Electronic Supplementary Material (ESI) for Chemical Communications

Synthesis of cyclic peptide/protein using the NEXT-A reaction followed by cyclization

Toshimasa Hamamoto, Masahiko Sisido, Takashi Ohtsuki, and Masumi Taki*

Supporting information

Materials and Methods

Contents:

1. General

Sinapinic acid (for protein measurements), α-Cyano-4-hydroxycinnamic acid (for peptides), and 2,5-dihydroxybenzoic acid (for other organic compounds) were used as matrices for the TOF-MS analysis. Before TOF-MS analysis, proteins and peptides were desalted, concentrated, and purified using ZipTipC_{18} silica resin (Millipore, Bedford, MA) according to the instruction manual. NMR spectra were recorded on a Varian Mercury 300 Spectrometer excepted tandem mass-mass analysis.

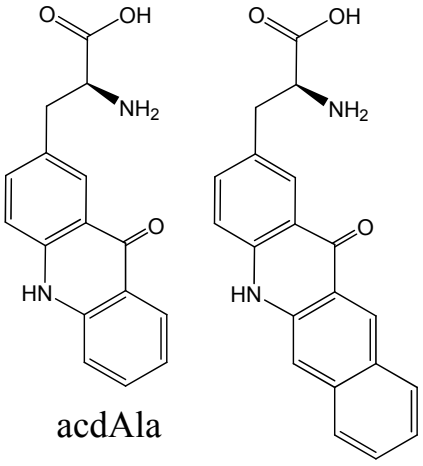
L/F-transferase (wild type) and doubly-mutated *E. coli* PheRS (αA294G&αT251A) were overexpressed and purified according to the reported procedure[1]; the PheRS mutant is known to induce misacylation with various non-natural amino acids. N-terminal sequence analysis on a protein sequencer and MALDI-TOF-MS analysis were performed as reported previously.[2]

E. coli $tRNA^{Phe}$ was obtained from Sigma; use of pure *E. coli* $tRNA^{Phe}$ (Sigma #R3143), instead of a crude tRNA mixture (Sigma #R1753 or R4251), is important to obtain consistent results.[1] Instead of using commercially available tRNA, overexpression and purification of the tRNA is an alternative approach for a cost-effective cyclization via the NEXT-A reaction. First, tRNA was overexpressed in *E. coli* (MRE600rif), and treated with conc. NaCl and phenol/chloroform according to the reported procedure.[1] Second, it was further purified with an anion exchange column.[3] Briefly, the tRNA was applied to a DE52 (Whatmann) column, washed with 10 mM Tris-HCl (pH 7.5) / 1 mM EDTA / 100mM NaCl buffer, then eluted with 10 mM Tris-HCl (pH 7.5) / 1 mM EDTA / 1M NaCl buffer. It was treated with phenol/chloroform solution, and precipitated with ethanol to obtain tRNA with

sufficient purity, which gave the same NEXT-A result obtained by the commercially available tRNA.

The expression plasmid (pH10UE), and deubiquitylating enzyme (Usp2-cc) encoding plasmid were kind gifts from Prof. A. Varshavsky and Dr. K. Piatkov, and Usp2-cc was expressed and purified as in the reported method.[4] Plasmid DNA encoding GFP (pGGFPH) was a kind gift from Prof. Takahiro Hohsaka.

Model peptide (UTIF; Urinary Tripsin Inhibitor Fragment) was purchased from Bachem (#H-2692). Fmoc-protected fluorescent amino acids, acrydonylalanine (acdAla) and benzoacrydonylalanine (badAla), were synthesized according to the reported procedures,[5, 6] and are currently available from Watanabe Chemical Industries (Hiroshima, Japan). Each chemical structure is shown below.



acdAla

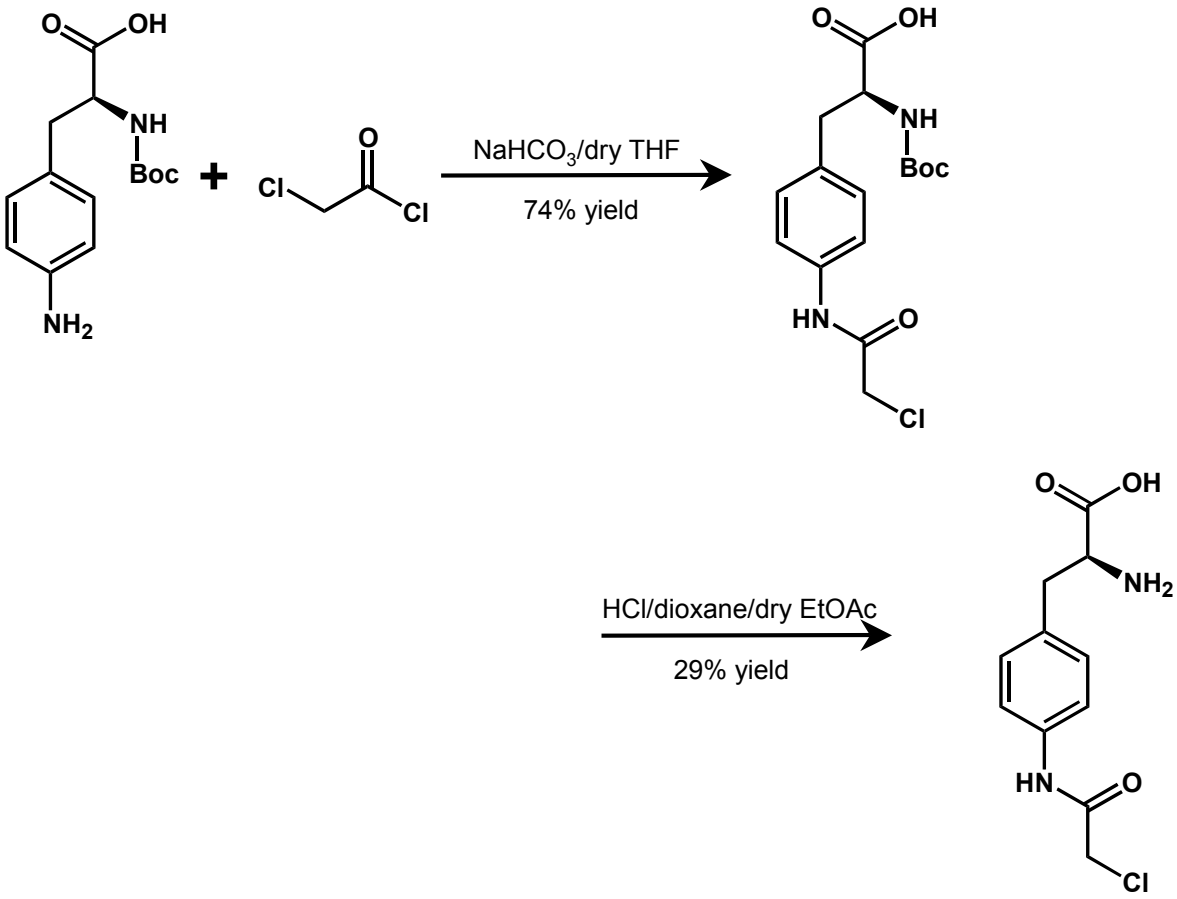
badAla

Other solvents and reagents are commercially available and used without further purifications.

All of the biochemical experiments were repeated at least twice or up to five times to confirm the reproducibility.

2. Synthesis of non-natural amino acids

2-1. *p*-chloroacetylamino-*L*-phenylalanine (caaPhe; 1)



2-1-1. Boc-caaPhe-OH

To a chilled suspension of Boc-aminophenylalanine (1.1 g, 4.0 mmol; Watanabe Chemicals) and $NaHCO_3$ (1.7 g, 20 mmol) in dry THF (30 mL), chloroacetyl chloride (2.0 mL, 25 mmol) in dry THF (10 mL) was added dropwise for 15 minutes under N_2 atmosphere. The mixture was stirred overnight initially at 0 ^{o}C and then at room temperature in the dark. The mixture was mixed with 4% aqueous $NaHCO_3$ and extracted with ethyl acetate. The extract was dried over magnesium sulfate, and concentrated under reduced pressure, to afford a white solid (1.1 g; 74% yield) which was used for the next reaction without further purification.

