Supporting Information

Bioinspired strategy for the ribosomal synthesis of thioether-bridged macrocyclic peptides in bacteria **

*Nina Bionda, Abby L. Cryan, and Rudi Fasan**

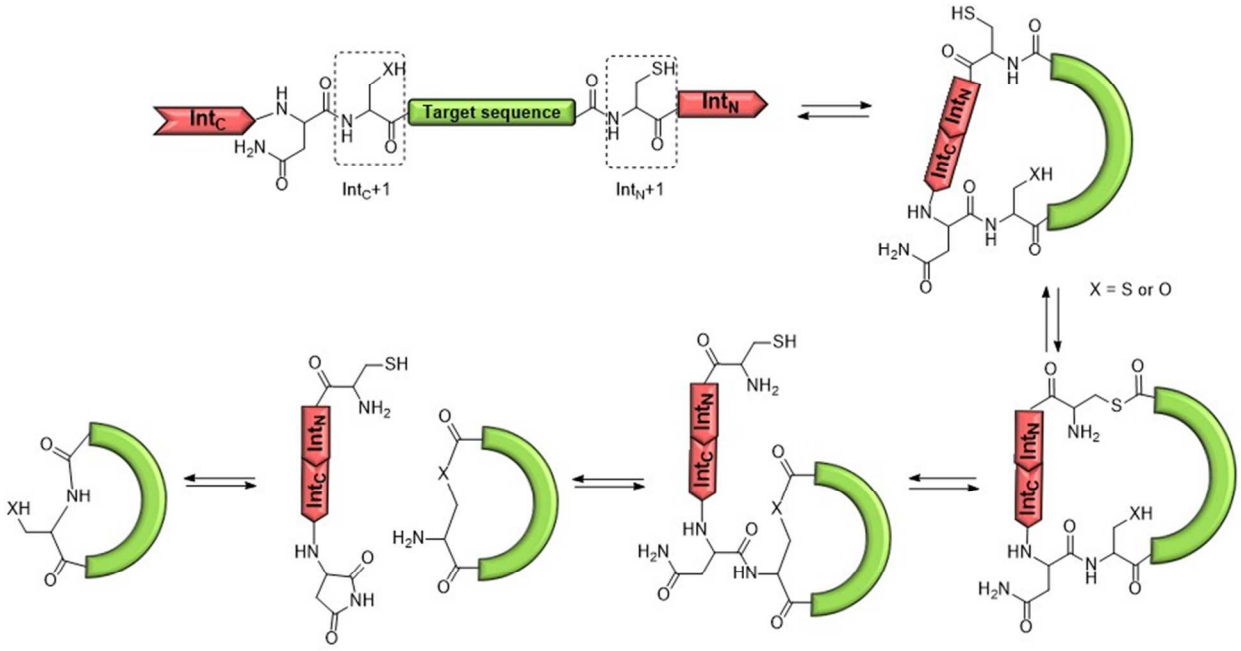
Department of Chemistry, University of Rochester, 14627 Rochester, New York, USA
Correspondence should be addressed to R.F. (fasan@chem.rochester.edu)

Table of contents:

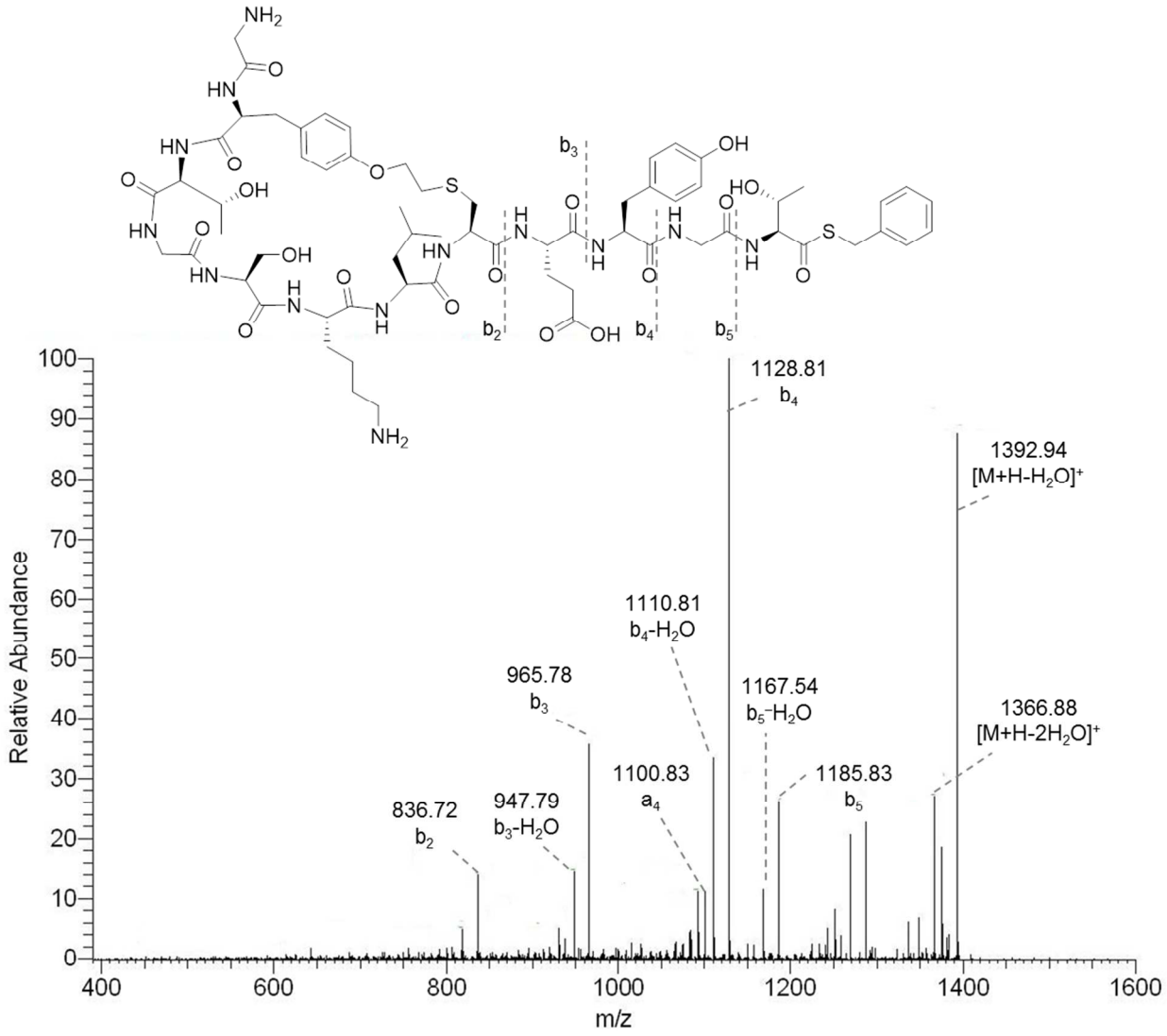
Supplementary Table S1. Oligonucleotide primers

