Electronic Supplementary Material (ESI) for ChemComm.

Supporting Information

Butelase-mediated synthesis of protein thioesters and its application for tandem chemoenzymatic ligation

Yuan Cao, Giang K T Nguyen, James P Tam* and Chuan-Fa Liu*

Supplementary Figure S1-S4

Materials and Methods

Material and Reagents

High Performance Liquid Chromatography (HPLC)

Mass spectrometry

Glycine thioester synthesis

Peptide synthesis

Cloning and expression of recombinant proteins

Dual labeling of ubiquitin

Mass spectra of the synthesized peptides

NMR spectra

Supplementary Figure S1-S4

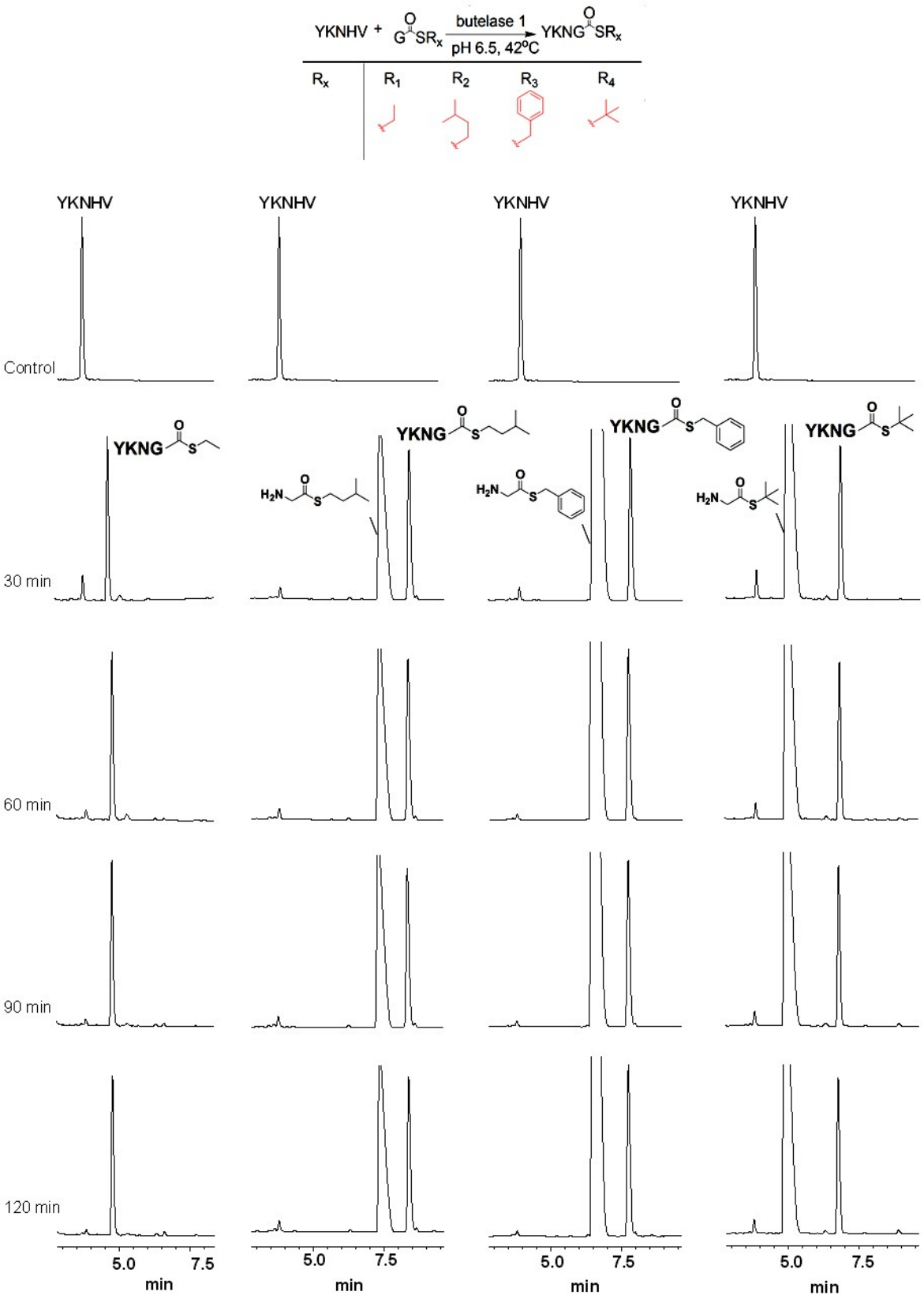


Figure S1. HPLC profile of the ligation between the four glycine thioesters **1a-d** and the model peptide YKNHV **2**. Glycine thioester **1a** eluted out before the displayed range.

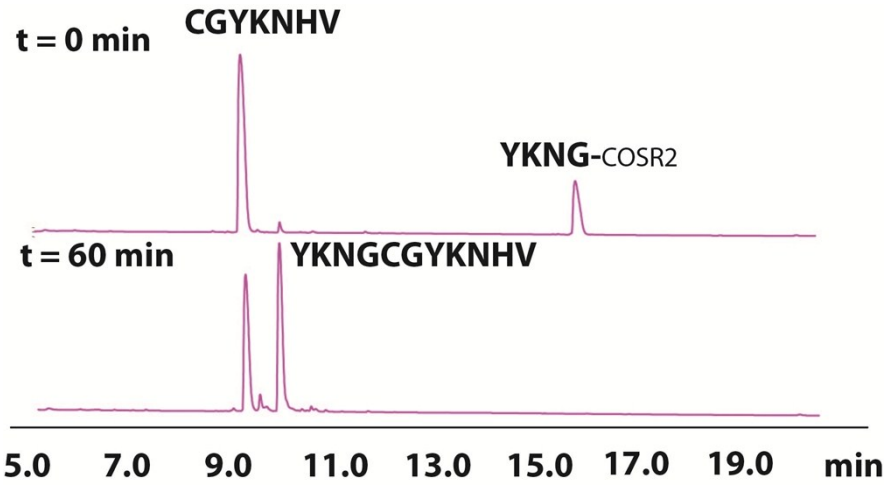


Figure S2. HPLC profile for the ligation between H-YKNG-COSR **3b** (1 mM) and H-CGYKNHV-NH_2 (3 mM) in the presence of 50 mM MESNa and 25 mM TCEP at pH 7.4 PBS, 42 °C.

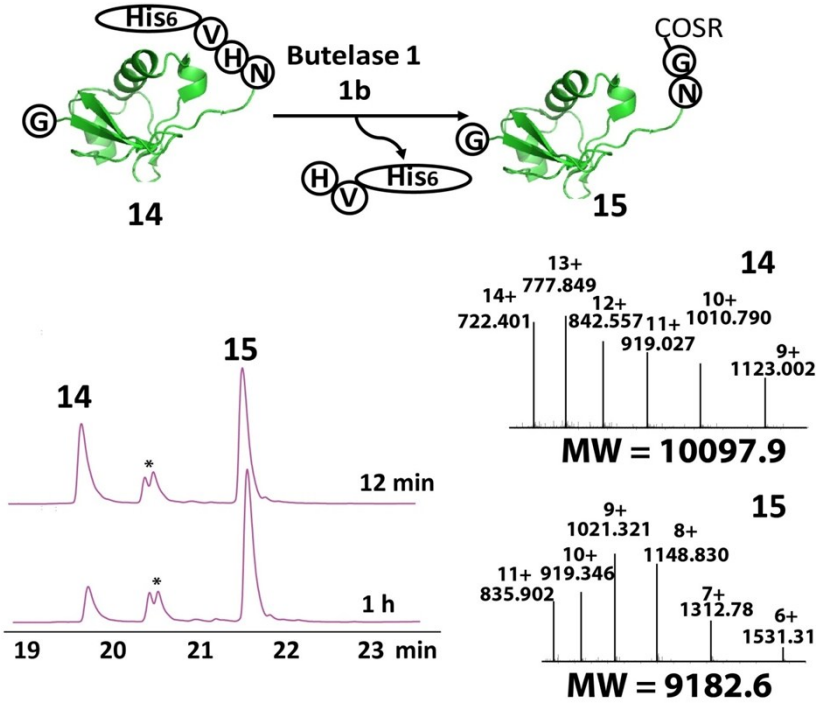


Figure S3. Schematic illustration of tagging G-Ub-NHV-His_6 with Gly-$COSR_2$.and its HPLC analysis. Reaction conditions: 100 μM **14**, 2 mM Gly-$COSR_2$ **1b**, 200 nM butelase 1, pH 6.5 20 mM phosphate buffer, 42 °C. Unmodified and modified ubiquitin **14** and **15** have observed masses of 10097.9 Da and 9182.6 Da, respectively. Asterisk-marked peaks are impurities associated with protein expression.