Boc-caaPhe-OH: TLC R_f = 0.68 (chloroform/methanol = 7:3); ^{1}H NMR (300 MHz, DMSO-d_6) spectrum is shown below.

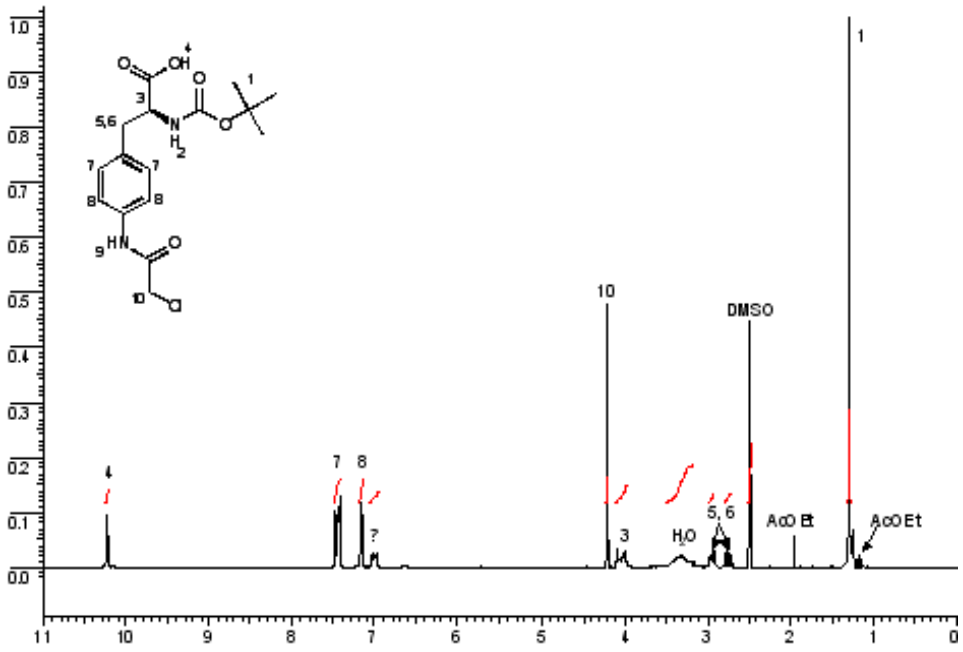


Figure S1. ^{1}H NMR spectrum of Boc-caaPhe-OH.

2-1-2. H-caaPhe-OH

To a chilled solution of Boc-caaPhe-OH (0.20 g, 0.56 mmol) in dry ethyl acetate (2.5 mL), 4 M HCl in dry dioxane (2.5 mL; Watanabe Chemicals) was added dropwise under N_2 atmosphere. The mixture was stirred for 2 hr at 0 ^{o}C and then the solvent was evaporated. The obtained white solid was resuspended in pure water, and the supernatant was lyophilized to afford a white powder (48 mg; 29% yield) which was used for the NEXT-A/Cyclization reaction without further purification.

1: TLC R_f= 0.43 (chloroform/methanol = 8:2); MALDI-TOF MS (m/z) found: 257.37 $[M+H]^+$; calcd. for $[M+H]^+$: 257.07; ^{1}H NMR (300 MHz, D_2O) spectrum is shown below.

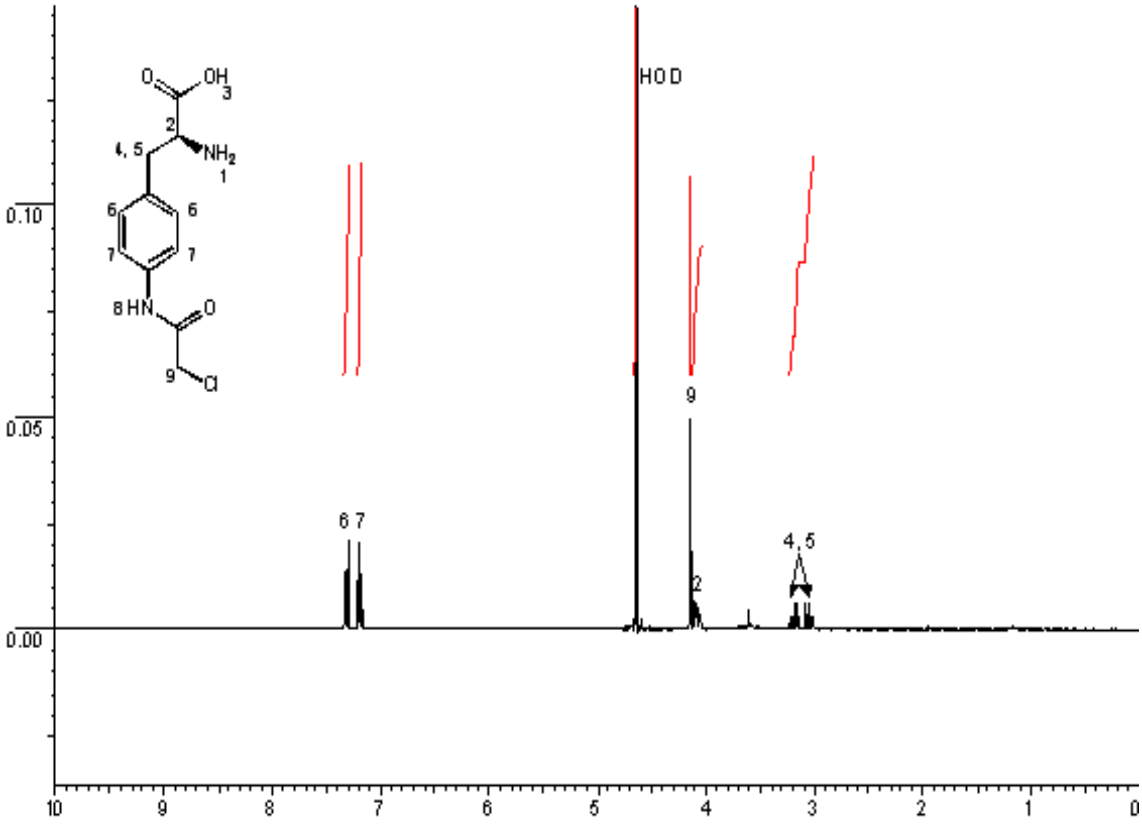
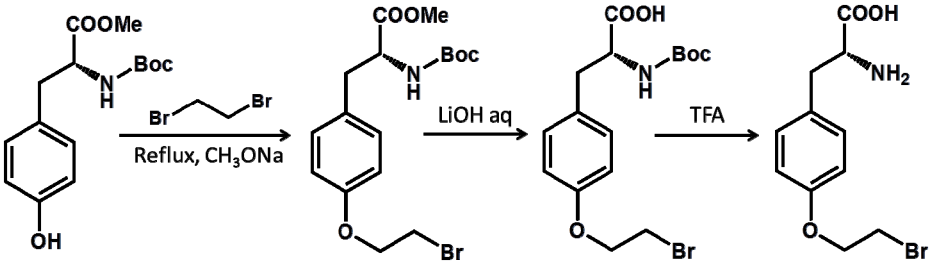


Figure S2. ^{1}H NMR spectrum of *p*-chloroacetylamino-*L*-phenylalanine (caaPhe; **1**).

2-2. *O*-bromoetyl-*L*-tyrosine (beTyr; 2)



2-2-1. Boc-beTyr-OMe

Dibromomethane (0.30 mL, 3.5 mmol), Boc-Tyr-OMe (1.0 g, 3.5 mmol, Watanabe Chemicals), and sodium methoxide (0.19 g, 3.5 mmol) were mixed in dry

acetonitrile (6 mL), refluxed for 12 hr, and brought to room temperature. The solvent was evaporated. The crude reaction product was mixed with saturated aqueous solution of NH_4Cl, and extracted with ethyl acetate. The extract was dried over magnesium sulfate, and concentrated under reduced pressure to afford amber-colored oil. Purification by silica-gel column chromatography (hexane/AcOEt = 3:1) gave 0.38 g (30% yield) of desired Boc-beTyr-OMe as a white powder.

