# The *Corynebacterium diphtheriae* shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds

Hae Joo Kang<sup>a</sup>, Neil G. Paterson<sup>a</sup>, Andrew H. Gaspar<sup>b</sup>, Hung Ton-That<sup>b,c,1</sup>, and Edward N. Baker<sup>a,1</sup>

<sup>a</sup>Maurice Wilkins Centre for Molecular Biodiscovery and School of Biological Sciences, University of Auckland, Auckland 1020, New Zealand; <sup>b</sup>Department of Molecular, Microbial, and Structural Biology, University of Connecticut Health Center, Farmington, CT 06030; and <sup>c</sup>Department of Microbiology and Molecular Genetics, University of Texas Health Science Center, Houston, TX 77030

Edited by David S. Eisenberg, University of California, Los Angeles, CA, and approved August 12, 2009 (received for review June 17, 2009)

Cell-surface pili are important virulence factors that enable bacterial pathogens to adhere to specific host tissues and modulate host immune response. Relatively little is known about the structure of Gram-positive bacterial pili, which are built by the sortase-catalyzed covalent crosslinking of individual pilin proteins. Here we report the 1.6-Å resolution crystal structure of the shaft pilin component SpaA from Corynebacterium diphtheriae, revealing both common and unique features. The SpaA pilin comprises 3 tandem Ig-like domains, with characteristic folds related to those typically found in non-pilus adhesins. Whereas both the middle and the C-terminal domains contain an intramolecular Lys-Asn isopeptide bond, previously detected in the shaft pilins of Streptococcus pyogenes and Bacillus cereus, the middle Ig-like domain also harbors a calcium ion, and the C-terminal domain contains a disulfide bond. By mass spectrometry, we show that the SpaA monomers are cross-linked in the assembled pili by a Lys-Thr isopeptide bond, as predicted by previous genetic studies. Together, our results reveal that despite profound dissimilarities in primary sequences, the shaft pilins of Gram-positive pathogens have strikingly similar tertiary structures, suggesting a modular backbone construction, including stabilizing intermolecular and intramolecular isopeptide bonds.

crystal structure | polymerization | bacterial pilus | mass spectrometry | pilin motif

**P**ili are long, thin protein assemblies that extend from the cell surface of many bacteria and play pivotal roles in colonization and pathogenesis. Like Gram-negative bacteria, many Gram-positive pathogens express pili on their surface (1-3). These structures have aroused great interest because of their direct roles in infection and pathogenesis and their importance as vaccine candidates (2, 3). They also use covalent isopeptide (amide) bonds, both intermolecular and intramolecular, to give strength and stability, and thus present a new paradigm among protein polymers (4-6).

Unlike Gram-negative pili, whose subunits associate via noncovalent interactions, these Gram-positive pili are formed by covalent polymerization of pilin subunits, orchestrated by transpeptidase enzymes called sortases (6, 7). The general principles of assembly were first established through studies on the SpaA pili expressed by Corynebacterium diphtheriae. These pili, encoded by the gene cluster spaA-spaB-srtA-spaC, comprise a polymeric shaft formed by SpaA, SpaC located at the tip, and SpaB found at the base and occasionally along the shaft (6, 8-10). In our current model, successive major pilin SpaA subunits are joined by the action of the pilus-specific sortase SrtA. Cleavage between Thr-494 and Gly-495 of the LPXTG motif near the SpaA C terminus is followed by presumed amide bond formation between the new C terminus and Lys-190 from a conserved YPKN pilin motif in the next subunit (6). Finally, the entire assembly is covalently attached to the cell wall peptidoglycan by a housekeeping sortase (9, 11). Presumably, the tip pilin SpaC is linked to the SpaA shaft via the same reaction as recently identified for the tip pilin BcpB of *Bacillus cereus* (12). It remains unclear how the minor pilin SpaB is incorporated into the pilus structure, although recent evidence indicates that SpaB forms the basal subunit for tethering the pilus to the cell wall and that it, too, is attached to the polymer via a specific Lys residue (9).

A major step forward in understanding Gram-positive pilus structure and assembly came with the structural analysis of the major pilin Spy0128 from *Streptococcus pyogenes* (5). Spy0128 does not have a recognizable pilin motif, but the crystal structure revealed columns of molecules resembling a putative polymer assembly and identified a candidate lysine; this was then confirmed by mass spectral analysis of native *S. pyogenes* pili. The structure also revealed unexpected internal crosslinks in the form of self-generated isopeptide bonds, 1 in each domain of the 2-domain structure, joining Lys and Asn side chains. These are strategically located to give strength and stability to the pilus assembly.

The major pilins of different Gram-positive bacteria show wide variations in size and sequence, making it difficult to predict whether the structural principles seen for *S. pyogenes* apply also to other Gram-positive pili. Here we present the high-resolution crystal structure of SpaA, the archetypal major pilin from *C. diphtheriae*. This reveals a modular structure comprising 3 tandem Ig-like domains, 2 of which contain internal Lys–Asn isopeptide bonds like those in Spy0128. We also confirm, by mass spectrometry, the identity of the lysine used in polymerization and note a pilus-like assembly of SpaA molecules in the crystal. These results point to a likely common architecture for the backbones of many Gram-positive pili and consolidate a new paradigm for the structure, stability, and assembly of these remarkable covalent polymers.

#### Results

**Structure Determination.** A construct comprising residues 53–486 of *C. diphtheriae* SpaA was expressed in *Escherichia coli*, purified, and crystallized. This lacks residues 1–52 encompassing the signal peptide and ends 4 residues before the sortase-recognition LPXTG motif. The crystal structure, with 1 SpaA molecule per asymmetric unit, was solved by single wavelength anomalous dispersion methods and refined at 1.6-Å resolution (R = 19.3%, R<sub>free</sub> = 22.0%) [supporting information (SI) Table S1]. Only the

Author contributions: H.J.K., N.G.P., A.H.G., H.T.-T., and E.N.B. designed research; H.J.K., N.G.P., and A.H.G. performed research; H.J.K. contributed new reagents/analytic tools; H.J.K., N.G.P., and E.N.B. analyzed data; and H.J.K., H.T.-T., and E.N.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 3HR6).

<sup>&</sup>lt;sup>1</sup>To whom correspondence may be addressed. E-mail: ton-that.hung@uth.tmc.edu or ted.baker@auckland.ac.nz.

This article contains supporting information online at www.pnas.org/cgi/content/full/ 0906826106/DCSupplemental.



Fig. 1. Crystal structure of SpaA. (A) End-to-end stacking of successive molecules, in which the C-domain of one molecule, SpaAn, packs against the N-domain of the next,  $SpaA_{n+1}$ . For each molecule, the N-domain is in gold, M-domain green, and C-domain blue. The M-domain Ca<sup>2+</sup> ion is shown as a gray sphere. Residues that form isopeptide bonds in the M- and C-domains are shown in red stick mode, and the disulfide bond in the C-domain is in yellow. The pilin motif lysine, Lys-190, in each molecule is labeled K. A black bar shows the distance between the C terminus of one molecule and Lys-190 on the next. (B) Schematic representation of the fold of each domain colored as in A. Helices are labeled 1 to 3, and  $\beta$ -strands A to W. The isopeptide bond crosslinks are shown with red bars and the disulfide bond with a yellow bar. An arrowhead points to Lys-190. Strands and loops connected by inter- or intramolecular isopeptide bonds are highlighted on a beige background. (C) Close view of the packing of adjacent molecules. Lys-190 projects upward, between the disordered AB loop (gray line) and the N-domain, toward the C terminus of the next molecule (C). A broken line shows where the 10 missing C-terminal residues would bridge the  $\approx$ 19-Å gap to Lys-190 on the next molecule. The conserved Trp-181 of the pilin motif is in the interface between molecules.

N-terminal Glu-53, an external loop 69–79, and the C-terminal residues 485–486 could not be modeled for lack of interpretable electron density.

