Supporting Information

Dearomatization-Rearomatization Strategy of Tyrosine for Peptides/Protein Modification through Thiol-Addition Reactions

Pengxin Wang,[a,b,] Yulian Cheng,[b] Chunlei Wu,[b] Ruixiang Luo,[b] Caibing Ma,[b] Yimin Zhou,[b] Zhilong Ma,[b] Rui Wang,[a,] Wu Su,[b,]* and Lijing Fang[b,]**

[a]Key Laboratory of Preclinical Study for New Drugs of Gansu Province, Institute of Drug Design & Synthesis, School of Basic Medical Sciences, Lanzhou University, Lanzhou 730000, Gansu, China.

[b]Guangdong Key Laboratory of Nanomedicine, Institute of Biomedicine and Biotechnology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, Guangdong, China.

wangrui@lzu.edu.cn, wu.su@siat.ac.cn and lj.fang@siat.ac.cn.

Table of contents

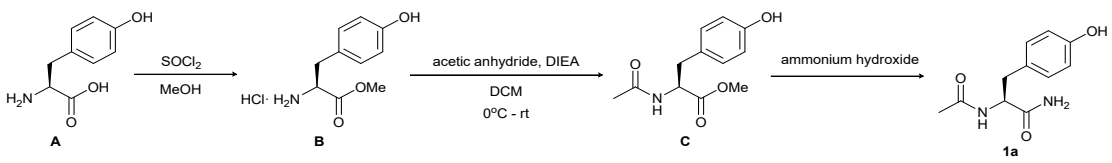
1. General Information

$PhI(OAc)_2$ (Iodosobenzene diacetate), HATU (2-(7-Azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate) and the HPLC grade solvents (CH_3CN and MeOH) were purchased from J&K Scientific. Bovine serum albumin was purchased from Macklin. 2-phenyl ethanethiol (**2a**), mercaptoethanol (**2b**), acetyl protected thioglucoside (**2j**), DIEA (N,N-Diisopropylethylamine), TFA (trifluoroacetic acid), piperidine and other chemicals were acquired from Energy Chemical and were used as received.

Analytical RP-HPLC was performed on the Agilent 1260 high-performance liquid chromatography (HPLC) instrument (UV-vis detector) with Poroshell 120, EC-C18 column (4.6 x 100 mm, 2.7 μm) maintained at 30 °C. The RP-HPLC gradient was started at 10% of B (MeCN), then increased to 100% of B over 20 min (A: 0.1% TFA in water) at a flow rate of 0.5 mL/min. Semi-preparative RP-HPLC was performed on the ULTIMAT 3000 Instrument (DIONEX). UV absorbance was measured using a photodiode array detector at 220 and 254 nm. The RP-HPLC gradient was started at 10% of B (MeCN), then increased to 100% of B over 30 min (A: 0.1% TFA in water). ^1H NMR (^{13}C NMR) spectra were recorded with a Bruker AV400 at 400 (100) MHz. Chemical shifts are referenced to either tetramethylsilane as an internal standard or the signals resulting from the residual solvent. High resolution mass spectra were measured with an ABI Q-star Elite.

2. Synthesis of Starting Materials

2.1 Synthesis of Ac-Tyr-NH_2 (1a)



To a solution of tyrosine (**A**) (7.24g, 40 mmol) in MeOH (100 mL) was added $SOCl_2$

(14.56 mL, 0.2 mmol) dropwise at 0°C. The resulting solution was warmed to room temperature and stirred overnight. The reaction mixture was concentrated to give tyrosine methyl ester hydrochloride (**B**), which was used directly without further purification.

Acetic anhydride (4.5 mL, 48 mmol) was added dropwise to the solution of **B** and DIEA (19.8 mL, 120 mmol) in dry DCM (100 mL) at 0°C and the resulting mixture was stirred at room temperature for 2h. The reaction was washed with saturated $NaHCO_3$ and brine, dried over anhydrous Na_2SO_4 and concentrated to dryness in vacuo. Purification of the residue by flash chromatography (petroleum ether/ethyl acetate=3/1 to 1/4) led to Ac-Tyr-OMe (**C**) as a white powder (7 g, 74%).

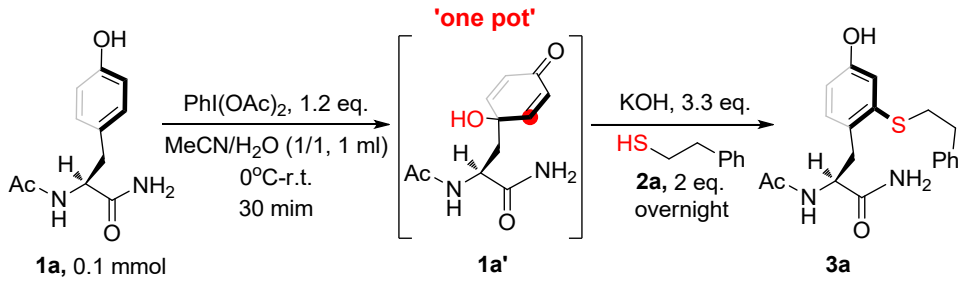
C (7 g) was dissolved into 50 ml ammonium hydroxide aq. solution. The mixture was allowed to stir at room temperature for 12 hours. Then the suspension was put under high vacuum to give the pure product **1a** as a white solid. This product was directly used without any further purification. All the characterization data are consistent with the previous report (Z. Qiu, L. Lv, J. Li, C.-C. Li and C.-J. Li, *Chemical Science*, **2019**, *10*, 4775.).

2.2 Synthesis of peptides 1 and 2

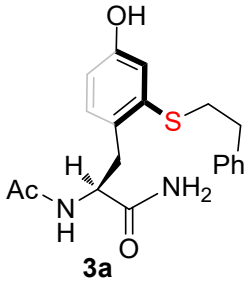
Peptides **1** and **2** were synthesized using Fmoc-based solid-phase peptide synthesis (SPPS) on Rink Amide AM resin. Fmoc (9-fluorenylmethoxycarbonyl) was deprotected with 20% piperidine in DMF. The resin was washed with DMF (5 x 6 mL) and dry DMF (1 x 6 mL). For peptide elongation, the protected amino acid (4 eq.) was activated using HATU (4 eq.) and DIEA (10 eq.) in dry DMF and then transferred to the deprotected resin. The extent of coupling was assessed by the Kaiser test.

3. General Procedure and characterization.

3.1 General Procedure for the synthesis of 3a-3d



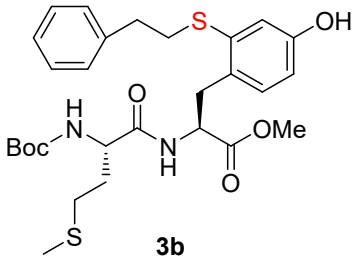
To a solution of **1a** (0.1 mmol) in MeCN/H_2O (1/1, 1 ml) was added $PhI(OAc)_2$ (1.2 equiv) at 0°C. The mixture was stirred at rt for 30 min before KOH (3.3 equiv) and **2a** (2.0 equiv) was added respectively. After stirring at rt overnight, the reaction mixture was purified by semi-preparative RP-HPLC to give **3a** as a brown solid (21.5 mg, 60%).



3a

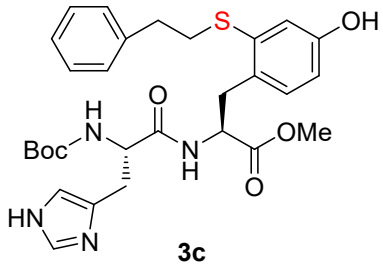
^{1}H NMR (400 MHz, CD_3OD): δ 7.34 – 7.16 (m, 5H), 7.06 (d, J = 8.3 Hz, 1H), 6.88 (d, J = 2.4 Hz, 1H), 6.59 (dd, J = 8.3, 2.4 Hz, 1H), 4.65 (dd, J = 9.0, 5.7 Hz, 1H), 3.26 – 3.13 (m, 3H), 2.98 – 2.85 (m, 3H), 1.91 (s, 3H) ppm; **^{13}C NMR** (100 MHz, CD_3OD): δ 175.1, 171.7, 156.5, 140.3, 136.7, 131.1, 128.2 (overlapped), 128.1 (overlapped), 127.4, 126.0, 115.1, 112.7, 53.5, 35.1, 34.8, 34.7, 21.1 ppm; **IR** (neat): 2963, 2926, 2852, 1647, 1417, 1264, 1095, 1025, 802 cm^{-1}; **HRMS** (ESI-TOF) m/z: calculated for $C_{19}H_{22}O_3N_2S$: 359.14239, $[M + H]^+$. Found: 359.14222, $[M + H]^+$.

Dipeptides **3b**, **3c** and **3d** were synthesized according to the general procedure above.



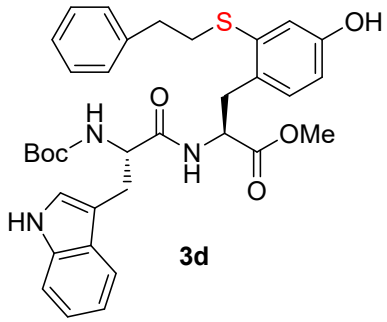
3b

^{1}H NMR (400 MHz, $CDCl_3$): δ 7.31 (dd, J = 13.7, 6.3 Hz, 2H), 7.23 (dd, J = 11.3, 7.2 Hz, 3H), 6.98 (d, J = 8.2 Hz, 1H), 6.84 (d, J = 2.0 Hz, 2H), 6.58 (d, J = 7.4 Hz, 1H), 5.24 (s, 1H), 4.85 (dd, J = 13.5, 8.1 Hz, 1H), 4.27 (s, 1H), 3.73 (s, 3H), 3.26 (dd, J = 14.2, 5.3 Hz, 1H), 3.19 – 3.11 (m, 2H), 3.07 (dd, J = 14.1, 8.5 Hz, 1H), 2.98 – 2.91 (m, 2H), 2.50 (s, 2H), 2.07 (s, 3H), 2.06 – 1.92 (m, 2H), 1.45 (s, 9H) ppm; **^{13}C NMR** (100 MHz, $CDCl_3$): δ 171.8, 171.4, 155.7, 155.5, 140.0, 137.4, 131.4, 128.6, 128.5, 127.1, 126.5, 115.7, 113.5, 80.4, 53.4, 53.1, 52.5, 35.3, 35.2, 35.1, 31.7, 30.0, 28.3, 15.2 ppm; **IR** (neat): 3751, 2930, 2362, 2929, 1685, 1559, 1507, 744, 668, 419 cm^{-1}; **HRMS** (ESI-TOF) m/z: calculated for $C_{28}H_{39}O_6N_2S_2$: 563.2244, $[M + H]^+$. Found: 563.2246, $[M + H]^+$.



