGCTTCTGATTT
MSP204	*sclA* forward TaqMan primer	TGAAAAGGTGCAACAAAGGGCGATA
MSP205	*sclA* reverse TaqMan primer	GTCCTGCTGGACCTTGTGCG
MSP206	Probe for *sclA* TaqMan transcript analysis. Contains 5'- FAM and 3'-BHQ1 labels	CTGGGCCTACTGGACCGGCT
MSP263	5'-end of *sclA* and excluded signal sequence used for cloning of *sclA* for expression; contains *Nde*I site	GCAGCGTTAAGCATATGACTCCTATG
MSP264	3'-end of *sclA* used for cloning of *sclA* for expression; contains *Xho*I site	CCTTGCTCCTCGAGTTAACCTCGTGGTCC
MSP275	5'-end – including native promoter – of *sclA* and used with MSP277 to amplify for cloning into pDC123; contains *EcoR*V site	TATCCAACGATATCAGAAGGTACAAG

MSP277	3'-end of *sclA* used with MSP275 to amplify for cloning into pDC123; contains *Hind*III site	ACCTTGAAAAGCTTGCTCACCGCG
232Up	5' end of *sclA* used for amplification of MGAS315 *sclA*Invasive allele	CTCCACAAAAGAGTGATCAGTC
ME7	Used with 232Up to generate *sclA*Invasive from MGAS315 for cloning into pJRS525	TCAGTGAATTCTCTTTAGAGGATTAG
pJRS525 F1	Used with ME6 to generate repaired full-length sclA fragment from MGAS315 (*sclA*M3-FL)	GGGTTTTCCCAGTCACG
ME6	Reverse primer used for PCR mutagenesis to repair null mutation in *sclA*Invasive	TCCAGCAGGACCTCGAGGTGAACGC
M3VF	5' end of sclA for cloning *sclA*M3-FL into pASK-IBA2 expression vector	GAGATGGCCGAGACTCCTATGACATCAAAGG
Scl1.28W MR	3' end of sclA containing StreptagII sequence for cloning *sclA*M3-FL into pASK-IBA2 expression vector	GTCAAGCTTATTATTTTTCGAACTGCGGGTGGCTCCAAGGTTTTTCTGGAGCTGGAGTTACC

References

1. Musser, J. M., Hauser, A. R., Kim, M. H., Schlievert, P. M., Nelson, K., and Selander, R. K. (1991) *Streptococcus pyogenes* causing toxic-shock-like syndrome and other invasive diseases: clonal diversity and pyrogenic exotoxin expression. *Proc Natl Acad Sci U S A* **88**, 2668-2672

2. Carroll, R. K., Shelburne, S. A., 3rd, Olsen, R. J., Suber, B., Sahasrabhojane, P., Kumaraswami, M., Beres, S. B., Shea, P. R., et al. (2011) Naturally occurring single amino acid replacements in a regulatory protein alter streptococcal gene expression and virulence in mice. *J Clin Invest* **121**, 1956-1968

3. Flores, A. R., Jewell, B. E., Olsen, R. J., Shelburne, S. A., 3rd, Fittipaldi, N., Beres, S. B., and Musser, J. M. (2014) Asymptomatic carriage of group A streptococcus is associated with elimination of capsule production. *Infect Immun* **82**, 3958-3967

4. Chaffin, D. O., and Rubens, C. E. (1998) Blue/white screening of recombinant plasmids in Gram-positive bacteria by interruption of alkaline phosphatase gene (*phoZ*) expression. *Gene* **219**, 91-99

5. McIver, K. S., and Scott, J. R. (1997) Role of *mga* in growth phase regulation of virulence genes of the group A *Streptococcus*. *J Bacteriol* **179**, 5178-5187

SUPPLEMENTARY MATERIAL FOR CHAPTER 6

REVERSIBLE PHOSPHOLIPID NANOGELS FOR DEOXYRIBONUCLEIC ACID FRAGMENT SIZE DETERMINATIONS UP TO 1,500 BASE PAIRS AND INTEGRATED SAMPLE STACKING

Brandon C. Durney, Beth A. Bachert, Hillary S. Sloane, Slawomir Lukomski, James P. Landers, and Lisa A. Holland
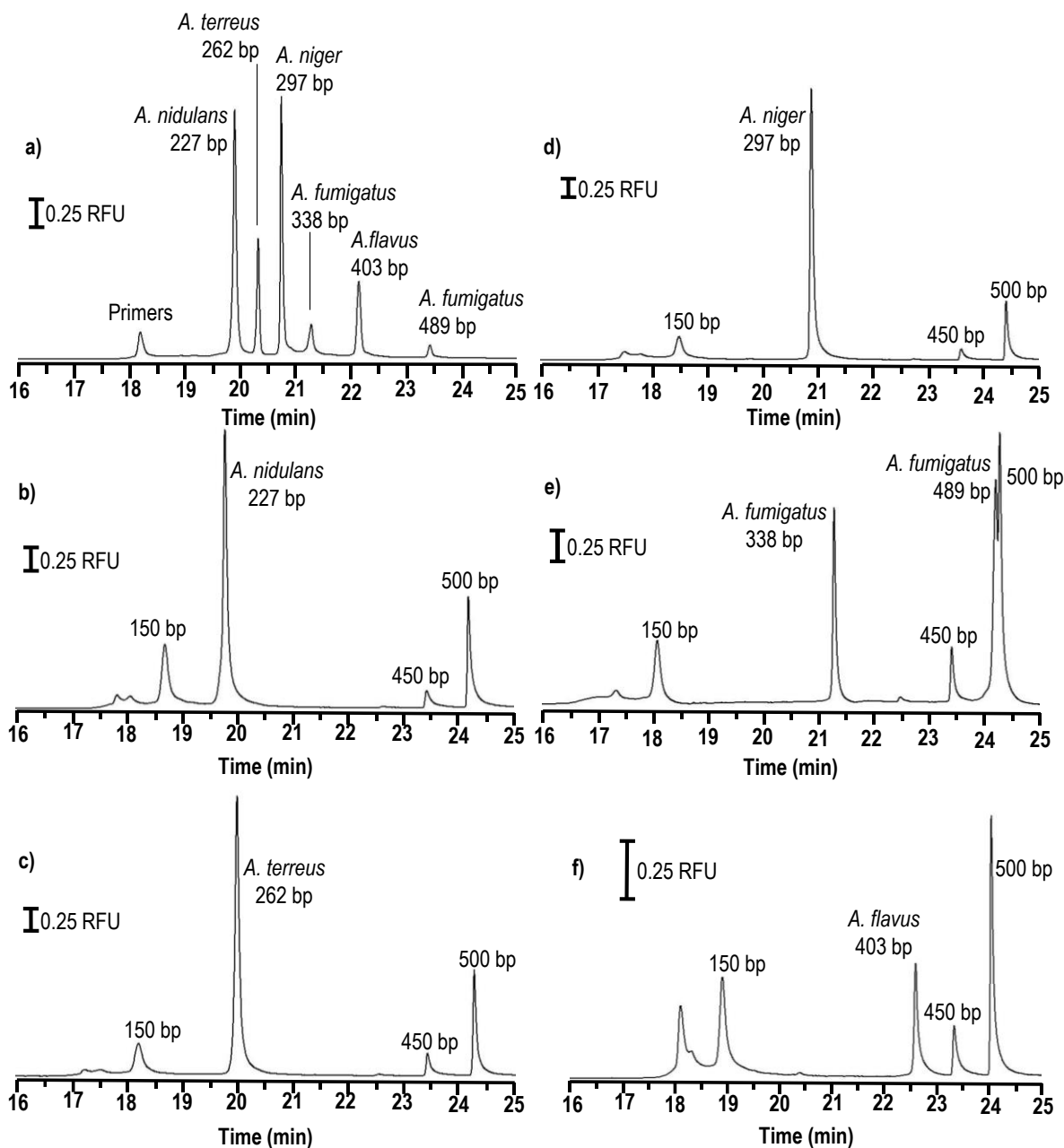
Figure S1. Separation of **(a)** 5 amplicons unique to different species of *Aspergillus* as well as electropherograms that contain internal standards of 150, 450, and 500 base pairs as well as a single amplicon for **(b)** *A. nidulans* (227 bp); **(c)** *A. terreus* (262 bp); **(d)** *A. niger* (297 bp); **(e)** *A. fumigatus* (338 and 489 bp); **(f)** *A. flavus* (403 bp).