**Modular Structure of SpaA.** SpaA is folded into 3 tandem Ig-type domains, giving an elongated molecule of  $\approx 105$  Å in length. Whereas the middle domain (M-domain; residues 193–351) and the C-terminal domain (C-domain; residues 352–484) are arranged linearly and share an extended strand P (Fig. 1), the N-terminal domain (N-domain; residues 54–192) sits on the M-domain at an angle of  $\approx 20^{\circ}$  to the long axis of the molecule. The SpaA molecules pack in columns through the crystal, in a manner resembling a pilus assembly; the N-domain of each molecule abuts against the C-domain of the next (Fig. 1*A*).



**Fig. 2.** Internal isopeptide bonds in SpaA. Residues involved in bond formation are in stick mode, colored by atom type, with surrounding hydrophobic residues also shown. Hydrogen bonds are shown with broken lines, distances in Å. The electron density is from an ( $F_{obs} - F_{cal}$ ) · Phi<sub>cal</sub> map, contoured at  $3\sigma$ . (A) The M-domain isopeptide bond formed between Lys-199 and Asn-321, with catalytic Asp-241. (B) The C-domain isopeptide bond between Lys-363 and Asn-482, with catalytic Glu-446.

The N- and C-domains both have the inverse IgG fold first described for the CnaB domains of the collagen binding protein Cna from *Staphylococcus aureus* (13). This comprises a  $\beta$ -sandwich of 7 strands. When superimposed, the 2 CnaB-type domains of SpaA show an rmsd in C $\alpha$  positions of 2.2 Å over 88 equivalent residues and share 14% sequence identity. A distinctive feature of the N-domain is the presence of 2 helices between strands B and C that partially cover one side of the core, whereas the C-domain uniquely contains an elongated  $\beta$ -ribbon, formed by strands S and T, running toward the M-domain. In contrast, the M-domain of SpaA has the CnaA fold, first seen in the N2 domain of *S. aureus* CnaA (14). This comprises 9  $\beta$ -strands that form a partially open  $\beta$ -barrel.

The closest structural homologues of the N- and C-domains of SpaA are the 2 CnaB-type domains of the Streptococcus agalactiae minor pilin GBS52 (15). In particular, the SpaA C-domain has significant sequence identity (25%) and structural similarity (rmsd 1.7 Å over 91 equivalent C $\alpha$  atoms) with the N2 domain of GBS52. The 2 domains of the S. pyogenes major pilin Spy0128 also share the same CnaB-type fold, albeit with some elaborations (5). Structural superpositions of the Spy0128 domains on to the SpaA N- and C-domains give rmsds in  $C\alpha$  positions ranging from 2.5 to 3.1 Å and sequence identities ranging from 3% to 17%. The M-domain shows strong structural homology with the N2 domain of CnaA, despite minimal sequence identity ( $\approx 8\%$ ); the rmsd is 3.4 Å over 123 equivalent C $\alpha$  atoms. Other similar CnaA-type domains include the N3 domain of S. aureus clumping factor A and the N2 domain of the Enterococcus faecalis collagen-binding protein Ace (16, 17).

**Internal Isopeptide Bonds and Other Stabilizing Features.** The M- and C-domains of SpaA both contain stabilizing internal isopeptide bonds, formed by intramolecular reaction between the Lys  $\varepsilon$ -amino group and the carboxyamide group of Asn. These were clearly apparent in the initial experimentally phased electron density map, in which continuous density linked the side chains of Lys-199 and Asn-321 in the M-domain and Lys-363 and Asn-482 in the C-domain (Fig. 2). The existence of these Lys–Asn isopeptide bonds was confirmed by electrospray ionization–time-of-flight mass spectrometry. The protein  $M_r$  was measured as 46,795.4 Da,  $\approx$ 34 Da less than that calculated from the amino acid sequence, 46,829.6 Da, consistent with loss of 2 units of NH<sub>3</sub> through formation of 2 isopeptide bonds. Confirmation was obtained by digestion of the recombinant protein and analysis by liquid chromatography–tandem mass spectrometry

(LC-MS/MS). Parent ions with mass-to-charge ratio (m/z) 996.8<sup>3+</sup> and 1067.5<sup>3+</sup> contained the M-domain isopeptide bond between Lys-199 and Asn-321, and a parent ion with m/z 751.7<sup>3+</sup> contained the C-domain linkage between Lys-363 and Asn-482 (Tables S2 and S3).

Similar Lys-Asn isopeptide bonds were first observed in the structure of Spy0128, where an associated Glu residue was shown to be essential for the intramolecular reaction to occur (5). In SpaA, the same role is played by Asp-241 for the M-domain isopeptide bond and Glu-446 for the C-domain bond (Fig. 2).

Two isomeric forms of isopeptide bond are found in SpaA. The M-domain bond (Lys-199-Asn-321) has a cis configuration, as for both isopeptide bonds of Spy0128, allowing its NH and O moieties to form a bidentate hydrogen-bonded interaction with the Asp-241 carboxyl group (Fig. 2). In contrast, the C-domain bond (Lys-363-Asn-482) has a trans configuration and only a single hydrogen bond with the carboxyl group of Glu-446. The hydrogen bonding patterns imply that both carboxyl groups are protonated. Both isopeptide bonds are located in the interior of their respective domains, surrounded by hydrophobic residues. Both also stack against aromatic residues, Tyr-219 in the M-domain and Phe-365 in the C-domain (Fig. 2). Other surrounding hydrophobic residues include Phe-306, Val-352, Val-221, and Leu-243 in the M-domain and Phe-378, Ile-361, Ile-480, and Ala-376 in the C-domain. This hydrophobic environment favors a nonprotonated Lys amino group and protonated Glu/Asp carboxyl group, thus facilitating the intramolecular reaction (5, 18).

The SpaA structure has 2 other notable stabilizing features. In the M-domain a metal binding site is formed by the loop joining strands H and I, with the metal ion coordinated by 8 oxygen atoms, from Asp-204, Asp-205, Gln-208, Gly-210, Glu-215, and 2 water molecules (Fig. 1). The coordination environment and average metal-ligand bond length (2.48 Å) are indicative of a  $Ca^{2+}$  ion, presumably cell derived. This is a unique feature, not seen before in any other CnaB- or CnaA-like domain, and its persistence despite the use of 1 mM EDTA in buffers implies a high affinity. In the C-domain, a disulfide bond joins Cys-383 on strand Q to Cys-443 on strand U (Fig. 1). The electron density shows that this bond is incompletely formed, with approximately 40% of molecules having both Cys reduced. This may result from the DTT needed for tag cleavage, and we anticipate that the disulfide would be fully formed in vivo, in the oxidizing extracellular environment.

Interestingly, whereas SpaA has no internal isopeptide bond in its N-domain, such a bond is present in the N-domain of the 3-domain major pilin BcpA from *B. cereus*, joining Lys-37 and Asn-163 (4). These residues are conserved in other pilins but replaced by Ala-61 and His-191, respectively, in SpaA (Fig. S1). In the SpaA structure, the Ala and His side chains are close enough such that if replaced by Lys and Asn, as in other pilins, an isopeptide bond could be formed. A conserved Glu that could catalyze Lys–Asn bond formation is present in the other pilins but in SpaA is replaced by Gln-153, positioned close to Ala-61 and His-191 (Fig. S2).

Sequence Elements Implicated in Pilus Assembly. Sequence comparisons and mutagenesis have identified 2 conserved sequence motifs that contain residues essential for pilus assembly. The pilin motif (consensus WxxxVxVYPK) contains the lysine that is joined to the C terminus of the next molecule during polymerization; Lys-190 in SpaA. We sought to confirm the role of Lys-190 in SpaA polymer formation, using mass spectrometry. Purified SpaA pilus polymers were separated by SDS-PAGE and digested in-gel with trypsin and AspN endopeptidase. The digestion products were analyzed by LC-MS/MS and the intersubunit amide bond identified from a peptide peak with m/z585.7<sup>3+</sup>. The  $M_r$  of this peptide corresponded exactly with that expected, allowing for the loss of 18 Da due to water elimination



**Fig. 3.** Intersubunit isopeptide bonds in SpaA. Fragmentation spectra of the parent ion at m/z 585.7<sup>3+</sup> containing the intersubunit bond between SpaA Lys-190 and Thr-494 are shown. Ion types are indicated, and internal ions are shown in italics. The structure of crosslinked peptide is shown above the spectra. Daughter ions produced during MS/MS of these peptides are summarized in Table S4.

during amide bond formation. The fragment ion spectra uniquely identified the peptides surrounding the pilin motif Lys-190 and the sortase-cleaved C-terminal Thr-494, respectively (Fig. 3 and Table S4).