3c

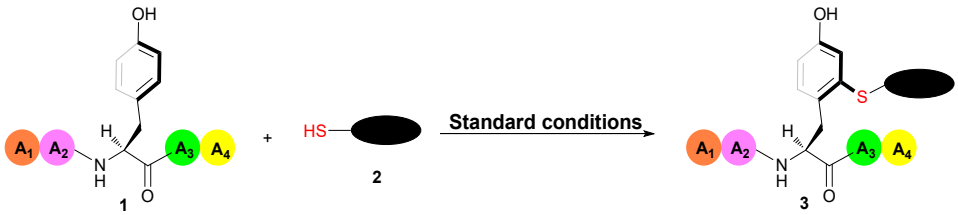
^{1}H NMR (400 MHz, CD_3OD) δ 8.79 (d, J = 17.2 Hz, 1H), 7.37 – 7.13 (m, 6H), 6.99 (d, J = 8.3 Hz, 1H), 6.88 (t, J = 3.6 Hz, 1H), 6.62 – 6.55 (m, 1H), 4.75 – 4.70 (m, 1H), 4.40 (s, 1H), 3.70 (d, J = 20.2 Hz, 3H), 3.28-3.23 (m, 1H), 3.20 – 3.13 (m, 3H), 3.01 – 2.90 (m, 4H), 1.42 (s, 9H) ppm; **^{13}C NMR** (100 MHz, CD_3OD): δ 172.2, 170.9, 156.7, 140.2, 136.7, 133.4, 131.3. 129.8, 128.2, 128.1, 126.7, 126.0, 117.1, 116.7, 115.4, 113.0, 78.2, 77.9, 77.5, 52.9, 51.4, 35.1, 34.8, 34.4, 27.2 ppm; **IR** (neat): 3868, 3743, 2929, 2360, 1684, 1507, 742, 668, 545, 457 cm^{-1}; **HRMS** (ESI-TOF) m/z: calculated for $C_{29}H_{37}O_6N_4S$: 569.2428, $[M + H]^+$. Found: 569.2427, $[M + H]^+$.



3d

^{1}H NMR (400 MHz, CD_3OD) δ 7.90 (d, J = 7.1 Hz, 1H), 7.59 (d, J = 7.8 Hz, 1H), 7.32 (t, J = 11.3 Hz, 1H), 7.30 – 7.21 (m, 2H), 7.20 – 7.09 (m, 4H), 7.08 (d, J = 4.0 Hz, 1H), 7.02 (t, J = 7.4 Hz, 1H), 6.86 (d, J = 7.6 Hz, 1H), 6.75 (d, J = 8.1 Hz, 1H), 4.63 (d, J = 6.3 Hz, 1H), 4.35 (t, J = 6.6 Hz, 1H), 3.62 (s, 3H), 3.18 (dd, J = 14.6, 5.7 Hz, 1H), 3.03 (dd, J = 22.9, 14.6 Hz, 3H), 2.94 (dd, J = 13.5, 5.9 Hz, 1H), 2.90 – 2.75 (m, 3H), 1.56 – 1.20 (m, 8H); **^{13}C NMR** (100 MHz, CD_3OD): δ 173.1, 171.7, 156.1, 155.5, 140.4, 136.7, 133.3, 129.4, 128.2, 128.0, 127.9, 127.4, 1 25.9, 123.2, 121.0, 120.7, 118.4, 118.0, 114.8, 110.9, 109.5, 79.35, 55.5, 53.8, 51.3, 36.2, 35.5, 35.7, 27.8, 27.2 ppm; **IR** (neat): 3733, 3040, 2830, 2361, 2337, 1686, 1511, 992, 670, 417 cm^{-1}; **HRMS** (ESI-TOF) m/z: calculated for $C_{34}H_{40}O_6N_3S$: 618.2632, $[M + H]^+$. Found: 618.2630, $[M + H]^+$.

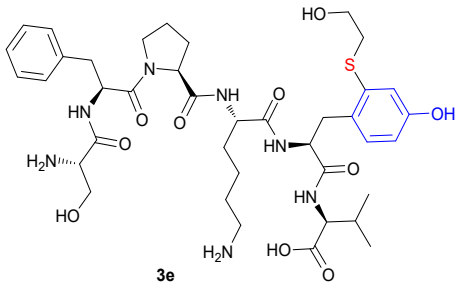
3.2 General Procedure for the synthesis of peptide conjugates 3e-3m.



To a solution of peptide **1** (0.005 mmol) in MeCN/H_2O (1/1, 0.2 ml) was added $PhI(OAc)_2$ (1.2 equiv) at 0°C. The mixture was stirred at rt for 30 min before KOH (3.3 equiv) and **2** (2.0 equiv) was added respectively. Then the reaction was kept stirring at rt overnight and was purified by semi-preparative RP-HPLC to give **3e-3m**.

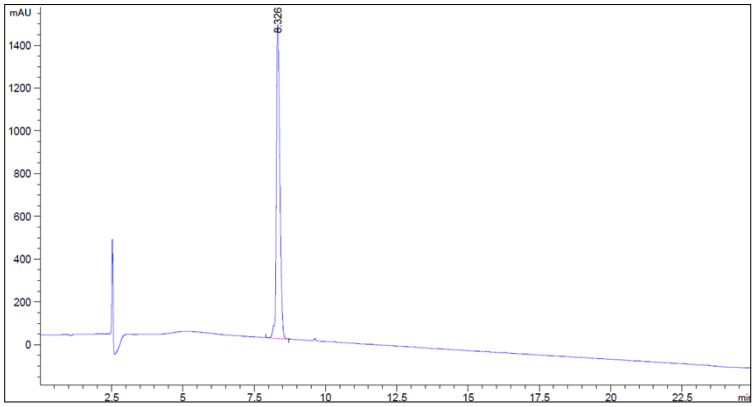
3.3 Chemical structures and characterization of peptides 3e-3m.

a)



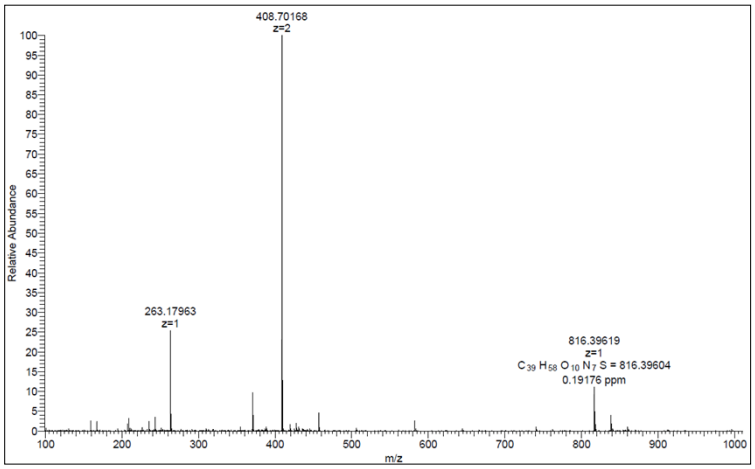
3e

b)



HPLC trace of peptide **3e**

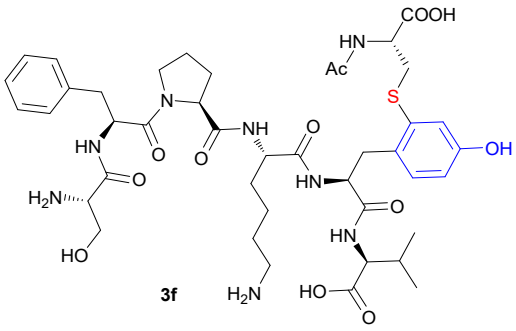
c)



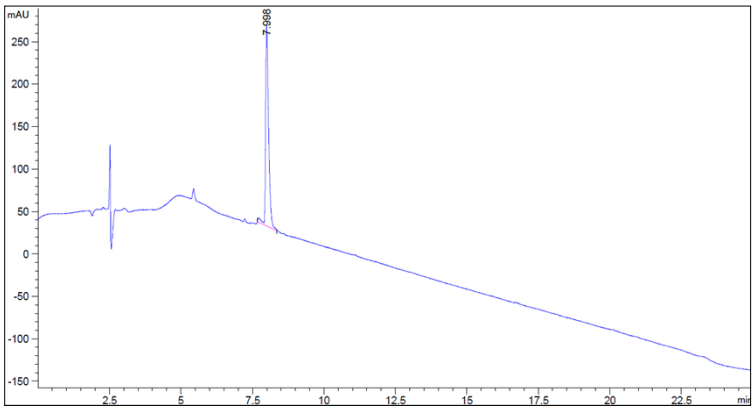
HRMS spectrum of peptide **3e**

Figure S1. a) Chemical structure of peptide **3e**, 26% isolated yield; b) Analysis of HPLC trace of peptide **3e**. (HPLC gradient is 10% to 100% of solution B in 30 min on the Analysis C18 column (λ = 220 nm); c) ESI-MS: m/z calculated for $C_{39}H_{57}O_{10}N_7S$: 816.39604, $[M + H]^+$. Found: 816.39619, $[M + H]^+$; 408.70168, $[M + 2H]^{2+}$.

a)

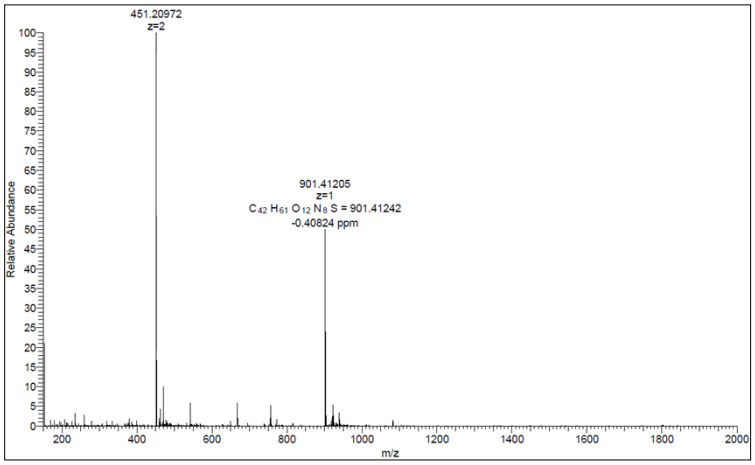


b)



