Electronic Supplementary Material (ESI) for RSC Chemical Biology.

Supporting Information

DNA encoded peptide library for SARS-CoV-2 3CL protease covalent inhibitor discovery and profiling

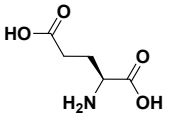
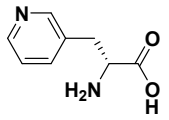
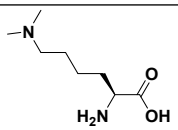
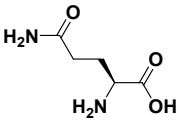
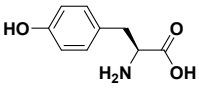
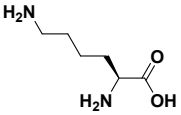
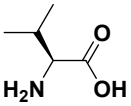
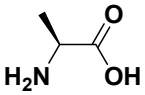
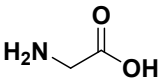
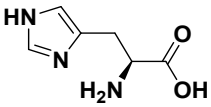
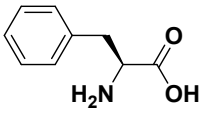
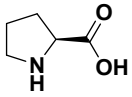
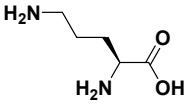
Yuyu Xing[1, 2], Huiya Zhang[1], Yanhui Wang[1], Zhaoyun Zong[1], Matthew Bogyo[3], Shiyu Chen[1, 2]*

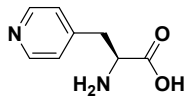
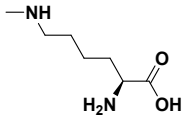
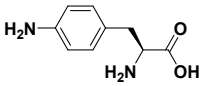
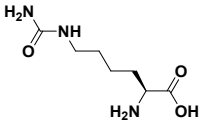
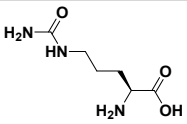
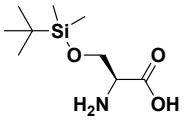
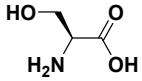
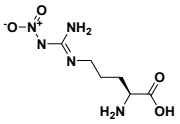
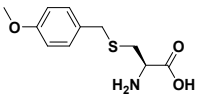
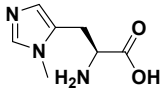
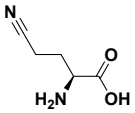
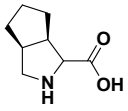
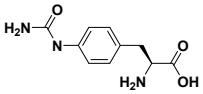
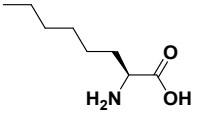
[1] Biotech Drug Research Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China
[2] University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing 100049, China
[3] Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA

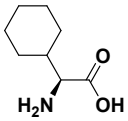
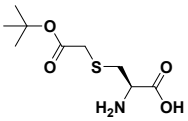
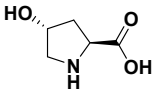
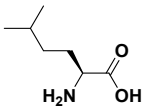
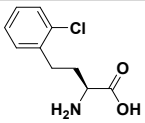
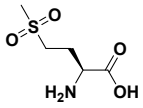
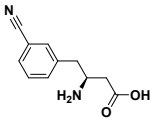
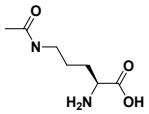
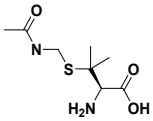
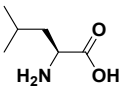
*Shiyu Chen
Email: chenshiyu@simm.ac.cn

Table of contents

Supplementary Table. 1. List of the amino acids and corresponding tag sequences utilized in the construction of DEPCIL.

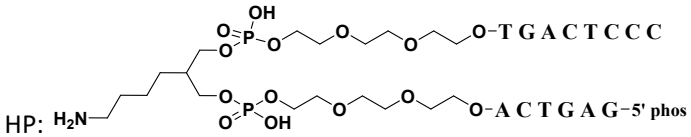
amino acid	structure	Tag1 Sequences	Tag2 Sequences	Tag3 Sequences	Tag4 Sequences
F01 Glu		AGGAGTTGTAG	GATGTGTACCA	AGAGCTGATGA	ACGAGTTCCCAA
F02 D-3-Pal		GTATTGCCAAG	AAGAGGAGTCA	AATCCTACGGA	CTAGATTCCCAA
F03 Lys(Me2)		TTCCAGAACAG	CTCCAATGACA	GATCTGGCTGA	AGCGTGCTACAA
F04 Gln		GCAGTCCAAAG	CAGTGCCTTCA	CTTACGCGAGA	CGACTTGTACAA
F05 Tyr		TGCGAACCAAG	CTGTTGCAACA	GGTATCATCGA	GAAGTCTGGCAA
F06 Lys		TACACGGAAAG	TCTTGTCGTCA	CCTTAGATCGA	TGTGGTAGCCAA
F07 Val		CAGTTGAGAAG	TCTGGACTGCA	ATTCACCAGGA	GGTCAGTTCCAA
F08 Ala		GATTGGTGAAG	ACGTGCAAGCA	TCCTACACGGA	GCCGTATAGCAA
F09 Gly		CATCGCAGTAG	GTGTCAGTCCA	CTACCGATAGA	ATAGCCAAGCAA
F10 His		TCATGGACCAG	CTCTTAGGTCA	ACGCAGTCAGA	CAAGCACCACAA
F11 Phe		AACTATGCGAG	CTGTAGACTCA	TTCGTTCGTGA	GAACTCCAGCAA
F12 Pro		TGATCGTGAAG	CCTGTTGGACA	TGCTACCGTGA	ATAACGCGACAA
F13 Orn		GGTAAGGCTAG	AAGCGATGGCA	GGTAATCTCGA	CTAACTCTGCAA

F14 4-Pal		TTACCGACGAG	ATCCAAGCACA	TGGAGAACAG A	TGCAGAAGTCAA
F15 Lys(Me)		GAGTCTAGAAG	GACGTCTGTCA	GCTATGCTAGA	CATCCGACTCAA
F16 Phe(4-NH2)		GAACTAAGAG	TTCGCTTGGCA	CTACCAACAGA	AATGGCGGACA A
F17 HoCit		TAATGCCTCAG	GAGATGTTCCA	TAGCGATGAGA	CGCTGTAAGCAA
F18 Cit		TAAGTGAGGAG	TCGACGTCTCA	GACGTACAGGA	AGACCATTGCAA
F19 Ser (TBDMS)		GCATGAACCAG	TCCAAGAAGCA	ATCGACAGAGA	ACTAGACAGCAA
F20 Ser		GCGCGATATAG	GTTGACTGGCA	CGGTATGGTGA	GTATTAGCGCAA
F21 Arg (NO_2)		CCGTATACCAG	CTGTCTGCACA	AACCTCTCCGA	ACGCACTGACAA
F22 Cys (pMeOBzl)		ACACCGGAAAG	CAACGCGATCA	CACGTCCTAGA	TACAACACCCAA
F23 His(3-Me)		CGAGAAGTGA G	GATCTACACCA	TGGTTGCTAGA	TCCGGTAGACAA
F24 l-cba		ATTGCACTGAG	AGCAAGACACA	GTCGCACTTGA	TGCCGACAACAA
F25 octa[c]pyrrol e		AGTAGCCACAG	GCACTAGTACA	TTATGAGGCGA	GCTCCGATTCAA
F26 Aph (Cbm)		CCATGACTCAG	GTTCGATCACA	TCCGAGTGAGA	GCTCGTTCACAA
F27 L-Aoc (2)		AAGACGGACA G	GGACGAATACA	CTATAGTCGGA	GAGGAGTACCA A

F28 Chg		ACATCCTACAG	GTTAGTTCGCA	GTGCACAGTGA	GGTCTCTAGCAA
F29 L-Cys (Boc-Methyl)		ACGCATACGAG	TGAGAGTGTCA	TCTTAGCGCGA	ATCCACGCTCAA
F30 Hyp		CTATCGCGTAG	GCAGTTATCCA	TATGACCTGGA	TCCGTTCAGCAA
F31 HomoLeu		GATGGTGTGAG	AGATCGTCCCA	CCTAGCTCAGA	GGCGGAATTCAA
F32 HomoPhe (2-Cl)		AACAGCAGAAG	TAAGCTGGCCA	AGTGGAACCGA	TCAACACTGCAA
F33 Met(O2)		TAGTGCCGAAG	GGATTCAGACA	CACAAGCATGA	CCTGGTCTTCAA
F34 3-cyano-phenyl		GACTGCTTGAG	TTCTCGTGACA	TTGGTAGCGGA	GGATCTGTCCAA
F35 Orn (Ac)		CCAATCAGTAG	CACCTAACACA	GCAATCTCAGA	TGTCGTTGGCAA
F36 Pen (Acm)		CGTGATGTAAG	TCTGATGCCCA	CTCAGATCGGA	GTGCGCATACAA
F37 Leu		TACAGAAGCAG	CGTCAGTACCA	GCGTTGCATGA	TGGAACATCCAA

As synthetic building blocks, all amino acids were N-terminally protected with Fmoc. The primary and secondary amino groups in the side chains were protected with a tert-butyloxycarbonyl (Boc) group. The carboxylic acid groups in the side chains were protected with Allyl oxycarbonyl group (OAll), and the hydroxyl groups were left unprotected. The urea group in the side chain of citrulline (Cit & HoCit) was not protected.

