Cell PRESS
Structure
Article

Structural Basis of Host Cell Recognition by the Pilus Adhesin from *Streptococcus pneumoniae*

Thierry Izoré,[1] Carlos Contreras-Martel,[1] Lamya El Mortaji,[1] Clothilde Manzano,[1] Rémy Terrasse,[1] Thierry Vernet,[1] Anne Marie Di Guilmi,[1,*] and Andréa Dessen[1,*]
[1]Institut de Biologie Structurale *Jean-Pierre Ebel*, UMR 5075 (CEA, CNRS, UJF); 41 rue Jules Horowitz, F-38027 Grenoble, France
*Correspondence: diguilmi@ibs.fr (A.M.D.G.), dessen@ibs.fr (A.D.)
DOI 10.1016/j.str.2009.10.019

SUMMARY

Pili are fibrous virulence factors associated directly to the bacterial surface that play critical roles in adhesion and recognition of host cell receptors. The human pathogen *Streptococcus pneumoniae* carries a single pilus-related adhesin (RrgA) that is key for infection establishment and provides protection from bacterial challenge in animal infection models, but details of these roles remain unclear. Here we report the high-resolution crystal structure of RrgA, a 893-residue elongated macromolecule whose fold contains four domains presenting both eukaryotic and prokaryotic origins. RrgA harbors an integrin I collagen-recognition domain decorated with two inserted "arms" that fold into a positively charged cradle, as well as three "stalk-forming" domains. We show by site-specific mutagenesis, mass spectrometry, and thermal shift assays that intradomain isopeptide bonds play key roles in stabilizing RrgA's stalk. The high sequence similarity between RrgA and its homologs in other Gram-positive microorganisms suggests common strategies for ECM recognition and immune evasion.

INTRODUCTION

Pili are elongated, flexible appendages directly associated to the bacterial surface that have been shown to participate in essential events for infection establishment, including adhesion to the extracellular matrix, DNA transfer, and biofilm formation. The well-studied Gram-negative pili consist of thin fibers formed by non-covalently associated building blocks, and binding to host cells is performed by adhesins, which are often localized on the fimbrial tip. Adhesin recognition of oligosaccharides of target cell glycoproteins or glycolipid receptors determines tissue tropism and is often essential for the initial phases of the infection process (Proft and Baker, 2009; Soto and Hultgren, 1999; Telford et al., 2006). Pili in Gram-positive pathogens have only been recently described, and consist of a backbone formed by covalently linked copies of a major pilin to which minor pilins are associated. Although the minor subunits are not required for formation of the pilus shaft, they have been shown to be essential for host cell recognition and adhesion of a variety of Gram-positive microorganisms (Abbot et al., 2007; Barocchi et al., 2006; Dramsi et al., 2006; Maisey et al., 2007; Mandlik et al., 2007; Nallapareddy et al., 2006; Nelson et al., 2007). In addition, pilus-related adhesins of Gram-positive bacteria confer protection against a potentially lethal bacterial challenge in animal disease models (Gianfaldoni et al., 2007; Maione et al., 2005; Mora et al., 2005), underlining the importance of these molecules in pathogenesis and as well as of their potential application as antigens in vaccination strategies (Proft and Baker, 2009; Telford et al., 2006).

The Gram-positive human pathogen *Streptococcus pneumoniae* accounts for a large number of cases of pneumonia, otitis, and meningitis, and is one of the major causes of community-acquired illnesses. Despite the availability of antibiotic treatment and vaccine strategies, it still claims yearly over 1.5 million victims worldwide, mostly due to the spread of antibiotic-resistant strains and the limited efficacy of vaccines (Bronzwaer et al., 2002; Levine et al., 2006). Thirty percent of *S. pneumoniae* strains isolated in clinical environments are piliated and carry the *rlrA* genetic element, which codes for all seven genes required for pilus formation: the *rlrA* regulating element, the backbone-forming pilin (RrgB), an adhesin (RrgA), a minor pilin (RrgC), and three sortases (SrtC-1, SrtC-2, SrtC-3) (Barocchi et al., 2006; Dramsi et al., 2008; LeMieux et al., 2006). Sortases are cysteine transpeptidases that recognize cell-wall sorting signals (CWSS) such as LPXTG-like motifs on target pilins and catalyze both the covalent association of pilus backbone subunits and the association of minor pilins to the fibrillar shaft (Mandlik et al., 2008a; Scott and Zahner, 2006; Ton-That and Schneewind, 2004). Although sortases have been shown to recognize preferentially one Rrg substrate, it is now known that their specificities can be interchangeable up to a certain extent, a fact that is potentially linked to the similarity of their active sites (Fälker et al., 2008; Manzano et al., 2008; Neiers et al., 2009; Manzano et al., 2009). Immunoelectron microscopy imaging has shown that the pneumococcal backbone fiber is formed by RrgB, whereas RrgA and RrgC are associated along the filament (Barocchi et al., 2006; Hilleringmann et al., 2008; LeMieux et al., 2006).

The only adhesin of the pneumococcal pilus, RrgA, plays a key role in bacterial adhesion both to respiratory and to epithelial cells (Nelson et al., 2007); in addition, it has been shown to interact with different elements of the extracellular matrix (ECM), namely fibronectin, collagen, and laminin (Hilleringmann et al., 2008). Interestingly, however, cellular recognition by pneumococcal pili also occurs in *ΔrrgBΔrrgC* strains, in which the RrgB backbone is notably absent, suggesting that the adhesive properties of RrgA are effective even when it is associated directly onto the bacterial surface (Nelson et al., 2007). However,

Table 1. Data Collection, Phasing, Molecular Replacement, and Structure Refinement Statistics

Data Collection				
Data set	Peak	Inflection Point	Low Remote	Native
Wavelength (Å)	0.9791	0.9793	0.9751	0.9810
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_12_12_1$
a (Å)	51.66	51.78	51.85	51.09
b (Å)	83.66	83.98	84.13	83.73
c (Å)	180.27	180.91	181.24	299.43
β (°)	89.79	89.82	89.83	
Resolution (Å)	2.41	2.41	2.41	1.90
No. observed/unique reflections	193,352/102,274	187,186/100,908	176,800/97,859	582,427/97,681
Completeness (%)	81 (60)	76.4. (50.7)	70.6 (41.8)	88.7 (64.4)
R_{sym} (last shell)	7.9 (36.3)	8.6 (59.4)	9.9 (82.4)	5.3 (35.9)
$I/\sigma(I)$ (last shell)	6.97 (2.13)	6.4 (1.41)	5.75 (0.97)	38.29 (5.96)
Phasing				
Phasing power ANO	0.828	0.606	0.256	
Phasing power ISO (acentric/centric)		0.104/0.109	0.192/0.189	
FOM_{obs}	0.3433			
FOM_{DM}	0.8759			
Molecular Replacement				
Phaser RFZ/TFZ				22.2/31.2
LLG				968/1090
Refinement				
Resolution (Å)				1.90
R_{work} (%)				19.80
R_{fire} (%)				23.80
No. of protein atoms				6486
No. of solvent atoms				819
No. of HEPES molecules				2
Rmsd, bond lengths (Å)				0.024
Rmsd, bond angles (°)				1.890
Mean B factor ($Å^2$)				18.56
Residues in most favored/allowed region of Ramachandran plot (%)				98.23

neither the identity of RrgA's cellular receptor(s) nor details regarding its multifaceted mode of ECM recognition are known.