Table S1. Effect of SYBR Green 1 concentration on migration time of 650 bp DNA[a]

Concentration	Injection size	Migration time in minutes	Chromatographic[2] resolution (CV)
2 µM	5 psi 7 sec	27.9	(4)
20 µM	5 psi 7 sec	27.8	(2)
100 µM	5 psi 7 sec	26.1	(5)

[a][DMPC]/[DHPC] = 2.5, 2.5% phospholipid. Separations ($n = 5$) are accomplished in a 25 µm id capillary with a total length of 40 cm capillary and effective length of 30.2 cm, with $E_{applied} = 100$ V/cm. The DNA base ladder (1Kb+ DNA ladder, Life Technologies) is detected with LIF at 520 nm using SYBR green 1 nucleic acid stain at different concentrations positioned before the detection window.

Table S2. Effect of phospholipid concentration on resolution[1]

| Hydration (%) | Temp (°C) | Resolution 450/475 base pair DNA[2] | | Resolution 900/950 base pair DNA[3] | |
		Chromatographic[2] resolution (CV)	In base pairs[3]	Chromatographic[2] resolution (CV)	In base pairs[3]
2.5	23	5.0 (6)	5	6.4 (7)	8
5.0	25	5.1 (10)	5	4.5 (10)	11
7.5	30	5.4 (7)	5	4.2 (20)	12
10.0	30	6.2 (7)	4	3.6 (20)	14

[1]Separations ($n = 5$) are achieved at the specified temperature with [DMPC]/[DHPC] = 2.5 in a 40 cm long, 25 μm id capillary. The effective length is 30.2 cm and E_{app} = 100 V/cm.
[2]Resolution is calculated as $0.589(\Delta t/w_{1/2av})$, where Δt is the difference in peak migration times and $w_{1/2av}$ is the average peak width at half height
[3]Resolution in base pairs is calculated as the difference in base pairs/resolution.

Table S3. Effect of [DMPC]/[DHPC] on resolution[1]

Hydration (%)	Temp (°C)	Resolution 450/475 base pair DNA[2]		Resolution 900/950 base pair DNA[3]	
		Chromatographic[2] resolution (CV)	In base pairs[3]	Chromatographic[2] resolution (CV)	In base pairs[3]
2.5	23	5.0 (6)	5	6.4 (7)	8
3.0	21	5.1 (8)	5	5.3 (10)	9
4.0	21	5.5 (10)	5	5.5 (10)	9
5.0	21	1.4 (20)	18	2.6 (20)	36

[1]Separations ($n = 5$) are achieved at the specified temperature with [DMPC]/[DHPC] = 2.5 in a 40 cm long, 25 μm id capillary. The effective length is 30.2 cm and E_{app} = 100 V/cm.

[2]Resolution is calculated as $0.589(\Delta t/w_{1/2av})$, where Δt is the difference in peak migration times and $w_{1/2av}$ is the average peak width at half height

[3]Resolution in base pairs is calculated as the difference in base pairs/resolution.

Table S4. Effect of temperature on resolution[1]

| Temp (°C) | Resolution 450/475 base pair DNA[2] | | Resolution 900/950 base pair DNA[3] | |
	Chromatographic[2] resolution (CV)	In base pairs[3]	Chromatographic[2] resolution (CV)	In base pairs[3]
19	1.2 (20)	22	1.8 (20)	27
20	3.2 (10)	8	5.3 (10)	9
21	4.5 (9)	6	5.7 (10)	9
22	4.5 (8)	6	5.8 (10)	9
23	5.0 (6)	5	6.4 (7)	8
24	2.7 (10)	9	4.2 (20)	12
25	2.5 (10)	10	3.3 (20)	15

[1]Separations ($n = 5$) are achieved at the specified temperature with [DMPC]/[DHPC] = 2.5 in a 40 cm long, 25 μm id capillary. The effective length is 30.2 cm and $E_{app} = 100$ V/cm.
[2]Resolution is calculated as $0.589(\Delta t/w_{1/2av})$, where Δt is the difference in peak migration times and $w_{1/2av}$ is the average peak width at half height
[3]Resolution in base pairs is calculated as the difference in base pairs/resolution.

SUPPLEMENTARY MATERIAL FOR CHAPTER 7

A UNIQUE SET OF THE *BURKHOLDERIA* COLLAGEN-LIKE PROTEINS PROVIDES INSIGHT INTO PATHOGENESIS, GENOME EVOLUTION AND NICHE ADAPTATION, AND INFECTION DETECTION

Beth A. Bachert**,** Soo J. Choi, Anna K. Snyder, Rita V.M. Rio, Brandon C. Durney, Lisa A. Holland, Kei Amemiya, Susan L. Welkos, Joel A. Bozue, Christopher K. Cote, Rita Berisio, and Slawomir Lukomski

Figure S1. Phylogenetic analyses of *bucl1* in *B. pseudomallei* and B. *mallei* strains.