HPLC trace of peptide **3f**

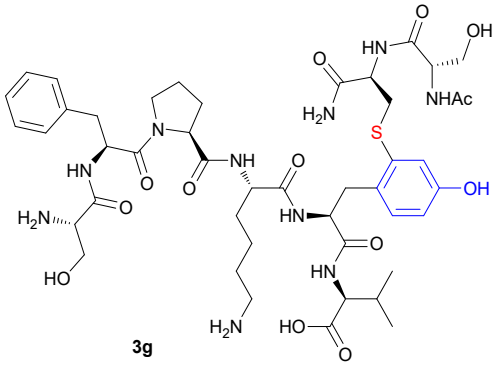
c)



HRMS spectrum of peptide **3f**

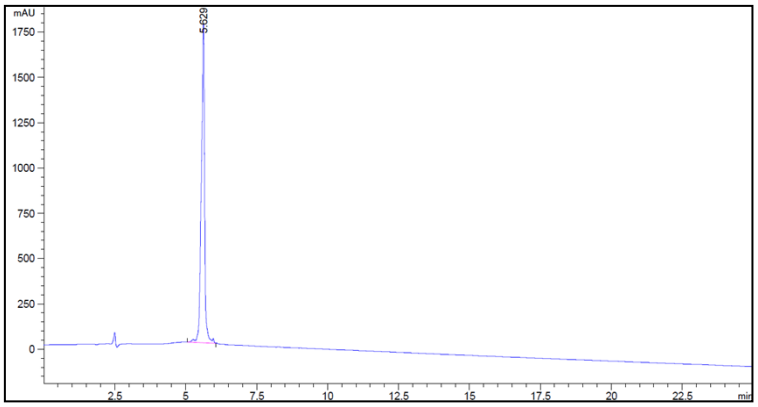
Figure S2. a) Chemical structure of peptide **3f**, 36% isolated yield; b) Analysis of HPLC trace of peptide **3f**. (HPLC gradient is 10% to 100% of solution B in 30 min on the Analysis C18 column (λ = 220 nm); c) ESI-MS: m/z calculated for $C_{42}H_{60}O_{12}N_8S$: 901.41242, $[M + H]^+$. Found: 901.41205, $[M + H]^+$; 451.20972, $[M + 2H]^{2+}$.

a)



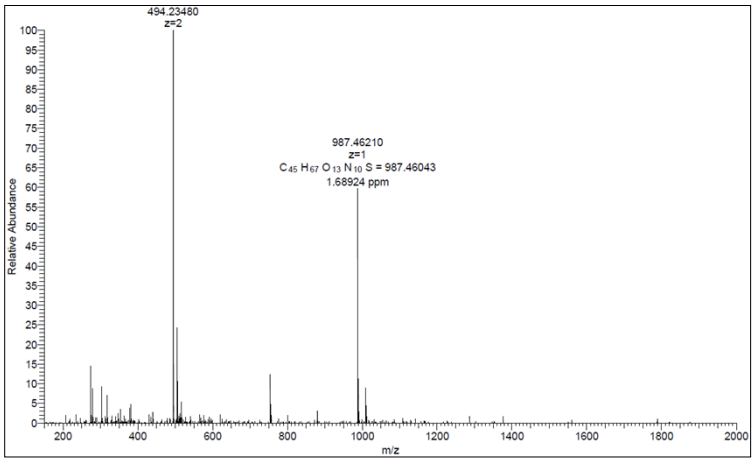
3g

b)



HPLC trace of peptide **3g**

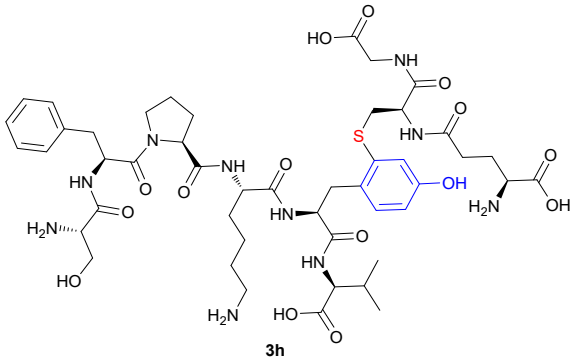
c)



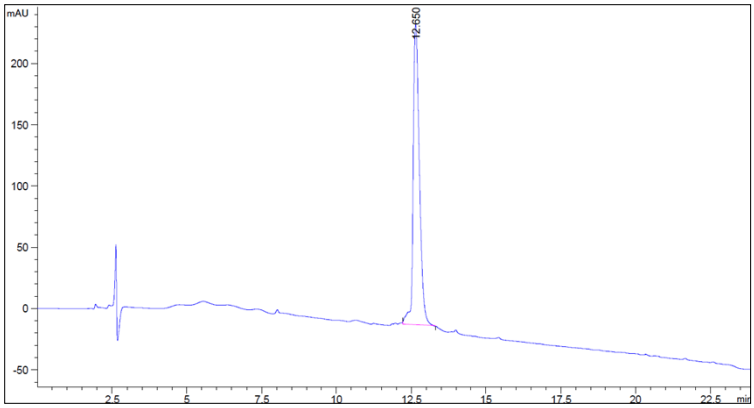
HRMS spectrum of peptide **3g**

Figure S3. a) Chemical structure of peptide **3g**, 34% isolated yield; b) Analysis of HPLC trace of peptide **3g**. (HPLC gradient is 10% to 100% of solution B in 30 min on the Analysis C18 column (λ = 220 nm); c) ESI-MS: m/z calculated for $C_{45}H_{66}O_{13}N_{10}S$: 987.46043, $[M + H]^+$. Found: 987.46210, $[M + H]^+$; 494.23480, $[M + 2H]^{2+}$.

a)

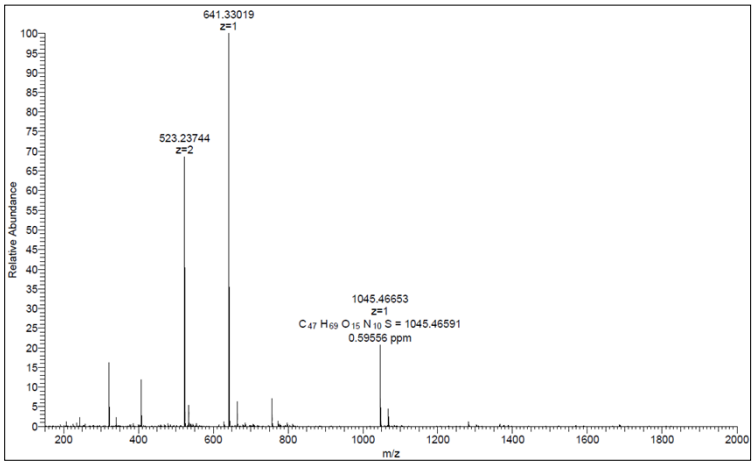


b)



HPLC trace of peptide **3h**

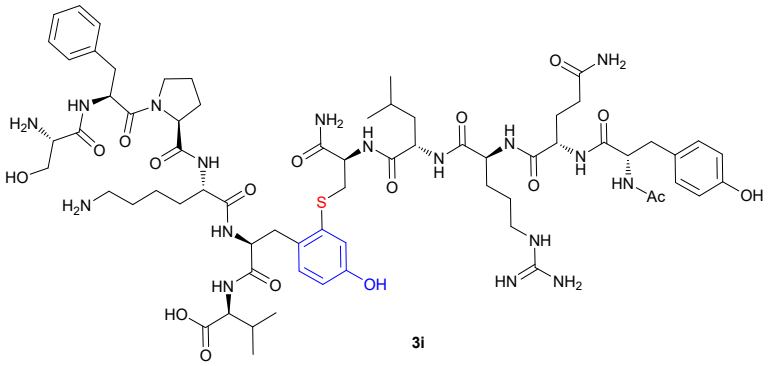
c)



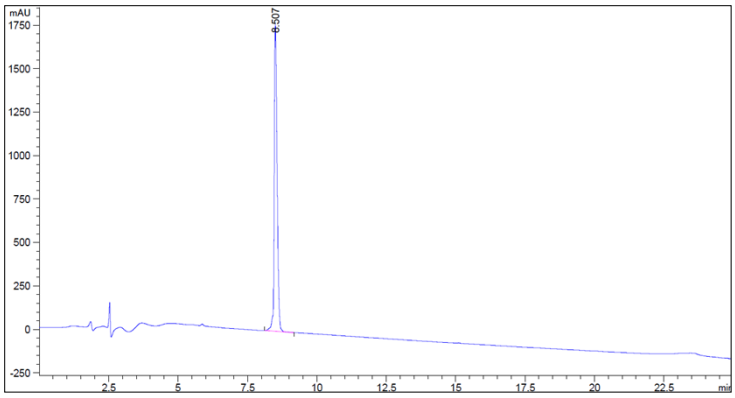
HRMS spectrum of peptide **3h**

Figure S4. a) Chemical structure of peptide **3h**, 30% isolated yield; b) Analysis of HPLC trace of peptide **3h**. (HPLC gradient is 10% to 100% of solution B in 30 min on the Analysis C18 column (λ = 220 nm); c) ESI-MS: m/z calculated for $C_{47}H_{68}O_{15}N_{10}S$: 1045.46591, $[M + H]^+$. Found: 1045.46653, $[M + H]^+$; 523.23744, $[M + 2H]^{2+}$.

a)

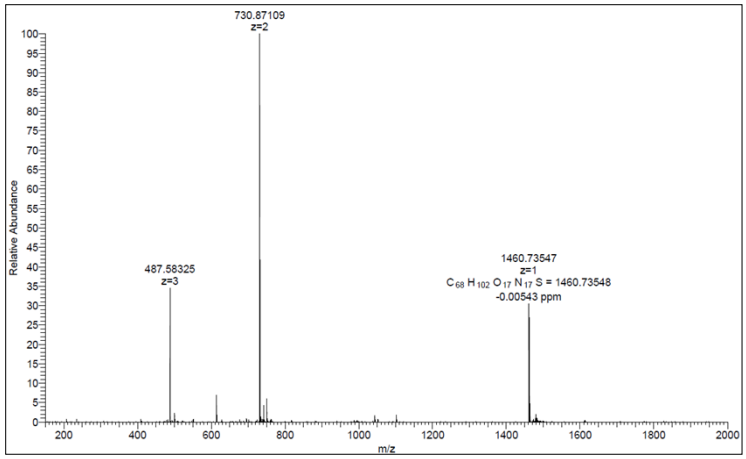


b)