In order to address this question, we solved the high-resolution crystal structure of RrgA from the highly infectious TIGR4 pneumococcal strain. In contrast to all adhesin structures solved to date, which display compact folds of limited size, RrgA is an 893-residue, elongated molecule that harbors four domains associated semilinearly as "beads on a string." The largest domain harbors a region that displays high similarity to the collagen-binding integrin I domain, whereas the three others display folds present in other bacterial systems, namely a region reminiscent of the CnaB adhesin of *S. aureus* and two immunoglobulin (IgG-like) domains. Both CnaB-like and IgG domains are stabilized by intermolecular isopeptide bonds, which we show, by mutagenesis and thermal shift assays, to be key for RrgA stabilization. The high sequence similarity between RrgA and other streptococcal pilus adhesins suggests that its complex, multifaceted fold, which involves elements from both eukaryotic and prokaryotic systems, will provide information regarding pilus-ECM interactions in a variety of Gram-positive pathogens.

RESULTS

RrgA, A Four-Domain Adhesin

RrgA is an 893-residue protein whose primary structure analysis predicts the presence of a von Willebrand factor-like domain and reveals an LPXTG-like CWSS at its C terminus; a construct that lacks the first 38 amino acids, predicted to be disordered, as well as the CWSS was used for crystallization. RrgA (39-868) crystallized as a monomer in space group $P2_1$, and the structure was solved by performing a three-wavelength anomalous dispersion (MAD) experiment on a selenomethionylated crystal at the ESRF synchrotron in Grenoble; this model was subsequently used to phase native data to 1.9 Å. Data collection, phasing, and refinement statistics are included in Table 1.

RrgA (residues 47-815 Figure 1; see Figure S1 available online) is an unusual four-domain (D1-D4), elongated molecule that is approximately 195 Å long and 70 Å wide; domains D1 and D2 carry elements from both N-terminal and C-terminal regions of RrgA into which other sequences have been inserted (domain D3 is inserted into D2; D2/D3 are inserted into D1). The four domains make few contacts with each other, being mostly associated through short linkers and hairpins, which suggests that the full-length protein, once associated to the RrgB backbone fiber, could display a considerable level of domain flexibility.

Domain D1 is highly reminiscent of an IgG domain (Figure S2). Five β strands are formed from the N terminus of the sequence (residues 47-141), whereas the last strand, which connects to D4, is formed by C-terminal residues 723-731 (dark green in Figure 1B). Domain D2 is formed by N-terminal residues 144-218 and C-terminal residues 593-722 that clasp into an 11-β-stranded sandwich (cyan and blue, Figure 1B). The two "halves" of the sandwich are locked together by an isopeptide bond between Lys191 and Asn695. Notably, the fold of D2 is reminiscent of that of the B domain of the collagen-binding adhesin Cna from *S. aureus* (Deivanayagam et al., 2000) (Figure S2); a comparison of the two structures reveals that it is mostly the core β strands that are structurally comparable, and RrgA harbors additional hairpins and loops located away from the core. The D2 domain is linked to a third, major domain (D3) through a short β-hairpin structure, itself also composed of residues from the N-terminal and C-terminal regions of the molecule (Thr214-Gln218, and Val587-Lys591).

RrgA's D3 domain (Lys219-Ile586) contains a region (yellow in Figure 1) whose fold highly resembles that of the human A3 domain of von Willebrand factor (VWA), a molecule shown to interact with collagens I and III (Bienkowska et al., 1997; Cruz et al., 1995). The VWA fold is found in many eukaryotic cell surface proteins involved in interactions with the extracellular matrix (Springer, 2006), including within the β domain of integrins, where it also has been shown to play a role in collagen interaction (Emsley et al., 2000; Kamata and Takada, 1994; Tuckwell et al., 1995). The classical VWA/integrin I fold consists of a central, six-stranded β sheet surrounded by α helices, resembling a dinucleotide-binding domain (Bienkowska et al., 1997; Emsley et al., 2000; Huizinga et al., 1997; Lee et al., 1995b; Qu and Leahy, 1995; Vorup-Jensen et al., 2003) and superposition of the structure of the α2β1 integrin onto the analogous domain in RrgA reveals high structural similarity (Figure S2) (albeit a sequence identity level that is below 10%). However, RrgA's integrin I-like domain displays a five-stranded parallel sheet and carries two main inserted "arms" (Figure S3). The first inserted region immediately follows β2 and involves residues Phe280-Asp318 (magenta in Figure 1); it folds into an elongated region carrying two β hairpins. The second, larger insertion region ranges from Met392 to Gly516 (red in Figure 1); although it is mostly composed of loops, it carries one short hairpin region and two helices. These two inserted arms extrude from the core of the integrin I-like domain and stretch out, away from the body of the molecule, making a large number of interactions throughout. Interestingly, sequence analyses and structural predictions of RrgA homologs in different pathogenic streptococcal species (Figure 2) reveal that the full-length molecules display a sequence identity level of approximately 50% (with

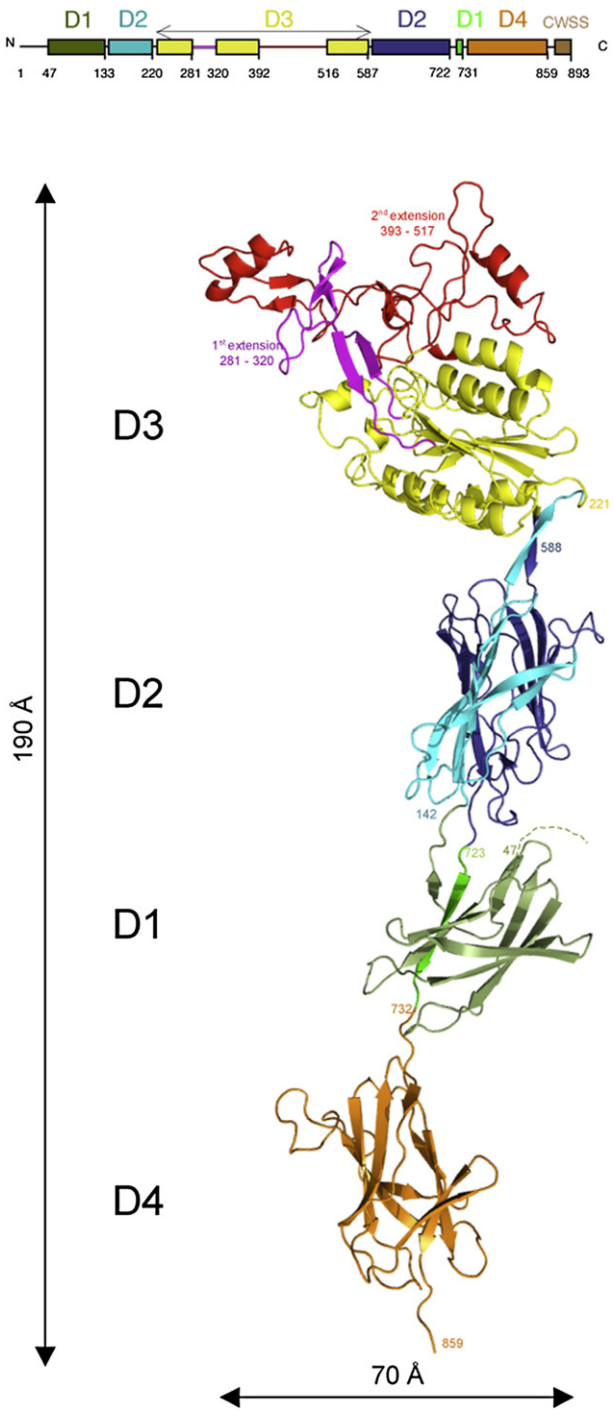

Figure 1. Schematic Arrangement and Tertiary Fold of RrgA

(Top) RrgA is composed of four domains, where D3 is inserted into D2, and D2/D3 are inserted into D1.