Nucleotide sequences encoding (A) the noncollagenous domain and (B) entire gene of *bucl1* alleles were used. Support values for each branch are shown as posterior probability from Bayesian analysis and bootstrap values from maximum parsimony analysis, respectively (PP/MP). Scale bar is representative of evolutionary distance in substitutions per nucleotide.
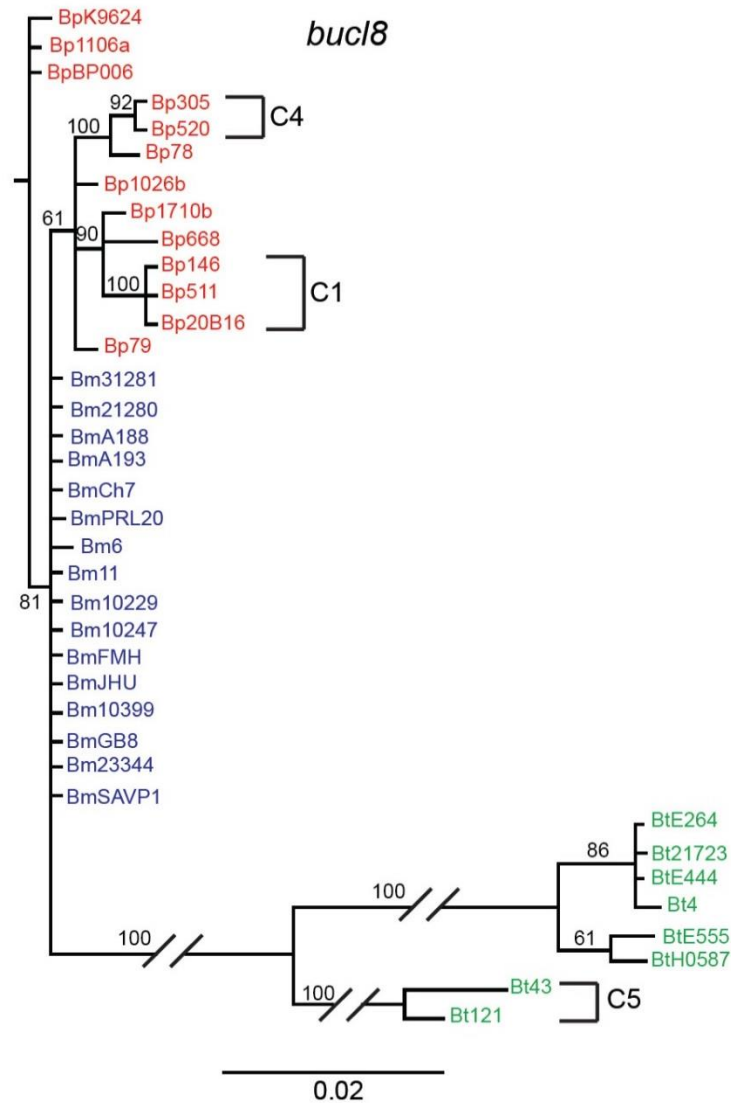
Figure S2. Phylogenetic analysis of *bucl8* among *Burkholderia* strains. Bayesian analysis was performed on nucleotide sequences of *bucl8* non-collagenous regions of a set of *Burkholderia* strains described in Table 3. Support values for each branch are shown as posterior probability from Bayesian analysis. Several clusters of strains, C1, C4, and C5, corresponding to those observed in the concatenated analysis were also observed. Scale bar is representative of evolutionary distance in substitutions per nucleotide.
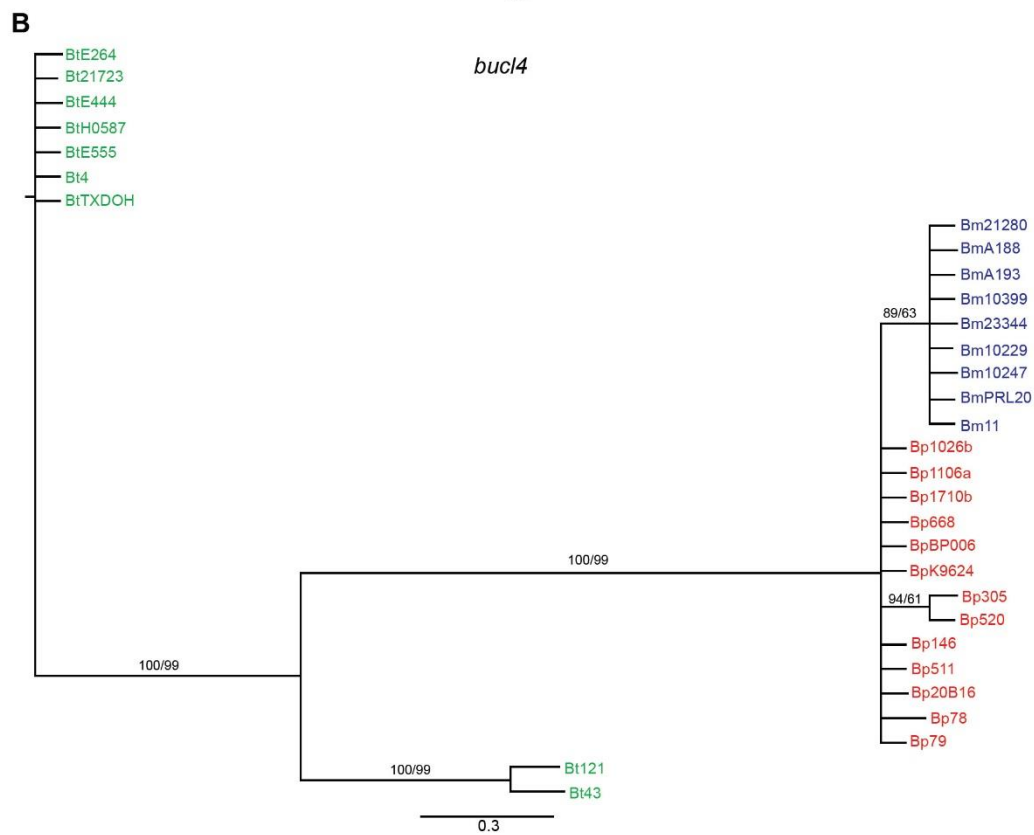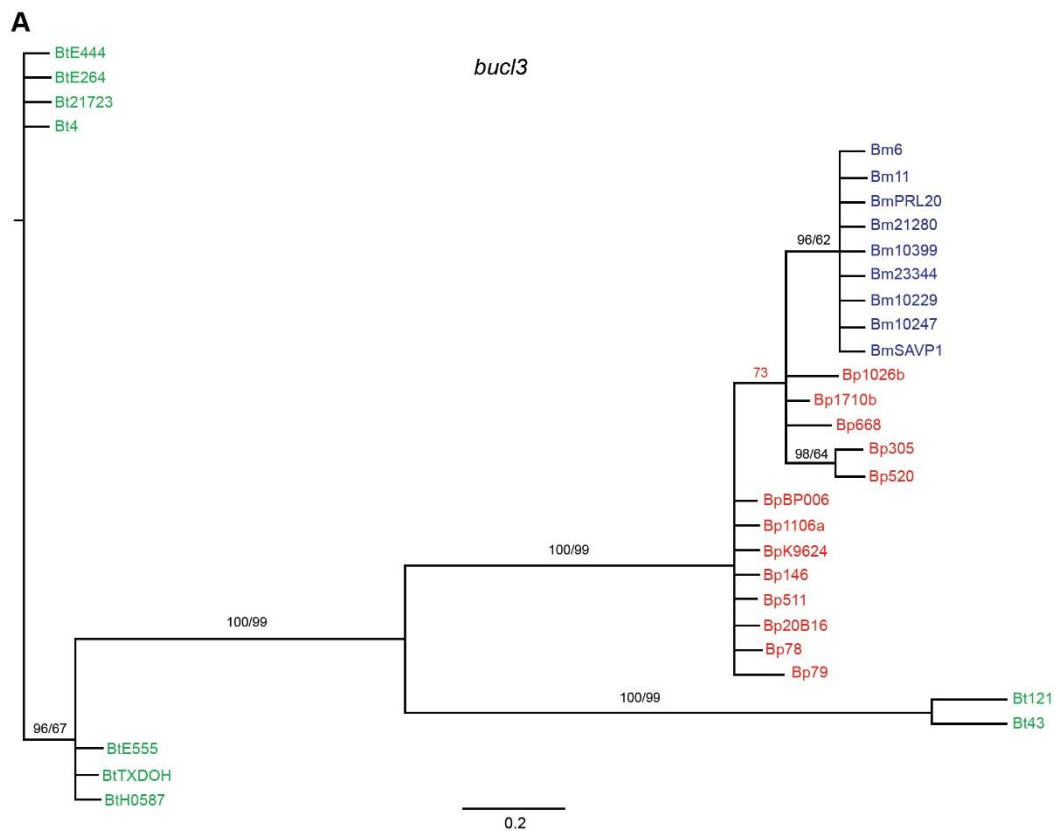
bucl3