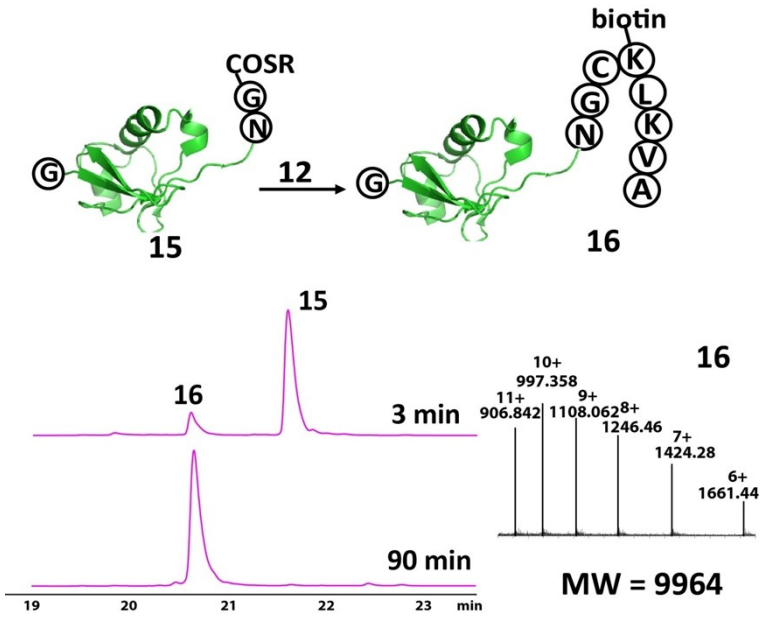


Figure S4. Schematic illustration of the C-terminal labeling of G-Ub-NG-$COSR_2$ **15** with CK(biotin)LKVA **12** and its HPLC analysis. Reaction conditions: 100 µM protein **15**, 500 µM CK(biotin)LKVA **12**, 20 mM MESNa, 5 mM TCEP in PBS, 42 °C. Biotin-labeled ubiquitin **16** has an observed mass of 9964 Da.

Materials and Methods

Materials and Reagents

Amino acids, coupling reagents and resins were obtained from chemimpex, Novabiochem and GL biochem. All other chemical reagents were of analytical grade, obtained from Sigma Aldrich, alfa aesar and Acros Organics. All solvents and chemicals were used as received without purification unless otherwise indicated. The plasmid of Sortase A was purchased from addgene. TLC analyses were carried out on plates pre-coated with silica gel. Visualization was facilitated with UV light and ninhydrin solution.

High Performance Liquid Chromatography (HPLC)

The analytical HPLC analyses were carried out on NexraX_2 LC-30AD with two analytical columns (Aeris peptide XB-C18, 4.6 x 250 mm for peptide analysis, Aeris widepore 3.6u C4, 4.6 x 150 mm for protein analysis). The peptide/protein purifications were performed using Shimadzu HPLC equipped with a semi-preparative HPLC column (Jupiter C18, 5 μM, 10 x 250 mm). All the HPLC runs were done using the mixture of two solutions, A (0.045% TFA in water) and B (0.039% TFA in acetonitrile/water (9/1) mixture). Flow rate was 0.8 ml/min for analysis and 2.5 ml/min for purification. UV detection was carried out at 220 nm.

Mass spectrometry

Small peptides ESI mass spectra data were obtained on a Thermo Finnigan LCQ DECA XP MAX with ESI ion source. Protein ESI mass spectra were measured on Q Exative hybrid Quadrupole-Orbitrap Mass Spectrometer. The deconvoluted data were obtained using the software of MagTran 1.03 and ESIProt 1.0.

Glycine thioesters synthesis

All the glycine thioesters used in this work were synthesized using standard Boc chemistry. A mixture containing one eq. of thiol compound, 1.2 eq. of Boc-Gly-OH, one eq. of benzotriazol-1-yl-oxytripyrrolidinophophonium hexafluorophosphate (PyBOP) were first dissolved in DCM. Two eq. of *N*,*N*-Diisopropylethylamine (DIEA) were added subsequently. The reaction was stirred for 2 h at room temperature. The solvent was removed under vacuum and the residue was purified using silica gel chromatography. The purified product was further treated with 30% TFA in DCM for 30 min to remove the Boc protecting group. The TFA and DCM were evaporated under vacuum followed by overnight lypholization. The final desired products were obtained in the form of TFA salt and kept as stock in the fridge or used without further purification. Thiols used were ethanethiol, 3-methyl-1-butanethiol, benzyl mercaptan, 2-methyl-2-propanethiol.

Peptide synthesis

The peptides H-**YXNHV**-NH_2 and H-**CK**(biotin)**LKVA**-NH_2 were synthesized using standard Fmoc chemistry. The syntheses of H-**YXNHV**-NH_2 and H-**CK**(biotin)**LKVA**-NH_2 were done on Rink amide MBHA resin. Before use, resin was pre-swelled using DCM for 10 min. Before the first coupling, an Fmoc deprotection was performed using 20% piperidine in dimethylformamide (DMF) for 2 min and a subsequent 20 min. After deprotection, the resin was washed with DMF, DCM and DMF alternatively. For the coupling reactions, 4 eq. of Fmoc-AA-OH, 4 eq. of PyBOP were first dissolved in DMF/DCM mixture. After the mixture was added to the resin, 8 eq. of DIEA was added. Coupling

reactions were carried out for 60 to 90 min. Coupling efficiency was examined by Kaiser test. The Fmoc amino acids used for the synthesis of YXNHV (X = K, V, L, S, F, (D)A, Nle) were Fmoc-Tyr(tBu)-OH, Fmoc-Lys(Boc)-OH, Fmoc-Val-OH, Fmoc-Leu-OH, Fmoc-Ser(tBu)-OH, Fmoc-Phe-OH, Fmoc-(D)Ala-OH, Fmoc-Nle-OH, Fmoc-Asn(Trt)-OH, Fmoc-His(Trt)-OH, Fmoc-Val-OH. For CK(biotin)LKVA, the building blocks used were Fmoc-Cys(Trt)-OH, Fmoc-Lys(biotin)-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Val-OH and Fmoc-Ala-OH. After the whole peptide was assembled, peptide was cleaved from the resin with a cocktail containing 95% TFA, 2.5% water and 2.5% TIS for one hour. After precipitated with cold ether, the peptide was purified using HPLC and characterized by mass spectrometry. Desired peptides were obtained in the white powder form after lyophilisation.

The two depsipeptides Nle-YLPET-glc-G and fluor-YLPET-glc-G were synthesized on rink amide MBHA resin. The synthesis procedure differed from the standard Fmoc chemistry for the coupling of glycolic acid and the next amino acid. As for glycolic acid coupling, 4 equivalents of glycolic acid and 4 equivalents of PyBOP were first dissolved in DCM/DMF and added to the resin. 8 eq. of DIEA was added subsequently. After coupling for 2 h and the resin washed, a 30 min treatment of 10% hydrazine in DMF was performed to remove the over-coupled glycolic acid through hydrazinolysis. For the next amino acid coupling to the hydroxyl group, DIC/HOBT/DMAP was used. The resin was first suspended in 9:1 ratio of DCM-DMF. 5 eq. of amino acid and 5 eq. of HOBt were dissolved in a minimum amount of DMF. After adding the solution to the resin, 5 eq. of DIC was added followed by 0.1 eq. of DMAP. The reaction was shaken for overnight. 5% acetic anhydride in pyridine was used after the coupling to cap the unreacted hydroxyl group. The following steps of coupling (using PyBOP), Fmoc deprotection and cleavage were done by standard Fmoc chemistry. The building blocks for the common sequence YLPET-glc-G were Fmoc-Tyr(tBu)-OH, Fmoc-Leu-OH, Fmoc-Pro-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Thr(tBu)-OH, glycolic acid, Fmoc-Gly-OH. Fmoc-Nle-OH and 5(6)-carboxyfluorescein were coupled to the peptidyl resin for the synthesis of the Nle-YLPET-glc-G and fluor-YLPET-glc-G, respectively. 5(6)-carboxyfluorescein was coupled through its carboxyl using also PyBOP as coupling reagent.

