Electronic Supplementary Information

Mechanistic study of CBT-Cys click reaction and its application for identifying bioactive N-terminal cysteine peptides in amniotic fluid

Zhen Zheng,[#a] Peiyao Chen,[#a] Gongyu Li,[#a] Yunxia Zhu,[#c] Zhonghua Shi,[d] Yufeng Luo,[a] Chun Zhao,[b] Ziyi Fu,[b] Xianwei Cui,[d] Chenbo Ji,[d] Fuqiang Wang,*[b] Guangming Huang,*[a] and Gaolin Liang,*[a]

[a]CAS Key Laboratory of Soft Matter Chemistry, Department of Chemistry, University of Science and Technology of China, Hefei, Anhui 230026, China

[b]State Key Laboratory of Reproductive Medicine, Analysis Center, Nanjing Medical University, Nanjing, Jiangsu 210093, China

[c]Key Laboratory of Human Functional Genomics of Jiangsu Province, Jiangsu Diabetes Center, Nanjing Medical University, Nanjing, Jiangsu, 210093, China

[d]Nanjing Maternal and Child Health Institute, Nanjing Maternal and Child Health Care Hospital Affiliated to Nanjing Medical University, Nanjing, Jiangsu 210093, China

[#]These authors contributed equally to this work.

Correspondence and requests for materials should be addressed to e-mail: F. W. (wangfq@njmu.edu.cn), G. H. (gmhuang@ustc.edu.cn) and G. L. (gliang@ustc.edu.cn).

Table of Contents

1. Methods

General Methods

All the starting materials were obtained from Adamas, Sangon Biotech or Bachem. Commercially available reagents were used without further purification, unless noted otherwise. All chemicals were reagent grade or better. 2-cyano-6-aminobenzothiazole (CBT) was purchased from Shanghai Chemical Pharm-Intermediate Tech. Co.. ^{15}N-labeled cysteine was bought from Creative Peptides (USA). All solvents used in this study were of HPLC grade. Purified water (conductivity of 18.2 MΩ.cm) was obtained from Milli-Q® Reference System (Millipore Corp., USA). HPLC analyses were performed on a Shimadzu UFLC system equipped with two LC-20AP pumps and an SPD-20A UV-vis detector using a Shimadzu PRC-ODS column, or on an Agilent 1200 HPLC system equipped with a G1322A pump and an in-line diode array UV detector using an Agilent Zorbax 300SB-C_{18} RP column, with CH_3CN (0.1% of trifluoroacetic acid (TFA)) and ultrapure water (0.1% of TFA) as the eluent. ^{1}H NMR and ^{13}C NMR spectra were obtained on a Bruker AV-300 MHz spectrometer.

Experimental settings for nano-ESI

All tandem MS experiments were carried out on a LTQ-Velos Pro (Thermo Fisher Scientific, CA, USA) mass spectrometer and high-resolution MS experiments were performed on a Exactive Plus (Thermo Fisher Scientific, CA, USA) mass spectrometer. MS operation conditions are as follows (if not otherwise specified): nano-spray voltage 1.2 kV, max. ion injection time 30 milliseconds, and 3 micro-scans for each individual scan. Isolation width was set as 1.0. MS inlet temperature of 275 ℃. In typical experiments, the size of spray emitter was kept ~10 μm. The emitters were pulled from borosilicate glass capillaries (1.2 mm o.d., 0.9 mm i.d.) by using a P-2000 laser-based micropipette puller (Sutter Instruments, Novato, CA, USA).

Induced nano-electrospray ionization.

Induced nano-electrospray ionization (InESI) was initiated by a home-made power supply system, similar with previous reports. For iESI, the starting induced voltage was typically 4 kV_{p-p}, 1 kHz, duty cycle 50%, pulse voltage input. This pulsed high voltage was originally generated from a 25 MHz function waveform generator (DG1022U, RIGOL, Beijing, China) and was then amplified by a household stereo power amplifier. As a result, the generated 100 V_{p-p} pulsed voltage was followed by a coil to further amplify the amplitude up to a final value of 0-20 kV_{p-p}. The high voltage was

monitored with a digital oscilloscope (DS1052E, RIGOL, Beijing, China) after attenuating (1: 1000) with an attenuation rod. Other operation parameters were kept the same with nanoESI.

Recovery test

The fitted calibration curves between corresponding HPLC peak area and the concentration of peptide **2** or product **3** were shown in Figure S15A and Figure S15B, respectively. Compound **1**-conjugated resin was then applied to fish peptide **2** at 10 nmol or 1 nmol. After an overnight reaction at room temperature, the remaining peptide **2** in the solution and the product **3** conjugated to the resin were collected for HPLC analysis. Their concentrations were determined with their corresponding calibration curves. The recovery rates were calculated according to the equation below.[1]

$$\text{Recovery \%} = (\text{sample value})/(\text{theoretical certified value}) \times 100$$

Amniotic fluid sample preparation

Amniotic fluid (AF) samples were obtained from pregnant women at 22^{nd} – 24^{th} week gestation who underwent amniocentesis at the second trimester of pregnancy in Nanjing Maternity and Child Health Care Hospital (Nanjing, Jiangsu, China). Samples were collected with written consent and ethics board approval. AF samples fell into pregnant women (n = 30). Samples were matched for gestational age and maternal age.

The AF was collected, stored on ice during transport to the laboratory (up to 1 h), and then centrifuged (1,000 g, 10 min, 4 ℃) to remove the cellular debris pellet. The supernatant was then added aliquot of protease inhibitor mixture (Complete Mini EDTA-free, Roche, Basel, Switzerland) and stored at -80 ℃.Samples were centrifuged at 12,000 g at 4 ℃ for 15 min after thawing on the ice, and the supernatant was collected. Then the samples were added with acetonitrile (20%, v/v), briefly vortexed and incubated for 20 minutes at room temperature. Molecular weight cut-off (MWCO) filters of 10 kDa were washed with 0.5 mL H_2O prior to use. The AF samples where centrifuged through the filters according to the manufacturer's recommendations. The flow-through from the filters containing the peptide fraction was recovered and lyophilized.

Thereafter, the supernatant was collected and the peptide concentration was measured by the bicinchonininc acid (BCA) protein assay reagent from Pierce (Pierce, Rockford, IL, USA). The supernatant was collected and vacuum-dried using a SpeedVac system (RVC 2−18, Marin Christ,

Osterod, Germany) as a neuropeptide sample and was stored at −80 °C for further analysis.

MALDI-TOF MS analyses

All MALDI-TOF MS analyses were performed on a Bruker Daltonics UltrafleXtreme MALDI TOF/TOF mass spectrometer. All mass spectra reported were obtained in the reflex positive ion mode with delayed ion extraction. Spectral data were obtained in the reflector mode with an extraction voltage of 20 kV. The laser power was adjusted to a value slightly above the detection threshold, and each mass spectrum was obtained from 100 sum scans across the sample surface. Sample aliquots (0.5 μL) and a-cyano-4-hydroxycinnamic acid (CHCA) matrix (0.5 μL) were spotted on the MALDI plate sequentially and dried at room temperature prior to MALDI-TOF MS analysis.

Peptide identification

Each acquired mass spectrum (*m*/*z* range 800 - 3,000) was processed using the software FlexAnalysis software (version 3.2, Bruker Daltonics, Bremen, Germany). The peak detection algorithm was Sort Neaten Assign and Place (SNAP); signal to noise (S/N) threshold, 3; quality factor threshold, 50. The standard MS/MS spectra were analyzed by Biotools software (version 3.2, Bruker Daltonics, Bremen, Germany).

The AF peptides MS/MS spectra were searched against database using the composite SWISS-PROT database containing 20,194 protein sequences of human (released in May 2015) using in-house MASCOT search program (http://www.matrixscience.com/) software (version 1.2, Matrix Science Inc, Boston, USA). The following fixed modifications were applied: N-terminal cysteine modifications (b_1 fragment ion of 363.034). The other parameters used were the following: parent ion mass tolerance, 15.0 ppm; fragment ion mass tolerance, 0.05 Da; enzyme, none; and maximum allowed variable post translational modification (PTM) per peptide. A fusion target-decoy approach was used for the estimation of the false discovery rate (FDR) and controlled at ≤ 1.0% ($-10 \log P \geq 20.0$) at peptide levels.

Bioinformatics analysis

Each peptide pI was calculated by on line pI/Mw tool (http://web.expasy.org/compute_pi/) for further research. In order to categorize the identified peptides, pathway and gene ontology (GO) analysis were carried out so as to uncover the potential roles that the peptide protein precursors played. The ontology contains three aspects: cellular component, biological process and molecular

function. The threshold of significance was defined by the P-value and FDR. Accurate sequence-based prediction of secondary structures and transmembrane domains for proteins of unknown structure was done as described (http://sable.cchmc.org/, http://polyview.cchmc.org/polyview_doc.html#VS_ShownInfo).

Cell culture and peptide syntheses

HEC-1-B cells and AA-CELL-24 cells were purchased from American Type Culture Collection and maintained in DMEM/F12 medium containing 10% fetal bovine serum (FBS) and 5% penicillin-streptomycin-amphotericin (Invitrogen, USA). All cells were incubated at 37 ℃ and 5% CO_2. The following peptides were synthesized by ChinaPeptides Co., Ltd (Shanghai, China).

First, we synthesized successive short peptides of N-terminal cysteine peptides, transactivator of transcription (TAT, YGRKKRRQRRR), and nuclear localization signal (NLS, RKRRK) protein to form a fusion peptide. The sequences of the fusion peptides are TAT-NLS-CGSQCTCQ-NH_2 (MUC19_HUMAN), TAT-NLS-CKTCNGSATLCTSCPK-NH_2 (PCSK5_HUAMN), TAT-NLS-CQHNTCGGTCDRC-NH_2 (LAMA5_HUMAN), TAT-NLS-CFLHLP-NH_2 (K0100_HUMAN), TAT-NLS-CHQMSFCL-NH_2 (FRITZ_HUMAN), TAT-NLS-CILLLEGGP-NH_2 (ZSWM5_HUMAN), TAT-NLS-CNGTKGERGPLGPPG-NH_2 (CO4A1_HUMAN) and TAT-NLS-CVVPIC-NH_2 (FBN3_HUMAN).

N-terminal cysteine peptides were derived from CBT-Fishing methods. TAT is a cell-penetrating peptide from the human immunodeficiency virus, and it can deliver proteins, DNA, RNA, and nanoparticles into the cytoplasm in a short time with extremely high efficiency.[2,3] However, because stabilized N-terminal cysteine peptides translocate into the nucleus, we synthesized TAT-NLS fusion peptides to analyze their effects on the nuclei of endometrial cancer cells.