In the SpaA structure, the pilin motif is located on  $\beta$ G, the last strand of the N-domain. Lys-190 is close to the point where  $\beta G$ crosses to the M-domain, becoming  $\beta$ H (Fig. 1). Nine residues before Lys-190 is the conserved Trp-181, the first residue of the pilin motif, and 9 residues after is Lys-199, which forms the M-domain isopeptide bond. The side chain of Lys-190 projects into a cleft between the main body of the N-domain and a mobile loop, residues 63-83 (Fig. 1C). Head-to-tail packing of molecules in the crystal places the  $\varepsilon$ -amino group of Lys-190  $\approx$ 19 Å from the C terminus of the next molecule, ample distance to accommodate the 10 missing residues between the last modeled residue, Lys-484, and the true sortase cleavage site, Thr-494. It also places Trp-181 in the interface between the 2 molecules (Fig. 1C): the conservation of this residue supports the idea that the crystal packing models the true biologic assembly. The fact that the M-domain isopeptide bond closely follows Lys-190, on the same extended  $\beta$ -strand, suggests why polymer formation is abrogated by deletion of the equivalent isopeptide bond in B. cereus BcpA (4); local structural destabilization could prevent proper presentation of the essential lysine to the sortase.

The second sequence motif implicated in assembly is the E-box motif (consensus YxLxETxAPxGY). This contains a conserved glutamate, Glu-446 in SpaA, which is essential for the incorporation of the minor pilins SpaB and SpaC (7). Intriguingly, Glu-446 proves to be the catalytic Glu that mediates formation of the Lys-363–Asn-462 intramolecular bond. Because Glu-446 is internal, we infer that it is the stability imparted by the intramolecular crosslink that is essential for minor pilin incorporation.

#### Discussion

The discovery of thin, hair-like pili on the surface of *C. diphtheriae* in 2003, and their characterization as covalent polymers, was a milestone in understanding colonization and infection by Gram-positive bacteria (6). Similar pilus assemblies are found for such important human pathogens as Group A and B streptococci, *Streptococcus pneumoniae*, and *B. cereus* (19–21), but the pilin subunits involved show large variations in size and sequence that mask any possible structural homology.

The present structure for the *C. diphtheriae* major pilin SpaA resolves this question, revealing a modular assembly that utilizes Ig-like domains similar to those used in the *S. pyogenes* major pilin Spy0128, despite very low sequence identity. These domains correspond to 2 types of Ig-like domain, the CnaB and CnaA folds that are widely used in the cell-surface adhesins known as MSCRAMMS (microbial surface components recognizing adhesive matrix molecules) (22). Prototype CnaB and CnaA domains are present in the multidomain *S. aureus* Cna protein, which has a structural B-region with repeating doublets of CnaB domains, preceded by a collagen-binding A-region with 2 CnaA domains (22, 23).

Spy0128, the major pilin of *S. pyogenes*, is one of the smallest shaft pilins (32.5 kDa) and comprises 2 tandem CnaB domains (5). SpaA, in contrast, is significantly larger (47 kDa) and has a single CnaA-type domain, the M-domain, inserted between 2 CnaB-type domains. This mosaic architecture suggests an evolutionary process in which copies or pieces of older genes are assembled to form new genes. It seems likely that all of the major pilins of sortase-assembled Gram-positive pili conform to the same structural principles. In many cases, for example the major pilins SpaD and SpaH that form the 2 other types of pilus produced by *C. diphtheriae*, sufficient sequence identity exists to infer similar structures; these share  $\approx 24\%$  identity with SpaA, including the intramolecular isopeptide bond-forming residues, the pilin motif, and the Cys residues. In others, such as Spy0128, there is much less sequence similarity.

Sequence comparisons with the major pilins from other Gram-positive bacteria, such as *S. pneumoniae*, *S. agalactiae*, and *B. cereus*, reveal N- and C-terminal regions that are similar to the N- and C-domains of SpaA, including the N-domain pilin motif and the C-domain isopeptide bond-forming residues. The middle regions are more variable; some may form CnaA-type domains like the SpaA M-domain, with substantial insertions/deletions, whereas others may have CnaB-type middle domains or adopt entirely different folds.

There is growing evidence that this modular architecture extends also to the minor pilin subunits. GBS52, a minor pilin from *S. agalactiae* that seems to correspond functionally to SpaB, comprises 2 CnaB-type domains (15), and its C-terminal N2 domain closely resembles the SpaA C-domain, including the internal isopeptide bond. Sequence comparisons show that the *S. pyogenes* minor pilin Cpa also has a C-terminal domain homologous with the C-domain of the *S. pyogenes* major pilin Spy0128, again including an internal isopeptide bond (5). This leads to the concept that incorporation of the minor pilins is facilitated by their structural resemblance to the pilins that comprise the polymeric shaft.

The 2 internal isopeptide bonds in SpaA confirm that such crosslinks are a common feature of Gram-positive pili. First discovered in the crystal structure of Spy0128 and confirmed by mass spectral analysis of native GAS pili (5), they have also been found in the *B. cereus* major pilin BcpA, which has 3 such bonds (4). The combined sequence and structural data from these 3 characterized major pilins now enable internal isopeptide bonds to be inferred from the sequences of other major pilins. Reevaluation of the crystal structures of GBS52 and the A- and B-domains of *S. aureus* Cna further showed that similar isopeptide bonds are also present in minor pilins and adhesins (5).

The SpaA structure further defines the determinants of intramolecular isopeptide bond formation. Mutagenesis has shown that a catalytic carboxyl group is essential for bond formation (4, 5, 24); this can be Asp or Glu, as shown by the use of Asp-241 in the M-domain and Glu-446 in the C-domain. The isopeptide moiety can have either *cis* or *trans* configuration, with

corresponding bidentate or monodentate interaction with the essential carboxyl side chain. The main requirement is proximity of the Lys–Asn pair and a hydrophobic environment in which both the Lys and Asp/Glu are uncharged. The locations of these isopeptide bonds seem to be characteristic of the folds of the domains in which they occur, and hence probably reflect their evolutionary history. In the CnaB-like C-domain of SpaA, the isopeptide bond joins the first and last  $\beta$ -strands, as in both CnaB-type domains of Spy0128. In contrast, in the CnaA-type M-domain, the isopeptide bond bridges the first and second-last strands, linking 2 opposing  $\beta$ -sheets.

Given the widespread occurrence of internal isopeptide crosslinks in the cell-surface proteins of Gram-positive bacteria, what is their structural and functional importance? We have shown that the intramolecular isopeptide bonds in Spy0128 strongly enhance thermodynamic stability and resistance to proteolysis (24). SpaA additionally contains a disulfide bond, conserved in the other C. diphtheriae major pilins, which might be expected to further enhance stability. The major pilins of S. pyogenes, B. cereus, and S. pneumoniae lack Cys residues, however, and given that many Gram-positive bacteria lack the disulfide formation machinery of Gram-negative bacteria (25, 26), we speculate that isopeptide bonds have evolved as an alternative means of stabilization. As amide bonds they would also be less prone to chemical disruption than disulfide bonds, a property that may be important for such thin, exposed assemblies, which do not seem to form higher-order bundles.

We further hypothesize that their strategic location gives mechanical (force-bearing) stability. Both in these pili, typified by Spy0128 and SpaA, and in multidomain adhesins such as Cna, an almost linear chain of covalent connectivity can be traced along the long axis (5, 27). In SpaA this begins with Lys-190, the site of attachment to the preceding subunit, and extends through the Mand C-domains to the next intermolecular linkage, possibly explaining why the N-domain does not require an isopeptide bond. The attachment of adhesins to host cells subjects them to significant tensile stress, and their structures are thought to have evolved both to withstand stress and to use it to optimize binding (28).