Primer	Sequence
12mer-Z1C_for	5'-GCATCCCATGGCTAGTGCGGATCC-3'
12mer-Z2C_for	5'-GCATCCCATGGCTAGACATGCTCCAAACTG-3'
12mer-Z3C_for	5'-GCATCCCATGGCTAGACAGGATGCAAACTG-3'
12mer-Z4C_for	5'-CAACAGGATCCTGCCTGGCCGAATACGG-3'
12mer-Z5C_for	5'-CAACAGGATCCAAATGCGCCGAATACGG-3'
12mer-Z6C_for	5'-CAACAGGATCCAAACTGTGCGAATACGG-3'
12mer-Z7C_for	5'-CAACAGGATCCAAACTGGCCTGCTACGG-3'
12mer-Z8C_for	5'-CAACAGGATCCAAACTGGCCGAATGCGG-3'
14mer-Z10C_for	5'-CAACAGGATCCAAATACCTGAACGCCGAATGCGG-3'
16mer-12C_1/2_for	5'-CAACAGGATCCAAATACCTGAACGCCGAATGCGG-3'
16mer-12C_2/2_for	5'-CAACAGGATCCCACAAATACCTGCGTAACGCCGA-3'
10mer-C4Z	5'-CCTCCCATATGGGCAGCGAAGCGGGCTGCAACATTGCGTAGTGC ATCACGGGAGATGCAC–3'
10mer-C6Z	5'-CCTCCCATATGGGCAGCGAATGCGGCACCAACATTGCGTAG TGCATCACGGGAGATGCAC–3'
10mer-C8Z	5'-CCTCCCATATGGGCTGCGAAGCGGGCACCAACATTGCGTAG TGCATCACGGGAGATGCAC-3'
Strep1-Z5C_for	5'-GGTTCCATATGGGCTAGCATCCGCAGTTCTGCGGCGATT GCATCACGGGAGATG-3'
Strep2-Z7C_for	5'-GGTTCCATATGGGCTAGCATCCGCAGGGTCCACCGTGCG GCGATTGCATCACGGGAGATG-3'
Strep3-11C_1/2_for	5'-CGTGCATCCGCAGTTCGCAAACTGCGATTGCATCACGGG AGATGC-3'
Strep3-11C_2/2_for	5'-GGTTCCATATGGGCTAGTTCACAAACGTGCATCCGCA GTTCGC-3'
SICLOPPS_for	5'-CAGGTCATATGGTTAAAGTTATCGGTCGTCGATCC -3'
SICLOPPS_rev	5'-CAACAGGTACCTTTAATTGTACCTGCGTCAAGTAATGGA AAG -3'
cStrep3(C)-Z3C_for	5'-CGCAGTTCGCGAACGCGTGCTTAAGTTTTGGCACCGAA ATT-3'
cStrep3(C)-Z3C_1/2_rev	5'-GATGGCAGTTGGTCTAGCAATTGTGGGCGATAGCACCATT AGC-3'
cStrep3(C)-Z3C_2/2_rev	5'-CGCGTTCGCGAACTGCGGATGGCAGTTGGTCTAGCAA TTG-3'
cStrep3(S)-Z3C_1/2_rev	5'-GGATGGCAGTTGGTCTAGCTATTGTGGGCGATAGCACCATT AGC-3'
cStrep3(S)-Z3C_2/2_rev	5'-CGCGTTCGCGAACTGCGGATGGCAGTTGGTCTAGCTATT GTG-3'
TyrRS_Y32G_for	5'-ATCTGCTGGTATAGGTTTTGAACCAAGTGG-3'
TyrRS_Y32G_rev	5'-CCACTTGGTTCAAAACCTATACCAGCAGAT-3'
YFP(stop)_for	5'-GGTTCCATATGGGTTAGGTGAGCAAGGGCGAGGAGC-3'
YFP_(XhoI)_rev	5'-CGTTGCTCGAGCTTGTACAGCTCGTCCATGCC-3'

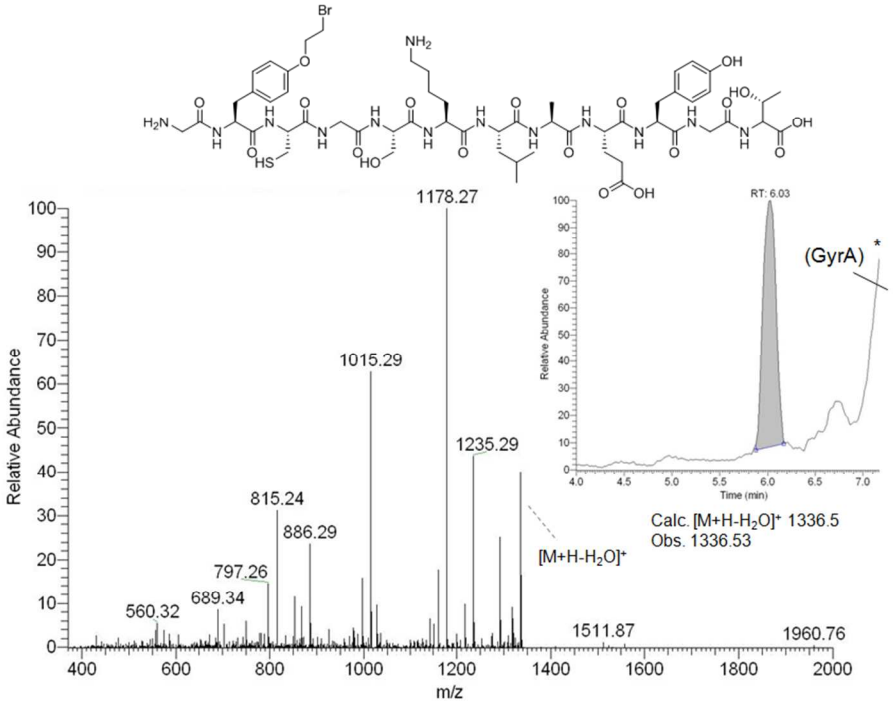
Supplementary Figure S1. Overview of the mechanism of split intein-mediated peptide circularization method (SICLOPPS).[1] Int_N and Int_C correspond to the N-domain and C-domain respectively, of *Synechocystis* sp. PCC6803 DnaE split intein.[2] The green segment corresponds to a variable peptide sequence. The Int_N+1 cysteine and Int_C+1 cysteine (or serine) are indicated.



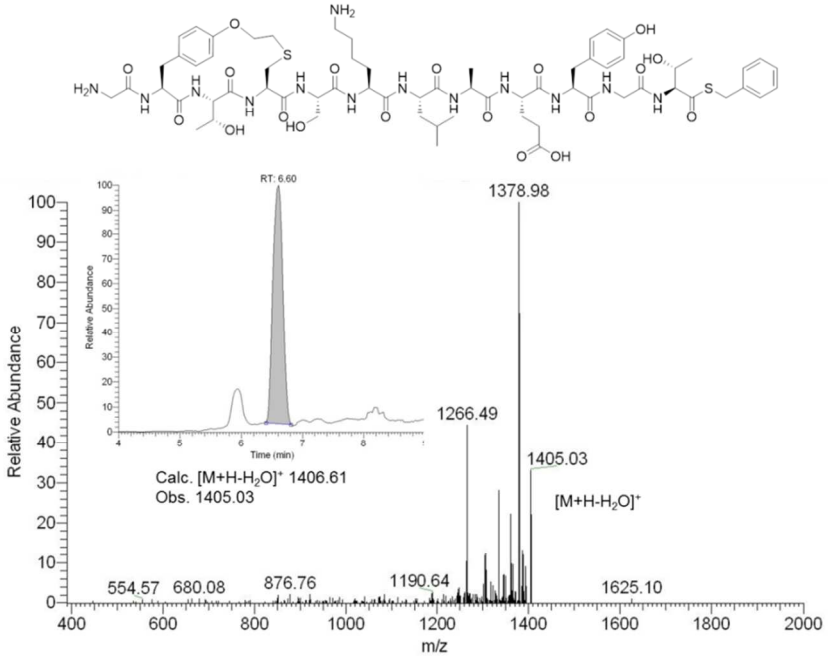
Supplementary Figure S2. Representative MS/MS spectrum corresponding to the macrocyclic peptide obtained from construct 12mer-Z6C (see Table 1 in main text). Assignment of the *a* and *b* fragments is indicated.



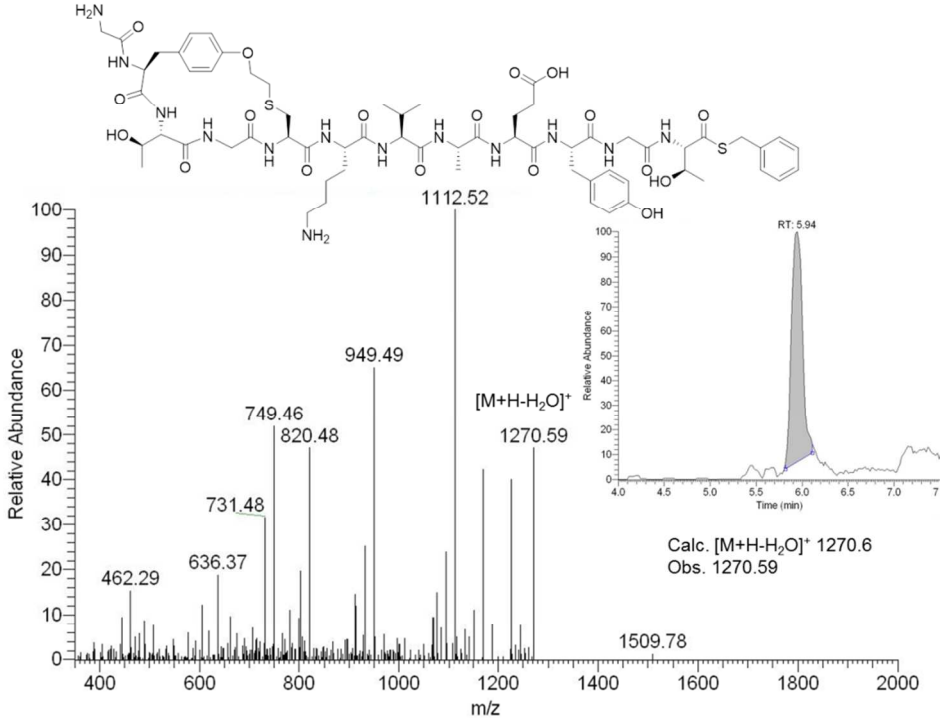
Supplementary Figure S3. Chemical structure, extracted-ion chromatogram, and MS/MS spectrum of the linear peptide obtained from construct 12mer-Z1C. * Unrelated multicharged ions from spliced GyrA intein.