HPLC trace of peptide **3i**

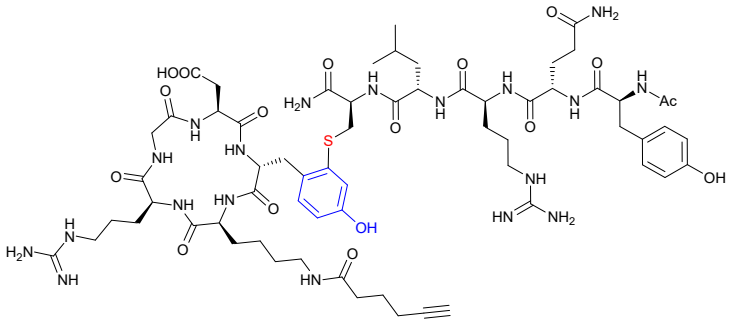
c)



HRMS spectrum of peptide **3i**

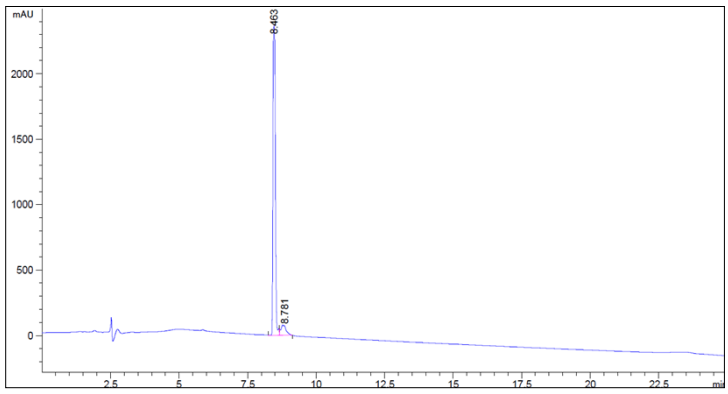
Figure S5. a) Chemical structure of peptide **3i**, 40% isolated yield; b) Analysis of HPLC trace of peptide **3i**. (HPLC gradient is 10% to 100% of solution B in 30 min on the Analysis C18 column (λ = 220 nm); c) ESI-MS: m/z calculated for $C_{68}H_{101}O_{17}N_{17}S$: 1460.73548, $[M + H]^+$. Found: 1460.73547, $[M + H]^+$; 730.87109, $[M + 2H]^{2+}$; 487.58325, $[M + 3H]^{3+}$.

a)



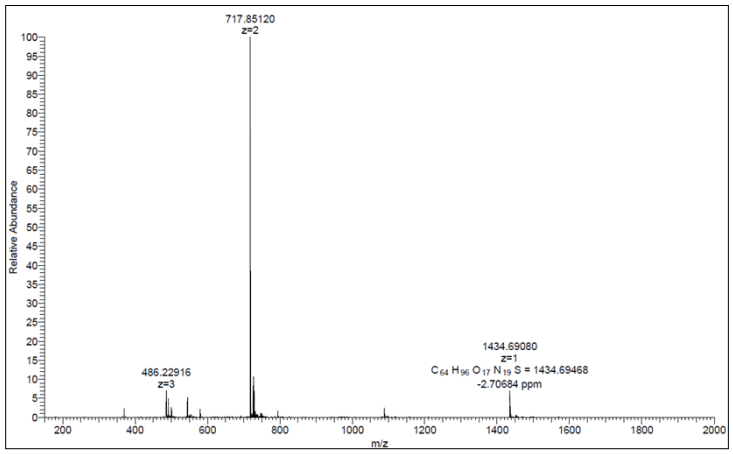
3j

b)



HPLC trace of peptide **3j**

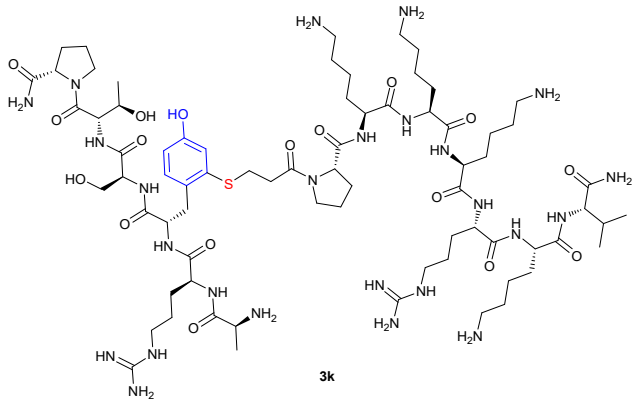
c)



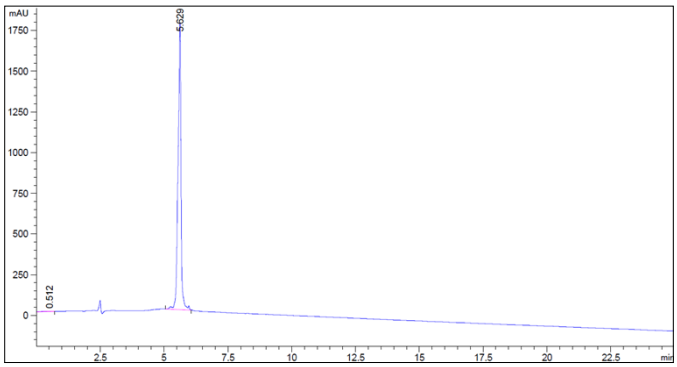
HRMS spectrum of peptide **3j**

Figure S6. a) Chemical structure of peptide **3j**, 48% isolated yield; b) Analysis of HPLC trace of peptide **3j**. (HPLC gradient is 10% to 100% of solution B in 30 min on the Analysis C18 column (λ = 220 nm); c) ESI-MS: m/z calculated for $C_{64}H_{95}O_{17}N_{19}S$: 1434.69468, $[M + H]^+$. Found: 1434.69080, $[M + H]^+$; 717.85120, $[M + 2H]^{2+}$; 486.22916, $[M + 3H]^{3+}$.

a)

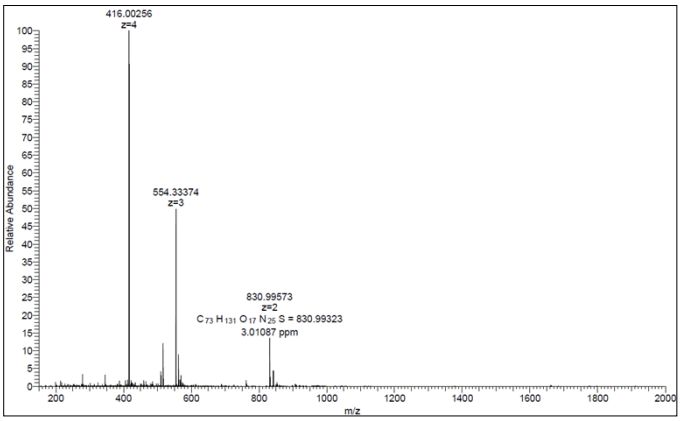


b)



HPLC trace of peptide **3k**

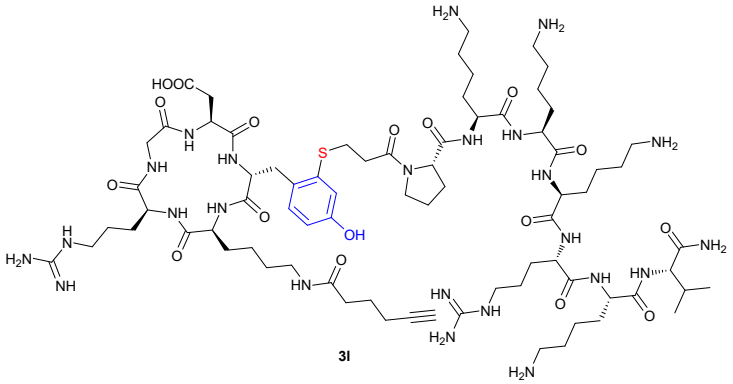
c)



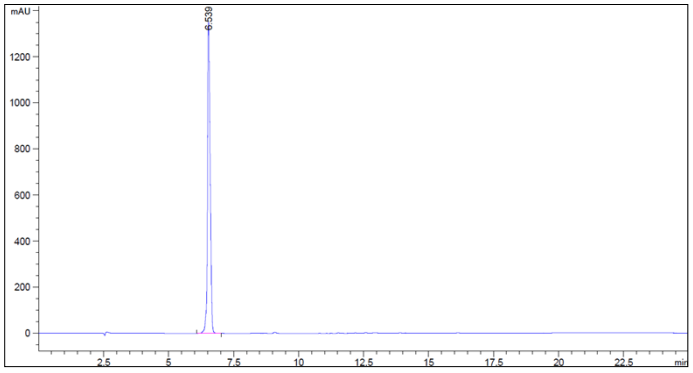
HRMS spectrum of peptide **3k**

Figure S7. a) Chemical structure of peptide **3k**, 40% isolated yield; b) Analysis of HPLC trace of peptide **3k**. (HPLC gradient is 10% to 100% of solution B in 30 min on the Analysis C18 column (λ = 220 nm); c) ESI-MS: m/z calculated for $C_{73}H_{129}O_{17}N_{25}S$: 830.99323, $[M + 2H]^{2+}$. Found: 830.99573, $[M + 2H]^{2+}$; 554.33374, $[M + 3H]^{3+}$; 416.00256, $[M + 4H]^{4+}$.

a)

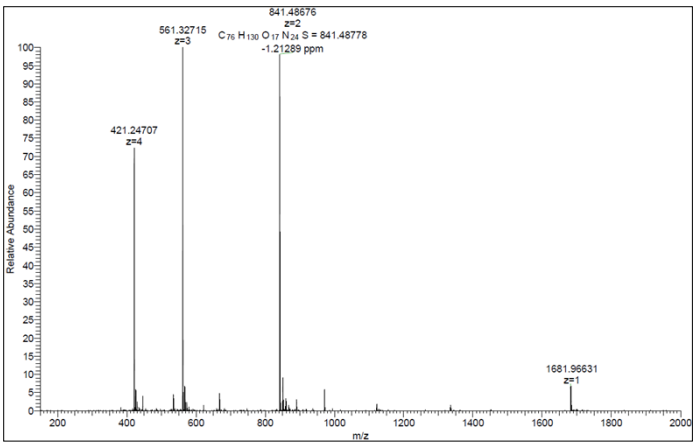


b)



