Supporting Information

Budzik *et al.* 10.1073/pnas.0803565105

SI Text

Plasmids and Strains. PCR with primer pairs P2 (encompassing an XbaI site) and P114 (NheI) amplified *bcpA* between residues 26 and 149. Primers P113 (NheI) and P103 (KpnI) amplified a fragment of *bcpA* beginning at residue 150 with the MH_6 tag and sortase D. Primer pairs P2/P114 and P113/103 generated PCR products that were digested with XbaI/NheI and NheI/KpnI and ligated into pLM5 digested with XbaI/KpnI to create pJB44. The following primers were used to create single amino acid substitutions in BcpA by site-directed mutagenesis as previously reported (1). Primers 73/74 created the K162A substitution in pJB44 thereby generating pJB57. Four substitutions were constructed in pJB44 to generate pJB103: primers 122/123 were used for the M159L substitution, 149/150 for the K151A substitution, 191/192 for the K166M substitution, and 181/182 for the I518M substitution. Primers 201/202 created the N163A substitution in pJB103, resulting in pJB112.

pJB7 encodes for BcpA-His_6 without the amino-terminal signal peptide or the carboxy-terminal sorting signal in the pET24-b expression vector (Novagen) (2). pJB7 was used as template to generate substitution mutations in *bcpA* via Quick-Change mutagenesis. Primers 183/184 introduced the E472A substitution of *bcpA*, thereby generating pJB97. Primers 185/186 introduced the E223A substitution of *bcpA* and generated pJB98. For pJB105, a plasmid harboring E223A and E472A substitutions, Quick-Change was performed with pJB97 template and primers 185/186.

pJB12 (1) encodes *bcpA-srtD-bcpB* under control of the IPTG-inducible P_{spac} promoter. Primers 183/184 introduced the E472A substitution in *bcpA* to generate pJB107. Primers 185/186 introduced the E223A substitution in BcpA, resulting in pJB108. To introduce both the E223A and E472A substitutions in *bcpA*, pJB107 was used as a template for mutagenesis with P185/186, thereby generating pJB123. Primers 197/198 introduced the N163A substitution in *bcpA* to produce pJB109.

pJB69 encodes for the translational fusion *bcpA-gst-srtD* under control of the IPTG-inducible P_{spac} promoter. Templates for amplification of *bcpA-srtD* and *gst* were *B. cereus* chromosomal DNA and pGEX2TK (Amersham Biosciences), respectively. Insertion of P2/P151 (*bcpA*), P152/P153 (*gst*), P154/P103 PCR products into pLM5, using XbaI, NheI, NcoI, and KpnI sites, resulted in pJB69. To generate pJB86, an isogenic plasmid lacking *srtD*, pJB69 was amplified with *Pfu* DNA polymerase and P88/P89. Amplified plasmids were first incubated for 3 h with DpnI to digest methylated parental plasmid DNA and were then digested with KpnI for 2 h. Digested plasmid DNA was ligated and transformed into *E. coli* DH5α.

pJB1 (3) was used for expression of an untagged version of PlyL amidase (4).

Pilus Assembly. *B. anthracis* strain Sterne F32 or its isogenic variant AHG263 (*srtA*::*ermC*) (5) harboring pLM5 or plasmids encoding for *B. cereus* pilus genes were grown overnight at 30°C on LB agar plates containing kanamycin and IPTG. Cells were suspended in 100 mM NaCl, vortexed briefly, and centrifuged for 2 min at 6,000 $\times$ *g*. The supernatant and sediment were separated and analyzed directly by immunoelectron microscopy with α-BcpA antisera and 10 nm gold anti-rabbit IgG conjugate as reported (1). For immunoblotting experiments, supernatant aliquots were precipitated with 7% TCA, incubated on ice for 30 min, and centrifuged at 16,000 $\times$ *g* for 10 min. The sediment was washed with 500 μl of acetone, air-dried, and suspended in 500 μl of 4% SDS, 50 mM Tris-HCl, pH 8.0. For immunoblotting experiments, cells were extracted for 10 min by boiling with 500 μl of 6 M urea, 1% SDS, 50 mM Tris-HCl, pH 9.5. Bacterial samples were washed in water, TCA extracted, and peptidoglycan was digested with mutanolysin (1).

Cyanogen Bromide Cleavage of Pili and Purification of Assembled Pilin Subunits. *B. anthracis* AHG263 (*srtA*::*ermC*) (5) harboring pJB103 or pJB112 was grown for 22 h at 30°C on 80 LB agar plates containing kanamycin and IPTG. Cells were suspended in 16 ml of water, vortexed briefly, and centrifuged for 10 min at 6,000 $\times$ *g*. Supernatant was retrieved and centrifuged again for 10 min at 6,000 $\times$ *g* to sediment remaining cells. Pili in the supernatant were digested with cyanogen bromide in 70% formic acid for 16 h at room temperature in the dark (6). The reaction mixture was dried under vacuum, washed twice with water, and dissolved in 30 ml of buffer A (6 M guanidine hydrochloride, 0.1 M NaH_2PO_4, 0.01 M Tris-HCl, pH 8.0). The sample was loaded on 2 ml of Ni-NTA Sepharose column preequilibrated with buffer A, washed with 15 ml of buffer A, and 15 ml of buffer B (8 M urea, 0.1 M NaH_2PO_4, 0.01 M Tris-HCl, pH 8.0). Cleavage products were eluted with 2 ml of buffer C (same as buffer B but pH 2.0). Proteins were subjected to RP-HPLC over C8 column with a linear gradient from 1% to 90% acetonitrile (CH_3CN) in 0.1% TFA in 90 min. Fractions (1 min) were concentrated ten-fold under vacuum to remove acetonitrile, and aliquots were subjected to SDS/PAGE, mass spectrometry, Edman degradation, and enzymatic cleavage with trypsin.