(Bottom) In 3D, RrgA's four domains are aligned much like beads on a string; D1 is shown in dark green (N-terminal region) and light green (C-terminal strand); the Cna-B-like domain of D2 is formed by residues from the N terminus (light blue) and C terminus (dark blue) of the molecule. The integrin I-like domain is shown in yellow, with the extended arms in red and magenta. The C terminus of RrgA is composed of domain D4 (orange).

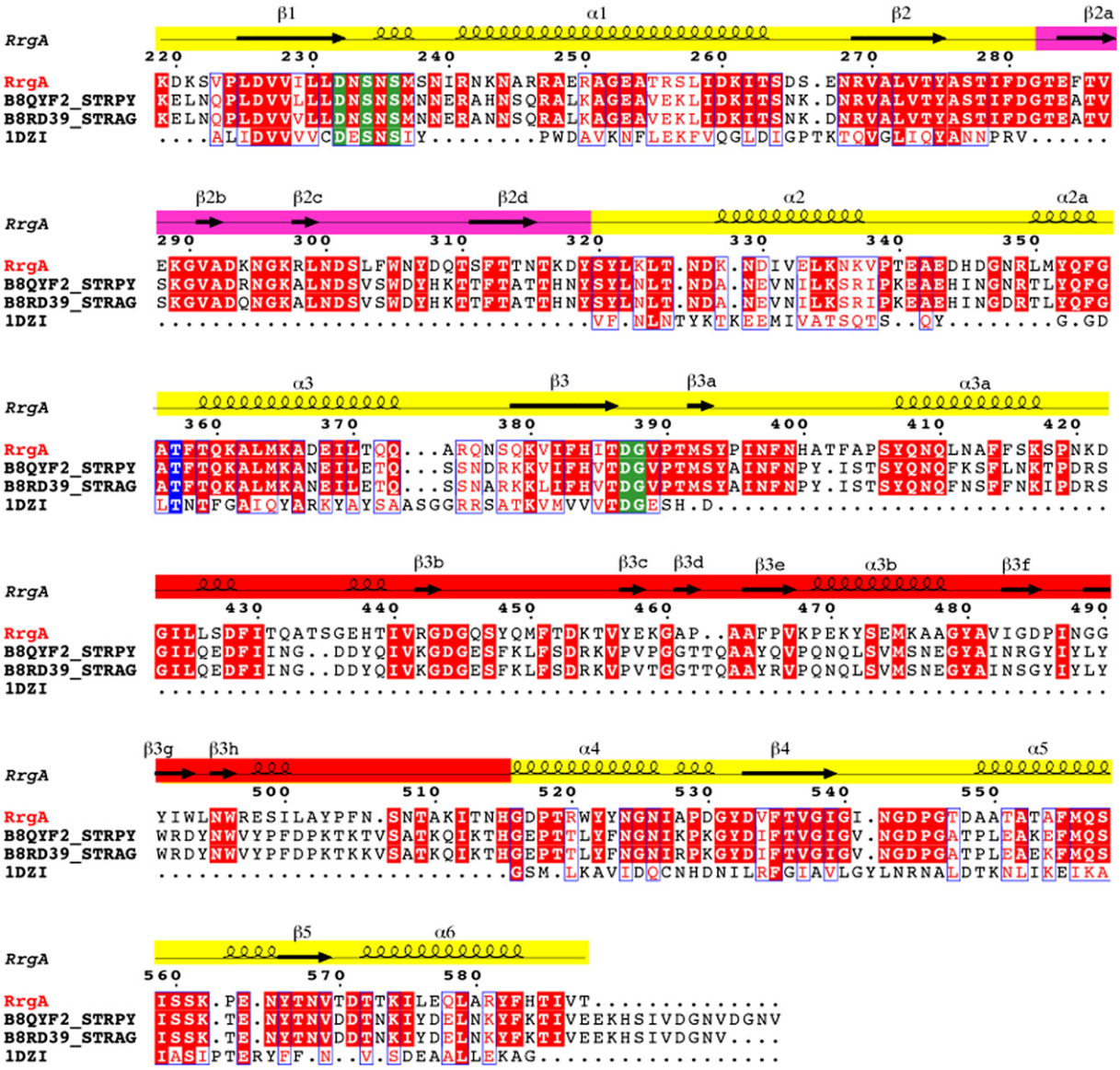

Figure 2. Streptococcal Pilus Adhesins Show Multiple Similar Structural Features

Sequence alignment of the D3 domain of RrgA with the analogous domain of the major pilus adhesins (RrgA homologs) from *Streptococcus agalactiae* (B8RD39) and *Streptococcus pyogenes* (B8QYF2), as well as the I domain of eukaryotic integrin α2β1. Identical residues are shown in white with a red background. Residues involved in the MIDAS motif are highlighted in green. Thr357, which could potentially bind the metal ion directly upon a ligand-induced conformational change, is highlighted in blue. Secondary structure elements corresponding to inserted arms are shown in magenta and red; note their clear absence in the sequence of integrin α2β1. The red arm region shows the highest level of sequence divergence.

similarity levels beyond 70%), and also harbor integrin I-like domains decorated with arms. However, it is precisely in the region of the inserted arms that the sequence identity level is the lowest (in the range of 40%); this is the most evident in the analysis of Figure 2, and suggests that this region could be employed for different roles in pathogenesis (see below).

Domain D4 (orange in Figure 1) precedes the CWSS C-terminal sequence and is separated from domain D3 through a 13-residue stretched linker region with high temperature factors (and thus flexibility). D4, like D1, is highly reminiscent of an IgG domain (Figure S2), and is composed uniquely of C-terminal residues 736-859. IgG domains have been identified within the structure of the minor adhesin GBS52 of the pilus of *S. agalactiae*, where they have been shown to be crucial for bacterial adhesion (Krishnan et al., 2007), and is notably present in structures of adhesins associated to Gram-negative pili (Soto and Hultgren, 1999). In addition, a second isopeptide bond is present between Lys742 and Asn854 (see below).

Multiple Extracellular Matrix-Binding Platforms

Integrin I domains carry a MIDAS (metal ion-dependent adhesion site) motif located largely on the C-termini of the central β sheet.