Butelase 1 preparation

Butelase 1 was isolated by a four-step chromatographic procedure, involving flash anion exchange, preparative anion exchange, size exclusion and analytical anion exchange. Briefly, freshly collected pods of Clitoria ternatea (500 gram) was blended with 1 l of extraction buffer (20 mM sodium phosphate, 1 mM EDTA, 1 mM PMSF, and 5 mM mercaptoethanol (β-ME), pH 6.0). After filtration of plant debris, ammonium sulfate was added to 15% saturation and the precipitated proteins were discarded. Ammonium sulfate was further added to 85% saturation and precipitate proteins were redissolved in 500 ml of extraction buffer. The dissolved protein extract was loaded to the flash chromatography column containing 300 ml slurry of Q Sepharose Fast Flow anion exchange resin (GE Healthcare). The column was washed with 2 l of extraction buffer and eluted with 500 ml of elution buffer (0.2 M NaCl, 20 mM sodium phosphate, 1 mM EDTA and 5 mM β-ME, pH 6.0). The eluent was dialyzed and fractionated by preparative anion exchange HPLC using 21 x 200 mm PolyWAX column (PolyLC). Fractions containing ligase activity were pooled and concentrated to a final volume of 3 ml using Amicon Ultra centrifugal filter (Millipore Ireland Ltd) with a 10-kDa MWCO. The concentrated solution was subjected to size exclusion chromatography using a 21.5 × 300 mm Biosuite HPLC column (Waters) and further purified by analytical anion-exchange chromatography using 4.6 x 200

mm PolyWax column (Poly LC). The purity of the isolated butelase 1 was determined by SDS-PAGE and silver staining. Approximately 0.4 mg butelase 1 can be obtained from 500 gram plant materials.

Cloning and expression of recombinant proteins

*Ubiquitin **6***

MGGMQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRG
GNHVHHHHHH

*Green fluorescent protein **8***

MHHHHHHSSGVDLGTENLYFQSMSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLP
VPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDF
KEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQS
ALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKNHV

*DARPin(ERK) **9***

MHHHHHHSSGVDLGTENLYFQSMGSDLGKKLLEAARAGQDDEVRILMANGADVNAHDDQGSTPLHLAAWIGH
PEIVEVLLKHGADVNARDTDGWTPLHLAADNGHLEIVEVLLKYGADVNAQDAYGLTPLHLAADRGHLEIVEVLLKH
GADVNAQDKFGKTAFDISIDNGNEDLAEILQKLNKNHV

*Ubiquitin **14***

MGGSGSGSQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLR
LRGGNHVHHHHHH

Ubiquitin **6** and **14** were cloned into the pET3b vector. Green fluorescent protein **8** and DARPin(ERK) **9** were cloned into the pNIC28-Bsa4 vector. The expressions of the proteins were done by transforming the constructed plasmids in to *E. coli* BL21 (DE3) competent cells. Single colony was inoculated into 10 ml LB culture containing ampicillin (for pET3b vector) or kanamycin (for pNIC28-Bsa4 vector) at 37 ^{o}C for overnight. The cultures were then transferred into 1 L antibiotics-containing LB culture and incubated at 37 ^{o}C until the OD600 reached ~ 0.6. A final concentration of ~ 0.5 mM IPTG was added to the culture to induce the protein expression. The culture was further incubated at 37 ^{o}C for another 3 hours. The cells were harvested by centrifugation at 9000 rpm for 15 min using Beckman coulter Avanti^{TM} centrifuge equipped with JA 10 rotor. The cell lysis was done using microfluidzer and the cell debris was removed by centrifugation at 15000 rpm for 30 min using JA 25.5 rotor. The supernatant were loaded to Ni-NTA column (Qiagen) and the desired proteins were eluted out using immidazole-containing buffer.

Dual labeling of ubiquitin

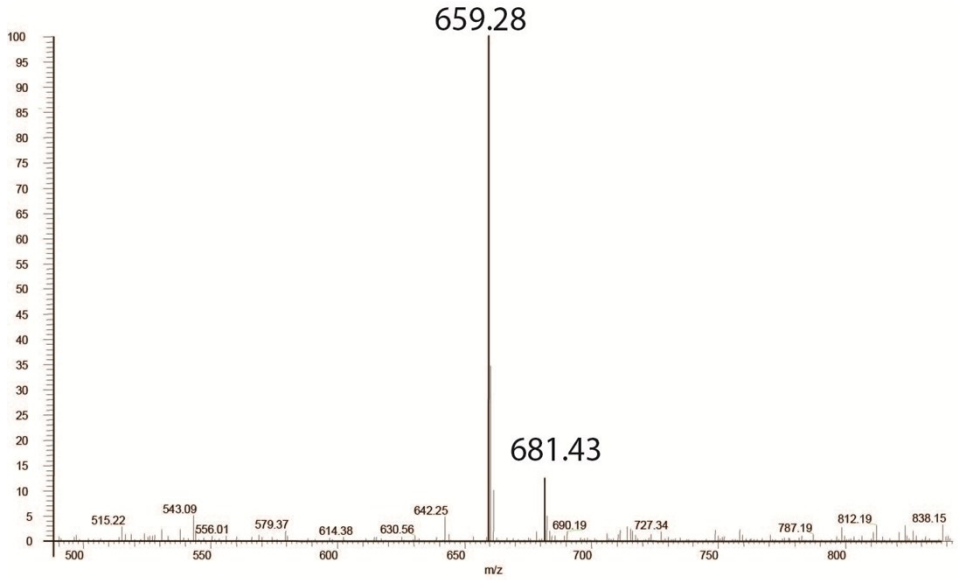
For the BML reaction, the ubiquitin (100 µM) **14** was incubated with glycine thioester **1b** (2 mM) and butelase 1 (200 nM) in pH 6.5 phosphate buffer (20 mM) at 42 °C. When the HPLC analysis showed that the reaction reached equilibrium, the desired ubiquitin thioester **15** was purified by RP-HPLC. Speedvac was subsequently used to remove as much acetonitrile in the HPLC collection as practical. Centrifugal filter was then used to concentrate the sample for the next reaction.

Subsequent NCL was performed in PBS reaction buffer containing 100 µM ubiquitin thioester **15**, 500 µM cysteinyl peptide **12**, 20 mM MESNa, 5 mM TCEP at 42 °C. The reaction was purified by RP-HPLC after analytical HPLC monitoring showed reaction completion. The acetonitrile in the collected HPLC fractions was removed by speedvac. The biotin-labeled ubiquitin **16** was concentrated using centrifugal filter for the next step.

For the SML reaction for N-terminal labeling, 50 µM biotin-labeled ubiquitin **15** was incubated in SrtA buffer (10 mM $CaCl_2$, 150 mM NaCl, 50 mM Tris, pH 7.5) with 100 µM depsipeptide **17a** or **17b**, 5 mM TCEP and 5 µM SrtA at 37 °C. Ni-NTA beads could be used for the separation of the His-tagged SrtA and the dual labeled ubiquitin **18a** or **18b**.

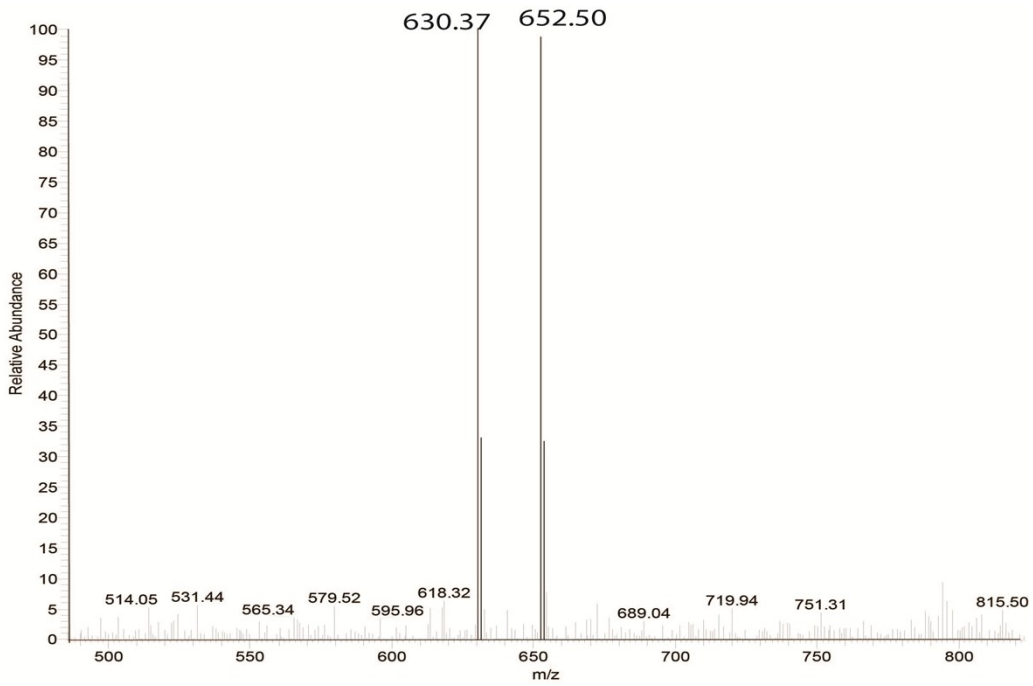
Mass spectra of the synthesized peptides

