Supporting Information

CLOSTRIDIOLYSIN S: A POST-TRANSLATIONALLY MODIFIED BIOTOXIN FROM *CLOSTRIDIUM BOTULINUM*

David J. Gonzalez [1], Shaun W. Lee[9], Mary E. Hensler [6], Andrew L. Markley[1] , Samira Dahesh [6], Douglas A. Mitchell[10], Nuno Bandeira[7], Victor Nizet[2, 6], Jack E. Dixon[1, 4, 5, 8] and Pieter C. Dorrestein[1, 2, 3, 4]

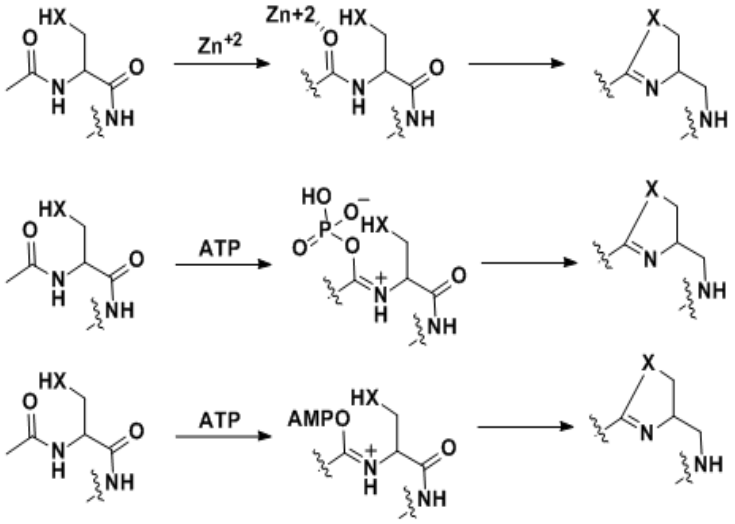
Department of Chemistry and Biochemistry[1], Skaggs School of Pharmacy and Pharmaceutical Sciences[2], Center for Marine Biotechnology and Biomedicine[3], Department of Pharmacology[4], Department of Cellular and Molecular Medicine[5], Department of Pediatrics[6], Division of Pharmacology and Drug Discovery, Department of Computer Science and Engineering[7], University of California at San Diego, La Jolla, CA 92093, Howard Hughes Medical Institute[8], Chevy Chase, MD 20815, Department of Biological Sciences[9], University of Notre Dame, Notre Dame, IN 46556, Departments of Chemistry[10], Microbiology and the Institute for Genomic Biology University of Illinois at Urbana-Champaign Urbana, IL 61801

