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 μ g/mL kanamycin. The supernatant of the overnight culture was collected by centrifugation and saturated with (NH₄)₂SO₄ 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 His₆-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 NH₄HCO₃ to remove the gel stain. The gel pieces were incubated with 100% acetonitrile and dried under vacuum before adding proteases dissolved in 25 mM NH₄HCO₃ 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 μ 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 0.3×150 -mm C₁₈ 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 I P E R T S I A V HALMG L P	T G Q P A N G T K L D S I	- G L P K V D G M S F T L Y R V N	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 FTKQGRIIINKHLHQE	DGSSAVGDPATGD	/ K G T P T E G L N G V T F K I Y K I K	100
FimA	43 T D A T G S L A I H K H L N	GGGKDI GNPTGTP-QNAD	SKGTP VQGVVFTAYPIT	90
FimP	42 P DAATTLTVHK CEOTDTN	GVKEGTGNEDPQA	- ECKP VSDVEFTITKLN	88
CPE156	37 APIKGSITISKKGATFTAYKLLD	AIKSGDAYEYSV	- N S D L K D F F N N S N Y G S Y S O E	90
EF1093	35 NG E SAO L V L HK K KMT D L P D	PLIONS GKEMSEE	DKYOGLADVTESLYNVT	83
SAC645	44 RPAKTTVNI YKLOADSYKSELTSNG	CLENKDGEVISNY-AKIG		102
Braß		ETCNYACNKYCVIPA		87
RenA				76
Bcer98 3114		ACAECDC SAKODV PK		86
DCE 5221				76
PCE_0341_1728				76
GE2457				104
CE2457			GSKGTEDVEFTATRVP	104
Sapв			SGEPAKGVKFKTELVQ	90
Spy0109	47 QP DK TTVNTYKLQGADF SK QP E	GIKNENGEPIDII-KLKDI	FUTAVIYLPGVKFKYYKVK	105
Spy0116	31 S SHK I DVV THKIKMI S LKGWPKEKNPDG I Y I G L	GDKNYNGERIDIIISYFGI	GALELDGVSFIYWSVD	98
Spal				144
Span				124
Spall				147
Span Firm A				147
FIMA Fim D	91 D-TNLKDPAGWDTTSDLSKAGVPDSAC	KACALK STTVO		149
CDELEC				120
CPEIDO	91 S-TQKLNGEQVKEFAT	NERKTIEENKK		129
EFT093	84 NEFYEQRAAGASVDAARQA-	VQSLIPGRPVAC	2GTTDANGNVTVQLPKK	130
SAG645		GITLEEGVSLPC	2KINA-QGLVVDALDSK	150
RrgB	88 NETIDENGQILGVNIDPQIF	KLSGAMPATAMK	KLTEAEG-AKFNTANL	134
ВсрА	// S-FEKISNDGKIVKED-	- VKPVMGATPNC	2VVIDDNGQAVLKDL	116
Bcer98_3114	87 S-FEKISSDGKIVKEA-	- VKPVEIAIPVF	(VIINEKGEAVLKEL	126
BCE_5331	77 T - FKV I ERNG EIVKEA -	VEKVDNATPISL	JVTGE-DGTVTFKDL	116
BCE_G9241_1728	77 S-FEKISNDGTVIKEE-		TKVTNAKGEAVFTNL	115
CE2457	105 VYKAGENGDEHPARTSAWQQATADIDLDMA	K S R V A E L A T P T E	EVALNPTNENGVTNWSA A	163
SapB	91 P L NN S G DWAKARE I T NP S EA	KAWTGVGAYSK	² GETGED <mark>G</mark> TLKFEEL	136
Spy0109	106 N Y S T S DD V L K S I K T V EQ A D S K -	T D L L DV A G A K E 1	FEA T DQ S G K V S I D L P S N	154
Spy0116	99 K E K Y K K L T K <mark>N</mark> P Q N Y D T V P K M K A F L Q G T E <mark>K N K A</mark> L	ENSSETIDGKT	GH TADK GGVKVKD L	157
<i>c</i>				
SpaA	145 TPALYLVVQELNGAEAVVKSQPFL	VAAPQINPIG	· DGWLQDVHVYPKHQA	193
SpaD		V S T P M V N E A S	· DAWNYNVVAYPKNTE	182
Spaн	148 SVAAYVVYETDAPADVTDKASPFV	VILPYPDIIANDDRLGERI	INTWLYDVNVFPKNGK	205
FIMA	ISO P VQAY L VCETTTP GNTVQKAKP F V	VIVPHPNIAAGAD	GQWTYDVHVYPKNEA	201
FimP	137 EVGAY LVSETRTP DKVI PAEDFV	VTLPMTNPQDT	· AKWNYNVHVYPKNT L	185
CPE156	130 D L G Y Y L V T E T S S D S E G A A V A S T P I I	V S V P Q V S G	· D SWNY DV T I NP KDNT	177
EF1093	131 QNGK DAVYT I K E E P K E G V V A A T NMV	VAFPVYEMIKQTDGSYKYO	JT E E LAVVH I Y PKNVVAND	192
SAG645	151 SNVRYLYVEDLKNSPSNITKAYAVPFV	LELPVANSTG	- T G F L S E I N I Y P K N V V	202
RrgB	135 PAAKYKIYEIHSLSTYVGEDGATLTGSKAVPIE	IELPLND	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 Y T	VDIPLTNKEG	- K V L N Y D V HM Y P <mark>K</mark> 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 Y T	VDIPLTNKDG	- NV L NY DV HMY P K NE I K R G	178
BCE_5331	117 P I G R Y L V K E V S G P D N V N I N T E P Y T	I D I P MT DK EG	- 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 P L G R Y E V K A I K G P E H I E L DM T T F I	VDIPMTSTNG	- K E L N Y D V H I Y P <mark>K</mark> N E T K R G	167
CE2457	164 AAGLYLIRETSTP AGVI PAQDFL	LAVPMTNPEG	- T G W L T D I Y V Y P <mark>K</mark> 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 F L	V Y V P M T N S E G - - - - - - - - - -	- K GW I Y D V H A Y P K N S K	186
Spy0109	155 DKVKYLFVESSNQDTVNKVVGYTAVPFI	L H L P V S N S N G	- K G Y Y D E V N V Y P <mark>K</mark> N T T – – –	207
Spy0116	158 A D G Y Y W F V E N S G S N I – – – A N G E T L S S S A A V P F G	LELPVYKADG	-	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 ₁ 2 ₁ 2 ₁		
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 _{svm} (%)	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 _{free} test set	2,898 (157)			
R _{work} /R _{free} (%)	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 ²⁺	-, 1 I		
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 996.8³⁺ containing Lys199-Asn321 isopeptide bond of SpaA