This site is formed by an aspartate and 2 serine residues organized within a Asp-X-Ser-X-Ser sequence, as well as a Thr and a second Asp residue, and has been shown to play a key role in the recognition of collagen (Emsley et al., 2000; Kamata and Takada, 1994). The analogous region in RrgA, located within domain D3, also carries a MIDAS motif composed of the Asp232-X-Ser234-X-Ser236 constellation of residues on loop1 and one acidic residue (Asp 387) on loop 3 (Figures 2 and 3A). In addition, as is the case for integrins, the MIDAS motif of RrgA coordinates a metal ion that we have modeled here as an Mg^{+2} (although no metal was included in the crystallization solution). Ser234, Ser236, and Asp387 form direct bonds to the metal through their hydroxyl oxygen atoms, whereas Asp232 and Thr357 contribute water-mediated contacts. It is of note that integrin I domains undergo large conformational changes upon ligand binding (Springer, 2006), and the conformation of the RrgA MIDAS domain, presented here in the absence of ligand, is closed. It is of interest, however, that the high structural similarity between the D3 domain of RrgA and beta-I integrin domains, including the presence of metal within the MIDAS motif, clearly points to collagen recognition as a potential function for RrgA; interestingly, this is confirmed by the studies of Hilleringmann and colleagues (Hilleringmann et al., 2008) who identified collagen as a major target for RrgA. One notable difference between the MIDAS motif of eukaryotic integrins and pneumococcal RrgA is that in the former, the motif is located superficially, on a shallow region of the protein surface, whereas in RrgA it is positioned within an a trench-like region cradled by the two inserted arms (Figure S4). Superposition of the structure of integrin $\alpha 2\beta 1$ in complex with a triple-stranded peptide from collagen III (PDB 1dzi) onto RrgA generated a large number of clashes (not shown), suggesting that accommodation of the collagen peptide into the trench would require localized movement of MIDAS residues, as shown for ''open'' conformations of many integrin I domains (Springer, 2006). Superposition of the structure of the RrgA MIDAS domain with the open and ''closed'' forms of the CR3 β2-integrin (Lee et al., 1995a) suggests that cleft opening could include movement of the Thr357 side chain, which could potentially interact directly with the metal ion in the liganded form (Figure S5). In addition, conformational changes could include movement or rotation of the arms. Such an entrenched recognition between a pilus adhesin and an ECM component could provide accrued stability to the bacterium on the matrix surface.

In addition to recognizing collagen, RrgA has also been shown to be important for *S. pneumoniae*'s recognition of laminin and fibronectin, common components of the eukaryotic extracellular matrix (ECM) (Hilleringmann et al., 2008). Glycosaminoglycans (GAGs) are linear polysaccharides attached covalently to proteins of the proteoglycan family, and are ubiquitously present on all animal cell surfaces. Interestingly, several proteins that recognize GAGs do so through elongated crevices rich in basic residues (Imberty et al., 2007). Notably, analysis of a surface potential map of RrgA reveals that the two insertion arms of D3 join on the region that is the farthest from the N and C termini of RrgA in order to form a large, cradle-shaped surface of highly basic character (Figure 4). The concentration of basic residues present in the U-shaped cradle suggests that the cradle could provide the binding site for negatively charged molecules (such as GAGs that are directly associated to fibronectin and laminin) and thus bridge the interaction between the bacterium and the ECM. Interestingly, only few of the basic residues identified in the U-shaped cradle of RrgA are strictly conserved (as Lys or

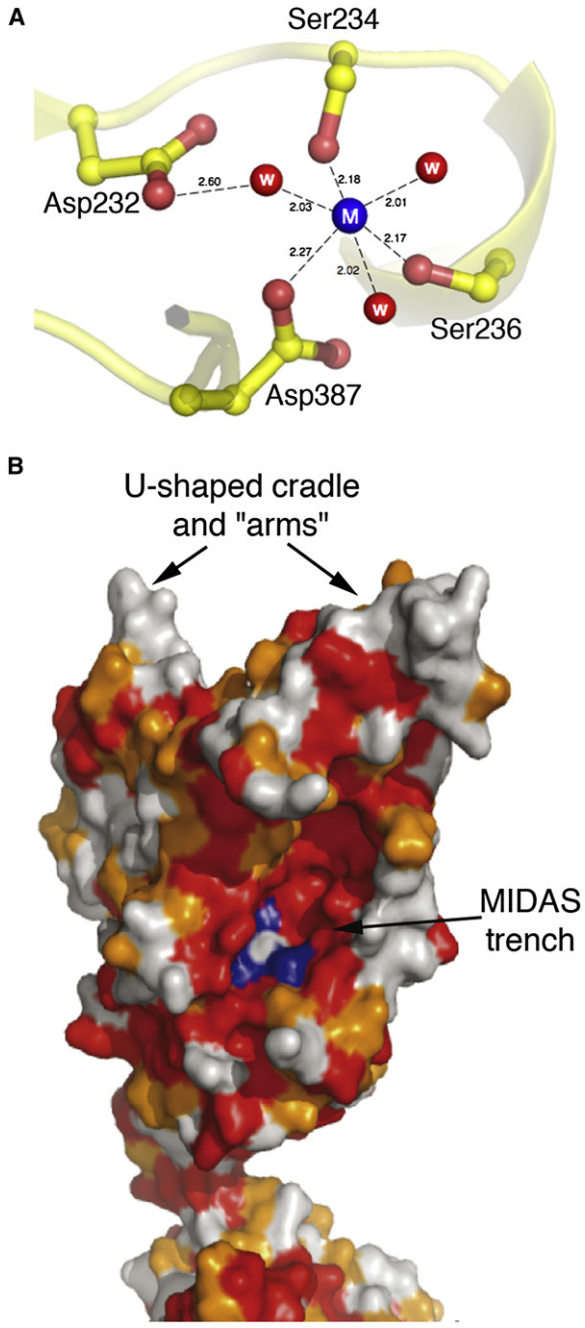

Figure 3. The MIDAS Motif of RrgA

(A) The metal ion, refined here as a Mg^{2+}, is stabilized by a classical constellation of residues also seen in MIDAS motifs of eukaryotic proteins.

(B) Mapping of identical (red) and similar (orange) residues in RrgA-like adhesins in Streptococcal spp. (*S. pneumoniae*, *S. agalactiae*, *S. pyogenes*) onto the structure of the D3 domain of RrgA. Residues involved in forming the MIDAS motif are in dark blue. The central region of the domain, which harbors the MIDAS motif, is the most conserved, while the arms of the structure display patches of nonconserved sequence.

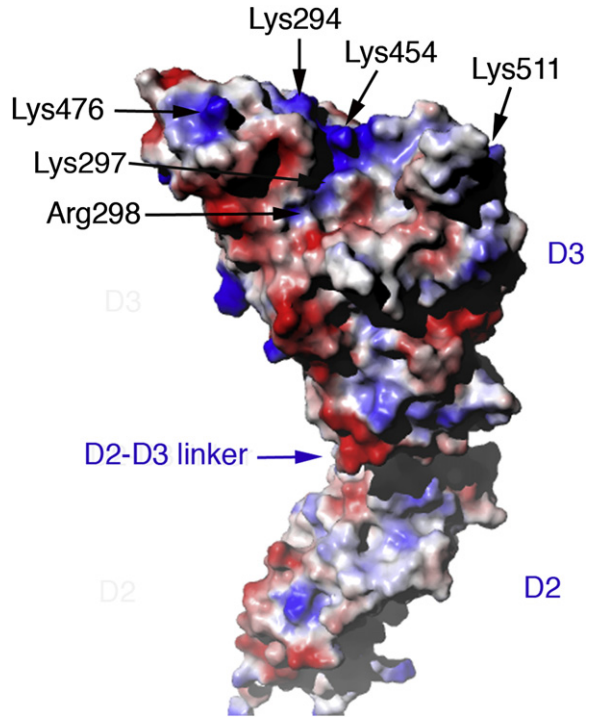

Figure 4. Another ECM-Binding Region Is Formed by the Inserted Arms

Surface potential diagram of the D3 domain of RrgA that shows the basic U-shaped cradle. This view is 180° away from the one shown in Figure 3B.