Peptide **2** H-**YKNHV**-NH_2



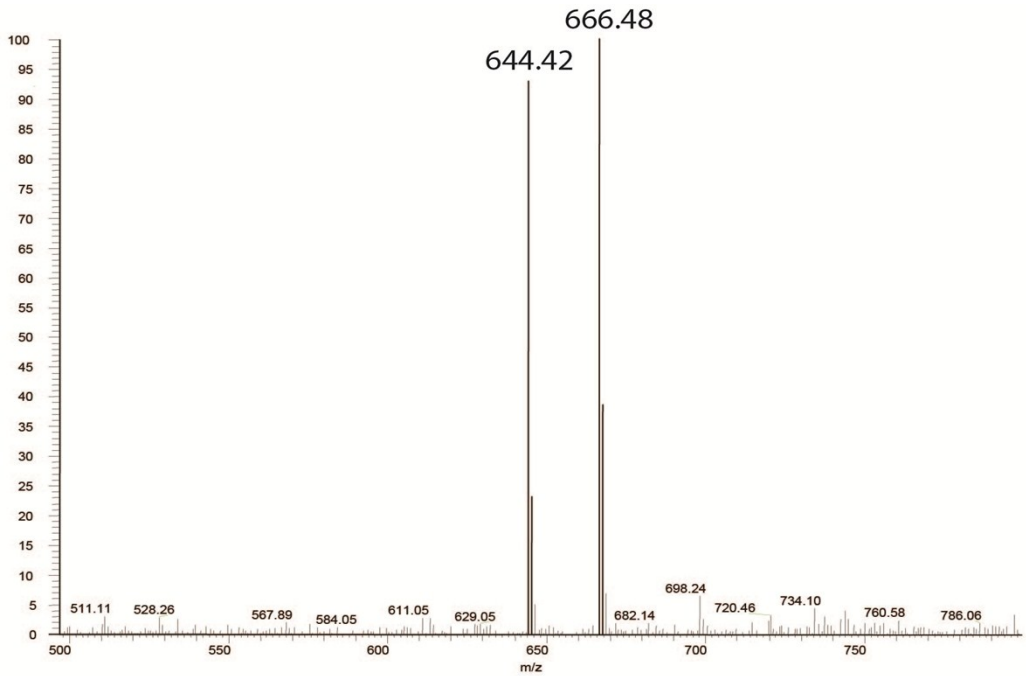
ESI-MS (positive) 659.28 (observed, M+H) 681.43 (observed, M + Na) 658.35 (calculated, M).

Peptide **4a** H-**YVNHV**-NH_2



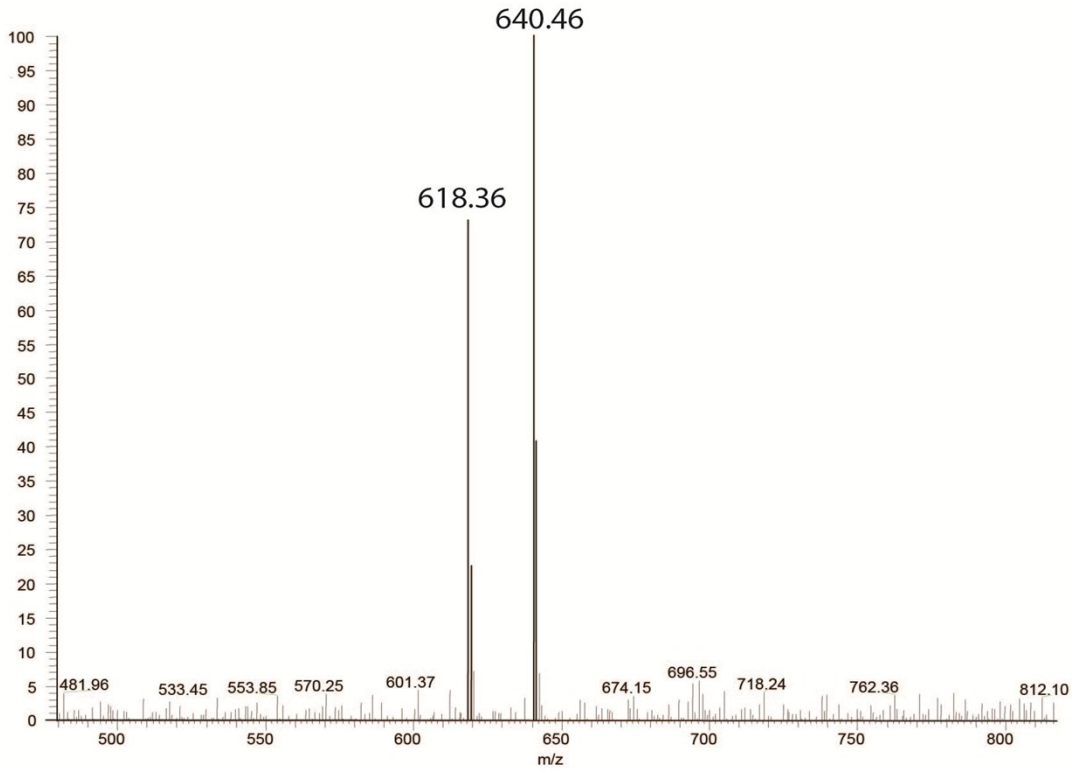
ESI-MS (positive) 630.37 (observed, M+H) 652.50 (observed, M + Na) 629.33 (calculated, M).

Peptide **4b** H-**YLNHV**-NH_2



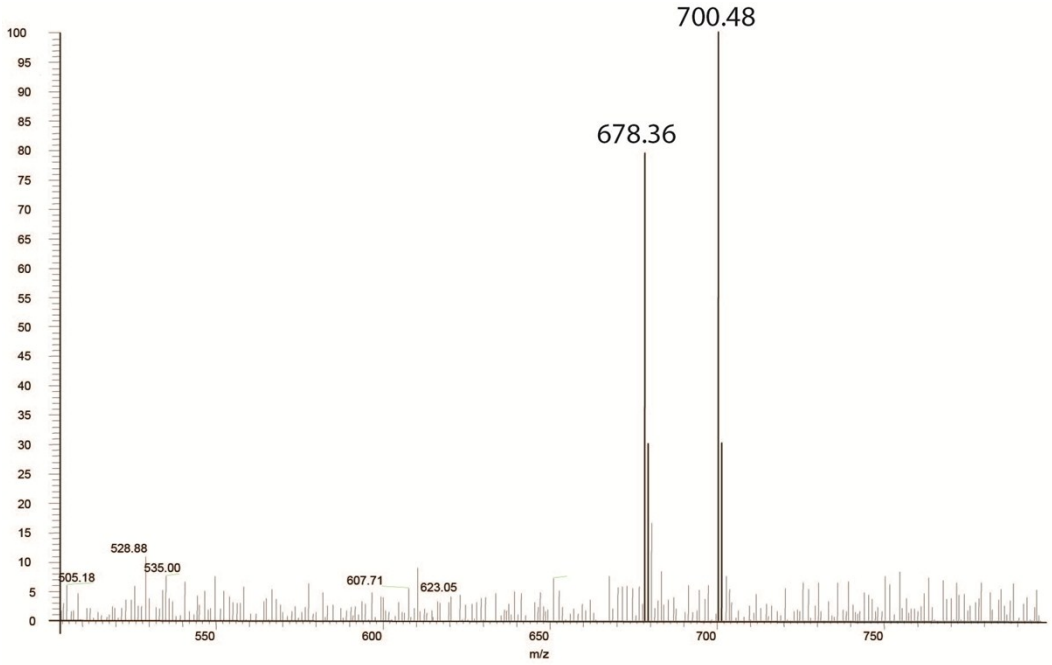
ESI-MS (positive) 644.42 (observed, M+H) 666.48 (observed, M + Na) 643.34 (calculated, M).