bucl4

Figure S3. Phylogenetic analysis of Bucl3 and Bucl4 amino acid sequences among *Burkholderia* strains. Bayesian analysis was performed on amino acid sequences of (A) Bucl3 and (B) Bucl4 non-collagenous regions of a set of *Burkholderia* strains described in Table 3. Support values for each branch are shown as posterior probability from Bayesian analysis and bootstrap values from maximum parsimony analysis, respectively (PP/MP). Posterior probability value, which was not supported by maximum parsimony analysis is shown in red. Scale bar is representative of evolutionary distance in substitutions per nucleotide.
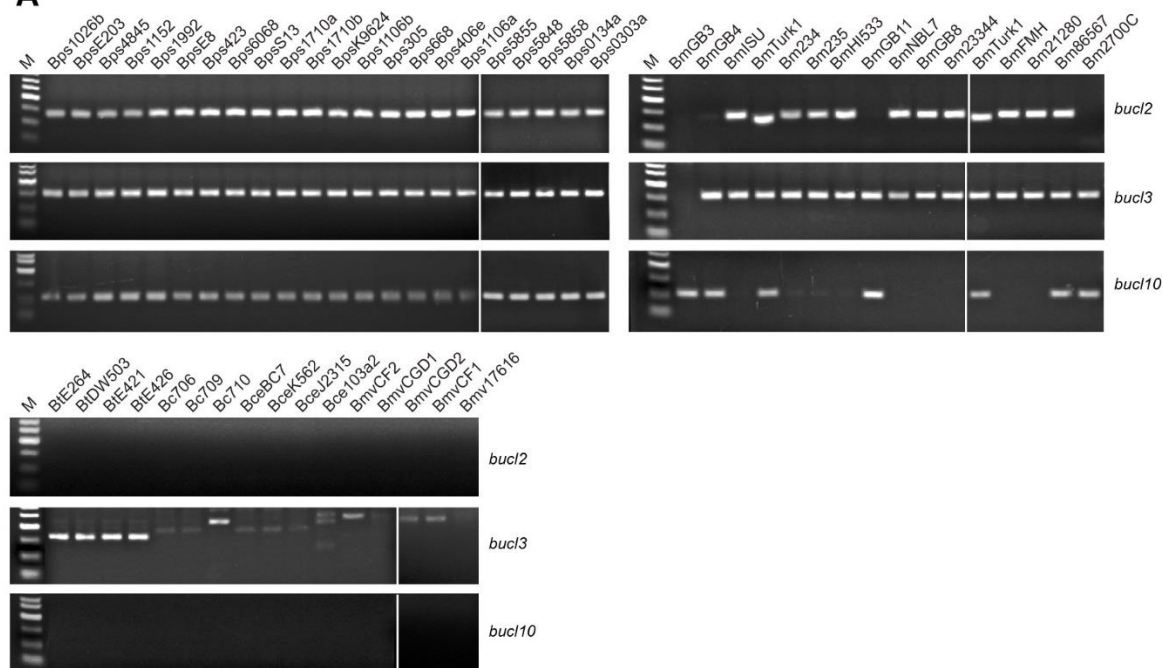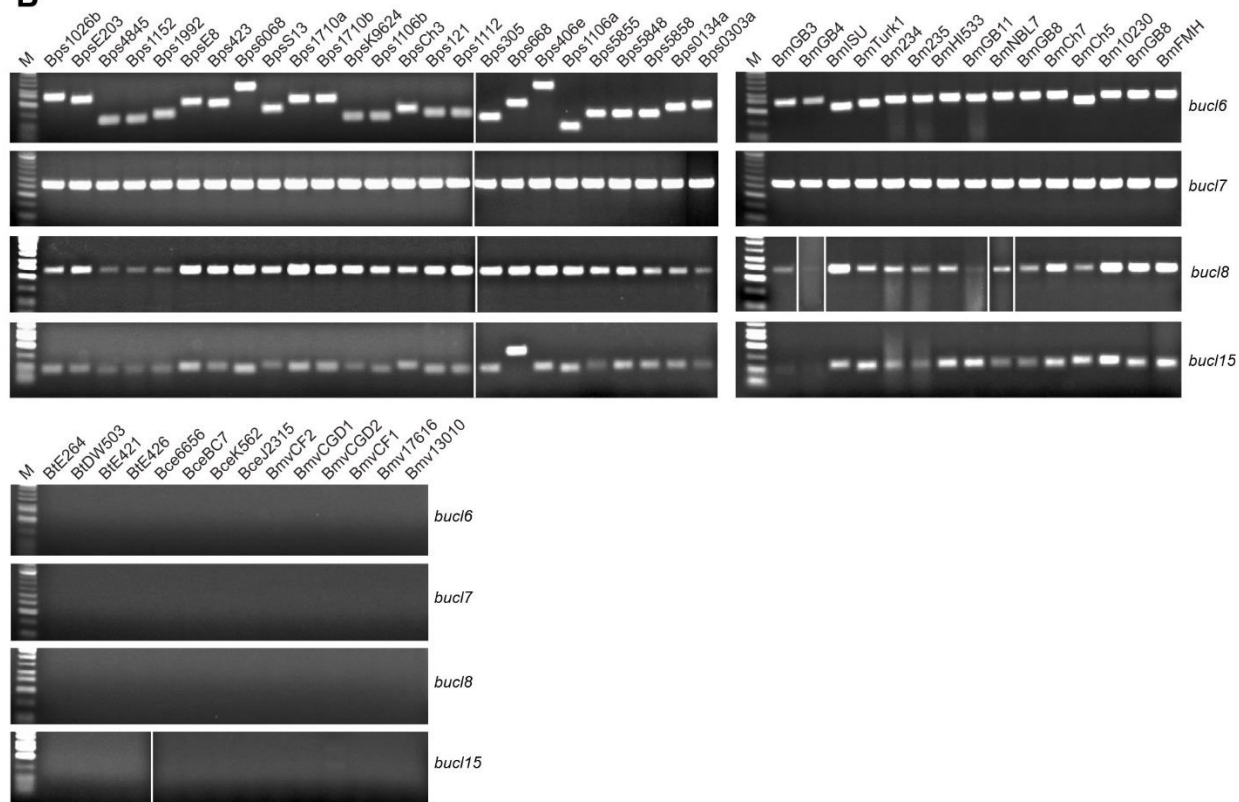
282