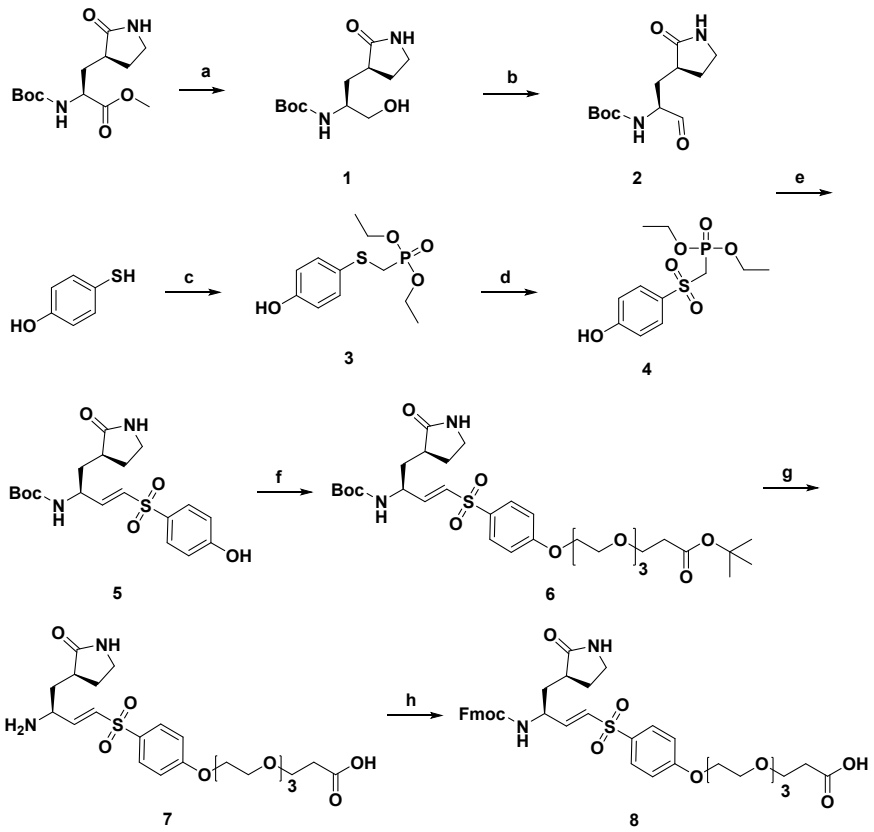
Supplementary Table. 2. Sequences of adapting DNA used in the generation of DEPCIL



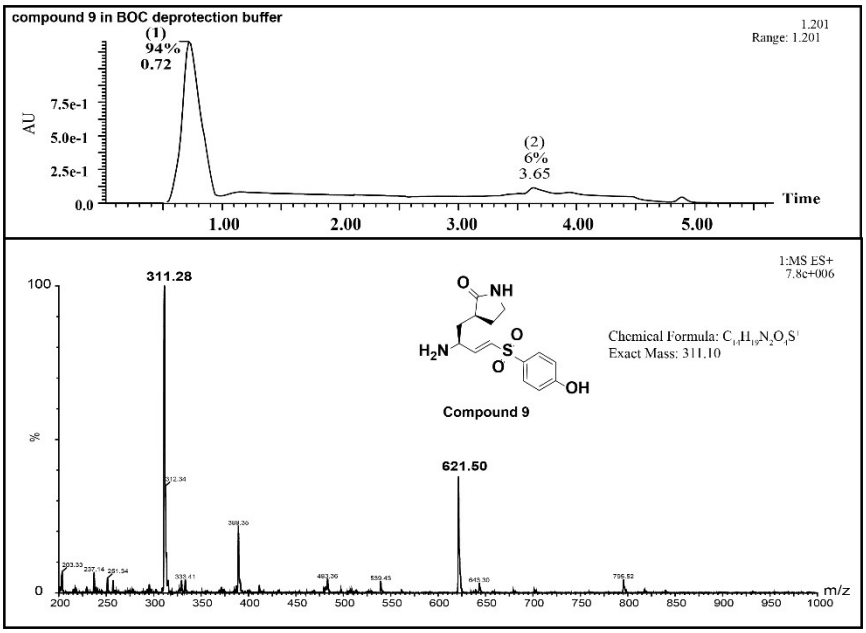
Sequences of Headpiece	5’-/5Phos/GAATCA/iSp9/iUniAmM/iSp9/TGACTCCC-3’.
Sequences of primer	5’-/5Phos/AAATCGATGTG/GGTTTAGCTAC/5Phos/5’.
Sequences of tag 1	5’-/5Phos/XXXXXXXXXAG/ACXXXXXXXXX/5Phos/5’.
Sequences of tag 2	5’-/5Phos/XXXXXXXXXCA/CTXXXXXXXXX/5Phos/5’.
Sequences of tag 3	5’-/5Phos/XXXXXXXXXGA/ACXXXXXXXXX/5Phos/5’.
Sequences of tag 4	5’-/5Phos/XXXXXXXXXCAA/TCXXXXXXXXX/5Phos/5’.
Sequences of closing tag	5’-/5Phos/TGAGCCGACTCTAAGTGCATTGCAC /GCAATGCACTTAGAGTCGGCTCATTG/5Phos/5’.

Supplementary Table 3. LCMS analysis of peptide inhibitors synthesized in this work

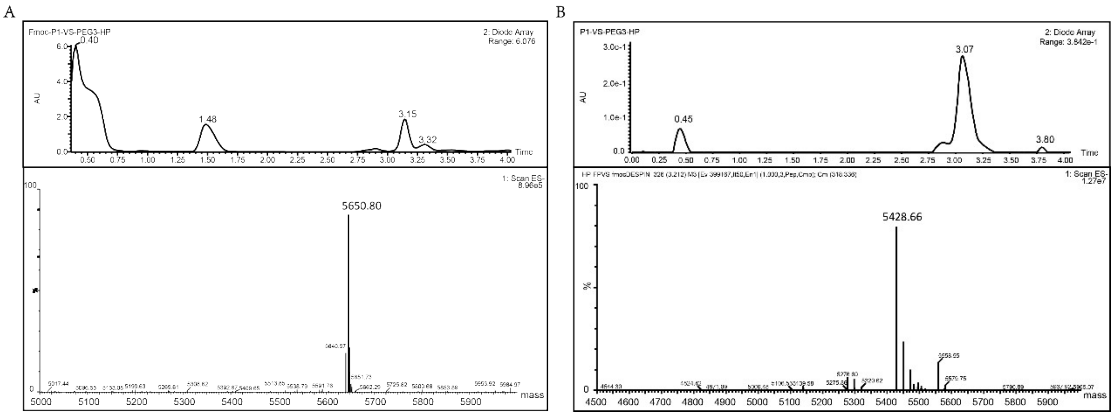
Name	Sequence	Calc. WT	Exp. m/z	Purity %
P3C-1	Gln-Val-Orn-Leu-VS	764.39	765.8	96%
P3C -2	Val-Pro-Aoc2-Phe-VS	794.40	795.8	94%
P3C -3	HoCit-Pen(Acm)-Aoc2-Val-VS	923.46	924.8	91%
P3C -4	Leu-Chg-Gly-Aoc2-VS	760.42	761.3	94%
P3C -5	4-Pal-HoCit-Ser-Ser-VS	803.33	804.7	94%
P3C -6	Leu-Ser-Met(O2)-Ser-VS	760.28	762.1	95%
P3C -7	HPhe(3-CN)-His-Orn-Leu-VS	860.40	861.8	95%
P3C -8	4-Pal-Lys-Glu-Ser-VS	842.36	843.7	97%
P3C -9	Aoc2-Ser-Gln-Aoc2-VS	807.42	809.8	94%
P3C -10	Aoc2-Cys(pMeOBzl)-Lys(Me)2-Cit-VS	987.49	989.0	93%
P3C -11	Gln-D-Val-Orn-Phe-4NH_2-VS	813.38	814.7	95%
P3C -12	Gln-Val-Orn-Met(O2)-VS	814.34	815.7	94%
P3C -13	Gln-Val-Orn-Ala-VS	722.34	723.8	97%
P3C -14	Gln-Val-Lys(Me)2-Leu-VS	806.44	807.7	96%
P3C -15	Gln-Val-Gln-Leu-VS	778.37	779.8	93%
P3C -16	Gln-Val-4-Pal-Leu-VS	798.37	799.8	91%
P3C -17	Gln-Val-Aoc2-Leu-VS	791.43	792.9	97%
P3C -18	Gln-D-3-Pal-Orn-Leu-VS	813.38	814.8	92%
P3C -19	Gln-Lys(Me2)-Orn-Leu-VS	821.45	822.8	94%
P3C -20	Gln-Orn(Ac)-Orn-Leu-VS	821.41	822.8	95%
P3C -21	Gln-Gly-Orn-Leu-VS	722.34	723.8	90%
P3C -22	Leu-Val-Orn-Leu-VS	749.41	750.8	94%
P3C -23	Ser-Val-Orn-Leu-VS	723.36	724.8	94%
P3C -24	Lys-Val-Orn-Leu-VS	764.43	765.8	96%
P3C -25	Gly-Val-Orn-Leu-VS	693.35	694.8	94%
P3C -26	Pro-Val-Orn-Leu-VS	733.38	734.8	96%
P3C -27	Orn-Val-Orn-Leu-VS	750.41	751.8	96%
S3C-1	Ser-Val-Ala-Leu-VS	680.32	682.1	95%



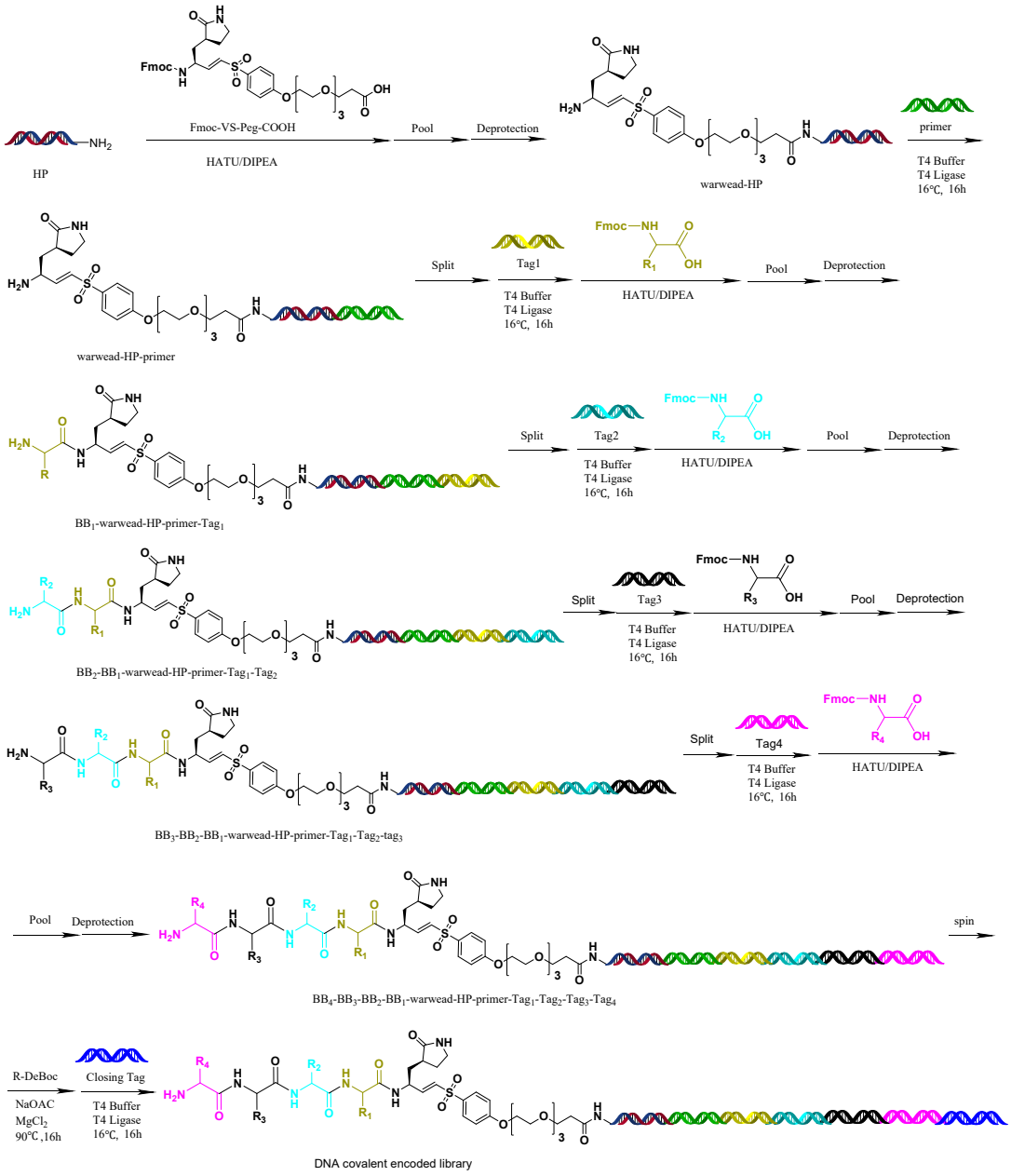
Supplementary Fig. 1. Synthetic procedure for Fmoc-(3S)-γ-lactam-VS-PEG-COOH. Reaction condition: (a) $NaBH_4$; (b) DMP, DCM; (c) NaH, THF, rt; (d) m-CPBA, DCM, rt; (e) NaH, THF, rt; (f) K_2CO_3, DMF, rt; (g) DCM: TFA=1:1, rt; (h) Na_2CO_3, H_2O, Fmoc-Cl, 1,4-Dioxane.