There is also growing evidence that the structural stabilization conferred by the internal isopeptide bonds can be critical for molecular recognition. Defective sortase recognition or failure to present the Lys residue of the pilin motif appropriately would explain the loss of pilus assembly when intramolecular isopeptide bonds in SpaA or BcpA are deleted by mutation of the catalytic Asp/Glu. The isopeptide bonds in CnaA and GBS52 may similarly influence the binding of partner molecules, because their isopeptide bond-containing domains are important for specific binding of collagen (CnaA) and capable of binding to human pulmonary epithelial cells (GBS52) (15, 29).

Finally, an intriguing feature of the crystal structures of both SpaA and Spy0128 is the way the pilin molecules pack end-to-end in columns in the crystal, resembling assembled pili. The intermolecular contacts do not seem to be particularly extensive (buried surface 850 Å<sup>2</sup> for Spy0128 and 814 Å<sup>2</sup> for SpaA), yet in each case the molecules pack such that the critical lysine residue (Lys-161 in Spy0128, Lys-190 in SpaA) is brought close to the C terminus of the next molecule. Importantly, in SpaA the conserved Trp-181 of the pilin motif forms part of this interface, consistent with a role in oligomer formation. The roles of the conserved Tyr-188 and Pro-189 of the pilin motif are less clear, but patches of electron density around them may represent parts of the unmodeled AB loop. These residues may interact transiently with this loop, which could in turn mediate the pilin–pilin and/or pilin–sortase interactions.

#### **Materials and Methods**

**Cloning and Protein Purification.** DNA encoding amino acids 53–486 of SpaA from *C. diphtheriae* was amplified by PCR from genomic DNA, cloned, over-

expressed in *E. coli* as an N-terminally His-tagged protein, and purified by nickel-affinity chromatography (30). After His-tag removal and final size-exclusion chromatography, the protein was concentrated to 100 mg/mL in 10 mM Tris-HCl (pH 8.0) and 50 mM NaCl. Selenomethionine (SeMet)-substituted SpaA was produced using the methionine biosynthesis inhibition method (31) and similarly purified, but with 5 mM DTT and 1 mM EDTA in the final gel filtration buffer.

**Crystallization and Structure Determination.** Crystals were grown in sitting drops comprising 100 nL protein (100 mg/mL) and 100 nL precipitant. The best native SpaA crystals were obtained with 20% PEG 3350, 0.1 M Nal, and 0.1 M NaF as precipitant and SeMet-SpaA crystals with 20% PEG 3350 and 0.2 M Na formate. Crystals were flash-cooled without further cryoprotection. X-ray diffraction data were collected on beamline PX1 at the Australian Synchrotron, to 1.6 Å and 1.8 Å resolutions, respectively, for native and SeMet-SpaA crystals. Data were processed and scaled with MOSFLM and SCALA (32). All 4 Se atoms were located by SHELX (33) with refinement and phase determination in autoSHARP (34). Density modification and model building with PHENIX (35, 36) placed 408 of 436 residues, and model building was completed using COOT (37). The model was refined using REFMAC (38). Data collection, phasing, and refinement statistics are in Table S1. Structural superpositions were done with SSM (39).

Isolation of SpaA Pili. Engineered SpaA pili were produced by transformation of the plasmid pAG153, encoding SpaA and SrtA, into a *C. diphtheriae* strain

- 1. Ton-That H, Schneewind O (2004) Assembly of pili in Gram-positive bacteria. *Trends Microbiol* 12:228–234.
- Proft T, Baker EN (2009) Pili in Gram-negative and Gram-positive bacteria—structure, assembly and their role in disease. *Cell Mol Life Sci* 66:613–635.
- Telford JL, Barocchi MA, Margarit I, Rappuoli R, Grandi G (2006) Pili in gram-positive pathogens. Nat Rev Microbiol 4:509–519.
- Budzik JM, et al. (2008) Amide bonds assemble pili on the surface of bacilli. Proc Natl Acad Sci USA 105:10215–10220.
- Kang HJ, Coulibaly F, Clow F, Proft T, Baker EN (2007) Stabilizing isopeptide bonds revealed in Gram-positive bacterial pilus structure. *Science* 318:1625–1628.
- Ton-That H, Schneewind O (2003) Assembly of pili on the surface of Corynebacterium diphtheriae. Mol Microbiol 50:1429–1438.
- Ton-That H, Marraffini LA, Schneewind O (2004) Sortases and pilin elements involved in pilus assembly of Corynebacterium diphtheriae. Mol Microbiol 53:251–261.
- Gaspar AH, Ton-That H (2006) Assembly of distinct pilus structures on the surface of Corynebacterium diphtheriae. J Bacteriol 188:1526–1533.
- Mandlik A, Das A, Ton-That H (2008) 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.
- Swierczynski A, Ton-That H (2006) Type III pilus of corynebacteria: Pilus length is determined by the level of its major pilin subunit. J Bacteriol 188:6318-6325.
- 11. Swaminathan A, et al. (2007) Housekeeping sortase facilitates the cell wall anchoring of pilus polymers in *Corynebacterium diphtheriae*. *Mol Microbiol* 66:961–974.
- Budzik JM, Oh SY, Schneewind O (2009) Sortase D forms the covalent bond that links BcpB to the tip of *Bacillus cereus* pili. J Biol Chem 284:12989–12997.
- 13. Deivanayagam CC, et al. (2000) Novel fold and assembly of the repetitive B region of the *Staphylococcus aureus* collagen-binding surface protein. *Structure* 8:67–78.
- Symersky J, et al. (1997) Structure of the collagen-binding domain from a Staphylococcus aureus adhesin. Nat Struct Biol 4:833–838.
- Krishnan V, et al. (2007) An IgG-like domain in the minor pilin GBS52 of Streptococcus agalactiae mediates lung epithelial cell adhesion. Structure 15:893–903.
- Deivanayagam CC, et al. (2002) A novel variant of the immunoglobulin fold in surface adhesins of *Staphylococcus aureus*: Crystal structure of the fibrinogen-binding MSCRAMM, clumping factor A. *EMBO J* 21:6660–6672.
- Liu Q, et al. (2007) The Enterococcus faecalis MSCRAMM ACE binds its ligand by the collagen hug model. J Biol Chem 282:19629–19637.
- Wikoff WR, et al. (2000) Topologically linked protein rings in the bacteriophage HK97 capsid. Science 289:2129–2133.
- 19. Barocchi MA, et al. (2006) A pneumococcal pilus influences virulence and host inflammatory responses. *Proc Natl Acad Sci USA* 103:2857–2862.
- 20. Budzik JM, Schneewind O (2006) Pili prove pertinent to enterococcal endocarditis. J Clin Invest 116:2582–2584.

that lacks *spaA* and *srtA*, followed by expression and purification of the polymers (6). These procedures are described more fully in *SI Materials and Methods*. The use of a His-tagged SpaA construct with a SrtA construct lacking 13 C-terminal residues means that the engineered SpaA pili are secreted into the culture medium and can be purified by nickel-affinity chromatography as previously described (6).

Proteolytic Digestion and Mass Spectral Analyses. Purified SpaA pili and recombinant SpaA protein were digested and analyzed according to previous protocols (5). Briefly, SDS-PAGE gel bands containing recombinant SpaA or SpaA pili were cut out and incubated with trypsin (Promega) followed by AspN endopeptidase (Roche). Peptides in the *m*/*z* range 300–1,600 were analyzed using a Q-STAR XL Hybrid MS/MS system (Applied Biosystems). Searches against SpaA sequence using Mascot search engine version 2.0.05 (Matrix Science) identified linear peptides, and the unmatched peptides were then searched manually to identify those containing noncontiguous peptides crosslinked by isopeptide bonds, either intramolecular or intermolecular. Full details are in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Tom Caradoc-Davies for help with data collection, Martin Middleditch for help with mass spectrometry, and Asis Das for critical insights. This work was supported by the Health Research Council and the Marsden Fund of New Zealand (E.N.B.) and National Institutes of Health Grant Al061381 (to H.T.-T). Data collection was undertaken on the PX1 beamline at the Australian Synchrotron, Victoria, Australia, with support from the New Zealand Synchrotron Group Ltd.