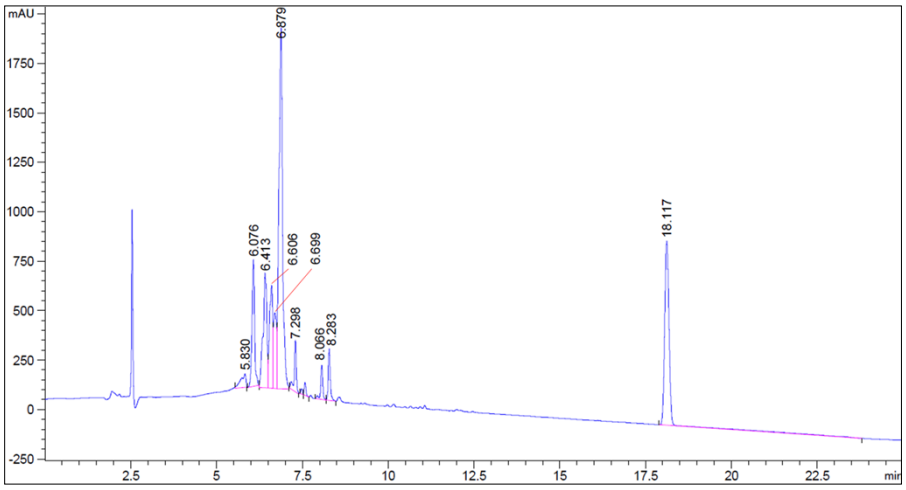
HPLC trace of peptide **3l**

c)

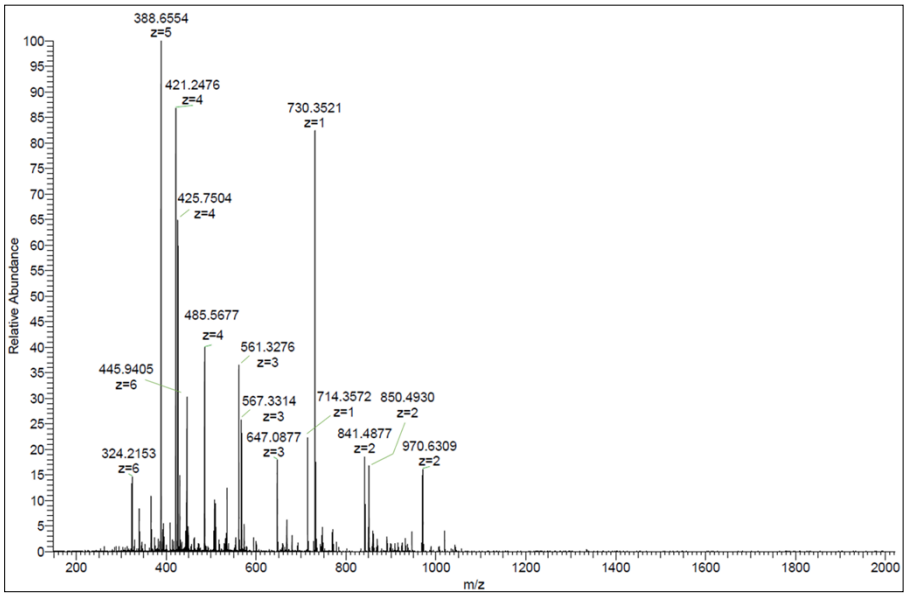


HRMS spectrum of peptide **3l**

Figure S8. a) Chemical structure of peptide **3l**, 45% isolated yield; b) Analysis of HPLC trace of peptide **3l**. (HPLC gradient is 10% to 100% of solution B in 30 min on the Analysis C18 column (λ = 220 nm). c) ESI-MS: m/z calculated for $C_{76}H_{128}O_{17}N_{24}S$: 1681.96882, $[M + H]^+$. Found: 1681.96631, $[M + H]^+$; 841.48676, $[M + 2H]^{2+}$; 561.32715, $[M + 3H]^{3+}$; 421.24707, $[M + 4H]^{4+}$.



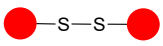
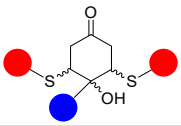
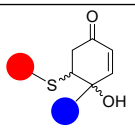
HPLC trace of the crude reaction of **3l**

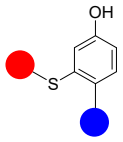
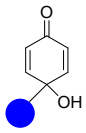


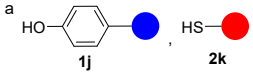
HRMS spectrum of the crude reaction of **3l**

Figure S9. The HPLC and MS analysis of the crude reaction mixture of **3l**

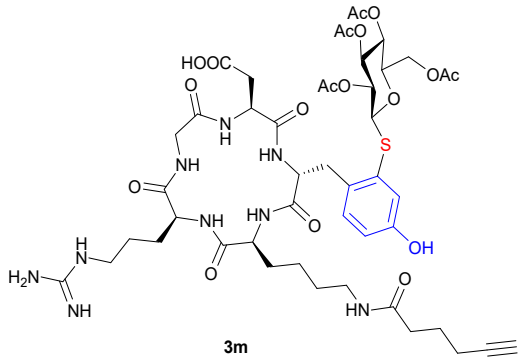
Table S1. Analysis of the crude reaction of **3l**

Retention time	**Structure**[a]	**ESI-MS (m/z)**
5.83 min		Found: 970.6365, $[M + H]^+$
6.07 min		Found: 1939.2411, $[M + H]^+$
6.41 min		Found: 1335.8052, $[M + 2H]^{2+}$
6.61 and 6.70 min		Found: 850.9963, $[M + 2H]^{2+}$

6.88 min		Found: 1681.96631, $[M + H]^+$
7.30 min		Found: 730.3514, $[M + H]^+$
18.12 min		

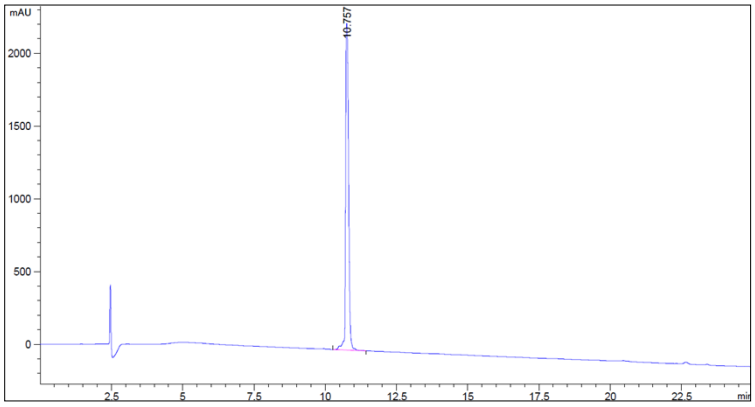
a 

a)



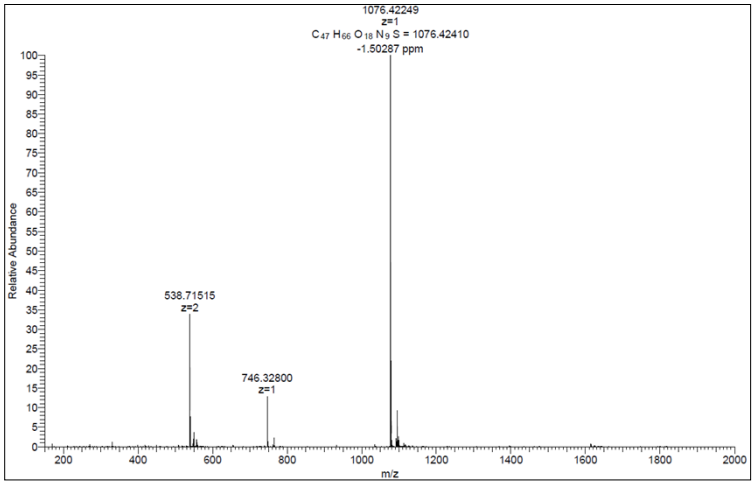
3m

b)



HPLC trace of peptide **3m**

c)



HRMS spectrum of peptide **3m**

Figure S10. a) Chemical structure of peptide **3m**, 43% isolated yield; b) Analysis of HPLC trace of peptide **3m**. (HPLC gradient is 10% to 100% of solution B in 30 min on the Analysis C18 column (λ = 220 nm); c) ESI-MS: m/z calculated for $C_{47}H_{65}O_{18}N_9S$: 1076.42410, $[M + H]^+$. Found: 1076.42249, $[M + H]^+$; 538.71515, $[M + 2H]^{2+}$.

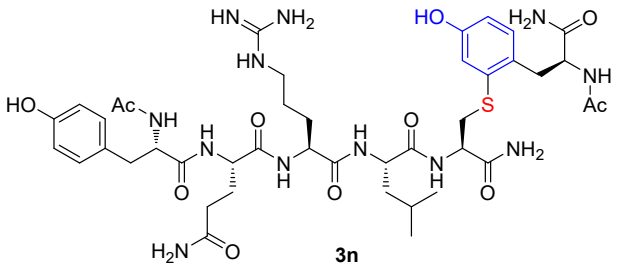
3.4 General procedure for the synthesis of 3n, 3o and 3p.

3n and **3o**: To a solution of **1** (0.01 mmol, 2 eq.) in MeCN/H_2O (1/1, 0.2 ml) was added $PhI(OAc)_2$ (2 eq.) at 0°C. The mixture was stirred at rt for 30 min before KOH (3.3 eq.) and **2i** (0.005 mmol, 1 eq.) was added respectively. After stirring at rt overnight, the reaction mixture was purified by semi-preparative RP-HPLC to give desired product **3n** and **3o**.