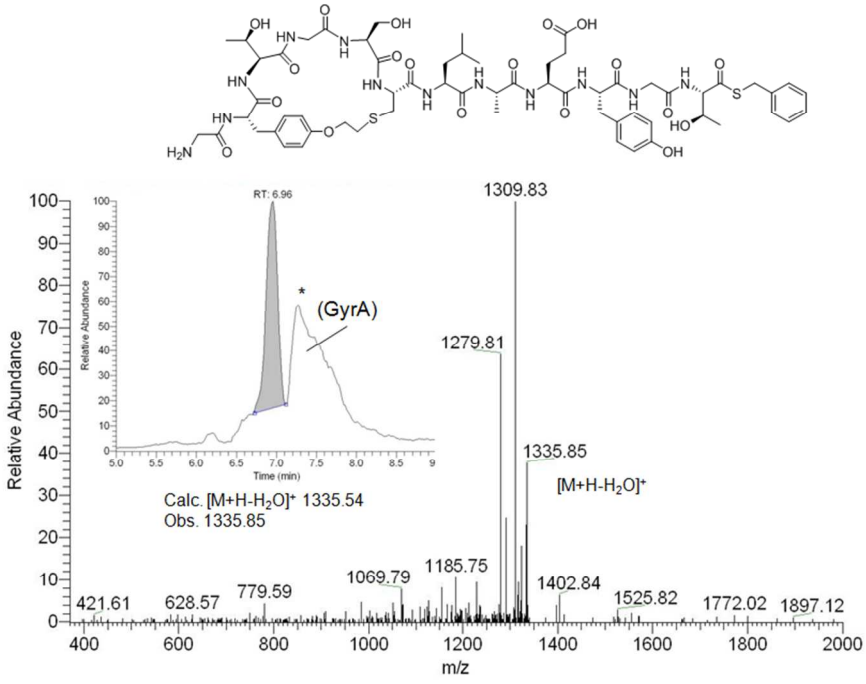
Supplementary Figure S4. Chemical structure, extracted-ion chromatogram, and MS/MS spectrum of the macrocyclic peptide obtained from construct 12mer-Z2C.



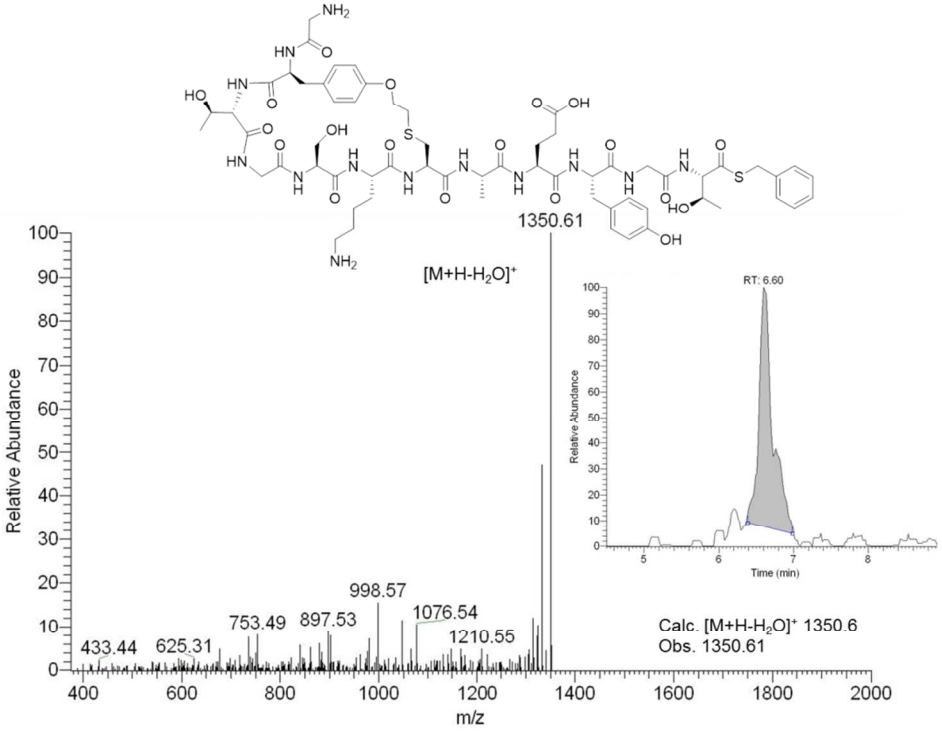
Supplementary Figure S5. Chemical structure, extracted-ion chromatogram, and MS/MS spectrum of the macrocyclic peptide obtained from construct 12mer-Z3C.



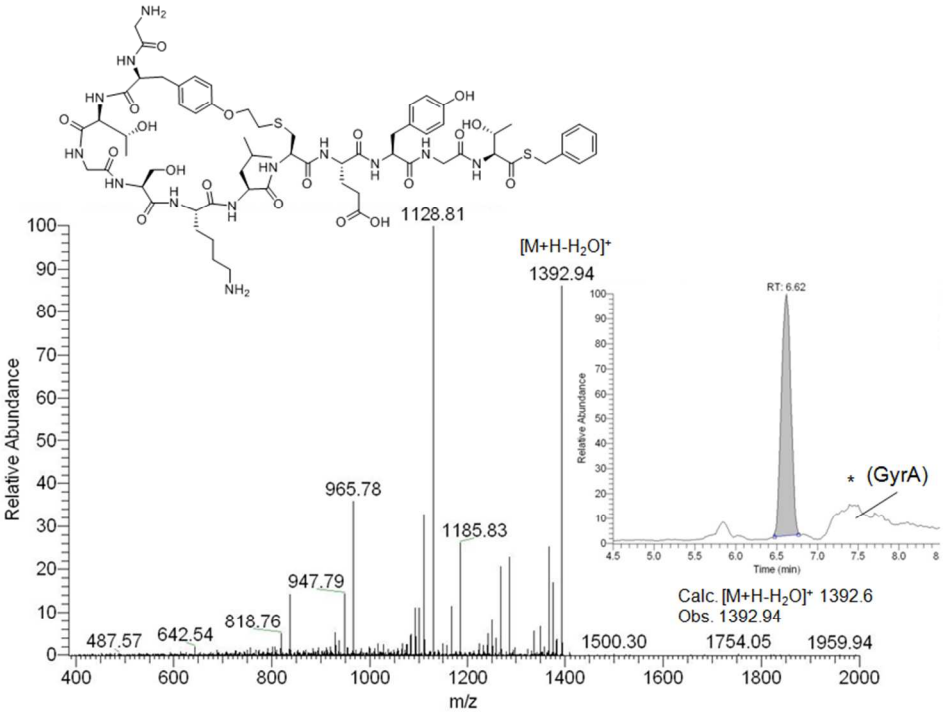
Supplementary Figure S6. Chemical structure, extracted-ion chromatogram, and MS/MS spectrum of the macrocyclic peptide obtained from construct 12mer-Z4C. * Unrelated multicharged ions from spliced GyrA intein.



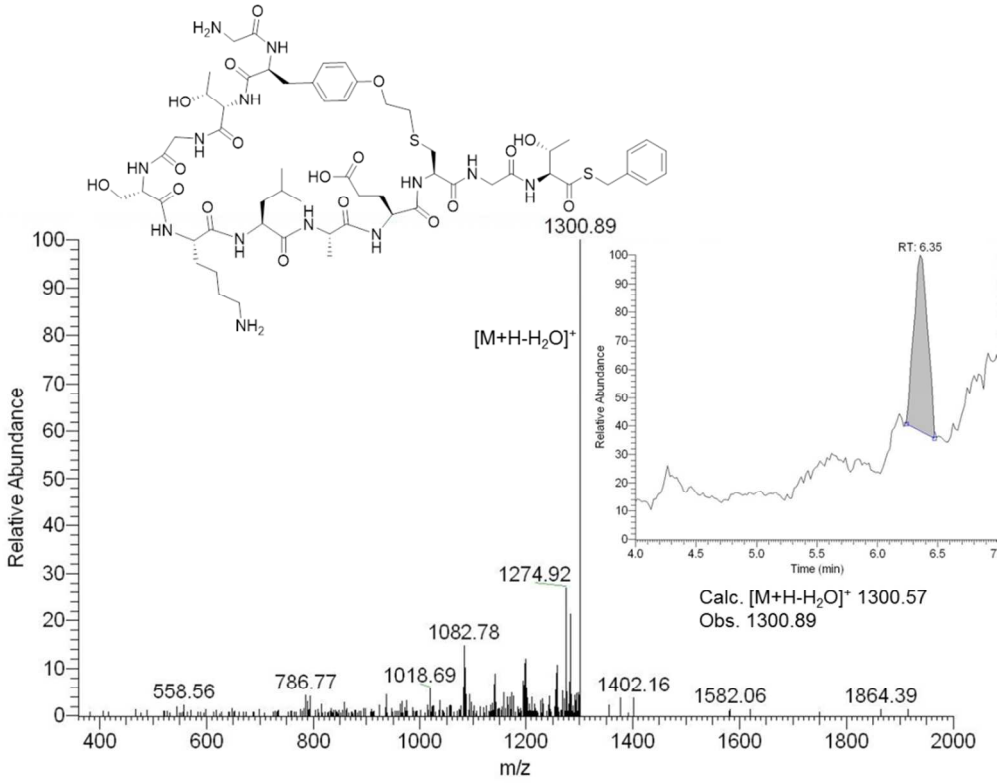
Supplementary Figure S7. Chemical structure, extracted-ion chromatogram, and MS/MS spectrum of the macrocyclic peptide obtained from construct 12mer-Z5C.