- 21. Lauer P, et al. (2005) Genome analysis reveals pili in Group B Streptococcus. Science 309:105.
- Patti JM, Allen BL, McGavin MJ, Hook M (1994) MSCRAMM-mediated adherence of microorganisms to host tissues. Annu Rev Microbiol 48:585–617.
- 23. Rich RL, et al. (1998) Domain structure of the *Staphylococcus aureus* collagen adhesin. *Biochemistry* 37:15423–15433.
- 24. Kang HJ, Baker EN (2009) Intramolecular isopeptide bonds give thermodynamic and proteolytic stability to the major pilin protein of *Streptococcus pyogenes*. J Biol Chem, in press.
- Dutton RJ, et al. (2008) Bacterial species exhibit diversity in their mechanisms and capacity for protein disulfide bond formation. *Proc Natl Acad Sci USA* 105:11933– 11938.
- 26. Heras B, et al. (2009) DSB proteins and bacterial pathogenicity. Nat Rev Microbiol 7:215–225.
- 27. Yeates TO, Clubb RT (2007) How some pili pull. Science 318:1558-1559.
- Sokurenko EV, Vogel V, Thomas WE (2008) Catch-bond mechanism of force-enhanced adhesion: counterintuitive, elusive, but...widespread? Cell Host Microbe 4:314–323.
- Zong Y, et al. (2005) A 'Collagen Hug' model for Staphylococcus aureus CNA binding to collagen. EMBO J 24:4224–4236.
- Kang HJ, Paterson NG, Baker EN (2009) Expression, purification, crystallization and preliminary crystallographic analysis of SpaA, a major pilin from Corynebacterium diphtheriae. Acta Crystallogr Sect F Struct Biol Cryst Commun, in press.
- Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J Mol Biol 229:105–124.
- 32. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50:760–763.
- Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr 58:1772–1779.
- Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. Methods Mol Biol 364:215–230.
- 35. Adams PD, et al. (2004) Recent developments in the PHENIX software for automated crystallographic structure determination. J Synchrotron Radiat 11:53–55.
- 36. Terwilliger TC (2003) Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr D Biol Crystallogr* 59:38–44.
- Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60:2126–2132.
- Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53:240–255.
- Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr 60:2256–2268.

#### BIOCHEMISTRY

Correction for "The *Corynebacterium diphtheriae* shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds," by Hae Joo Kang, Neil G. Paterson, Andrew H. Gaspar, Hung Ton-That, and Edward N. Baker, which appeared in issue 40, October 6, 2009, of *Proc Natl Acad Sci USA* (106:16967–16971; first published September 21, 2009; 10.1073/pnas.0906826106).

The authors note that, due to a printer's error, on page 16967, the keyword "**mas spectromrtry**" should instead have appeared as "**mass spectrometry**." The online version has been corrected.

www.pnas.org/cgi/doi/10.1073/pnas.0911293106

#### **CELL BIOLOGY**

Correction for "CBP and p300 are cytoplasmic E4 polyubiquitin ligases for p53," by Dingding Shi, Marius S. Pop, Roman Kulikov, Ian M. Love, Andrew Kung, and Steven R. Grossman, which appeared in issue 38, September 22, 2009, of *Proc Natl Acad Sci USA* (106:16275–16280; first published September 4, 2009; 10.1073/pnas.0904305106).

The authors note that the author name Andrew Kung should instead have appeared as Andrew L. Kung. The online version has been corrected. The corrected author line and author contributions footnote appear below.

#### Dingding Shi, Marius S. Pop, Roman Kulikov, Ian M. Love, Andrew L. Kung, and Steven R. Grossman

Author contributions: D.S., M.S.P., R.K., and S.R.G. designed research; D.S., M.S.P., R.K., and I.M.L. performed research; A.L.K. contributed new reagents/analytic tools; D.S., M.S.P., R.K., and S.R.G. analyzed data; and S.R.G. wrote the paper.

www.pnas.org/cgi/doi/10.1073/pnas.0910983106

#### ECOLOGY

Correction for "Adaptive shell color plasticity during the early ontogeny of an intertidal keystone snail," by Patricio H. Manríquez, Nelson A. Lagos, María Elisa Jara, and Juan Carlos Castilla, which appeared in issue 38, September 22, 2009, of *Proc Natl Acad Sci USA* (106:16298–16303; first published September 2, 2009; 10.1073/pnas.0908655106).

The authors note that on page 16298, right column, first full paragraph, the first sentence appeared incorrectly in part. "In the rocky intertidal habitats dominated by mussel beds (dark colored) and barnacle stands (light colored), we found that more than 95% of early postmetamorphic stages of C. concholepas ( $\approx 2$ to 20 mm periostomal length, PL) showed a striking colormatching with the most abundant prey (Fig. 2A and B)" should instead have appeared as "In the rocky intertidal habitats dominated by mussel beds (dark colored) and barnacle stands (light colored), we found that more than 95% of early postmetamorphic stages of C. concholepas ( $\approx 2$  to 20 mm peristomal length, PL) showed a striking color-matching with the most abundant prey (Fig. 2 A and B)." Additionally, the authors note that on page 16301, right column, first full paragraph, the second sentence did not appear in full due to a printer's error. The sentence should instead have appeared as "This modulation of the shell coloration is restricted to snails of small size (ca. less than 3 cm), coincidentally the size at which C. concholepas appear to escape predation by crabs (author's personal observations)." These errors do not affect the conclusions of the article.

www.pnas.org/cgi/doi/10.1073/pnas.0910777106

# **Supporting Information**

# Kang et al. 10.1073/pnas.0906826106

### **SI Materials and Methods**

Construction of Plasmid pAG153 Encoding SpaA and SrtA. Primers SpaA-498-5 (aaaagatcttgagttcgattggcttttttc) and SpaA-KpnI-3 (aaaggta ccgttctttttcttgttgtcgatc) were used to PCR-amplify the promoter and coding sequence of spaA (devoid of the sequence encoding the last 39 C-terminal residues) from plasmid pHTT11 DNA while appending BglII and KpnI sites for cloning purposes (1). Primers Kpn-MH6-5 (aaaggtaccatgcatcaccatcaccatcacgcc ggatttgaactgccac) and SrtA $_{\Delta 13}$ -Bg-3 (cgcagatctttaggtgatgatgatgatgatg gcgatg) were used to amplify a DNA fragment containing the coding sequence for 6 histidine residues and the last 39 residues of *spaA* fused to the coding sequence of *srtA* lacking the coding sequences for the last 13 residues, while appending KpnI and BgIII sites. Both DNA fragments were digested with BgIII and KpnI restriction enzymes and ligated with the cleaved BglII site of the E. coli/Corynebacterium shuttle vector pCGL0243 (1) to generate pAG153. This recombinant plasmid was transformed into a *C. diphtheriae* strain that lacks *spaA* and *srtA*.

**Isolation of SpaA Pili.** Overnight cultures of *C. diphtheriae* harboring pAG153 were diluted 1:10 in a defined medium supplemented with 50  $\mu$ g/mL kanamycin. The supernatant of the overnight culture was collected by centrifugation and saturated with (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> to 30% and then 75% for 1 h at 4 °C. Precipitated proteins were solubilized in 15 mL of EQ buffer (150 mM NaCl and 50 mM Tris-HCl, pH 7.5) and dialyzed twice in EQ buffer for 24 h at 4 °C. Dialyzed material was centrifuged to

1. Ton-That H, Schneewind O (2003) Assembly of pili on the surface of *Corynebacterium* diphtheriae. Mol Microbiol 50:1429–1438.

remove insoluble proteins, and soluble His<sub>6</sub>-tagged SpaA pilus proteins were purified using nickel-affinity chromatography as previously described (1).