Observed m/z*	Charge	Calculated <i>m/z</i> ⁺	$\Delta_{\mathrm{obs-calc}}^{\ddagger}$	Proposed structure	lon type
189.08	+1	189.12	-0.04	AV	y2
205.12	+1	205.10	0.02	W	У ₁ ,
266.09	+1	266.12	-0.03	HQ	b ₂
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 _{2'}
337.16	+1	337.16	0	HQA	b₃
398.21	+1	398.20	0.01	PTAQ [§]	Internal
404.19	+1	404.19	0	AQW	У _{З'}
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 [,]
666.28	+1	666.32	-0.04	HQALSE	b ₆
696.34	+1	696.33	0.01	<i>PSNPTAQ</i> [§]	Internal
716.31	+1	716.34	-0.03	NPTAQW	У6'
810.41	+2	810.40	0.01	HQALSEPVKTAV and DNQ (-NH₃) [¶]	Parent-y ₁₂
846.46	+2	846.43	0.03	HQALSEPVKTAV and DNQA (-NH ₃) [¶]	Parent-y _{11'}
900.51	+1	900.42	0.09	PSNPTAQW	¥8'
939.48	+2	939.47	0.01	HQALSEPVKTAV and DNQAW (-NH ₃) [¶]	Parent-y _{10'}
989.09	+2	989.01	0.08	HQALSEPVKTAV and DNQAWV (-NH ₃) [¶]	Parent-y _{9'}
1,045.66	+2	1045.55	0.11	HQALSEPVKTAV and DNQAWVL (-NH ₃) [¶]	Parent-y _{8'}

*Monoisoptic masses of observed ions.

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

[‡]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 751.7³⁺ containing Lys363-Asn482 isopeptide bond of SpaA.

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

*Monoisoptic masses of observed ions.

PNAS PNAS

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

[‡]Difference between observed ion mass and calculated ion mass.

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

 $\ensuremath{^{\mbox{l}}\xspace{-1.5}}\xspace{-1.5}$ 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 ³⁺) 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 ⁺	$\Delta_{\text{obs-calc}}^{\ddagger}$	Proposed structure	lon type
106.05	+1	106.05	0	S	У1
187.15	+1	187.11	0.04	DV	a ₂
211.14	+1	211.16	-0.02	PL [§]	Internal
215.11	+1	215.10	0.01	DV	b ₂
215.11	+1	215.14	-0.03	EL	a _{2'}
219.13	+1	219.13	0	LS	У2
243.13	+1	243.13	0	EL	b _{2'}
312.22	+1	312.20	0.02	PLT §-H ₂ O	Internal
340.22	+1	340.19	0.03	ELP	b _{3'}
352.15	+1	352.16	-0.01	DVH	b ₃
418.22	+1	418.23	-0.01	QALS	У4
451.21	+1	451.23	-0.02	DVHV	b ₄
472.57	+3	472.58	-0.01	DVHVYPKHQ and PLT (- H ₂ O) [¶]	Parent-y ₃ -b ₂
496.26	+3	496.27	-0.01	DVHVYPKHQA and PLT (- H ₂ O) [¶]	Parent-y ₂ -b ₂
533.97	+3	534.96	0.01	DVHVYPKHQAL and PLT (- H ₂ O) [¶]	Parent-y ₁ -b ₂
555.32	+1	555.29	0.03	HQALS	y ₅
568.99	+3	568.97	0.02	DVHVYPKHQALS and PLT (- $H_2O)^{1}$	Parent-b _{2'}
614.33	+1	614.29	0.04	DVHVY	b ₅

*Monoisoptic masses of observed ions.

PNAS PNAS

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

[‡]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