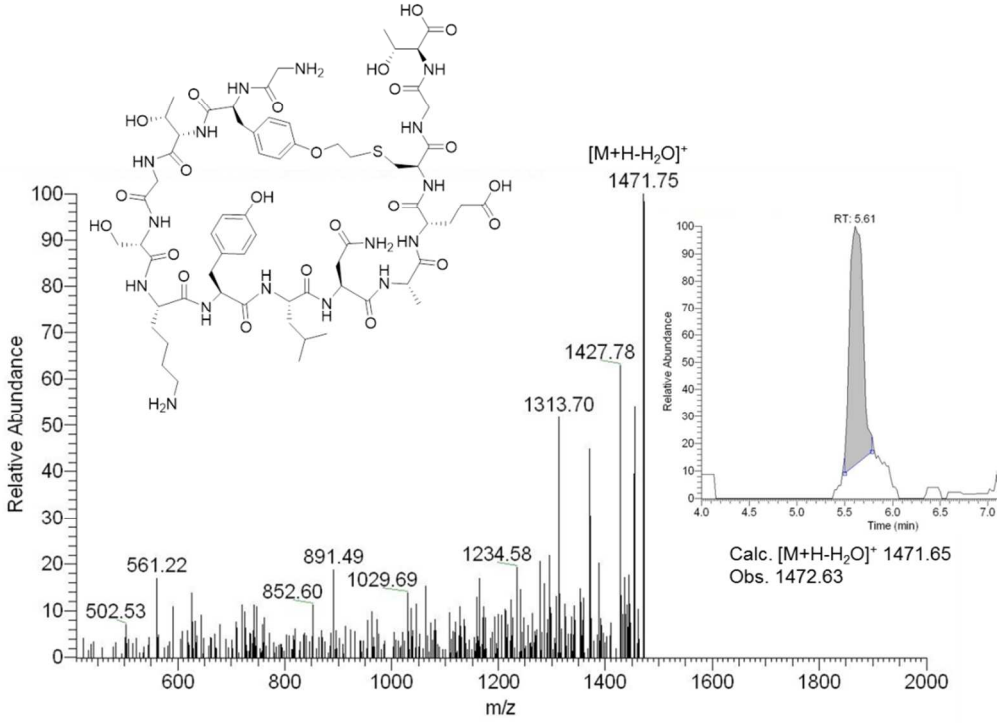
Supplementary Figure S8. Chemical structure, extracted-ion chromatogram, and MS/MS spectrum of the macrocyclic peptide obtained from construct 12mer-Z6C. * Unrelated multicharged ions from spliced GyrA intein.



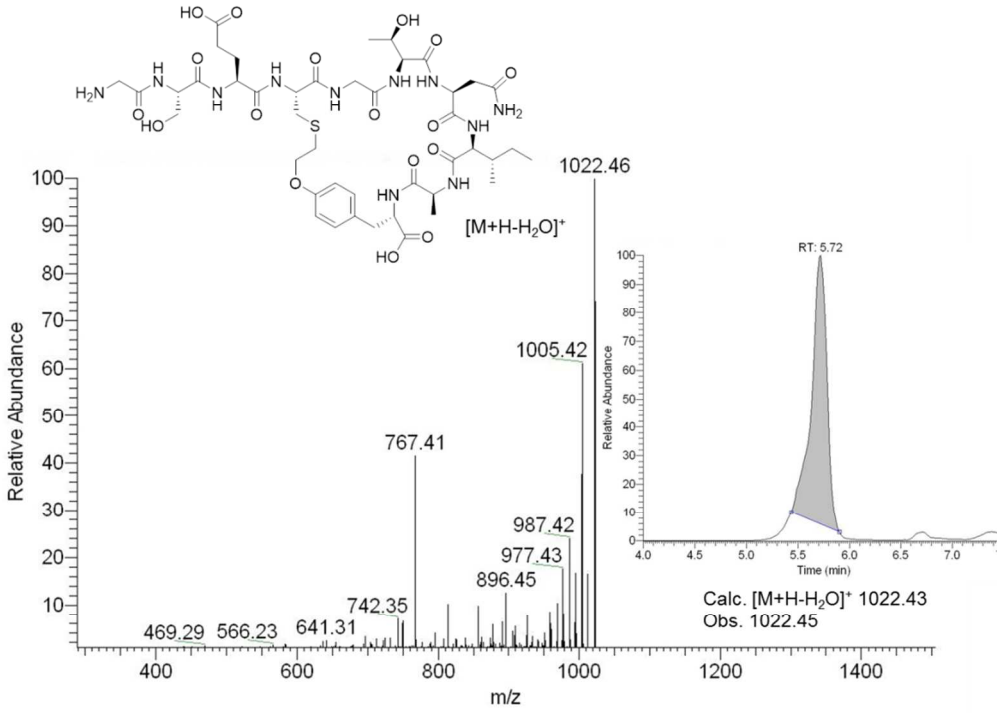
Supplementary Figure S9. Chemical structure, extracted-ion chromatogram, and MS/MS spectrum of the macrocyclic peptide obtained from construct 12mer-Z8C.



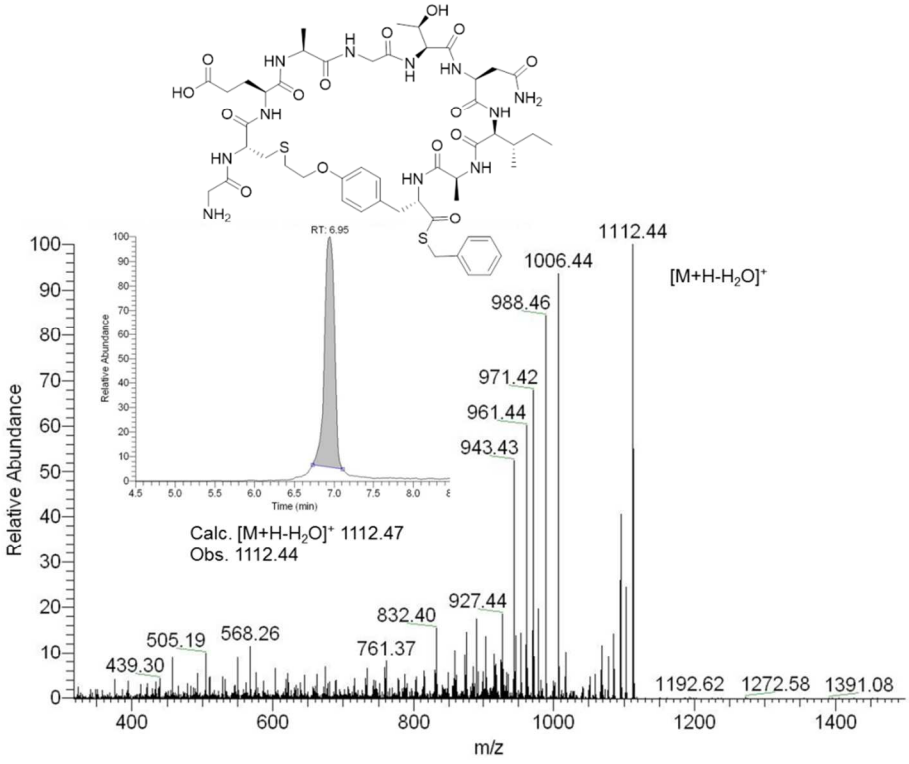
Supplementary Figure S10. Chemical structure, extracted-ion chromatogram, and MS/MS spectrum of the macrocyclic peptide obtained from construct 14mer-Z10C.