Running title: Clostridiolysin S: A post-translationally modified biotoxin

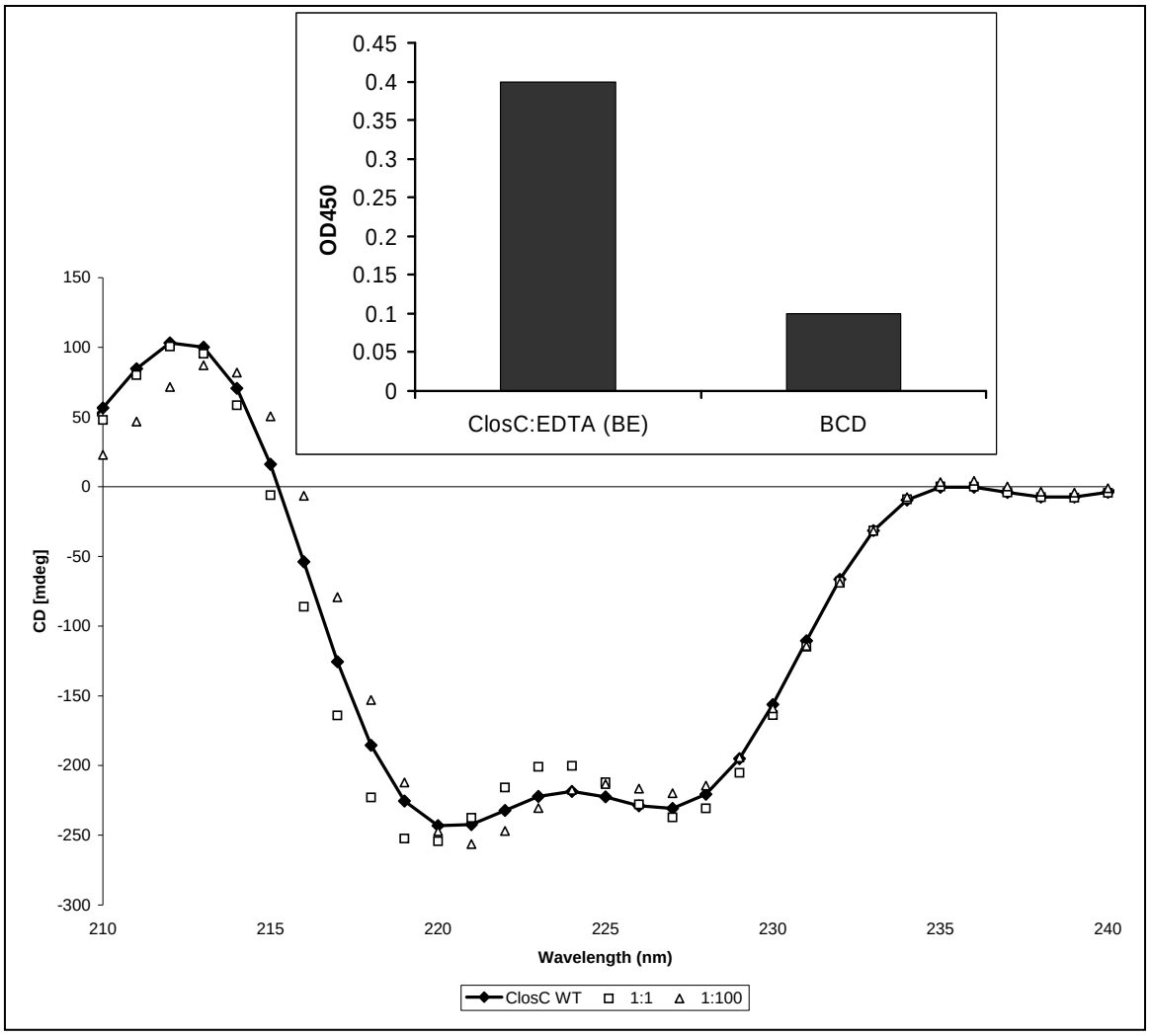
To whom correspondence should be addressed: Pieter C. Dorrestein Skaggs School of Pharmacy and Pharmaceutical Sciences and Departments of Pharmacology, Chemistry and Biochemistry, Biomedical Science Building (BSB) Room 4090 9500 Gilman Drive, MC 0636 La Jolla, CA 92093-0636 Office: 858-534-6607

