Supplementary Information

Introducing Aldehyde Functionality to Proteins Using Ligand-directed Affinity Labeling

Yinan Song,[#] Feng Xiong,[#] Jianzhao Peng, Yi Man Eva Fung, Yiran Huang, and Xiaoyu Li *

[†] Department of Chemistry and State Key Laboratory of Synthetic Chemistry, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China

[#] These authors contributed equally to this work.

Email: xiaoyuli@hku.hk

Content

Supplementary Information

1. Abbreviations

ACN: acetonitrile

CBS: carboxybenzene sulfonamide

DCC: *N*, *N*'-dicyclohexylcarbodiimide

DCM: dichloromethane

DIPEA: *N*, *N*'-diisopropylethylamine

DMAP: 4-dimethylaminopyridine

DMEM: Dulbecco's Modified Eagle Medium

DMF: *N*, *N*'-dimethylformamide

DMP: Dess–Martin periodinane

DMSO: dimethyl sulfoxide

DSC: *N*, *N*'- disuccinimidyl carbonate

EtOAc: ethyl acetate

EDC: 1-ethyl-3- (3-dimethylaminopropyl) carbodiimide hydrochloride

ESI-TOF-MS: electron-spray ionization time-of-flight mass spectrometry

FA: folic acid

FAM: carboxyfluorescein

FBS: fetal bovine serum

Fmoc: 9−fluorenylmethyloxycarbonyl

FOLR: folate receptor

GSH: glutathione

HOBt: *N*-hydroxylbenzotriazole

IAA: iodoacetamide

MALDI: Matrix-assisted laser desorption/ionization

mCPBA: meta-chloroperoxybenzoic acid

NHS: *N*-hydroxysuccinimide

PAGE: polyacrylamide gel electrophoresis

PBS: phosphate-buffered saline

PS: penicillin-streptomycin

SDS: sodium dodecyl sulfate

THF: tetrahydrofuran

TEA: trimethylamine

TFA: trifluoroacetic acid

UV: ultraviolet

XTT: 2,3-bis(2-methoxy-4-nitro-5-sulfophenyl)-5-[(phenylamino)carbonyl]-hydroxide

2. Materials and general methods

a) General.

Unless otherwise noted, all reagents and solvents were purchased from commercial sources and used as received. Common chemical reagents were purchased from Meyer (Shanghai) Chemical Technology Co., Ltd., J&K Scientific Ltd., Dieckmenn Co., Ltd. and Sigma-Aldrich.

Cell Culture. Cell lines were purchased from ATCC. Hypoxia cell cultivation was generated by AnaeroPack from Mitsubishi Gas Chemical Company. Cell culture reagents and cell tracking products were purchased from Thermo Fisher Scientific.

Biological experiments. Deglycosylation Kit and enzymes were purchased from New England Biolabs. Antibodies and proteins were purchased from Sino Biologicals, Santa Cruz, and Abcam. Dynabeads™ Biotin Binder was purchased from Thermo Fisher Scientific. Water was purified with a Thermo Scientific Barnstead Nanopure system. Concentrations of proteins were determined based on the absorbance at 280 nm using BCA protein assay Kit (Merck, 71285-M). Gel images were captured by a Bio-Rad Chemidoc system and processed by ImageJ software. Confocal images were acquired with a confocal laser microscope (Carl Zeiss LSM 710 NLO). Flow cytometer analysis was carried out with a Flow Cytometer and Cell Sorter Workstation (BD FACS AriaIII), and data analysis was performed using FlowJo software. Error bars in all data plots are standard deviation (SD) with the experiments individually repeated in triplicates.

Compound synthesis. Synthesized small molecules were characterized by ^{1}H and ^{13}C NMR (500 MHz or 400 MHz, Bruker AVAMCE III) and HRMS (Bruker APEX IV FTMS).

b) Information of the proteins and antibodies used in this study.

Protein/Antibody	Vendor	Catalog #
FR	Sino Biologics	11241-H08H
CA-II	Sigma	C2522
FKBP12	Abcam	Ab167985
anti-CA-12 antibody	Abcam	Ab195233
anti-FR antibody	Sino Biologics	81073-T40
streptavidin HRP	Abcam	Ab7403
anti-actin antibody	Sino Biologics	100166-MM10

3. Supplementary figures and tables referenced in the main text.

Figure S1.

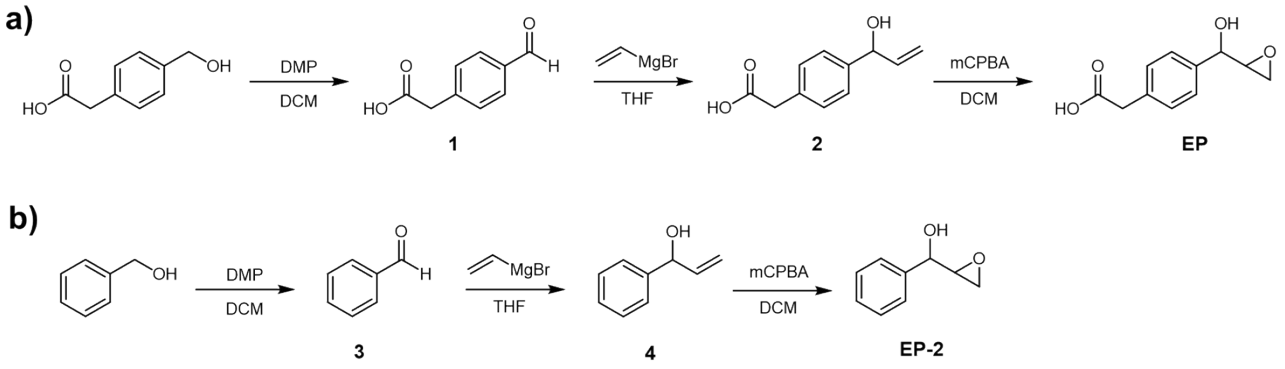


Figure S1: a) Synthesis scheme for the epoxy alcohol precursor **EP;** b) **EP-2** was synthesized similarly and used to test the epoxide opening regioselectivity (see Figure S2).

Experimental details:

Compound **1**

Dess-Martin periodinane (3.831 g, 9.03 mmol) in DCM (15 mL) was added dropwise to the solution of 4-(carboxymethyl)benzyl alcohol (1 g, 8.02 mmol) in DCM (15 mL, 100 μL H_2O) at 0 °C. The reaction mixture was then stirred at r.t. for another 4 hours. After washed with Na_2SO_3 (aq.) and extracted with DCM, the solvent was removed under vacuum. The residue was purified by flash column chromatography on silica column (hexane: EtOAc = 1 : 1) to afford compound **1** (921 mg, 92 % yield) as light yellow solid. ^{1}H NMR (500 MHz, $CDCl_3$) δ 9.99 (s, 1H), 7.84 (m, 2H), 7.45 (m, 2H), 3.74 (m, 2H).

Compound **2**

To a well-stirred, argon-protected solution of **1** (600 mg, 3.6 mmol) in dry THF (20 mL) at -20 °C, 1 M vinyl magnesium bromide THF solution (7.2 mL) was added dropwise in 30 min. Then the reaction mixture was slowly warmed up to r.t. for another 2 hours. The reaction mixture was washed with saturated NH_4Cl (aq.) and extracted with DCM. The residue was purified by silica column chromatography (DCM: MeOH= 20: 1) after the removal of organic solvents under vacuum to afford compound **2** (601 mg, 87% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.29 (m, 4H), 6.01 (m, 1H), 5.32 (m, 1H), 5.18 (m, 2H), 3.62 (m, 2H).

Compound **EP**

Compound **2** (200 mg, 1.04 mmol) in dry DCM (5 mL) was added with *m*CPBA (358 mg, 2.08 mmol) at 0 °C. The reaction mixture was then returned to r.t. for another 12 hours. The organic solvent was removed under vacuum after washed with saturated $Na_2S_2O_3$ (aq.) and extracted with DCM. The residue was purified through

column chromatography (DCM: MeOH= 20: 1) giving **EP** (142 mg, 66 % yield). Mixture of the diastereomers (dr ≈ 1 : 1): ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.25 (m, 8H), 4.74 (d, *J* = 3.3 Hz, 1H), 4.34 (d, *J* = 5.9 Hz, 1H), 3.57 (m, 4H), 3.13 (m, 2H), 2.85 (q, *J* = 4.8, 2.8 Hz, 1H), 2.75 (t, *J* = 4.4 Hz, 1H), 2.72 (m, 1H), 2.68 (t, *J* = 4.5 Hz, 1H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 176.31, 176.27, 138.85, 138.50, 133.52, 129.60, 129.54, 126.75, 126.61, 74.27, 70.91, 56.13, 55.19, 45.48, 44.08, 40.61.

Compound **3**

Dess-Martin periodinane (3.18 g, 7.5 mmol) in DCM (10 mL) was added dropwise to the solution of benzyl alcohol (540 mg, 5 mmol) in DCM (10 mL, 50 μL H_2O) at 0 °C. The reaction mixture was then stirred at r.t. for another 4 hours. After washed with saturated Na_2SO_3 (aq.) and $NaHCO_3$ (aq.) extracted with DCM, the solvent was removed under vacuum. The residue was purified by flash column chromatography on silica column (hexane: EtOAc = 15 : 1) to afford compound **3** (492 mg, 91 % yield), which was further used in the next step of synthesis without further characterization.

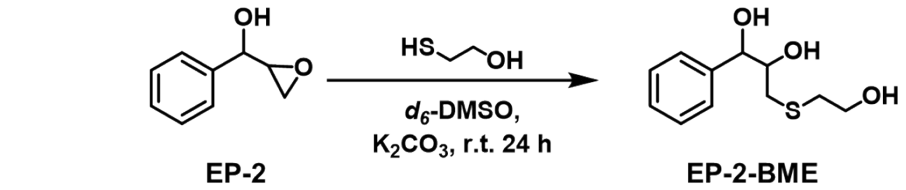
Compound **4**

To a well-stirred, argon-protected solution of **3** (306 mg, 2.88 mmol) in dry THF (20 mL) at -20 °C, 1 M vinyl magnesium bromide THF solution (3.18 mL) was added dropwise in 30 min. Then the reaction mixture was returned to r.t. for another 2 hours. The reaction mixture was washed with saturated NH_4Cl (aq.) and extracted with EA. The residue was purified by silica column chromatography (hexane: EtOAc= 15: 1) after the removal of organic solvents under vacuum to afford compound **4** (311 mg, 80 % yield), which was directly used in the next step of synthesis without further characterization.

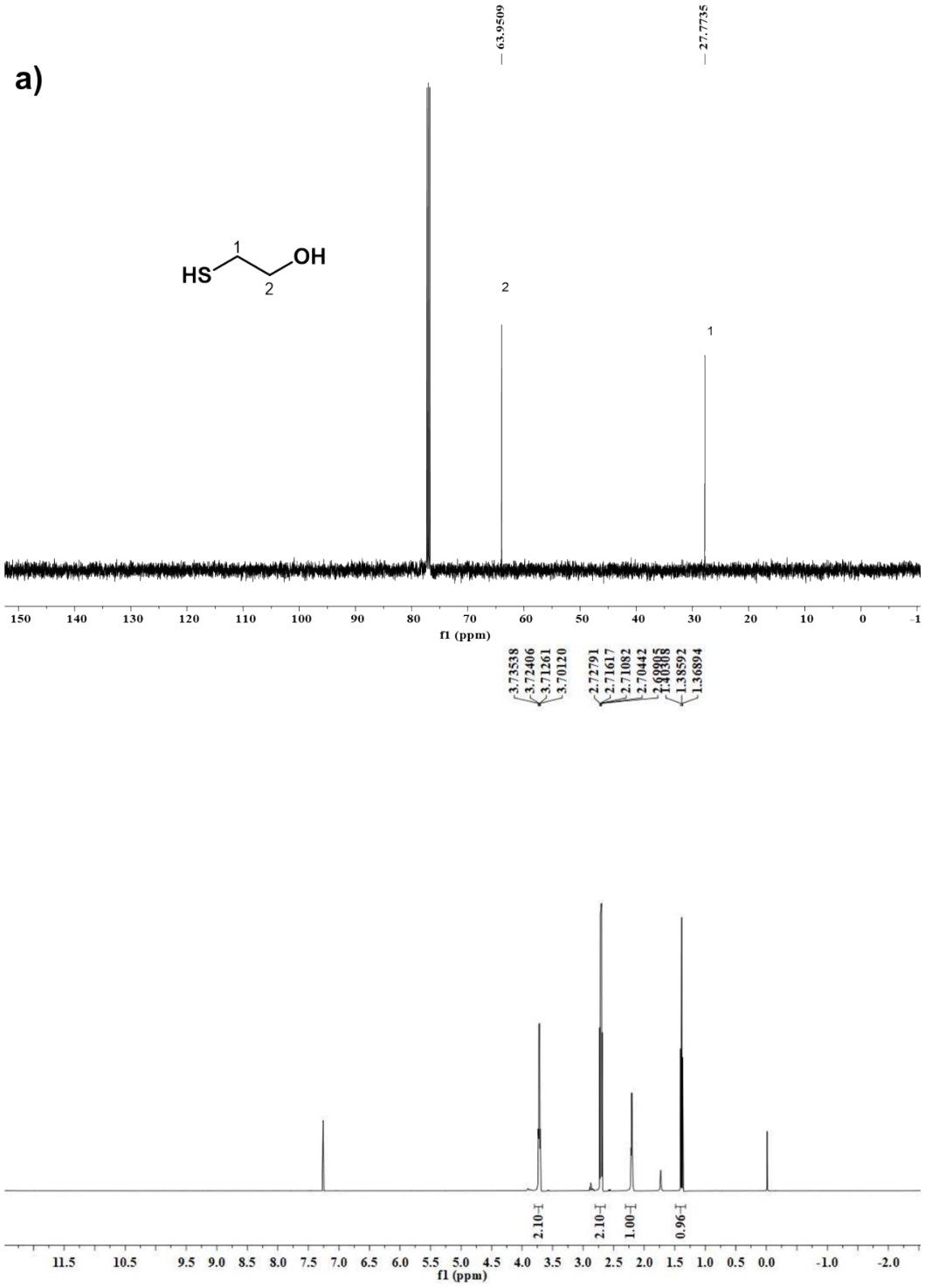
Compound **EP-2**

Compound **4** (191 mg, 1.5 mmol) in dry DCM (5 mL) was added *m*CPBA (367 mg, 2.13 mmol) at 0 °C. The reaction mixture was then return to r.t. for another 12 hours. The organic solvent was removed under vacuum after washed with saturated Na_2SO_3 (aq.) and $NaHCO_3$ (aq.) extracted with DCM. The residue was purified through column chromatography (hexane: EtOAc= 8: 1) after removing the organic solvent under vacuum to afford the compound **EP-2** (169 mg, 75 % yield). Mixture of the diastereomers (dr ≈ 1 : 1): ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.41 (m, 10H), 4.85 (m, 1H), 4.44 (m, 1H), 3.23 (m, 2H), 2.95 (m, 1H), 2.83 (m, 2H), 2.76 (m, 1H). ^{13}C NMR (100 MHz, $CDCl_3$) δ 140.23, 139.84, 128.59, 128.54, 128.14, 128.12, 126.50, 126.36, 74.64, 71.13, 56.22, 55.25, 45.40, 43.90.