MALDI-MS. HPLC fractions were mixed in a 1:1 ratio with CH_3CN-water-trifluoroacetic acid (TFA; 60:40:0.1). Samples (0.4 μl) were cospotted with matrix (0.4 μl; α-cyano-4-hydroxycinnamic acid; Sigma) at 10 mg/ml in CH_3CN-water-trifluoroacetic acid (TFA; 30:40:0.1), 4 mM $(NH_4)_2HPO_4$. Matrix-assisted laser desorption ionization-mass spectrometry (MALDI-MS) spectra were obtained in a reflectron time-of-flight (TOF) instrument (ABI Biosystems) in linear and reflectron modes. Spectra were acquired using external calibration with bovine insulin.

ESI-FTMS. For direct infusion, sequential HPLC fractions were individually loaded into 2 μm i.d. externally coated nanospray emitters (New Objective Inc.) and desorbed using a spray voltage of between 1.2 and 1.4 kV (vs. the inlet of the mass spectrometer) (7). These conditions produced a flow rate of 20–50 nl/min. Protein and peptide samples were analyzed using a hybrid linear ion-trap/FTICR mass spectrometer (LTQ FT Ultra, Thermo Electron) at the University of California, Los Angeles (UCLA), as described (7). Fragmentation experiments included collision activated dissociation (CAD), electron capture dissociation (ECD), activated ion electron capture dissociation (aiECD), and infrared multiphoton dissociation (IRMPD). All FTICR fragmentation spectra were analyzed using the Manual Extract software program (Thermo Electron), which generated a list of monoisotopic mass values from the multiply-charged ion datas. Theoretical fragmentation spectra were produced with ProteinProspector v4.27.2 MS-Product web-based program (University of California, San Francisco; http://prospector.ucsf.edu). Observed average molecular weight calculations were performed with the XTractAll for Qual Browser program (Thermo Electron).

Trypsin Digestion and Purification of Branched Tryptic Peptides. Protein concentration in HPLC fractions was determined by bicinchoninic assay (Pierce), and residual TFA was neutralized by concentrated sodium hydroxide and 66 mM ammonium bicarbonate, pH 8.0. Sequencing grade modified trypsin (Promega) was prepared according to the manufacturer's instructions and added at an enzyme:substrate weight ratio of 1:20. The reaction was allowed to proceed for 16 h at 37°C. Tryptic peptides were combined with 12 ml of denaturing buffer A and purified by Ni-NTA affinity chromatography as described above. Purified peptides were subjected to RP-HPLC over a C18 column with a linear gradient from 1% to 90% acetonitrile (CH_3CN) in 0.1% formic acid in 90 min. Samples were concentrated under vacuum, and aliquots subjected to mass spectrometry and Edman degradation. MALDI-MS was performed as previously described except that the samples and matrix were diluted in CH_3CN-water-formic acid.

Edman Degradation. Samples were subjected to SDS/PAGE, transferred to PVDF membranes, stained with amido black, and submitted for Edman sequencing at the University of Illinois at Urbana-Champaign (UIUC) Biotechnology Center Protein Sciences Facility. Peptide RP-HPLC samples were dried under vacuum and submitted for Edman sequencing.

Purification of BcpA-GST Cleavage Products. *B. anthracis* strains Sterne and AHG263 (*srtA*::*ermC*) harboring pJB69 (*bcpA-gst-srtD*) or pJB86 (*bcpA-gst*) were grown in 6 liters of media containing IPTG and kanamycin at 30°C. Sedimented cells were suspended in 50 ml of PlyL buffer (0.5 M sucrose, 1 mM DTT, 1 mM PMSF, 1 mM EDTA, 5 mM sodium phosphate, pH 6.0) and subjected to 10 1-min pulses with a bead beater. The cells were combined with 15 ml of an *E. coli* BL21(DE3)(pJB1) lysate (3). PlyL digestion of the cell wall was allowed to occur for 16 h at 22°C. Protoplasts were sedimented by centrifugation at 9,200 × *g* and resuspended in TSE-triton [20 mM Tris, 200 mM NaCl, 10 mM EDTA, 10 mM β-mercaptoethanol, 1% Triton X-100, pH 7.5, with 1 mg DNase I, 150 μg RNase A, and 1 mM phenylmethanesulphonylfluoride (PMSF)] (8). The protoplasts were disrupted by sonication 10 times with 10-sec pulses. Protoplasts were incubated for 4 hours at 4°C with stirring, and insoluble material was removed by centrifugation. After adjusting the pH of the supernatant to 8.0 with 2 M sodium phosphate, BcpA-GST cleavage products were subjected to 1.5 ml of glutathione-Sepharose column preequilibrated with TSE-triton, washed with TSE-triton, and eluted with 20 mM glutathione (pH 8.0). Aliquots were subjected to SDS/PAGE and stained with Coomassie R-250. Protein species were excised and submitted for trypsin digestion and LC-ESI-MS/MS at the Taplin Mass Spectrometry facility, Harvard Medical School. The purified cleavage products were concentrated 5-fold under vacuum and prepared for Edman degradation.

Trypsin Digestion and Mass Spectrometry of Recombinant BcpA-His_6 Proteins. Recombinant $BcpA_{His\text{-}6}$ (1) and its variants were digested with sequencing grade modified trypsin (Promega) at 37°C with enzyme:substrate ratio of 1:50 for 24 h. Aliquots were subjected to SDS/PAGE and stained with Coomassie R-250.