Proteolytic Digestion and Mass Spectral Analyses. Purified SpaA pili and recombinant SpaA protein were digested and analyzed in a similar way to that previously described (2). SDS-PAGE gel bands containing  $\approx 10 \ \mu g$  of SpaA or SpaA pili were diced and washed with 50% acetonitrile and 25 mM NH<sub>4</sub>HCO<sub>3</sub> to remove the gel stain. The gel pieces were incubated with 100% acetonitrile and dried under vacuum before adding proteases dissolved in 25 mM NH<sub>4</sub>HCO<sub>3</sub> and 10% acetonitrile. The gel pieces were incubated with trypsin (Promega) for 4 h followed by AspN endopeptidase (Roche) for 16 h at 37 °C. For SpaA pili, 1.5 μg of trypsin and 2  $\mu$ g of AspN were used. For recombinant SpaA, 500 ng of trypsin and 200 ng of AspN were used. The digested samples were analyzed using a Q-STAR XL Hybrid MS/MS system (Applied Biosystems) and on a  $0.3 \times 150$ -mm C<sub>18</sub> column with 0.1% aqueous formic acid (A) and 0.1% formic acid in acetonitrile (B), in which a linear gradient of 10-35% B was applied to separate the peptides. Data within the m/z range 300-1,600 were acquired, and 3 product ion scans were performed on the 3 most abundant multiply charged precursors. Initial searches against SpaA sequence using Mascot search engine version 2.0.05 (Matrix Science) identified linear peptides only. Unmatched peptides were manually searched to identify peptides crosslinked by the isopeptide bonds, either intramolecular or intermolecular.

 Kang HJ, Coulibaly F, Clow F, Proft T, Baker EN (2007) Stabilizing isopeptide bonds revealed in Gram-positive bacterial pilus structure. Science 318:1625–1628.

| SpaA           | 51 IPERTSIAVHALMGLP  | T GOPANGTKLDSI                                       | - GLPKVDGMSFTLYRVN                                 | 95  |
|----------------|--|--|--|-----|
| SpaD           | 48 F E R K G S L T L H K K K G A E                               | SEKRATGKEMDDV  | - AGEP LNGVT FKITKLN                               | 92  |
| SpaH           | 52 FTKOGRIJINKHLHOE  | DGSSAVGDPATGD  | YKGTPTEGLNGVTFKIYKIK                               | 100 |
| FimA           | 43 T D A T G S L A L HK H L N                                    | GGGKDIGNPTGTP-ONAD                                   | SKGTP VOGVV FTAYP IT                               | 90  |
| FimP           | 42 P D A A T T L T V HK C FO T D T N                             | GVKEGTGNEDPOA  | - ECKP VSDVEETITKIN                                | 88  |
| CPE156         | 37 AP LKCS LT LSKKCATETAYK                                       |  |  | 90  |
| EE1002         |  |  |  | 92  |
| 546645         |  |  |  | 102 |
| Brap           |  |  |  | 07  |
| RIGB           |  |  |  | 70  |
| BCDA           |  |  |  | 70  |
| BCer98_3114    | 38 - KTKGSLTTHKTEQEKDKT  | AGAEGDGSARQDVP                                       |  | 80  |
| BCE_5331       | 28 - PTKGSLTTHKYAHEKDGQ  | QGEAGIGEPGQKVP                                       | EGATP TKGVKYEVVKVA                                 | 76  |
| BCE_G9241_1728 | 28 - K STGTLTTHKFEQSK DFK  | I I I EGNGKEGQI VP                                   | EGAKPLEGVIFEIKRIE                                  | 76  |
| CE2457         | 58 VP ETANVVITKLEQPTV  | AGSVASGQSQTLP  | EGSKGIEDVEFTAYRVP                                  | 104 |
| SapB           | 47 NDQQ L S L T L HK R I GA                                      | EGEQGDGTEITNP  | - SGEP AKGVKYKIELVQ                                | 90  |
| Spy0109        | 47 Q P D K T T V N I Y K L Q G A D F S K Q P E                   | GIKNENGEPIDIT - KLKD                                 | T F G T A V T Y L P G V K F K Y Y K V K            | 105 |
| Spy0116        | 31 S S H K T D V V I H K I K M T S L K G W P K E K N P D G T Y T | GLGDKNYNGEKIDTITSYFG                                 | EGAEELDGVSFTYWSVD                                  | 98  |
|                |  | _  |  |     |
| SpaA           | 96 E - I D L T T Q A G W D A A S K I K L E E L                   | YTNGHPTDKVTKV  | 'ATKKTEG <mark>G</mark> VAKFDNL                    | 144 |
| SpaD           | 93 FDL Q NGD - WA K F P K T                                      | A A D A K G H E T S T T K                            | EVETSGN <mark>G</mark> TAVFDNL                     | 134 |
| SpaH           | 101 N - I D L S K Q E G W E A V N K Y N I P A D                  | P T <mark>K</mark> D D N – – – – – – – – – L K E V D | )TVTTANG <mark>G</mark> KATTKDL                    | 147 |
| FimA           | 91 D - INL K DPAGWDT I SDL SKA GVP DS                            | A C T N P A A P T L G A H K F G T P K V              | SPATNDE <mark>G</mark> LATITEM                     | 149 |
| FimP           | 89 VDL TTY DGWKTLA DLK GD  | V V K A G A L K S T T V Q                            | KITTGANGLASFTDAQT                                  | 136 |
| CPE156         | 91 S - IQK LNGE QVKEF  | AINLHKYILENKK  | SGQELKDGQKNTV                                      | 129 |
| EF1093         | 84 N E F Y E Q R A A G A S V D A A K Q                           | A – VQ S L T P – – – – – – – G K P V A               | QGT TDANGNVTVQLPKK                                 | 130 |
| SAG645         | 103 TDI SVDELKKLTTVEAADA   | KVGTILEEGVSLP  | OKT NA - OGLVVDALDSK                               | 150 |
| RraB           | 88 NEIIDENGOTLGVNIDPO  | T F K L S G AM P A T AM                              | IKKLTEAEG-AKENTANL                                 | 134 |
| BcnA           | 77 S - E E K I S N D G K I V K E                                 | DVKPVMGATPN  | NOVVTOONGOAVIKDI                                   | 116 |
| Bcer98 3114    | 87 S - F F K I S S D G K I V K F                                 | A V K P V F T A T P V                                | KVTTNEKGEAVIKEI                                    | 126 |
| BCE 5331       |  |  |  | 116 |
| BCE C9241 1728 |  |  |  | 115 |
| CE2457         |  |  |  | 163 |
| CL2437         |  |  |  | 126 |
| 5apb           |  | K TOLLOV   |  | 150 |
| Spy0109        |  |  |  | 154 |
| SpyUTI6        | 99 KEKTKKLIK <mark>N</mark> PQNTDIVPKMKAFLQGIEKNK                | ALENSSEI   | GHL  | 157 |
| <i>~</i> .     |  |  |  |     |
| SpaA           | 145 TPALYLVVQELNGAEAVVRSQP                                       | FLVAAPQTNPTG   |  | 193 |
| SpaD           |  | FIVSIPMVNEAS   |  | 182 |
| SpaH           | 148 SVAAYVVYET DAPADVIDRASP                                      | FVVTLPYPDTTANDDRLGER                                 | . TNTWLYDVNVFPKNGK                                 | 205 |
| FimA           | 150 P VQAY L V CET T T P G N I VQ K A K P                        | FVVTVPHPNTAAGAD                                      | - GQWIYDVHVYPKNEA                                  | 201 |
| FimP           | 137 EVGAYLVSETRTP DKVI PAED                                      | FVVTLPMTNPQDT  | - AKWNYNVHVYPKNT L                                 | 185 |
| CPE156         | 130 DLGYYLVTETSSDSEGAAVASTP                                      | 1 I V S V P Q V S G                                  | - DSWNYDVT I NPKDNT                                | 177 |
| EF1093         | 131 QNGK DAVYT I K E E P K E G V V A A T N                       | MVVAFPVY EMIKQTDGSYKY                                | GTEELAVVHIYPKNVVAND                                | 192 |
| SAG645         | 151 SNVRYLYVEDLKNSPSNITKAYAVP                                    | FVLELPVANSTG   | - T G F L S E I N I Y P K N V V                    | 202 |
| RrgB           | 135 PAAKYKIYEIHSLSTYVGEDGATLTGSKAVP                              | IEIELPLND  | VVDAHVYPKNTE                                       | 186 |
| BcpA           | 117 P L G R Y E V K E V A G P P H V N L N P N T                  | YTVDIPLTNKEG   | – K V L N Y D V HM Y P K N E I K R G               | 168 |
| Bcer98_3114    | 127 P L G R Y E V K E V A G P P H V N L N P N T                  | YT <mark>VDIP</mark> LTNKDG                          | – NV LNY DV HMY P K NE I K R G                     | 178 |
| BCE_5331       | 117 PIGRYLVKEVSGPDNVNINTEP                                       | YTIDIPMTDKEG   | – K K L N Y D V H I Y P <mark>K</mark> N E I K R G | 168 |
| BCE_G9241_1728 | 116 PLGRYEVKAIKGPEHIELDMTT                                       | FIVDIPMTSTNG   | – K E L N Y D V H I Y P K N E T K R G              | 167 |
| CE2457         | 164 A A G L Y L I R E T S T P A G V I P A Q D                    | FLLAVPMTNPEG   | – T G W L T D I Y V Y P K G S T – – –              | 211 |
| SapB           | 137 P K G L Y K V T E T E V P A N S G L V P G A P                | FLVYVPMTNSEG   | – K GW I Y DV HAY P K N S K – – –                  | 186 |
| Spy0109        | 155 DKVKYLFVESSNQDTVNKVVGYTAVP                                   | FILHLPVSNSNG   | – K G Y Y D E <mark>V N V Y P K</mark> N T T – – – | 207 |
| Spy0116        | 158 ADGYYWFVENSGSNI ANGETLSSSAAVP                                | FGLELPVYKADG   | – – STITELHVYPKNTT – – –                           | 211 |
|                |  |  |  |     |