Figure S2.

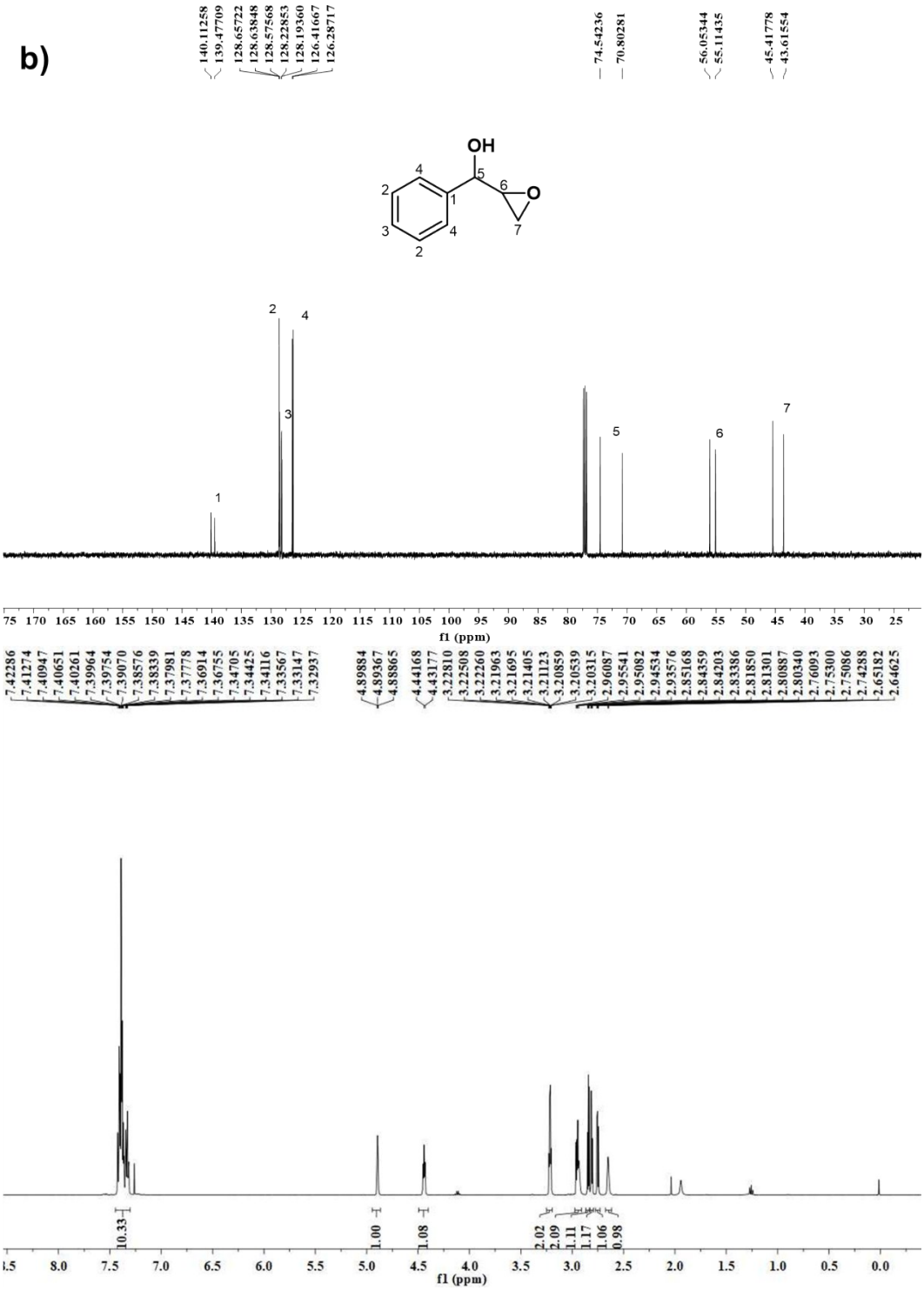


EP-2 EP-2-BME



continue on next page

b)



continued on next page

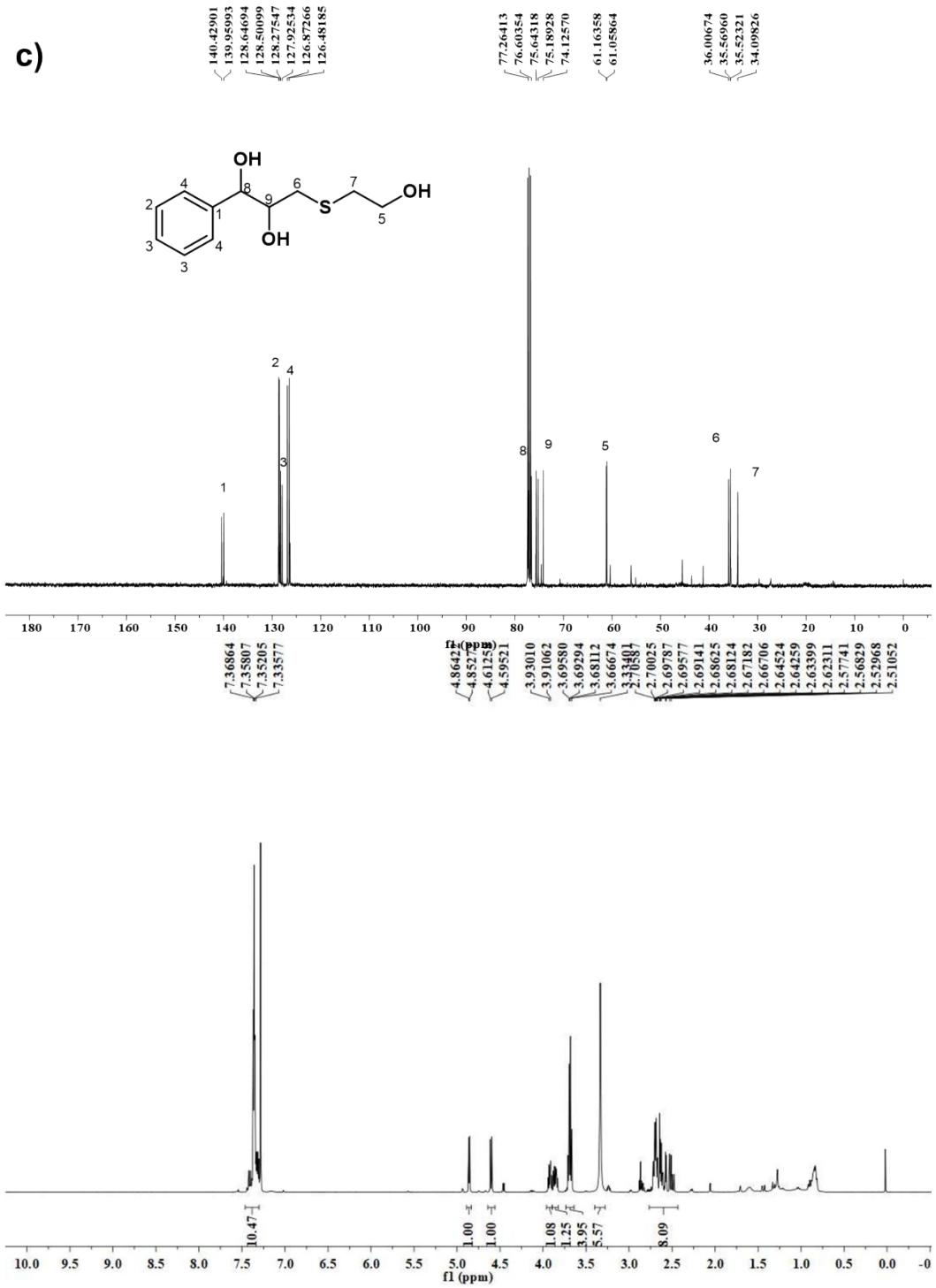


Figure S2. EP-2 reacted with 2-mercaptoethanol to examine the regioselectivity of epoxide opening. a)-b) ^{1}H and ^{13}C NMR spectra of the two reactants. c) ^{1}H and ^{13}C NMR spectra of the reaction mixture (react for 24 hours at room temperature in DMSO-d_6), which showed the formation of the desired diol product.

Figure S3.

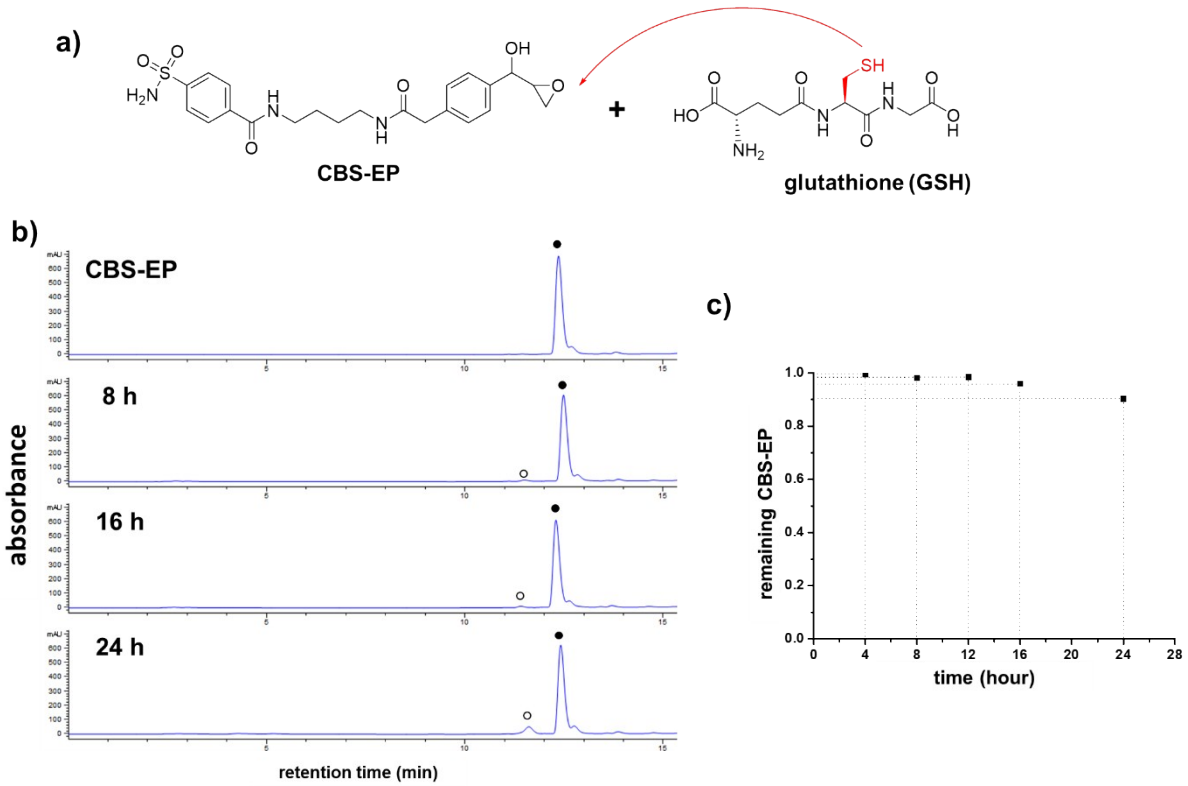


Figure S3. a) **CBS-EP** (100 μM) in 1x PBS buffer (pH 7.4) reacted with 5 mM GSH at 37 °C. b) The reaction was monitored with HPLC (absorbance: 254 nm) at different time points. ●: **CBS-EP**; ○: the epoxide opening product. c) Plot of the remaining **CBS-EP** at different time points.

Figure S4.