Peptide **4c** H-**YSNHV**-NH_2



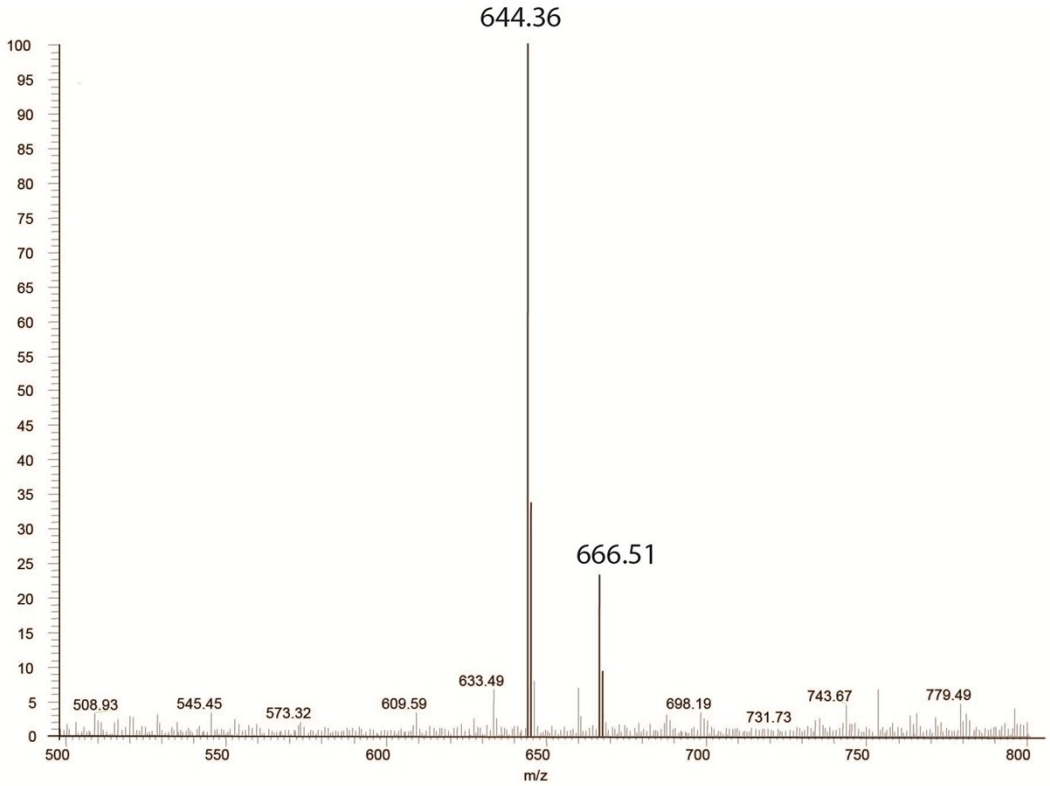
ESI-MS (positive) 618.36 (observed, M+H) 640.46 (observed, M + Na) 617.29 (calculated, M).

Peptide **4d** H-**YFNHV**-NH_2



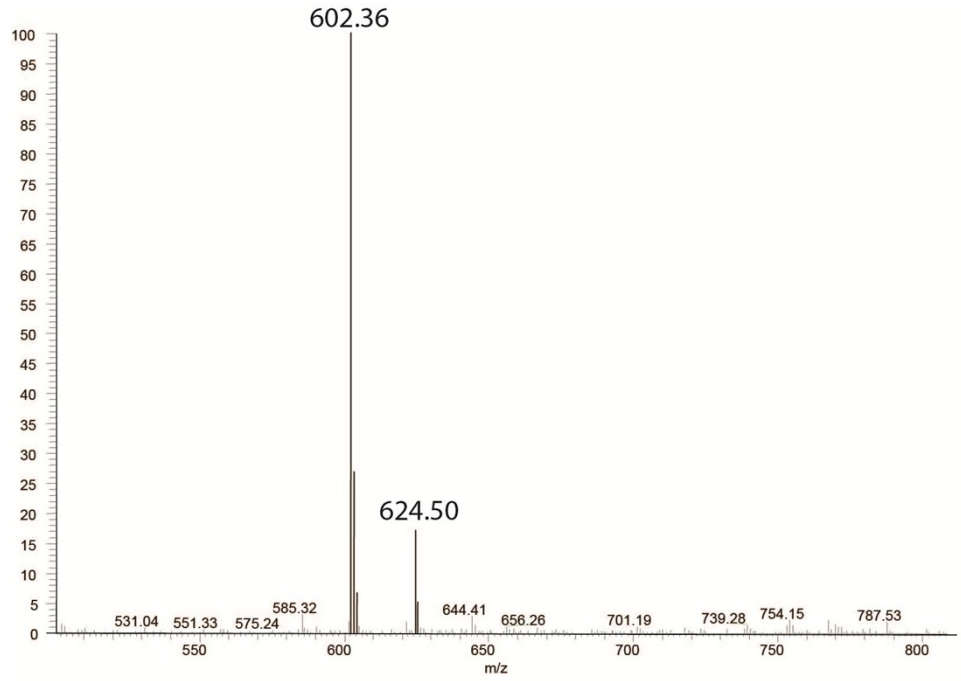
ESI-MS (positive) 678.36 (observed, M+H) 700.48 (observed, M + Na) 677.33 (calculated, M).

Peptide **4e** H-**Y**-**Nle**-**NHV**-NH_2



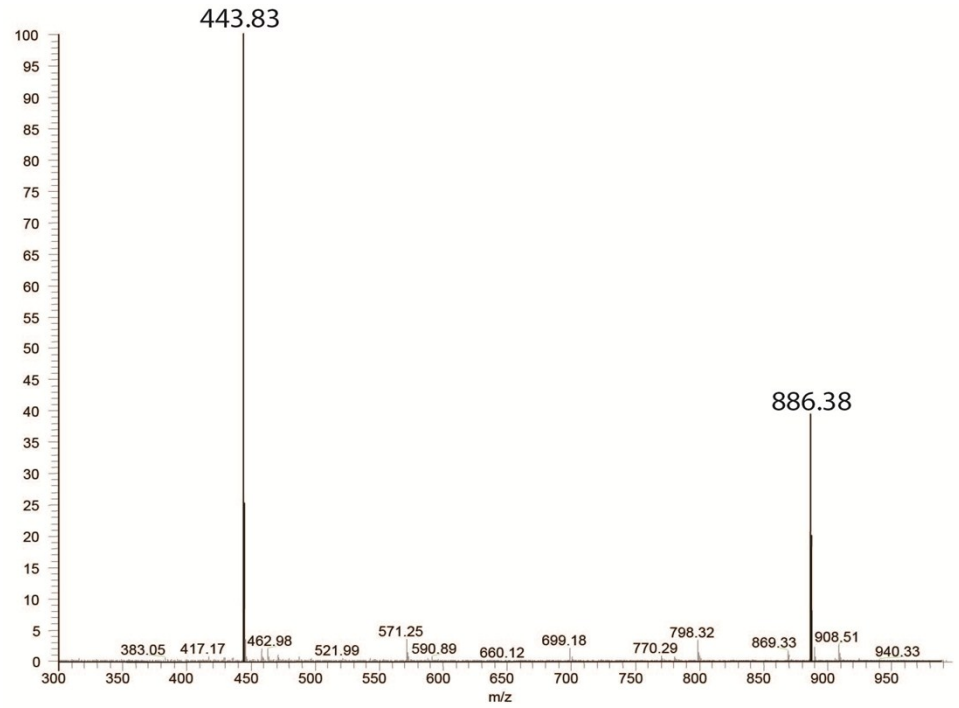
ESI-MS (positive) 644.36 (observed, M+H) 666.51 (observed, M + Na) 643.34 (calculated, M).