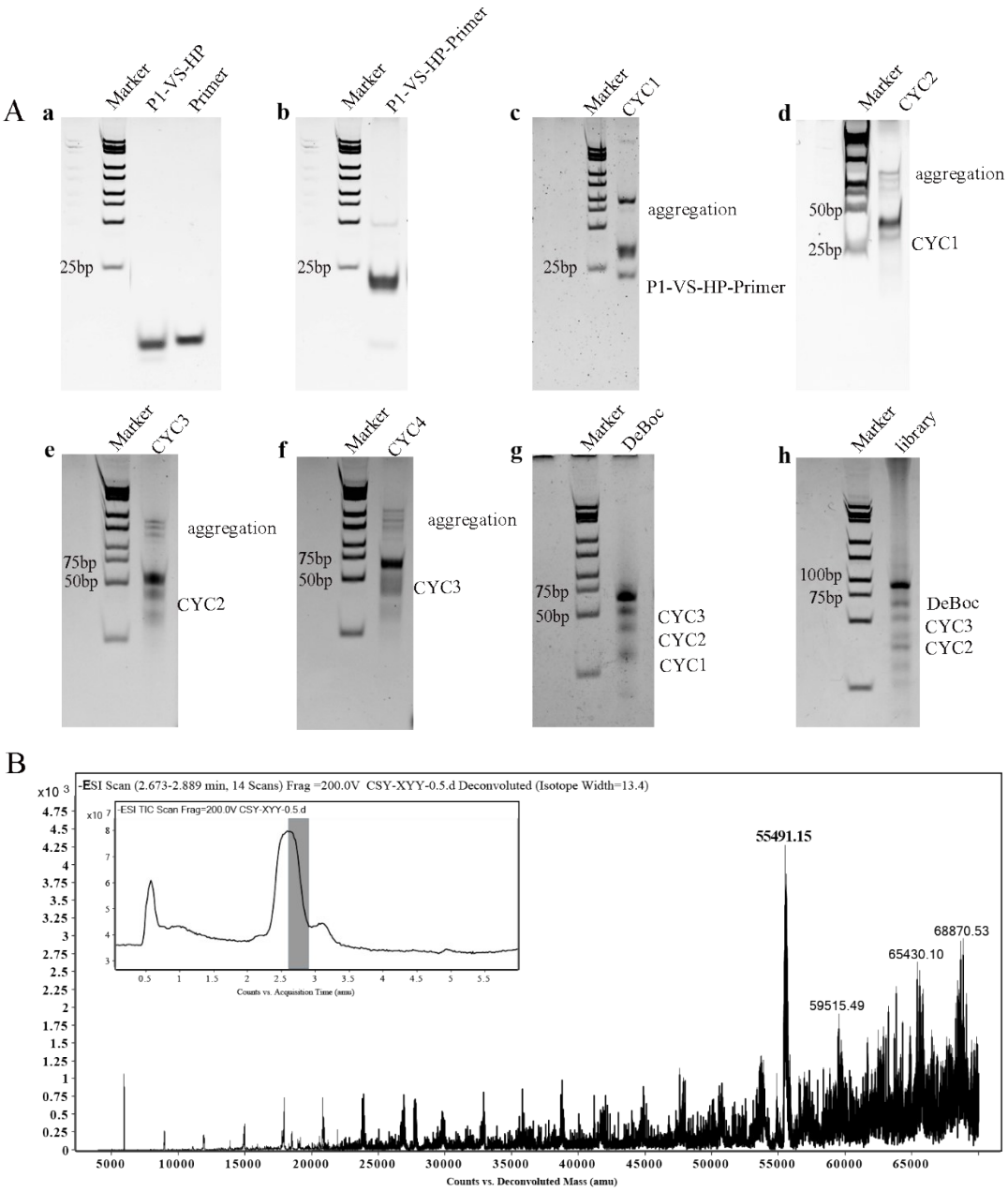
Supplementary Fig. 2. Determination of the stability of P1-VS. Compound 9 was incubated in the BOC deprotection buffer at 95°C for 16 hours, and LC-MS analysis of the product was performed. The results indicated that Compound 9 is stable, and the MS profile remained unchanged.



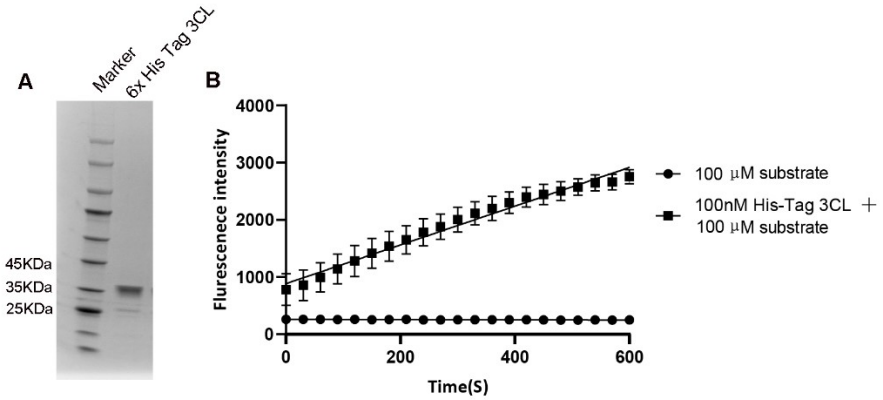
Supplementary Fig. 3. LCMS analysis of Fmoc deprotection. (A) LCMS analysis of Fmoc-P1-VS-PEG3-HP. (B) LCMS analysis of P1-VS-PEG3-HP.



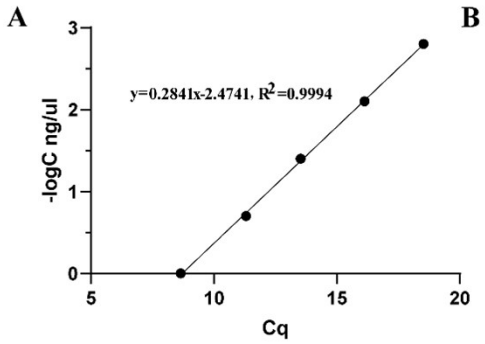
Supplementary Fig. 4. Procedure of DEPCIL construction. DEPCIL construction involves several cycles, including DNA splitting in a 96-well plate, DNA tag ligation, amino acid coupling, pooling, Fmoc deprotection, precipitation, and ultrafiltration purification. Following the deprotection of BOC from side chains, the library is then conjugated with the closing tag.



Supplementary Fig. 5. Analysis of intermediates for DEPCIL construction. (A) PAGE analysis. a). P1-VS-HP, 7bp. b). P1-VS-HP -Primer, 18bp. c). cycle1, AA1-P1-VS-HP -Primer-Tag1, 29bp. d). cycle2, AA2-AA1-P1-VS-HP -Primer-Tag1-Tag2, 40bp. e) cycle3, AA3-AA2-AA1-P1-VS-HP -Primer-Tag1-Tag2-Tag3, 51bp. f). cycle4, AA4-AA3-AA2-AA1-P1-VS-HP -Primer-Tag1-Tag2-Tag3-Tag4, 62bp. g) PAGE analysis of BOC deprotection, 62bp. h). final DEPCIL, AA4-AA3-AA2-AA1-P1-VS-HP -Primer-Tag1-Tag2-Tag3-Tag4-closing Tag, 87bp. (B) LCMS analysis of final DEPCIL in the format of AA4-AA3-AA2-AA1-P1-VS-HP -Primer-Tag1-Tag2-Tag3-Tag4-closing Tag.



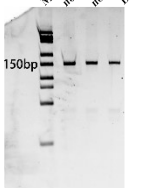
Supplementary Fig. 6. SDS-PAGE analysis of the 6x His-tagged 3CL protease and its hydrolysis activity in cutting a fluorogenic substrate generated from the native peptide sequence. (A) SDS-PAGE analysis of 6x His-tag 3CL protease (35.167 KDa). (B) Activity assay of the achieved 3CL protease for cutting 100μM substrate.



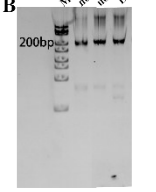
sample	Cq	c(ng/μl)
Protein-free control group	18.874	0.00194
Screening group	17.107	0.00411

Supplementary Fig. 7. Plot for the standard curve of qPCR analysis: (A) The plot of Cq value versus DNA concentration. The Cq value of qPCR is shown on the X-axis, and the negative logarithm of DNA concentration is shown on the Y-axis. qPCR was performed with DNA concentrations of 1, 0.2, 0.04, 0.008, and 0.0016 ng/μl, and the relationship between Cq value and DNA concentration was calculated. (B) The table of qPCR results comparing the screening and control groups.