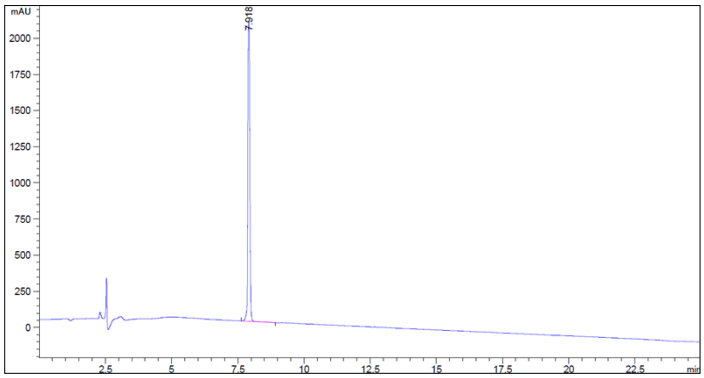
3p: To a solution of **1a** (0.005 mmol, 10 eq.) in MeCN/H_2O (1/1, 0.2 ml) was added $PhI(OAc)_2$ (10 equiv) at 0°C. The mixture was stirred at rt for 45 min before KOH (33 equiv) and BSA (0.0005 mmol, 1 eq.) was added respectively. After stirring at rt overnight, the reaction mixture was purified by ultrafiltration to afford the desalinated Tyr-BSA conjugate **3p**. 1 mg/ml solution of Tyr-BSA conjugate **3p** in water was prepared to accomplish the protein mass spectrometry.

3.5 Chemical structures and characterization of 3n, 3o and 3p.

a)

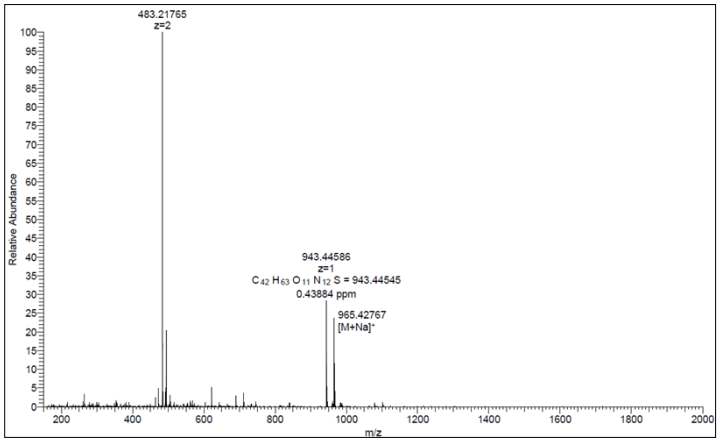


b)



HPLC trace of peptide **3n**

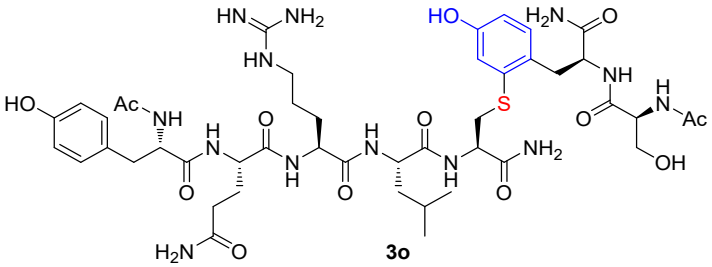
c)



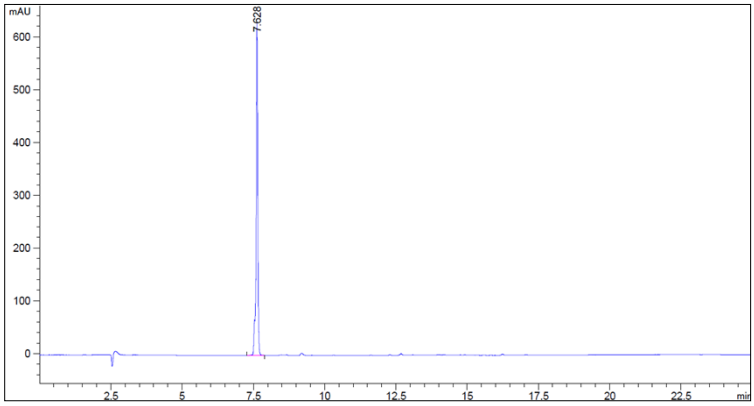
HRMS spectrum of peptide **3n**

Figure S11. a) Chemical structure of peptide **3n**, 41% isolated yield; b) Analysis of HPLC trace of peptide **3n**. (HPLC gradient is 10% to 100% of solution B in 30 min on the Analysis C18 column (λ = 220 nm); c) ESI-MS: m/z calculated for $C_{42}H_{62}O_{11}N_{12}S$: 943.44545, $[M + H]^+$. Found: 943.44586, $[M + H]^+$; 965.42767, $[M + Na]^+$; 483.21765, $[M + 2H]^{2+}$.

a)

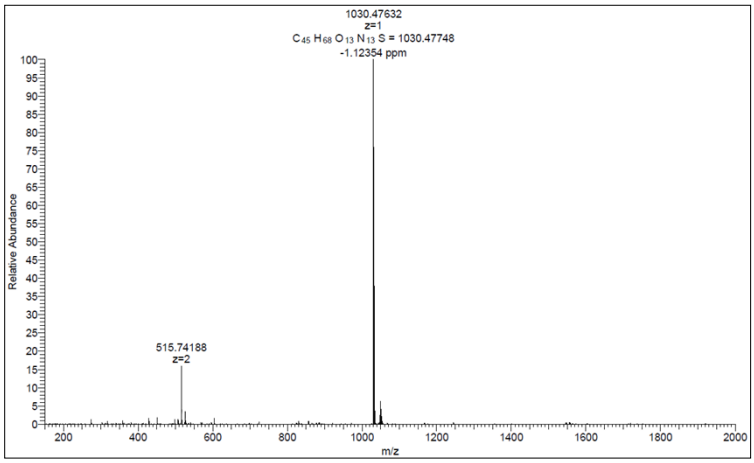


b)



HPLC trace of peptide **3o**

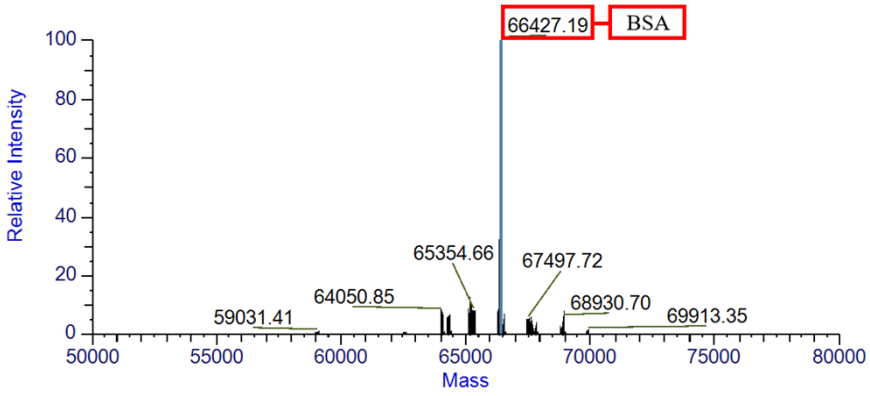
c)



HRMS spectrum of peptide **3o**

Figure S12. a) Chemical structure of peptide **3o**, 37% isolated yield; b) Analysis of HPLC trace of peptide **3o**. (HPLC gradient is 10% to 100% of solution B in 30 min on the Analysis C18 column (λ = 220 nm); c) ESI-MS: m/z calculated for $C_{45}H_{67}O_{13}N_{13}S$: 1030.47748, $[M + H]^+$. Found: 1030.47632, $[M + H]^+$; 515.74188, $[M + 2H]^{2+}$.

a)



b)