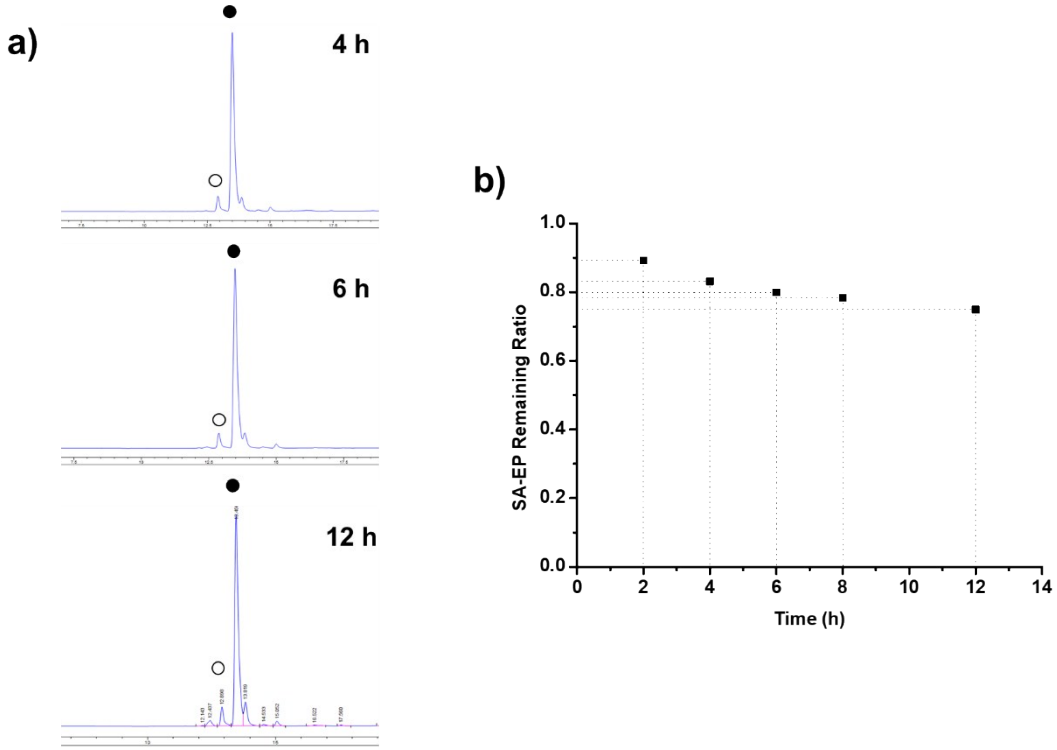


Figure S4. Stability test of **CBS-EP** in a mildly acidic condition. a) **CBS-EP** (100 μM) in 1x PBS buffer (pH 5.8) reacted with 5 mM GSH at 37 °C. b) The reaction was monitored with HPLC (absorbance: 254 nm) at different time points. ●: **CBS-EP**; ○: the epoxide opening product. c) Plot of the remaining **CBS-EP** at different time points.

Figure S5.

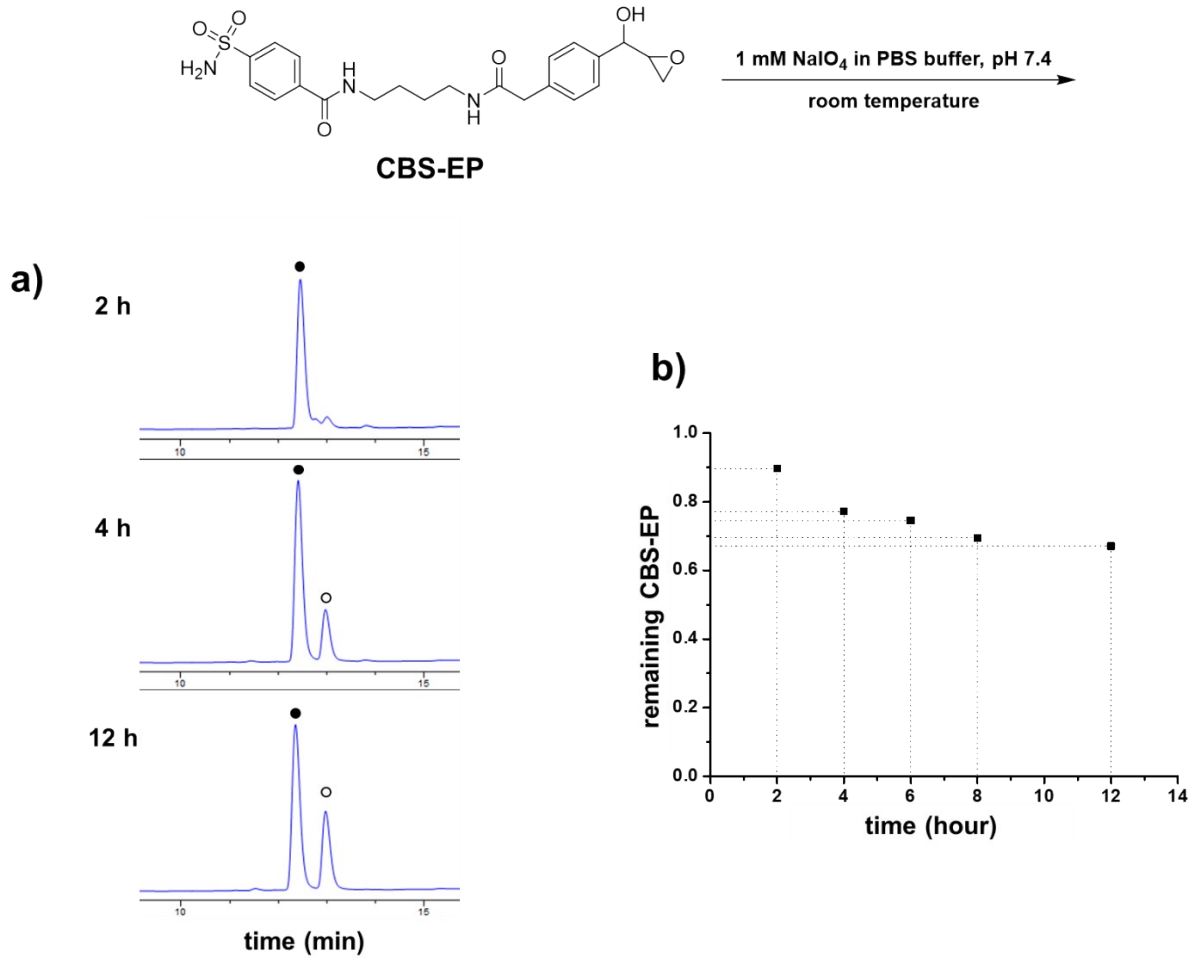


Figure S5. CBS-EP stability test under oxidative diol cleavage condition. **CBS-EP** (100 μM) and $NaIO_4$ (1 mM) in 1x PBS buffer (pH 7.4) was maintained at room temperature. a) The reaction was analyzed with HPLC (λ= 254 nm) at different time intervals. ●: **CBS-EP**; ○: cleavage product. b) Plot of the remaining **CBS-EP** at different time points.

Figure S6.

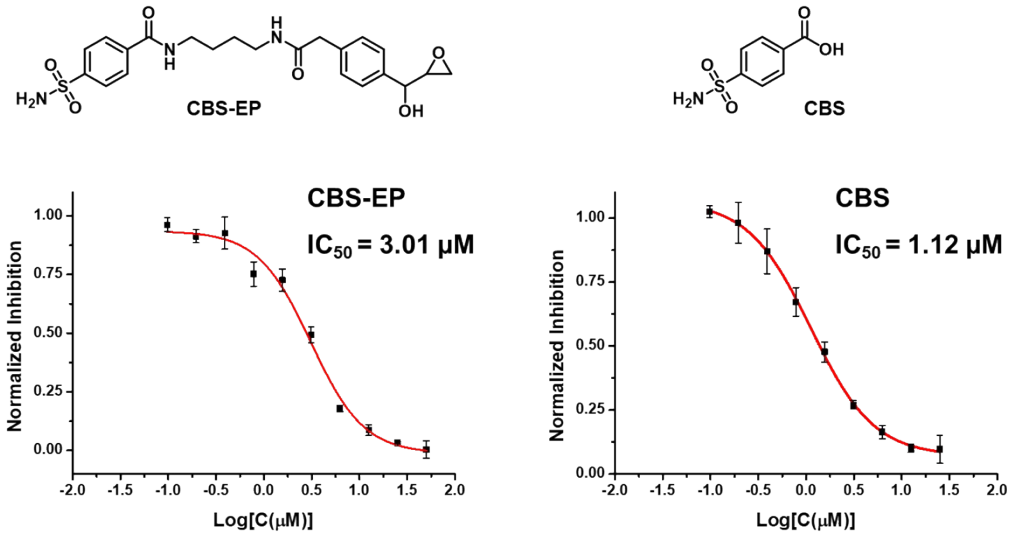


Figure S6. CA-II inhibition assay for **CBS-EP** and **CBS**. The assays were conducted following a reported procedure.[1] In brief, the esterase activity of CA-II was measured by spectrometric monitoring the increase of the absorbance at 348 nm, resulting from the hydrolysis of CA-II substrate 4-nitrophenyl acetate. In a 96-well plate, a series of inhibitor solutions with gradient concentrations were added to 1 μM CA-II solutions in HEPES buffer (50 mM, pH 7.4). The increase of absorbance at 348 nm indicating the generation of 4-nitrophenolate was detected after the addition of 4-nitrophenyl acetate (0.45 mM) at room temperature. After 3 repeats, the IC_{50} values were obtained by non-linear curve fitting analysis of the plots of the initial rate against inhibitor concentration using the OriginPro software.

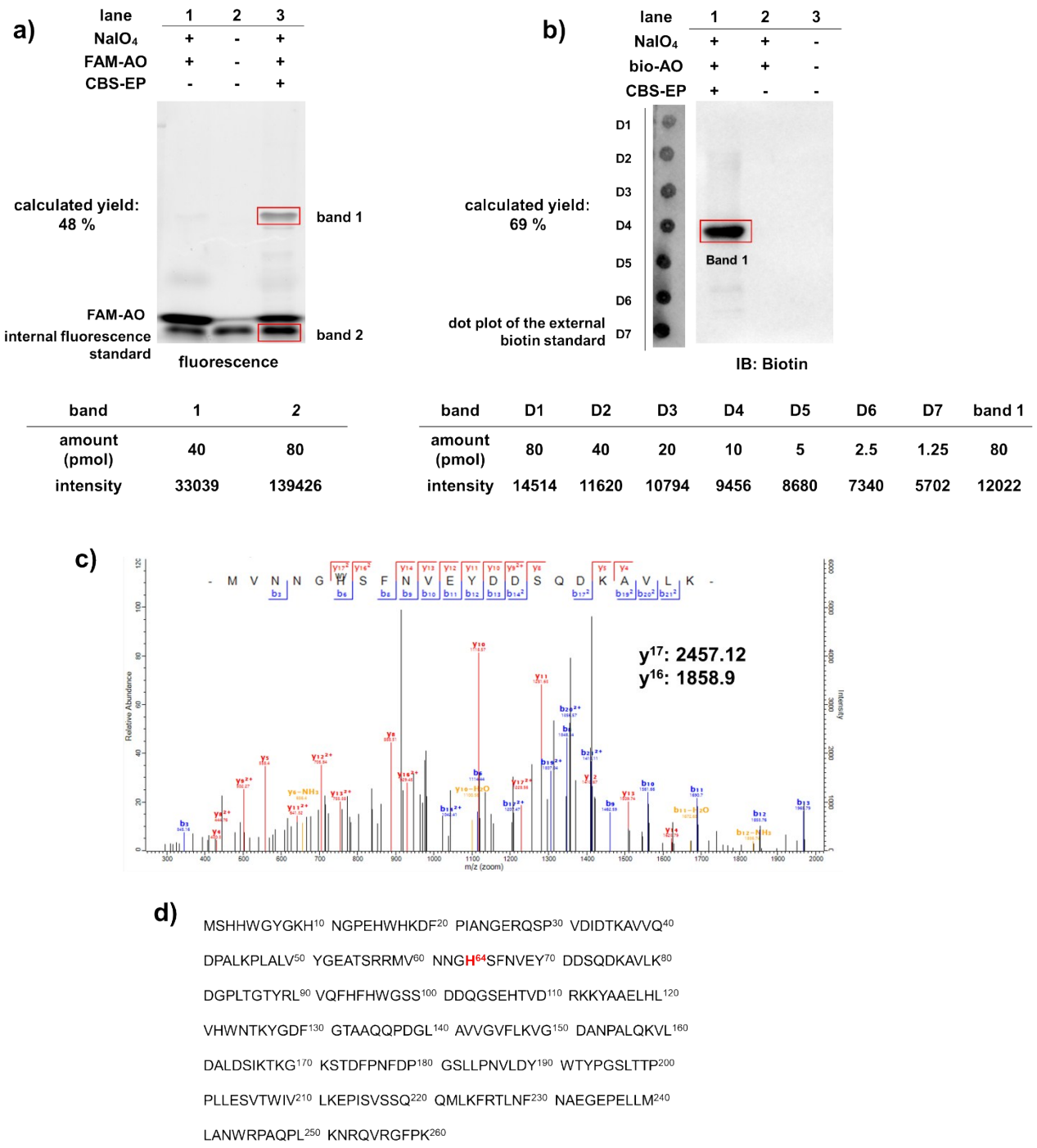


Figure S7.