E-mail: pdorrestein@ucsd.edu

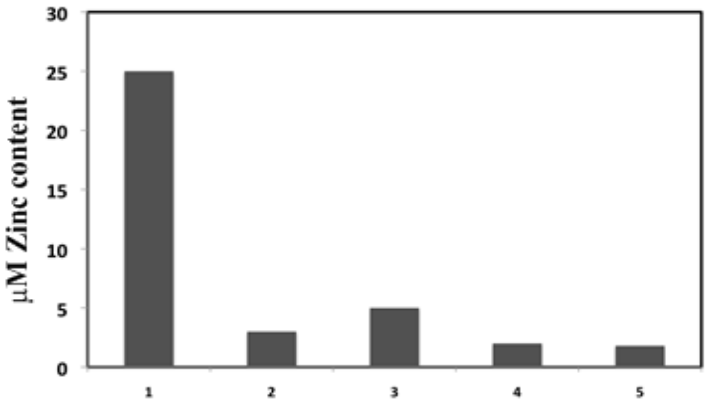
Contents



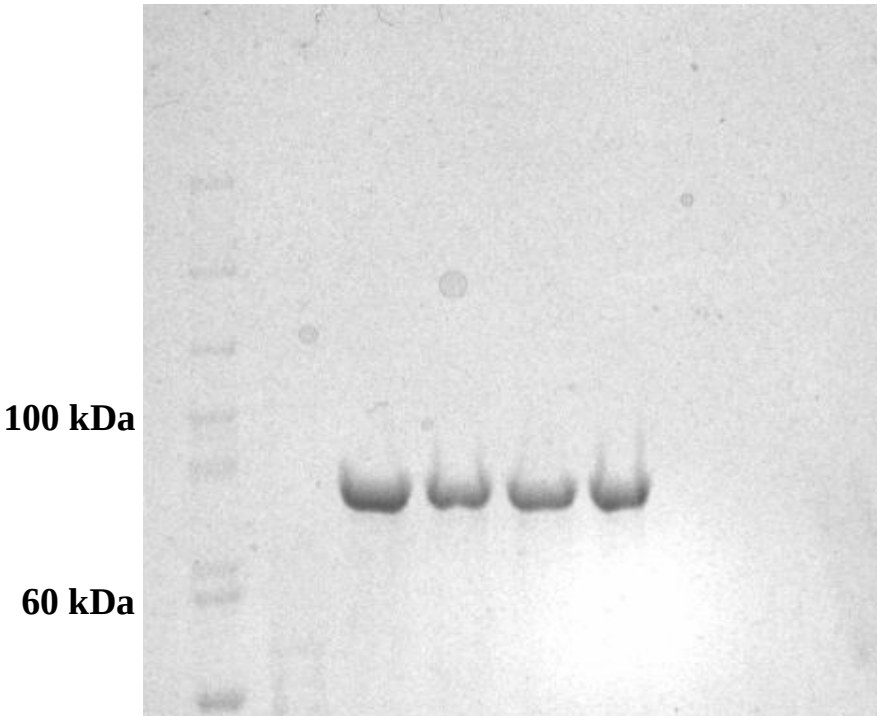
Supplemental Fig 1. Alternative mechanisms for the formation of thiazolines or (methyl)-oxazolines using Zn^{+2} or ATP as Lewis acid for carbonyl destabilization.



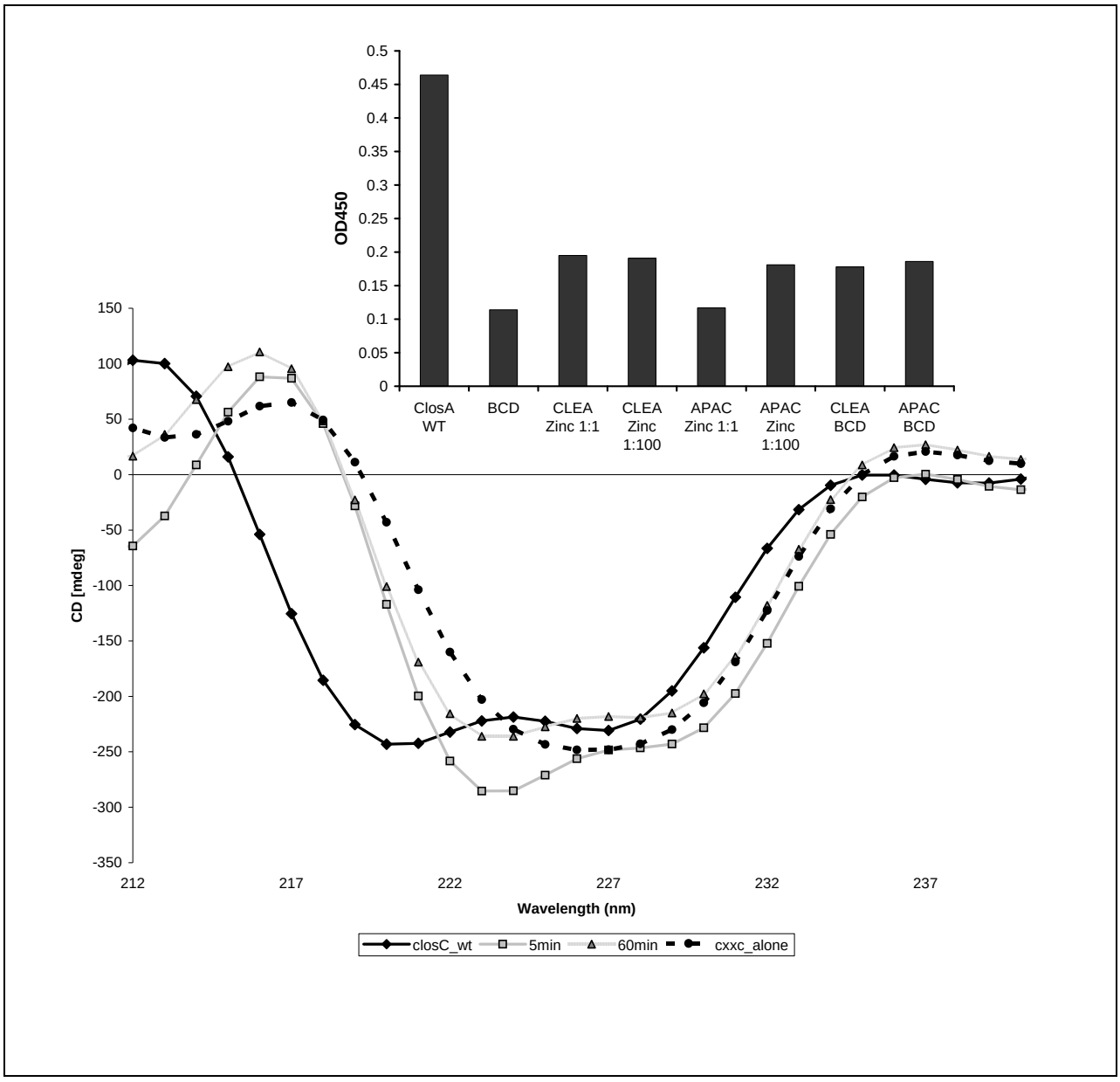
Supplemental Fig 2. MBP-ClosC was incubated with 1:1 and 1:100 protein-to-EDTA molar ration then buffer exchanged (BE). The CD shows the MBP-ClosC remains unperturbed by the EDTA and is active in the bioassays (insert, y-axis shows OD450 monitoring heme release upon erythrocyte lysis).