Recombinant proteins were subjected to RP-HPLC with a PLRP-S polymeric reverse phase column (Polymer Labs; dimensions 150 × 2.1 mm, 300 Å and 5 μm particle size) and a linear gradient from 5% to 60% acetonitrile (CH_3CN) in 0.1% formic acid in 45 min. Mass measurements of RP-HPLC fraction 38 were performed by direct infusion of sample loaded into nanospray emitters in the LTQ FT Ultra mass spectrometer.

1. Budzik JM, Marraffini LA, Schneewind O (2007) Assembly of pili on the surface of *Bacillus cereus* vegetative cells. *Mol Microbiol* 66:495–510.
2. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW (1990) Use of T7 polymerase to direct expression of cloned genes. *Methods Enzymol* 185:60–89.
3. Marraffini LA, Schneewind O (2007) Sortase C-mediated anchoring of BasI to the cell wall envelope of *Bacillus anthracis*. *J Bacteriol* 189:6425–6436.
4. Low LY, Yang C, Perego M, Osterman A, Liddington RC (2005) Structure and lytic activity of a *Bacillus anthracis* prophage endolysin. *J Biol Chem* 280:35433–35439.
5. Gaspar AH, *et al*. (2005) Bacillus anthracis sortase A (SrtA) anchors LPXTG motif-containing surface proteins to the cell wall envelope. *J Bacteriol* 187:4646–4655.
6. Ton-That H, Labischinski H, Berger-Bächi B, Schneewind O (1998) Anchor structure of staphyococcal surface proteins. III. The role of the FemA, FemB, and FemX factors in anchoring surface proteins to the bacterial cell wall. *J Biol Chem* 273:29143–29149.
7. Zabrouskov V, Whitelegge JP (2007) Increased coverage in the transmembrane domain with activated-ion electron capture dissociation for top-down Fourier-transform mass spectrometry of integral membrane proteins. *J Proteome Res* 6:2205–2210.
8. Navarre WW, Schneewind O (1994) Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in gram-positive bacteria. *Mol Microbiol* 14:115–121.