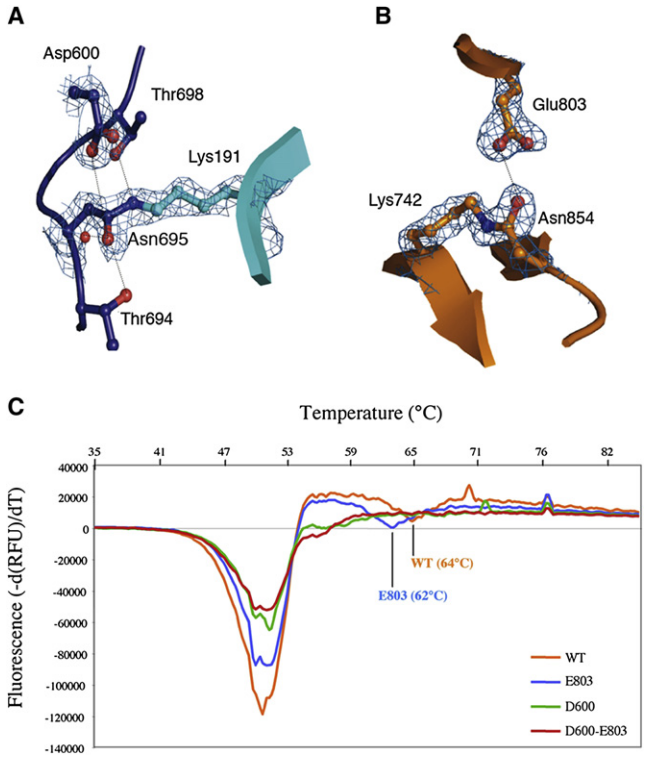

Figure 5. Domains D2 and D4 Are Stabilized by Intramolecular Isopeptide Bonds

(A) Domain D2 carries a bond between the side chains of Lys191 and Asn695 that, in addition to Asp600, is also stabilized by two Thr hydroxyl groups.

(B) The isopeptide bond in domain D4 is stabilized by nearby Glu803. In both (A) and (B), Fo-Fc simulated annealing omit maps are shown.

(C) Thermal shift analysis results for RrgA isopeptide bond mutants. Wild-type RrgA (orange) displays maxima at 51°C and 64°C, while mutants involving the D2 domain isopeptide bond have lost the 64°C peak, indicating that the RrgA stalk could be compromised.

Arg) among the sequences of the pilus adhesins of *S. agalactiae* and *S. pyogenes* (Figures 2 and 3B), suggesting that the crevice formed by the arms in these adhesins could potentially recognize other host cell molecules. This observation could provide a partial explanation for the distinct tissue tropism observed for piliated Gram-positive species.

The "Stalk" Is Stabilized by Intramolecular Isopeptide Bonds

The pilus-forming biosynthetic machinery in Gram-positive organisms requires the formation of intermolecular isopeptide bonds, notably between the sortases and the LPXTG-like motif on their target proteins (Mandlik et al., 2008b; Scott and Zahner, 2006). In addition, isopeptide bonds formed within a single domain have also been identified in the structure of Sp0128, the major pilin subunit from *S. pyogenes* (Kang et al., 2007) and of a minor accessory pilus molecule from the *S. agalactiae* pilin (Krishnan et al., 2007). These bonds have been shown to be the product of an intramolecular reaction between the side chains of a Lys and an adjacent Asp residue, and the NZ and OD1 atoms of the resulting bond are both stabilized by hydrogen bonding with an acidic residue in the immediate vicinity (Kang et al., 2007). RrgA harbors two intramolecular isopeptide bonds, one located in the Cna-B-like D2 domain, and a second one in D4. The bond within D2 is formed between the side chain of Lys191 (in the N-terminal region of the RrgA sequence) and Asn695 (in the C terminus of the structure). Although an acidic group (Asp600) is located in its vicinity, the side chains of Thr694 and Thr698 also provide additional stabilization for the isopeptide linkage (Figure 5A). The isopeptide bond present in D4 is formed by proximal elements of the C-terminal sequence, namely Lys742 and Asn854; the hydrogen bonding stabilization profile is more classical, being provided by Glu803 (Figure 5B). As is the case for other isopeptide linkages, the intramolecular isopeptide bonds in RrgA are also surrounded by hydrophobic residues: Ile206, Leu208, and Phe713 for the D2 bond and Leu751, Phe756, and Ile852 for the D4 bond.

We sought to confirm the importance of the two isopeptide bonds for RrgA's stability using site-specific mutagenesis, mass spectrometry, and thermal shift analyses. The molecular mass of wild-type RrgA, measured by mass spectrometry, was consistent with the absence of two NH_3 groups (thus confirming the formation of two isopeptide bonds; see Table 2). We initially studied the isopeptide bond in D2, and generated single mutants in which Lys191, Asn695, and Asp600 were mutated into alanines. For each mutant, the measured mass was compatible with the loss of one isopeptide bond; thus, mutants K191A and N695A were no longer capable of generating the isopeptide bond in D2 (uniquely in D4). Interestingly, an RrgA variant carrying a D600A mutation behaved similarly to the K191A and N695A mutants, also displaying the inability to generate one

Table 2. Mass Measurements of RrgA Variants by Electrospray Ionization Mass Spectrometry Time-of-Flight Mass

	CalculatedMass (Da)	MeasuredMass (Da)	Mass Difference (Calculated − Measured)	Number of Isopeptide Bonds Present
WT RrgA	94 197.5	94,163.7	33.8	2
E803A	94 139.5	94,123.8	15.7	1
K191A	94 140.4	94,124.8	15.6	1
N695A	94 154.5	94,139.3	15.2	1
D600A	94 153.5	94,139.3	14.2	1
D600A/E803A	94 095.5	94,096.0	−0.5	0

The presence of one isopeptide bond corresponds to the loss of one NH_3 group (17Da).

isopeptide bond even in the presence of intact Lys191 and Asn695 side chains. These results suggest that the function of the acidic side chain in isopeptide bond formation is dual: it provides hydrogen bonding stabilization to the bond once it is formed, but must also play a key role in its formation during protein folding. In light of these results and in order to study the isopeptide bond located within domain D4, we generated an RrgA variant carrying a single E803A mutation. Mass spectrometric measurements of RrgA-E803A confirmed that one isopeptide bond was also missing in this form, thus confirming that the presence of an acidic residue in the immediate environment of an isopeptide bond is an essential factor for its formation. Notably, no isopeptide bonds could be identified in an RrgA-D600A-E803A double mutant (Table 2).

We then performed thermal shift analyses of all RrgA variants. Proteins were gradually heated in an IQ5 Bio-Rad apparatus and thermal unfolding curves were monitored through the detection of changes in fluorescence of a Sypro Orange probe. This analysis (Figure 5c) reveals that wild-type RrgA displays two unfolding maxima: the first at approximately 51°C (±0.5), which is shared with all other RrgA variants, and a second, less pronounced peak, whose T_m value is approximately 64°C (±0.2). Because the major peak is present in experiments with all RrgA forms independent of the presence or absence of isopeptide bonds, it could represent the unfolding of the D3 domain. In the RrgA-E803A variant, which lacks the isopeptide bond in domain D4, the T_m of the second peak is 2°C lower than for the wild-type form; this observation suggests that the minor peak corresponds to the unfolding of the smaller domains. Notably, in the cases of both RrgA-D600A and RrgA-D600A-E803A, the second peak is not observed; the absence of this second transition point suggests that absence of the isopeptide bond in domain D2 (Cna-B-like) engenders a destabilization of the entire RrgA stalk. Recently, it has been shown that the isopeptide bonds carried by *S. pyogenes* Sp0128 also play a key role in stabilization of the fiber-forming molecule (Kang and Baker, 2009) whereas isopeptide bonds in the analogous molecule of *B. cereus*, BcpA, are crucial for pilus assembly (Budzik et al., 2008). These results confirm that such bonds play a universal stabilizing role in the pilus-formation processes of Gram-positive species.