Cell growth

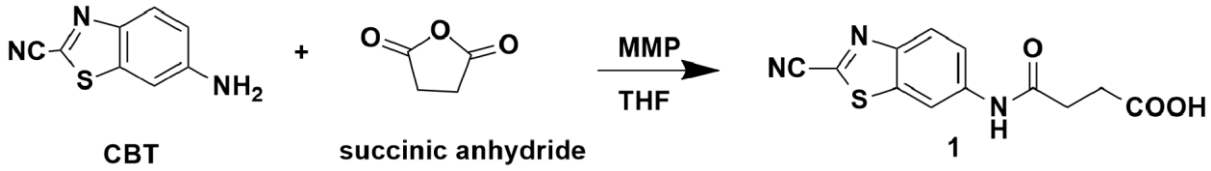
Cell growth was analyzed using 2-(2-methoxy-4-nitrophenyl)-3-(4-nitrophenyl)-5-(2, 4-disulfophenyl)-2H-tetrazolium, monosodium salt, and the cell counting kit-8 (CCK-8, Sigma). HEC-1-B cells and AA-CELL-24 cells were seeded in 96-well plates and incubated with above eight peptides. After culturing for another 48 h, cell growth was analyzed by CCK-8 and the optical density was detected at 450 nm. The growth ratio was normalized to the cells without treatment.

Cell cycle analysis

HEC-1-B and AA-CELL-24 cells were plated in 6-well plates for 24 h, and the medium was replaced with fresh culture medium containing 80 μM peptide. After incubation for 24 h, the cells were harvested by trypsinization and then fixed with 70% ethanol at 4 ℃. Intracellular DNA of the cells was stained with 50 ng/mL propidium iodide in the dark for 30 min at room temperature, and the percentages of sub-G1 cells were determined by flow cytometry (BD LSRII analyzer, BD Biosciences).

2. Syntheses and characterizations of compound 1 and peptide 2

Scheme S1. Synthetic route for compound **1**.



*Synthesis of compound 4-((2-cyanobenzo[d]thiazol-6-yl)amino)-4-oxobutanoic acid (**1**)*:

The 2-cyano-6-aminobenzothiazole (CBT, 35 mg, 0.2 mmol) was added to a mixture of succinic anhydride (80 mg, 0.8 mmol) and 4-methylmorpholine (MMP, 101 mg, 1.0 mmol) in THF (8.0 mL) at 0 ℃ and the reaction mixture was stirred for 2 h then overnight at room temperature. The pure product **1** was obtained after HPLC purification. HPLC trace: one peak (Figure S1). MS: calc. M^+ = 276.04, obsvd. ESI-MS: *m*/*z* 276.06. ^{1}H NMR of compound **1** (d_6-DMSO, 300 MHz, Figure S2) δ (ppm): 10.51 (s, 1 H), 8.74 (s, 1 H), 8.17 (d, *J* = 9.0 Hz, 1 H), 7.71 (d, *J* = 9.0 Hz, 1 H), 2.66 (t, *J* = 6.0 Hz, 2 H), 2.58 (t, *J* = 5.9 Hz, 2 H). ^{13}C NMR of compound **1** (d_6-DMSO, 75 MHz, Figure S3) δ (ppm): 173.8, 170.8, 147.4, 139.7, 136.7, 134.7, 124.7, 120.4, 113.6, 110.8, 31.13, 28.62.

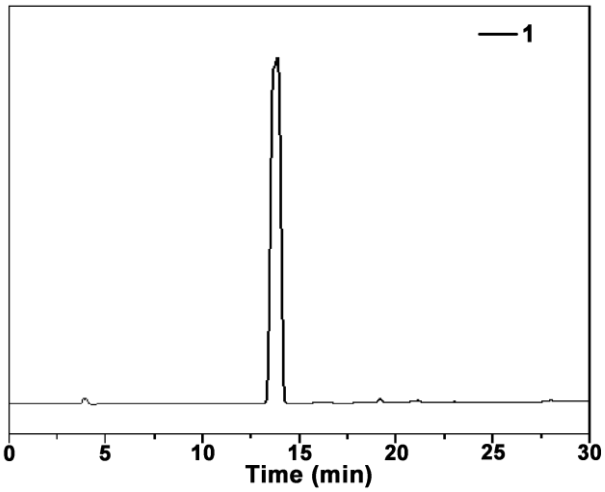


Figure S1. HPLC trace of compound **1**.

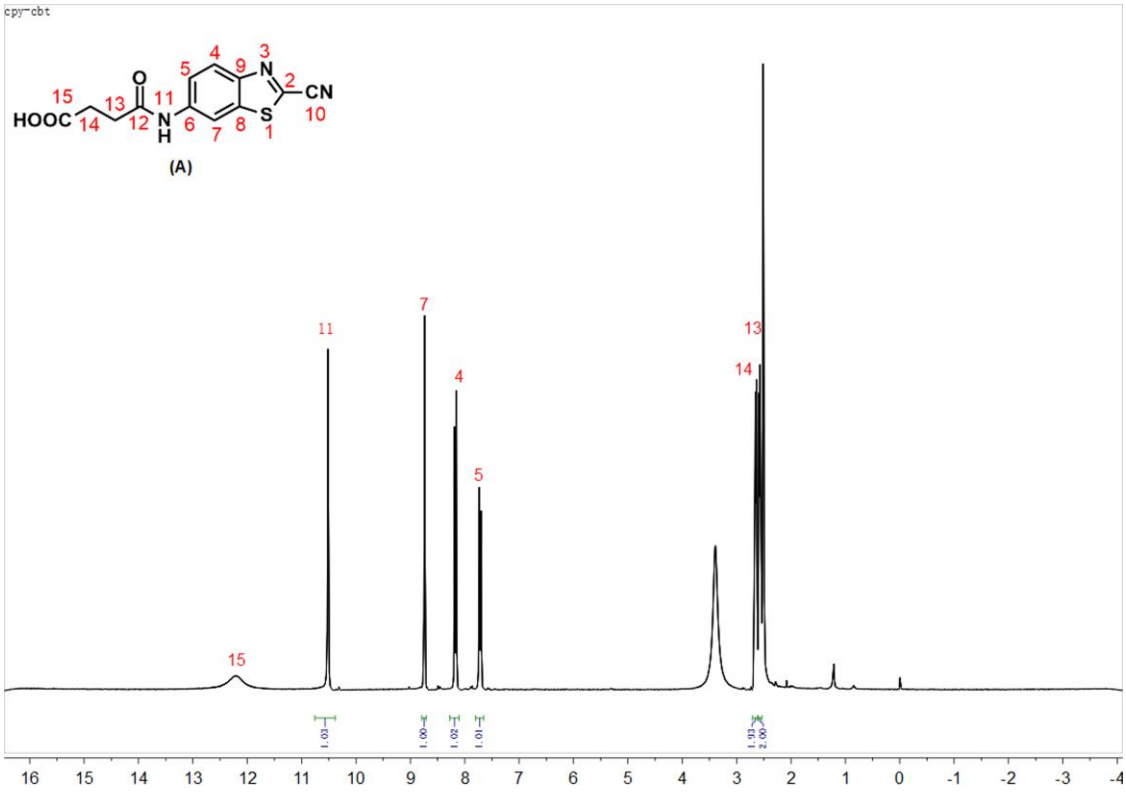


Figure S2. [1]H NMR spectrum of compound **1**.

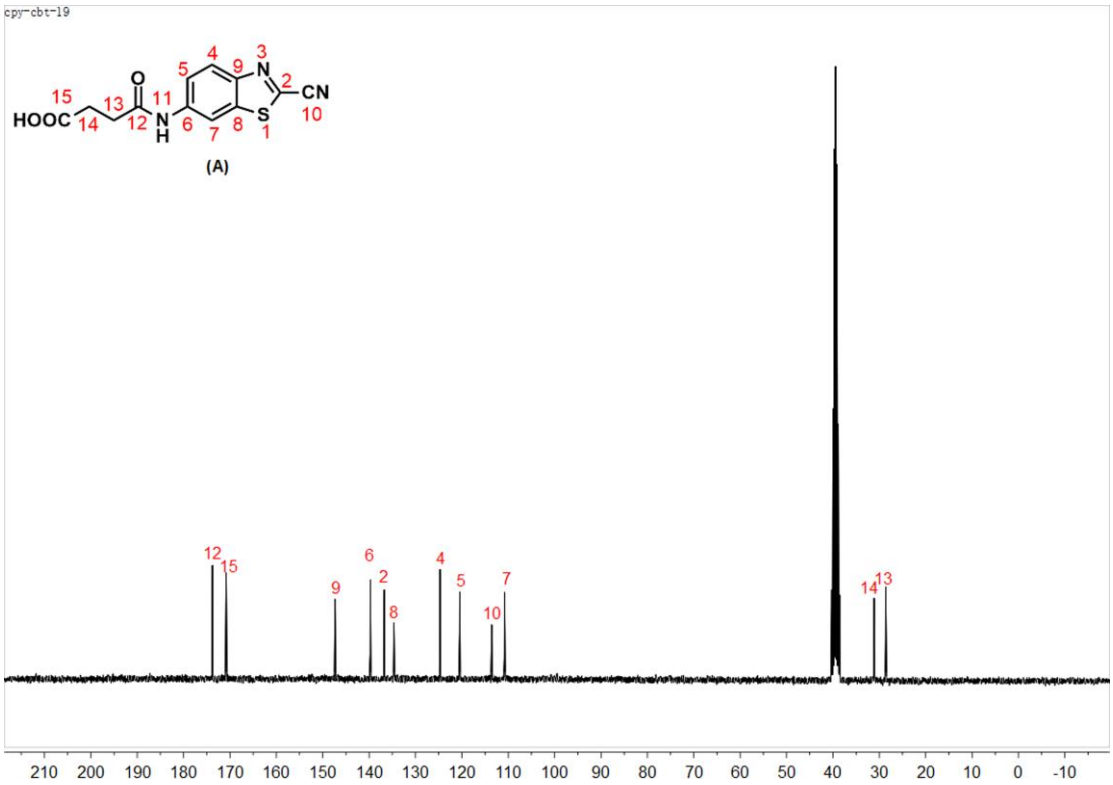
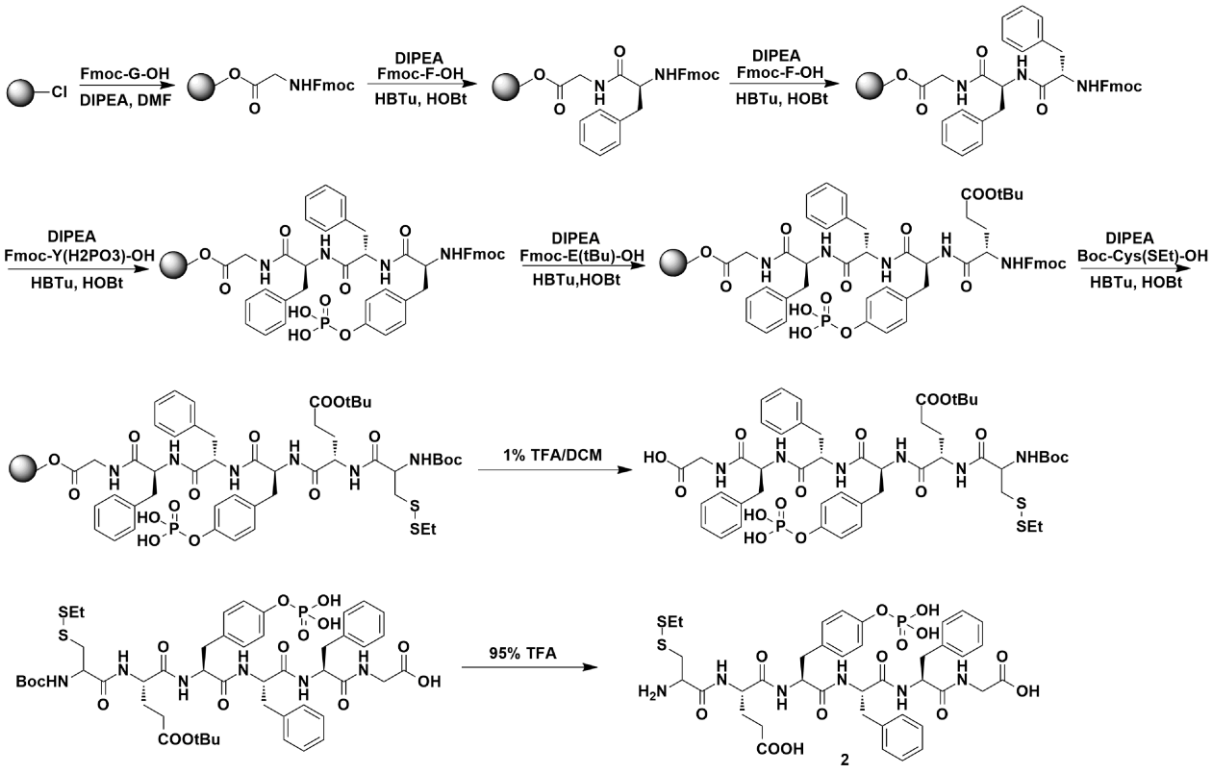