Figure S7. a)-b) The labeling experiments in Figure 2b were repeated to quantify the labeling yield. a) **FAM-AO** labeling. An internal fluorescence standard was used in the electrophoresis as the reference for quantitation. Red boxes indicated the areas that were quantified for yield calculation. b) **Bio-AO** labeling. The dot plot of a biotinylated standard sample was used as the reference for quantitation. The other experimental conditions are the same as in Figure 2. The Image Lab software was used for calculation. c) Tandem mass spectral analysis of the labeling site on CA-II using **CBS-EP**. d) The primary amino acids sequence of CA-II. The labeled histidine-64 is highlighted in red.

Figure S8.

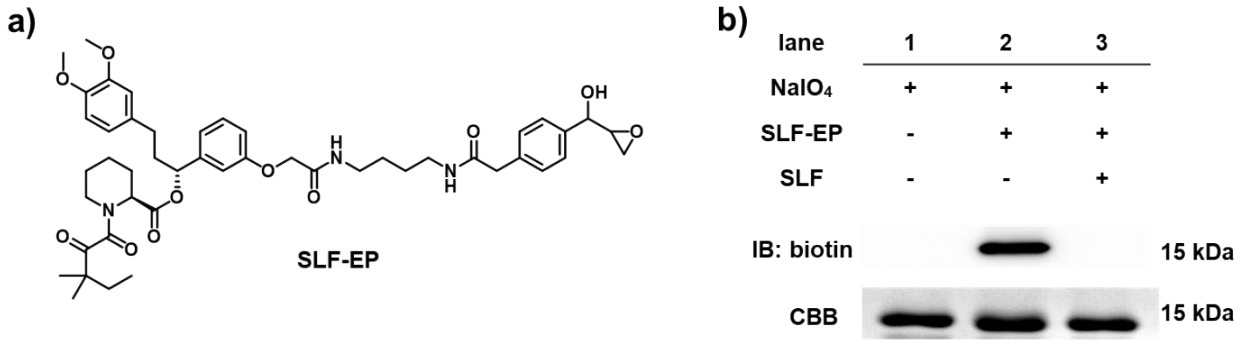


Figure S8. a) Structure of **SLF-EP** for FKBP12 labeling. b) FKBP12 (10 μM) was labeled with **SLF-EP** (20 μM; 16 h, 37 °C), cleaved with $NaIO_4$ (1 mM, 4 °C, 30 min), and tagged with **bio-AO** (40 μM; r.t. 90 min). Two control experiments were performed: without **SLF-EP** or with free **SLF** competition. Samples were analyzed with SDS-PAGE followed by Western blot analysis. In the competition experiments, 250 μM free **SLF** was added. CBB: Coomassie brilliant blue staining.

Figure S9.

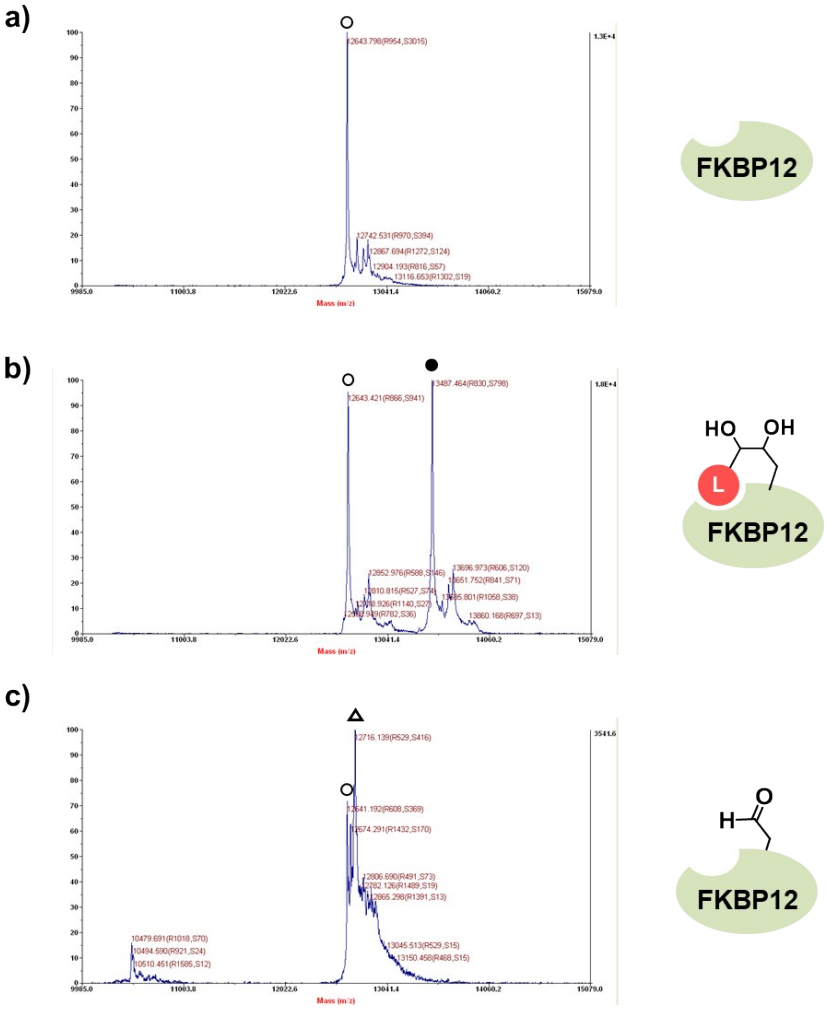


Figure S9. The FKBP12 labeling with **SLF-EP** in Figure S7 was characterized with MALDI-TOF MS. a) FKBP protein; b) after **SLF-EP** labelling; c) after periodate cleavage. ○: FKBP12 m/z = 12644; ●: FKBP12 with the **SLF-EP** probe, m/z = 13487; Δ: after periodate cleavage (the aldehyde-modified FKBP12), m/z = 12716. L: the **SLF** ligand.

Figure S10.

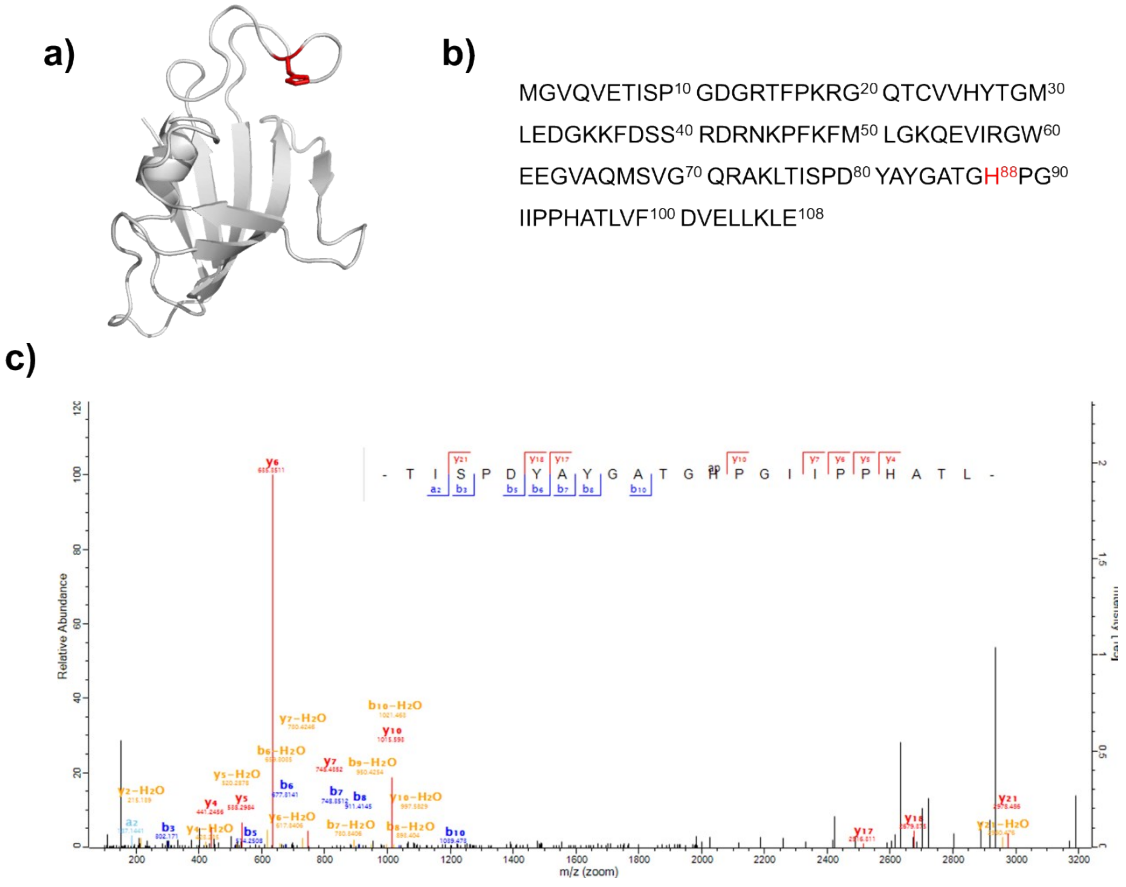


Figure S10. a) Crystal structure of FKBP12 (PDB: 2PPN) with the modification site histidine-88 highlighted. b) The primary amino acids sequence of FKBP12. The labeled His^{88} is highlighted in red. c) Tandem mass spectral analysis of the labeling site of FKBP12 using **SLF-EP**.

Figure S11.

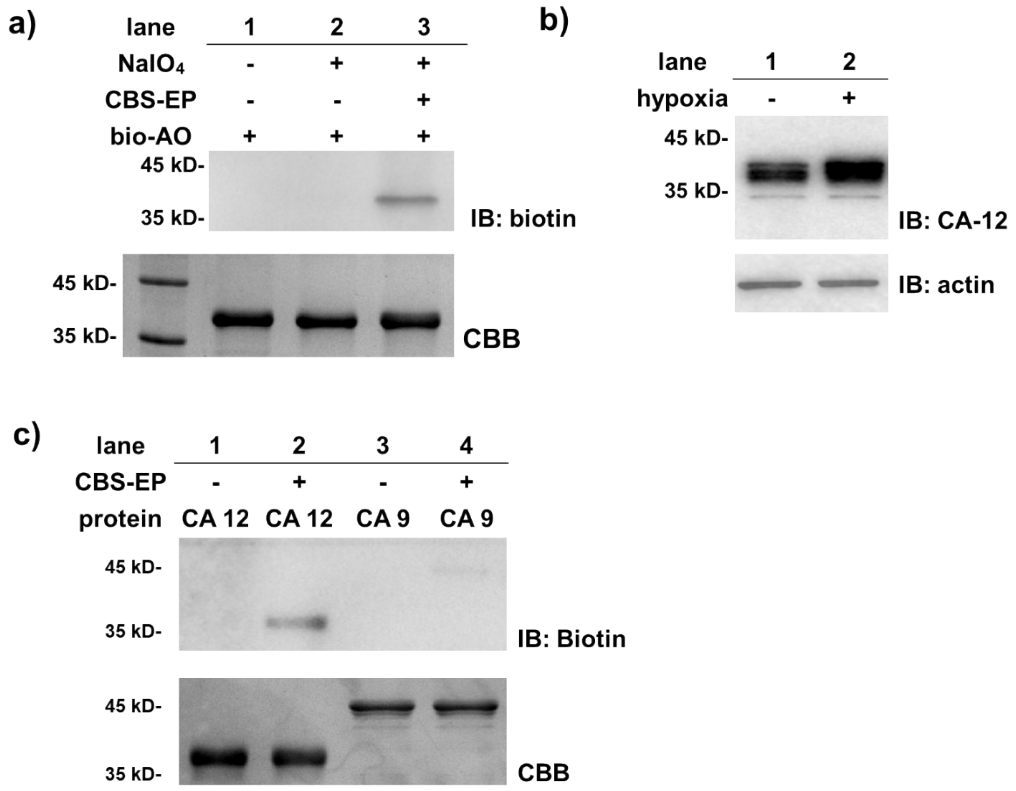


Figure S11. a) The purified CA-12 (10 μM) was labeled with **CBS-EP** (20 μM; 16 h, 37 °C), cleaved with $NaIO_4$ (1 mM, 4 °C, 30 min), and tagged with **bio-AO** (40 μM; r.t. 90 min). Samples were analyzed with SDS-PAGE followed by Western blot analysis. CBB: Coomassie brilliant blue staining. b) A549 cells were cultivated in DMEM (10% FBS, 100 unit/mL penicillin and 100 μg/mL streptomycin) (+) with or (-) without hypoxia condition for 36 hours. After cell lysis, the cell lysates were analyzed with Western-blot. c). The purified CA-12 and CA-9 were labeled in parallel and the labeling efficiencies were compared. The labeling procedure and conditions are the same as in a). The results showed that, albeit with the similar amount of protein input, very little amount of CA-9 was labeled. IB: immunoblotting; CBB: Coomassie Brilliant Blue.

Figure S12.

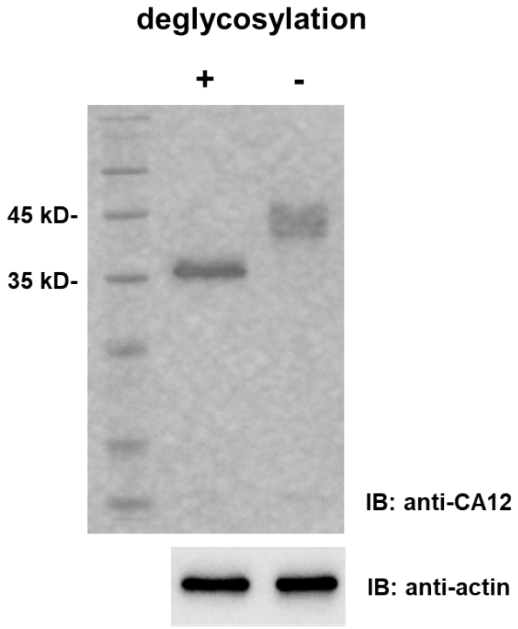


Figure S12. A549 cells were incubated with 5 μL Mix (II) mix enzymes (Deglyco-Mix; New England Biolabs. Inc, P6044) per 100 μL PBS buffer (pH 7.4) at 37 °C. The cells were then lysed with RIPA buffer (Beyotime, P0013B) and analyzed with Western-blot using an anti-CA-12 antibody to examine the deglycosylation efficiency.