DISCUSSION

Adhesion to target host cells by surface-exposed virulence factors, such as pili, is a key step in infection establishment by a large number of bacterial pathogens. The well-studied pili from Gram-negative pathogens have been shown to display specific adhesins on their tips that play important roles in adherence to both the extracellular matrix and to glycoprotein or glycolipid receptors, often determining tissue tropism (Proft and Baker, 2009; Soto and Hultgren, 1999). In Gram-positive pathogens, however, clarification of the details of the pilus-mediated adhesion process has only recently started. In this work, we report the structural characterization of a totally novel adhesin that is essential for the infectivity process of a major human pathogen and whose tertiary fold is composed of linearly aligned domains observed in both eukaryotic and prokaryotic organisms.

RrgA is composed of four independent domains linked by flexible regions. This organization can potentially allow for interdomain flexibility, which could be important for the recognition of different host targets during the infection process, while the presence of two isopeptide bonds guarantees stability of the individual domains. The composite nature of RrgA includes two IgG domains (D1/D4), and their potential function could include enhancement of binding of the adhesin to host tissue receptors and/or providing stability for the macromolecule. These functions are observed in a number of bacterial adhesins; PapG, for example, is located at the distal end of the P pilus of uropathogenic *E. coli* (UPEC) and folds into two IgG-like domains, one of which is key for receptor recognition (Dodson et al., 2001). The GBS52 pilus adhesin of *S. agalactiae* also harbors two IgG-like regions, one of which is key in promoting adherence to lung tissue, while the second one provides structural support (Krishnan et al., 2007). In RrgA, the latter function could be shared by its Cna-B-like domain (D2), because the B domain of collagen-binding protein Cna of *S. aureus* has also been suggested as being the stalk that projects the A collagen-binding domain away from the bacterial surface and toward the ECM (Deivanayagam et al., 2000). In RrgA, domains D1, D2, and D4 are associated linearly, with D3 located at the edge (Figures 1; Figure S1). This arrangement suggests that one of the roles of domains D1, D2, and D4 could be to project the ECM-recognition D3 domain away from the pilus fiber in order to reach out toward the host cell as much as possible. Association of RrgA to the RrgB pilus backbone is catalyzed through recognition of dedicated sortases of the LPXTG-like CWSS motif on the C terminus of D4, lending support to this hypothesis.

The D3 domain of RrgA harbors two different potential ECM-recognition platforms: the integrin I domain, shown to participate in collagen binding in many eukaryotic proteins, and the basic

U-shaped cradle, located at the very edge of the adhesin, which could potentially be the binding site for negatively charged glucosaminoglycans or other acidic molecules of the ECM. The high sequence similarity between RrgA and pilus adhesins from *S. pyogenes* and *S. agalactiae* points to the possibility that the three-dimensional fold of these proteins could be highly similar (notably with the strict conservation of the MIDAS motif in all three molecules, Figures 2 and 3B); the identification of collagen-binding capabilities by a *S. pyogenes* strain (Kreikemeyer et al., 2005) provides support for this proposition. In addition, the presence of the integrin I domain, which is widespread in a number of eukaryotic proteins, in pilus-associated adhesins of all three pathogenic species suggests that this could be a strategy that allows the pathogen to evade host defense mechanisms. Albeit the structural similarity of the core regions, there are notable charge differences within the arms (notably within residues forming the cradle) that could be an indication of the differential interaction of such adhesins with the host, which is to be expected from molecules that target different tissues.

RrgA and its counterparts in *S. pyogenes* and *S. agalactiae* have been shown to provide protection in animal models of acute disease and could thus represent interesting potential antigens for vaccine development (Abbot et al., 2007; Nelson et al., 2007; Dramsi et al., 2006; Maisey et al., 2007). Notably, strategies for the generation of novel vaccines include the rational engineering of antigenic targets through the understanding of domain boundaries and epitope structures (Dormitzer et al., 2008). The identification of individual domains with distinct host cell recognition functions in the structure of the pilus adhesin reported here is an essential step in this process, and should be exploited in the development of novel anti-streptococcal vaccines. In addition, the predicted structural similarity between the adhesins of these three major piliated microorganisms suggests that the RrgA structure could provide a basis for the understanding of ECM-recognition features by different Gram-positive pathogens.

EXPERIMENTAL PROCEDURES

Expression and Purification of Recombinant Proteins

The region of the *S. pneumoniae* TIGR4 *rrgA* gene (SP0462) coding for amino acids 39 to 868 was amplified using conventional polymerase chain reaction (PCR) methodologies and cloned into vector pLIM (Protein Expert, Grenoble). The resulting vector was transformed into *E. coli* RIL cells (Invitrogen) to produce recombinant protein with an N-terminal hexahistidine tag. Protein expression was induced in Terrific Broth with 1 mM IPTG at 37°C for 3 hr. Bacteria were collected by centrifugation, resuspended in a minimal volume of buffer (50 mM Tris-HCl, 150 mM NaCl [pH 8.0], 20 mM Imidazole, 1 mM PMSF, 0.1μm aprotinin, and 1 μM pepstatin) and lysed by sonication. The lysate was clarified by centrifugation and applied onto an immobilized metal chelate affinity chromatography column (HisTrapHP; GE HealthCare) pre-equilibrated in lysis buffer. Protein elution was performed by an imidazole step-gradient. Fractions containing RrgA were pooled and dialyzed 3 hr into 50 mM Tris-HCl, 150 mM NaCl (pH 8.0) at 4°C, and subsequently incubated with 1/5 (w:w) TEV protease at 15°C. The protein was reloaded onto an HisTrapHP column and the eluted cleaved products dialyzed 3 hr into the same buffer as above. Subsequently, RrgA was loaded onto a MonoQ (5/50GL GE HealthCare) column in 50 mM Tris-HCl (pH 8.0) and eluted in the flow through; fractions were pooled, concentrated, and injected onto a HiLoad 16/60 Superdex 200 (GE HealthCare) column in 50 mM Tris-HCl, 150 mM NaCl (pH 8.0). Pooled and concentrated fractions were used for crystallization tests. All RrgA variants were expressed using the same methodology, except that the MonoQ step was not employed.

Selenomethionylated RrgA was expressed in *E. coli* RIL cells in minimal medium supplemented with thiamine (0.2 mg/ml), leucine (50 mg/ml), valine (50 mg/ml), isoleucine (50 mg/ml), lysine (100 mg/ml), phenylalanine (100 mg/ml), threonine (100 mg/ml), and selenomethionine (60 mg/ml). Expression and purification was performed as for the native protein.

The pLIM-RrgA expression plasmid served as a template for the introduction of single amino or double acid substitutions by PCR using the QuikChange XL Site Directed Mutagenesis Kit (Stratagene). All the mutant constructs were subsequently sequenced; the results showed that only expected mutations were introduced during PCR. Expression and purification of all RrgA mutants were performed as for the wild-type protein.