Boc-beTyr-OMe: TLC R_f = 0.33 (hexane/AcOEt = 3:1); ^{1}H NMR ($CDCl_3$) spectrum is shown below.

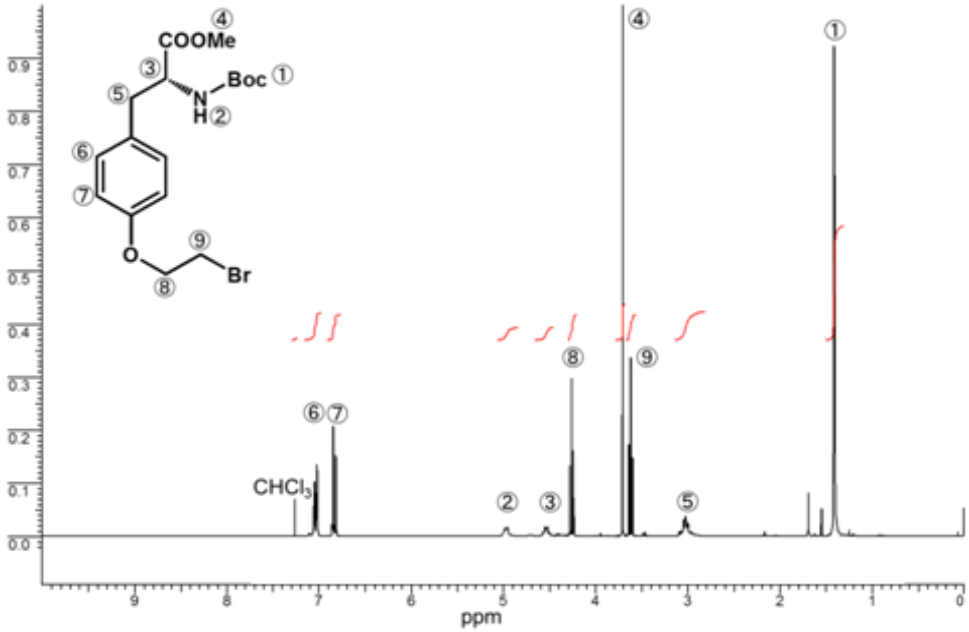


Figure S3. ^{1}H NMR spectrum of Boc-beTyr-OMe.

2-2-2. Boc-beTyr-OH

To a chilled solution of Boc-beTyr-OMe (0.38 g, 0.94 mmol) in THF (0.3 mL), 1 M LiOH in water was added dropwise. The mixture was stirred for 2 hr at 0 ^{o}C and then the solvent was evaporated. The mixture was acidified to pH 1 with 5% aqueous

$KHSO_4$ and extracted with ethyl acetate. The extract was dried over sodium sulfate, and concentrated under reduced pressure to afford desired Boc-beTyr-OH quantitatively which was used for the next reaction without further purification.

Boc-beTyr-OH: TLC R_f = 0.38 (chloroform/methanol = 9:1); ^{1}H NMR ($CDCl_3$) spectrum is shown below.

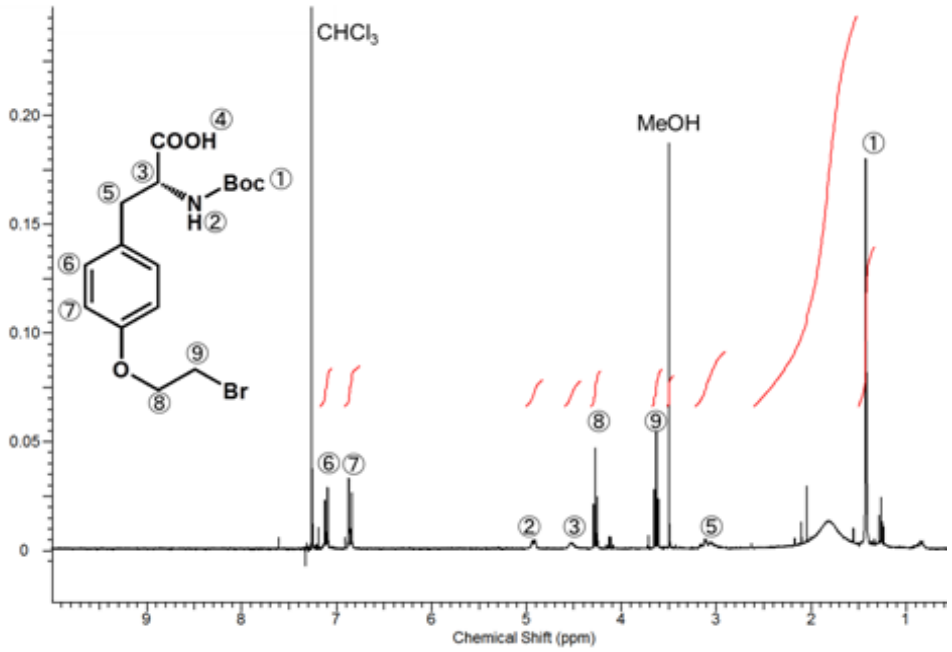


Figure S4. ^{1}H NMR spectrum of Boc-beTyr-OH.

2-2-3. H-beTyr-OH

The deprotection of Boc group of Boc-beTyr-OH with trifluoroacetic acid (TFA) was carried out by following the same procedure as previously described.[7]

2: MALDI-TOF MS (m/z) found: 288.34 $[M+H]^+$; calcd. for $[M+H]^+$: 288.03; ^{1}H NMR (300 MHz, D_2O) spectrum is shown below.

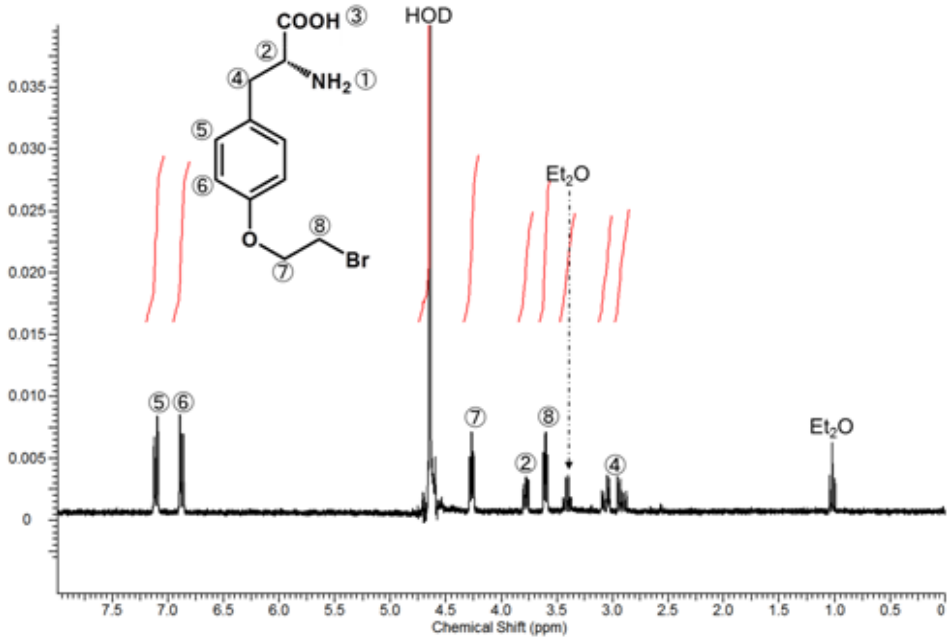
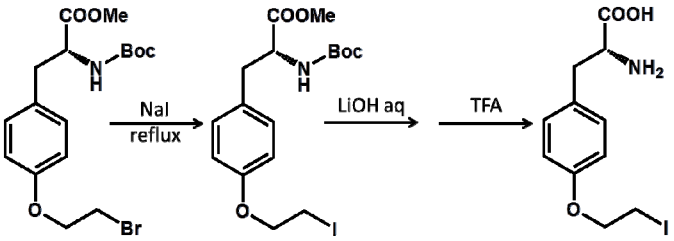


Figure S5. ^{1}H NMR spectrum of *O*-bromoetyl-*L*-tyrosine (beTyr; **2**).