Figure S4. Distribution of *bucl* genes among *Burkholderia* spp. select agents by PCR. Presence of (A) *bucl* genes *2*, *3*, and *10* and (B) *bucl* genes *6*, *7*, *8*, and *15*, was assessed by PCR on a collection of genomic DNA from *B. pseudomallei* and *B. mallei* select agents (top panels), as well as in control strains of *B. thailandensis*, *B. cepacia*, *B. cenocepacia*, and *B. multivorans* (bottom panels). Amplicon sizes based on Bp K96243: In A) *bucl2*, 133 bp; *bucl3*, 166 bp; and *bucl10*, 109 bp; In B) *bucl6*, 115 bp; *bucl7*, 264 bp; *bucl8*, 243 bp; and *bucl15*, 95 bp.M, 50-bp DNA ladder. PCR data shown in panels A and B for 25 Bp strains come from two merged gel images.

GENERAL REFERENCES

1. Berisio, R., Vitagliano, L., Mazzarella, L., and Zagari, A. (2002) Crystal structure of the collagen triple helix model [(Pro-Pro-Gly)$_{10}$]$_3$. *Protein Sci* **11**, 262-270
2. Ramachandran, G. N. (1988) Stereochemistry of collagen. *Int J Pept Protein Res* **31**, 1-16
3. Bella, J. (2016) Collagen structure: new tricks from a very old dog. *Biochem J* **473**, 1001-1025
4. Shoulders, M. D., and Raines, R. T. (2009) Collagen structure and stability. *Annu Rev Biochem* **78**, 929-958
5. Brodsky, B., and Persikov, A. V. (2005) Molecular structure of the collagen triple helix. *Adv Protein Chem* **70**, 301-339
6. Exposito, J. Y., Cluzel, C., Garrone, R., and Lethias, C. (2002) Evolution of collagens. *Anat Rec* **268**, 302-316
7. Engel, J. (1997) Versatile collagens in invertebrates. *Science* **277**, 1785-1786
8. Suhre, M. H., Gertz, M., Steegborn, C., and Scheibel, T. (2014) Structural and functional features of a collagen-binding matrix protein from the mussel byssus. *Nat Commun* **5**, 3392
9. Kramer, J. M., Cox, G. N., and Hirsh, D. (1982) Comparisons of the complete sequences of two collagen genes from *Caenorhabditis elegans*. *Cell* **30**, 599-606
10. Lukomski, 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
11. Sylvestre, P., Couture-Tosi, E., and Mock, M. (2002) A collagen-like surface glycoprotein is a structural component of the *Bacillus anthracis* exosporium. *Mol Microbiol* **45**, 169-178
12. Pizarro-Guajardo, M., Olguin-Araneda, V., Barra-Carrasco, J., Brito-Silva, C., Sarker, M. R., and Paredes-Sabja, D. (2014) Characterization of the collagen-like exosporium protein, BclA1, of *Clostridium difficile* spores. *Anaerobe* **25**, 18-30
13. Paterson, G. K., Nieminen, L., Jefferies, J. M. C., and Mitchell, T. J. (2008) PclA, a pneumococcal collagen-like protein with selected strain distribution, contributes to adherence and invasion of host cells. *FEMS Microbiol Lett* **285**, 170-176
14. Duncan, C., Prashar, A., So, J., Tang, P., Low, D. E., Terebiznik, M., and Guyard, C. (2011) Lcl of *Legionella pneumophila* is an immunogenic GAG binding adhesin that promotes interactions with lung epithelial cells and plays a crucial role in biofilm formation. *Infect Immun* **79**, 2168-2181
15. Bachert, B. A., Choi, S. J., Snyder, A. K., Rio, R. V. M., Durney, B. C., Holland, L. A., Amemiya, K., Welkos, S. L., et al. (2015) A unique set of the *Burkholderia* collagen-like proteins provides insight into pathogenesis, genome evolution and niche adaptation, and infection detection. *PLoS ONE* **10**, e0137578
16. Chopra, R. K., and Ananthanarayanan, V. S. (1982) Conformational implications of enzymatic proline hydroxylation in collagen. *Proc Natl Acad Sci U S A* **79**, 7180-7184

17. Vitagliano, L., Berisio, R., Mazzarella, L., and Zagari, A. (2001) Structural bases of collagen stabilization induced by proline hydroxylation. *Biopolymers* **58**, 459-464

18. Berisio, R., and Vitagliano, L. (2012) Polyproline and triple helix motifs in host-pathogen recognition. *Curr Protein Pept Sci* **13**, 855-865

19. 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**, 348-356

20. Xu, Y., Keene, D. R., Bujnicki, J. M., Hook, M., and Lukomski, S. (2002) Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. *J Biol Chem* **277**, 27312-27318

21. Han, R., Zwiefka, A., Caswell, C. C., Xu, Y., Keene, D. R., Lukomska, E., Zhao, Z., Hook, M., et al. (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

22. Chan, V. C., Ramshaw, J. A., Kirkpatrick, A., Beck, K., and Brodsky, B. (1997) Positional preferences of ionizable residues in Gly-X-Y triplets of the collagen triple-helix. *J Biol Chem* **272**, 31441-31446

23. Leikina, E., Mertts, M. V., Kuznetsova, N., and Leikin, S. (2002) Type I collagen is thermally unstable at body temperature. *Proc Natl Acad Sci U S A* **99**, 1314-1318

24. Mohs, A., Silva, T., Yoshida, T., Amin, R., Lukomski, S., Inouye, M., and Brodsky, B. (2007) Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline. *J Biol Chem* **282**, 29757-29765

25. Rasmussen, M., Jacobsson, M., and Bjorck, L. (2003) Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J Biol Chem* **278**, 32313-32316

26. Doxey, A. C., and McConkey, B. J. (2013) Prediction of molecular mimicry candidates in human pathogenic bacteria. *Virulence* **4**, 453-466

27. Frickey, T., and Lupas, A. (2004) CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* **20**, 3702-3704

28. Lukomski, S., Nakashima, K., Abdi, I., Cipriano, V. J., Shelvin, B. J., Graviss, 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

29. Whatmore, A. M. (2001) *Streptococcus pyogenes sclB* encodes a putative hypervariable surface protein with a collagen-like repetitive structure. *Microbiology* **147**, 419-429

30. Rasmussen, M., Eden, A., and Bjorck, L. (2000) SclA, a novel collagen-like surface protein of *Streptococcus pyogenes*. *Infect Immun* **68**, 6370-6377

31. Rasmussen, M., and Bjorck, L. (2001) Unique regulation of SclB - a novel collagen-like surface protein of *Streptococcus pyogenes*. *Mol Microbiol* **40**, 1427-1438