**Fig. S1.** Sequence alignment of N-terminal regions of Gram-positive bacterial major pilins. Residues that are aligned with the isopeptide bond forming residues of BcpA (K37 and Asn-163) [Budzik JM et al. (2008) Amide bonds assemble pili on the surface of bacilli. *Proc Natl Acad Sci USA* 105:10215–20] and predicted catalytic residue (E125 of BcpA) are outlined with red boxes. The pilin motif lysines are highlighted in orange. SpaA, SpaD, and SpaH from *C. diphtheriae*, FimA and FimP from *Actinomyces naeslundii*, CPE156 from *Clostridium perfringens*, EF1093 from *Enterococcus faecalis*, SAG645 from GBS, RrgB from *S. pneumoniae*, BcpA, Bcer98\_3114, BCE\_5331, and BCE\_G9241\_1728 from *B. cereus*, CE2457 from *Corynebacterium efficiens*, SapB from *Corynebacterium jeikeium*, Spy0109 from *S. pyogenes* strain M2, and Spy0116 from *S. pyogenes* strain M4.

PNAS PNAS



Fig. S2. Ribbon representation of SpaA N-domain. Residues of Ala-61, Gln-153, His-191, and Lys-190 of SpaA are shown in stick mode.

# Table S1. Data collection, phasing, and refinement statistics

| Parameters   | Native SpaA                       | SeMet-SpaA                                    |  |  |
|--|-----------------------------------|---|--|--|
| Space group  | P212121                           | P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> |  |  |
| Unit cell dimensions (Å, deg)                          | a = 34.76, b = 63.68, c = 198.11, | a = 34.78, b = 64.04, c = 199.19,             |  |  |
|  | $lpha=eta=\gamma=$ 90.0           | $lpha=eta=\gamma=$ 90.0                       |  |  |
| Data collection*                                       |                                   |   |  |  |
| Resolution range (Å)                                   | 40.00-1.60 (1.69-1.60)            | 40.00-1.80 (1.90-1.80)                        |  |  |
| Wavelength (Å)   | 0.773                             | 0.980   |  |  |
| Total reflections                                      | 745,330 (54,546)                  | 545,315 (43,503)                              |  |  |
| No. of unique reflections                              | 58,024 (7,383)                    | 41,950 (5,546)                                |  |  |
| Redundancy   | 12.8 (7.4)                        | 6.9 (4.1)                                     |  |  |
| Completeness (%)                                       | 97.7 (87.1)                       | 98.2 (89.7)                                   |  |  |
| Mean $I/\sigma(I)$                                     | 13.9 (2.6)                        | 23.8 (4.2)                                    |  |  |
| R <sub>svm</sub> (%)                                   | 11.4 (50.3)                       | 6.1 (35.1)                                    |  |  |
| Phasing statistics for SeMet-SpaA                      |                                   |   |  |  |
| Resolution range (Å)                                   | 40.00-1.80                        |   |  |  |
| Figure of Merit (acentric/centric)                     | 0.45/0.24                         |   |  |  |
| Phasing power (anomalous)                              | 1.47                              |   |  |  |
| Refinement statistics for Native SpaA                  |                                   |   |  |  |
| Resolution range (Å)                                   | 40.00-1.60                        | (1.64–1.60)                                   |  |  |
| No. of reflections                                     | 54,856                            | (3,418)                                       |  |  |
| No. of reflections for R <sub>free</sub> test set      | 2,898 (157)                       |   |  |  |
| R <sub>work</sub> /R <sub>free</sub> (%)               | 19.3/22.0 (27.1/29.0)             |   |  |  |
| RMS deviations from standard bond lengths/angles (Å/°) | 0.014                             | /1.53   |  |  |
| Average B-factor (Ų)                                   | 12.                               | 57  |  |  |
| Protein atoms  | 3,2                               | 46  |  |  |
| lons   | 1 Ca <sup>2+</sup>                | -, 1 I <sup></sup>                            |  |  |
| Water molecules  | 57                                | '1  |  |  |
| Ramachandran plot by MOLPROBITY (1)                    |                                   |   |  |  |
| Most favored/allowed (%)                               | 98.6/                             | 100.0   |  |  |
| Generously allowed or outliers (%)                     | 0.                                | 0   |  |  |

\*Items in parentheses are for the outermost shell of data.

PNAS PNAS

1. Richardson JS, Arendall WB, Richardson DC (2003) New tools and data for improving structures, using all-atom contacts. Methods Enzymol 374:385–412.

## Table S2. MS/MS of a peptide at m/z 996.8<sup>3+</sup> containing Lys199-Asn321 isopeptide bond of SpaA

| Observed m/z* | Charge | Calculated <i>m/z</i> <sup>+</sup> | $\Delta_{\mathrm{obs-calc}}^{\ddagger}$ | Proposed structure  | lon type                |
|---------------|--------|------------------------------------|---|---|-------------------------|
| 189.08        | +1     | 189.12                             | -0.04                                   | AV  | y2                      |
| 205.12        | +1     | 205.10                             | 0.02                                    | W   | У <sub>1</sub> ,        |
| 266.09        | +1     | 266.12                             | -0.03                                   | HQ  | b <sub>2</sub>          |
| 290.17        | +1     | 290.17                             | 0                                       | TAV   | Уз                      |
| 299.11        | +1     | 299.13                             | -0.02                                   | PSN§  | Internal                |
| 333.14        | +1     | 333.16                             | -0.02                                   | QW  | У <sub>2'</sub>         |
| 337.16        | +1     | 337.16                             | 0                                       | HQA   | b₃                      |
| 398.21        | +1     | 398.20                             | 0.01                                    | PTAQ <sup>§</sup>   | Internal                |
| 404.19        | +1     | 404.19                             | 0                                       | AQW   | У <sub>З'</sub>         |
| 450.28        | +1     | 450.25                             | 0.03                                    | HQAL  | b4                      |
| 537.31        | +1     | 537.28                             | 0.03                                    | HQALS   | b₅                      |
| 568.28        | +1     | 568.27                             | 0.01                                    | PSNPTA§   | Internal                |
| 602.33        | +1     | 602.30                             | 0.03                                    | PTAQW   | У5 <sup>,</sup>         |
| 666.28        | +1     | 666.32                             | -0.04                                   | HQALSE  | b <sub>6</sub>          |
| 696.34        | +1     | 696.33                             | 0.01                                    | <i>PSNPTAQ</i> <sup>§</sup>                               | Internal                |
| 716.31        | +1     | 716.34                             | -0.03                                   | NPTAQW  | У6'                     |
| 810.41        | +2     | 810.40                             | 0.01                                    | HQALSEPVKTAV and DNQ (-NH₃) <sup>¶</sup>                  | Parent-y <sub>12</sub>  |
| 846.46        | +2     | 846.43                             | 0.03                                    | HQALSEPVKTAV and DNQA (-NH <sub>3</sub> ) <sup>¶</sup>    | Parent-y <sub>11'</sub> |
| 900.51        | +1     | 900.42                             | 0.09                                    | PSNPTAQW  | ¥8'                     |
| 939.48        | +2     | 939.47                             | 0.01                                    | HQALSEPVKTAV and DNQAW (-NH <sub>3</sub> ) <sup>¶</sup>   | Parent-y <sub>10'</sub> |
| 989.09        | +2     | 989.01                             | 0.08                                    | HQALSEPVKTAV and DNQAWV (-NH <sub>3</sub> ) <sup>¶</sup>  | Parent-y <sub>9'</sub>  |
| 1,045.66      | +2     | 1045.55                            | 0.11                                    | HQALSEPVKTAV and DNQAWVL (-NH <sub>3</sub> ) <sup>¶</sup> | Parent-y <sub>8'</sub>  |