2-3. *O*-iodoethyl-*L*-tyrosine (ieTyr; 3)



2-3-1. Boc-ieTyr-OMe

Sodium iodide (60 mg, 0.40 mmol) and Boc-beTyr-OMe (89 mg, 0.22 mmol) were mixed in dry acetone (1.5 mL), refluxed for 6 hr, and brought to room temperature. The solvent was evaporated to afford yellowish white solid. The crude reaction product was mixed with water, and extracted with ethyl acetate. The extract was washed with saturated aqueous NaCl, dried over sodium sulfate, and concentrated under reduced pressure. The obtained yellowish white solid (6.4 mg; 10% yield) was used for the next reaction without further purification.

Boc-ieTyr-OMe: TLC R_f = 0.38 (hexane/AcOEt = 3:1); ^{1}H NMR ($CDCl_3$) spectrum is shown below.

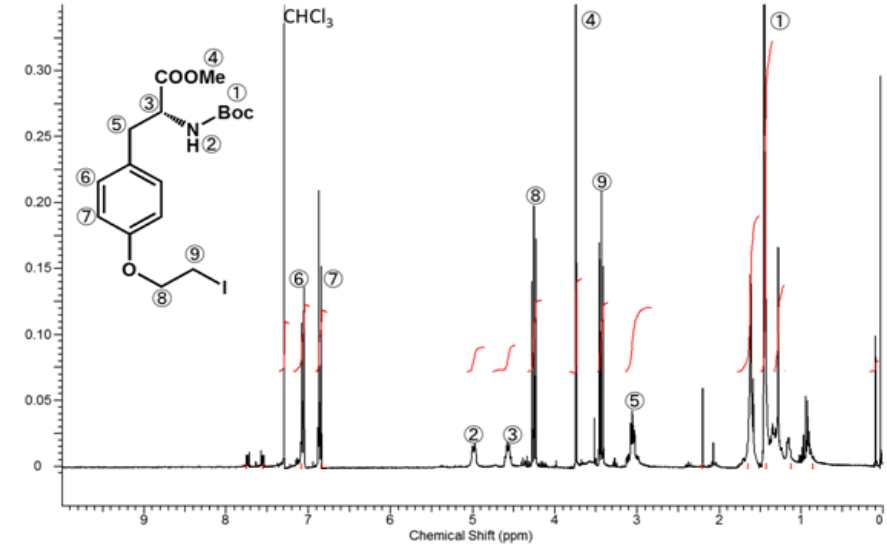


Figure S6. ^{1}H NMR spectrum of Boc-ieTyr-OMe.

2-3-2. H-ieTyr-OH

To a chilled solution of Boc-ieTyr-OMe (6.4 mg, 0.014 mmol) in THF (0.1 mL), 1 M LiOH in water was added dropwise. The mixture was stirred for 3 hr at 0 ^{o}C and then the solvent was evaporated. The mixture was acidified to pH 2 with 5% aqueous $KHSO_4$ and extracted with ethyl acetate. The extract was dried over sodium sulfate, and concentrated under reduced pressure to afford desired Boc-ieTyr-OH. The deprotection of Boc group of Boc-ieTyr-OH with trifluoroacetic acid (TFA) was carried out by following the same procedure as previously described.[7]

Boc-ieTyr-OH:TLC R_f = 0.33 (chloroform/methanol = 9:1).

3: MALDI-TOF MS (m/z) found: 336.39 $[M+H]^+$; calcd. for $[M+H]^+$: 336.01; ^{1}H NMR (300 MHz, D_2O) spectrum is shown below.

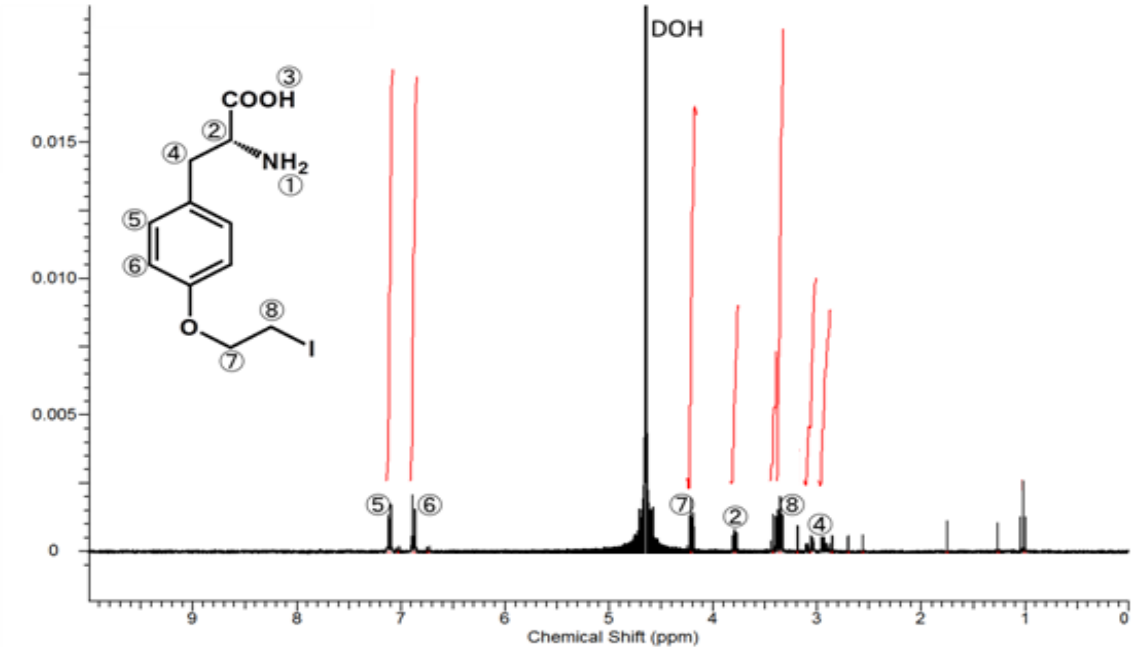


Figure S7. ^{1}H NMR spectrum of *O*-iodoethyl-*L*-tyrosine (ieTyr; **3**).

3. Introduction of non-natural amino acids at the N-terminus of a peptide via the NEXT-A reaction followed by intramolecular cyclization.

In vitro N-terminal-specific introduction of non-natural amino acids followed by the cyclization, namely the NEXT-A/Cyclization, was carried out in a 5 μL reaction mixture containing 50 mM HEPES-KOH (pH 7.6), 20 mM KCl, 1 mM spermidine, 2.5 mM ATP, 10 mM $MgCl_2$, 2 mM dithiothreitol (DTT), each amino acid (1 nmol; 0.2 mM), target peptide (180 pmol; 35 μM), $tRNA^{Phe}$ (39 pmol; 7.7 μM), mutant ARS (5.0 pmol; 1.0 μM), and wild-type L/F-transferase (7.3 pmol; 1.5 μM). The mixture was incubated in the dark at 37 ˚C for 1 hr. The mixture was desalted, concentrated, and purified using ZipTipC_{18}, and analyzed by MALDI-TOF-MS.

In case of a model peptide (UTIF; RGP**C**RAFI) in which *O*-haloethyl-*L*-tyrosine was introduced, quantitative cyclization was achieved when the incubation time was prolonged to 5 hr.