32. Almengor, A. C., and McIver, K. S. (2004) Transcriptional activation of *sclA* by Mga requires a distal binding site in *Streptococcus pyogenes*. *J Bacteriol* **186**, 7847-7857

33. Almengor, A. C., Walters, M. S., and McIver, K. S. (2006) Mga is sufficient to activate transcription in vitro of *sof-sfbX* and other Mga-regulated virulence genes in the group A *Streptococcus*. *J Bacteriol* **188**, 2038-2047

34. Timoney, J. F. (2004) The pathogenic equine streptococci. *Vet Res* **35**, 397-409

35. Karlstrom, A., Jacobsson, K., Flock, M., Flock, J., and Guss, B. (2004) Identification of a novel collagen-like protein, SclC, in *Streptococcus equi* using signal sequence phage display. *Vet Microbiol* **104**, 179-188

36. Beres, S., Sesso, R., Pinto, S., Hoe, N., Porcella, S., Deleo, F., and Musser, J. (2008) Genome sequence of a lancefield group C *Streptococcus zooepidemicus* strain causing epidemic nephritis: new information about an old disease. *PLoS ONE* **3**, e3026

37. Karlstrom, A., Jacobsson, K., and Guss, B. (2006) SclC is a member of a novel family of collagen-like proteins in *Streptococcus equi* subspecies *equi* that are recognised by antibodies against SclC. *Vet Microbiol* **114**, 72-81

38. Flock, M., Karlström, Å., Lannergård, J., Guss, B., and Flock, J. I. (2006) Protective effect of vaccination with recombinant proteins from *Streptococcus equi* subspecies *equi* in a strangles model in the mouse. *Vaccine* **24**, 4144-4151

39. Waller, A., Flock, M., Smith, K., Robinson, C., Mitchell, Z., Karlström, Å., Lannergård, J., Bergman, R., et al. (2007) Vaccination of horses against strangles using recombinant antigens from *Streptococcus equi*. *Vaccine* **25**, 3629-3635

40. Guss, B., Flock, M., Frykberg, L., Waller, A. S., Robinson, C., Smith, K. C., and Flock, J. I. (2009) Getting to grips with strangles: an effective multi-component recombinant vaccine for the protection of horses from *Streptococcus equi* infection. *PLoS Pathog* **5**, e1000584

41. Lannergård, J. (2006) *Potentially virulence-related extracellular proteins of Streptococcus equi.* Doctoral thesis, Swedish University of Agricultural Sciences

42. Bateman, A., Holden, M. T., and Yeats, C. (2005) The G5 domain: a potential N-acetylglucosamine recognition domain involved in biofilm formation. *Bioinformatics* **21**, 1301-1303

43. Ruggiero, A., Tizzano, B., Pedone, E., Pedone, C., Wilmanns, M., and Berisio, R. (2009) Crystal structure of the resuscitation-promoting factor (DeltaDUF)RpfB from *M. tuberculosis*. *J Mol Biol* **385**, 153-162

44. Ruggiero, A., Squeglia, F., Romano, M., Vitagliano, L., De Simone, A., and Berisio, R. (2016) Structure and dynamics of the multi-domain resuscitation promoting factor RpfB from Mycobacterium tuberculosis. *J Biomol Struct Dyn*, 1-9

45. Christner, M., Franke, G. C., Schommer, N. N., Wendt, U., Wegert, K., Pehle, P., Kroll, G., Schulze, C., et al. (2010) The giant extracellular matrix-binding protein of *Staphylococcus epidermidis* mediates biofilm accumulation and attachment to fibronectin. *Mol Microbiol* **75**, 187-207

46. Schroeder, K., Jularic, M., Horsburgh, S. M., Hirschhausen, N., Neumann, C., Bertling, A., Schulte, A., Foster, S., et al. (2009) Molecular characterization of a

novel *Staphylococcus aureus* surface protein (SasC) involved in cell aggregation and biofilm accumulation. *PLoS ONE* **4**, e7567

47.   Imai, S., Ito, Y., Ishida, T., Hirai, T., Ito, I., Yoshimura, K., Maekawa, K., Takakura, S., et al. (2011) Distribution and clonal relationship of cell surface virulence genes among *Streptococcus pneumoniae* isolates in Japan. *Clin Microbiol Infect* **17**, 1409-1414

48.   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

49.   Squeglia, F., Bachert, B., Romano, M., Lukomski, S., and Berisio, R. (2013) Crystallization and preliminary X-ray crystallographic analysis of the variable domain of Scl2.3, a streptococcal collagen-like protein from invasive M3-type *Streptococcus pyogenes. Acta Crystallogr Sect F Struct Biol Cryst Commun* **69**, 1023-1025

50.   Squeglia, F., Bachert, B., De Simone, A., Lukomski, S., and Berisio, R. (2014) The crystal structure of the streptococcal collagen-like protein 2 globular domain from invasive M3-type group A *Streptococcus* shows significant similarity to immunomodulatory HIV protein gp41. *J Biol Chem* **289**, 5122-5133

51.   Han, R., Caswell, C. C., Lukomska, E., Keene, D. R., Pawlowski, M., Bujnicki, J. M., Kim, J. K., and Lukomski, S. (2006) Binding of the low-density lipoprotein by streptococcal collagen-like protein Scl1 of *Streptococcus pyogenes. Mol Microbiol* **61**, 351-367

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

53.   Ffrench-Constant, C. (1995) Alternative splicing of fibronectin--many different proteins but few different functions. *Exp Cell Res* **221**, 261-271

54.   Oliver-Kozup, H., Martin, K. H., Schwegler-Berry, D., Green, B. J., Betts, C., Shinde, A. V., Van De Water, L., and Lukomski, S. (2013) The group A streptococcal collagen-like protein-1, Scl1, mediates biofilm formation by targeting the extra domain A-containing variant of cellular fibronectin expressed in wounded tissue. *Mol Microbiol* **87**, 672-689

55.   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

56.   Jarnagin, W. R., Rockey, D. C., Koteliansky, V. E., Wang, S. S., and Bissell, D. M. (1994) Expression of variant fibronectins in wound healing: cellular source and biological activity of the EIIIA segment in rat hepatic fibrogenesis. *J Cell Biol* **127**, 2037-2048

57.   Shinde, A. V., Bystroff, C., Wang, C., Vogelezang, 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 $\alpha9\beta1$-dependent cellular activities. *J Biol Chem* **283**, 2858-2870

58.   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 $\alpha9\beta1$

and one of its ligands, the EIIIA segment of fibronectin, in cutaneous wound healing. *J Invest Dermatol* **123**, 1176-1181

59. Yamaguchi, M., Terao, Y., and Kawabata, S. (2013) Pleiotropic virulence factor - *Streptococcus pyogenes* fibronectin-binding proteins. *Cell Microbiol* **15**, 503-511

60. Caswell, C. C., Oliver-Kozup, H., Han, R., Lukomska, E., and Lukomski, S. (2010) Scl1, the multifunctional adhesin of group A *Streptococcus*, selectively binds cellular fibronectin and laminin, and mediates pathogen internalization by human cells. *FEMS Microbiol Lett* **303**, 61-68