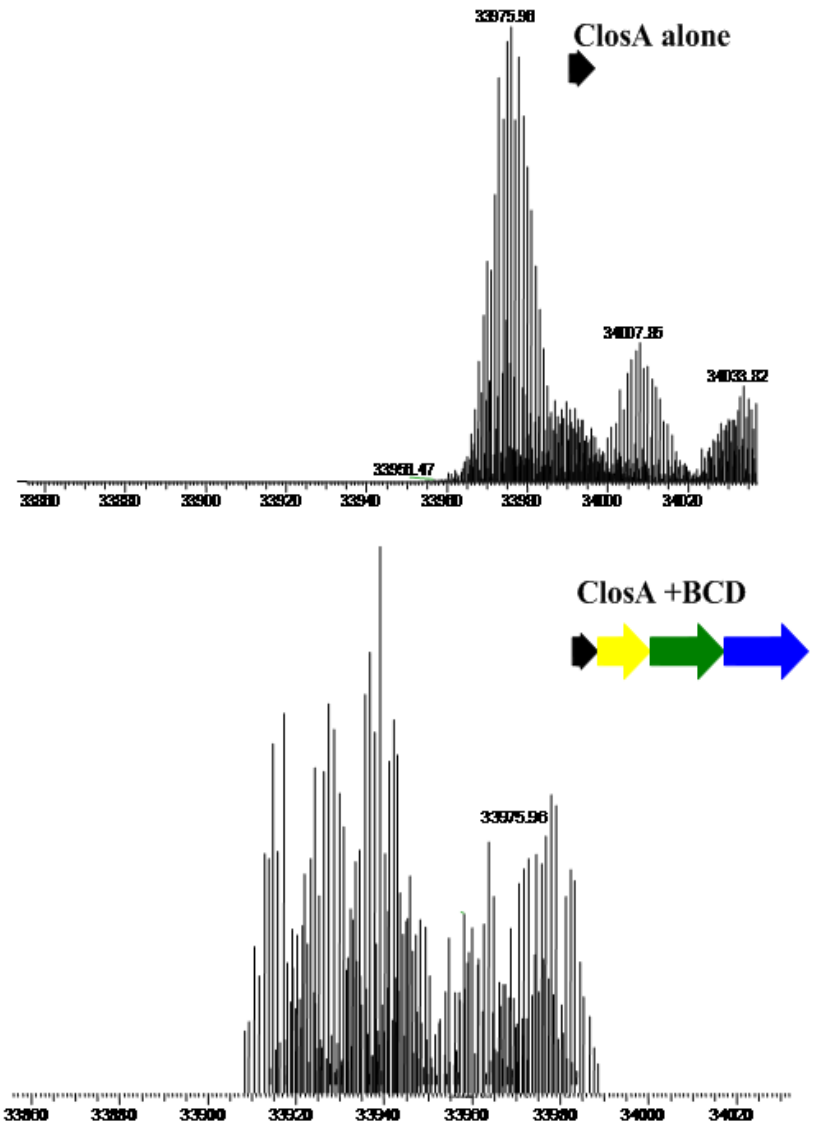
Supplemental Fig 3. proteins ICP-MS data of wt ClosC versus CxxC mutants. In order: 1. ClosC 2. ALEC 3. CLEA 4. APAC 5. CPAA. ICP-MS experiments were performed in duplicate with two different protein preparations.