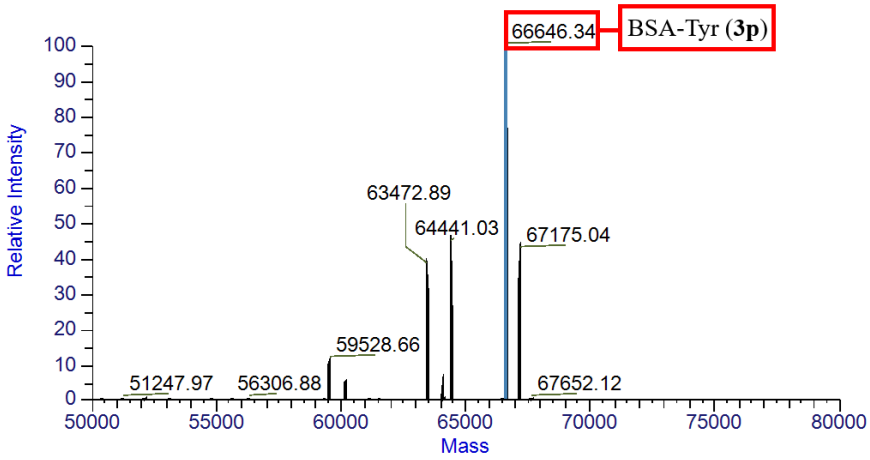
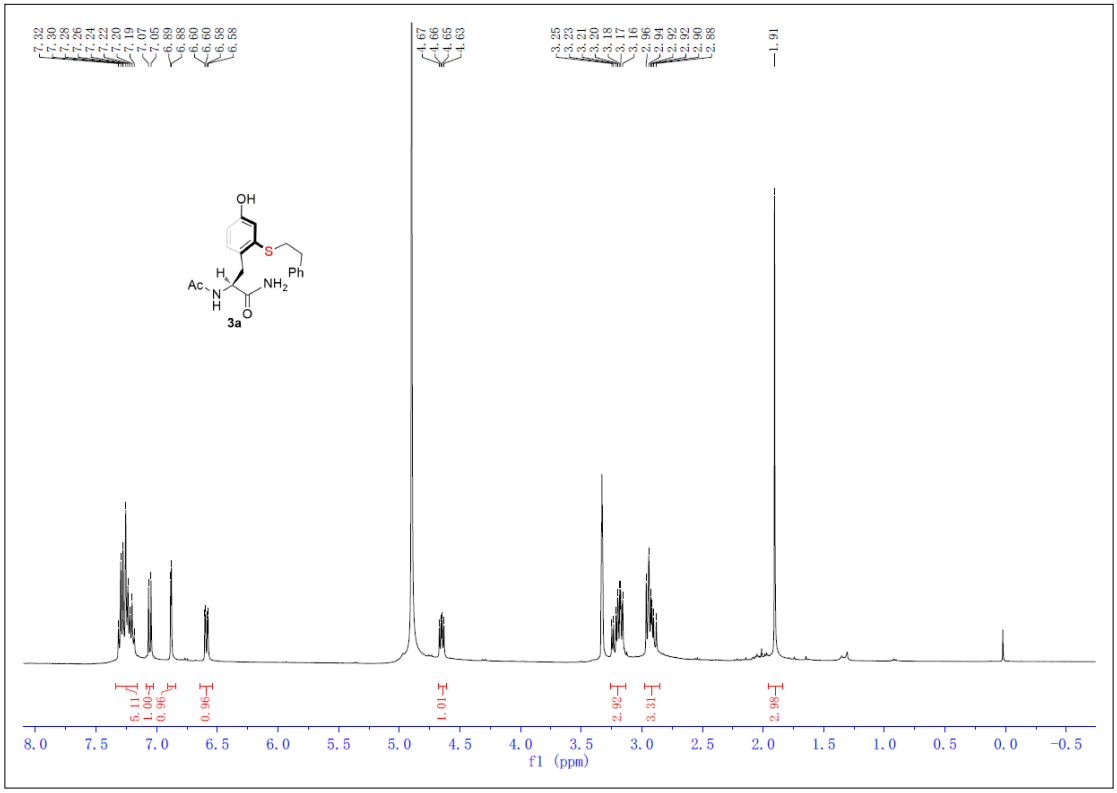
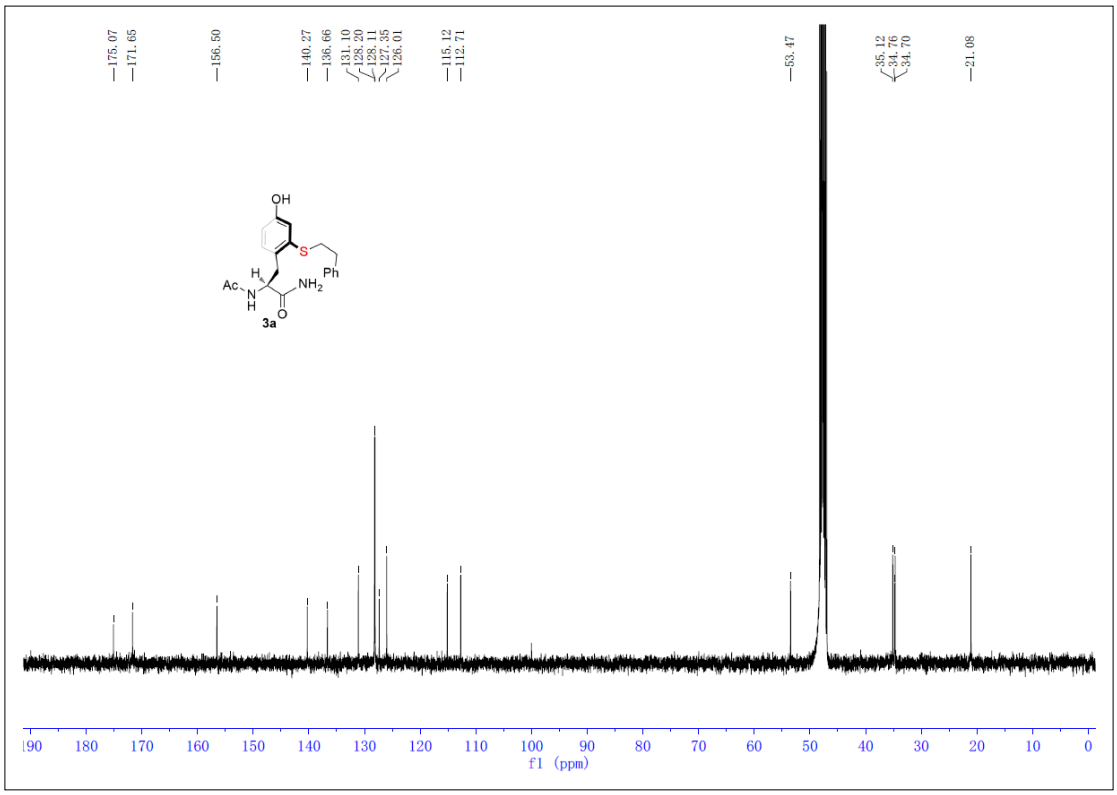


Figure S13. a) Mass spectrum of Bovine albumin (BSA). b) Mass spectrum of **3p**.

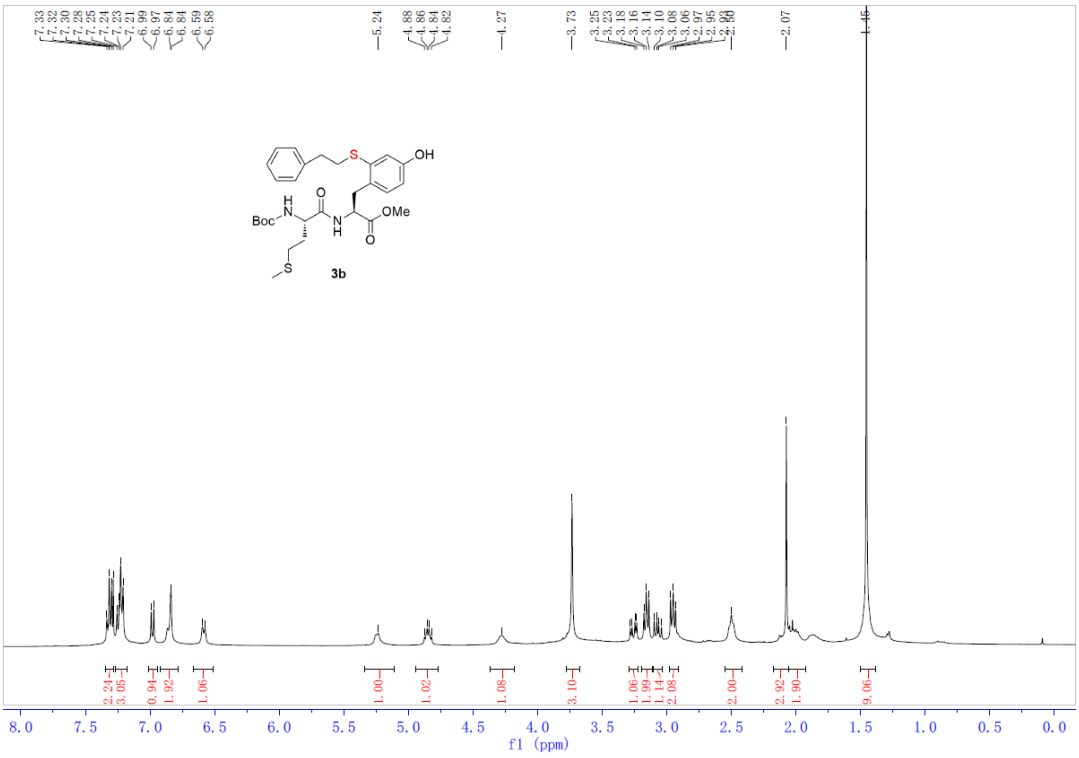
4. NMR Spectra



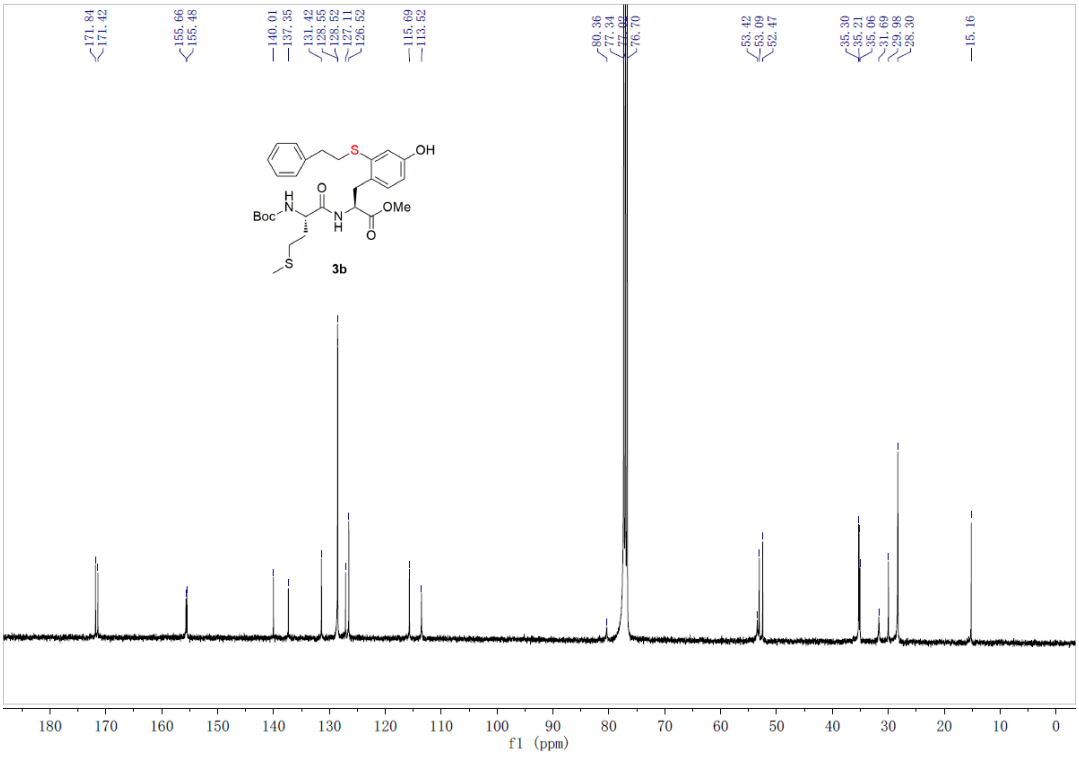
^{1}H NMR spectrum (400 MHz, CD_3OD) of compound **3a**