A



B



C

Category	order	Sequences
Expected format		XXXXXXXXXAGXXXXXXXXXGTXXXXXXXXXGAXXXXXXXXXCAA
Actual format	1	TACAGAAGCAGAAGCGATGGGTATTCACCAGGACGACTTGTACAA
	2	AACTATGCGAGCGACGAATAGTTGCTACCGTGAGGTCAGTTCCAA
	3	CAGTTGAGAAGGGACGAATAGTCTCAGATCGGACGCTGTAAGCAA

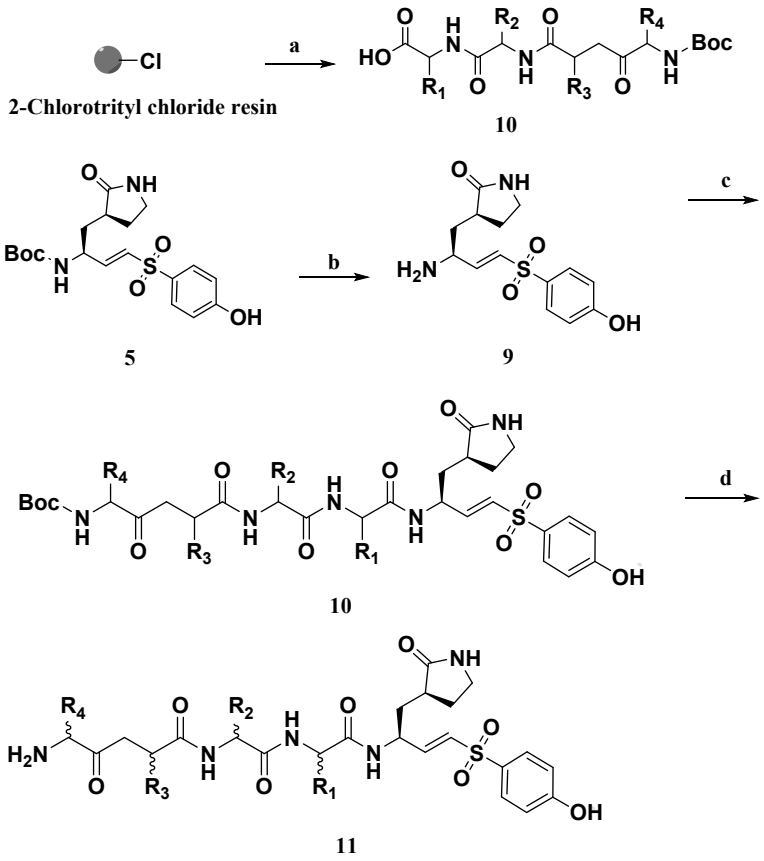
D

Sequences	BB4	BB3	BB2	BB1	Frgc. %
1	Gln	Val	Orn	Leu	0.00283
2	Val	Pro	Aoc_2	Phe	0.00248
3	HoCit	Pen(Acm)	Aoc_3	Val	0.00166
4	Leu	Chg	Gly	Aoc_2	0.00160
5	4-Pal	HoCit	Ser	Ser	0.00152
6	Leu	Ser	$Met(O_2)$	Ser	0.00152
7	HPhe(3-CN)	His	Orn	Leu	0.00150
8	4-Pal	Lys	Gln	Ser	0.00150
9	Aoc_2	Ser	Gln	Aoc_2	0.00147
10	Aoc_2	Cys(pMeOBzl)	$Lys(Me)_2$	Cit	0.00144

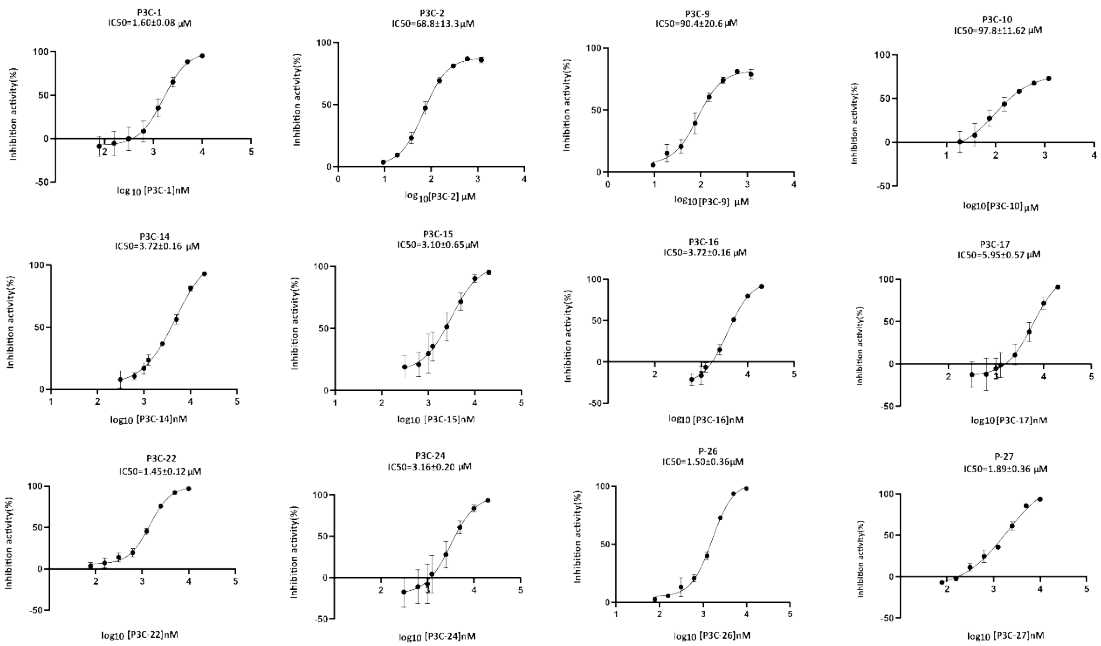
E

Sequences	BB4	BB3	BB2	BB1	Frgc. %
1	HoCit	Lys(Me,Boc)	Cys(pMeOBzl)	HomoLeu	0.00141
2	HoCit	Ser(TBDMS)	$Arg(NO_2)$	4-Pal	0.00138
3	Lys(Boc)	I-eba	$Met(O_2)$	4-Pal	0.00132
4	L-Cys(Boc-Methyl)	Pro	Phe	HomoLeu	0.00128
5	4-Pal	D-3-Pal	Gly	Ser(TBDMS)	0.00128
6	Lys(Me,Boc)	HomoLeu	His(Boc)	Orn(Ac)	0.00128
7	D-3-Pal	HoCit	Glu(OAll)	$Arg(NO_2)$	0.00125
8	4-Pal	Phe	HoCit	3-cyano-phenyl	0.00125
9	Cit	Gln	$Lys(Me_2)$	Hyp	0.00125
10	Gly	$Lys(Me_2)$	I-eba	Hyp	0.00125

Supplementary Fig. 8. PCR amplification of the DEPCIL library for generating samples for NGS sequencing and subsequent analysis of the discovered peptides. (A) PAGE analysis of step 1 PCR, illustrating the attachment of a pair of partial adaptors to the DEPCIL library. (B) PAGE analysis of step 2 PCR, demonstrating the attachment of a pair of complete adaptors to the DEPCIL library. (C) Presentation of DNA sequences in the expected format for the DEPCIL library and representative sequences discovered through DEPCIL screening. (D) The top 10 sequences discovered from negative screening conducted in the absence of 3CL protease. (E) The first 10 sequences discovered in the DEPCIL screening against 3CL protease.



Supplementary Fig. 9. The synthetic procedure for the discovered covalent inhibitors of the 3CL protease as VS (vinyl sulfone) conjugates. Reaction condition: (a) Fmoc-AA-OH, HBTU, HOBT, DIPEA; 20% HFIP/DCM; (b) hydrochloric acid 1,4-dioxane; (c) PyBOP, HOBT, DIPEA; (d) TFA: H_2O: Thioanisole: EDT: Phenol = 82.5: 5: 5: 2.5: 5.



Supplementary Fig. 10. Kinetic analysis of discovered covalent inhibitors of 3CL protease for determining the IC_{50}s. The corresponding sequences of inhibitors were indicated in table **3**.

Supplement Methods

Reagents

Thioanisole, phenol and 1,2-Ethanedithiol are purchased from GL Biochem or Shanghai bidepharm; 4-Mercaptophenol, Diethyl iodomethylphosphonate, 3-Chloroperbenzoic acid (m-CPBA) and Dess-Martin periodinane(DMP) have bought from Shanghai bidepharm; Commonly used solvents such as N,N-dimethyl formamide (DMF) were purchased from Sinoreagent Chemical Reagent Co., Ltd. MeCN, MeOH and other reagents used in analytical experiments were chromatographic pure and were purchased from Baker Company or Fisher Company. Other routine reagents were analytically pure, and the water used in the experiment was double distill water. Silica gel (300-400 mesh) used for column chromatography was from Sinoreagent Chemical Reagent Co., Ltd. TLC Silica gel plate (HSFG-254) was obtained from Shanghai Haohong Scientific Co., Ltd.

Stability of comound 9 P1-VS:

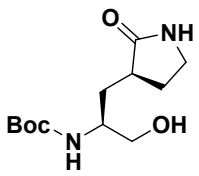
A 100 μL solution of 1 mM compound **9** (P1-VS) in 75 mM NaOAc, 1mM $MgCl_2$ was incubated in a PCR machine at 90 °C for 16 hours. At the end of the reaction, the sample was analyzed in a LCMS machine and the molecular weight of the product was detected.

Time-dependent inhibition of 3CL protease by P3C-1

In 384-well black plates, 20 μL of 3CL protease was mixed with 20 μL of different concentrations of the P3C-1 peptide (ranging from 0 to 10 μM final concentration at 2-fold serial dilutions). The plates were then incubated at 37°C for 10 minutes, 1 hour, and 12 hours. Subsequently, 20 μL of a 150 μM substrate (Dabcyl-KTSAVLQ↓SGFRKM-Edans) was added, and the fluorescent signal (excitation at 340 nm, emission at 490 nm) was detected at 1-minute intervals for a duration of 10 minutes. The acquired data were plotted against time to calculate the residual activity of the 3CL protease and analyzed using GraphPad Prism software version 9.0. The data were fitted to the equation Y = Bottom + (Top-Bottom) / (1 + 10^((Log IC_{50}-X) * HillSlope)) to calculate the IC_{50} values.