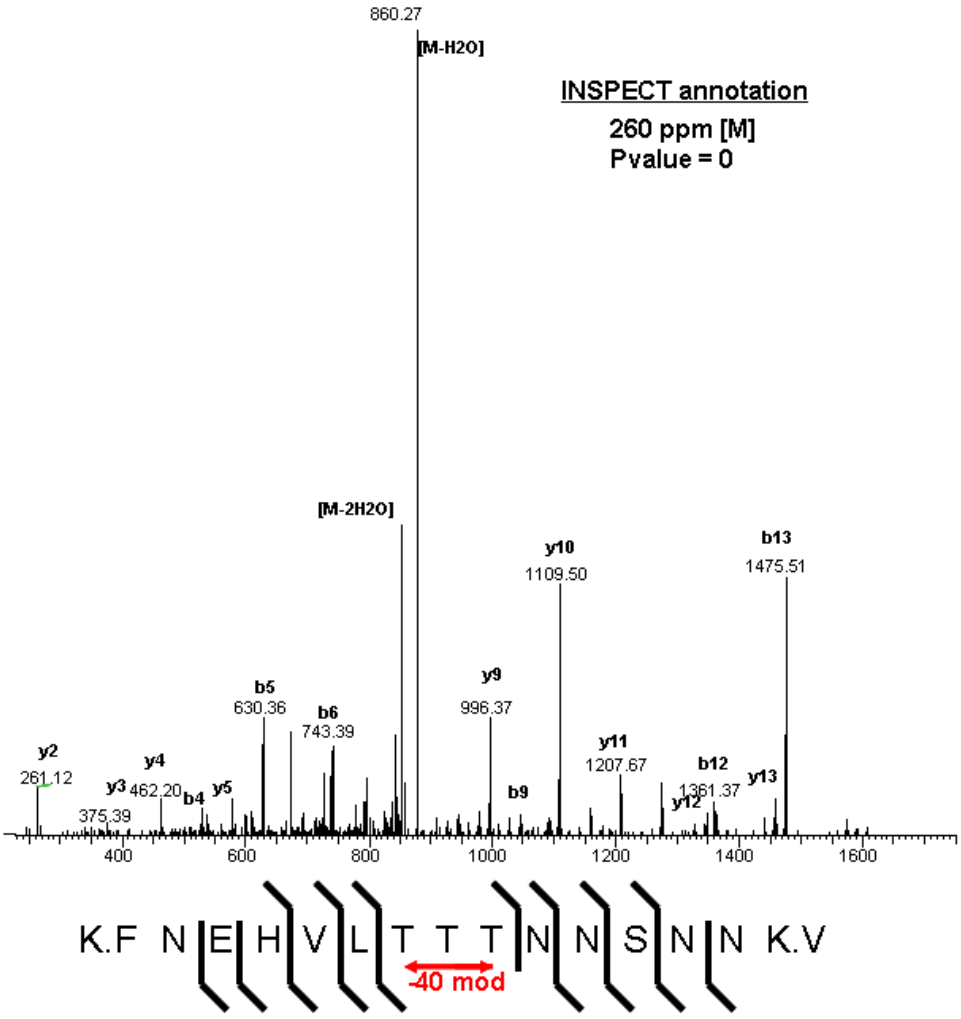
Supplemental Fig 4. SDS-PAGE of the proteins used for CD analysis. Proteins were concentrated (40kDa cut-off) and buffer exchanged. Molecular weight of MBP-ClosC is ~89kDa. 1. ALEC 2. CLEA 3. APAC 4. CPAA.