Crystallization, Data Collection, and Structure Solution

Crystals of native RrgA were grown by the hanging-drop vapor diffusion method at 20°C in 20% PEG 3350, 0.2 M disodium phosphate dihydrate, while selenomethionine crystals were produced in 50 mM MES (pH 6.5), 5% PEG 4000, 10% glycerol, and 100mM NaCl. Both native and selenomethyonylated crystals were cryoprotected by successive brief incubations in mother liquor containing increasing concentrations of glycerol (up to 20%), and were subsequently flash-cooled in liquid nitrogen until data collection. A three-wavelength MAD experiment was performed on the Se edge on the ID-29 beamline (2.4 Å, space group $P2_1$). The three data sets from the MAD experiment were processed and scaled with XDS (Kabsch, 1993); identification of selenium atom positions, refinement, and phasing were performed with AutoSHARP (Bricogne et al., 2003; de la Fortelle and Bricogne, 1997) and automatic model building was performed with ARP/wARP 7.0.1 (Perrakis et al., 1999). PHASER (Storoni et al., 2004) was subsequently used to perform molecular replacement using the model generated by the MAD experiment in order to phase a native data set (1.9 Å, space group $P2_12_12_1$) collected at the European Synchrotron Radiation Facility (ESRF) ID23-EH1 beamline (Grenoble, France). Cycles of automatic and manual building were performed using ARP/wArp and Coot (Emsley and Cowtan, 2004), respectively, coupled with refinement cycles by REFMAC 5.4 (Murshudov et al., 1997).

Fluorescence-Based Assays

Assays were conducted in an IQ5 96-well format real-time PCR instrument (Bio-Rad) in the presence of a Sypro Orange (Molecular Probes) probe. The total volume was to 25 μl. Samples were heat-denatured from 20°C to 100°C at a rate of 1°C per minute. At each step, excitation was performed at 470 nm, while emission of Sypro Orange fluorescence was monitored at 570 nm. Plotting of the fluorescence versus temperature curves, followed by the calculation of the first derivative at each point allowed the identification of each inflection point; the minima were referred to as the melting temperatures (T_m). Data result from two independent experiments performed in triplicate.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and can be found with this article online at doi:10.1016/j.str.2009.10.019.

ACKNOWLEDGMENTS

The coordinates of RrgA have been deposited at the Protein Data Bank with code 2WW8 and are available prerelease from the authors. The authors wish to thank Richard Kahn (IBS) and the staff of ESRF beamlines ID29 and ID23 for help with data collection, Izabel Bérard and Eric Forest from the IBS mass spectrometry facility (LSMP, IBS) for analyses, J. Marquez and the HTX Lab team (Partnership for Structural Biology) for access to and help with high-throughput crystallization, and Hugues Lortat-Jacob (IBS) for many helpful discussions. This work was partly supported by a Young Investigator (ANR Jeunes Chercheurs) grant to A.M.D.G. (n° 05-JCJC-0049) and by EC grant LSHM-CT-2004-512138 (to A.D.).

Received: August 21, 2009
Revised: September 29, 2009
Accepted: October 22, 2009
Published: January 12, 2010

REFERENCES

Abbot, E.L., Smith, W.D., Siou, G.P.S., Chiriboga, C., Smith, R.J., Wilson, J.A., Hirst, B.H., and Kehoe, M.A. (2007). Pili mediate specific adhesion of Streptococcus pyogenes to human tonsil and skin. Cell. Microbiol. *9*, 1822–1833.

Barocchi, M.A., Ries, J., Zogaj, X., Hemsley, C., Albiger, B., Kanth, A., Dahlberg, S., Fernebro, J., Moschioni, M., Masignani, V., et al. (2006). A pneumococcal pilus influences virulence and host inflammatory responses. Proc. Natl. Acad. Sci. USA *103*, 2857–2862.

Bienkowska, J., Cruz, M., Atiemo, A., Handin, R., and Liddington, R. (1997). The von Willebrand Factor A3 domain does not contain a metal ion-dependent adhesion site motif. J. Biol. Chem. *272*, 25162–25167.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallogr. D Biol. Crystallogr. *59*, 2023–2030.

Bronzwaer, S.L., Cars, O., Buchholz, U., Molstad, S., Goettsch, W., Veldhuijzen, I.K., Kool, J.L., Sprenger, M.J.W., and Degener, J.E. (2002). A European study on the relationship between antimicrobial use and antimicrobial resistance. Emerg. Infect. Dis. *8*, 278–282.

Budzik, J.M., Marraffini, L.A., Souda, P., Whitelegge, J.P., Faull, K.F., and Schneewind, O. (2008). Amide bonds assemble pili on the surface of bacilli. Proc. Natl. Acad. Sci. USA *105*, 10215–10220.

Cruz, M.A., Yuan, H., Lee, J.R., Wise, R.J., and Handin, R.I. (1995). Interaction of the von Willebrand factor (vWF) with collagen. J. Biol. Chem. *270*, 10822–10827.

de la Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Methods Enzymol. *276*, 472–494.

Deivanayagam, C.C.S., Rich, R.L., Carson, M., Owens, R.T., Danthuluri, S., Bice, T.W., Hook, M., and Narayana, S.V.L. (2000). Novel fold and assembly of the repetitive B region of the *Staphylococcus aureus* collagen-binding surface protein. Structure *8*, 67–78.

Dodson, K.W., Pinkner, J.S., Rose, T., Magnusson, G., Hultgren, S.J., and Waksman, G. (2001). Structural basis of the interaction of the pyelonephritic E. coli adhesin to its human kidney receptor. Cell *105*, 733–743.

Dormitzer, P.R., Ulmer, J.B., and Rappuoli, R. (2008). Structure-based antigen design: a strategy for next generation vaccines. Trends Biotechnol. *26*, 659–667.

Dramsi, S., Caliot, E., Bonne, I., Guadagnini, S., Prevost, M.-C., Kojadinovic, M., Lalioui, L., Poyart, C., and Trieu-Cout, P. (2006). Assembly and role of pili in group B streptococci. Mol. Microbiol. *60*, 1401–1413.

Dramsi, S., Magnet, S., Davison, S., and Arthur, M. (2008). Covalent attachment of proteins to peptidoglycan. FEMS Microbiol. Rev. *32*, 307–320.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Emsley, J., Knight, C.G., Farndale, R.W., Barnes, M.J., and Liddington, R. (2000). Structural basis of collagen recognition by integrin $\alpha2\beta1$. Cell *101*, 47–56.

Fälker, S., Nelson, A.L., Morfeldt, E., Jonas, K., Hultenby, K., Ries, J., Melefors, O., Normark, S., and Henriques-Normark, B. (2008). Sortase-mediated assembly and surface topology of adhesive pneumococcal pili. Mol. Microbiol. *70*, 595–607.

Gianfaldoni, C., Censini, S., Hilleringmann, M., Moschioni, M., Facciotti, C., Pansegrau, W., Masignani, V., Covacci, A., Rappuoli, R., Barocchi, M.A., and Ruggiero, P. (2007). *Streptococcus pneumoniae* pilus subunits protect mice against lethal challenge. Infect. Immun. *75*, 1059–1062.

Hilleringmann, M., Giusti, F., Baudner, B.C., Masignani, V., Covacci, A., Rappuoli, R., Barocchi, M.A., and Ferlenghi, I. (2008). Pneumococcal pili are composed of protofilaments exposing adhesive clusters of RrgA. PLoS Pathog. *4*, e1000026.

Huizinga, E.G., van der Plas, R.M., Kroon, J., Sixma, J.J., and Gros, P. (1997). Crystal structure of the A3 domain of human von Willebrand factor: implications for collagen binding. Structure *5*, 1147–1156.

Imberty, A., Lortat-Jacob, H., and Perez, S. (2007). Structural view of glycosaminoglycan-protein interactions. Carbohydr. Res. *342*, 430–439.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Crystallogr. *26*, 795–800.

Kamata, T., and Takada, Y. (1994). Direct binding of collagen to the I domain of integrin $\alpha2\beta1$ (VLA-2, CD49b/CD29) in a divalent cation-independent manner. J. Biol. Chem. *269*, 26006–26010.