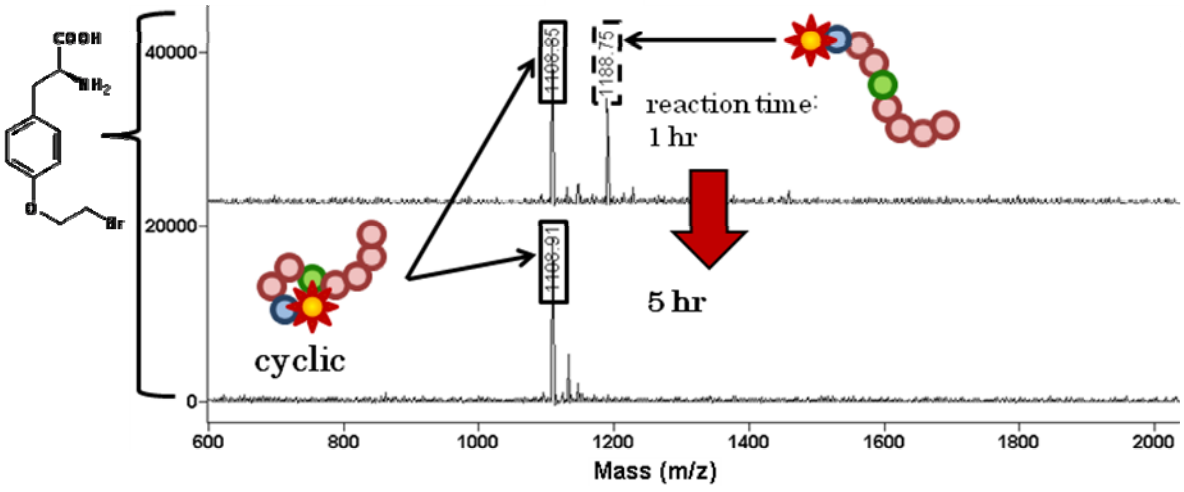


Figure S8. MALDI-TOF mass spectra of a model peptide (UTIF; RGP**C**RAFI) observed after the NEXT-A/Cyclization reaction. The incubation time was 1 hr (upper), and 5 hr (lower).

The NEXT-A/Cyclization reaction is also carried in a 1 mL reaction scale containing a target peptide (RVC-acdAla; 22 µg, 35 nmol), an amino acid (caaPhe), and other reagents with the same molar concentration shown above. The molar concentrations of target peptide, tRNA, and amino acid have been optimized and should not be altered when changing the reaction scale. After the reaction, the solution was acidified with TFA and precipitate was removed by centrifugation. The supernatant was filtered, and purified by reversed-phase HPLC (Waters, XTerra Prep, 5µC18, 10 x 50) eluting with 0.1% aq. TFA-acetonitrile gradient mixture (a linear gradient of 0-100 % acetonitrile over 10 min at flow rate 5 mL/min was used) to afford pure cyclic peptide (17 µg, 20 nmol; 57% yield).

Cyclized RV<u>C</u>-acdAla: MALDI-TOF MS (m/z) found: 842.45 $[M+H]^+$; calcd. for $[M+H]^+$: 842.35; HPLC (XTerra Prep) R_f = 4.2 min.

Note: tRNA was recycled from the precipitate by phenol/chloroform extraction, followed by ethanol precipitation. The recycled tRNA showed full catalytic activity for the NEXT-A/Cyclization.

4. Solid-phase synthesis of target peptides

Syntheses of the target peptides were carried out on the solid phase. Fmoc amino acids including fluorescent ones (Watanabe Chemicals) were used as the monomers. Virtually no side reaction took place during the peptide synthesis. The solid phase synthesis was performed on an Fmoc-NH-SAL-PEG resin (super acid-labile polyethyleneglycol resin from Watanabe Chemicals). Chain elongation was achieved by using 1-Hydroxy-7- azabenzotriazole (HOAt) and *N,N,N′,N′*-tetramethyl-*O*-(7-azabenzo-triazol-1-yl)uronium hexafluorophosphate (HATU) as the coupling agent in the presence of diisopropylethylamine (DIPEA) / *N,N*-dimethylformamide (DMF) / N-Methylpyrrolidone (NMP). The coupling time was set to 40 ˚C for 12 and 3 hrs when fluorescent non-natural and other amino acids were elongated, respectively. The coupling efficiency at each elongation step was almost quantitative, as estimated from UV absorption of the fulvene-adduct formed upon removal of the N_{δ}-Fmoc protecting group with 20% piperidine / DMF for 7 min at room temperature (ε_{290}= 4950 M^{-1} cm^{-1}). The resin-bound peptide was cleaved off from the resin with trifluoroacetic acid (TFA) / water / 1,2-ethandithiol (EDT) / triisopropylsilane (TIS) (94/2.5/2.5/1 v/v/v/v) for 1 hr at room temperature. Each obtained peptide was analyzed by a reversed-phase HPLC (C18 column) eluting with 0.1% aq. TFA-acetonitrile gradient mixture (a linear gradient of 0-100 % acetonitrile over 60 min at flow rate 0.6 mL/min was used), and only a single peak was found. Each was also identified by

MALDI-TOF mass spectroscopy before and after the NEXT-A/Cyclization reaction with **1**.

RPC-acdAla: MALDI-TOF MS (m/z) found: 638.54 $[M+H]^+$; calcd. for $[M+H]^+$: 638.29.

Cyclized RPC-acdAla: MALDI-TOF MS (m/z) found: 840.47 $[M+H]^+$; calcd. for $[M+H]^+$: 840.37.

RVC-acdAla: MALDI-TOF MS (m/z) found: 640.61 $[M+H]^+$; calcd. for $[M+H]^+$: 640.31.

Cyclized RVC-acdAla: MALDI-TOF MS (m/z) found: 842.67 $[M+H]^+$; calcd. for $[M+H]^+$: 842.35.

RGDCGGSDYKDDDDK-badAla: MALDI-TOF MS (m/z) found: 1958.56 $[M+H]^+$; calcd. For $[M+H]^+$: 1958.76.

Cyclized RGDCGGSDYKDDDDK-badAla: MALDI-TOF MS (m/z) found: 2160.47 $[M+H]^+$; calcd. for $[M+H]^+$: 2160.84.

5. Tandem mass-mass analysis of the peptides

All of the above peptides were subjected to tandem mass-mass analysis (Shimadzu, AXIMA-QIT (MALDI-QIT-TOF)) for further identification.

