Supporting Information

Kwon et al. 10.1073/pnas.1316855111

SI Materials and Methods

Protein Expression and Purification. C1 and C2 constructs were transformed into *Escherichia coli* BL21 (DE3) cells. Cells were expressed in ZYP-5052 autoinduction media (1) for 20 h at 28 °C. Cells were harvested by centrifugation (4,000 × *g* at 277 K for 20 min) and resuspended in lysis buffer [50 mM Tris·HCl (pH 8.0), 300 mM NaCl, 20 mM imidazole] supplemented with Mini EDTA-free protease inhibitor (Roche). Cells were lysed by sonication, and the lysate was clarified by centrifugation (13,000 × *g* for 30 min). The supernatant was then filtered (0.2 μm) and loaded onto a 5-mL HiTrap Chelating column (GE Healthcare) charged with Ni^{2+} ions and equilibrated with the lysis buffer. After washing the loaded column with the lysis buffer, proteins were eluted by a gradient of 20–500 mM imidazole. The proteins were further purified by size-exclusion chromatography using a Superdex 75 16/60 column equilibrated with 50 mM Tris·HCl (pH 8.0) and 150 mM NaCl. Fractions containing the target proteins were concentrated to 100 mg/mL in the elution buffer, 50 mM Tris·HCl (pH 8.0), and 150 mM NaCl.

For selenomethionine (SeMet) C2 expression, the C2 construct was transformed into DL41 (DE3) cells (Stratagene) and expressed in autoinduction media according to the method of Sreenath et al. (2). SeMet C2 protein was purified similar to native C2.

The C1 mutant proteins were expressed and purified as for the WT protein but eluted earlier than the WT from the size-exclusion column, and in broad peaks suggestive of multiple and possibly aggregated species. The mutant proteins could be concentrated, but only to ∼50 mg/mL.

Crystallization. Crystallization experiments of trypsin-treated C2 and native C1 were carried out by sitting drop vapor diffusion (100 nL of protein and 100 nL of precipitant) at 291 K using a Cartesian nanoliter dispensing robot (Genomic Systems) and a locally compiled crystallization screen with 480 conditions (3). Subsequent fine screens were performed in hanging drop vapor diffusion experiments with varying concentrations of protein. In these optimizations, reservoir solution (1 μL) was added to 1 μL of protein on a siliconized coverslip (Hampton Research) and the drops were incubated at 291 K over the reservoir solution.

Crystals for both C1 and trypsin-treated SeMet C2 were seeded by transferring small pieces of the respective native crystals into the hanging drops that had been equilibrated overnight. Multiple rounds of seeding improved the quality of the crystals, especially for the C2 construct.

The final trypsin-treated native and SeMet C2 crystals were obtained using 100 mg/mL protein [in 50 mM Tris·HCl (pH 8.0), 150 mM NaCl] equilibrated with 12.5% (vol/vol) PEG 1000, 12.5% (vol/vol) PEG 3350, 12.5% (vol/vol) 2-Methyl-2,4-pentanediol, 0.03 M $MgCl_2$, 0.03 M $CaCl_2$, and 0.1 M bicine/Trizma base (pH 8.5). C1 crystals were obtained from 100 mg/mL protein [in 50 mM Tris·HCl (pH 8.0), 150 mM NaCl] equilibrated with 10% (vol/vol) PEG 20,000, 20% (vol/vol) PEG monomethyl ether 550, 0.02 M amino acids (0.02 M sodium L-glutamate, 0.02 M DL-alanine, 0.02 M glycine, 0.02 M DL-lysine, 0.02 M DL-serine), and 0.1 M 3-morpholinopropane-1-sulfonic acid (MOPS)/Hepes (pH 7.5).

Bioinformatics Search for Internal Ester Bonds in Other Proteins. The amino acid sequence of C1 Cpe0147 was blasted against all known nonredundant sequences using the blastp (protein–protein BLAST) algorithm on the National Center for Biotechnology Information BLAST server (http://blast.ncbi.nlm.nih.gov). The sequence matches were visually inspected for conservation of the ester bond-forming residues (equivalent to Cpe0147 residues Thr-11, Asp-41, Glu-108, His-133, Asp-138, and Gln-141).

1. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41(1):207–234.
2. Sreenath HK, et al. (2005) Protocols for production of selenomethionine-labeled proteins in 2-L polyethylene terephthalate bottles using auto-induction medium. *Protein Expr Purif* 40(2):256–267.
3. Moreland N, et al. (2005) A flexible and economical medium-throughput strategy for protein production and crystallization. *Acta Crystallogr D Biol Crystallogr* 61(Pt 10): 1378–1385.

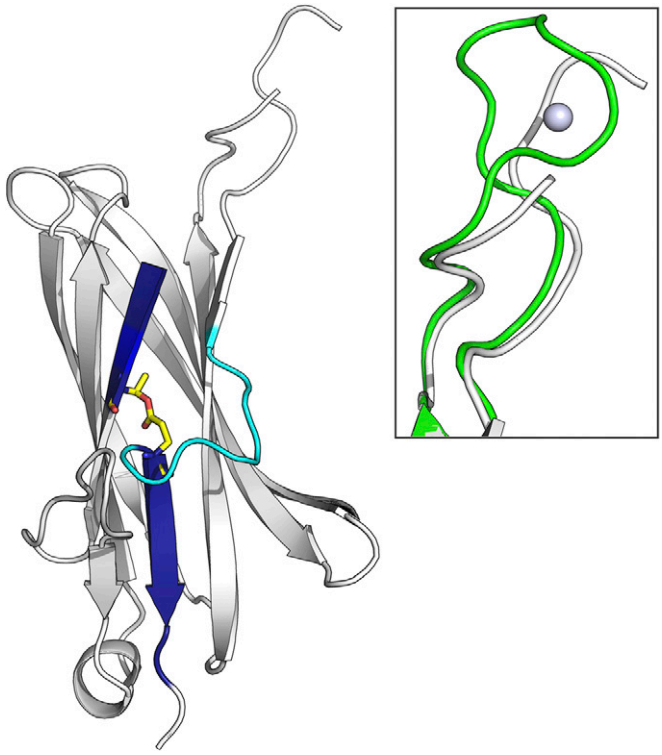

Fig. S1. Structure of the C1 construct. A single-domain C1 structure highlights sections of strands A and G (blue) linked by the internal ester bond (yellow sticks). The loop interrupting the G strand is shown in cyan. (*Inset*) Close-up view of a structural superimposition of the calcium binding F-G loop of C2 (calcium is shown as a sphere) and the same loop in C1, which shows a different conformation and disorder, such that the tip of the C1 loop could not be modeled.