Peptide **4f** H-**Y**(D)**ANHV**-NH_2



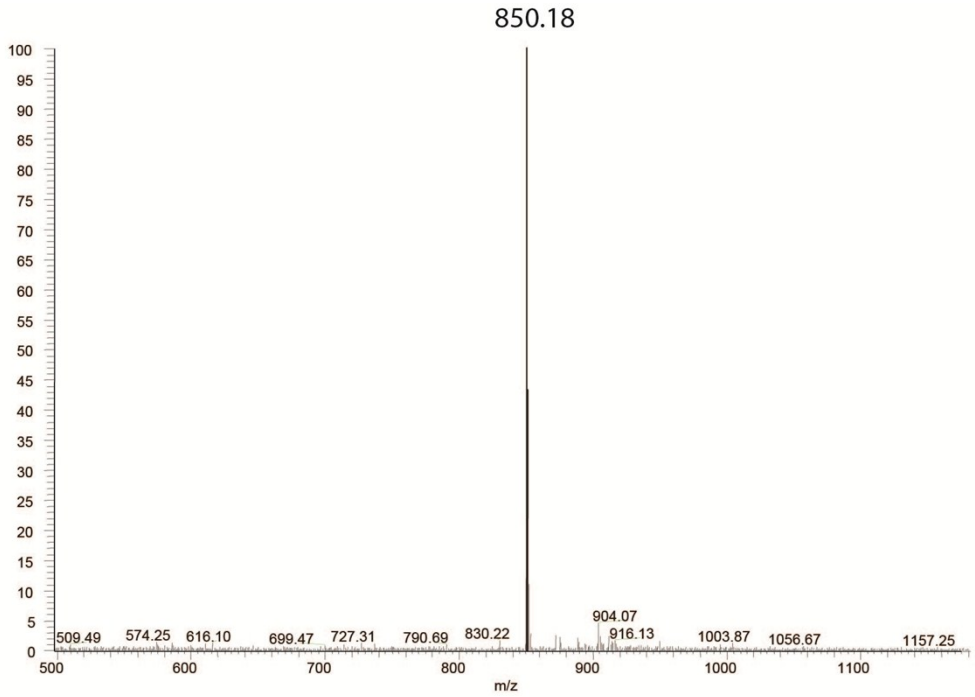
ESI-MS (positive) 602.36 (observed, M+H) 624.50 (observed, M + Na) 601.30 (calculated, M).

Peptide **12** H-**CK**(biotin)**LKVA**-NH_2



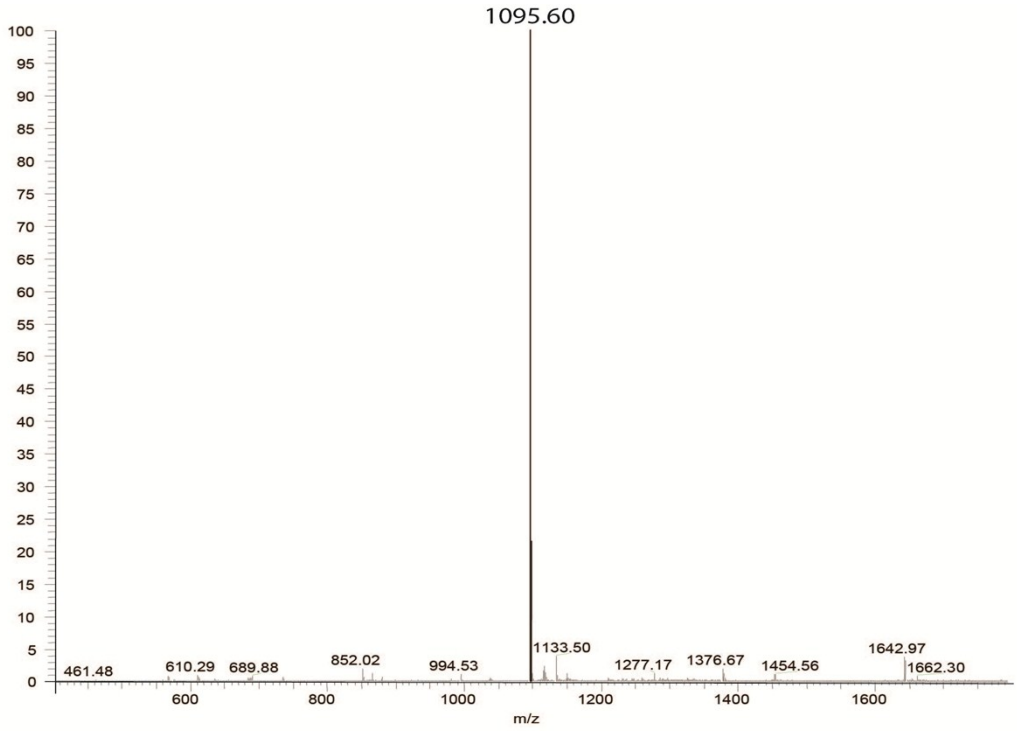
ESI-MS (positive) 886.38 (observed, M+H) 885.49 (calculated, M).

Depsipeptide **17a** Nle-YLPET-glc-G



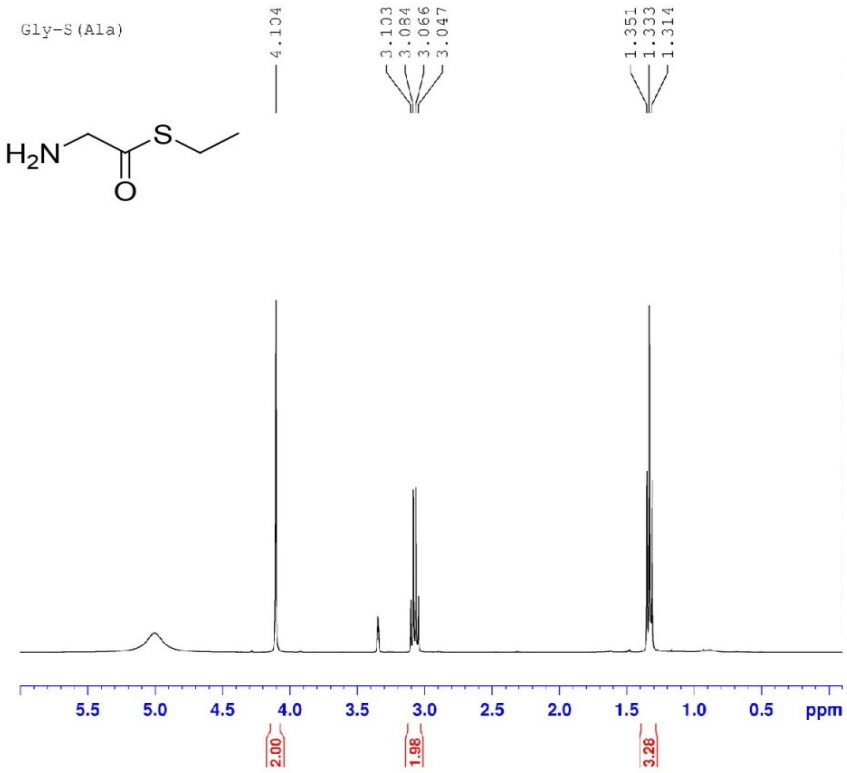
ESI-MS (positive) 850.18 (observed, M+H) 848.94 (calculated, M).

Depsipeptide **17b** fluor-YLPET-glc-G

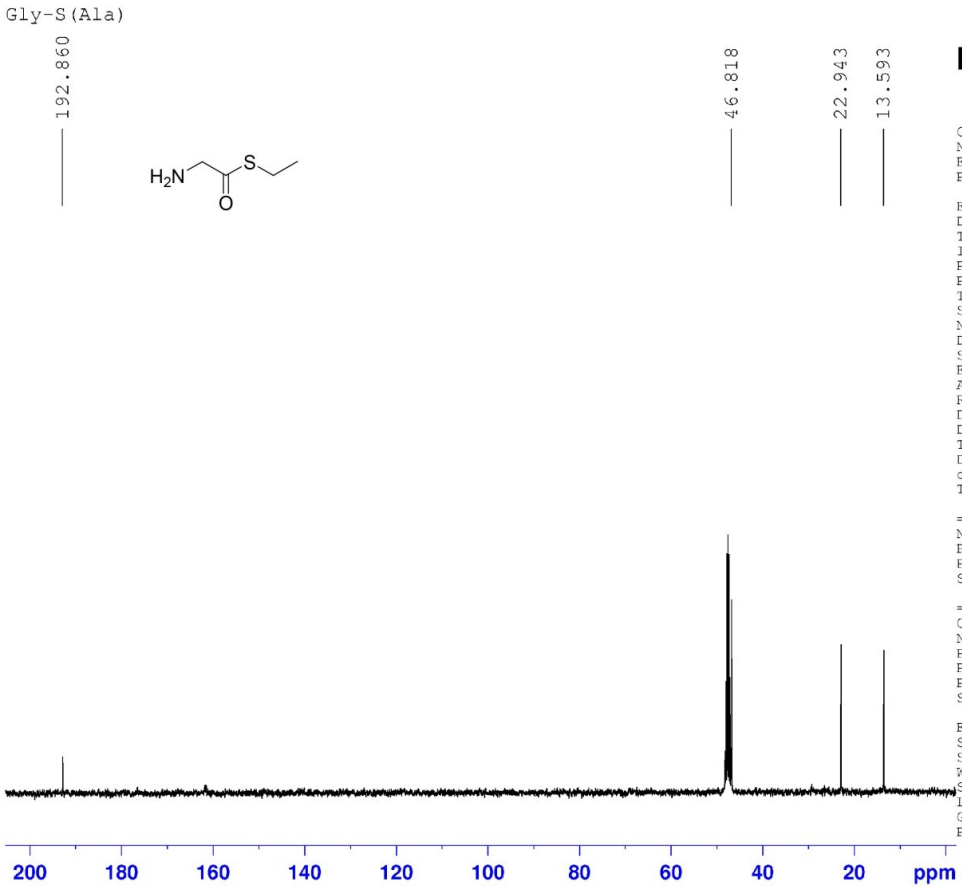


ESI-MS (positive) 1095.60 (observed, M+H) 1094.08 (calculated, M).