Figure S3. [13]CNMR spectrum of compound **1**.

Scheme S2. Synthetic route for peptide **2**.



*Synthesis of peptide Cys(Set)-Glu-Tyr(H_2PO_3)-Phe-Phe-Gly (**C(SEt)EYpFFG**, **2**)*:

Peptide Cys(Set)-Glu-Tyr(H_2PO_3)-Phe-Phe-Gly with protecting groups was synthesized with solid phase peptide synthesis (SPPS). The tBu protecting groups were then removed with dichloromethane (DCM, 0.5 mL) and triisopropylsilane (TIPS, 40 μL) in TFA (9.5 mL) for 3 h. Peptide **2** was obtained after HPLC purification using water-acetonitrile added with 0.1% TFA as the eluent (from 7:3 to 1:9). HPLC trace: one peak (Figure S4). UV-vis spectrum (Figure S5). High resolution mass (HR-MS): calculated for $C_{39}H_{50}N_6O_{13}PS_2$ $[(M+H)^+]$ = 905.26149, obsvd. TOF/MS: *m*/*z* 905.25905.

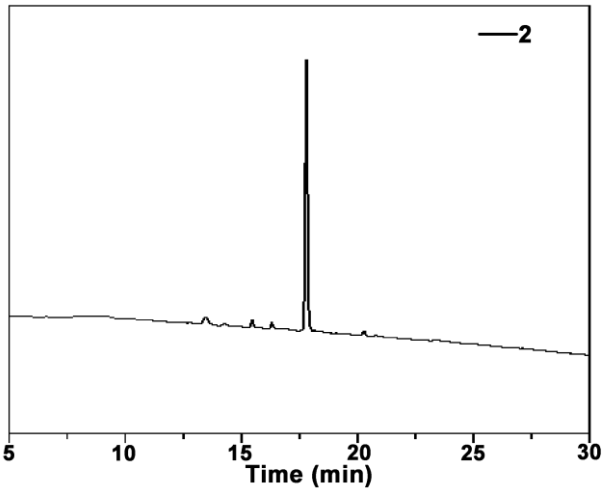


Figure S4. HPLC trace of peptide **2**.

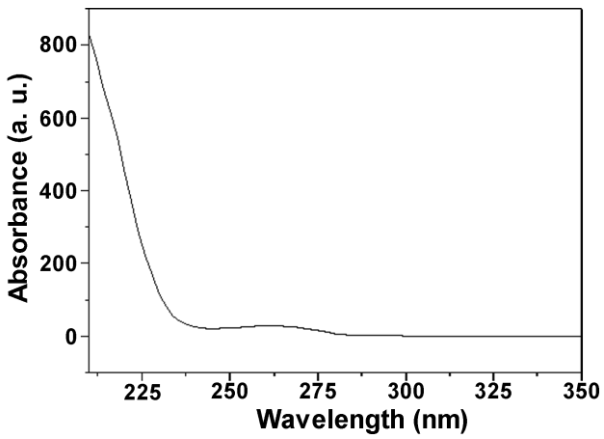


Figure S5. UV-vis spectrum of **2**.

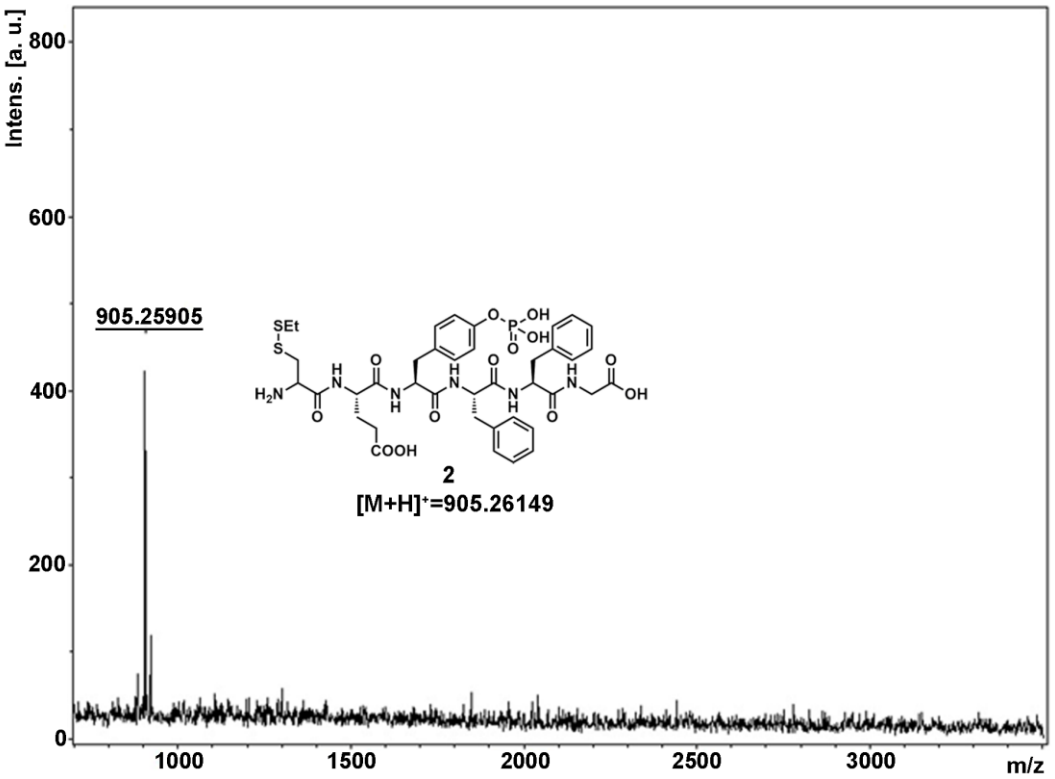
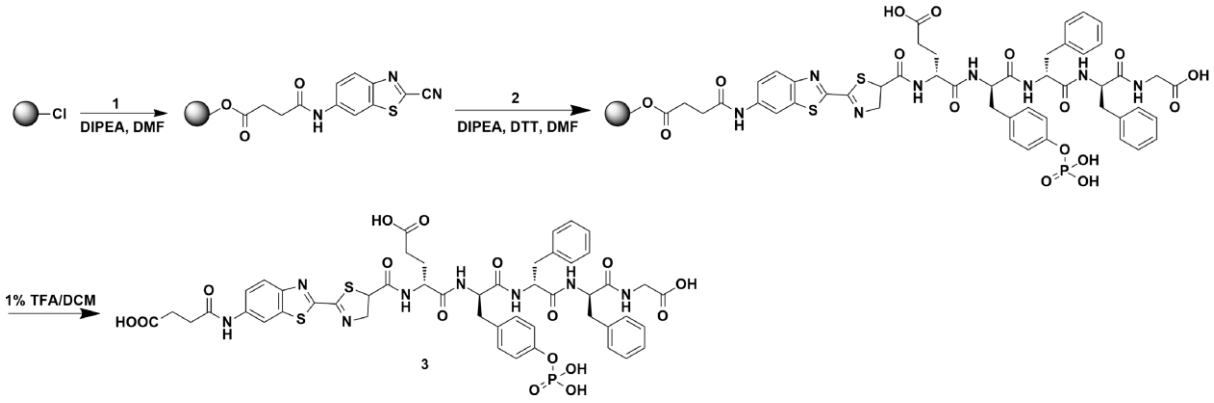


Figure S6. HR-MALDI-TOF MS spectrum of **2**.

Scheme S3. Schematic illustration of the reaction between **1** and **2** to yield **3**.



Compound **1** (137.5 µg, 500 nmol) was conjugated to 5 mg 2-chlorotrityl chloride resin pre-activated with N,N-dimethylformamide (DMF). Then, different amount of **2** (from 100 amol to 10 nmol), together with 20 equiv. of DL-Dithiothreitol (DTT), in DMF was shaken with the resin overnight at room temperature. After the reaction completed, the resin was thoroughly washed with DMF and product **3** was cleaved from the resin using 1% TFA in DCM for matrix-assisted laser desorption ionization/time of flight mass spectrometry (MALDI/TOF MS) analyses.