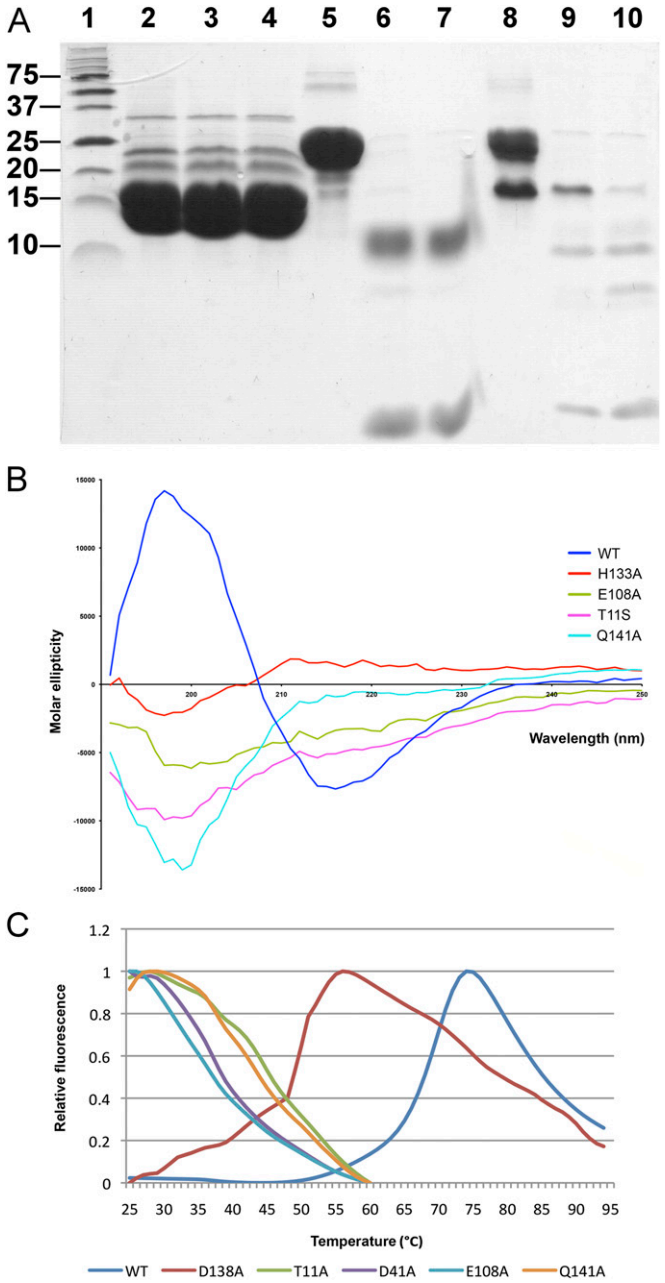

Fig. S2. Proteolytic and thermal stability data for the C1 construct. (*A*) SDS/PAGE analysis of WT, Q141A, and D138A proteins. Lane 1 shows the mass ladder (size in kilodaltons is indicated to the left), lane 2 shows untreated WT, lane 3 shows trypsin-treated WT (6-h incubation), lane 4 shows trypsin-treated WT (24-h incubation), lane 5 shows untreated Q141A, lane 6 shows trypsin-treated Q141A (6-h incubation), lane 7 shows trypsin-treated Q141A (24-h incubation), lane 8 shows untreated D138A, lane 9 shows trypsin-treated D138A (6-h incubation), and lane 10 shows trypsin-treated D138A (24-h incubation). The results show that WT protein is stable to trypsin proteolysis, whereas the Q141A (minus an internal ester bond) is readily proteolyzed. The D138A sample shows a mixed population of ester bond-formed and -unformed protein. (*B*) CD data for WT and mutant C1 proteins. The WT protein curve is indicative of a folded all β-sheet protein and is consistent with the crystal structure. The mutant proteins (H133A, E108A, T11S, and Q141A) produce curves that are indicative of random coils. (*C*) Differential scanning fluorimetry plot of SYPRO orange fluorescence vs. temperature for WT, D138A, T11A, D41A, E108A, and Q141A proteins. The WT curve is indicative of stable, folded protein. The curves corresponding to all mutants except D138A are indicative of unfolded protein. The curve for the D138A sample suggests a mixed population of ester bond-formed and -unformed protein and is consistent with the SDS/PAGE results.