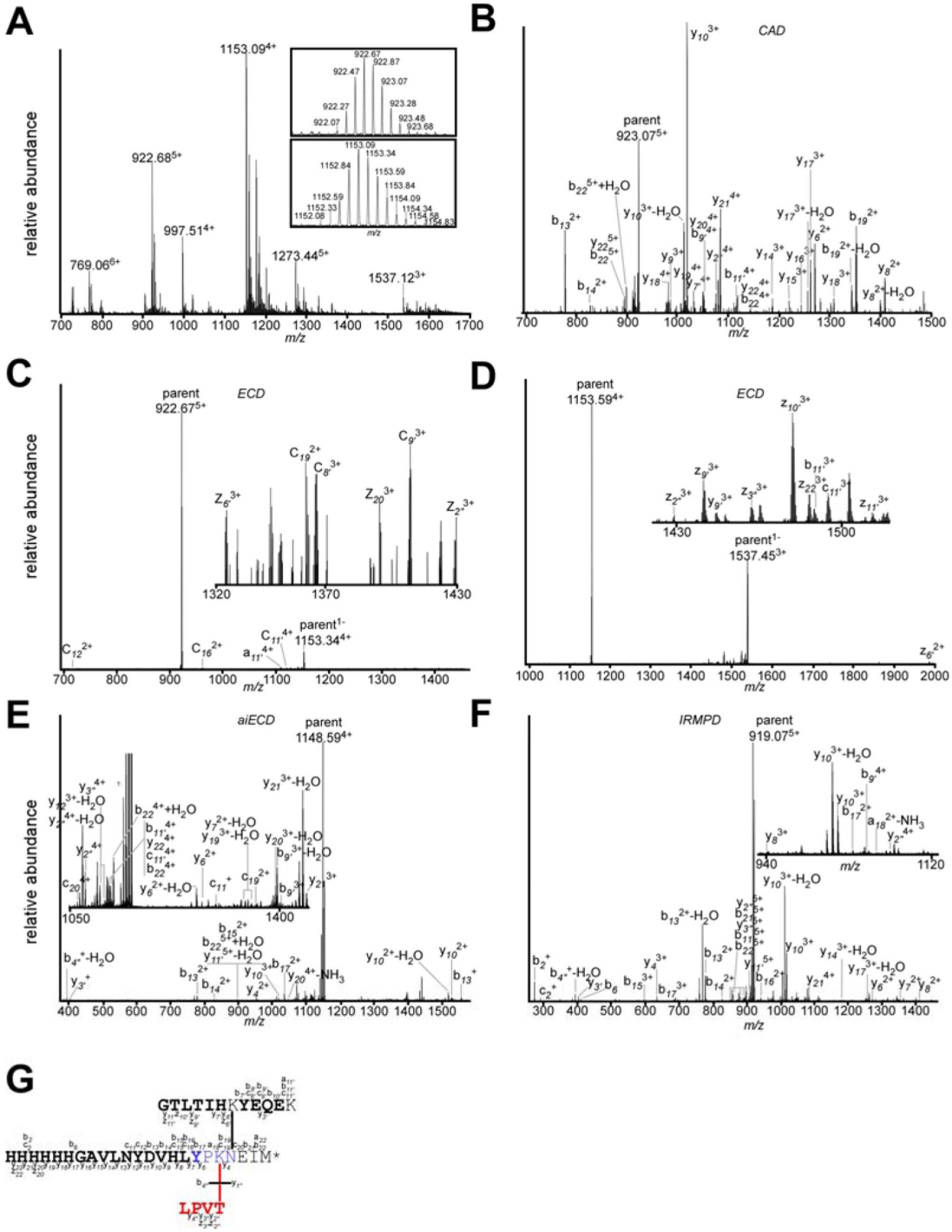

Fig. S1. ESI-FTMS and MS/MS fragmentation of compound A1. (*A*) ESI-FTMS spectrum (*m/z*) of compound A1 (1537.12^{3+}, 1153.09^{4+}, 922.68^{5+}, 769.06^{6+}). The estimated average molecular weight of compound A1 measured by FTMS (4608.34) was calculated from all of the observed charge states with XTractAll for Qual Browser. Insets display ion isolations of 1153.09^{4+} and 922.68^{5+} ions. (*B*) Collision activated dissociation spectrum (*m/z*) of parent ion 922.68^{5+}. (*C*) Electron capture dissociation spectrum (*m/z*) of parent ion 922.68^{5+}. (*D*) Electron capture dissociation spectrum (*m/z*) of parent ion 1153.09^{4+}. (*E*) Activated ion electron capture dissociation spectrum (*m/z*) of parent ion 1153.09^{4+}. (*F*) Infrared multiphoton dissociation spectrum (*m/z*) of parent ion 919.07^{5+}. Activated ion electron capture dissociation spectrum of parent ion 922.68^{5+} and infrared multiphoton dissociation spectrum of parent ion 1153.09^{4+} are not shown. (*G*) Proposed structure of compound A1. The fragment ions obtained by CAD, ECD, aiECD, and IRMPD fragmentation methods are labeled. The bond colored red denotes the intermolecular isopeptide bond catalyzed by sortase. The bond colored black shows the intramolecular isopeptide bond. ′ denotes fragment ions that arose from fragmentation of the GTLTIHKYEQEK branch of the peptide. ′′ denotes fragment ions that arose from fragmention of the LPVT branch of the peptide. The other fragment ions resulted from fragmentation of the HHHHHHGAVLNYDVHLYPKNEIM* branch of the peptide. M* denotes a homoseryl residue. Residues in bold were identified by Edman degradation.

Table S1. Edman degradation of BcpA

Compound A of BcpA pili

Cycle	Amino acid (pM)
1	H (17.05), L (35.04), D (13.47)
2	H (13.79), P (27.85), S (6.37)
3	H (13.87), V (20.65), P (21.58)
4	H (13.22), T (23.07), S (4.30)
5	H (12.64), K (2.04)
6	H (10.4), G (4.39)
7	G (9.37), T (8.17)
8	A (7.01), L (2.59)
9	V (6.98), T (5.32)
10	L (5.91), I (2.95)
11	N (4.40)
12	Y (3.66)
13	D (4.19)
14	V (5.20)
15	H (2.26)
16	L (3.15)
17	Y (2.80)

Compound A1 of BcpA pili

Cycle	Amino acid (pM)
1	H (44.30), L (89.11), G (57.91)
2	H (36.51), P (46.50), T (82.49)
3	H (41.30), V (52.06), L (56.11)
4	H (40.76), T (127.40)
5	H (42.86), I (42.20)
6	H (68.46),
7	G (19.39)
8	A (22.82), Y (27.37)
9	V (21.07), E (18.72)
10	L (18.81), Q (14.94)
11	N (16.45), E (14.80)
12	Y (15.81)
13	D (13.69)
14	V (12.47)
15	H (6.16)
16	L (8.45)
17	Y (7.08)

Compound A* of BcpA pili

Cycle	Amino acid (pM)
1	H (1320.96), L (1721.18)
2	H (1246.33), P (975.19)
3	H (1243.34), V (1047.13)
4	H (1308.06), T (1281.12)
5	H (1268.51)
6	H (1090.57)
7	G (614.07)
8	A (723.29)
9	V (543.92)
10	L (497.28)
11	N (417.33)
12	Y (388.16)
13	D (323.97)
14	V (259.14)
15	H (147.38)
16	L (107.27)
17	Y (66.50)
18	P (23.87)
19	K (8.84)
20	A (8.53)

$BcpA_{His6}$

Cycle	Amino acid (pM)
1	A (28.25)
2	S (14.82)
3	M (11.17)
4	T (28.27)
5	G (21.45)
6	G (25.48)
7	Q (9.44)

BcpA-GST product of SrtD

Cycle	Amino acid (pM)
1	G (2.18)
2	G (2.65)
3	I (2.03)
4	G (2.3)
5	T (1.39)
6	T (1.39)
7	L (1.55)
8	F (1.19)

BcpA-GST product of SrtA

Cycle	Amino acid (pM)
1	G (1.06)
2	G (1.22)
3	I (0.57)
4	G (1.15)
5	T (0.50)
6	T (0.45)
7	L (0.88)
8	F (0.69)

BcpA-GST product of SrtAD

Cycle	Amino acid (pM)
1	G (3.54)
2	G (3.15)
3	I (3.51)
4	G (2.59)
5	T (2.52)
6	T (2.32)
7	L (2.27)
8	F (1.40)