3. Supporting figures and tables

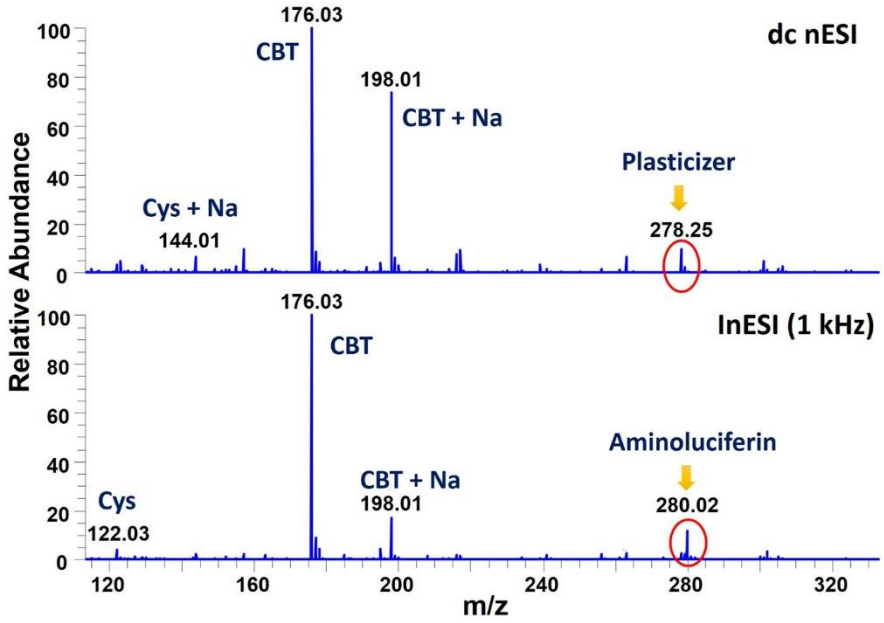


Figure S7. Representative mass spectra (positive mode) of CBT-Cys reaction mixture using conventional nESI (top) and Induced-nESI (bottom).

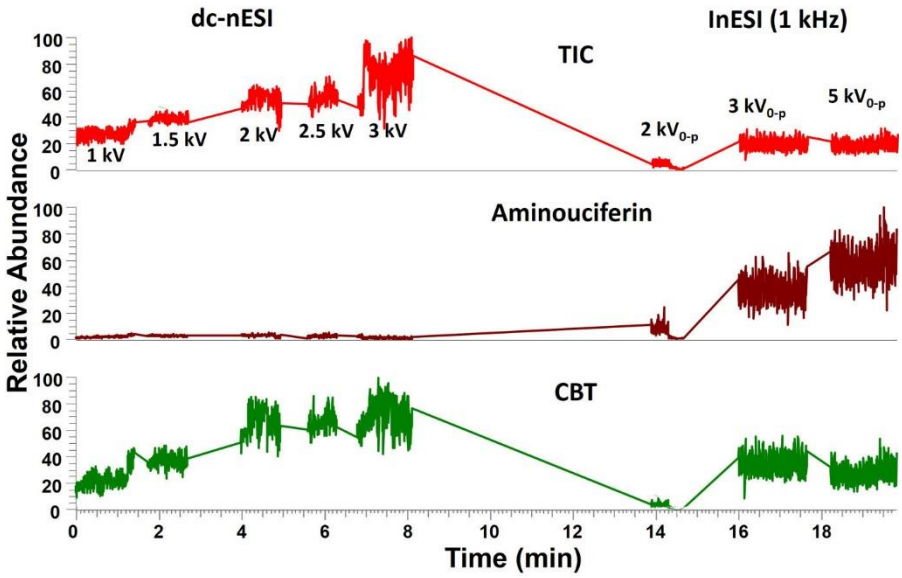


Figure S8. Ion chromatography comparison between conventional nESI and InESI for product **Aminoluciferin** (middle) and reactant CBT (bottom). In conventional nESI, although the reactant CBT is readily obseved, no significant product **Aminoluciferin** peak could be observed even by adjusting the dc voltage from 1.0 kV to 3.0 kV. However, during vlotage-varied InESI operations, both reactant CBT and product **Aminoluciferin** could be readily observed.

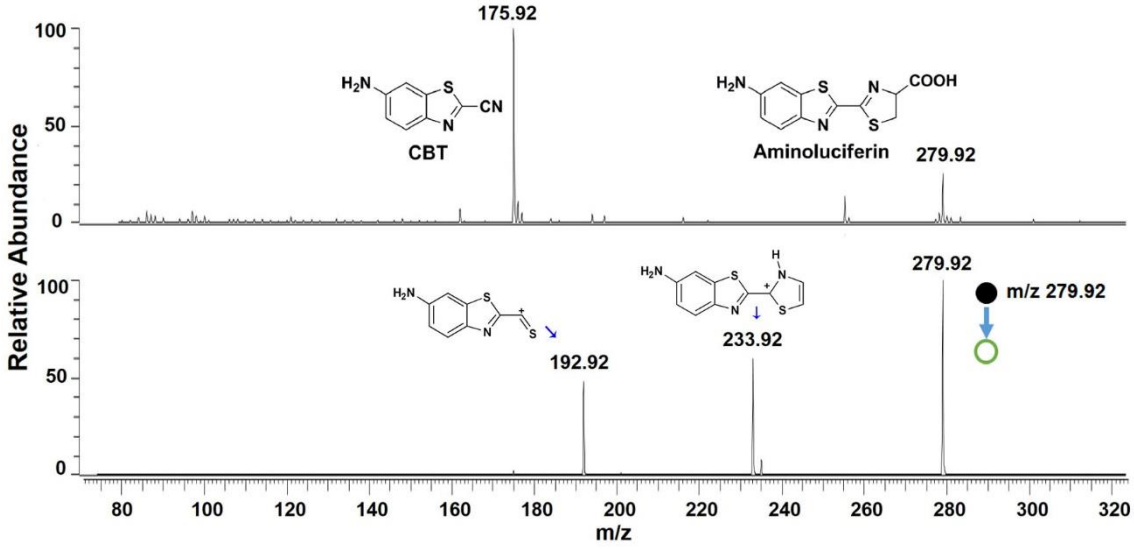


Figure S9. InESI (+)-MS (top) for the CBT-Cys click reaction and InESI (+)-MS/MS (bottom) for the identification of the product **Aminoluciferin**. Reaction conditions: 50 μM CBT with 50 μM Cys under pH 7.4, CH_3OH : H_2O = 1 : 1.

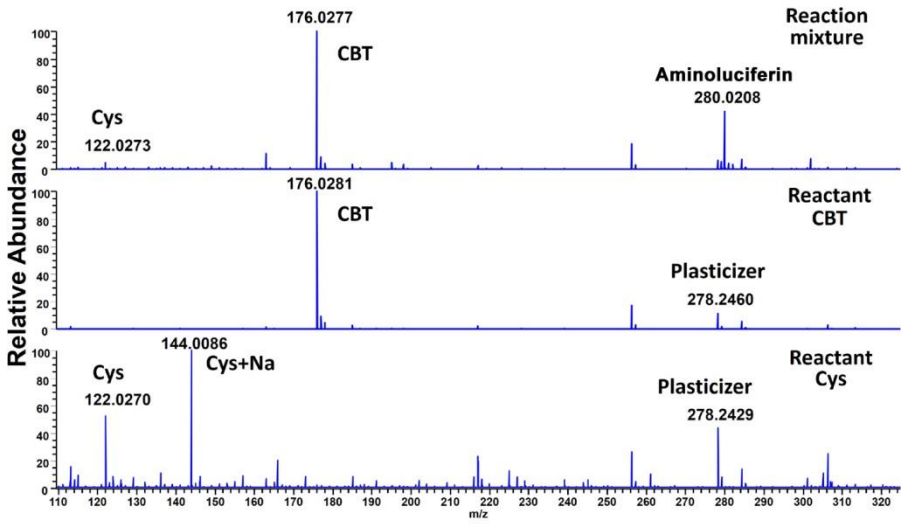


Figure S10. High resolution InESI-MS of the reaction mixture (top), reactant CBT (middle) and reactant Cys (bottom) for the identification of the product **Aminoluciferin**. Reaction conditions: 50 μM CABT with 50 μM Cys under pH 7.4, CH_3OH : H_2O = 1 : 1.

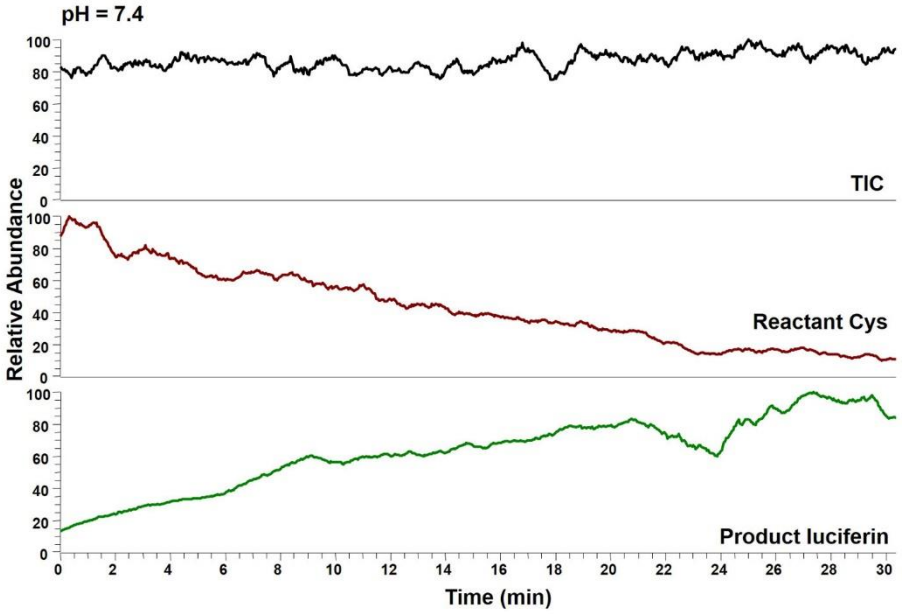


Figure S11. Selected ion chronograms of reactant Cys (red curve, *m*/*z* 122) and the product **Aminoluciferin** (green curve, *m*/*z* 280) for continuously monitoring of the click condensation reaction between CBT and Cys at pH 7.4.