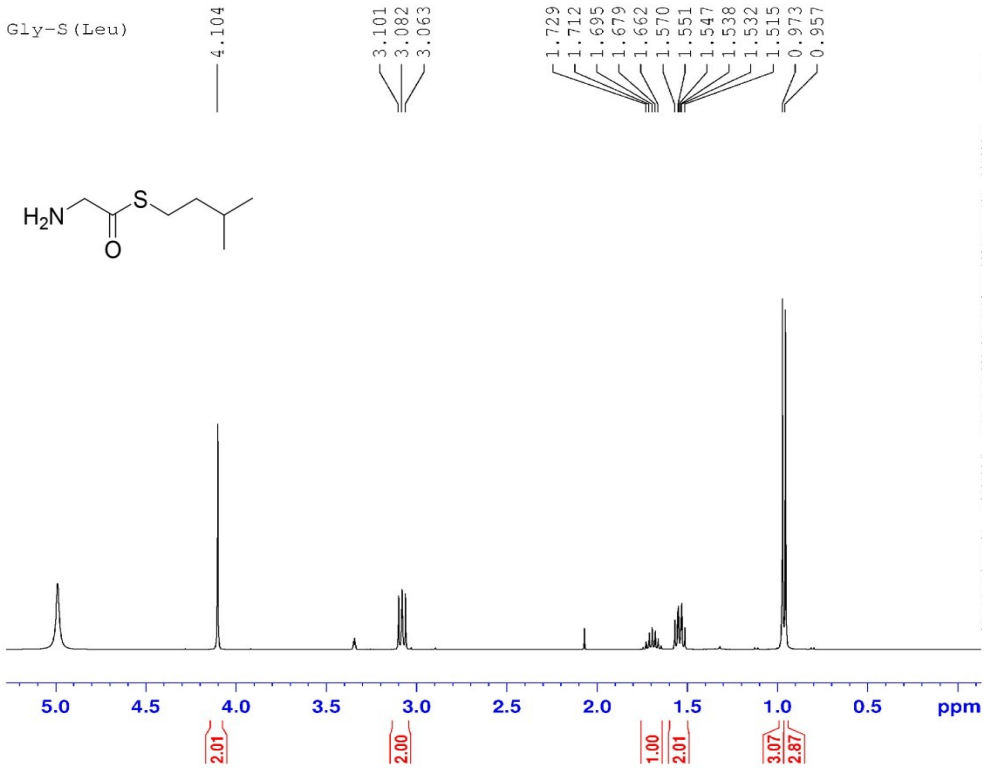
NMR spectra



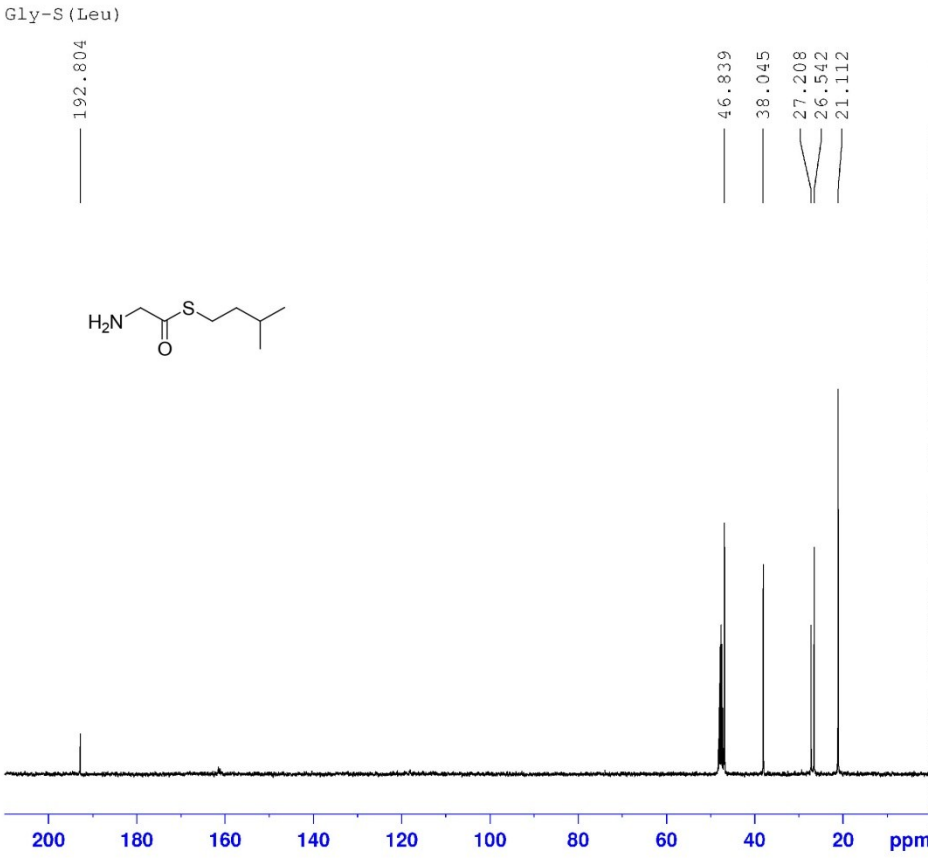
^{1}H NMR (400 MHz, $CDCl_3$): δ in ppm 4.104 (s, 2H, NH_2-C<u>H</u>$_2$), 3.103-3.047 (q, 2H, S-CH_2), 1.351-1.314 (t, 3H, CH_3)



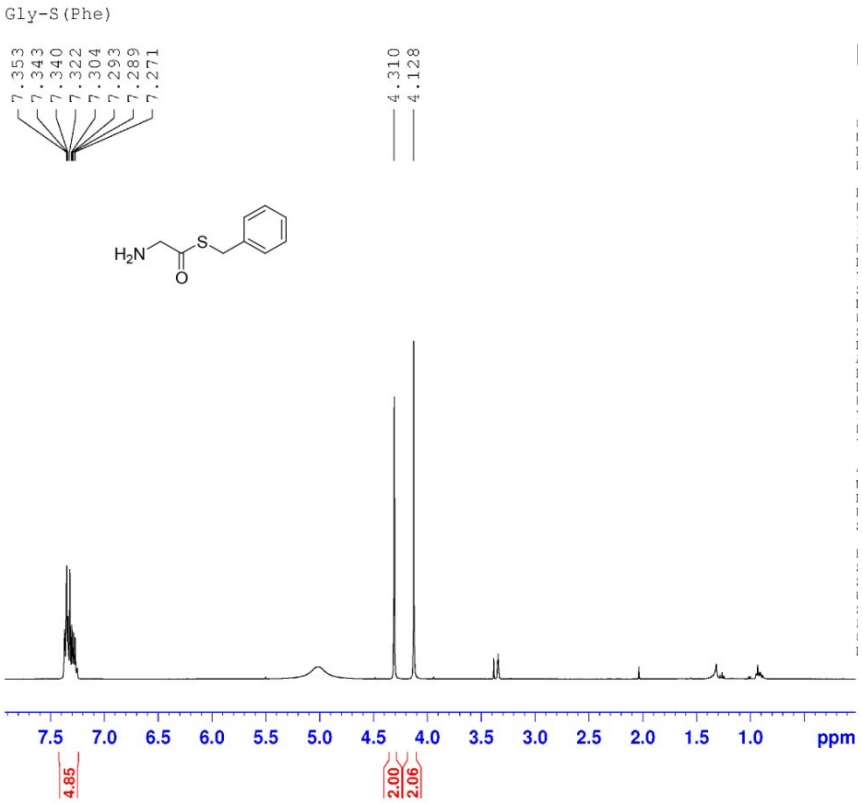
^{13}C NMR (400 MHz, $CDCl_3$): δ in ppm 192.860 (C=O), 46.818 (CH_2), 22.943(CH_2), 13.593 (CH_3)