Figure S13.

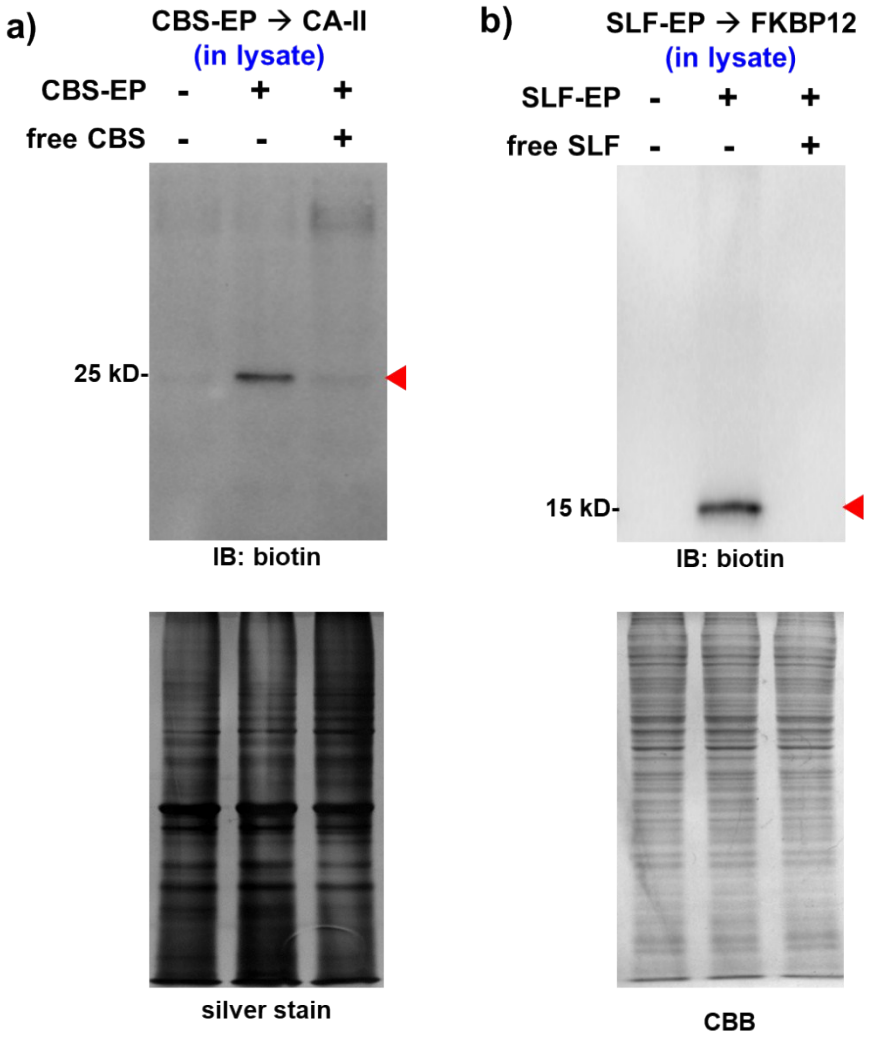


Figure S13. Loading control of Figure 3a. Labeling of CA-II (2% w/w) and FKBP 12 (2% w/w) in deglycosylated Hela cell lysates. Condition: CA-II or FKBP 12 (10 µM) was labeled with **CBS-EP** or **SLF-EP** (20 µM; 16 h, 37 °C), cleaved with $NaIO_4$ (1 mM, 4 °C, 30 min), and tagged with **FAM-AO** or **bio-AO** (40 µM; r.t. 90 min). Samples were analyzed with SDS-PAGE followed by Western blot (above) and silver staining (bottom).

Figure S14.

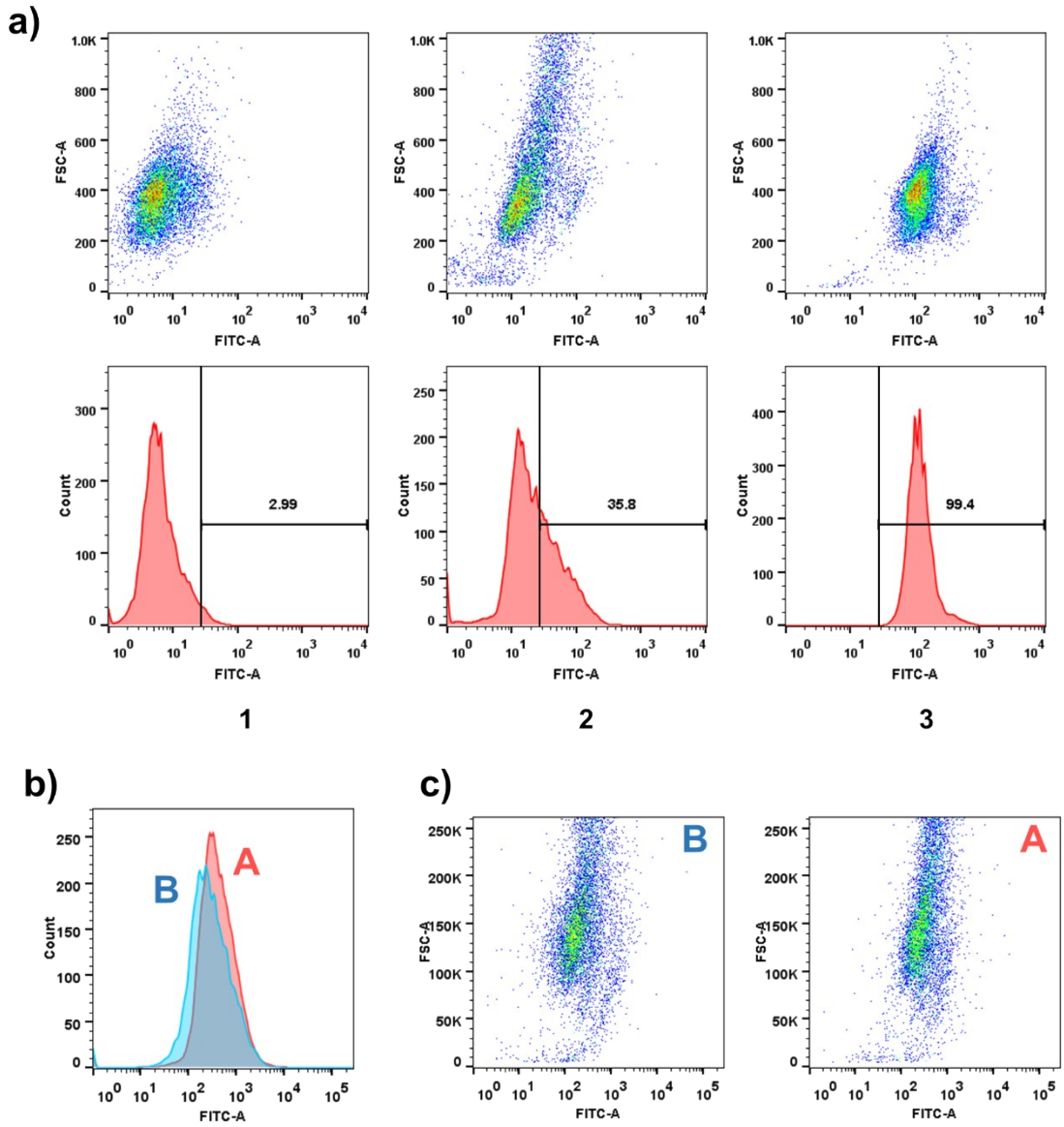


Figure S14. a) Flow cytometry histograms of the cell labeling experiments in Figure 3d. CA-12 on A549 cells were labeled with **CBS-EP**, tagged with **FAM-AO**. Condition: **CBS-EP** (4 μM), 37 °C, 4 h; oxidation: $NaIO_4$ (1 mM), 4 °C, 5 min; tagging: **FAM-AO** (40 μM), aniline (5 mM), 4 °C, 90 min. b) Similar to Figure 3d, CA-12 on A549 cells was labeled with **CBS-EP** and tagged with **FAM-AO** but using the completely "no-wash" procedure as described in details in **Section 4d**; the other experimental conditions are the same as in Figure 3d.The labeled cells were analyzed with flow cytometry. The histograms are shown in c). **A**: the labeled A549 cells; **B**: the labeled A549 cells but with 100x free CBS (200 μM) completion after the labeling and before the analysis. The results showed a slight decrease of the cell fluorescence signal.

Figure S15.

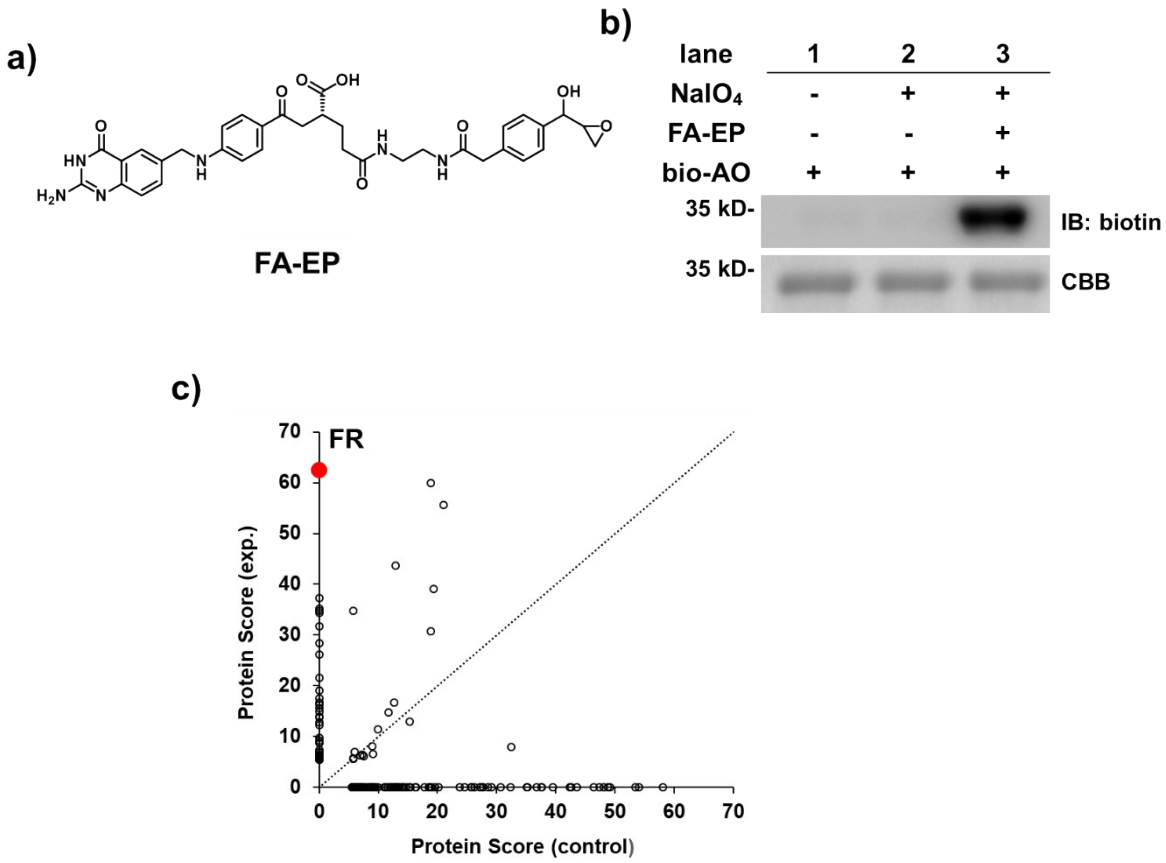


Figure S15. a) Structure of **FA-EP**. b) Purified FR (10 μM) was labeled with **FA-EP** (20 μM; 16 h, 37 °C), cleaved with $NaIO_4$ (1 mM, 4 °C, 30 min), and tagged with **bio-AO** (40 μM; r.t. 90 min). Samples were analyzed with SDS-PAGE followed by Western blot analysis. CBB: Coomassie brilliant blue staining. c) FR on Hela cells were labeled with **FA-EP**, tagged with **bio-AO**, affinity-purified, and analyzed with tandem MS. *y*-axis: with **FA-EP**; *x*-axis: without **FA-EP** (negative control).

Figure S16.