Kang, H.J., and Baker, E.N. (2009). Intramolecular isopeptide bonds give thermodynamic and proteolytic stability to the major pilin protein of Streptococcus pyogenes. J. Biol. Chem. *284*, 20729–20737.

Kang, H.J., Coulibaly, F., Clow, F., Proft, T., and Baker, E.N. (2007). Stabilizing isopeptide bonds revealed in Gram-positive bacterial pilus structure. Science *318*, 1625–1628.

Kreikemeyer, B., Nakata, M., Oehmcke, S., Gschwendtner, C., Normann, J., and Podbielski, A. (2005). *Streptococcus pyogenes* collagen type I-binding Cpa surface protein. Expression profile, binding characteristics, biological function, and potential clinical impact. J. Biol. Chem. *280*, 33228–33239.

Krishnan, V., Gaspar, A.H., Ye, N., Mandlik, A., Ton-That, H., and Narayana, S.V.L. (2007). An IgG-like domain in the minor pilin GBS52 of *Streptococcus agalactiae* mediates lung epithelial cell adhesion. Structure *15*, 893–903.

Lee, J.-O., Bankston, L.A., Arnaout, M.A., and Liddington, R.C. (1995a). Two conformations of the integrin A-domain (I-domain): a pathway for activation? Structure *3*, 1333–1340.

Lee, J.-O., Rieu, P., Arnaout, M.A., and Liddington, R. (1995b). Crystal structure of the A domain from the α subunit of integrin CR3 (CD11b/CD18). Cell *80*, 631–638.

LeMieux, J., Hava, D.L., Basset, A., and Camilli, A. (2006). RrgA and RrgB are components of a multisubunit pilus encoded by the *Streptococcus pneumoniae* rlrA pathogenicity islet. Infect. Immun. *74*, 2453–2456.

Levine, O.S., O'Brien, K.L., Knoll, M., Adegbola, R.A., Black, S., Cherian, T., Dagan, R., Goldblatt, D., Grange, A., Greenwood, B.M., et al. (2006). Pneumococcal vaccination in developing countries. Lancet *367*, 1880–1882.

Maione, D., Margarit, I., Rinaudo, C.D., Masignani, V., Mora, M., Scarselli, M., Tettelin, H., Brettoni, C., Iacobini, E.T., Rosini, R., et al. (2005). Identification of a universal group B *Streptococcus* vaccine by multiple genome screen. Science *309*, 148–150.

Maisey, H.C., Hensler, M., Nizet, V., and Doran, K.S. (2007). Group B streptococcal pilus proteins contribute to adherence to and invasion of brain microvascular endothelial cells. J. Bacteriol. *189*, 1464–1467.

Mandlik, A., Swierczynski, A., Das, A., and Ton-that, H. (2007). Corynebacterium diphtheriae employs specific minor pilins to target human pharyngeal epithelial cells. Mol. Microbiol. *64*, 111–124.

Mandlik, A., Das, A., and Ton-That, H. (2008a). The molecular switch that activates the cell wall anchoring step of pilus assembly in gram-positive bacteria. Proc. Natl. Acad. Sci. USA *105*, 14147–14152.

Mandlik, A., Swierczynski, A., Das, A., and Ton-That, H. (2008b). Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. Trends Microbiol. *16*, 33–40.

Manzano, C., Contreras-Martel, C., El Mortaji, L., Izore, T., Fenel, D., Vernet, T., Schoehn, G., Di Guilmi, A.M., and Dessen, A. (2008). Sortase-mediated pilus fiber biogenesis in *Streptococcus pneumoniae*. Structure *16*, 1838–1848.

Manzano, C., Izoré, T., Di Guilmi, A.M., and Dessen, A. (2009). Sortase activity is controlled by a flexible lid in the pilus biogenesis mechanism of Gram-positive pathogens. Biochemistry *48*, 10549–10557.

Mora, M., Bensi, G., Capo, S., Falugi, F., Zingaretti, C., Manetti, A.G.O., Maggi, T., Taddei, A.R., Grandi, G., and Telford, J.L. (2005). Group A *Streptococcus* produce pilus-like structures containing protective antigens and Lancefield T antigens. Proc. Natl. Acad. Sci. USA *102*, 15641–15646.

Murshudov, G., Vagin, A., and Dodson, E. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Nallapareddy, S.R., Singh, K.V., Sillanpaa, J., Garsin, D.A., Hook, M., Erlandsen, S.L., and Murray, B.E. (2006). Endocarditis and biofilm-associated pili of Enterococcus faecalis. J. Clin. Invest. *116*, 2799–2807.

Neiers, F., Madhurantakam, C., Fälker, S., Manzano, C., Dessen, A., Normark, S., Henriques-Normark, B., and Achour, A. (2009). Two crystal structures of pneumococcal pilus sortase C provide novel insights into catalysis and substrate specificity. J. Mol. Biol. *393*, 704–716.

Nelson, A.L., Ries, J., Bagnoli, F., Dahlberg, S., Falkner, S., Rounioja, S., Tschop, J., Morfeldt, E., Ferlenghi, I., Hilleringmann, M., et al. (2007). RrgA is a pilus-associated adhesin in *Streptococcus pneumoniae*. Mol. Microbiol. *66*, 329–340.

Perrakis, A., Morris, R.M., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. *6*, 458–463.

Proft, T., and Baker, E.N. (2009). Pili in Gram-negative and Gram-positive bacteria–structure, assembly, and their role in disease. Cell. Mol. Life Sci. *66*, 613–635.

Qu, A., and Leahy, D.J. (1995). Crystal structure of the I-domain from the CD11a/CD18 (LFA-1, αLβ2) integrin. Proc. Natl. Acad. Sci. USA *92*, 10277–10281.

Scott, J.R., and Zahner, D. (2006). Pili with strong attachments: Gram-positive bacteria do it differently. Mol. Microbiol. *62*, 320–330.

Soto, G.E., and Hultgren, S.J. (1999). Bacterial adhesins: common themes and variations in architecture and assembly. J. Bacteriol. *181*, 1059–1071.

Springer, T.A. (2006). Complement and multifaceted functions of VWA and integrin I domains. Structure *14*, 1611–1616.

Storoni, L.C., McCoy, A.J., and Read, R.J. (2004). Likelihood-enhanced fast rotation functions. Acta Crystallogr. D Biol. Crystallogr. *60*, 432–438.

Telford, J.L., Barocchi, M.A., Margarit, I., Rappuoli, R., and Grandi, G. (2006). Pili in Gram-positive pathogens. Nat. Rev. Microbiol. *4*, 509–519.

Ton-That, H., and Schneewind, O. (2004). Assembly of pili in Gram-positive bacteria. Trends Microbiol. *12*, 228–234.

Tuckwell, D., Calderwood, D.A., Green, L.J., and Humphries, M.J. (1995). Integrin α2 I-domain is a binding site for collagens. J. Cell Sci. *108*, 1629–1637.

Vorup-Jensen, T., Ostermeier, C., Shimaoka, M., Hommel, U., and Springer, T. (2003). Structure and allosteric regulation of the αxβ2 integrin I domain. Proc. Natl. Acad. Sci. USA *100*, 1873–1878.

Note Added in Proof

As this manuscript was going to press, Hilleringman and coworkers published work describing the electron microscopy anaylsis of streptococcal pili, where RrgA is clearly located at the tip of the pilus fibril (Hilleringman et al. (2009) EMBO J. 28, 3921-3930.).