Synthesis of Fmoc-P1-VS-PEG-COOH

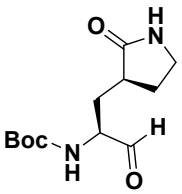
Synthesis of tert-butyl ((S)-1-hydroxy-3-((S)-2-oxopyrrolidin-3-yl) propan-2-yl) carbamate (1):



To a solution of (S)-Methyl 2-((tert-butoxycarbonyl)amino)-3-((S)-2-oxopyrrolidin-3-yl)propanoate (2.0 mmol) in dry THF, NaBH4 (0.45 g, 12 mmol) was added portion-wise at 0°C, followed by the addition of CH3OH (2 mL). The reaction mixture was then stirred at room temperature for 3 hours. The completion of the reaction was confirmed by TLC, and then the reaction was quenched with saturated NH_4Cl solution (20 mL). The reaction mixture was extracted with ethyl acetate (50 mL × 3), and the organic layers were combined and washed with saturated NH_4Cl solution (50 mL × 3) and brine (50 mL × 3). The organic phase was dried over Na_2SO_4, concentrated, and the residue was purified by column chromatography (DCM:MeOH = 20:1 v/v) to afford the pure product 1 (0.82 g, 90%) as a light solid. ^{1}H NMR (600 MHz, DMSO-d6) δ 7.53 (s, 1H), 6.56 (d, J = 9.0 Hz, 1H), 3.45 – 3.39 (m, 1H), 3.31 – 3.30 (m, 1H), 3.25 – 3.19 (m, 1H),

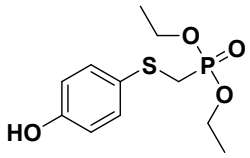
3.15 – 3.14 (m, 1H), 3.13 – 3.10 (m, 2H), 2.24 – 2.11 (m, 1H), 1.68 (ddd, *J* = 13.7, 11.3, 3.3 Hz, 2H), 1.64 – 1.54 (m, 2H), 1.37 (s, 9H). LCMS: m/z $[M+H]^+$ calcd for $C_{12}H_{23}N_2O_4^+$ 259.17; found 259.20.

Synthesis of tert-butyl ((S)-1-hydroxy-3-((S)-2-oxopyrrolidin-3-yl) propan-2-yl) carbamate (2):



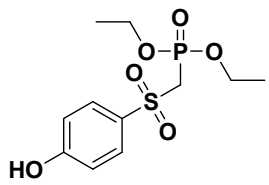
To a solution of compound 1 (1.0 mmol) in DCM, DMP (0.55 g, 1.2 mmol) was added slowly, and the reaction mixture was stirred at room temperature. Upon consumption of the reactant, the reaction was filtered and washed with saturated Na2S2O3 solution (50 mL ×3). The reaction mixture was then extracted with ethyl acetate (50 mL ×3), and the organic layers were combined and washed with saturated NaHCO3 solution (50 mL ×3) and brine (50 mL ×3). The organic phase was dried over Na2SO4, concentrated, and the residue was purified by flash column chromatography (DCM: MeOH = 15:1 v/v) to afford the pure product 2 (0.34 g, 76%) as a colorless oily liquid. ^{1}H NMR (600 MHz, Chloroform-d) *δ* 9.56 (d, *J* = 3.9 Hz, 1H), 6.10 (d, *J* = 5.6 Hz, 1H), 5.69 (s, 1H), 4.20 (d, *J* = 9.1 Hz, 1H), 3.38 – 3.33 (m, 2H), 2.51 – 2.42 (m, 1H), 2.06 – 1.97 (m, 2H), 1.95 – 1.82 (m, 2H), 1.45 (s, 9H). LCMS: m/z $[M+Na]^+$ calcd for $C_{12}H_{20}N_2NaO_4^+$ 279.13; found 279.11.

Synthesis of 4-hydroxy-thiophenyl-methyl-diethylphosphonate sulfone (3):



To a dried flask flushed with N_2, 4-hydroxythiophenol (2.52 g, 20.0 mmol) was added in 30 mL THF at 0°C. Sodium hydride (0.84 g, 21 mmol) was slowly added, leading to the generation of gas. The reaction was stirred at room temperature for 30 min, and diethyl iodomethyl phosphonate (5.56 g, 20 mmol) was added as a solution in 10 mL THF. The reaction was stirred at room temperature for 4 hours and quenched with 20 mL of 5% potassium hydrogen sulfate. The resulting aqueous phase was extracted with ethyl acetate (100 mL ×3). The organic layers were washed with saturated $NaHCO_3$ solution (50 mL ×3) and brine (50 mL ×3). The residue was purified by flash column chromatography (PE:EA = 5:1 v/v) to afford the pure product 3 (4 g, 58%) as a yellow oil. ^{1}H NMR (600 MHz, Chloroform-*d*) *δ* 8.28 (s, 1H), 7.42 – 7.37 (m, 2H), 6.74 – 6.68 (m, 2H), 4.21 – 4.12 (m, 4H), 3.03 (d, *J* = 12.4 Hz, 2H), 1.33 (t, *J* = 7.1 Hz, 6H). LCMS: m/z $[M+H]^+$ calcd for $C_{11}H_{18}O_4PS^+$ 277.07; found 277.32.

Synthesis of diethyl (((4-hydroxypheny) sulfonyl) methyl) phosphonate (4):

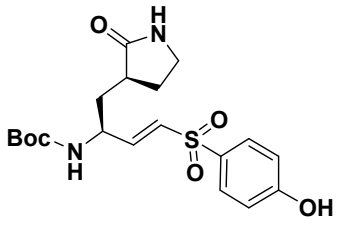


To a CH_2Cl_2 solution of **3** (4.0g，14.4 mmol) was added m-chloroperoxybenzoic acid (5.0g，28.96 mmol). The reaction mixture was stirred at room temperature for 2h and then quenched with aqueous saturated sodium thiosulfite (Na_2SO_3). The resulting solution was dilute ethyl acetate (300 mL×3) and washed with H_2O (300 mL×3) and brine (300 mL×2). The organic layer was dried with anhydrous Na_2SO_4 and concentrated in vacuo. The residue was purified by flash column chromatography (PE:EA=1:1 v/v) to afford the pure product **4** (4.5 g, 73%) as a white solid.

^{1}H NMR (600 MHz, Chloroform-*d*) *δ* 9.11 (s, 1H), 7.81 – 7.77 (m, 2H), 6.78 – 6.72 (m, 2H), 4.26 – 4.17 (m, 4H), 3.81 (d, *J* = 16.7 Hz, 2H), 1.36 (t, *J* = 7.1 Hz, 6H). LCMS: m/z $[M+H]^+$ calcd for $C_{11}H_{18}O_6PS^+$ 309.06; found 309.53.

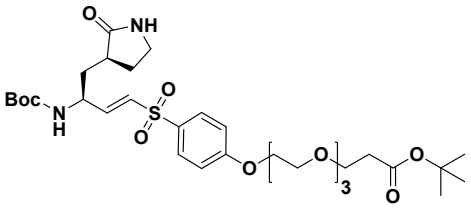
Synthesis of tert-butyl ((S, E)-4-((4-hydroxyphenyl) sulfonyl)-1-((S)-2-oxopyrrolidin-3yl) but-3-en-

2-yl) carbamate (5):



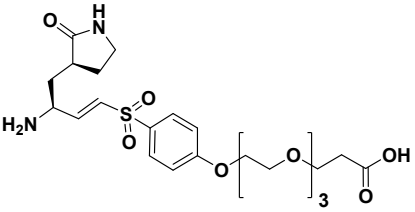
To a cold solution of compound 4 (242.3 mg, 0.8 mmol) in THF (10 mL), sodium hydride (45.2 mg, 1.2 mmol) was added under N_2. After 30 minutes, the intermediate 2 (241.9 mg, 0.9 mmol) in 10 mL THF was added and mixed at room temperature for 1.5 hours. The reaction mixture was then quenched with 20 mL of 5% $KHSO_4$ solution. The resulting solution was diluted with ethyl acetate (100 mL × 3) and washed with water (100 mL × 3) and brine (100 mL × 2). The organic layer was dried with anhydrous Na_2SO_4 and concentrated in vacuo. The residue was purified by flash column chromatography (DCM:CH_3OH = 10:1 v/v) to afford the pure product 5 (255.2 mg, 79%) as a white solid. 1H NMR (600 MHz, Chloroform-d) δ 7.64 (d, *J* = 8.4 Hz, 2H), 6.84 – 6.80 (m, 2H), 6.78 (d, *J* = 6.4 Hz, 1H), 6.47 (d, *J* = 15.6 Hz, 1H), 6.03 (d, *J* = 16.3 Hz, 1H), 5.82 (t, *J* = 7.8, 6.1 Hz, 1H), 4.36 (s, 1H), 3.42 – 3.30 (m, 2H), 2.55 – 2.47 (m, 1H), 2.00 – 1.91 (m, 1H), 1.83 (t, *J* = 10.7 Hz, 2H), 1.63 (s, 1H), 1.40 (s, 9H). LCMS: m/z $[M+H]^+$ calcd for $C_{19}H_{27}N_2O_6S^+$ 411.16; found 411.26.

Synthesis of tert-butyl 3-(2-(4-(((S, E)-3-((tert-butoxy carbonyl) amino)-4-((S)-2-oxopyrrolidin-3-yl) but-1-en-1-yl) sulfonyl) phenoxy) ethoxy) propanoate (6):



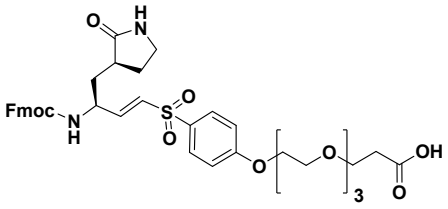
To a solution of compound 5 (185.7 mg, 0.45 mmol) in DMF (13.3 mL), K_2CO_3 (125.2 mg, 0.9 mmol) was added. After 5 minutes, Bromo-PEG3-t-butyl ester (169.4 mg, 0.5 mmol) was added, and the reaction was stirred at room temperature for 72 hours. The mixture was quenched with water and extracted with DCM (30 mL × 3) and water (30 mL × 3). The organic layer was washed with an aqueous solution of brine (30 mL × 3). The organic layers were combined, dried over Na_2SO_4, filtered, and concentrated in vacuo. The resulting crude product was purified by flash chromatography (DCM:MeOH 50:1) to afford the title compound (116.2 mg, 59%) as a yellowish oily liquid. ^{1}H NMR (500 MHz, Chloroform-d) δ 7.79 – 7.74 (m, 2H), 7.00 (d, *J* = 8.6 Hz, 2H), 6.81 (dd, *J* = 17.1, 14.6 Hz, 1H), 6.45 (dd, *J* = 17.0, 14.4 Hz, 1H), 6.10 (s, 1H), 4.38 (s, 1H), 4.18 (t, *J* = 4.8 Hz, 2H), 3.87 (t, *J* = 4.8 Hz, 2H), 3.74 – 3.69 (m, 4H), 3.69 – 3.65 (m, 2H), 3.64 – 3.59 (m, 4H), 3.50 (s, 1H), 3.34 (dd, *J* = 8.9, 6.0 Hz, 2H), 2.49 (t, *J* = 6.5 Hz, 2H), 2.44 (s, 1H), 2.05 (s, 1H), 1.96 (ddd, *J* = 15.5, 11.2, 4.5 Hz, 1H), 1.84 – 1.76 (m, 1H), 1.41 (s, 18H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 179.95, 171.05, 163.02, 155.49, 145.12, 131.85, 131.24, 130.02, 115.25, 80.69, 80.13, 77.41, 77.16, 76.91, 71.03, 70.77, 70.69, 70.50, 69.52, 68.03, 67.03, 49.90, 40.57, 38.30, 36.40, 35.56, 29.82, 28.38, 28.23. LCMS: m/z $[M+H]^+$ calcd for $C_{32}H_{51}N_2O_{11}S^+$ 671.32; found 671.64.