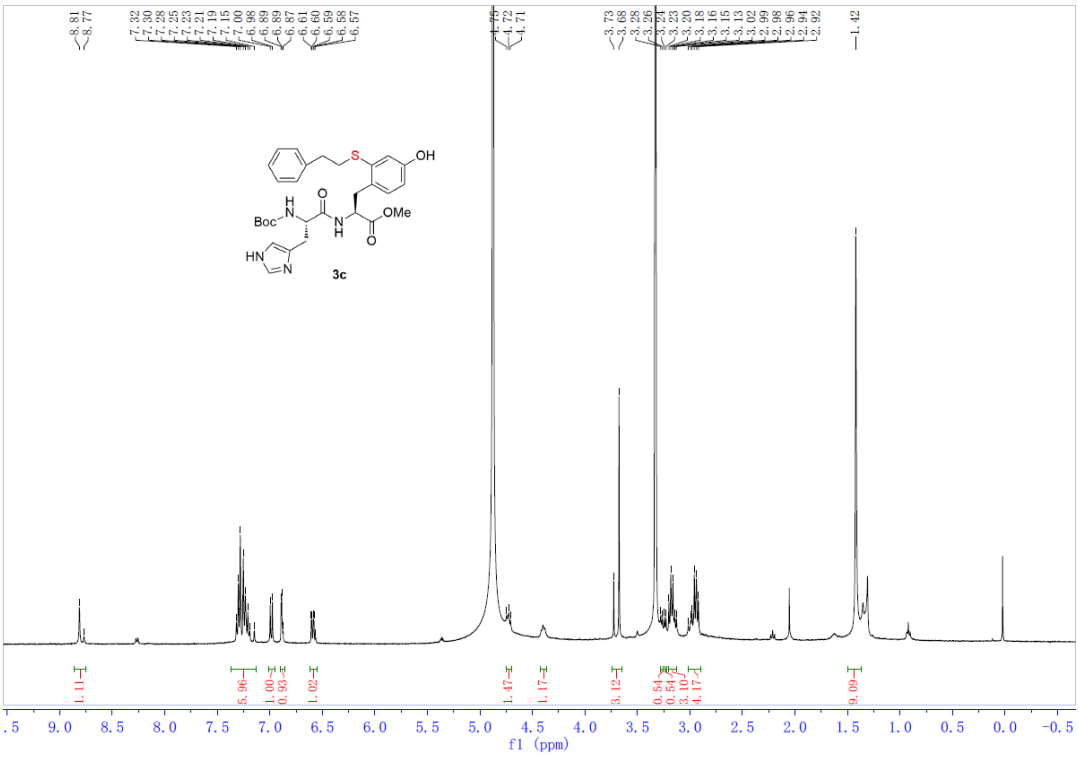
^{13}C NMR spectrum (100 MHz, CD_3OD) of compound **3a**



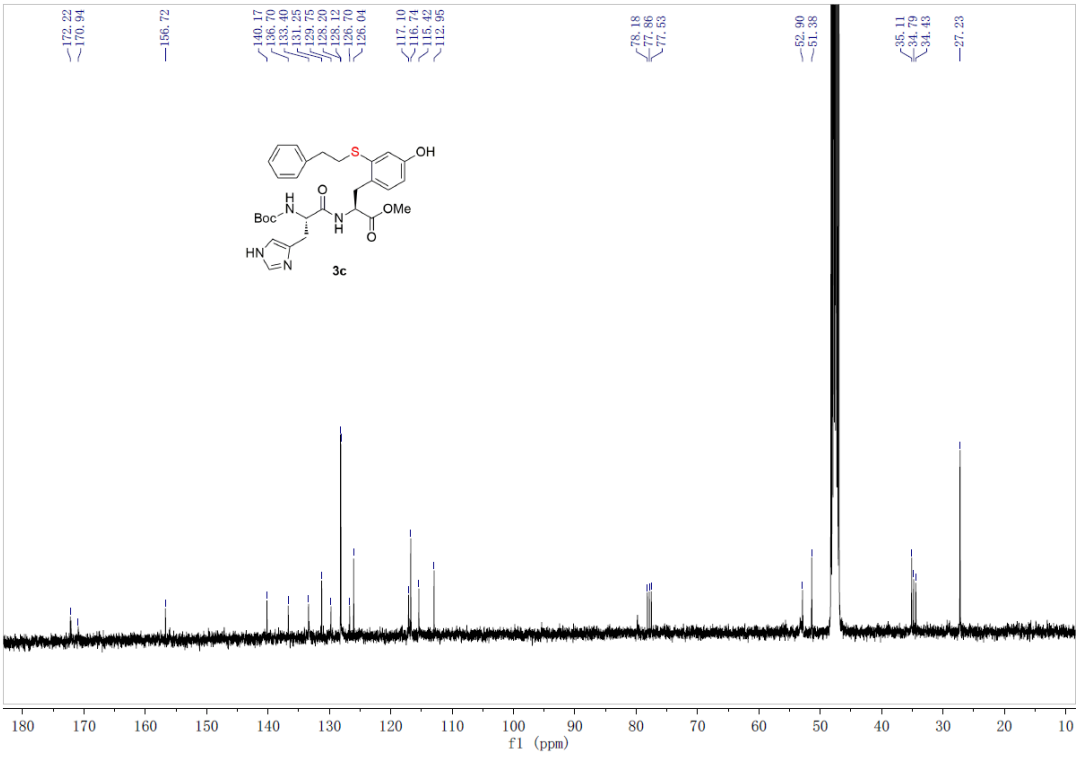
^{1}H NMR spectrum (400 MHz, $CDCl_3$) of compound **3b**



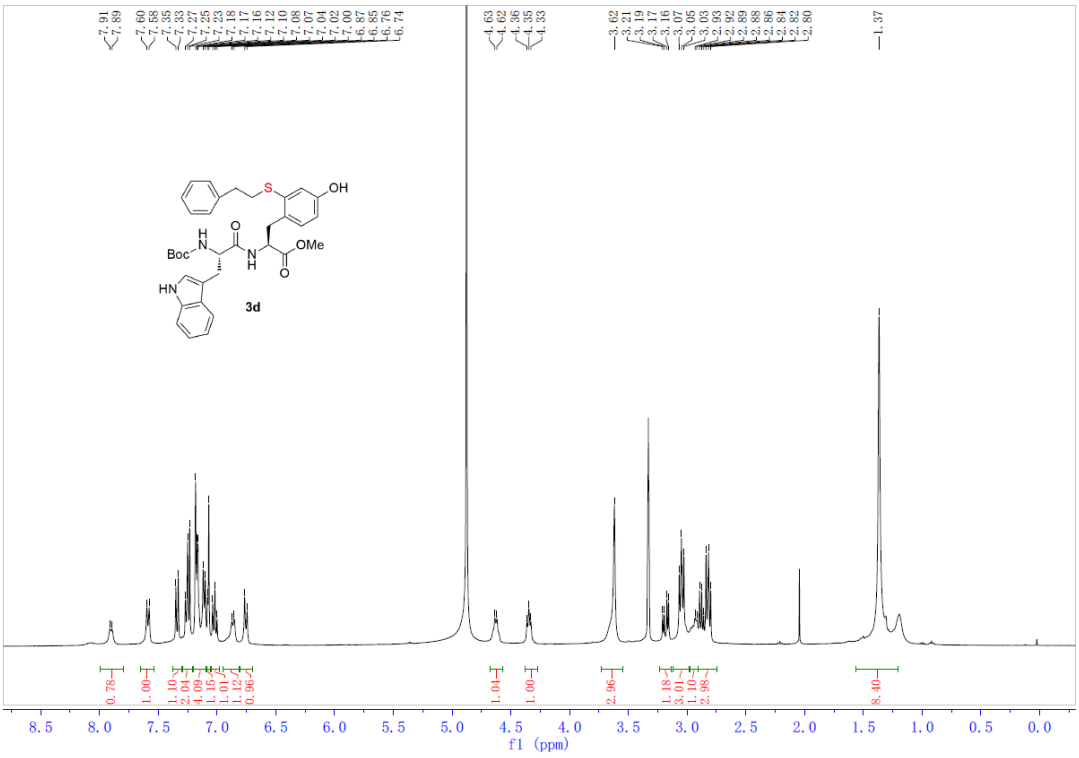
^{13}C NMR spectrum (100 MHz, $CDCl_3$) of compound **3b**



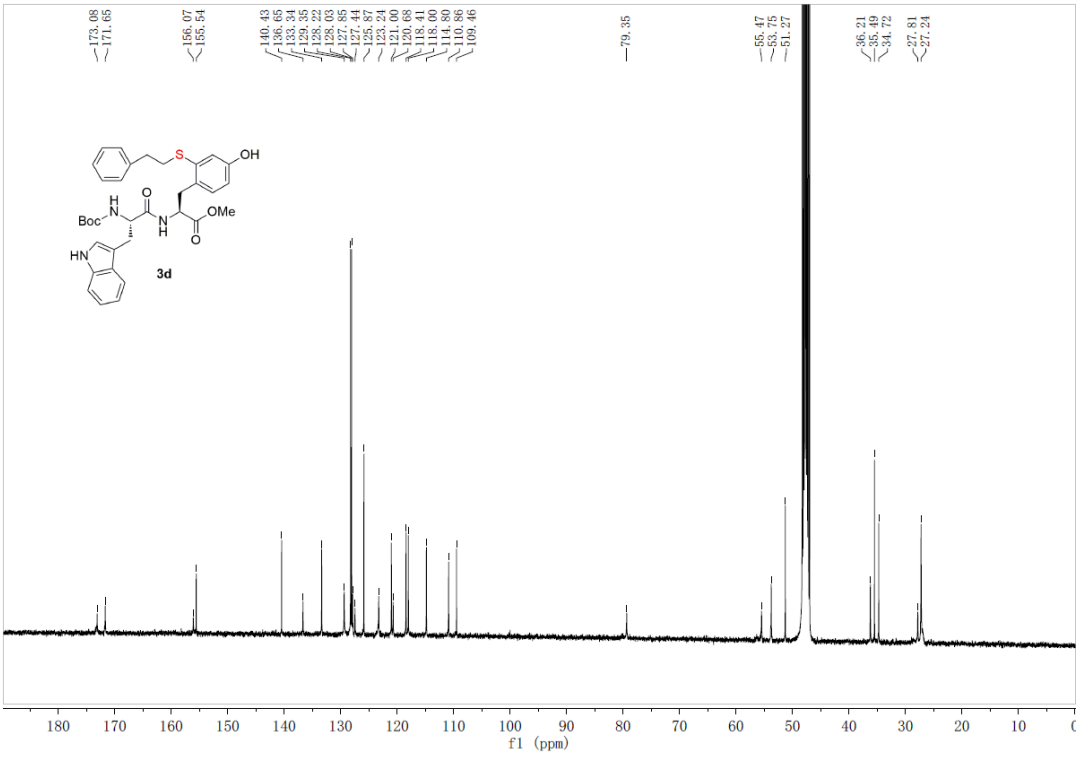
^{1}H NMR spectrum (400 MHz, CD_3OD) of compound **3c**



^{13}C NMR spectrum (100 MHz, CD_3OD) of compound **3c**



^{1}H NMR spectrum (400 MHz, CD_3OD) of compound **3d**



^{13}C NMR spectrum (100 MHz, CD_3OD) of compound **3d**