61. Fisher, M., Huang, Y.-S., Li, X., McIver, K. S., Toukoki, C., and Eichenbaum, Z. (2008) Shr is a broad-spectrum surface receptor that contributes to adherence and virulence in group A *Streptococcus. Infect Immun* **76**, 5006-5015

62. Terao, Y., Kawabata, S., Kunitomo, E., Nakagawa, I., and Hamada, S. (2002) Novel laminin-binding protein of *Streptococcus pyogenes*, Lbp, is involved in adhesion to epithelial cells. *Infect Immun* **70**, 993-997

63. Linke, C., Caradoc-Davies, T. T., Young, P. G., Proft, T., and Baker, E. N. (2009) The laminin-binding protein Lbp from *Streptococcus pyogenes* is a zinc-receptor. *J Bacteriol*

64. Humtsoe, J. O., Kim, J. K., Xu, Y., Keene, D. R., Hook, M., Lukomski, S., and Wary, K. K. (2005) A streptococcal collagen-like protein interacts with the $\alpha2\beta1$ integrin and induces intracellular signaling. *J Biol Chem* **280**, 13848-13857

65. Caswell, C. C., Barczyk, M., Keene, D. R., Lukomska, E., Gullberg, D. E., and Lukomski, S. (2008) Identification of the first prokaryotic collagen sequence motif that mediates binding to human collagen receptors, integrins $\alpha2\beta1$ and $\alpha11\beta1$. *J Biol Chem* **283**, 36168-36175

66. Kim, J., Xu, Y., Xu, X., Keene, D., Gurusiddappa, S., Liang, X., Wary, K., and Hook, M. (2005) A novel binding site in collagen type III for the integrins, $\alpha1\beta1$ and $\alpha2\beta1$. *J Biol Chem* **280**, 32512-32520

67. Akiyama, H., Morizane, S., Yamasaki, O., Oono, T., and Iwatsuki, K. (2003) Assessment of *Streptococcus pyogenes* microcolony formation in infected skin by confocal laser scanning microscopy. *J Dermatol Sci* **32**, 193-199

68. 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

69. Lembke, C., Podbielski, A., Hidalgo-Grass, C., Jonas, L., Hanski, E., and Kreikemeyer, B. (2006) Characterization of biofilm formation by clinically relevant serotypes of group A streptococci. *Appl Environ Microbiol* **72**, 2864-2875

70. Courtney, H. S., Ofek, I., Penfound, T., Nizet, V., Pence, M. A., Kreikemeyer, B., Podbielski, A., Hasty, D. L., et al. (2009) Relationship between expression of the family of M proteins and lipoteichoic acid to hydrophobicity and biofilm formation in *Streptococcus pyogenes*. *PLoS ONE* **4**, e4166

71. Oliver-Kozup, H. A., Elliott, M., Bachert, B. A., Martin, K. H., Reid, S. D., Schwegler-Berry, D. E., Green, B. J., and Lukomski, S. (2011) The streptococcal collagen-like protein-1 (Scl1) is a significant determinant for biofilm formation by group A *Streptococcus. BMC Microbiol* **11**, 262

72. Bachert, B. A., Choi, S. J., LaSala, P. R., Harper, T. I., McNitt, D. H., Boehm, D. T., Caswell, C. C., Ciborowski, P., et al. (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

73. Flores, 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

74. Caswell, C. C., Lukomska, E., Seo, N. S., Hook, M., and Lukomski, S. (2007) Scl1-dependent internalization of group A *Streptococcus* via direct interactions with the $\alpha2\beta1$ integrin enhances pathogen survival and re-emergence. *Mol Microbiol* **64**, 1319-1331

75. Popova, S. N., Lundgren-Akerlund, E., Wiig, H., and Gullberg, D. (2007) Physiology and pathology of collagen receptors. *Acta Physiol (Oxf)* **190**, 179-187

76. Zutter, M. M., and Santoro, S. A. (1990) Widespread histologic distribution of the $\alpha2\beta1$ integrin cell-surface collagen receptor. *Am J Pathol* **137**, 113-120

77. Pfaff, M., Aumailley, M., Specks, U., Knolle, J., Zerwes, H. G., and Timpl, R. (1993) Integrin and Arg-Gly-Asp dependence of cell adhesion to the native and unfolded triple helix of collagen type VI. *Exp Cell Res* **206**, 167-176

78. Dohrmann, S., Anik, S., Olson, J., Anderson, E. L., Etesami, N., No, H., Snipper, J., Nizet, V., et al. (2014) Role for streptococcal collagen-like protein 1 in M1T1 group A *Streptococcus* resistance to neutrophil extracellular traps. *Infect Immun* **82**, 4011-4020

79. Caswell, C. C., Han, R., Hovis, K. M., Ciborowski, P., Keene, D. R., Marconi, R. T., and Lukomski, S. (2008) The Scl1 protein of M6-type group A *Streptococcus* binds the human complement regulatory protein, factor H, and inhibits the alternative pathway of complement. *Mol Microbiol* **67**, 584-596

80. Rodriguez de Cordoba, S., Esparza-Gordillo, J., Goicoechea de Jorge, E., Lopez-Trascasa, M., and Sanchez-Corral, P. (2004) The human complement factor H: functional roles, genetic variations and disease associations. *Mol Immunol* **41**, 355-367

81. Jozsi, M., and Zipfel, P. F. (2008) Factor H family proteins and human diseases. *Trends Immunol* **29**, 380-387

82. Reuter, M., Caswell, C. C., Lukomski, S., and Zipfel, P. F. (2010) Binding of the human complement regulators CFHR1 and factor H by streptococcal collagen-like protein 1 (Scl1) via their conserved C termini allows control of the complement cascade at multiple levels. *J Biol Chem* **285**, 38473-38485

83. Kraiczy, P., Skerka, C., Kirschfink, M., Brade, V., and Zipfel, P. F. (2001) Immune evasion of *Borrelia burgdorferi* by acquisition of human complement regulators FHL-1/reconectin and Factor H. *Eur J Immunol* **31**, 1674-1684

84. Kraiczy, P., Skerka, C., Brade, V., and Zipfel, P. F. (2001) Further characterization of complement regulator-acquiring surface proteins of *Borrelia burgdorferi*. *Infect Immun* **69**, 7800-7809

85. Gao, Y., Liang, C., Zhao, R., Lukomski, S., and Han, R. (2010) The Scl1 of M41-type group A *Streptococcus* binds the high-density lipoprotein. *FEMS Microbiol Lett* **309**, 55-61

86.   Pahlman, L. I., Marx, P. F., Morgelin, M., Lukomski, S., Meijers, J. C., and Herwald, H. (2007) Thrombin-activatable fibrinolysis inhibitor binds to *Streptococcus pyogenes* by interacting with collagen-like proteins A and B. *J Biol Chem* **282**, 24873-24881

87.   Valls Seron, M., Plug, T., Marquart, J. A., Marx, P. F., Herwald, H., de Groot, P. G., and Meijers, J. C. (2011) Binding characteristics of thrombin-activatable fibrinolysis inhibitor to streptococcal surface collagen-like proteins A and B. *Thromb Haemost* **106**, 609-616