Synthesis of 3-(2-(4-(((S, E)-3-amino-4-((S)-2-oxopyrrolidin-3-yl) but-1-en-1-yl) sulfonyl) phenoxy) ethoxy) propanoic acid (7):



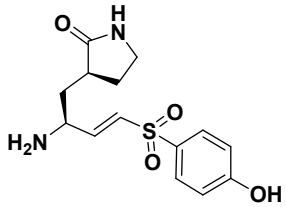
The intermediate 6 (78.2 mg, 0.1 mmol) was added to 2 mL DCM:TFA (1:1), and the reaction was stirred at room temperature for 1.5 hours. The reaction mixture was concentrated in vacuo to afford the title compound as a yellowish liquid. LCMS: m/z $[M+H]^+$ calcd for $C_{23}H_{35}N_2O_9S^+$ 515.21; found 515.20.

Synthesis of 3-(2-(4-(((S, E)-3-((((9H-fluoren-9-yl) methoxy) carbonyl) amino)-4-((S)-2-oxopyrrolidin-3-yl) but-1-en-1-yl) sulfonyl) phenoxy) ethoxy) propanoic acid (8):



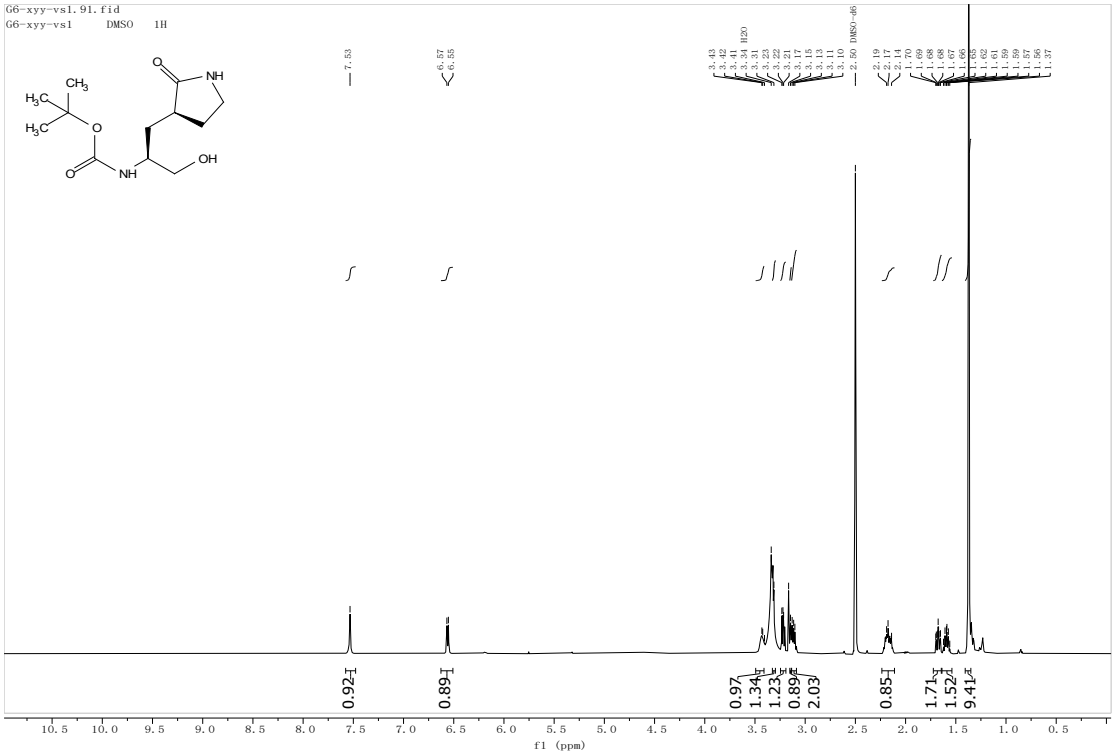
To a solution of compound 7 in H_2O (5 mL), Na_2CO_3 (37.12 mg, 0.35 mmol) was added under N_2. The solution of Fmoc-Cl (33.2 mg, 0.12 mmol) in 5 mL dioxane was added at 0°C, and the reaction was stirred at room temperature for 12 hours. The mixture was quenched with water and extracted with DCM (10 mL × 3) and water (10 mL × 3). The organic layer was washed with an aqueous solution of brine (10 mL × 3). The organic layers were combined, dried over Na_2SO_4, filtered, and concentrated in vacuo. The resulting crude product was purified by flash chromatography (DCM:MeOH 13:1) to afford the title compound (33.5 mg, 38%) as a colorless oily liquid. ^{1}H NMR (500 MHz, Chloroform-*d*) *δ* 7.75 (t, *J* = 7.4 Hz, 4H), 7.66 (d, *J* = 7.3 Hz, 1H), 7.57 (d, *J* = 7.6 Hz, 2H), 7.54 – 7.45 (m, 1H), 7.38 (t, *J* = 7.5 Hz, 2H), 7.32 – 7.27 (m, 2H), 6.98 (d, *J* = 8.5 Hz, 2H), 6.79 (dd, *J* = 20.5, 10.1 Hz, 1H), 6.47 (dd, *J* = 22.6, 15.0 Hz, 1H), 6.20 (d, *J* = 8.1 Hz, 1H), 4.42 (s, 1H), 4.34 (dd, *J* = 27.0, 7.2 Hz, 2H), 4.15 (s, 2H), 3.84 (s, 3H), 3.69 (dd, *J* = 9.5, 5.8 Hz, 4H), 3.66 – 3.56 (m, 6H), 3.34 (t, *J* = 8.8 Hz, 2H), 2.54 (t, *J* = 6.2 Hz, 2H), 2.48 – 2.42 (m, 2H), 2.38 – 2.30 (m, 1H), 1.98 (d, *J* = 20.7 Hz, 1H), 1.91 – 1.74 (m, 1H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 181.29, 174.88, 163.30, 156.54, 144.83, 144.25, 144.04, 141.68, 135.10, 134.55, 131.74, 130.31, 130.23, 129.48, 128.12, 127.48, 125.53, 124.73, 120.71, 120.36, 115.61, 77.67, 77.41, 77.16, 71.29, 71.00, 70.82, 70.67, 69.79, 68.32, 67.21, 66.91, 50.57, 47.55, 41.26, 38.71, 35.27, 30.09, 28.53. LCMS: m/z $[M+H]^+$ calcd for $C_{38}H_{45}N_2O_{11}S^+$ 737.27; found 737.34.

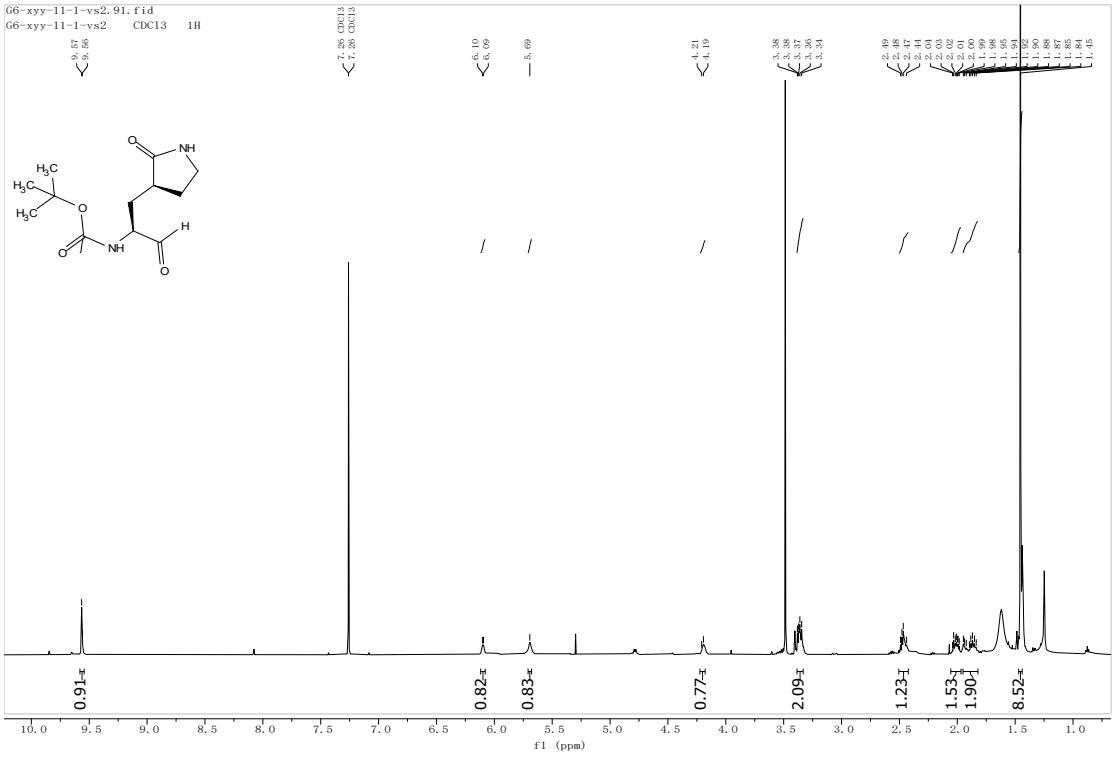
Synthesis of (S)-3-((R, E)-2-amino-4-((4-hydroxyphenyl) sulfonyl) but-3-en-1-yl) pyrrolidin-2-one hydrochloride (9):