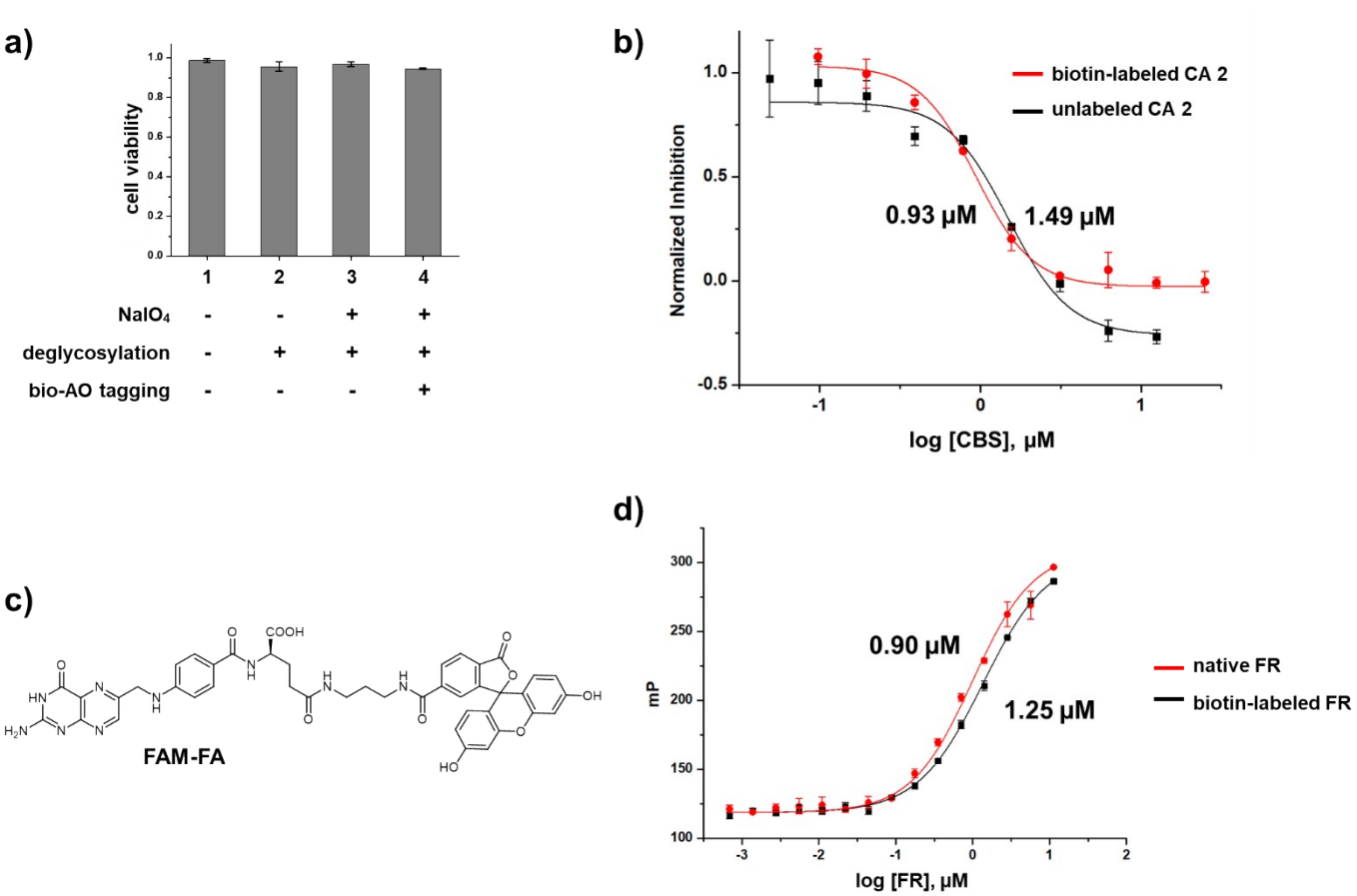


Figure S16. a) Cell viability test of periodate oxidation, deglycosylation and oxime ligation. The viability of A549 cells after deglycosylation with or without periodate oxidation (pH 7.4 PBS buffer and 1 mM $NaIO_4$ at 4 ℃ for 5 min), and oxime ligation (**bio-AO** 40 μM and 5 mM aniline at 4 ℃ for 90 min in 1x PBS buffer (pH 6.7)). Analyzed by hemocytometer with trypan blue. b) **CBS** inhibition assay of the unlabeled and biotin-labeled CA-2. The assays were conducted following a reported procedure and the details are described in the caption of Figure S6.[1] CA-2 was labeled with the same procedure as in Figure 2. c) The structure of FAM-FA. d) Fluorescence polarization (FP) assay results using FAM-FA against the unlabeled and biotin-labeled folate receptor (FR). The assay was performed based on previous reports.[2,5] The experimental details of the FP assay are provided in Section 4j.

Figure S17.

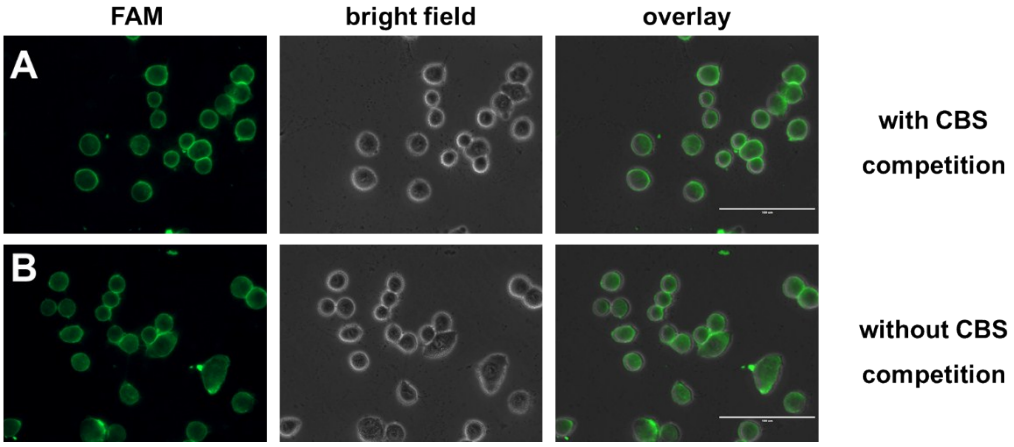
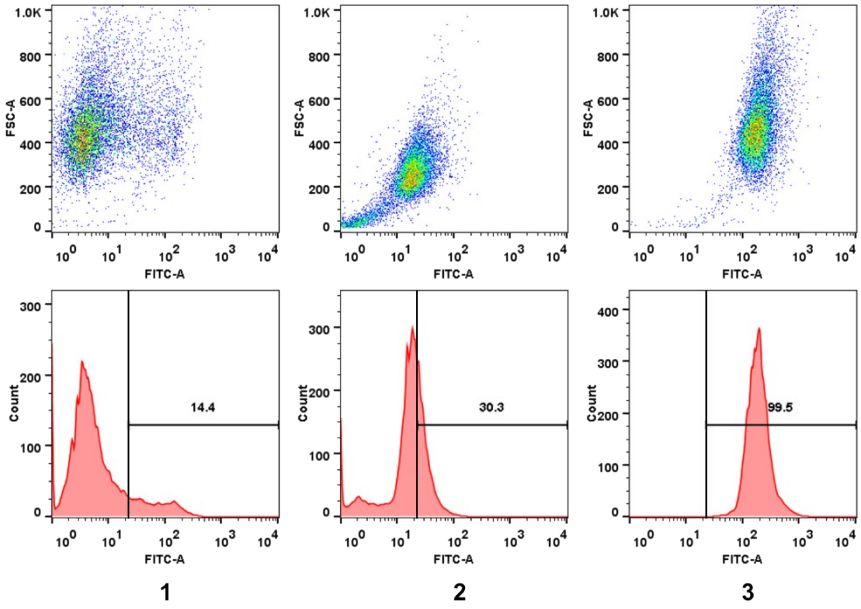


Figure S17. The experiments in Figure 4c, the labeling of CA-12 with the fluorogenic TPNF probe on A549 cells, was conducted with 100x free **CBS** (200 μM) added after the labeling and before TPNF addition. The other experimental conditions are the same as in Figure 4c but with only one time point (20 min.) was captured as the example. **A**: with **CBS** competition; **B**: without competition.

Figure S18.



	1	2	3
CBS-EP	-	-	+
deglycosylation	+	-	+
TPNF	+	+	+

Figure S18. CA-12 on A549 cells was labelled with an aldehyde group and then tagged with **TPNF** following the procedure as shown in Figure 4. Conditions: **TPNF** (10 μM), aniline (5 mM), 1x PBS buffer (pH 6.7) at 4 ℃, 90 min. The labelled cells were analyzed with flow cytometry and the histograms are shown above.

Figure S19.

Figure 2b, top panel

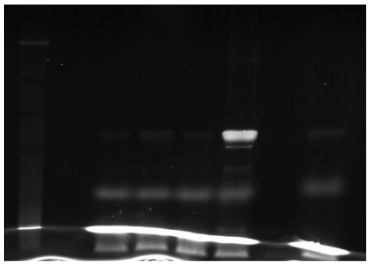


Figure 2b, middle panel

Figure 2b, bottom panel

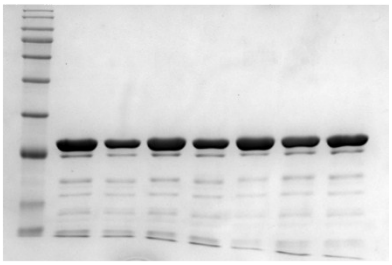


Figure 3a, left panel

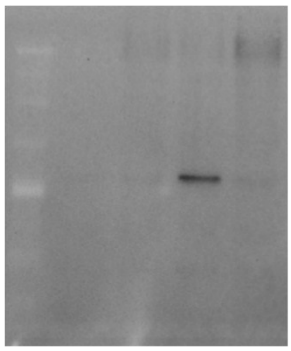


Figure 3a, right panel

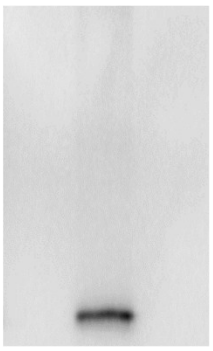


Figure 3b, top panel

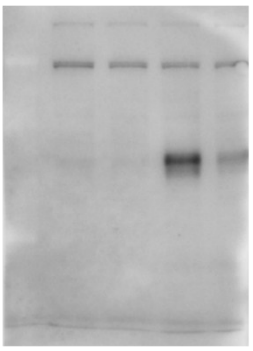


Figure 3b, middle panel

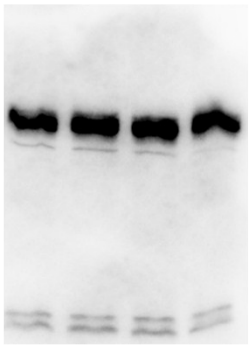
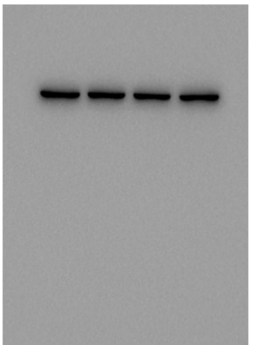


Figure 3b, bottom panel



continued on next page

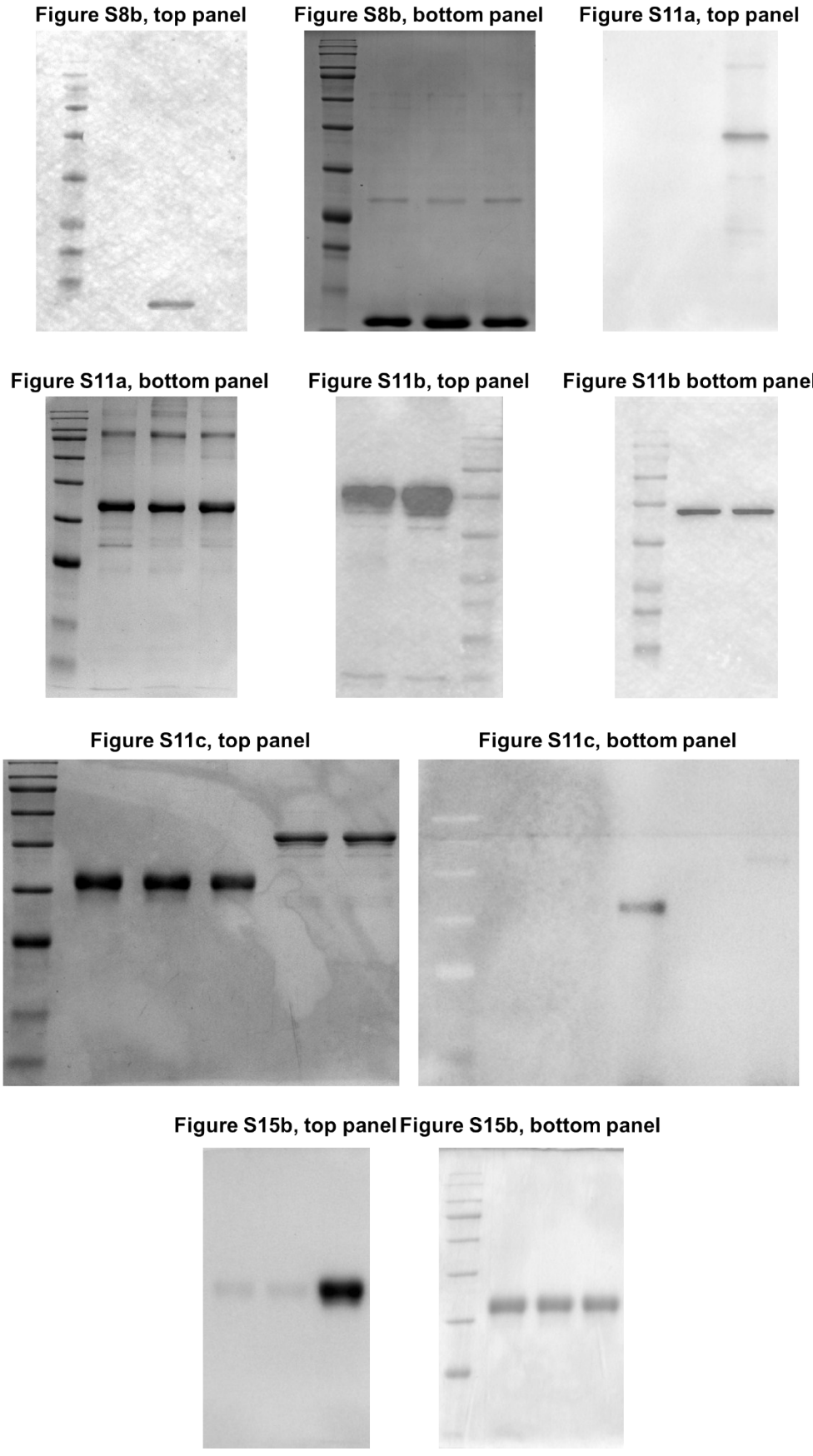


Figure S19. Uncropped, full sized gel images for all the electrophoresis experiments in this study.