Table S2. ESI-FTMS fragmentation analysis of compound A1

Monoisotopic mass			Daughter ion				
Calculated*	Observed^{+}	$\Delta_{calc-obs}$	proposed structure	Name●	Charge (z)	Method(s)	Parent ion(s)
274.1251	274.1175	0.008	HH	b_2	1	IRMPD	923
291.1516	291.1944	−0.04	HH	c_2	1	IRMPD	923
392.2496	393.1658	−0.92	LPVT	$b_{4''}$ -H_2O	1	IRMPD	923, 1153
	393.1658	−0.92			1	aiECD	1153
403.214	403.2064	0.008	QEK	$y_{3'}$	1	IRMPD	923, 1153
					1	aiECD	1153
822.3607	822.3524	0.008	HHHHHH	b_6	2	IRMPD	923
1293.6413	1293.634	0.008	HHHHHHGAVLN	c_{11}	1	aiECD	1153
1456.7046	1456.696	0.009	HHHHHHGAVLNY	c_{12}	2	ECD	923
1536.6944	1536.689	0.006	HHHHHHGAVLNYD	b_{13} -H_2O	2	IRMPD, aiECD	923
					1,2	IRMPD, aiECD	1153
1554.705	1554.698	0.008	HHHHHHGAVLNYD	b_{13}	1,2	aiECD	1153
					2	CAD, IRMPD, aiECD	923
					1	CAD, IRMPD	1153
1653.7734	1653.765	0.008	HHHHHHGAVLNYDV	b_{14}	2	CAD, IRMPD, aiECD	1153, 923
1790.8323	1790.824	0.009	HHHHHHGAVLNYDVH	b_{15}	2	CAD, IRMPD aiECD	1153
					2,3	IRMPD	923
1807.8589	1807.849	0.01	HHHHHHGAVLNYDVH	c_{15}	2	ECD	923
1885.9058	1885.902	0.004	HHHHHHGAVLNYDVHL	b_{16} -H_2O	2	IRMPD	1153
1885.9494	1885.902	0.048	N(GTLTIHKYEQEK)EIM*	y_4 -H_2O	2	IRMPD	1153
1903.9164	1903.907	0.009	HHHHHHGAVLNYDVHL	b_{16}	2	IRMPD	923
1903.96	1903.909	0.051	N(GTLTIHKYEQEK)EIM*	y_4	3	IRMPD	923
					2	CAD, IRMPD, aiECD	1153
1920.9429	1920.932	0.011	HHHHHHGAVLNYDVHL	c_{16}	2	ECD	923
2048.9692	2048.963	0.006	HHHHHHGAVLNYDVHLY	b_{17} -H_2O	2	IRMPD	1153
2066.9797	2066.971	0.009	HHHHHHGAVLNYDVHLY	b_{17}	2,3	IRMPD	923
					2	CAD, IRMPD, aiECD	1153
2119.011	2118.134	0.877	HHHHHHGAVLNYDVHLYP	a_{18} -NH_3	2	IRMPD	923
2521.3501	2521.339	0.009	PK(LPVT)N(GTLTIHKYEQEK)EIM*	y_6 -H_2O	2	CAD, IRMPD, aiECD	1153
					2	IRMPD	923
2539.3606	2539.353	0.008	PK(LPVT)N(GTLTIHKYEQEK)EIM*	y_6	2	CAD, IRMPD	923
					2	CAD, IRMPD, aiECD	1153
2684.3698	2684.4	−0.03	HHHHHHGAVLNYDVHLYPK(LPVT)	b_{19} -H_2O	2	CAD	923
2684.4134	2684.404	0.009	YPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_7 -H_2O	2	CAD, IRMPD, aiECD	1153
					2	IRMPD	923
2702.3804	2702.413	−0.03	HHHHHHGAVLNYDVHLYPK(LPVT)	b_{19}	2	CAD	923
2702.424	2702.415	0.01	YPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_7	2	IRMPD	923
					2	CAD	1153
2719.4069	2719.396	0.011	HHHHHHGAVLNYDVHLYPK(LPVT)	c_{19}	2	ECD	923
					2	aiECD	1153
2797.4975	2797.496	0.002	LYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_8 -H_2O	2,3	IRMPD	923
					2	CAD	923
2815.508	2815.496	0.012	LYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_8	2,3	IRMPD	923
					2	CAD	923, 1153
2934.5564	2934.559	−0.003	HLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_9 -H_2O	2	CAD, IRMPD	1153
2952.5669	2952.559	0.008	HLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_9	2	CAD	1153
					3	CAD	923
3033.6248	3033.618	0.007	VHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{10} -H_2O	3	CAD, IRMPD, aiECD	923
					2,3	CAD, IRMPD, aiECD	1153
3051.6354	3051.627	0.009	VHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{10}	3	CAD, IRMPD, aiECD	923
					2,3	CAD, IRMPD, aiECD	1153
3166.6623	3166.664	−0.002	DVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{11}	2	CAD	1153
3311.7151	3311.713	0.002	YDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{12} -H_2O	3	IRMPD	1153
3425.758	3425.75	0.008	NYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{13} -H_2O	2	CAD	1153
3443.7686	3443.744	0.025	NYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{13}	2	CAD	1153
3538.8421	3538.842	0	LNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{14} -H_2O	3	IRMPD	923
3556.8526	3556.842	0.011	LNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{14}	3	CAD	923
					2	CAD	1153
3655.921	3655.906	0.015	VLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{15}	3	CAD	923
3726.9582	3726.943	0.015	AVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{16}	3	CAD	923
3765.9691	3765.953	0.016	GAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{17} -H_2O	3	CAD, IRMPD	923
3783.9796	3783.972	0.007	GAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{17}	3	CAD	923, 1153