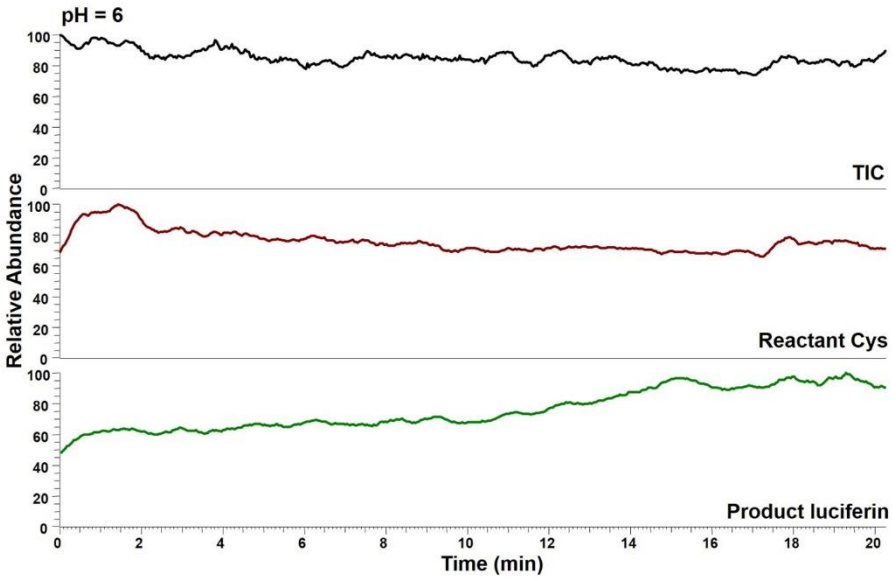


Figure S12. Selected ion chronograms of reactant Cys (red curve, *m*/*z* 122) and the product **Aminoluciferin** (green curve, *m*/*z* 280) for continuously monitoring of the condensation reaction between CBT and Cys at pH 6.0.

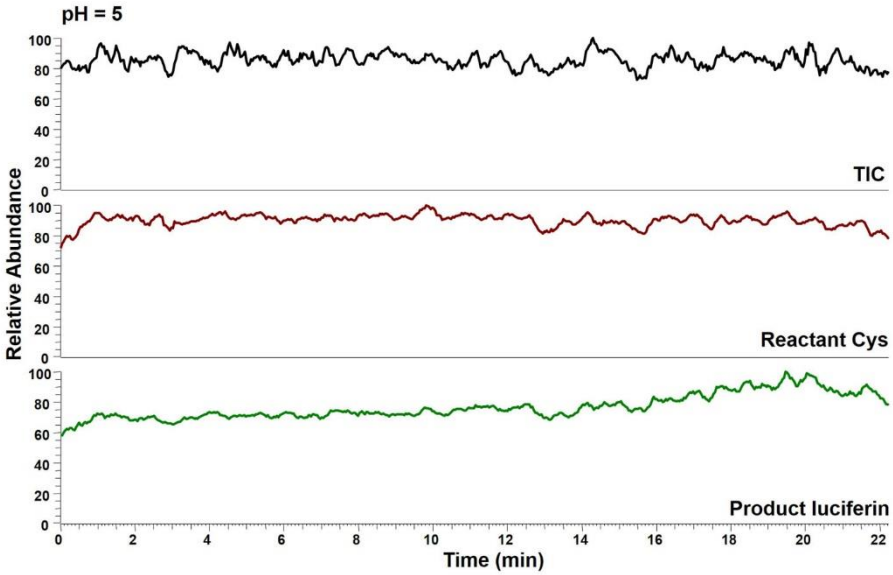


Figure S13. Selected ion chronograms of reactant Cys (red curve, *m*/*z* 122) and the product **Aminoluciferin** (green curve, *m*/*z* 280) for continuously monitoring of the condensation reaction between CBT and Cys at pH 5.0.

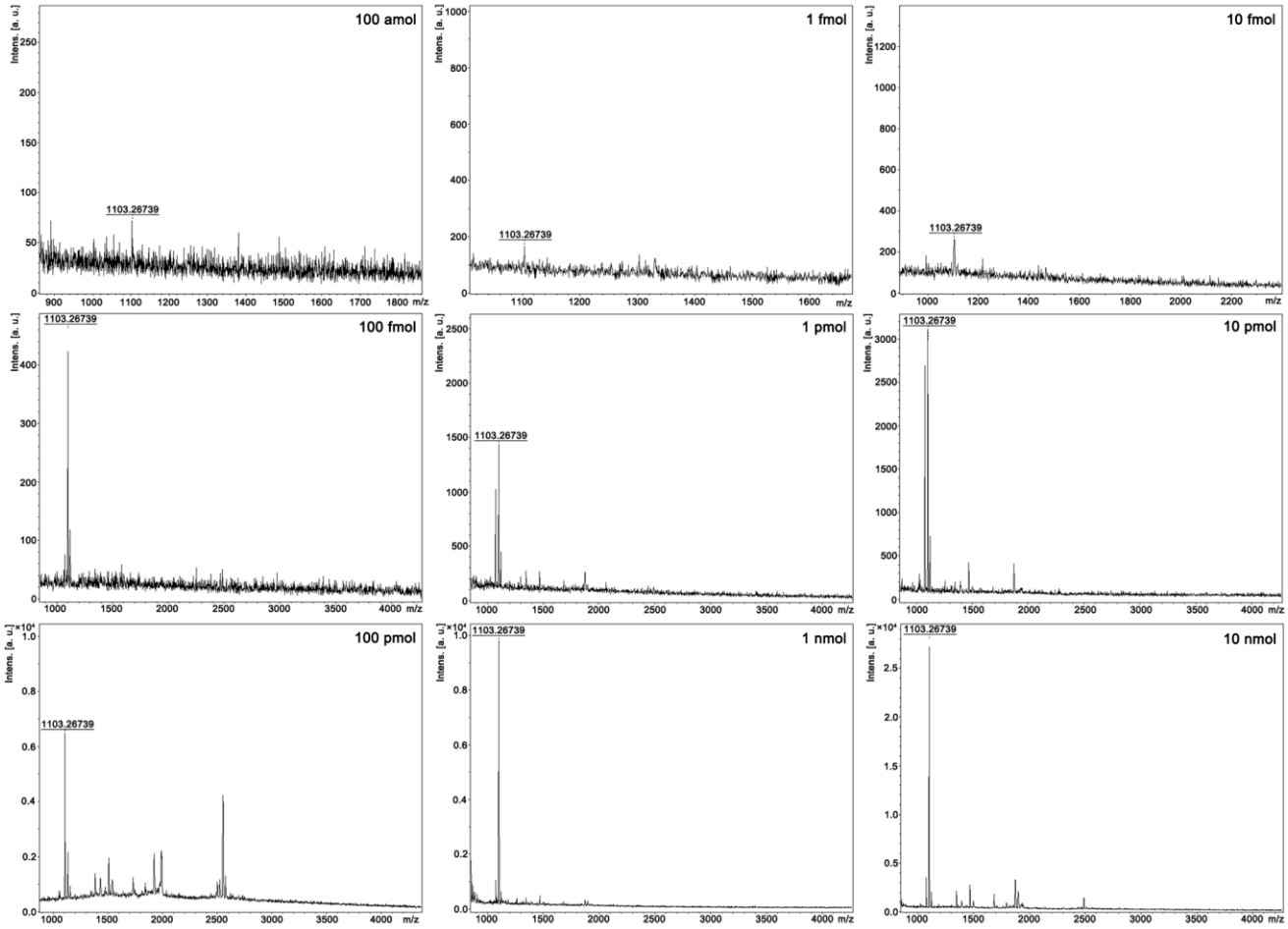


Figure S14. MALDI-TOF MS spectra of the product **3** obtained from the reactions that 500 nmol **1** reacted with peptide **2** at 100 amol to 10 nmol.

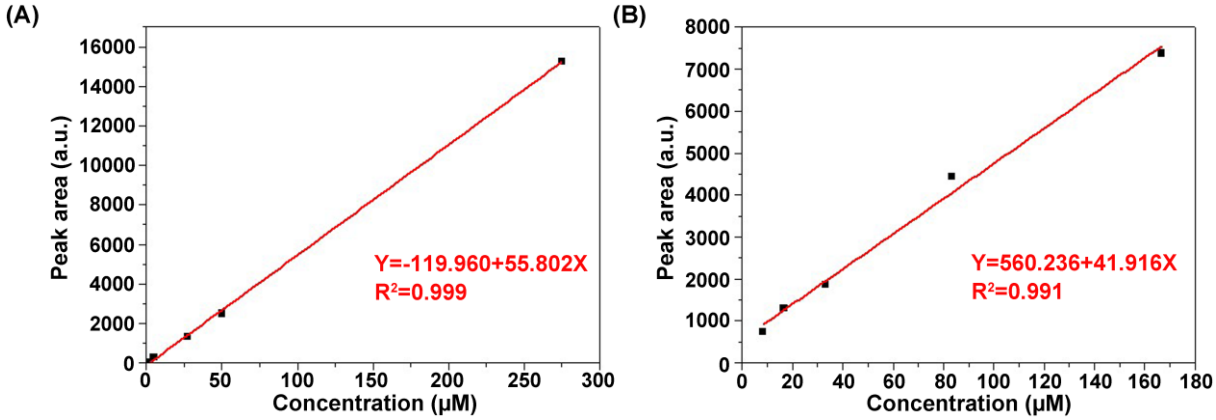


Figure S15. The fitted calibration curves between the HPLC peak area and the concentration of peptide **2** (A) or product **3** (B).