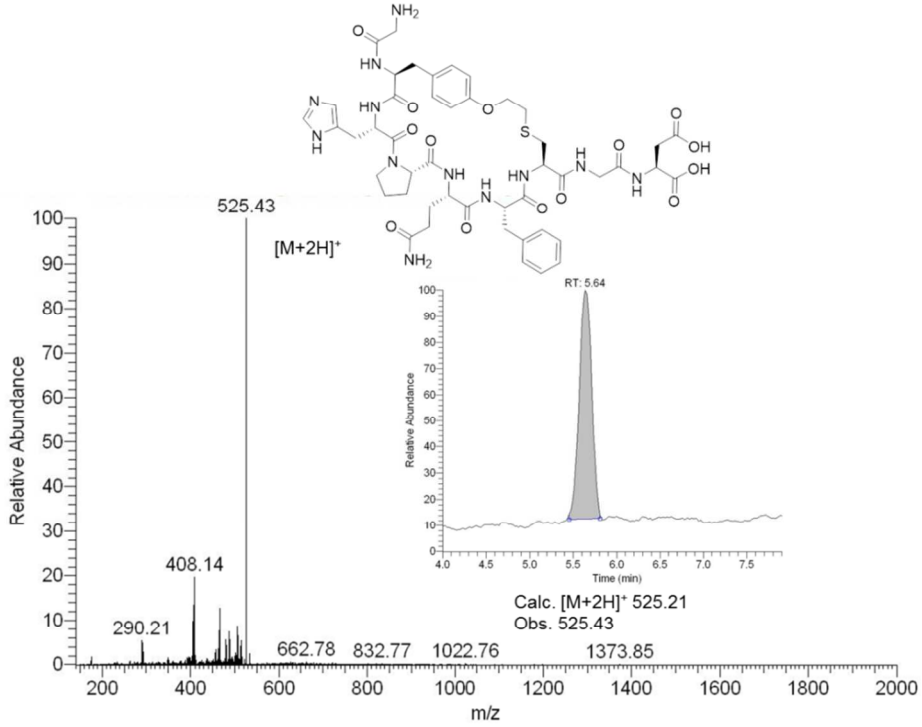
Supplementary Figure S11. Chemical structure, extracted-ion chromatogram, and MS/MS spectrum of the macrocyclic peptide obtained from construct 10mer-C6Z.



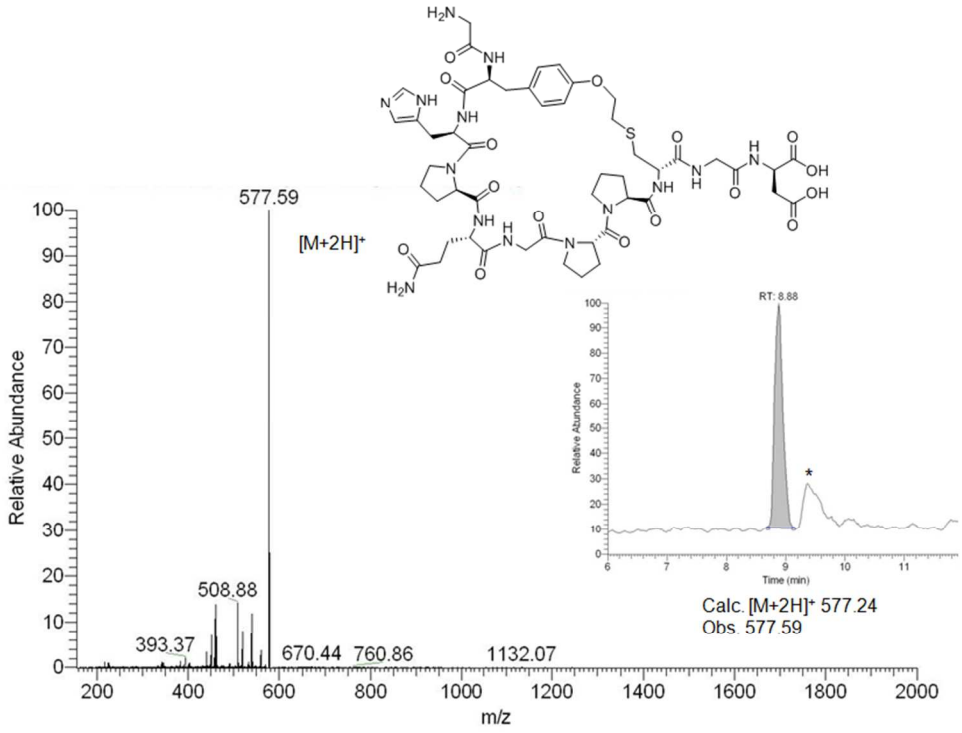
Supplementary Figure S12. Chemical structure, extracted-ion chromatogram, and MS/MS spectrum of the macrocyclic peptide obtained from construct 10mer-C8Z.



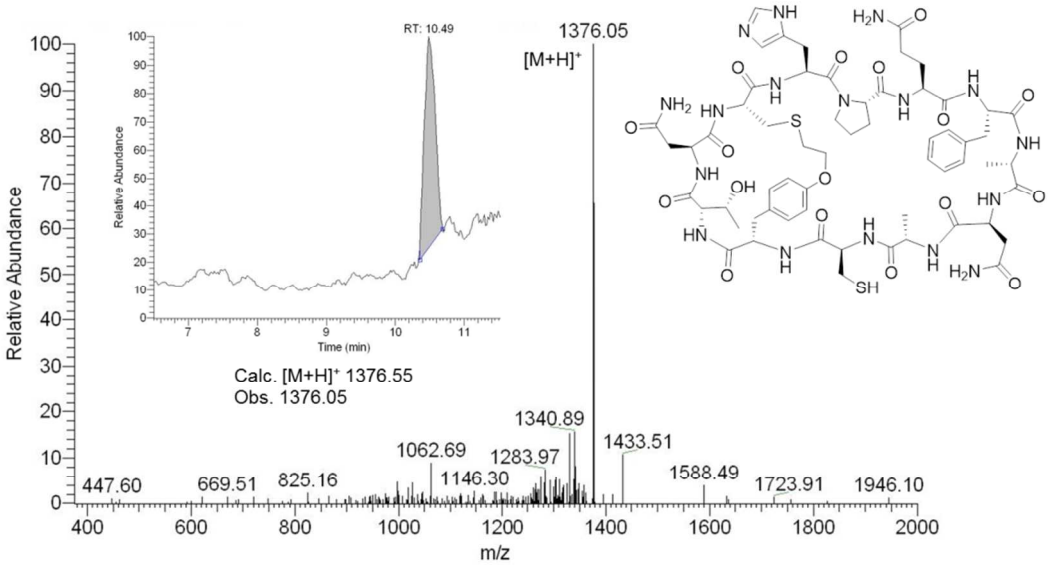
Supplementary Figure S13. Chemical structure, extracted-ion chromatogram, and MS/MS spectrum of the macrocyclic peptide obtained from construct Strep1-Z5C.



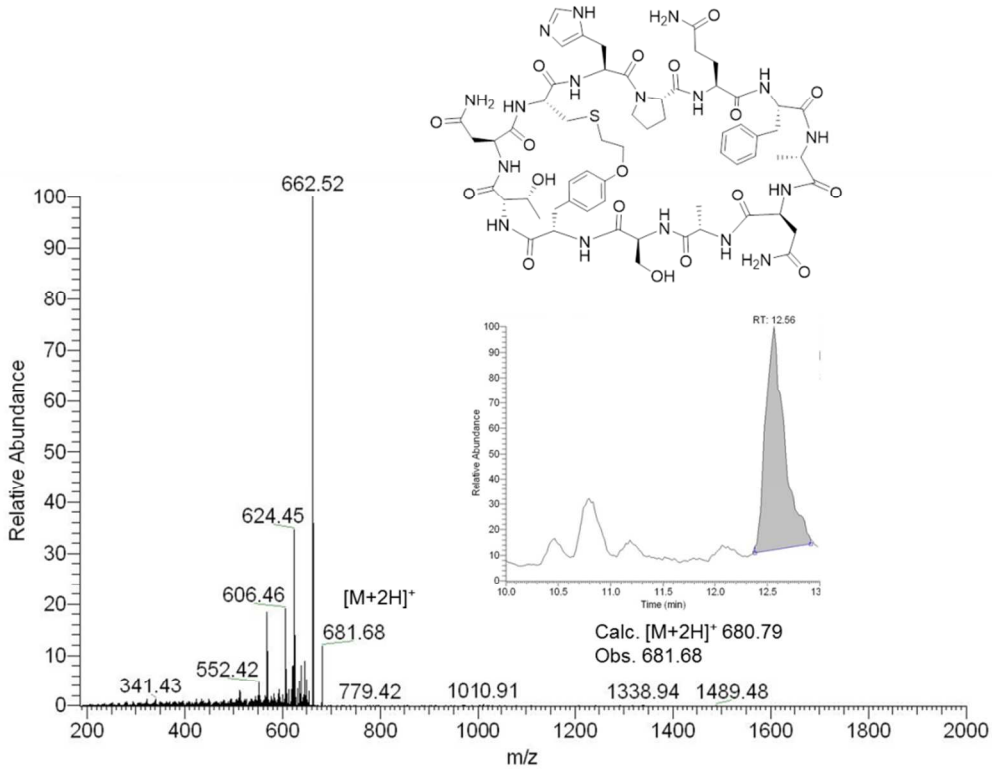
Supplementary Figure S14. Chemical structure, extracted-ion chromatogram, and MS/MS spectrum of the macrocyclic peptide obtained from construct Strep2-Z7C. * Unrelated multicharged ions from protein impurity.