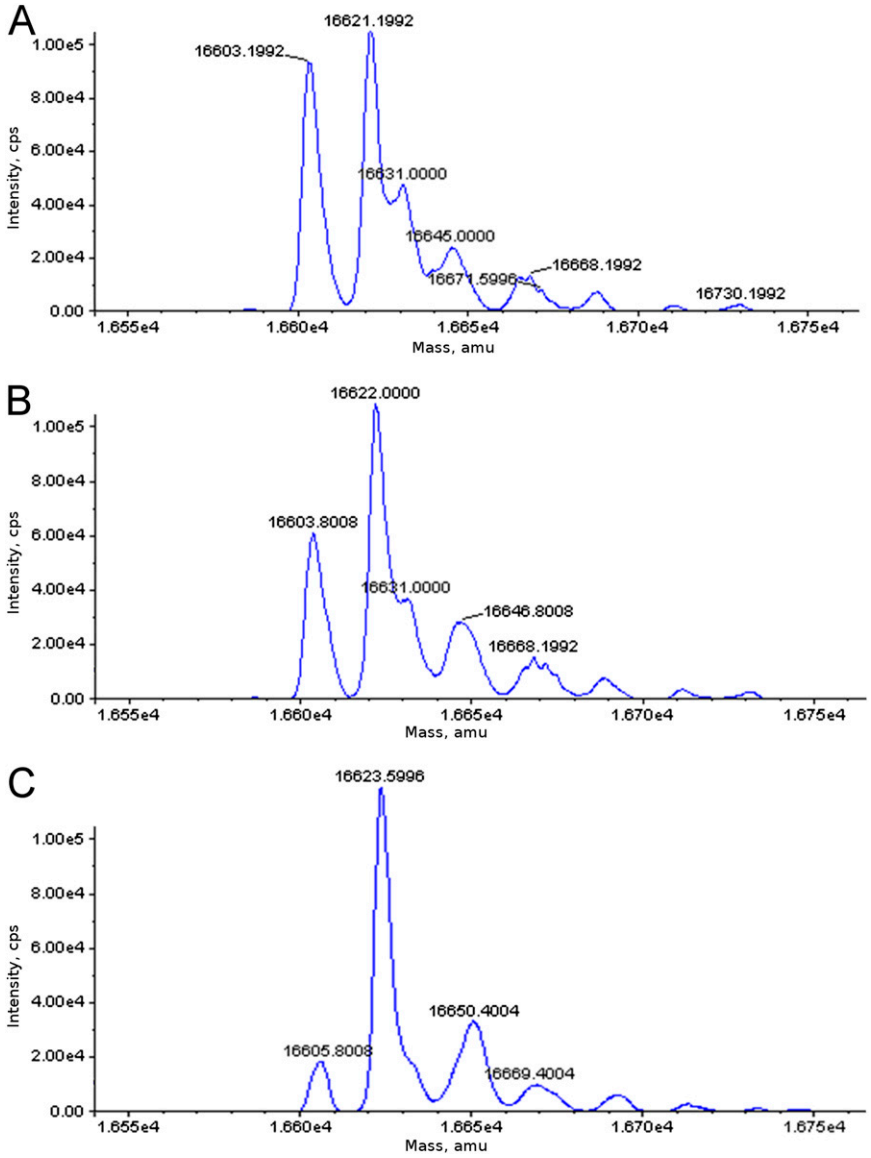

Fig. S3. Time course of ester bond hydrolysis in D138A C1 protein. (*A*) Immediately after purification, a mixed population of ester bond-formed (16,603 Da, calculated and observed) and ester bond-unformed (16,620 Da calculated, 16,621 Da observed) protein is observed at 50:50 ratio. (*B*) After incubation at 37 °C for 48 h, the formed/unformed ratio is now 40:60. (*C*) After further incubation for 150 h, the ratio has dropped further to 20:80. Ratios were calculated from integrated curve data. The time course is suggestive of a slow hydrolysis mechanism in D138A protein, as shown in Fig. S4. amu, atomic mass units.