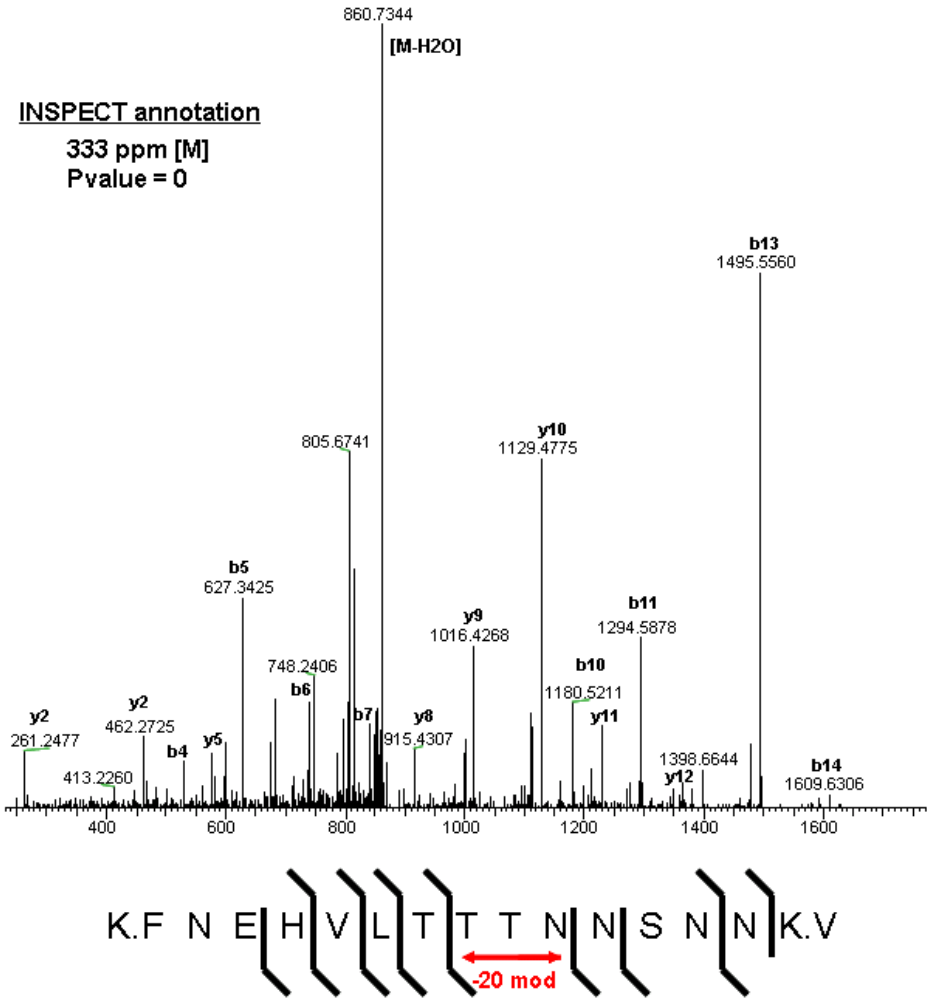
Supplemental Fig 5. CxxC proteins CLEA and APAC were incubated with zinc at a 1:1 and 1:100 protein-to-zinc molar ratio then buffer exchanged (BE). The CD shows the zinc could not restore alpha-helical content comparable to wt. MBP-ClosC. Insert shows bioactivity of the CxxC proteins upon zinc incubation (y-axis shows OD450 monitoring heme release upon erythrocyte lysis).