Monoisotopic mass			Daughter ion				Parent ion(s)
Calculated*	Observed+	$\Delta_{calc-obs}$	proposed structure	Name●	Charge (z)	Method(s)	
3903.028	3903.018	0.01	HGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{18} -H_2O	3	CAD, IRMPD	1153
					3	CAD	923
					3,4	IRMPD	923
3911.0205	3911.007	0.014	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHK)	$b_{7'}$	3	CAD	1153
3921.0385	3921.023	0.016	HGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{18}	3	CAD	1153
					3,4	CAD	923
3965.9787	3965.968	0.011	HHHHHHGAVLNYDVHLYPK(LPVT)N(KYEQEK)	$y_{6'}$ -H_2O	3	CAD, IRMPD	1153
					4	IRMPD	923
3967.9705	3967.95	0.02	HHHHHHGAVLNYDVHLYPK(LPVT)N(KYEQEK)	$z_{6'}$	2	ECD	1153
					3	ECD	923
3983.9892	3983.981	0.008	HHHHHHGAVLNYDVHLYPK(LPVT)N(KYEQEK)	$y_{6'}$	3	CAD	1153
					2	ECD	1153
4040.0869	4040.076	0.011	HHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{19} -H_2O	3	ECD, IRMPD, aiECD	1153
					3,4	IRMPD	923
4057.0572	4057.064	−0.01	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKY)	$b_{8'}$ -NH_3	4	IRMPD	923
					3	IRMPD	1153
4058.0974	4058.072	0.026	HHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{19}	3	CAD	1153
					4	CAD	923
4074.0838	4074.064	0.02	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKY)	$b_{8'}$	3	CAD	1153
4091.1103	4091.103	0.007	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKY)	$c_{8'}$	3	ECD	923
4103.0376	4103.022	0.016	HHHHHHGAVLNYDVHLYPK(LPVT)N(HKYEQEK)	$y_{7'}$ -H_2O	3	CAD	1153
					4	IRMPD	923
4121.0481	4121.037	0.011	HHHHHHGAVLNYDVHLYPK(LPVT)N(HKYEQEK)	$y_{7'}$	4	CAD	923
					3	CAD	1153
4177.1458	4177.139	0.007	HHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{20} -H_2O	4	IRMPD	923
					3	CAD, IRMPD, aiECD	1153
4178.0696	4178.054	0.016	HHHHHHGAVLNYDVHLYPKN(GTLTIHKYEQEK)EIM*	$y_{1''}$ -H_2O	4	IRMPD	1153
4178.1298	4178.059	0.071	HHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{20} -NH_3	4	IRMPD, aiECD	1153
4179.0536	4179.129	−0.08	HHHHHHGAVLNYDVHLYPKN(GTLTIHKYEQEK)EIM*	$y_{1''}$ -NH_3	3	ECD	923
4179.1376	4179.129	0.009	HHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	z_{20}	3	ECD	923
4185.1158	4185.104	0.012	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYE)EIM*	$b_{9'}$ -H_2O	3	CAD, IRMPD, aiECD	1153
					4	IRMPD	923
4195.1564	4195.145	0.011	HHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{20}	3,4	CAD, IRMPD	1153
4203.1264	4203.113	0.013	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYE)EIM*	$b_{9'}$	3	CAD, IRMPD, aiECD	1153
					4	CAD, IRMPD	923
4220.1529	4220.148	0.005	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYE)EIM*	$c_{9'}$	3	ECD	923
4244.1642	4244.069	0.095	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)	c_{20}	4	IRMPD, aiECD	1153
4279.1173	4279.103	0.015	HHHHHHGAVLNYDVHLYPK(T)N(GTLTIHKYEQEK)EIM*	$y_{2''}$ -H_2O	3,4	CAD, IRMPD, aiECD	1153
					4,5	IRMPD	923
4281.1091	4281.104	0.005	HHHHHHGAVLNYDVHLYPK(T)N(GTLTIHKYEQEK)EIM*	$z_{2''}$	3	ECD	1153, 923
4297.1278	4297.11	0.018	HHHHHHGAVLNYDVHLYPK(T)N(GTLTIHKYEQEK)EIM*	$y_{2''}$	3,4	CAD	1153
					4	CAD	923
					4,5	IRMPD	923
					4	IRMPD, aiECD	1153
4314.1584	4314.188	−0.03	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQ)EIM*	$b_{10'}$ -NH_3	4	CAD	923
4314.2047	4314.194	0.011	HHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{21} -H_2O	4	IRMPD	923
					3	CAD, IRMPD, aiECD	1153
4319.1612	4319.151	0.01	HHHHHHGAVLNYDVHLYPK(LPVT)N(TIHKYEQEK)EIM*	$z_{9'}$	3	ECD	1153
4332.2153	4332.203	0.012	HHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{21}	4	CAD, IRMPD	923
					3	CAD, IRMPD, aiECD	1153
					4	CAD	1153
4335.1799	4335.169	0.011	HHHHHHGAVLNYDVHLYPK(LPVT)N(TIHKYEQEK)EIM*	$y_{9'}$	3	ECD	1153
4348.2115	4347.188	1.024	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQ)EIM*	$c_{10'}$	3	ECD	1153
4356.1802	4356.169	0.011	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)E	b_{21}	5	IRMPD	923
4374.1908	4374.179	0.012	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)E	b_{21} +H_2O	4	IRMPD	1153
4378.1857	4378.18	0.005	HHHHHHGAVLNYDVHLYPK(VT)N(GTLTIHKYEQEK)EIM*	$y_{3''}$ -H_2O	3	CAD	1153
					4	IRMPD, aiECD	1153
					5	IRMPD	923
4380.1775	4380.17	0.007	HHHHHHGAVLNYDVHLYPK(VT)N(GTLTIHKYEQEK)EIM*	$z_{3''}$	3	ECD	1153
4396.1963	4396.188	0.008	HHHHHHGAVLNYDVHLYPK(VT)N(GTLTIHKYEQEK)EIM*	$y_{3''}$	5	IRMPD, aiECD	923
					4	CAD, IRMPD, aiECD	1153
					3	CAD	1153