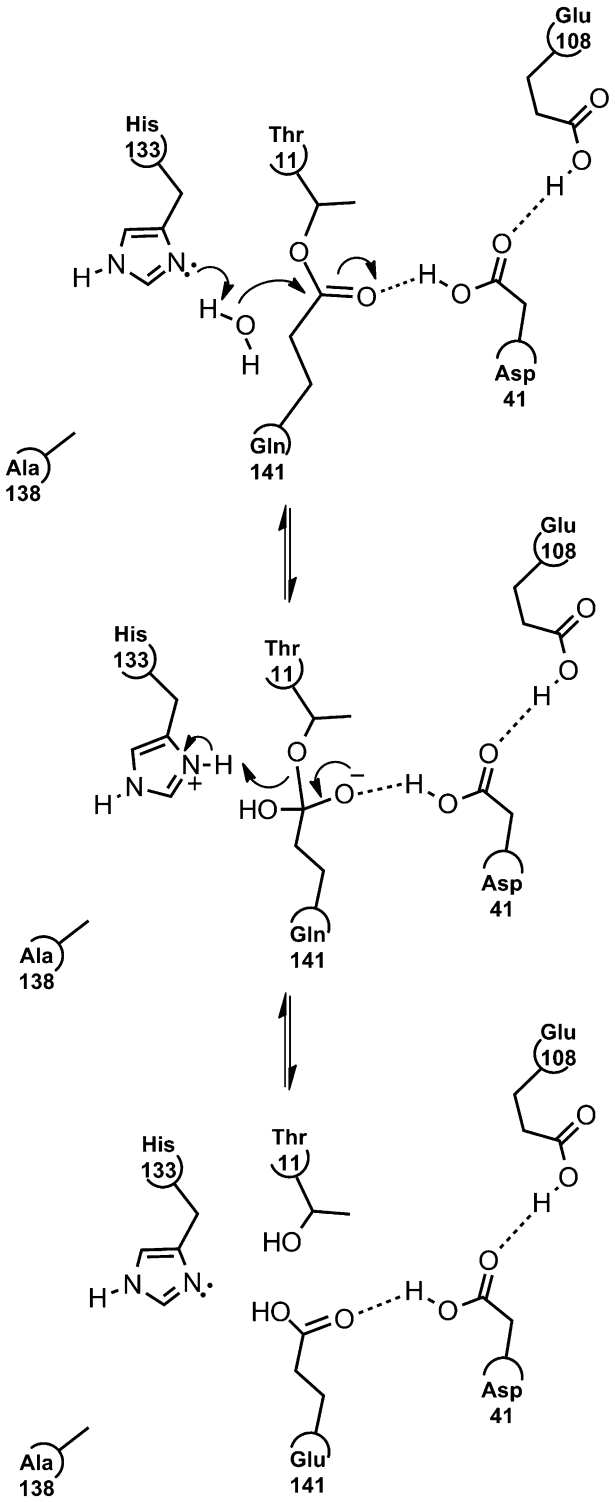

Fig. S4. Proposed D138A hydrolysis mechanism. In the first step, a water molecule initiates nucleophilic attack in concert with protein abstraction by the His-133 base. The resulting high-energy tetrahedral intermediate containing an oxyanion species is stabilized by the Asp-41/Glu-108 hydrogen bond pair. The tetrahedral species spontaneously collapses, and a proton is transferred from His-133, now acting as an acid, to Thr-11. The Thr-11 side chain is regenerated, and the Gln-141 side chain has been converted to a glutamic acid.