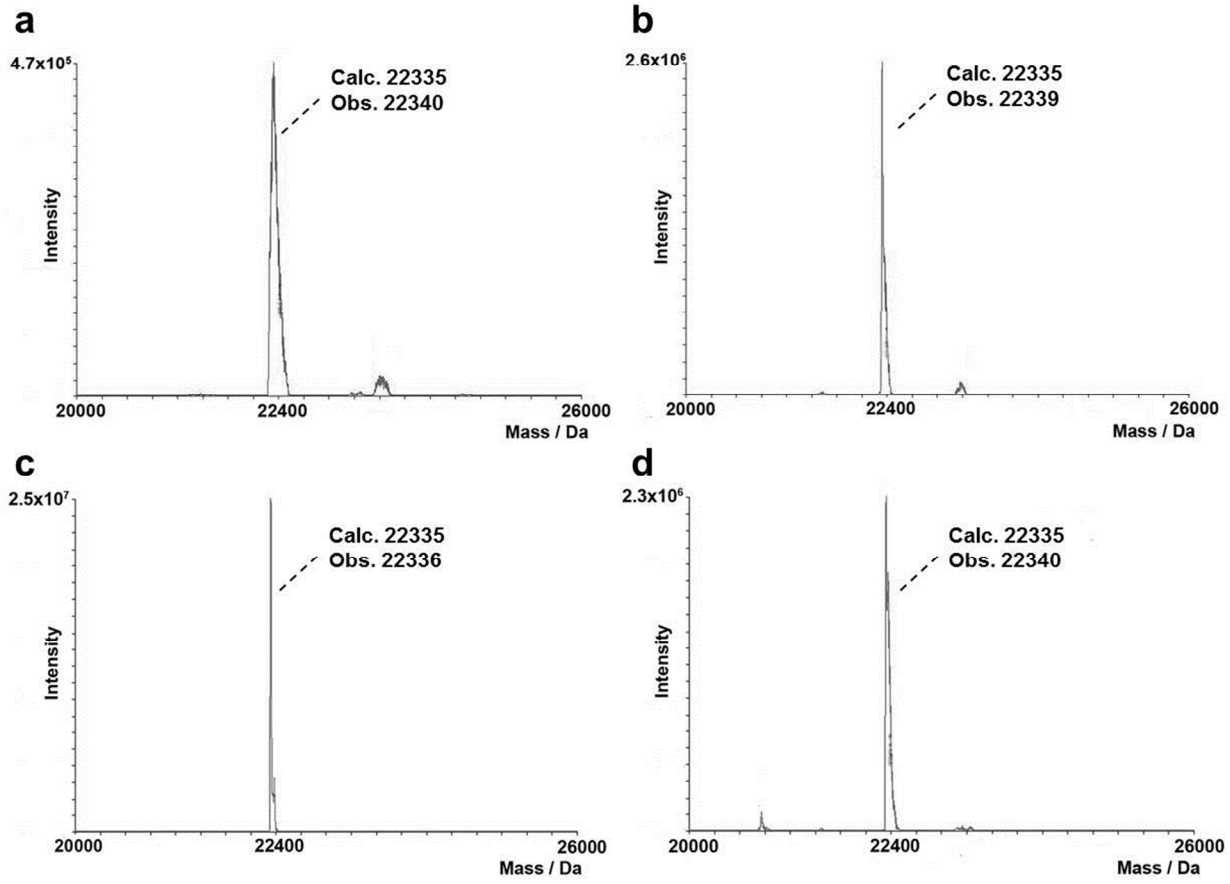
Supplementary Figure S15. Chemical structure, extracted-ion chromatogram, and MS/MS spectrum of the macrocyclic peptide obtained from construct cStrep3(C)-Z3C.



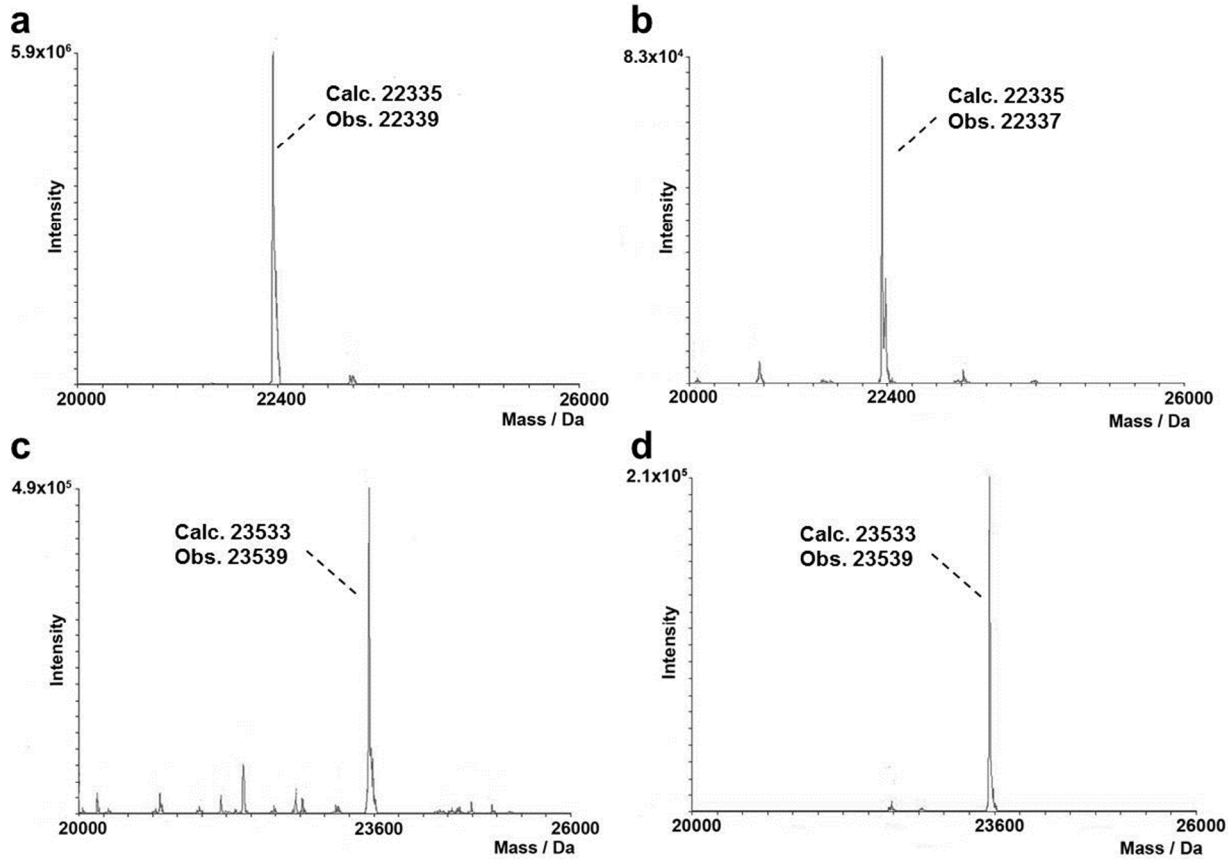
Supplementary Figure S16. Chemical structure, extracted-ion chromatogram, and MS/MS spectrum of the macrocyclic peptide obtained from construct cStrep3(S)-Z3C.



Supplementary Figure S17. Deconvoluted LC-MS mass spectra of proteins isolated after benzyl mercaptan-induced splicing of purified construct (a) 12mer-Z1C, (b) 12mer-Z4C, (c) 10mer-C6Z, and (d) 10mer-C8Z.