88.   Zipfel, P. F., Skerka, C., Hellwage, J., Jokiranta, S. T., Meri, S., Brade, V., Kraiczy, P., Noris, M., et al. (2002) Factor H family proteins: on complement, microbes and human diseases. *Biochem Soc Trans* **30**, 971-978

89.   Gustafsson, M. C., Lannergard, J., Nilsson, O. R., Kristensen, B. M., Olsen, J. E., Harris, C. L., Ufret-Vincenty, R. L., Stalhammar-Carlemalm, M., et al. (2013) Factor H binds to the hypervariable region of many *Streptococcus pyogenes* M proteins but does not promote phagocytosis resistance or acute virulence. *PLoS Pathog* **9**, e1003323

90.   Horstmann, R. D., Sievertsen, H. J., Knobloch, J., and Fischetti, V. A. (1988) Antiphagocytic activity of streptococcal M protein: selective binding of complement control protein factor H. *Proc Natl Acad Sci U S A* **85**, 1657-1661

91.   Wurfel, M. M., Kunitake, S. T., Lichenstein, H., Kane, J. P., and Wright, S. D. (1994) Lipopolysaccharide (LPS)-binding protein is carried on lipoproteins and acts as a cofactor in the neutralization of LPS. *J Exp Med* **180**, 1025-1035

92.   Bhakdi, S., Tranum-Jensen, J., Utermann, G., and Fussle, R. (1983) Binding and partial inactivation of *Staphylococcus aureus* alpha-toxin by human plasma low density lipoprotein. *J Biol Chem* **258**, 5899-5904

93.   Murch, O., Collin, M., Hinds, C. J., and Thiemermann, C. (2007) Lipoproteins in inflammation and sepsis. I. Basic science. *Intensive Care Med* **33**, 13-24

94.   Ravnskov, U. (2003) High cholesterol may protect against infections and atherosclerosis. *QJM : monthly journal of the Association of Physicians* **96**, 927-934

95.   Netea, M. G., Demacker, P. N., Kullberg, B. J., Boerman, O. C., Verschueren, I., Stalenhoef, A. F., and van der Meer, J. W. (1996) Low-density lipoprotein receptor-deficient mice are protected against lethal endotoxemia and severe gram-negative infections. *J Clin Invest* **97**, 1366-1372

96.   Netea, M. G., Demacker, P. N., de Bont, N., Boerman, O. C., Stalenhoef, A. F., van der Meer, J. W., and Kullberg, B. J. (1997) Hyperlipoproteinemia enhances susceptibility to acute disseminated *Candida albicans* infection in low-density-lipoprotein-receptor-deficient mice. *Infect Immun* **65**, 2663-2667

97.   Makoveichuk, E., Cherepanov, P., Lundberg, P., Lundberg, S., Forsberg, A., and Olivecrona, G. (2003) pH6 antigen of *Yersinia pestis* interacts with plasma lipoproteins and cell membranes. *J Lipid Res* **44**, 320-330

98.   Liu, L., Zhou, L., Li, Y., Bai, W., Liu, N., Li, W., Gao, Y., Liu, Z., et al. (2015) High-density lipoprotein acts as an opsonin to enhance phagocytosis of group A *Streptococcus* by U937 cells. *Microbiol Immunol* **59**, 419-425

99.   Zhou, L., Liu, L., Yang, J., Li, Y., Bai, W., Liu, N., Li, W., Gao, Y., et al. (2016) LDL acts as an opsonin enhancing the phagocytosis of group A *Streptococcus* by monocyte and whole human blood. *Med Microbiol Immunol* **205**, 155-162

100. Courtney, H. S., Zhang, Y. M., Frank, M. W., and Rock, C. O. (2006) Serum opacity factor, a streptococcal virulence factor that binds to apolipoproteins A-I and A-II and disrupts high density lipoprotein structure. *J Biol Chem* **281**, 5515-5521

101. Bajzar, L. (2000) Thrombin activatable fibrinolysis inhibitor and an antifibrinolytic pathway. *Arterioscler Thromb Vasc Biol* **20**, 2511-2518

102. Plug, T., and Meijers, J. C. (2016) Structure-function relationships in thrombin-activatable fibrinolysis inhibitor. *Journal of Thrombosis and Haemostasis* **14**, 633-644

103. Bengtson, S. H., Sanden, C., Morgelin, M., Marx, P. F., Olin, A. I., Leeb-Lundberg, L. M., Meijers, J. C., and Herwald, H. (2009) Activation of TAFI on the surface of *Streptococcus pyogenes* evokes inflammatory reactions by modulating the kallikrein/kinin system. *J Innate Immun* **1**, 18-28

104. Fiedler, T., Koller, T., and Kreikemeyer, B. (2015) *Streptococcus pyogenes* biofilms-formation, biology, and clinical relevance. *Front Cell Infect Microbiol* **5**, 15

105. 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

106. Sanchez, C. J., Kumar, N., Lizcano, A., Shivshankar, P., Dunning Hotopp, J. C., Jorgensen, J. H., Tettelin, H., and Orihuela, C. J. (2011) *Streptococcus pneumoniae* in biofilms are unable to cause invasive disease due to altered virulence determinant production. *PLoS ONE* **6**, e28738

107. Blanchette-Cain, K., Hinojosa, C. A., Akula Suresh Babu, R., Lizcano, A., Gonzalez-Juarbe, N., Munoz-Almagro, C., Sanchez, C. J., Bergman, M. A., et al. (2013) *Streptococcus pneumoniae* biofilm formation is strain dependent, multifactorial, and associated with reduced invasiveness and immunoreactivity during colonization. *MBio* **4**, e00745-00713

108. Marks, L. R., Davidson, B. A., Knight, P. R., and Hakansson, A. P. (2013) Interkingdom signaling induces *Streptococcus pneumoniae* biofilm dispersion and transition from asymptomatic colonization to disease. *MBio* **4**, e00438-00413

109. Marks, L. R., Mashburn-Warren, L., Federle, M. J., and Hakansson, A. P. (2014) *Streptococcus pyogenes* biofilm growth *in vitro* and *in vivo* and its role in colonization, virulence, and genetic exchange. *J Infect Dis* **210**, 25-34

110. Connolly, K. L., Roberts, A. L., Holder, R. C., and Reid, S. D. (2011) Dispersal of Group A streptococcal biofilms by the cysteine protease SpeB leads to increased disease severity in a murine model. *PLoS ONE* **6**, e18984

111. Jorgensen, R., Purdy, A. E., Fieldhouse, R. J., Kimber, M. S., Bartlett, D. H., and Merrill, A. R. (2008) Cholix toxin, a novel ADP-ribosylating factor from *Vibrio cholerae*. *J Biol Chem* **283**, 10671-10678

112. LaPenta, D., Rubens, C., Chi, E., and Cleary, P. P. (1994) Group A streptococci efficiently invade human respiratory epithelial cells. *Proc Natl Acad Sci USA* **91**, 12115-12119