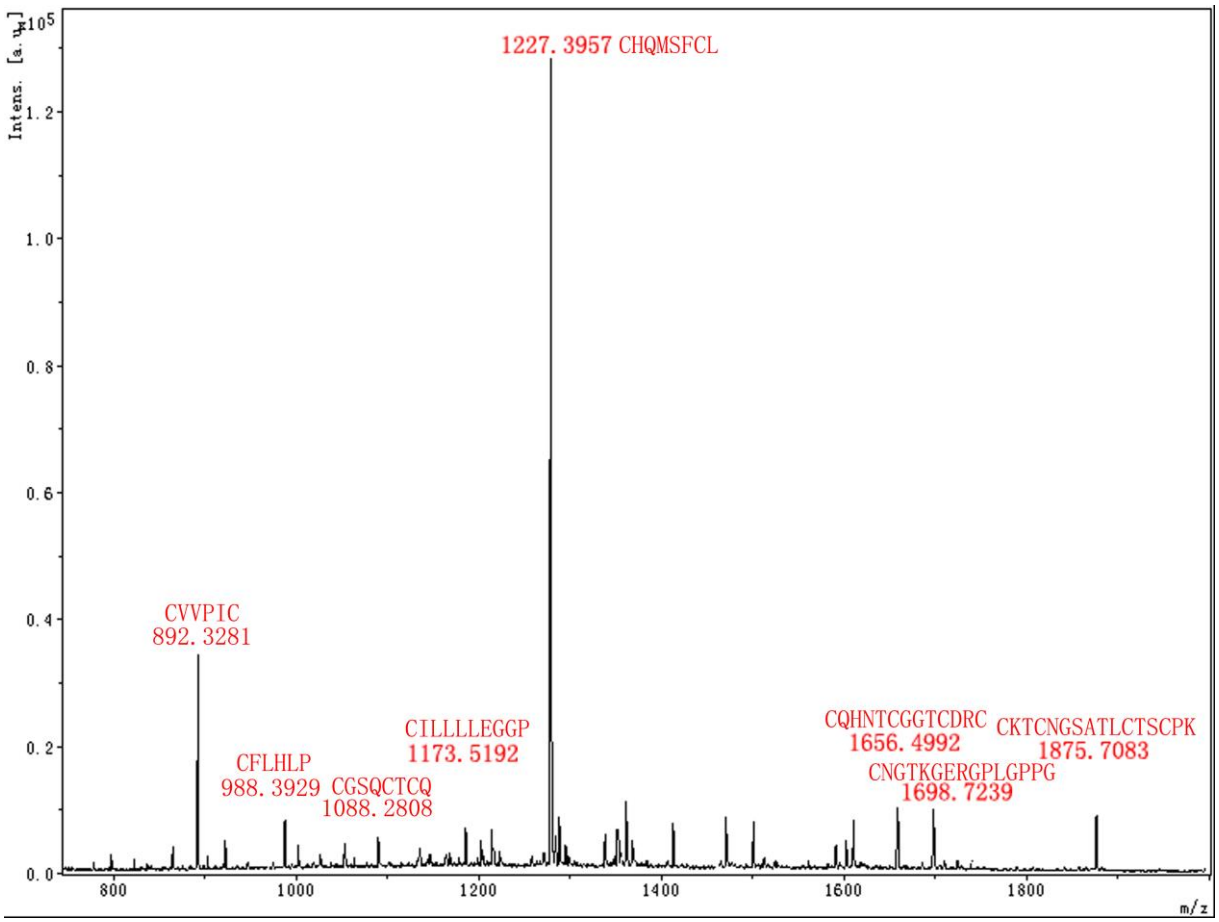


Figure S16. MALDI-TOF MS spectrum of the peptides enriched from Amniotic fluid in this work.

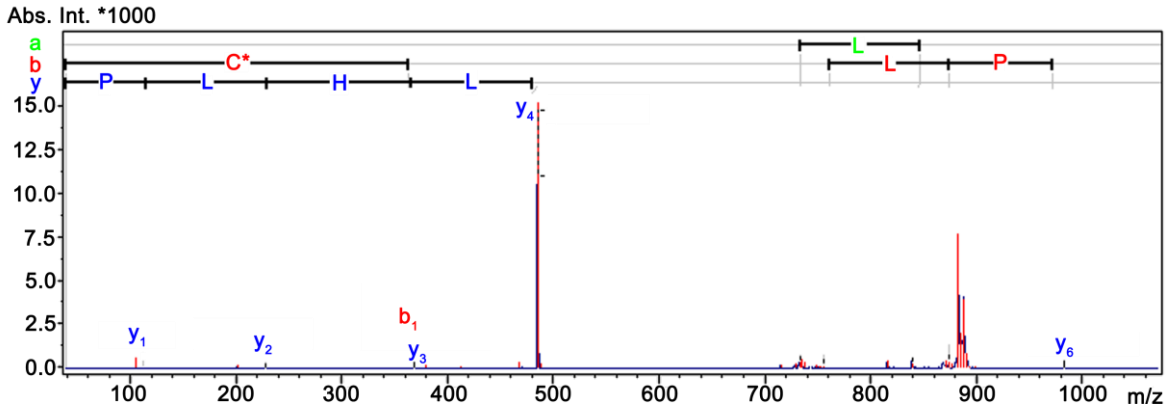


Figure S17. MALDI-TOF/TOF MS spectrum of one peptide enriched from Amniotic fluid in this work. The identified peptide sequence is CFLHLP. C* referred to the N-terminal Cys motif reacting with **1**.

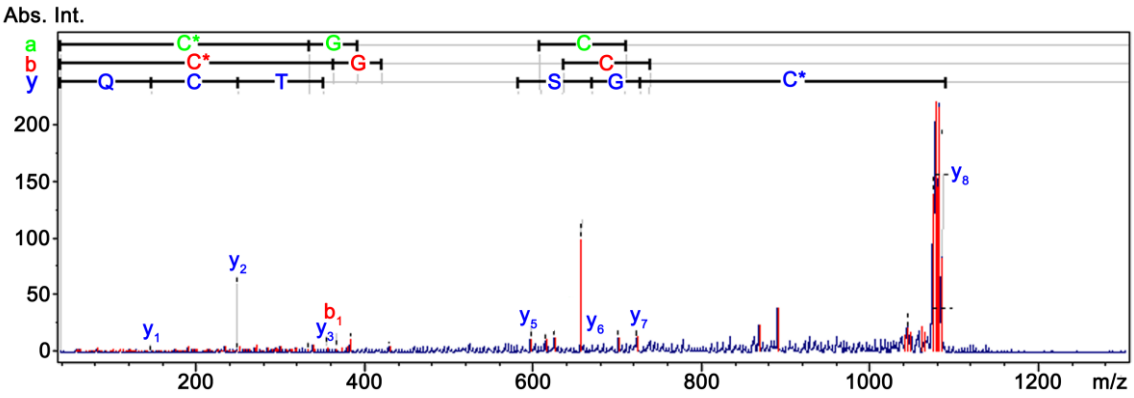


Figure S18. MALDI-TOF/TOF MS spectrum of one peptide enriched from Amniotic fluid in this work. The identified peptide sequence is CGSQCTCQ. C* referred to the N-terminal Cys motif reacting with **1**.

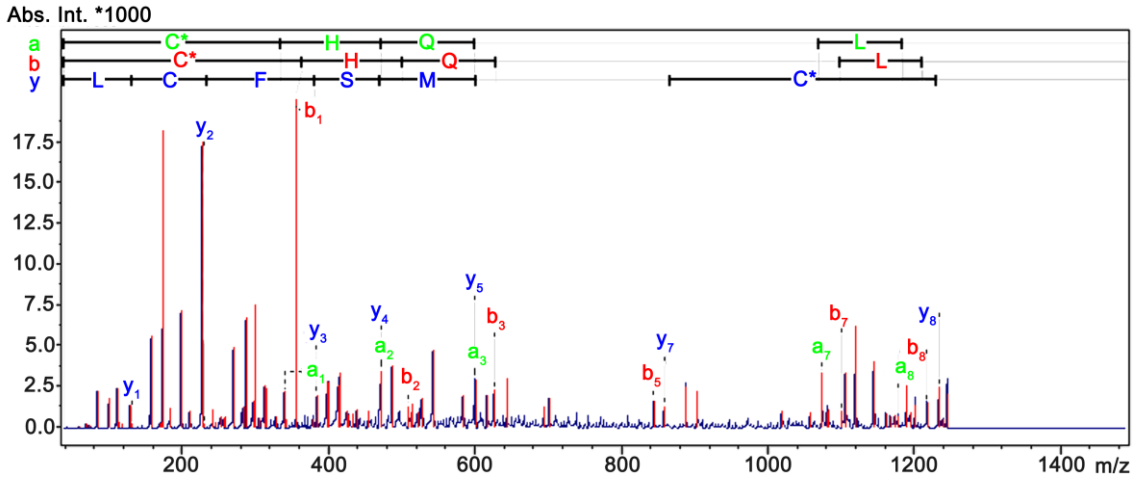


Figure S19. MALDI-TOF/TOF MS spectrum of one peptide enriched from Amniotic fluid in this work. The identified peptide sequence is CHQMSFCL. C* referred to the N-terminal Cys motif reacting with **1**.

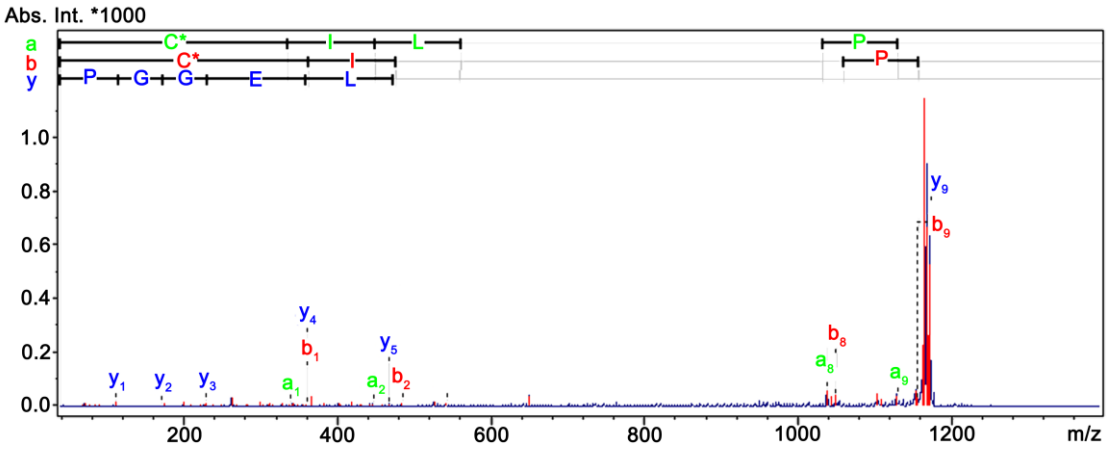


Figure S20. MALDI-TOF/TOF MS spectrum of one peptide enriched from Amniotic fluid in this work. The identified peptide sequence is CILLLEGGP. C* referred to the N-terminal Cys motif reacting with **1**.