Supplemental Fig 6. HPLC purified recombinantly expressed ClosA was introduced into the FTICR-MS and visualized at 21ppm. Incubating ClosA with the BCD synthetase complex, followed by HPLC produced a series of overlapping ions. Manual annotation of the resulting spectra produced ambiguous mass assignments; therefore a bottom-up approach was taken to visualize the PTM occurring on ClosA.



Supplemental Fig 7 Tandem mass spectrometry of T11/T12 heterocyclic peptide.



Supplemental Fig 8 Tandem mass spectrometry of T12 heterocyclic peptide.

Additional description of Mass Spectrometry

To observe the C-terminal end of ClosA, the digested 2-bromoethylamine alkylated samples were subjected to nanocapillary–LCMS/MS analysis on an LTQ instrument. Because of the heterologous nature of the converted ClosA substrate observed in the top down analysis, it was expected that similar heterologous mixtures of modified and unmodified peptides would result in the bottom-up approach. Therefore the data was subjected to the tandem mass spectral processing program Spectral Networks. Spectral Networks processed 576,339 total spectra and identified 133,435 clusters. Of the indentified clusters 42 were observed in the C-terminal portion of ClosA (CCVSVS region).

In conjunction with Spectral Networks, the tandem MS/MS data was analyzed by InSpecT. The resulting MS/MS data analysis of ClosA outputted a series of sequence tags, which covered the C-terminal ends in a ladder type fashion that had a -20 and/or +43 Da modification. All residues in the ClosA primary sequence were annotated except C26 and C27. Each of the peptides identified to contain a 20 Da loss were manually confirmed. With exception of an -18 Da modification on T11 and the expected +43 Da modifications on Cys, Ser, Thr that were found, no other modifications were recovered in this search. We believe that the 18 Da loss found in the ClosA sample was due to a low level of source fragmentation as this ion is found to be nearly undetectable species in the broadband scan.
Figure 4 shows the tandem mass spectrum of a CLS peptide with a (methyl)-oxazole at position T46. Manual inspection of the PTM peptide shows the parental mass fragmented has a Δm= 0.42 Da (230 ppm mass error) relative to the theoretical mass. The b and y-ion series ppm error range from 5 ppm-200 ppm. Within the b and y-ion series 14/22 y-ions were annotated with 7 contingent y-ions capturing the -20 Da modification. The b-ion series annotated 7/22 with 5 contingent b-ions capturing the -20 Da modification. Because the y^{12}/b^{11} ions were not annotated and y^{13}/b^{10} were annotated, we infer the 20 Da loss is localized at T46 and thus represents methyloxazole formation. Our conclusion is supported by the chemistry which takes place upon methyloxazole formation, since the amide bond between S45 and A44 would have been lost upon heterocyclization and therefore y^{12}/b^{11} would be nonexistent. In addition to the PTM peptide, the unmodified peptide was also captured with a Δm= 0.52 (288 ppm mass error) and b and y-ion series ppm error range from 5 ppm-300 ppm. Furthermore, the analysis of the tandem mass spectrometry data identified several other peptides with multiple 20 Da losses. The first localizes heterocyclized residues at positions: T11 and T12 (Peptide 3 Supplemental Fig. 7). The parent mass of the fragmented peptide with heterocycles at T11 and T12 had a mass error of 260 ppm mass error and a b and y-ion series with ppm error ranging from 10 ppm-250 ppm. The identified peptide displayed 9 contingent y-ions all of which contained the modification. The b-ion series annotated 7/14 ions. Peptide 4 is a similar peptide as Peptide 3 but has one less heterocycle. The mass error of this peptide was 333 ppm. Our method was able to identify an SLS peptide with a 40 Da loss relative to the unmodified peptide (installation of two oxazoles by SagBCD enzymes). The identified peptide (Peptide 2) had a mass error of 71 ppm.

Primers used in studies

Intergrational mutagenesis of closA and closC

(i) ClosA_For_*EcoRI* GCGGAATTCTAATGAACACGTACTGACAACTAC) (ii) ClosA_Rev_*BamHI* GCCGGATTCGCTGCTCCGCCCACCTGTT . ClosC_For_*EcoRI* GCGGAATTCGCGGAATTCGCCCCAAAAAGTATTATTTTATATCAG (iii) ClosC_Rev_*BamHI* GCCGGATTCCACTCTAGACATCCCGTTTTAGG, respectively.