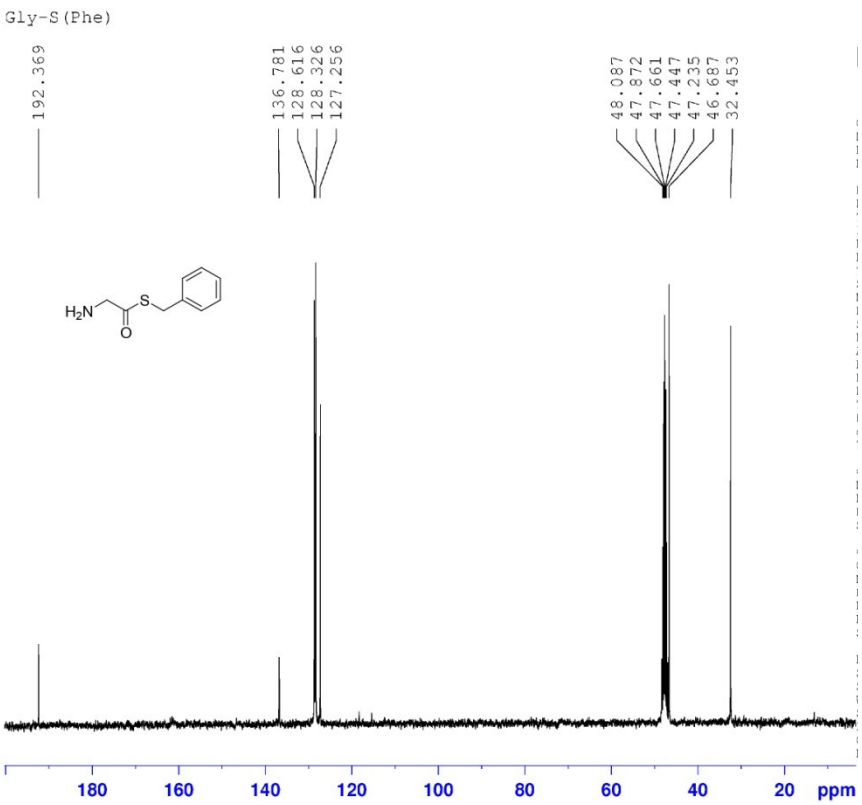
^{1}H NMR (400 MHz, $CDCl_3$): δ in ppm 4.104 (s, 2H, NH_2-C<u>H</u>$_2$), 3.101-3.063 (t, 2H, S-CH_2), 1.729-1.662 (m, 1H, CH), 1.570-1.515 (m, 2H, CH_2), 0.973 (s, 3H, CH_3), 0.957 (s, 3H, CH_3)



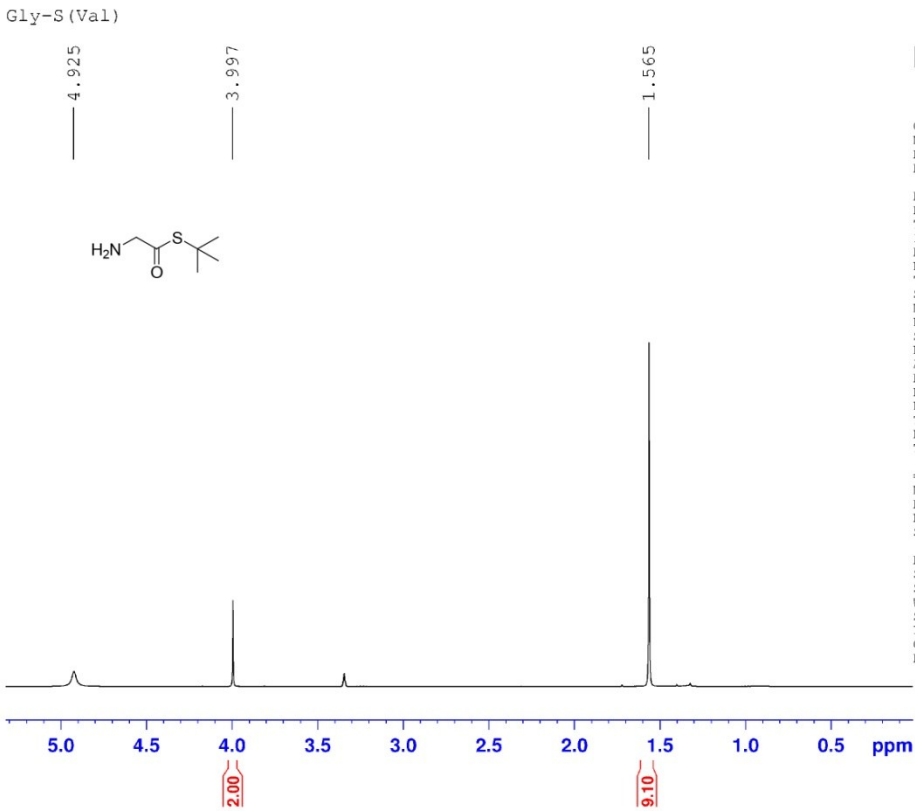
^{13}C NMR (400 MHz, $CDCl_3$): δ in ppm 192.804 (C=O), 46.839 (NH_2-CH_2), 38.045 (S-CH_2), 27.208 (CH_2), 26.542 (CH), 21.112 (CH_3)



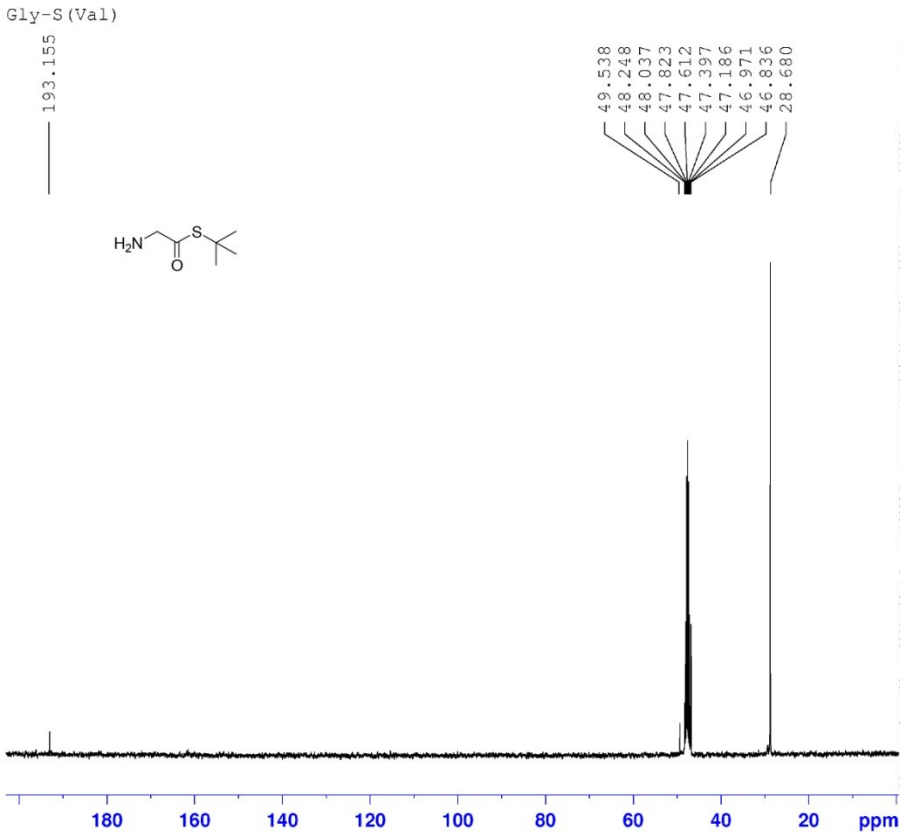
^{1}H NMR (400 MHz, $CDCl_3$): δ in ppm 7.353-7.271 (m, 5H, Ph), 4.310 (s, 2H, CH_2-Ph), 4.128 (s, 2H, NH_2-CH_2)



^{13}C NMR (400 MHz, $CDCl_3$): δ in ppm 192.369 (C=O), 136.781 (Ph), 128.161(Ph), 128.326 (Ph), 127.256 (Ph), 46.687 (S-CH_2), 32.453 (NH_2-CH_2)



^{1}H NMR (400 MHz, $CDCl_3$): δ in ppm 3.997 (s, 2H, CH_2), 1.565 (s, 9H, CH_3)



^{13}C NMR (400 MHz, $CDCl_3$): δ in ppm 193.155 (C=O), 28.680 (CH_3)