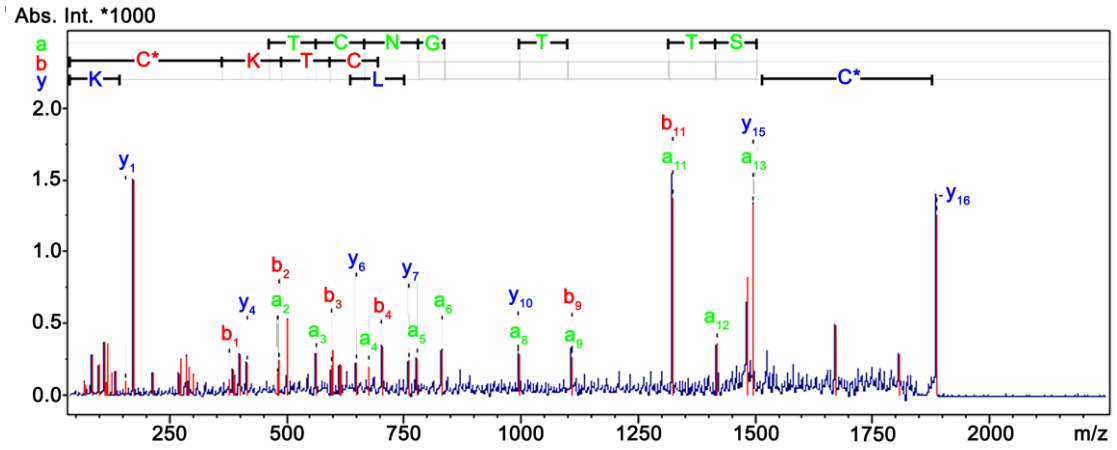


Figure S21. MALDI-TOF/TOF MS spectrum of one peptide enriched from Amniotic fluid in this work. The identified peptide sequence is CKTCNGSATLCTSCPK. C* referred to the N-terminal Cys motif reacting with **1**.

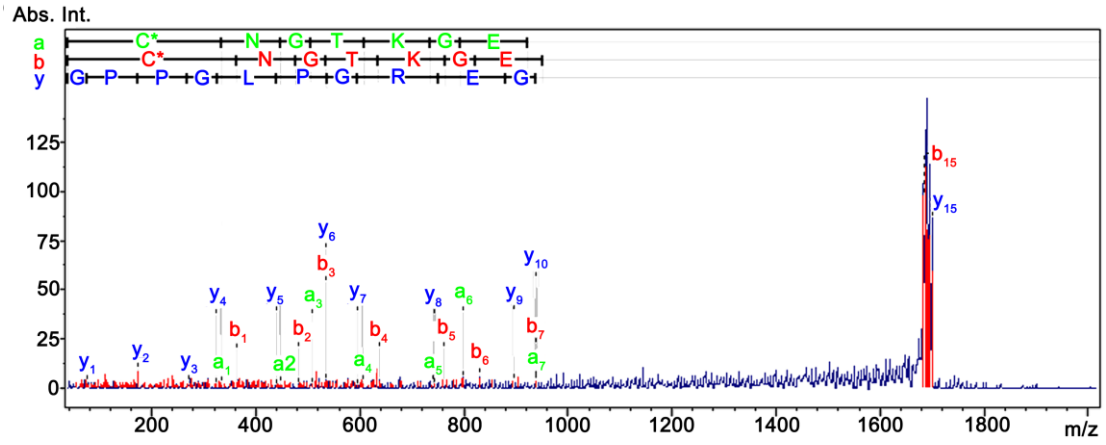


Figure S22. MALDI-TOF/TOF MS spectrum of one peptide enriched from Amniotic fluid in this work. The identified peptide sequence is CNGTKGERGPLGPPG. C* referred to the N-terminal Cys motif reacting with **1**.

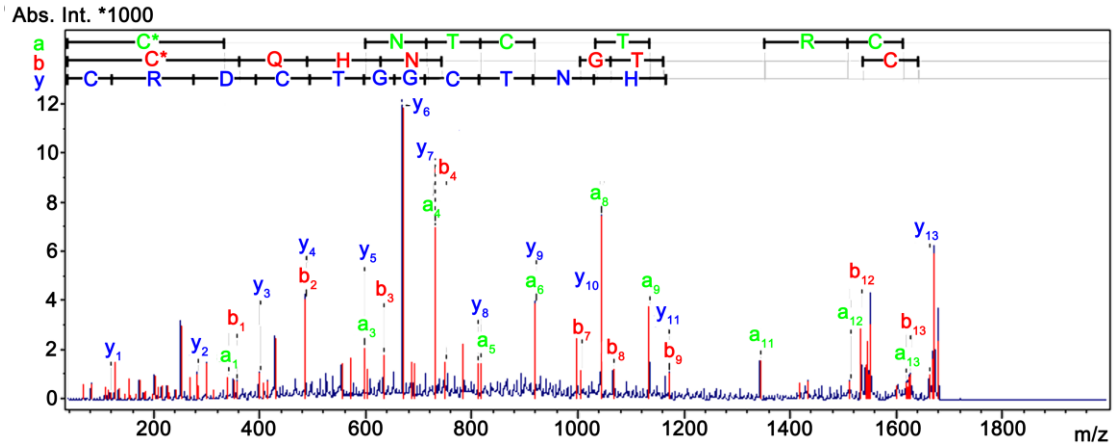


Figure S23. MALDI-TOF/TOF MS spectrum of one peptide enriched from Amniotic fluid in this work. The identified peptide sequence is CQHNTCGGTCDRC. C* referred to the N-terminal Cys motif reacting with **1**.

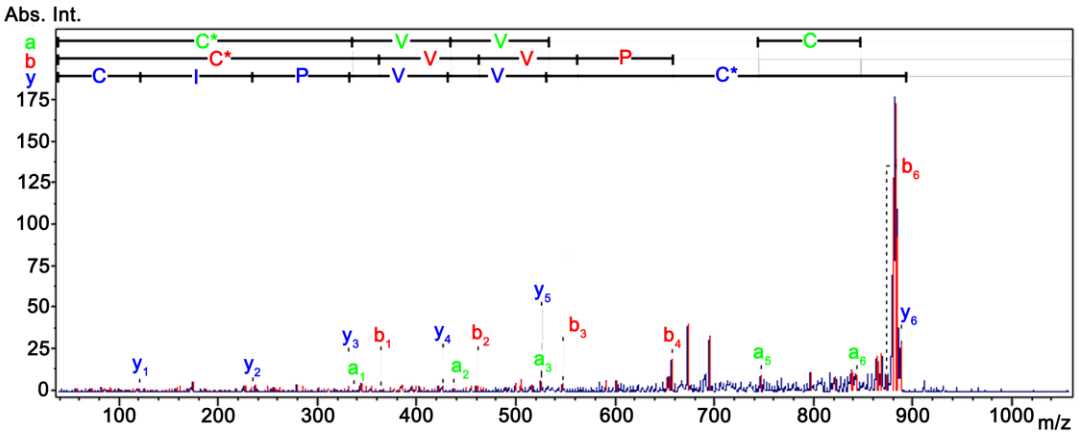


Figure S24. MALDI-TOF/TOF MS spectrum of one peptide enriched from Amniotic fluid in this work. The identified peptide sequence is CVVPIC. C* referred to the N-terminal Cys motif reacting with **1**.

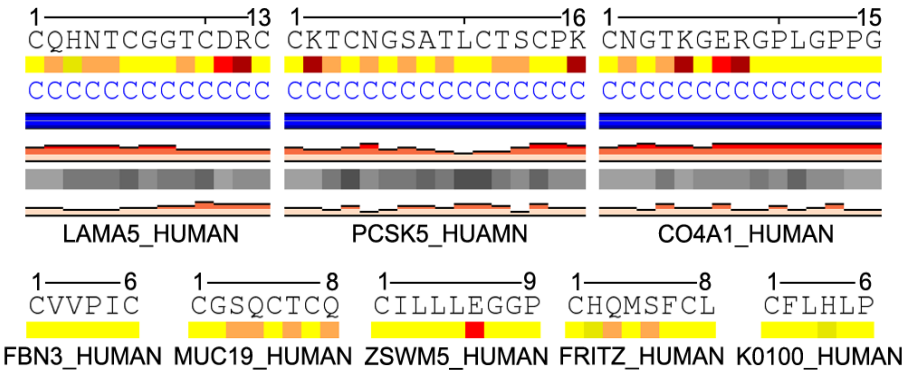


Figure 25. The sequence-based prediction of the secondary and supersecondary structures of the eight identified peptides. Blue C, C-coil structure. Hydrophobic and hydrophilic are labeled with yellow and red colors, respectively.

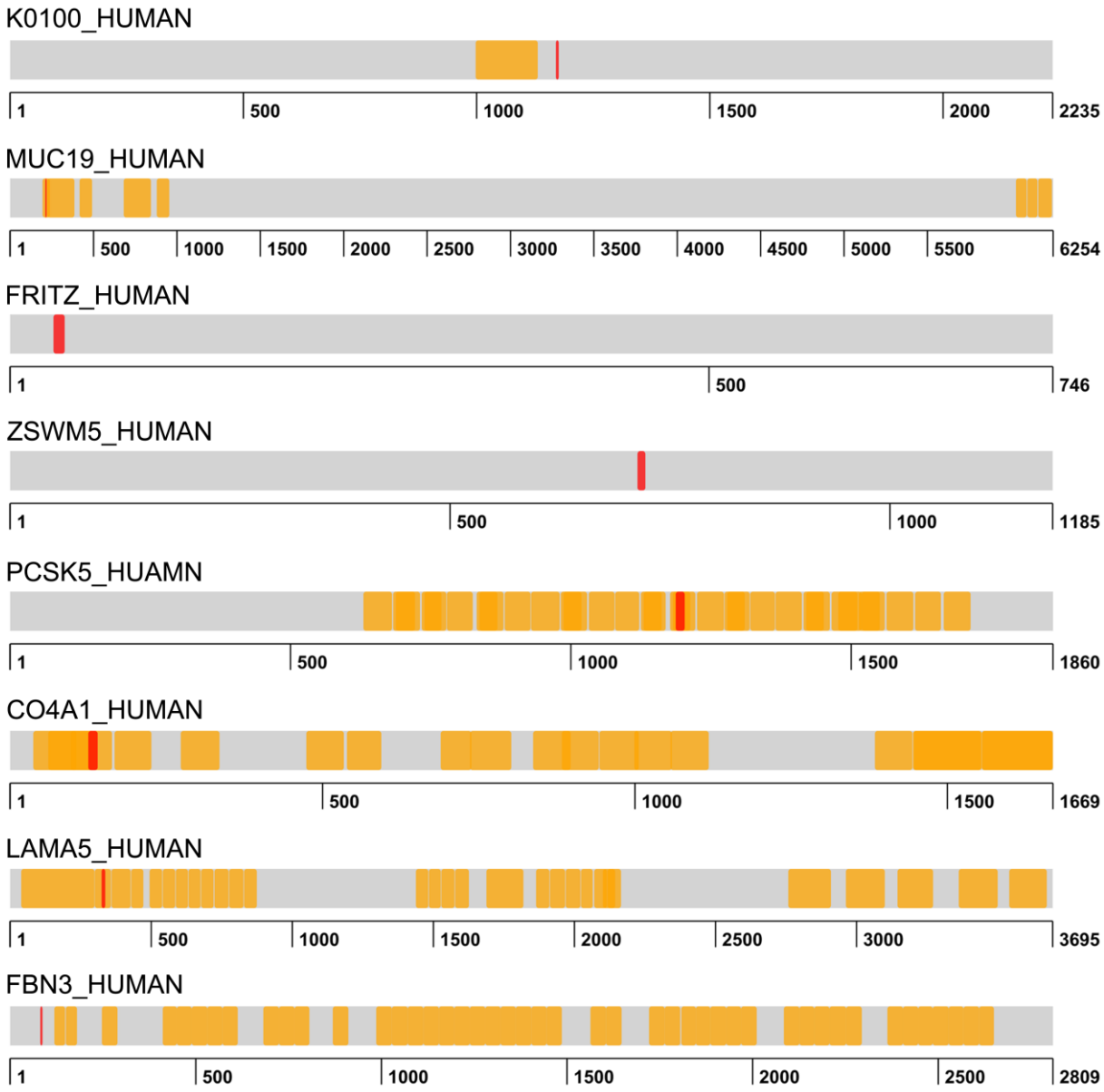


Figure S26. The positions of the identified peptides in their corresponding proteins. For each protein, domain regions were highlighted with orange rectangles, while the identified peptide was highlighted with red rectangle.