Supplementary Figure S18. Deconvoluted LC-MS mass spectra of proteins isolated from the cell lysate using Ni-NTA beads: (a) Strep1-Z5C construct, (b) Strep2-Z7C construct; and using chitin beads: (c) cStrep3(C)-Z3C construct, (d) cStrep3(S)-Z3C construct.



Experimental procedures

Reagents and analytical methods. Chemical reagents and solvents were purchased from Sigma-Aldrich, Acros Organics and Chem-Impex. Silica gel chromatography was carried out using AMD Silica Gel 60 (230-400 mesh). Streptavidin-binding agarose beads were purchased from Pierce and chitin beads were obtained from BioLabs. ^{1}H and ^{13}C NMR spectra were recorded on the Bruker Avance spectrometers (400 and 125 MHz, respectively) using solvent peaks as reference. The NMR data are reported as chemical shifts (δ ppm). LC-MS analyses were performed on Thermo Scientific LTQ Velos ESI/ion-trap mass spectrometer coupled to an Accela U-HPLC. Peptides and proteins were analyzed using Thermo Scientific HyPurity C4 column (particle size 5μm, 100 x 2.1 mm) and a linear gradient 5% to 95% ACN (with 0.1% formic acid) in water (with 0.1% formic acid) over 13 min. MALDI-TOF spectra were acquired on the Bruker Autoflex III spectrometer using α–cyano-4-hydroxycinnamic acid and sinapinic acid as matrix for peptides and proteins, respectively.

Cloning and plasmid construction. Oligonucleotides were purchased from Integrated DNA Technologies and their sequences are provided in Supplementary Table S1. Genes encoding 12mer-Z1C, 12mer-Z2C, and 12mer-Z3C were prepared by PCR using forward primers 12mer-Z1C_for, 12mer-Z2C_for, and 12mer-Z3C_for, respectively, T7_term as reverse primer. For these constructs, the template was pMG-G8T,[3] a pET22b(+) (Novagen) derived plasmid encoding for MG(amber stop)TGSAEYGT peptide, followed by *Mxe* GyrA(N198A) and a His tag. PCR products (0.75 Kbp) were digested with *Nde* I and *Xho* I and cloned into pMG-G8T plasmid to produce p12mer-Z1C, p12mer-Z2C, and p12mer-Z3C. Genes encoding 12mer-Z4C, 12mer-Z5C, 12mer-Z6C, 12mer-Z8C were prepared by PCR using forward primers 12mer-Z4C_for, 12mer-Z5C_for, 12mer-Z6C_for, 12mer-Z8C_for, respectively, T7_term as reverse primer, and template pMG-G8T. The PCR products (0.75 Kbp) were digested with *BamH* I and *Xho* I and cloned into the pMG-G8T vector to produce p12mer-Z4C, p12mer-Z5C, p12mer-Z6C, p12mer-Z8C. The gene encoding 14mer-Z10C was prepared by PCR using p12mer-Z8C as template, MG10_for as the forward primer, and T7_term as the reverse primer. The PCR product (0.6 Kbp) was digested with *BamH* I and *Xho* I and cloned into pMG-G8T to provide p14mer-Z10C vector. The gene encoding 16merZ12C was prepared through consecutive PCR amplifications using 16merZ12C_1/2_for and 16merZ12C_2/2_for as forward primers,

respectively, and T7_term as the reverse primer. The PCR product (0.6 Kbp) was digested with *BamH* I and *Xho* I into pMG-G8T to provide p16mer-Z12C. Genes encoding 10mer-C4Z, 10mer-C6Z, 10mer-C8Z were prepared by PCR using pMG-G8T as template, 10mer-C4Z_for, 10mer-C6Z_for, 10mer-C8Z_for, respectively, as the forward primers, and T7_term as the reverse primer. The PCR products (0.6 Kbp) were digested with *Nde* I and *Xho* I into pMG-G8T to provide p10mer-C4Z, p10mer-C6Z, and p10mer-C8Z, respectively.

Genes encoding for Strep1-Z5C and Strep2-Z7C were prepared by PCR using pMG-G8T as template, Strep1-Z5C_for and Strep2-Z7C_for, respectively, as forward primers, and T7_term as reverse primer. The PCR products (0.6 Kbp) were digested with *Nde* I and *Xho* I into pMG-G8T to provide pStrep1-Z5C and pStrep2-Z7C, respectively. In all vectors the genes encoding for the biosynthetic precursor are under the control of an IPTG-inducible T7 promoter. The genes encoding the *N*-terminal and *C*-terminal DnaE inteins from *Synechocystis sp*. were extracted from pSFBAD09 and pJJDuet30 plasmids[4] (Addgene #11963 and #11962), respectively. DnaE-Int_N was amplified using primers SICLOPPS_for as forward primers and cStrep3(C)-Z3C_1/2_rev and cStrep3(S)-Z3C, respectively, as reverse primers. The PCR products (0.6 Kbp) were used as templates for a second PCR reaction using forward primer SICLOPPS_for and cStrep3(C)-Z3C_2/2_rev and cStrep3(S)-Z3C_2/2_rev as reverse primers, respectively. The DnaE-Int_C was amplified using primers cStrep-Z3C_for as forward primer and SICLOPPS_rev. The fragments were fused to generate a gene encoding the DnaE-Int_N fused to the DnaE-Int_C through the desired target sequence. The resulting products (0.5 Kbp) were digested with *Nde* I and *Kpn* I and cloned into pSFBAD09 to provide the plasmids p_cStrep3(C)-Z3C and p_cStrep3(S)-Z3C. Ligation at the *Kpn* I site also introduces the chitin-binding domain (CBD). In all vectors excluding p_cStrep3(C)-Z3C and p_cStrep3(S)-Z3C, the genes encoding for the biosynthetic precursor are under the control of an IPTG-inducible T7 promoter. In p_cStrep3(C)-Z3C and p_cStrep3(S)-Z3C, the genes encoding for the SICLOPPS construct are under the control of an arabinose-inducible AraC promoter.

The vector encoding for the O2beY-RS synthetase was prepared by Quick Change mutagenesis using plasmid pEVOL_OpgY (*vide infra*) as template and oligonucleotides TyrRS_Y32G_for and TyrRS_Y32G_rev as the mutagenizing primers. Plasmid vectors for the expression of Yellow Fluorescent Protein were prepared as follows. The YFP encoding gene was PCR amplified from pEYFP-N1 (BDbiosciences) using primers YFP(stop)_for and

YFP_(*Xho*I)_rev and the PCR product (0.7 Kbp) was cloned into pET22b using *Nde*I and *Xho*I, giving pET22_YFP(stop) plasmid. This process introduced an amber stop codon (TAG) in the second position of the protein sequence and fused a His tag to the C-terminus of the YFP gene. Using a similar procedure, a pET22b-based vector for the expression of His-tagged YFP but lacking the amber stop codon was prepared (called pET22_YFP). The sequences of all the plasmid constructs described above were confirmed by DNA sequencing at the Functional Genomics Center of the University of Rochester.

Aminoacyl-tRNA synthetase screening and characterization. The plasmid pEVOL_OpgY encoding for an engineered $tRNA_{CUA}$ (*Mj*$tRNA_{CUA}$) and aminoacyl-tRNA synthetase (*Mj*Tyr-RS) from *Methanococcus jannaschii* for amber codon suppression with *O*-propargyl-tyrosine (OpgY)[5] was kindly provided by the Schultz group. Compared to wild-type *Mj*Tyr-RS, OpgY-RS contains the amino acid mutations Y32A, E107P, D158A, L162A. As described above, a pEVOL-based vector for the expression of O2beY-RS (pEVOL_O2beY) was prepared by introducing a A32G mutation into each of the synthetase gene. Incorporation of the unnatural amino acid O2beY using OpgY-RS and O2beY-RS synthetases was evaluated as follows. *E. coli* BL21(DE3) cells were co-transformed with the pET22_YFP(stop) plasmid and the pEVOL plasmid encoding for the appropriate aminoacyl-tRNA synthetase, and then grown in Luria-Bertani (LB) media containing ampicillin (50 mg/L) and chloramphenicol (26 mg/L) at 37°C overnight. The overnight cultures were used to inoculate (50 μL) 96-deep well plates containing M9 media (840 μL / well) supplemented with the same concentrations of antibiotics. At an OD_{600} of about 0.6, the appropriate unnatural amino acid (OpgY or O2beY) at the final concentration of 2 mM, arabinose (0.06% m/v), and IPTG (0.2 mM) were added. After overnight (12 hours) growth at 27°C, the cell cultures were diluted (1:1) with 100 μL of potassium phosphate buffer (50 mM, 150 mM NaCl, pH 7.5) into black, flat-bottomed 96-well plates. To measure background incorporation by the AARS, cell cultures containing no unnatural amino acid were included as a control. Fluorescence intensity was determined using a λ_{ex} of 514 nm and λ_{em} of 527 nm. Each sample was measured in quadruplicate. The results were normalized to the final cell density based on OD_{600}.