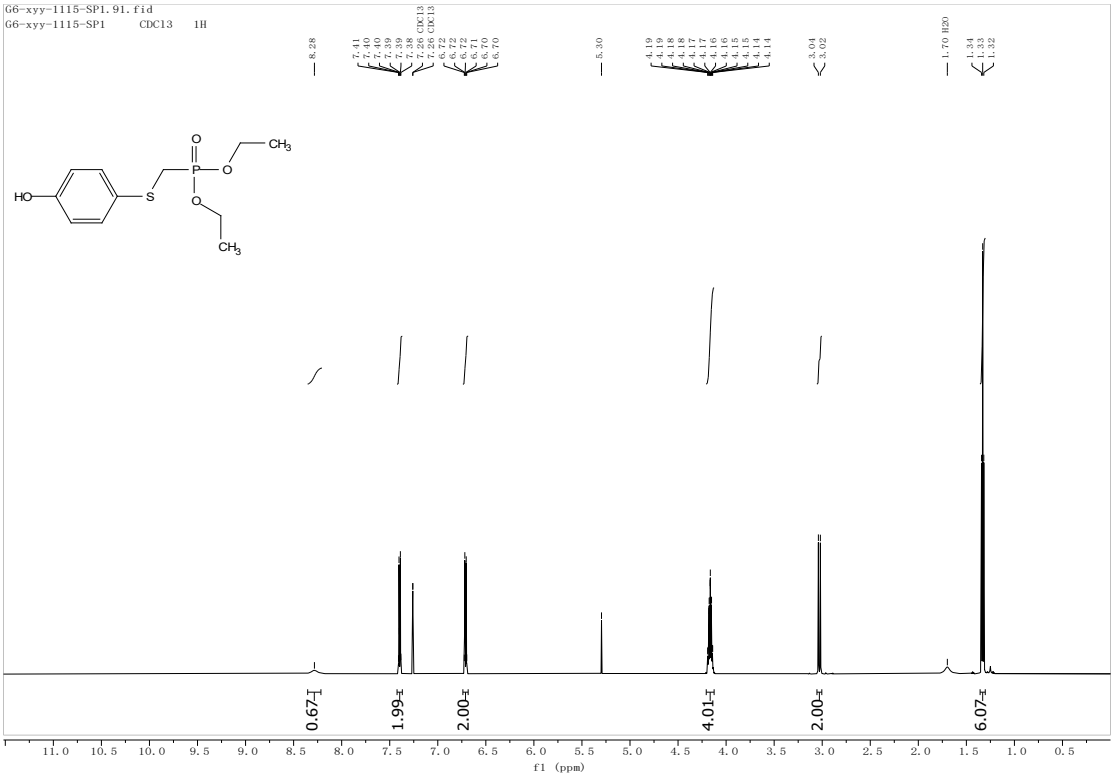


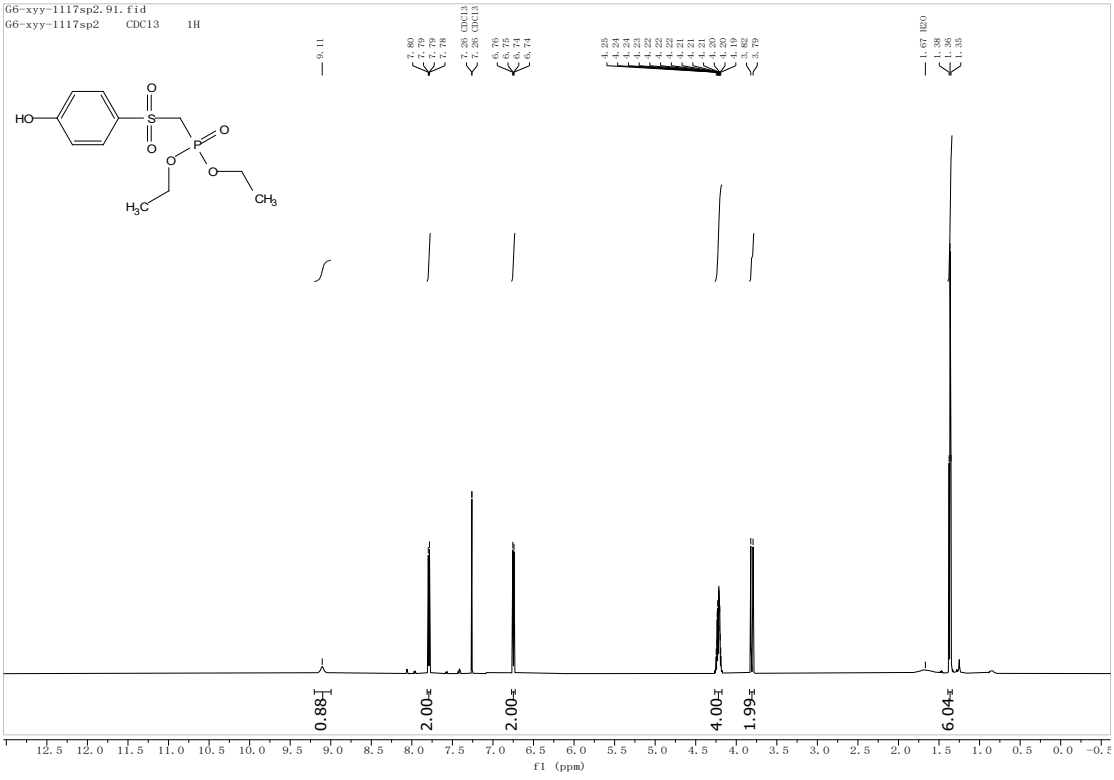
To a solution of compound 5 (300 mg, 0.73 mmol) in DCM (10 mL), 4 M hydrogen chloride in 1,4-dioxane solution (50 mL) was added, and the reaction was stirred at room temperature for 30 minutes. The reaction mixture was then concentrated in vacuo to afford the title compound (200 mg, 88%) as a yellowish solid. LCMS: m/z 311.1 $[M+H]^+$