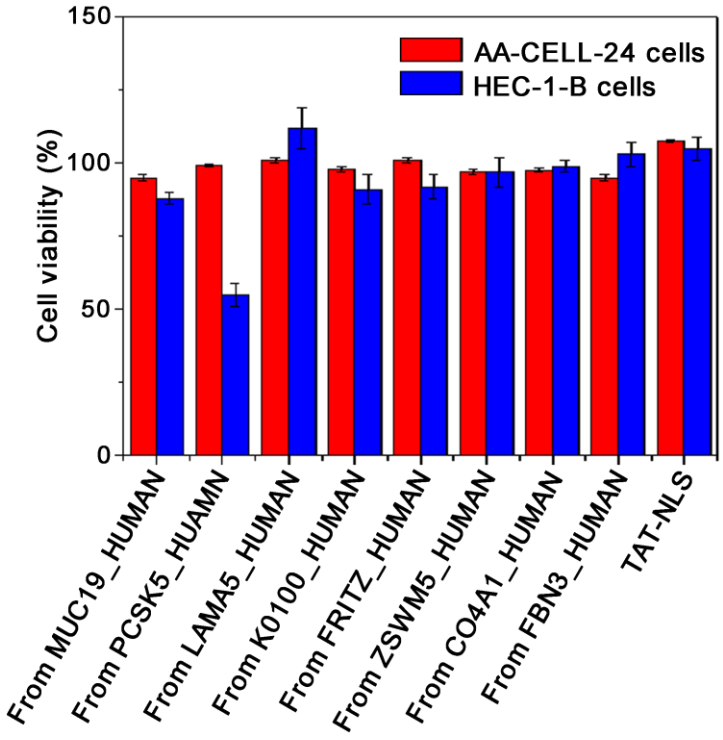


Figure S27. Cell viability of AA-CELL-24 cells (red) and HEC-1-B cells (blue) treated with 80 μM the fusion peptides (from each protein) or TAT-NLS for 48 h, respectively. Each value represents the average value from three independent experiments.

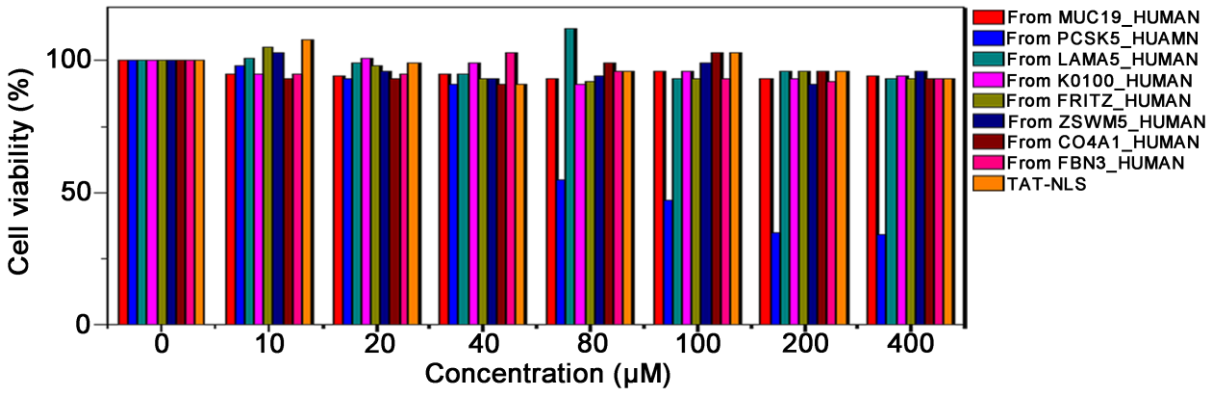


Figure S28. Cell viability of HEC-1-B cells treated with TAT-NLS-CGSQCTCQ-NH_2 (from MUC19_HUMAN, red), TAT-NLS-CKTCNGSATLCTSCPK-NH_2 (from PCSK5_HUAMN, blue), TAT-NLS-CQHNTCGGTCDRC-NH_2 (from LAMA5_HUMAN, dark cyan), TAT-NLS-CFLHLP-NH_2 (from K0100_HUMAN, magenta), TAT-NLS-CHQMSFCL-NH_2 (from FRITZ_HUMAN, dark yellow), TAT-NLS-CILLLEGGP-NH_2 (from ZSWM5_HUMAN, navy), TAT-NLS-CNGTKGERGPLGPPG-NH_2 (from CO4A1_HUMAN, wine), TAT-NLS-CVVPIC-NH_2 (from FBN3_HUMAN, pink) or TAT-NLS (orange) for 48 h, respectively.

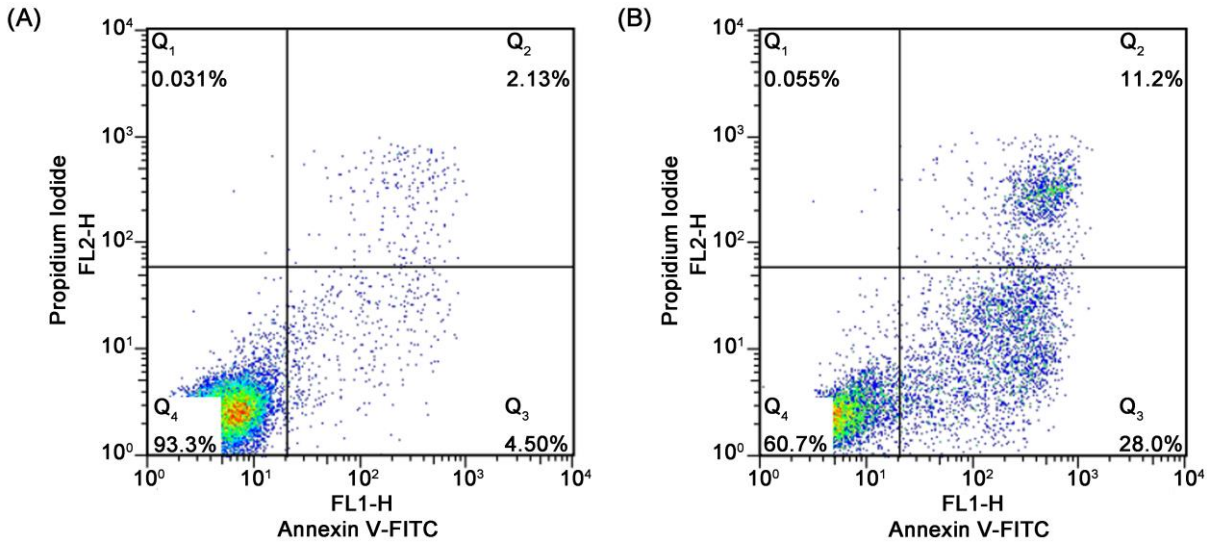


Figure S29. Flow cytometric analysis of HEC-1-B cells after incubated with (B) or without (A) 80 μM TAT-NLS-CKTCNGSATLCTSCPK-NH_2 for 24 h and stained with annexin V-FITC and propidium iodide. The percentages of live cells, cells at early stage of apoptosis, cells at late stage of apoptosis, and cells of necrosis were expressed in each quadrant.

Table S1. The recovery test of using **1**-conjugated resin to fish **2** in solution.

Amount of **2** (nmol)	Reacted (%)	Unreacted (%)	Total recovery (%)
10	40.7 ± 13.1	46.2 ± 3.6	86.8 ± 9.6
1	24.0 ± 16.4	81.7 ± 0.9	105.7 ± 15.5

Table S2. The PCSK5_HUAMN-degrading enzymes predicted by MEROPS

Peptidase	Enzyme class
Ser_{1178}-Cys_{1179} cleavage site	
Cathepsin S	Cysteine
Cathepsin E	Cysteine
Lysosomal 66.3 kDa protein ({Mus musculus}.-type)	Cysteine
S26 homologue	Serine
Matrix metallopeptidase-2	**Metallo**
Inter-containing DNA polymerase Ⅱ large subunite DP2 precursor Mername–AA281	Asparagine
Signalase (animal) 21 kDa component	Serine

Signal peptidase complex (animal)	Serine
Lys_{1121}-Gly_{1122} cleavage site	
Matrix metallopeptidase-2	**Metallo**
Matrix metallopeptidase-9	Metallo
Matrix metallopeptidase-13	Metallo
Granzyme A	Serine
Trypsin 1	Serine
Lysyl endopeptidase (bacteria)	Serine
Oligopeptidase B	Serine

Matrix metallopeptidase-2 (in bold) is the only enzyme that was pretided to cleave both Ser_{1178}-Cys_{1179} and Lys_{1121}-Gly_{1122} sites of protein PCSK5_HUAMN to yield peptide CKTCNGSATLCTSCPK.

4. References

(1) Duong, S.; Strobel, N.; Buddhadasa, S.; Stockham, K.; Auldist, M.; Wales, B.; Orbell, J.; Cran, M. *Food Chem.* **2016,** *211*, 570.

(2) Nagahara, H.; Vocero-Akbani, A. M.; Snyder, E. L.; Ho, A.; Latham, D. G.; Lissy, N. A.; Becker-Hapak, M.; Ezhevsky, S. A.; Dowdy, S. F. *Nat. Med.* **1998**, *4*, 1449.

(3) Gump, J. M.; Dowdy, S. F. *Trends Mol. Med.* **2007**, *13*, 443.