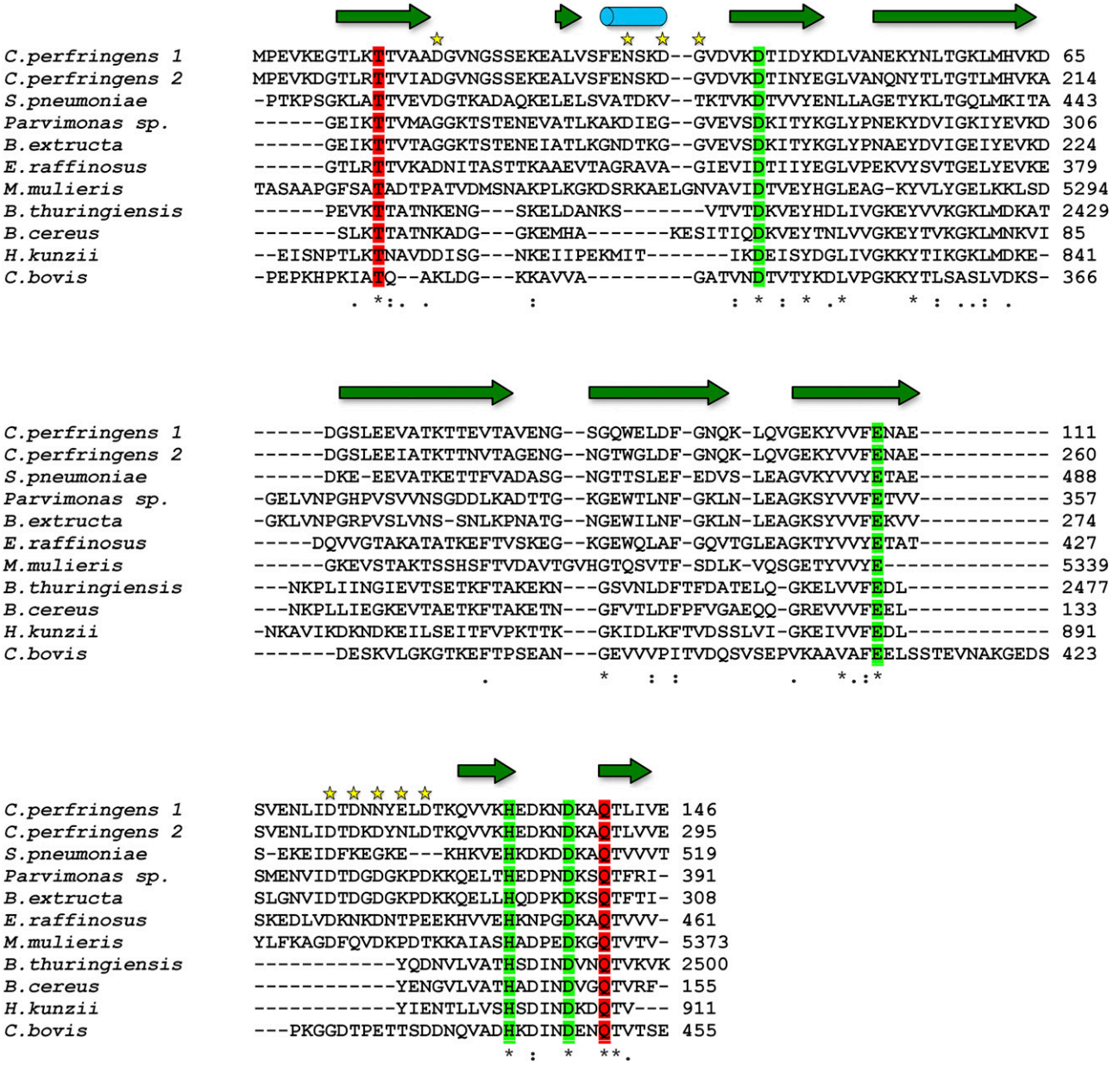

Fig. S5. Multiple sequence alignment of proteins containing potential ester bonds from various bacterial species. In each of these proteins, the cross-linking Thr and Gln residues are conserved (red), as are the accessory residues equivalent to Asp-41, Glu-108, His-133, and Asp-138 (green). The locations of secondary structure elements (β-strands, green arrows; α-helix, blue cylinder) assigned from the crystal structure and calcium-binding residues (yellow stars) are indicated above the sequences. Sequences shown are as follows: *Clostridium perfringens* (EDT72456 putative surface-anchored protein) repeat domain 1, *C. perfringens* (EDT72456 putative surface-anchored protein) repeat domain 2, *Streptococcus pneumoniae* (EHD91134 Cna B-type domain protein, partial), *Parvimonas* sp. oral taxon 393 strain F0440 (EGV09726 Cna B-type domain protein), *Bulleidia extructa* (EFC05496 Cna B-type domain protein), *Enterococcus raffinosus* (WP_010744397 LPXTG domain-containing protein cell wall anchor domain), *Mobiluncus mulieris* (EFN93563 LPXTG motif cell wall anchor domain protein), *Bacillus thuringiensis* (AFQ30314 collagen adhesin protein), *Bacillus cereus* (EJR64346 LPXTG domain-containing protein cell wall anchor domain), *Helcococcus kunzii* (EHR34835 hypothetical protein HMPREF9709_00577), and *Corynebacterium bovis* (ZP_08516882 hypothetical protein CbovD2_04897).

Table S1. Data collection, phasing, and refinement statistics

Parameters	SeMet C2	Native C1
Space group	$P2_12_12_1$	$P2_1$
Unit cell, Å	48.19, 63.45, 104.98	48.86, 28.98, 49.31
Data collection statistics		
Resolution, Å	19.6–1.9 (1.94–1.90)	19.4–1.10 (1.16–1.10)
Wavelength, Å	0.95370	0.72930
No. of unique reflections	26,100 (1,733)	49,641 (7,065)
Redundancy	14.5 (14.6)	14.9 (14.8)
Completeness, %	99.9 (99.7)	97.7 (95.9)
Mean I/σ	23.3 (1.5)	21.6 (2.2)
$CC_{1/2}$*	1.00 (0.78)	1.00 (0.89)
Phasing statistics for SeMet		
No. of sites	2 Se	N/A
Figure of merit	0.26	
Autobuilt residues	286	
CC	0.78	
Refinement statistics		
Resolution range, Å	19.6–1.90	19.4–1.10
No. of reflections	26,023	47,054
No. of reflections for R_{free}	1,332	2,571
R/R_{free}, %†	19.1/21.9	18.2/21.0
rmsd from standard bond length/angles, Å/°	0.010/1.23	0.013/1.34
Ramachandran favored, %	98.3	97.8
MOLPROBITY score	99th percentile	93rd percentile

Figures in parentheses are for the high-resolution shell. N/A, ■■■.