Monoisotopic mass			Daughter ion				Parent ion(s)
Calculated*	Observed+	$\Delta_{calc-obs}$	proposed structure	Name●	Charge (z)	Method(s)	
4424.2428	4424.193	0.05	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EI	a_{22} -NH_3	4	IRMPD, aiECD	1153
4432.2326	4432.227	0.005	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQE)EIM*	$a_{11'}$	4	ECD	923
					3	ECD	1153
4432.2452	4432.241	0.004	HHHHHHGAVLNYDVHLYPK(LPVT)N(LTIHKYEQEK)EIM*	$z_{10'}$	5	IRMPD	923
					3	ECD	1153
4442.217	4442.206	0.011	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQE)EIM*	$b_{11'}$ -H_2O	4,5	IRMPD	923
					4	CAD	1153
					3,4	IRMPD, aiECD	1153
4451.2537	4451.251	0.003	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EI	b_{22} -H_2O	4	aiECD	1153
4451.2636	4451.256	0.008	HHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{22} -H_2O	4	CAD, IRMPD	1153
					5	IRMPD	923
4453.2555	4453.249	0.006	HHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	z_{22}	3	ECD	1153
4460.2276	4460.223	0.004	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQE)EIM*	$b_{11'}$	3,4	CAD	1153
					4	IRMPD, aiECD	1153
					4	CAD	923
					4,5	IRMPD	923
					3	ECD	1153
4469.2643	4469.256	0.008	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EI	b_{22} or	5	IRMPD	923
4469.2742	4469.256	0.018	HHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EIM*	y_{22}	4	CAD, IRMPD, aiECD	1153
					4,5	CAD	923
4477.2541	4477.242	0.012	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQE)EIM*	$c_{11'}$	4	aiECD	1153
					4	ECD	923
					3	ECD	1153
4478.2381	4478.229	0.01	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQE)EIM*	$b_{11'}$ +H_2O	5	IRMPD	923
					4	CAD, IRMPD	1153
4487.2748	4487.266	0.009	HHHHHHGAVLNYDVHLYPK(LPVT)N(GTLTIHKYEQEK)EI	b_{22} +H_2O	5	CAD, IRMPD, aiECD	923
					4	CAD, IRMPD, aiECD	1153
4493.249	4492.274	0.975	HHHHHHGAVLNYDVHLYPK(PVT)N(GTLTIHKYEQEK)EIM*	$y_{4''}$	4	CAD	1153
4531.3011	4531.275	0.026	HHHHHHGAVLNYDVHLYPK(LPVT)N(TLTIHKYEQEK)EIM*	$y_{11'}$ -H_2O	5	IRMPD, aiECD	923
4533.2929	4533.289	0.004	HHHHHHGAVLNYDVHLYPK(LPVT)N(TLTIHKYEQEK)EIM*	$z_{11'}$	3	ECD	1153
4549.3116	4549.271	0.041	HHHHHHGAVLNYDVHLYPK(LPVT)N(TLTIHKYEQEK)EIM*	$y_{11'}$	5	IRMPD	923

*Internal ions are not included in this table.

+Observed mass values from each fragmentation method were averaged.

●''Daughter ions that arose from the LPVT branch. ' denoted daughter ions that arose from the GTLTIHKYEQEK branch. Other ions denote daughter ions that arose from the HHHHHHGAVLNYDVHLYPKNEIM* branch. M* denotes a homoseryl residue.

Table S3. ESI-FTMS of $BcpA_{His6}$ and its variants

Plasmid	Allele	Average molecular weight		$\Delta_{calc-obs}$	Isopeptide bonds
		Calculated	Observed		
pJB7	wt	56697.27	56648.60417	48.66583	3
pJB97	E472A	56639.24	56603.81769	35.42231	2
pJB98	E223A	56639.24	56603.83043	35.40957	2
pJB105	E472A, E223A	56581.2	56564.34909	16.85091	1