Protein expression and purification. Chemically competent BL21(DE3) *E. coli* cells were co-transformed with pEVOL_O2beY and the appropriate pET22b-based vector encoding for the

desired precursor polypeptide. Overnight cultures were grown in Luria-Bertani (LB) medium supplemented with 50 mg/L of ampicillin and 26 mg/L of chloramphenicol and used to inoculate 0.2 L of M9 media containing the same concentration of antibiotics and supplemented with 1% glycerol. Bacterial cultures were grown at 37°C until OD_{600} reached 0.6, at which point O2beY (2 mM), L-arabinose (0.06% m/v) and IPTG (0.2 mM) were added to induce protein expression. Cultures were grown for additional 12 hours at 27°C. Cells were harvested by centrifugation at 4,000 rpm for 25 min, resuspended in 50 mM Tris, 300 mM NaCl, 20 mM imidazole buffer (pH 7.5) and lysed by sonication. Upon centrifugation at 13,000 rpm for 30 min, cell lysate was loaded on a Ni-NTA affinity column and the protein of interest was eluted with 50 mM Tris, 150 mM NaCl, 300 mM imidazole buffer (pH 7.5). Fractions containing the protein were combined and concentrated followed by buffer exchange with potassium phosphate buffer (50 mM, 150 mM NaCl, pH 7.5). Protein concentration was determined using the extinction coefficient at 280 nm (ε_{280}) calculated based on the protein sequence and aliquots of the protein solutions were frozen and stored at -80°C. Typical expression yields for the O2beY-containing proteins were 40-50 mg/L of culture. The identity of the isolated proteins was confirmed using MALDI-TOF MS and LC-MS.

Amber stop codon suppression efficiency. Chemically competent BL21(DE3) *E. coli* cells were transformed with pET22_YFP alone or, alternatively, co-transformed with pEVOL_O2beY plasmid and pET22_YFP(stop). Protein expression was performed in parallel as described above with the exception in the case of wild-type YFP only ampicillin (50 mg/L) and no unnatural amino acid was added to the culture medium. The proteins were purified using Ni-NTA chromatography. Under these conditions, the calculated expression yield for wild-type YFP was 31 mg/L while that for YFP(O2beY) was 26.4 mg/L, indicating that the efficiency of amber stop codon suppression by means of the O2beY-RS synthetase and cognate *Mj* tRNA is about 85%.

Protein splicing experiments and peptide analysis. The protein constructs corresponding to Entries 1-11 of Table 1 were expressed and isolated by Ni-affinity chromatography as described above. The purified proteins (200 μM, 20 μL final volume) were then incubated with 15 mM benzyl mercaptan and 20 mM TCEP in phosphate buffer (50 mM, pH 8). After 2-4 hours, the reaction mixtures were analyzed directly by LC-MS under the conditions specified above (Reagents and Analytical Methods section). For measuring the relative amount of the desired macrocyclic product and that of the linear (uncyclized) peptide, both the thioester (peptide-(CO)-

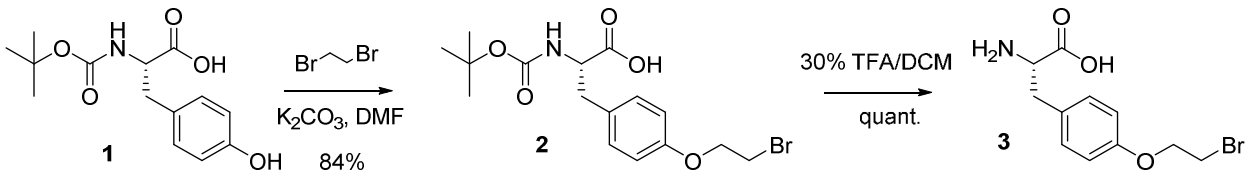
SBn) and free carboxy form (peptide-COOH) of these peptides were searched and taken into consideration for determining the extent of cyclization. For the linear peptide, the LC-MS method also included the search for potential adducts resulting from nucleophilic displacement of O2beY side-chain 2-bromoethyl group with BnSH, cysteine, glutathione, or water, in both thioester and free carboxy form. However, none of the latter adducts was observed under the applied experimental conditions.

Isolation of HPQ-containing cyclic peptides via streptavidin affinity. Protein expression of constructs 12-13 was performed as described above. After centrifugation, cells were resuspended in 50 mM Tris, 300 mM NaCl and 20 mM imidazole (pH 7.5) and lysed by sonication. Cell lysates were incubated with streptavidin-coated beads for 1 hour under gentle shaking on ice. Beads were washed two times with the same buffer followed by incubation with acetonitrile:H_2O (70:30 v/v) for one minute to release any streptavidin-bound peptides. Eluates were lyophilized and the identity of the peptides determined by LC-MS analysis.

Preparation and isolation of bicyclic macrocycles. Protein expression of constructs 14-15 was performed as described above with the difference that cells were incubated for additional 3 hours at 37°C after overnight growth. Cells were harvested, lysed and the cell lysate treated as described above to isolate and analyze the streptavidin-bound peptides by LC-MS. To analyze the amount of protein splicing occurred *in vivo*, the same cell lysate samples were incubated with chitin beads for 1h on ice. Beads were washed two times with buffer followed by incubation with acetonitrile:H_2O (70:30 v/v) for one minute to release any chitin-bound protein. Eluates were analyzed by LC-MS.

Synthetic procedures

Scheme S1. Synthesis of *O*-(2-bromoethyl)-tyrosine (O2beY)



***O*-(2-bromoethyl)-tyrosine (3).** To a reaction flask containing *N-tert*-butoxycarbonyl-tyrosine **1** (2 g, 7.1 mmol) and potassium carbonate (2.94 g, 21.3 mmol) in dry DMF (20 mL) dibromoethane (1.83 mL, 21.3 mmol) was added dropwise over 20 min. The reaction mixture was stirred at room temperature for 18 h after which the reaction mixture was filtered, diluted with 60 mL of water, acidified with acetic acid to pH 4 and extracted with 2 x 100 mL of EtOAc. Organic layers were combined and dried over sodium sulfate. The solvent was removed under reduced pressure yielding yellow oil as crude product which was purified by flash column chromatography using 10:9:1 hexane:EtOAc:HOAc acid as solvent system. Fractions of interest were combined and solvent removed under reduced pressure yielding off-white powder as product **2** (2.3 g, 84%). ^{1}H NMR (400 MHz, CD_3OD) δ 1.39 (s, 9H), 2.8-3.05 (m, 2H), 3.3 (t, 2H), 3.51 (t, 2H), 4.37 (t, 2H), 6.69 (d, 2H), 7.02 (d, 2H); ^{13}C NMR (125 MHz, CD_3OD) δ 28.73, 29.49, 37.92, 56.82, 65.77, 80.69, 116.27, 128.84, 131.32, 157.39, 157.77, 173.414. MS (ESI) calculated for $C_{14}H_{19}NO_5$ $[M]^+$: *m/z* 387.07, found 387.17. Purified **2** was treated with 20 mL of 30% TFA/DCM to remove the *N*-terminal protection. Upon completed reaction (determined by TLC), the solvent was removed under reduced pressure, crude residue dissolved 2 x in 10 mL of HOAc followed by solvent evaporation yielding the final product **3** as off-white solid in quantitative yield (1.7 g). ^{1}H NMR (400 MHz, CD_3OD) δ 3.05-3.25 (m, 2H), 3.58 (t, 2H), 4.28 (t, 1H), 4.51 (t, 2H), 6.77 (d, 2H), 7.09 (d, 2H); ^{13}C NMR (125 MHz, CD_3OD) δ 29.1, 36.9, 55.35, 66.92, 116.92, 125.54, 131.59, 158.41, 169.93. MS (ESI) calculated for $C_{11}H_{14}BrNO_3$ $[M+H]^+$: *m/z* 288.02, found 288.51.

References

[1] C. P. Scott, E. Abel-Santos, M. Wall, D. C. Wahnon, S. J. Benkovic, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13638-13643.

[2] H. Wu, Z. Hu, X. Q. Liu, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9226-9231.

[3] J. M. Smith, F. Vitali, S. A. Archer, R. Fasan, *Angew. Chem. Int. Ed.* **2011**, *50*, 5075-5080.

[4] S. Zuger, H. Iwai, *Nat Biotechnol* **2005**, *23*, 736-740.

[5] A. Deiters, P. G. Schultz, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1521-1524.