*Mn(I) half-set correlation $CC_{1/2}$ as calculated by SCALA.

†$R = \Sigma_{hkl} \mid |F_o(hkl)| - |F_c(hkl)| \mid / \Sigma_{khl} |F_o(hkl)|$. The R value is calculated using 95% of the data selected randomly and used in refinement. R_{free} is calculated from the remaining 5% of the data not used in refinement.

Table S2. Mutant MS details

Mutant	Molecular mass observed, Da	Molecular mass calculated, Da	$\Delta_{obs-calc}$, Da	Bond formation
WT C1	16,647	16,664	−17	Yes
WT C2	33,265	33,298	−33	Yes
T11A	16,633	16,634	−1	No
D41A	16,619	16,620	−1	No
H133A	16,598	16,598	0	No
D138A	16,619/16,603	16,620	−1/−17	Mixed
Q141A	16,606	16,607	−1	No

$\Delta_{obs-calc}$, difference between observed and calculated molecular mass.

Table S3. Tandem MS of a peptide at 676.9^{3+} containing Thr11-Gln141 ester bond of Cpe0147

Observed, m/z*	Charge	Calculated, m/z†	$\Delta_{obs-calc}$	Proposed structure‡
716.37	+3	716.37	0	AQTLIVEKPLEHHHHHH and T (−NH_3)
750.05	+3	750.05	0	AQTLIVEKPLEHHHHHH and TT (−NH_3)
783.06	+3	783.07	−0.01	AQTLIVEKPLEHHHHHH and TTV (−NH_3)
806.74	+3	806.75	−0.01	AQTLIVEKPLEHHHHHH and TTVA (−NH_3)
830.43	+3	830.43	0	AQTLIVEKPLEHHHHHH and TTVAA (−NH_3)
868.77	+3	868.77	0	AQTLIVEKPLEHHHHHH and TTVAAD (−NH_3)
887.77	+3	887.78	−0.01	AQTLIVEKPLEHHHHHH and TTVAADG (−NH_3)

*Monoisotropic masses of observed ions.

†Calculated ions. Monoisotropic masses were calculated using the fragment ion calculator (http://db.systemsbiology.net:8080/proteomicsToolkit/FragIonServlet.html).

‡Loss of 17 Da from losing NH_3 is shown in parentheses.

Table S4. Primer sequences

Primer name	Primer direction	Sequences
C1 cloning	Fw	TAT AAT GGC CAG **CAT ATG** ACC CTG AAA ACC ACC GTT GCA GC
	Rev	GGC TAT **CTC GAG** CGG TTT TTC CAC AAT CAG GGT CTG G
T11A	Fw	**GCA** ACC GTT GCA GCA GAT GGT G
	Rev	TTT CAG GGT CAT ATG TAT ATC TCC TTC TTA AAG
D41A	Fw	**GCA** ACC GTT GCA GCA GAT GGT G
	Rev	TTT CAG GGT CAT ATG TAT ATC TCC TTC TTA AAG
H133A	Fw	**GCA** GAA GAT AAA AAT GAT AAA GCC CAG AC
	Rev	TTT CAC CAC CTG TTT GGT ATC CGA TTC
D138A	Fw	**GCA** AAA GCC CAG ACC CTG ATT GTG
	Rev	ATT TTT ATC TTC GTG TTT CAC CAC CTG TTT GG
Q141A	Fw	**GCA** ACC CTG ATT GTG GAA AAA CCG
	Rev	GGC TTT ATC ATT TTT ATC TTC GTG TTT CAC

Bold letters indicate restriction enzyme sites or mutation sites. Fw, forward; Rev, reverse.