4. General Methods.

a) Labeling of the purified proteins (in buffer or in cell lysates).

The purified protein (10-20 μM) was incubated with the corresponding **EP** probe (20 μM) for 16 h in PBS buffer (pH 7.4) or in the corresponding cell lysates at 37 ℃. After incubation, periodate oxidation was carried out by the addition of $NaIO_4$ (final concentration: 1 mM) at 4 ℃ for 30 min. Without quenching, 1 μL 10 % formic acid per 20 μL reaction mixture was added to adjust the pH of the solution to be slightly acidic, which has been reported to be optimal for the subsequent aminooxy tagging reaction.[3] Samples were then incubated with either **FAM-AO** (40 μM) or **bio-AO** (40 μM) at r.t. for 90 min.

b) Cell culture.

A549 cells with elevated CA-12 expression were obtained with hypoxia cultivation (AnaeroPack; Mitsubishi Gas Chemical) at 37 °C for 36 h in DMEM (10 % FBS, 100 unit/mL penicillin, and 100 μg/mL streptomycin). For FR overexpressed cells, $HeLa^{FR}$ cells were cultured in folic-acid-depleted RPMI 1640 medium (10 % FBS, 100 unit/mL penicillin, and 100 μg/mL streptomycin) to induce the upregulation of folate receptor expression on cell membrane. After at least 6 passages, the cells were ready for experiments. All other cells were cultured in DMEM medium supplemented with 10% FBS, 100 unit/mL penicillin, and 100 μg/mL streptomycin in a 5% CO_2 humidified incubator at 37 °C.

c) Labeling of endogenous proteins.

Cells were treated with the corresponding **EP** probe (4 μM) for 4 h in DMEM (10 % FBS, 100 unit/mL penicillin, and 100 μg/mL streptomycin) at 37 ℃. The DMEM was removed by gentle aspiration and replaced with the solution containing the deglycosylase enzyme mix (in the deglycosylation mix buffer provided by the manufacturer), and the deglycosylation was conducted following the manufacturer's instructions. After deglycosylation, a buffer exchange was conducted to 1x PBS (pH 7.4) containing $NaIO_4$ (1 mM). The cleavage reaction was maintained at 4 ℃ for 5 min. Next, another buffer exchange was conducted and the cells were incubated with **FAM-AO** or **bio-AO** (40 μM) and aniline (5 mM) in PBS buffer (pH 6.7) at room temperature for 90 min.

d) Labeling of endogenous proteins (the "no-wash" labeling procedure; Figure S14b and S17).

The DMEM cell medium was replaced with the deglycosylation mix buffer (1x PBS at pH ~7.0, provided by the manufacturer of the Deglyco-Mix kit) without the enzymes before the labeling probe was added. The subsequent target binding, crosslinking, deglycosylation, oxidation, and tagging steps were conducted the same as described above in **c)** but without any buffer exchange. In the experiments with free CBS competition, 100x **CBS** (200 μM) was added to the cell after the labeling.

e) Live cell imaging.

Cells were incubated with corresponding **EP** probe (4 μM) for 4 h at 37 °C in DMEM (10 % FBS, 100 unit/mL penicillin, and 100 μg/mL streptomycin). After deglycosylation, periodate oxidation was carried out with $NaIO_4$ (1 mM) at 4 °C for 5 min in PBS buffer (pH 7.4). Next, **FAM-AO** (40 μM) or **TPNF** (10 μM) and aniline (5 mM) was added at 4 °C in DMEM (10 % FBS, 100 unit/mL penicillin, and 100 μg/mL streptomycin) or PBS buffer (5 % FBS, pH 6.7) for cell imaging.

f) Deglycosylation.

Deglycosylation was conducted by using the Deglyco-Mix kit (New England Biolabs. Inc, P6044) following the manufacturer's instruction.

g) Flow cytometry analysis.

Cells were washed with 1x PBS for three times and fixed in 4 % PFA at room temperature for 15 min, followed by washing two more times with 1x PBS buffer by gentle resuspension and centrifugation at 500× g for 5 min. Flow cytometry analysis was carried out with a BD FACS Aria III system following the standard protocol.

h) Western blot.

The gel was electro-transferred onto immune-blot PVDF membranes. The membrane was blocked with 5 % non-fat milk in TBST buffer, incubated with the corresponding antibodies, followed by incubation with HRP-conjugated goat anti-rabbit IgG antibody or goat anti-mouse IgG antibody. The membranes were developed with Clarity™ Western ECL Substrate.

i) Tandem MS spectrometry.

The samples were first resolved with SDS-PAGE electrophoresis. After stained with GelCode™ Blue (Thermo Fisher, 24590), the gel bands were excised and then subjected to in-gel digestion following a previous report.[4] In brief, after washing and destaining of the gel bands, the samples were reduced with 25 mM DTT in 50 mM NH_4HCO_3 buffer at 55 ℃ for 45 min. Then alkylation was carried out by incubating the gel pieces with 55 mM IAA in 50 mM NH_4HCO_3 buffer at r.t. for 30 min in the dark. The gel pieces were saturated with 0.5 μg mass spectrometry grade trypsin in 25 mM NH_4HCO_3 at 37 °C overnight to digest the proteins. The resulting peptides were enriched with StageTips. The eluted peptides were dried by SpeedVac and submitted for LC-MS/MS analysis.

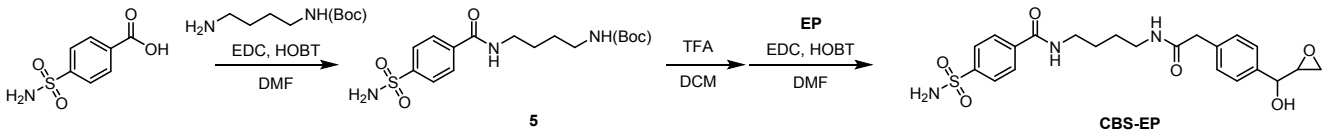
LC-MS/MS analysis was performed with an LCQ Orbitrap Elite System by the Analytical Instrumentation Center at Peking University, Beijing, China or with an Orbitrap Fusion mass spectrometer at the Core Research Facility Center of SUSTC (Southern University of Science and Technology China) in Shenzhen, China.

j) Fluorescence polarization (FP) assay to measure the binding affinity of folic acid (FA) with the unlabeled and labeled folate receptor (FR).

The preparation of the FAM-conjugated folic acid probe (FAM-FA) followed a previously reported synthesis procedure, except fluorescein was used as the fluorophore.[2] The FP assay was performed on a PerkinElmer VICTOR X5 multi-label plate reader following previous reports.[5] In brief, FAM-FA was dissolved in XXX. The solution was then combined with the protein at various concentrations, incubated for 90 min at 23 ºC in a 384-well microplate, and then subjected to FP measurement immediately. Conditions: ligand, 5 nM; protein, varied; solution volume, 70 μL. All measurements were conducted for three times. The obtained anisotropy data were fitted and plotted following a previously reported procedure.[5-7]

5. Synthetic schemes.

a) **CBS-EP.**



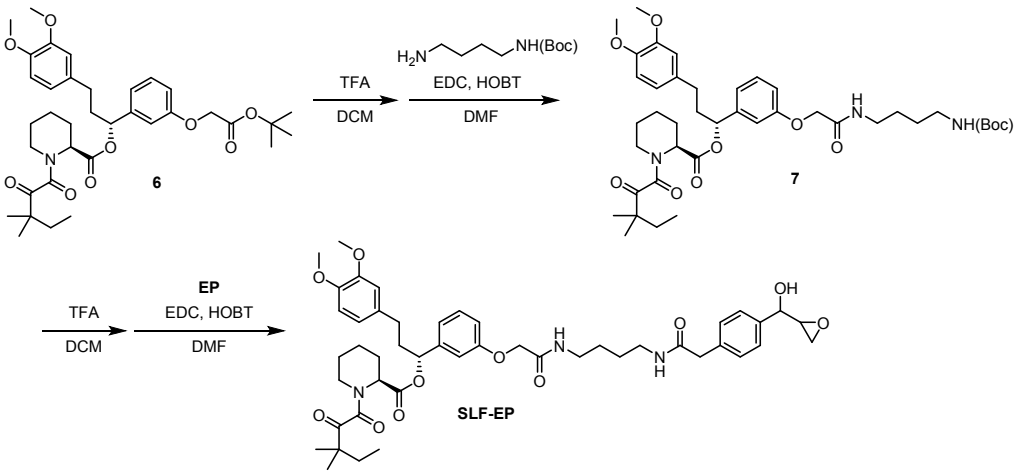
Compound 5

EDC (697 mg, 4.5 mmol), HOBT (607 mg, 4.5 mmol), and carboxybenzene sulfonamide (600 mg, 3 mmol) were dissolved in dry DMF (5 mL). After stirring at r.t. for half an hour, *tert*-butyl-*N*-(4-aminobutyl)carbamate (678 mg, 3.6 mmol) was added and then the solution was stirred at r.t. for another 16 h. The reaction mixture was washed with saturated brine solution and extracted with EA. After removing the solvent by evaporation, the residue was purified by flash column chromatography (DCM: MeOH= 50:1) to afford compound **5** as a light yellow oil (1.033 g, 93 % yield), which was directly used in the next step of synthesis after being characterized with MS (m/z, $C_{16}H_{26}N_3O_5S$): calculated $[M+H]^+$= 371.1; observed MS= 371.2.

CBS-EP

Boc-deprotection was carried out by adding 95 % TFA/DCM (3 mL) to compound **5** (215 mg, 0.58 mmol). After stirring at r.t. for 4 hours, the solvent and TFA were removed under vacuum. Without further purification, the residue was re-dissolved in dry DMF (1 mL) and added to a well-stirred **EP** (100 mg, 0.48 mmol) in DMF (3 mL) solution with EDC (111 mg, 0.72 mmol) and HOBT (97 mg, 0.72 mmol). The reaction was maintained for 8 hours. After the removing the solvents under vacuum, the residue was purified by column chromatography (DCM: MeOH = 15:1) to afford the final product **CBS-EP** (211 mg, 95 % yield) as a colorless oil. HR-ESI MS (m/z, $C_{22}H_{27}N_3O_6S$): calculated $[M+H]^+$: 462.1693; observed MS: 462.1687. Mixture of the diastereomers (dr ≈ 1 : 1): ^{1}H NMR (400 MHz, Methanol-*d*4) δ 8.01 (m, 4H), 7.37 (m, 4H), 4.59 (m, 0.5H), 4.39 (m, 0.5H), 3.55 (m, 2H), 3.44 (m, 2H), 3.40 (m, 1H), 3.27 (m, 2H), 3.15 (m, 1H), 2.79 (m, 1H), 1.65 (m, 4H). ^{13}C NMR (100 MHz, Methanol-*d*4) δ 172.60, 167.35, 146.17, 139.49, 137.75, 135.33, 128.75, 128.65, 127.59, 126.57, 126.33, 125.93, 74.41, 72.26, 55.77, 54.79, 54.42, 44.18, 42.40, 42.23, 39.33, 38.82, 26.46, 26.34.

b) **SLF-EP.**



Compound 7

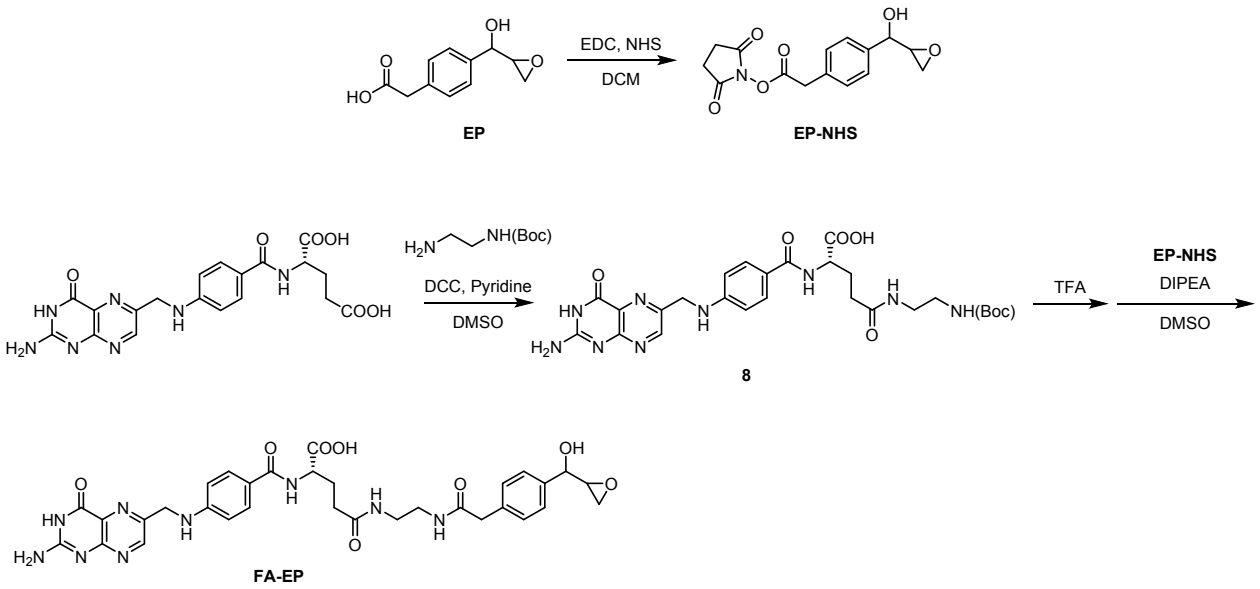
The precursor compound **6** was prepared following a previous report.[8] Compound **6** (82 mg, 0.14 mmol) in DCM (6 mL) was added with TFA (2 mL) at 0 ℃. The reaction mixture was then slowly warmed to r.t. and stirred for another 3 hours. After removing the solvents under vacuum, the residue was dissolved in DMF (3 mL) and the resulting deprotected compound **6** was directly used by for next step without further purification.