\*Monoisoptic masses of observed ions.

<sup>†</sup>Calculated ions. Monoisotopic masses were calculated using the Fragment Ion Calculator (http://db.systemsbiology.net:8080/proteomicsToolkit/ FragIonServlet.html).

<sup>‡</sup>Difference between observed ion mass and calculated ion mass.

§Internal ions are shown in italics.

PNAS PNAS

 $^{\P}\text{Loss}$  of 17 Da from losing  $NH_3$  is shown in parentheses.

# Table S3. MS/MS of a peptide at m/z 751.7<sup>3+</sup> containing Lys363-Asn482 isopeptide bond of SpaA.

| Observed <i>m/z</i> * | Charge | Calculated m/z <sup>+</sup> | $\Delta_{obs-calc}^{\ddagger}$ | Proposed structure                    | lon type                            |
|-----------------------|--------|-----------------------------|--------------------------------|---------------------------------------|-------------------------------------|
| 147.12                | +1     | 147.08                      | 0.04                           | GA                                    | У2                                  |
| 147.12                | +1     | 147.11                      | 0.01                           | К                                     | У <sub>1</sub> ,                    |
| 155.07                | +1     | 155.08                      | -0.01                          | PG§                                   | Internal                            |
| 177.11                | +1     | 177.10                      | 0.01                           | FG                                    | b <sub>2</sub>                      |
| 205.12                | +1     | 205.10                      | 0.02                           | FG                                    | a <sub>2</sub>                      |
| 244.12                | +1     | 244.13                      | -0.01                          | PGA                                   | Уз                                  |
| 244.12                | +1     | 244.15                      | -0.03                          | NK (-NH₃) <sup>¶</sup>                | y <sub>2'</sub> (-NH <sub>3</sub> ) |
| 333.10                | +1     | 333.15                      | -0.05                          | FGQ                                   | b <sub>3</sub>                      |
| 345.15                | +1     | 345.18                      | -0.03                          | TPGA                                  | У4                                  |
| 418.25                | +1     | 418.25                      | 0                              | FGQI                                  | a4                                  |
| 446.26                | +1     | 446.24                      | 0.02                           | FGQI                                  | b <sub>4</sub>                      |
| 472.27                | +1     | 472.27                      | 0                              | IDNK (-NH <sub>3</sub> ) <sup>¶</sup> | y <sub>4′</sub> (-NH <sub>3</sub> ) |
| 516.26                | +1     | 516.24                      | 0.02                           | GNTPGA                                | У6                                  |
| 547.33                | +1     | 547.29                      | 0.04                           | FGQIT                                 | b <sub>5</sub>                      |
| 631.26                | +1     | 631.27                      | -0.01                          | DGNTPGA                               | У7                                  |

\*Monoisoptic masses of observed ions.

<sup>†</sup>Calculated ions. Monoisotopic masses were calculated using the Fragment Ion Calculator (http://db.systemsbiology.net:8080/proteomicsToolkit/ FragIonServlet.html).

<sup>‡</sup>Difference between observed ion mass and calculated ion mass.

 $\ensuremath{{}^{\$}}\xspace$  Internal ions are shown in italics.

PNAS PNAS

 $\ensuremath{^{\mbox{l}}\xspace{-1.5}}\xspace{-1.5}$  Loss of 17 Da from losing  $NH_3$  is shown in parentheses.

| Table S4. Daughter ions produced during MS/MS of a peptide (m/z 585.7 <sup>3+</sup> ) that was generated from trypsin and AspN digestion of | эf |
|---|----|
| SpaA pili (this peptide contains the intersubunit isopeptide bond between SpaA Lys190 and Thr494)   |    |

| Observed m/z* | Charge | Calculated m/z <sup>+</sup> | $\Delta_{\text{obs-calc}}^{\ddagger}$ | Proposed structure                                    | lon type                              |
|---------------|--------|-----------------------------|---------------------------------------|---|---------------------------------------|
| 106.05        | +1     | 106.05                      | 0                                     | S   | У1                                    |
| 187.15        | +1     | 187.11                      | 0.04                                  | DV  | a <sub>2</sub>                        |
| 211.14        | +1     | 211.16                      | -0.02                                 | PL <sup>§</sup>                                       | Internal                              |
| 215.11        | +1     | 215.10                      | 0.01                                  | DV  | b <sub>2</sub>                        |
| 215.11        | +1     | 215.14                      | -0.03                                 | EL  | a <sub>2'</sub>                       |
| 219.13        | +1     | 219.13                      | 0                                     | LS  | У2                                    |
| 243.13        | +1     | 243.13                      | 0                                     | EL  | b <sub>2'</sub>                       |
| 312.22        | +1     | 312.20                      | 0.02                                  | PLT §-H <sub>2</sub> O                                | Internal                              |
| 340.22        | +1     | 340.19                      | 0.03                                  | ELP   | b <sub>3'</sub>                       |
| 352.15        | +1     | 352.16                      | -0.01                                 | DVH   | b <sub>3</sub>                        |
| 418.22        | +1     | 418.23                      | -0.01                                 | QALS  | У4                                    |
| 451.21        | +1     | 451.23                      | -0.02                                 | DVHV  | b <sub>4</sub>                        |
| 472.57        | +3     | 472.58                      | -0.01                                 | DVHVYPKHQ and PLT (- H <sub>2</sub> O) <sup>¶</sup>   | Parent-y <sub>3</sub> -b <sub>2</sub> |
| 496.26        | +3     | 496.27                      | -0.01                                 | DVHVYPKHQA and PLT (- H <sub>2</sub> O) <sup>¶</sup>  | Parent-y <sub>2</sub> -b <sub>2</sub> |
| 533.97        | +3     | 534.96                      | 0.01                                  | DVHVYPKHQAL and PLT (- H <sub>2</sub> O) <sup>¶</sup> | Parent-y <sub>1</sub> -b <sub>2</sub> |
| 555.32        | +1     | 555.29                      | 0.03                                  | HQALS   | y <sub>5</sub>                        |
| 568.99        | +3     | 568.97                      | 0.02                                  | DVHVYPKHQALS and PLT (- $H_2O)^{1}$                   | Parent-b <sub>2'</sub>                |
| 614.33        | +1     | 614.29                      | 0.04                                  | DVHVY   | b <sub>5</sub>                        |

\*Monoisoptic masses of observed ions.

PNAS PNAS

<sup>†</sup>Calculated ions. Monoisotopic masses were calculated using the Fragment Ion Calculator (http://db.systemsbiology.net:8080/proteomicsToolkit/ FragIonServlet.html).

<sup>‡</sup>Difference between observed ion mass and calculated ion mass.

§Internal ions are shown in italics.

 $^{\P}\text{Loss}$  of 18 Da from losing  $H_2O$  is shown in parentheses