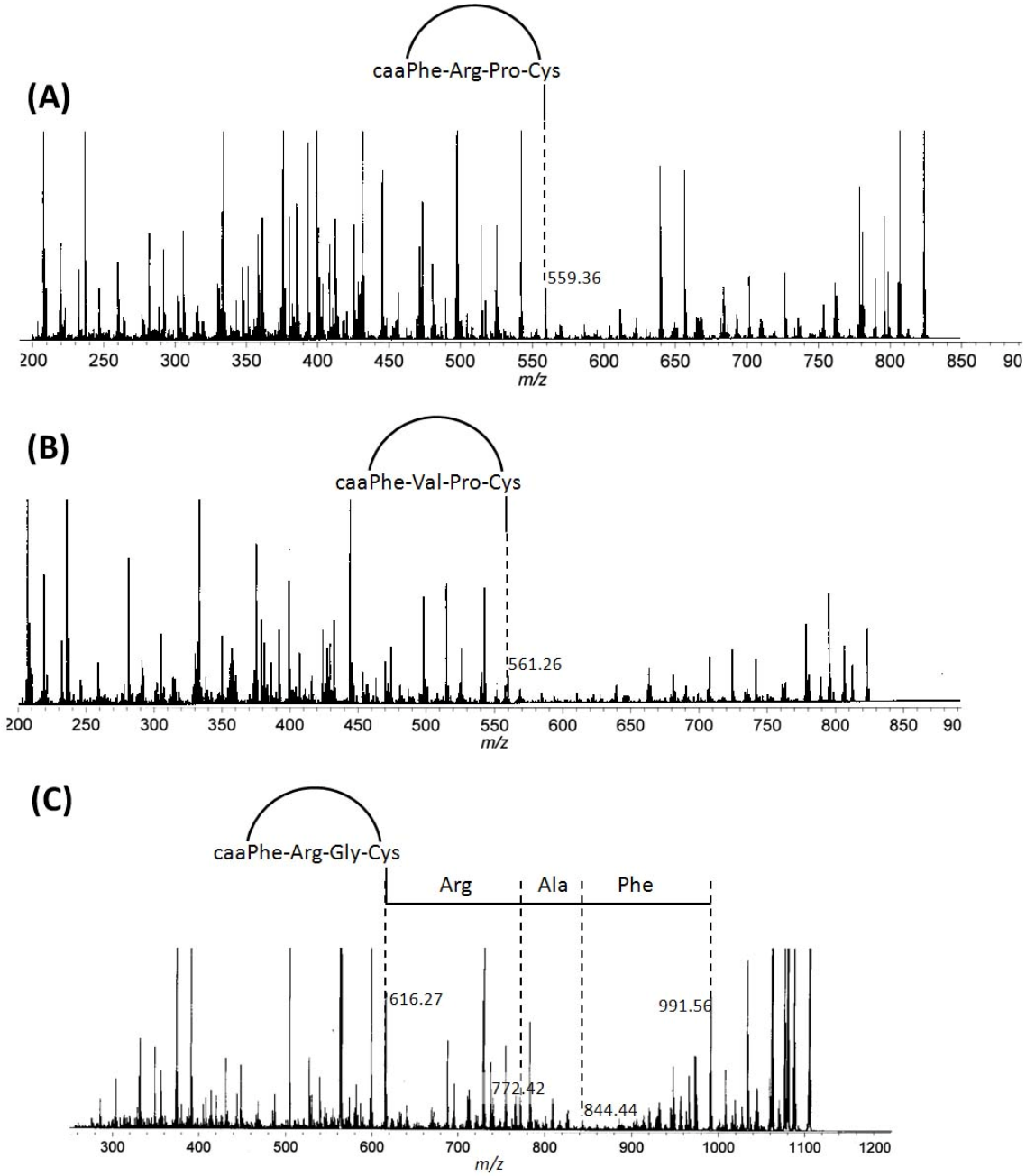


Figure S9. Tandem mass-mass spectra of (A) cyclized RPC-acdAla, (B) cyclized RVC-acdAla, and (C) cyclized UTIF. Peaks correspond to a set of assigned sequential fragment ions, generated by cleavage of amide bonds, are shown.

6. Cloning and expression of linear peptide-fused green fluorescent protein (GFP[8]).

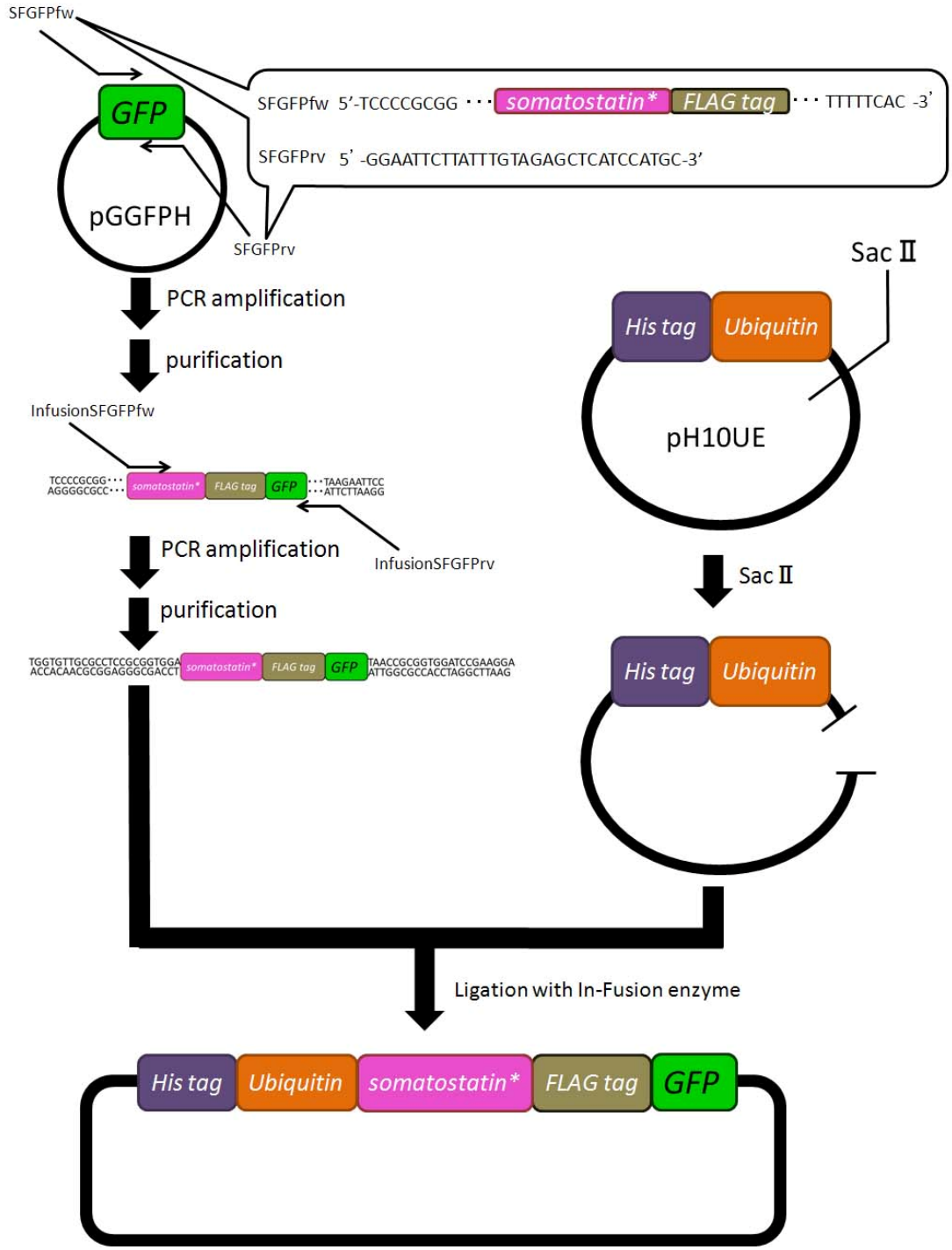


Figure S10.

Cloning of the expression vector (pH10UE-Lys-somatostatin*-FLAG-GFP).

To expose a single lysine (Lys) moiety at the N-terminus, we cloned a linear peptide (Lys-somatostatin*-FLAG)-fused GFP coding sequence into a plasmid vector pH10UE that contained a His10 tag and a ubiquitin (Ub)-coding region.[1] When polymerase chain reaction (PCR) amplification of the Lys-somatostatin*-FLAG-GFP coding sequence originated from pGGFPH plasmid[9], the 3' and 5' regions of homology were generated by adding 19 bp extensions to both PCR primers (underlined) that precisely match the ends of the linearized vector of pH10UE. The PCR primers (Hokkaido System Science, Japan) used in this study are shown below;

(1st PCR amplification)

SFGFPfw:

TCCCCGCGGTGGAAAAAATTTCTTCTGGAAAACCTTCACCAGCTGTGATTAT

AAAGATGA TGATGATAAA AGTAAAGGAG AAGAACTTTT CAC

SFGFPrv:

GGAATTCTTATTTGTAGAGC TCATCCATGC

(2nd PCR amplification)

InfusionSFGFPfw:

TGGTGTTGCGCCTCCGCGGTGGAAAAAATTTCTTCTGGAAAAC

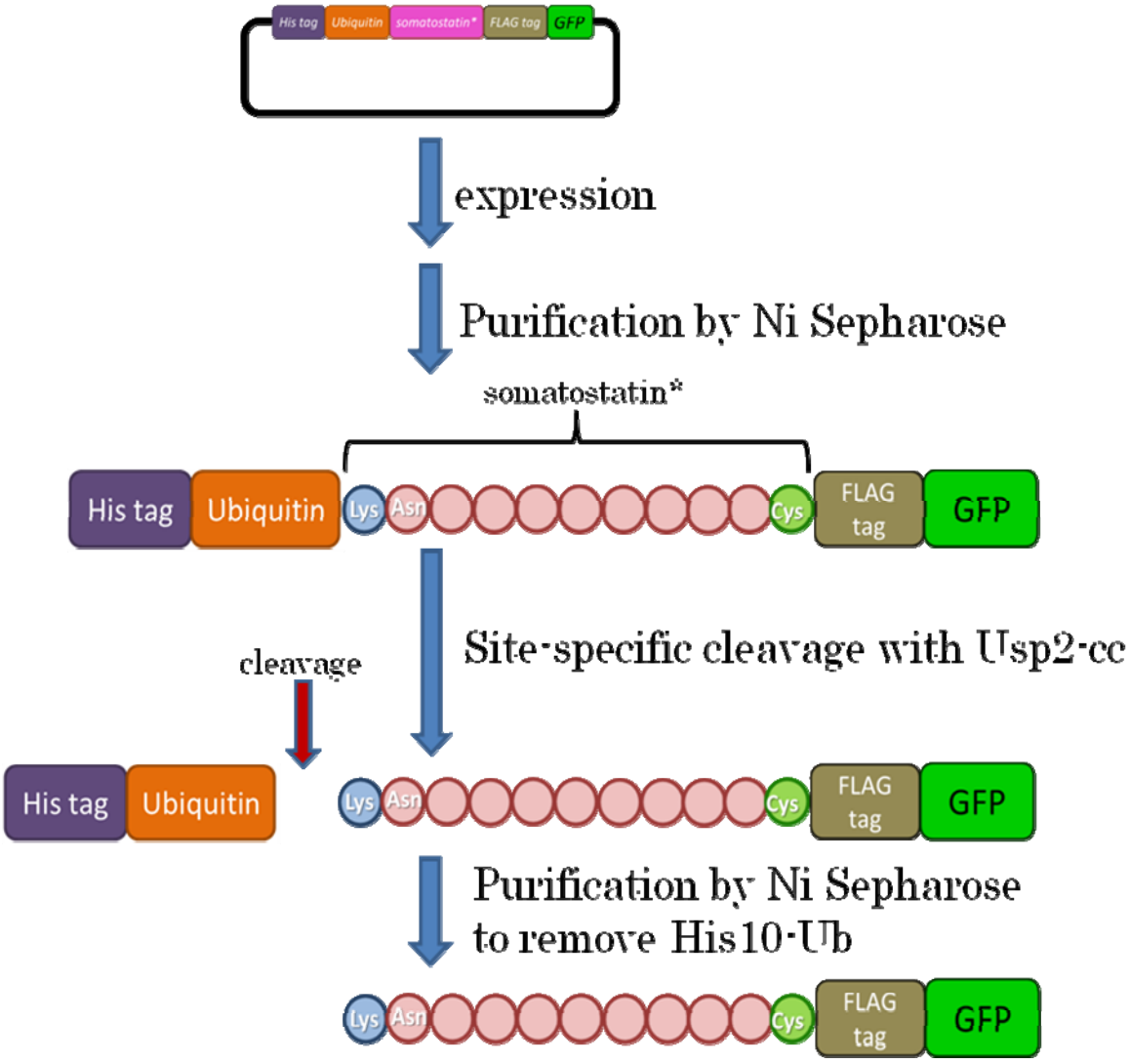
InfusionSFGFPrv:

GAATTCGGATCCACCGCGGTTATTTGTAGAGCTCATCCATG

The resulting PCR fragment was ligated into *Sac*II digested pH10UE in the presence of the In-Fusion® enzyme (Clontech) according to the instruction manual, to obtain a plasmid pH10UE-Lys-somatostatin*-FLAG-GFP. The plasmid was sequenced by an automated DNA sequencer. The results indicated that the

Lys-somatostatin*-FLAG-GFP encoding region was successfully inserted just after the 3' end of ubiquitin-encoding region.

The plasmid was transformed into BL21 (DE3) *E. coli* competent cells (Takara, Japan). The transformant was cultivated in 500 mL of LB medium (Nakalai, Japan) containing ampicillin at 37 ˚C, with shaking to optical density at 600 nm of 0.85. After addition of isopropyl-1-thio-*β*-*D*-galactoside (IPTG) to a final concentration of 10 mM, the cells were incubated at 25 ˚C for 12 h.



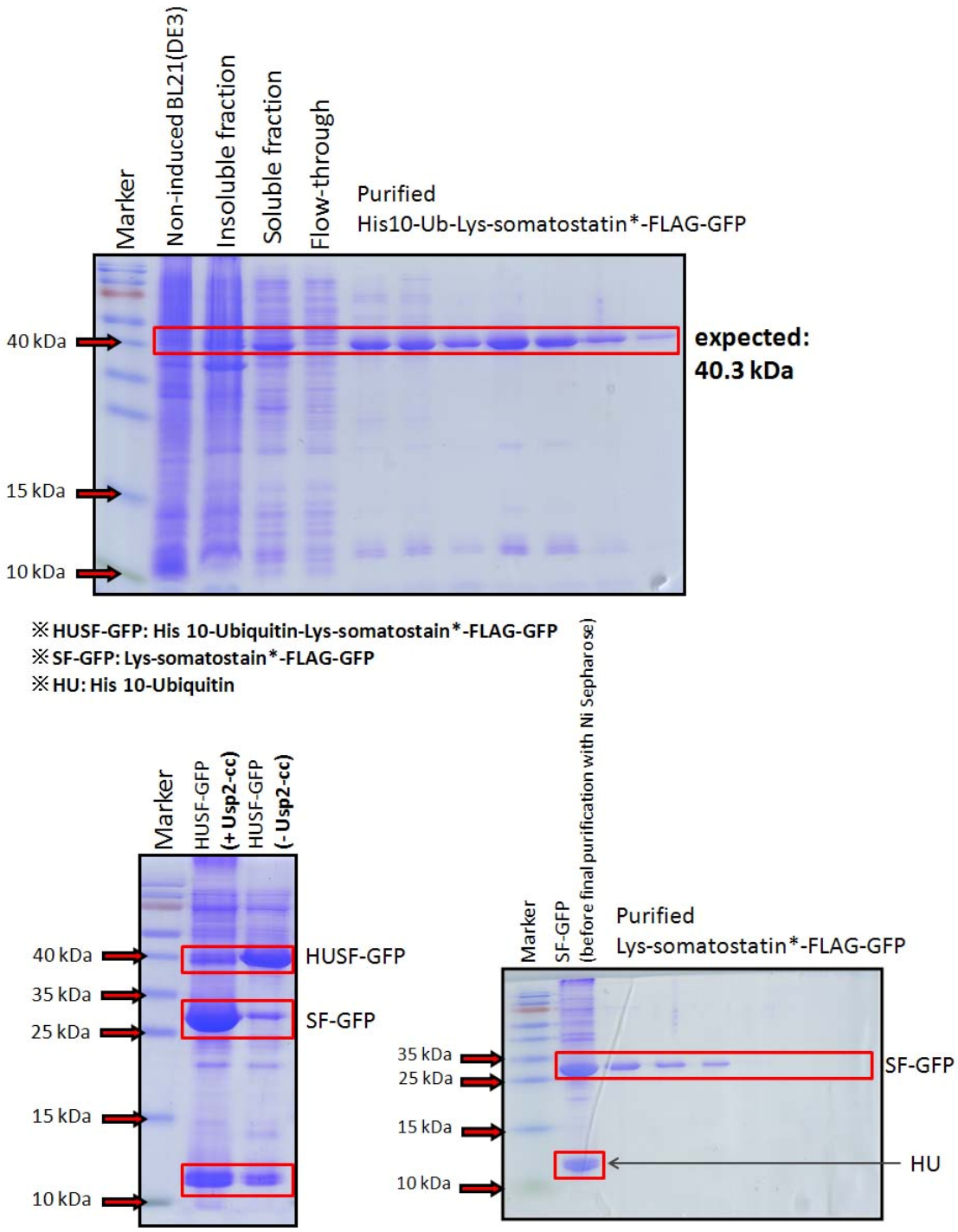


Figure S11. Purification of Lys-somatostatin*-FLAG-GFP. Each protein fraction was subjected to SDS-PAGE (15%) and the gels were stained with Coomassie Brilliant Blue (CBB).

The expressed His10-Ub-Lys-somatostatin*-FLAG-GFP was purified from *E. coli* using Ni Sepharose 6 Fast Flow (GE) according to the instruction manual. Amicon Ultra centrifugal filter (Millipore) was used for the concentration and buffer exchange toward 20 mM Tris-HCl (pH 8.0) / 150 mM NaCl / 10 mM 2-mercaptoethanol. Ub moiety was then cleaved in the same buffer with a purified deubiquitylating enzyme (Usp2-cc)[1] followed by final purification with Ni Sepharose 6 Fast Flow, to obtain Lys-somatostatin*-FLAG-GFP. Amicon Ultra centrifugal filter was used for the concentration and N-terminal amino-acid sequence of the Lys-somatostatin*-FLAG-GFP was analyzed by an amino-acid sequencer. The obtained protein was also analyzed by 15% SDS-PAGE (Fig. S11), and confirmed on MALDI-TOF-MS [Lys-somatostatin*-FLAG-GFP: (m/z) found 29045, calculated for $[M + H]^+$: 29050].