Table S4. MALDI-MS/MS CAD fragmentation analysis of compound A

m/z				
Calculated	Observed	$\Delta_{calc-obs}$	Proposed structure	Ion type*
86.0964	86.1496	−0.0532	L	Immonium
110.0713	110.1216	−0.0503	H	Immonium
247.1302	247.1691	−0.0389	HH	a_2
275.1251	275.174	−0.0489	HH	b_2
365.1932	365.2299	−0.0367	HGAV	i
384.1891	384.2241	−0.035	HHH	a_3
393.2496	394.1976	−0.948	LPVT	$b_{4''}$-H_2O
412.184	412.2247	−0.0407	HHH	b_3
502.2521	502.2945	−0.0424	HHGAV	i
502.3024	502.2945	0.0079	LYPK	i
521.248	521.265	−0.017	HHHH	a_4
549.2429	549.226	0.0169	HHHH	b_4
577.298	577.2341	0.0639	VLNYD	i-28
			LNYDV	i-28
639.311	639.3176	−0.0066	HHHGAV	i
639.3613	639.3176	0.0437	HLYPK	i
658.3069	658.2913	0.0156	HHHHH	a_5
686.3018	686.2844	0.0174	HHHHH	b_5
729.3791	729.3043	0.0748	HHGAVLN	i
714.357	714.3141	0.0429	LNYDVH	i-28
			NYDVHL	i-28
742.3519	742.3034	0.0485	LNYDVH	i
			NYDVHL	i
795.3658	795.3447	0.0211	HHHHHH	a_6
823.3607	823.3509	0.0098	HHHHHH	b_6
866.438	866.3511	0.0869	HHHGAVLN	i
870.4104	870.3921	0.0183	HGAVLNYD	i
880.3822	880.3527	0.0295	HHHHHHG	b_7
923.4244	923.3816	0.0428	HHHHHHGA	a_8
951.4193	951.3538	0.0655	HHHHHHGA	b_8
969.4789	969.4863	−0.0074	GAVLNYDVH	i
969.4789	969.4863	−0.0074	HGAVLNYDV	i
1007.4694	1007.4189	0.0505	HHGAVLNYD	i
1022.4928	1022.4398	0.053	HHHHHHGAV	a_9
1050.4877	1050.5219	−0.0342	HHHHHHGAV	b_9
1106.5378	1106.5409	−0.0031	HHGAVLNYDV	i
1144.5283	1144.5115	0.0168	HHHGAVLNYD	i
1163.5718	1163.533	0.0388	HHHHHHGAVL	b_{10}
1202.5953	1202.4943	0.101	HGAVLNYDVHL	i-NH_3
1245.6262	1245.6366	−0.0104	GAVLNYDVHLY	i
1277.6147	1277.5817	0.033	HHHHHHGAVLN	b_{11}
1356.6376	1356.6807	−0.0431	HHGAVLNYDVHL	i
1440.6781	1440.6433	0.0348	HHHHHHGAVLNY	b_{12}
1493.7397	1493.7289	0.0108	HHHGAVLNYDVHL	i
1555.705	1555.696	0.009	HHHHHHGAVLNYDV	b_{14}
1791.8323	1791.8667	−0.0344	HHHHHHGAVLNYDVH	b_{15}
1797.9183	1797.666	0.2523	HHGAVLNYDVHLYPKA	i-H_2O
1904.9164	1904.9685	−0.0521	HHHHHHGAVLNYDVHL	b_{16}
2039.9848	2040.0516	−0.0668	HHHHHHGAVLNYDVHLY	a_{17}
2068.1114	2068.0454	0.066	NYDVHLYPK(-LPVT)AEIM*	y_{14}
2295.2383	2295.312	−0.0737	AVLNYDVHLYPK(-LPVT)AEIM*	y_{17}
2314.1628	2314.2729	−0.1101	HHHHGAVLNYDVHLYPKAEI	i-H_2O
2432.2973	2432.3677	−0.0704	GAVLNYDVHLYPK(-LPVT)AEIM*	y_{18}
2569.3562	2569.4868	−0.1306	HGAVLNYDVHLYPK(-LPVT)AEIM*	y_{19}
2707.3389	2707.4758	−0.1369	HHHHHHGAVLNYDVHLYPKAEIM*	$y_{1''}$
2843.474	2843.4099	0.0641	HHHGAVLNYDVHLYPK(-LPVT)A	y_{21}

*''

Daughter ions that arose from the LPVT branch. No mark denotes daughter ions that arose from the HHHHHHGAVLNYDVHLYPKAEIM* branch. M* denotes a homoserine lactone residue. i denotes internal ions.

Table S5. Primers used in this study

Primer	Restriction site	Nucleic acid sequence (5′-3′)
P2	XbaI	AAAtctagaGCACACTATTGCTTTTAAGAAGG
P73	none	GATGTTCACATGTATCCC**GCA**AATGAGATTAAACGTGG
P74	none	CCACGTTTAATCTCATT**TGC**GGGATACATGTGAACATC
P88	KpnI	AAAggtaccAATACCTCCCAAAACAAGATTTC
P89	KpnI	AAAggtaccATCCTATTGTTATGTGTGTTTCTA
P103	KpnI	AAAggtaccTTATCTTTGAATTTCCGGTCCC
P113	NheI	AAAgctagcATGCATCACCATCACCATCACGGTAAGGTGCTAAACTATGATG
P114	NheI	AAAgctagcCTCTTTATTCGTTAATGGGATATC
P122	none	CTAAACTATGATGTTCAC**CTG**TATCCCAAAAATGAGATT
P123	none	AATCTCATTTTTGGGATA**CAG**GTGAACATCATAGTTTAG
P149	none	CACCATCACCATCACGGT**GCG**GTGCTAAACTATGATGTTC
P150	none	GAACATCATAGTTTAGCAC**CGC**ACCGTGATGGTGATGGTG
P151	NheI	AAAgctagcATTTCTAGCAAACTTTTTTCTAAAG
P152	NheI	AAAgctagcTCCCCTATACTAGGTTATTGG
P153	NcoI	AAAccatggTTAGTCACGATGAATTCCCGGGG
P154	NcoI	AAAccatggGATTTTAGGAACATGTAACCATAT
P181	none	AATAATAAAAGTGGATGG**ATG**CTTCCGGTAACGGGTGG
P182	none	CCACCCGTTACCGGAAG**CAT**CCATCCACTTTTATTATT
P183	none	GAGACTATCAAATTTTT**GCA**ACGAAAGCACCGACATAT
P184	none	ATATGTCGGTGCTTTCGT**TGC**AAAAATTTGATAGTCTC
P185	none	GAGAATACTATTTCCAA**GCA**ACAAAAGCTCCGAAAGG
P186	none	CCTTTCGGAGCTTTTGT**TGC**TTGGAAATAGTATTCTC
P191	none	GTATCCCAAAAATGAGATT**ATG**CGTGGTGCAGTGGATTTAA
P192	none	TTAAATCCACTGCACCACG**CAT**AATCTCATTTTTGGGATAC
P197	none	GTTCACATGTATCCCAAAG**CTG**AGATTAAACGTGGTGCA
P198	none	TGCACCACGTTTAATCTC**AGC**TTTGGGATACATGTGAAC
P201	none	GTTCACCTGTATCCCAAA**GCT**GAGATTATGCGTGGTGCA
P202	none	TGCACCACGCATAATCTC**AGC**TTTGGGATACAGGTGAAC

Noncapitalized letters indicate the sequence of the restriction site inserted. Bold letters indicate nucleotide changes introduced. Underlined nucleotides encode for the MH_6 tag.