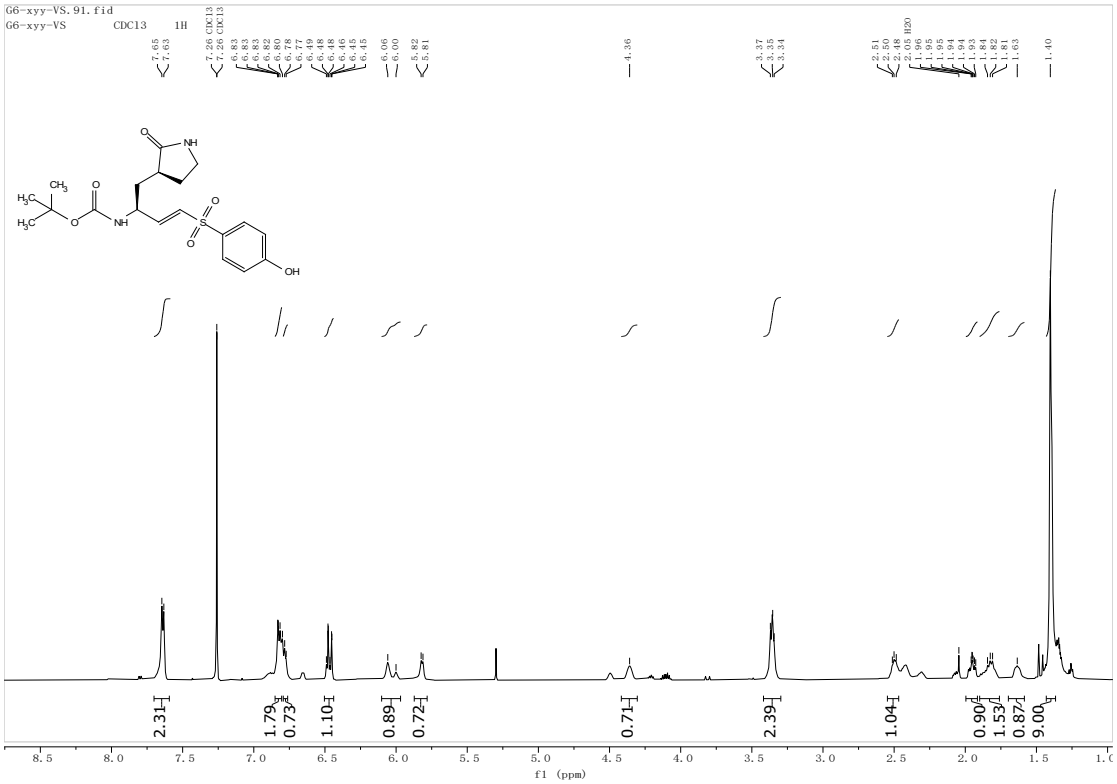
NMR spectrums of intermediate compound

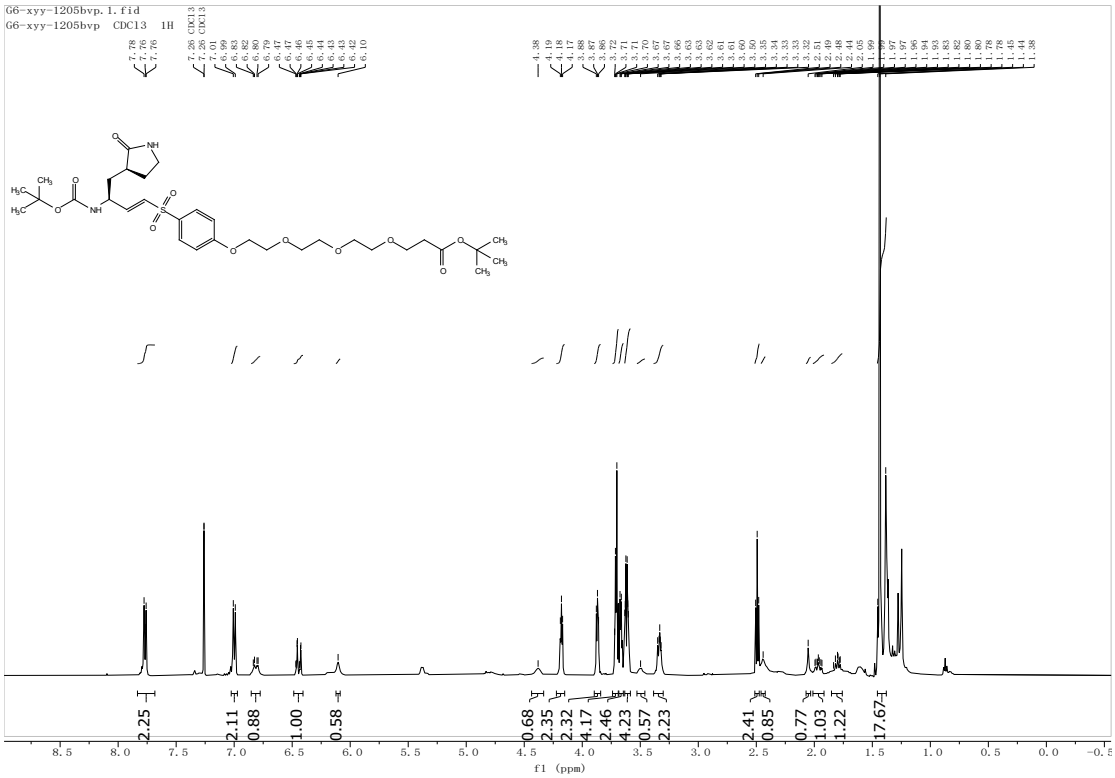


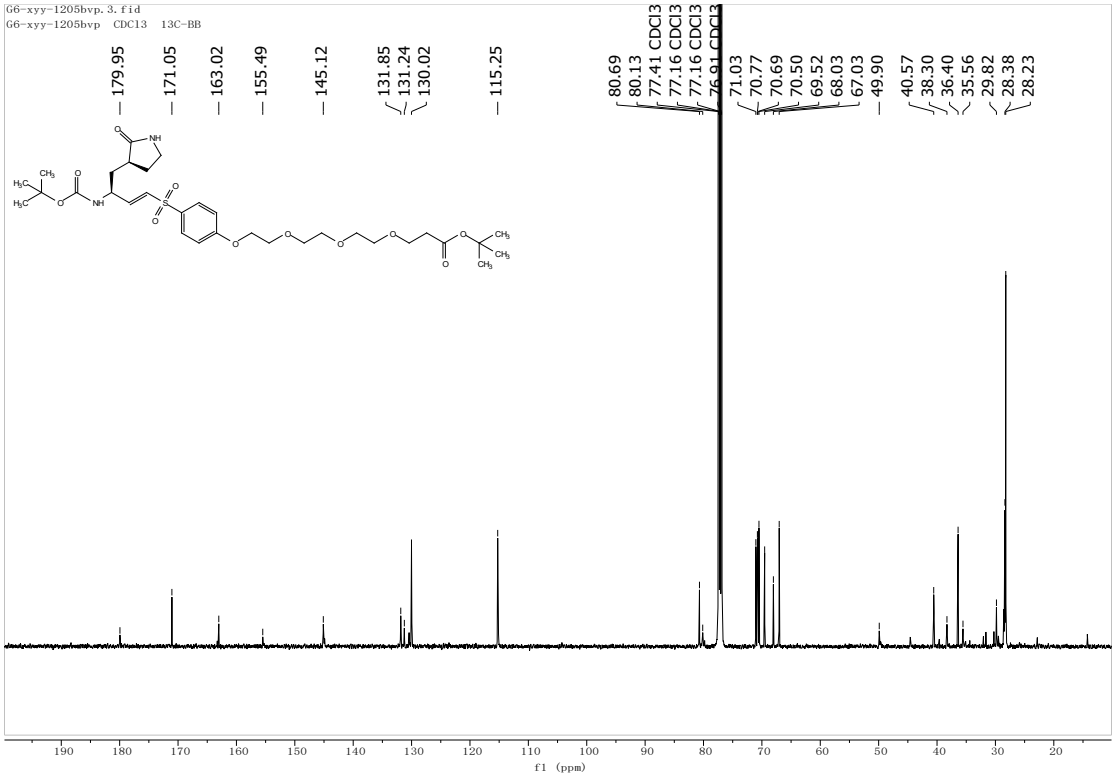


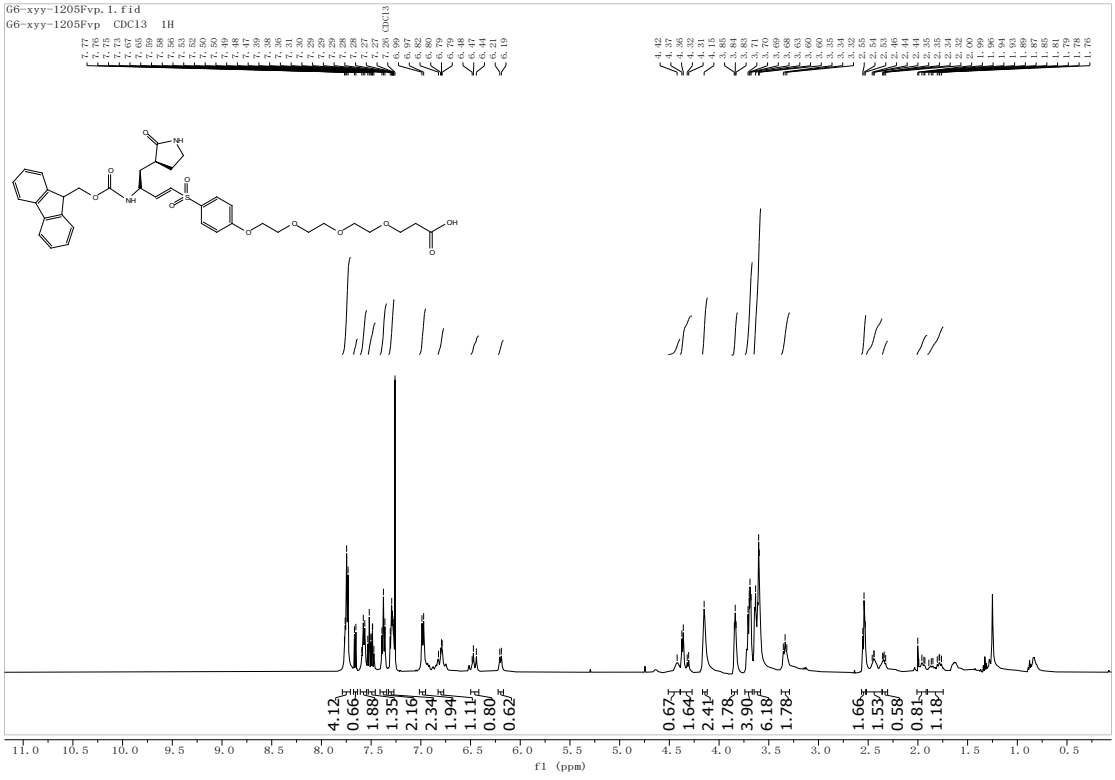

G6-xyy-1115-SP1.91.fid
G6-xyy-1115-SP1 CDCl3 1H
HO
S
O
P
O
O
CH3
CH3
f1 (ppm)

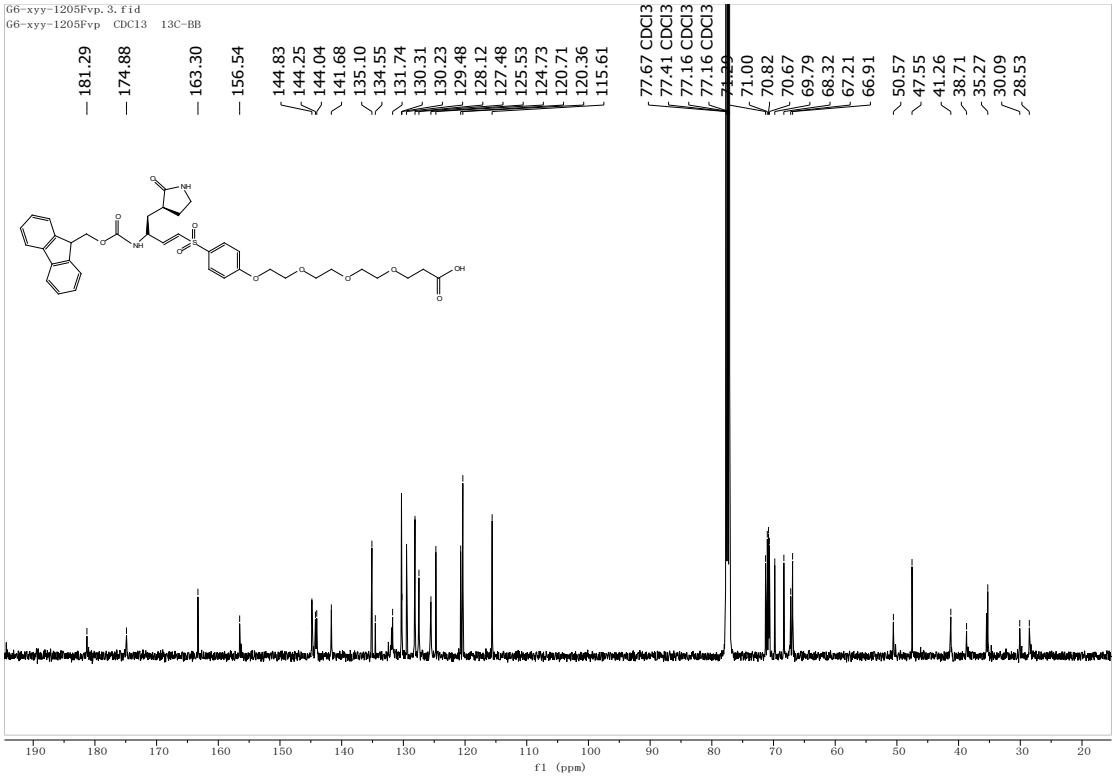

G6-xyy-1117sp2.91.fid
G6-xyy-1117sp2 CDCl3 1H
HO
S
O
O
P
O
O
H3C
CH3
f1 (ppm)


G6-xyy-VS.91.fid
G6-xyy-VS CDCl3 1H
f1 (ppm)


G6-xyy-1205bvp.1.fid
G6-xyy-1205bvp CDCl3 1H
f1 (ppm)


G6-xyy-1205bvp.3.fid
G6-xyy-1205bvp CDCl3 13C-BB
179.95
171.05
163.02
155.49
145.12
131.85
131.24
130.02
115.25
80.69
80.13
77.41 CDCl3
77.16 CDCl3
77.16 CDCl3
76.91 CDCl3
71.03
70.77
70.69
70.50
69.52
68.03
67.03
49.90
40.57
38.30
36.40
35.56
29.82
28.38
28.23
f1 (ppm)


G6-xyy-1205Fvp.1.fid
G6-xyy-1205Fvp CDCl3 1H
CDCl3
4.12
0.66
1.88
1.35
2.16
2.34
1.94
1.11
0.80
0.62
0.67
1.64
2.41
1.78
3.90
6.18
1.78
1.66
1.53
0.58
0.81
1.18
f1 (ppm)


G6-xyy-1205Fvp.3.fid
G6-xyy-1205Fvp CDCl3 13C-BB
181.29
174.88
163.30
156.54
144.83
144.25
144.04
141.68
135.10
134.55
131.74
130.31
130.23
129.48
128.12
127.48
125.53
124.73
120.71
120.36
115.61
77.67 CDCl3
77.41 CDCl3
77.16 CDCl3
77.16 CDCl3
71.00
70.82
70.67
69.79
68.32
67.21
66.91
50.57
47.55
41.26
38.71
35.27
30.09
28.53
f1 (ppm)