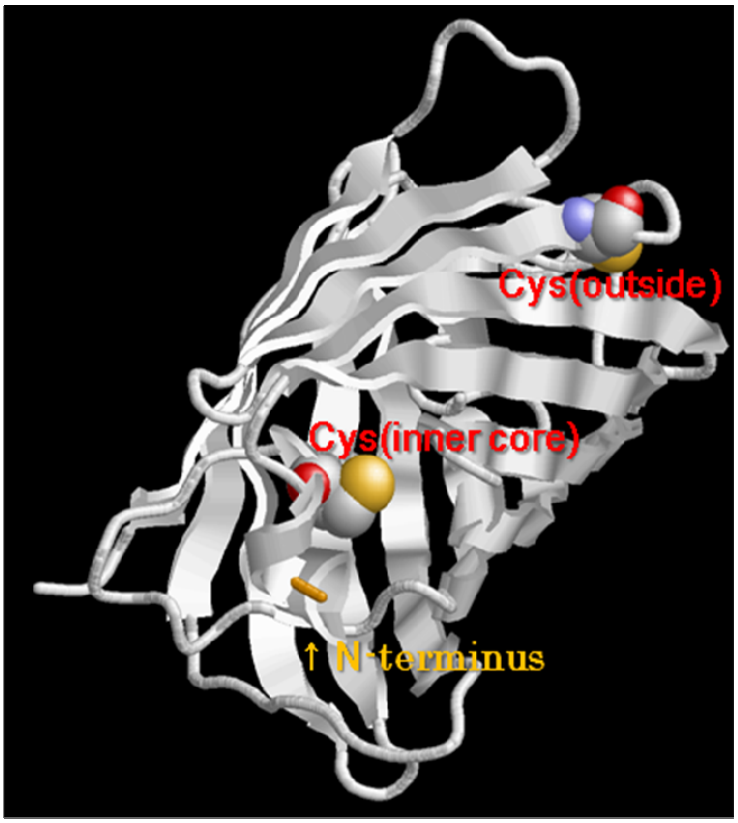


Figure S12. Structure of GFP (PDB: 1b9c).

7. caaPhe introduction at the N-terminus of Lys-somatostatin*-FLAG-GFP followed by cyclization via the NEXT-A/Cyclization reaction

Typically, *in vitro* N-terminal specific introduction of caaPhe (**1**) was carried out in a 30 μL reaction mixture containing 50 mM HEPES-KOH (pH 7.6), 20 mM KCl, 1 mM spermidine, 2.5 mM ATP, 10 mM $MgCl_2$, caaPhe (3 nmol; 0.1 mM), Lys-somatostatin*-FLAG-GFP (1.1 nmol; 37 μM), $tRNA^{Phe}$ (5.4 nmol; 7.7 μM), ARS (αA294G/αT251A mutant; 60 pmol; 2.0 μM), and wild-type L/F-transferase (0.17 nmol; 5.8 μM). The mixture was incubated in the dark at 37 ˚C for 1 hr. Introduction of caaPhe followed by cyclization was confirmed by Edman degradation (Fig. S13) or MALDI-TOF-MS analysis.

Enterokinase digestion was also performed for further identification according to the instruction manual obtained from Novagen. The digested fragments were purified using ZipTipC_{18}, and analyzed by MALDI-TOF-MS.

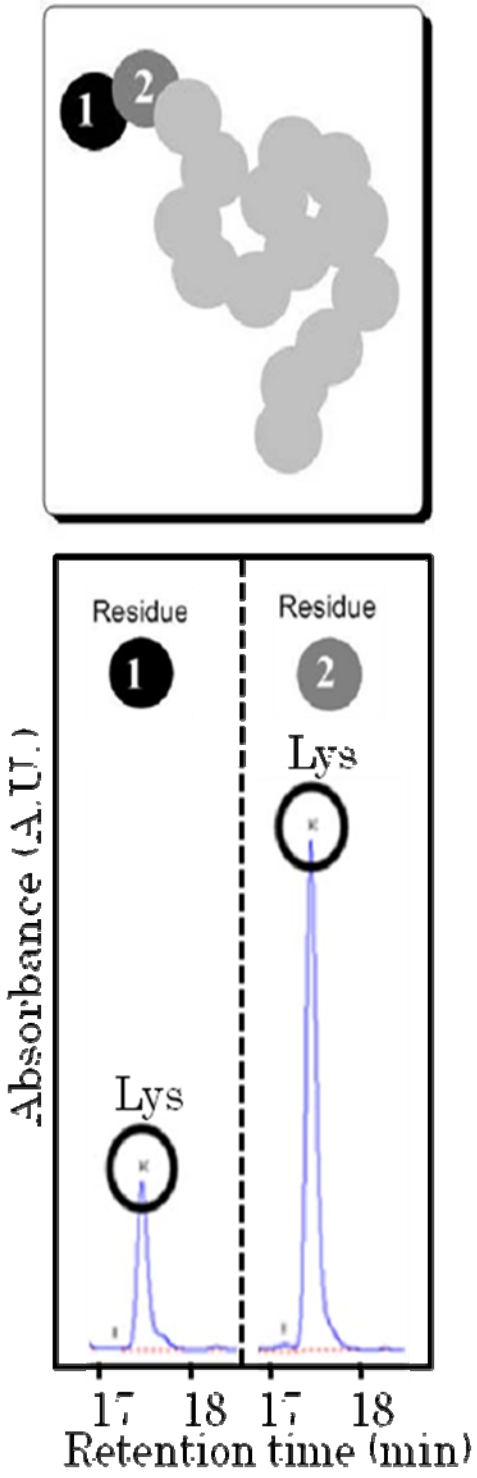


Figure S13. Quantitative analysis of the caaPhe-incorporation efficiency[1, 2] under the standard reaction condition. When caaPhe was successfully introduced to the N-terminus of Lys-somatostatin*-FLAG-GFP, a Lys unit will be detected as the 2nd residue. Otherwise, it will be detected as the 1st residue. Chromatographic analysis of the phenylthiohydantoin derivatives of the N-terminal amino acid separated from the remaining portion of the protein identifies the residues on the basis of Edman degradation. Changing molar ratio of the reagents, addition of ATP re-generating system (creatine phosphate / creatine phosphokinase), or elongation of the reaction time did not affect the incorporation efficiency.

17. Supporting references:

1. K. Ebisu, H. Tateno, H. Kuroiwa, K. Kawakami, M. Ikeuchi, J. Hirabayashi, M. Sisido and M. Taki, *Chembiochem*, 2009, 10, 2460-2464.
2. M. Taki, A. Kuno, S. Matoba, Y. Kobayashi, J. Futami, H. Murakami, H. Suga, K. Taira, T. Hasegawa and M. Sisido, *Chembiochem*, 2006, 7, 1676-1679.
3. T. Hohsaka, D. Kajihara, Y. Ashizuka, H. Murakami and M. Sisido, *J. Am. Chem. Soc.*, 1999, 121, 34-40.
4. R. E. Connor, K. Piatkov, A. Varshavsky and D. A. Tirrell, *Chembiochem*, 2008, 9, 366-369.
5. A. Szymanska, K. Wegner and L. Lankiewicz, *Helvetica Chimica Acta*, 2003, 86, 3326-3331.
6. M. Taki, Y. Yamazaki, Y. Suzuki and M. Sisido, *Chemistry Letters*, 2010, 39, 818-819.
7. M. Taki, T. Hohsaka, H. Murakami, K. Taira and M. Sisido, *FEBS Lett*, 2001, 507, 35-38.
8. R. Y. Tsien, *Annual Review of Biochemistry*, 1998, 67, 509-544.
9. M. Taki, S. Y. Sawata and K. Taira, *J. Biosci. Bioeng.*, 2001, 92, 149-153.