Cloning and protein purification of ClosA, ClosB, ClosC, ClosD, SagB

(i) ClosA(ForBamHI) AAG GAT CCA TGC TGA AAT TTA ACG AAC ATG TGC TGA CC ClosA(RevNotI) AAG CGG CCG CTT AGT TGC CGC CCT GAC CCG C (ii) ClosB(ForBamHI) AAG GAT CCA TGC TGC TGA AAA ACC TGA AAA AAC AGA AAG TC ClosB(RevNotI) AAG CGG CCG CTT AAT CAC CCT GTT CCC AGC CCA

CG (iii) ClosC(ForBamHI) AAA GGA TCC ATG AAA AAT AAT ACC ATC TAT CGT CTG AGC ClosC(RevNotI) AAA GCG GCC GCT TAG CTA TTA ATA TCT TCC AGA ATA CCA TCG (iv) ClosD(ForBamHI) AAA GGA TCC ATG ATC AAA TTT AGT CCG TCA TTT AAT AAT ATT CTG G ClosD(RevNotI) AAA GCG GCC GCT TAT GGC ATC GGA TGC GGG TAT TC.

In vivo Lytic activity: Clos single-gene complementation into Group A Streptococcus (GAS) Δsag A-D strains

(i) ClosA (ForXbaI) TAA TCT AGA TAA ATG CTG AAA TTT AAC GAA CAT GTG CT ClosA(RevBamHI) TAA GGA TCC TAA TTA GTT GCC GCC CTG ACC CGC CGC CGC A (ii) ClosB(ForXbaI) TAA TCT AGA TAA ATG CTG CTG AAA AAC CTG AAA AAA CAG A ClosB(RevBamHI) TAA GGA TCC TAA TTA ATC ACC CTG TTC CCA GCC CAC GAC T (iii) ClosC(ForXbaI) TAA TCT AGA TAA ATG AAA AAT AAT ACC ATC TAT CGT CTG A ClosC(RevBamHI) TAA GGA TCC TAA TTA GCT ATT AAT ATC TTC CAG AAT ACC A (iv) ClosD (ForXbaI) TAA TCT AGA TAA ATG ATC AAA TTT AGT CCG TCA TTT AAT ClosD (RevBamHI) TAA GGA TCC TAA TTA TGG CAT CGG ATG CGG GTA TTC GTT.

Sited Directed Mutagenesis of ClosC CxxC mutants

(i) ClosC_C332A TTC CGT TTT GTC CGG CGG CCG GCA ATA TTA GCA AAG C ClosC_C332A GCT TTG CTA ATA TTG CCG GCC GCC GGA CAA AAC GGA A (ii) ClosC_C329A TCT GCT GCG CGT TCC GTT TGC TCC GGC GTG C ClosC_C329A GCA CGC CGG AGC AAA CGG AAC GCG CAG CAG A (iii) ClosC_C245A GAA AAC CGG GTG CCT GGA AGC TTT TGA GCA GCG TAT TCT G ClosC_C245A CAG AAT ACG CTG CTC AAA AGC TTC CAG GCA CCC GGT TTT C (iv) ClosC_C242A GCC GCC GAA AAC CGG GGC CCT GGA ATG TTT TGA G ClosC_C242A CTC AAA ACA TTC CAG GGC CCC GGT TTT CGG CGG C ClosC_C242A_C245A GCC GCC GAA AAC CGG GGC CCT GGA AGC TTT TGA GCA GCG TAT T ClosC_C242A_C245A AAT ACG CTG CTC AAA AGC TTC CAG GGC CCC GGT TTT CGG CGG C (v) ClosC_C329A_C332A GAT CTG CTG CGC GTT CCG TTT GCT CCG GCG GCC GGC AAT ATT A ClosC_C329A_C332A TAA TAT TGC CGG CCG CCG GAG CAA ACG GAA CGC GCA GCA GAT C.

Point mutation analysis of identified heterocyclized sites on clostriodiolysin S

(i)T46A_fwd GCA GCG CGA GCG CCG GTG GTG GT T46A_rev ACC ACC ACC GGC GCT CGC GCT GC.