The residue was dissolved in DMF (4 mL). After the addition of EDC (33 mg, 0.21 mmol) and HOBT (29 mg, 0.21 mmol), the solution was stirred at r.t. for 0.5 h, before *tert*-butyl-*N*-(4-aminobutyl)carbamate (32 mg, 0.17 mmol) was added. The reaction was maintained for another 16 h. After washed with saturated brine solution and extracted with EtOAc, the solvent was removed by vacuum. The residue was subjected to column chromatography (DCM: MeOH= 20:1) to afford the compound **7** (99 mg, 93 % yield) as a colorless oil, which was directly used in the next step of synthesis after being characterized with MS (m/z, $C_{41}H_{59}N_3O_{10}$): calculated $[M+Na]^+$: 775.4; observed MS: 775.8.

SLF-EP

Deprotection of the boc group was carried out by adding 50 % TFA/DCM (3 mL) to compound **7** (90 mg, 0.13 mmol) at 0 ℃. After stirred at r.t. for 3 hours, the solvent and acid were removed by vacuum. The residue was re-dissolved in dry DMF (1 mL) and added to a well-stirred **EP** (25 mg, 0.12 mmol) in DMF (1 mL) with EDC (28 mg, 0.18 mmol) and HOBT (24 mg, 0.18 mmol). The reaction was maintained for another 12 h. After removing the solvents under vacuum, the residue was purified by column chromatography (DCM: MeOH = 20:1) to give the final product **SLF-EP** (64 mg, 63 % yield) as a colorless oil. HR-ESI MS (m/z, $C_{47}H_{61}N_3O_{11}$): calculated $[M+H]^+$: 844.4379; observed MS: 844.4389. Mixture of the diastereomers (dr ≈ 1 : 1): ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.06 (m, 11H), 5.85 (m, 1H), 5.33 (m, 1H), 4.45 (m, 3H), 3.91 (m, 8H), 3.61 (m, 2H), 3.35 (m, 1H), 3.23 (m, 4H), 2.94 (m, 2H), 2.63 (m, 1H), 2.33 (m, 1H), 2.08 (m, 2H), 1.75 (m, 2H), 1.48 (m, 7H), 1.27 (m, 9H), 0.92 (m, 3H). ^{13}C NMR (100 MHz, $CDCl_3$) δ 201.23, 171.19, 169.66, 168.36, 167.42, 158.83, 142.12, 139.03, 135.14, 133.32, 130.01, 129.54, 127.10, 126.98, 120.16, 120.08, 113.81, 113.51, 111.78, 111.39, 74.03, 71.01, 67.22, 55.92, 55.85, 54.98, 51.27, 46.69, 44.15, 43.92, 43.41, 39.14, 38.63, 38.16, 32.44, 32.19, 31.71, 31.23, 29.67, 26.90, 26.60, 26.39, 24.91, 23.40, 23.19, 21.13, 8.73.

c) **FA-EP**



EP-NHS

N-hydroxysuccinimide (55 mg, 0.48 mmol) was added to a well-stirred **EP** (84 mg, 0.4 mmol) solution with EDC (75 mg, 0.48 mmol) in DCM (3 mL). The reaction mixture was allowed to stir at r.t. for 6 h. The reaction mixture was then washed with saturated NH_4Cl (a.q.), $NaHCO_3$ (a.q.), and brine solution extracted with EtOAc. After removing the solvents under vacuum, the residue was purified by silica column chromatography (hexane: EA= 2:1) to give the product **EP-NHS** (102 mg, 84 % yield). Mixture of the diastereomers (dr ≈ 1 : 1): ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.38 (m, 8H), 4.91 (d, *J* = 2.9 Hz, 1H), 4.46 (d, *J* = 5.2 Hz, 1H), 3.93 (m, 4H), 3.21 (m, 2H), 2.93 (m, 1H), 2.83 (m, 10H), 2.74 (m, 1H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 169.00, 166.65, 139.79, 139.11, 131.36, 129.59, 129.53, 126.88, 126.78, 74.11, 70.43, 60.38, 55.80, 54.92, 53.41, 45.34, 43.51, 37.26, 25.54, 21.01, 14.16.

Compound 8

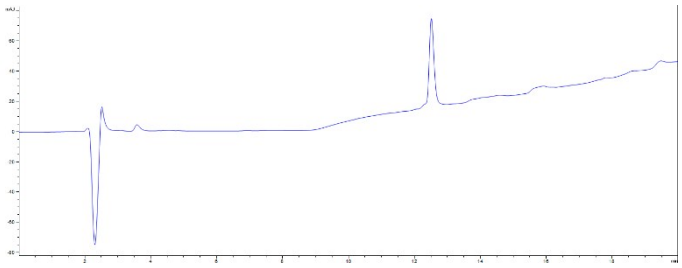
N-Boc-ethylenediamine (200 mg, 1.25 mmol) and DCC (386 mg, 1.88 mmol) were added to a solution of **FA** (550 mg, 1.25 mmol) in dry DMSO (15 mL) and pyridine (8 mL) under Ar protection. The reaction mixture was stirred at r.t. for 24 h. After filtration, the solution was diluted by Et_2O (400 mL) and stirred vigorously at 0 ℃. The orange precipitate was collected by vacuum filtration and washed with Et_2O to afford compound **8** (524 mg, 69 % yield) as a light orange powder, which was directly used in the next step of synthesis after being characterized with MS (m/z, $C_{26}H_{33}N_9O_7$): calculated $[M+Na]^+$: 605.2; observed MS: 604.9.

FA-EP

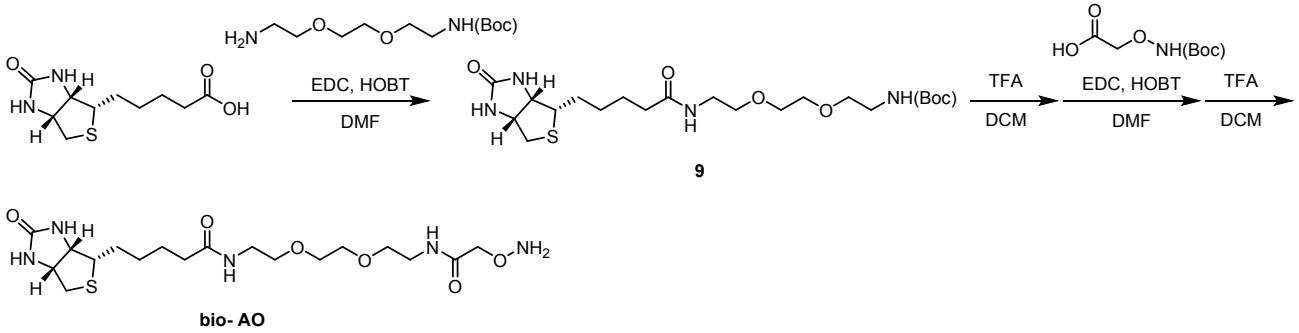
Compound **8** (300 mg, 0.51 mmol) was dissolved in 95 % TFA/DCM (2 mL) and stirred for 3 hours. The acid and solvent were removed by vacuum and the dark red residue was re-dissolved in DMF (1 mL). After the addition of TEA (1 mL), light yellow precipitate was formed. After centrifugation and washing with acetone and

ACN several times, the deprotected compound **8** (74 mg, 30% yield) was acquired and used directly for next step without further purification.

EP-NHS (26 mg, 0.12 mmol) and DIPEA (1 mL) were added to a solution in DMSO (0.5 mL) containing the deprotected compound **8** (38 mg, 0.08 mmol). The reaction mixture was stirred till clarity (~0.5 h). Then the reaction mixture was poured into a mixture of 20 % acetone/Et_2O solution. The yellow precipitate was centrifuged and washed with acetone and Et_2O for four times to afford the compound **FA-EP** (18 mg, 33 % yield) as a yellow powder analyzed by HR-ESI MS (m/z, $C_{32}H_{35}N_9O_8$): calculated $[M+H]^+$: 674.2681; observed MS: 674.2702 and HPLC (spectrum as below, absorbance at 220 nm).



d) Synthesis of **bio-AO.**

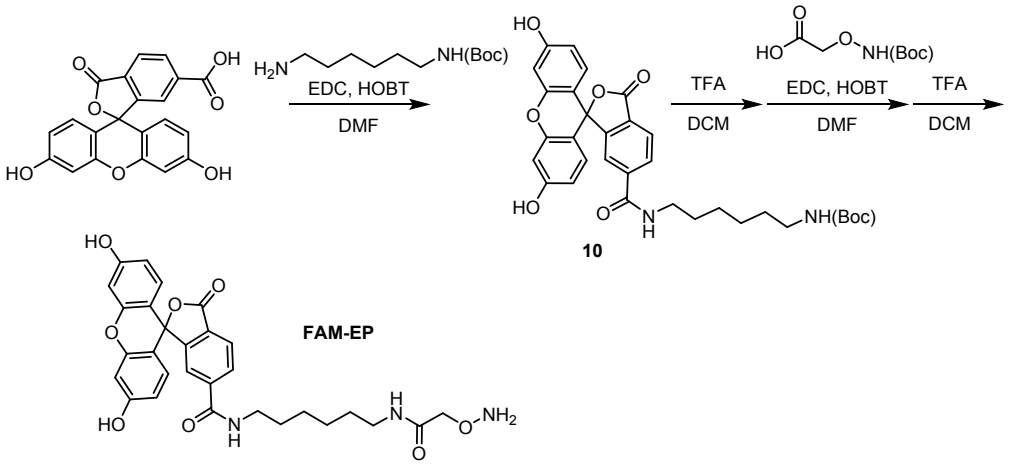


Compound 9

To a solution containing biotin (200 mg, 0.82 mmol) in dry DMF (5 mL), EDC (190 mg, 1.23 mmol) and HOBT (160 mg, 1.23 mmol) were added. After stirred at r.t. for half an hour, boc-1-amino-3,6-dioxa-8-octanediamine (244 mg, 1 mmol) was added and the mixture was stirred at r.t. for another 16 h. The solution was then washed with saturated brine solution and extracted with EA. After removing the solvents under vacuum, the residue was purified by flash column chromatography (DCM: MeOH= 10:1) to give compound **9** (375 g, 96 % yield) as a colorless oil, which was directly used in the next step of synthesis after being characterized with MS. (m/z, $C_{21}H_{38}N_4O_6S$): calculated $[M+Na]^+$= 497.2; observed MS= 497.1.

bio-AO

Deprotection of boc was carried out by adding 95 % TFA/DCM (5 mL) to compound **9** (397 mg, 0.8 mmol). After stirred at r.t. for 4 hours, the solvent and acid were removed by vacuum. Without further purification, the residue was re-dissolved in dry DMF (2 mL) and added to a well-stirred solution in DMF (3 mL) containing (boc-aminooxy)acetic acid (126 mg, 0.66 mmol), EDC (155mg, 1 mmol), and HOBT (135 mg, 1 mmol). The reaction was maintained for another 16 hours, washed with saturated brine solution, and extracted with EtOAc. After removing the organic solvents under vacuum, the residue was purified with column chromatography (DCM: MeOH = 10:1). The protection group was removed by the addition of 95 % TFA/DCM (5 mL). The acid and solvent were then removed by vacuum to give the final product **bio-AO** (220 mg, 74 % yield) as a colorless oil. HR-ESI MS (m/z, $C_{18}H_{33}N_5O_6S$): calculated $[M+H]^+$: 448.2224; observed MS: 448.2213. ^{1}H NMR (400 MHz, DMSO-d_6) δ 4.42 (m, 2H), 4.31 (dd, J = 7.6, 4.8 Hz, 1H), 4.13 (dd, J = 7.7, 4.4 Hz, 1H), 3.51 (s, 1H), 3.42 (m, 3H), 3.28 (q, J = 5.7 Hz, 2H), 3.19 (m, 2H), 2.82 (dd, J = 12.5, 5.1 Hz, 1H), 2.51 (m, 1H), 2.07 (t, J = 7.4 Hz, 2H), 1.48 (m, 6H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 172.68, 167.49, 163.24, 71.85, 69.99, 69.95, 69.58, 69.22, 67.12, 61.53, 59.69, 55.88, 38.87, 38.82, 35.55, 28.65, 28.49, 25.72.



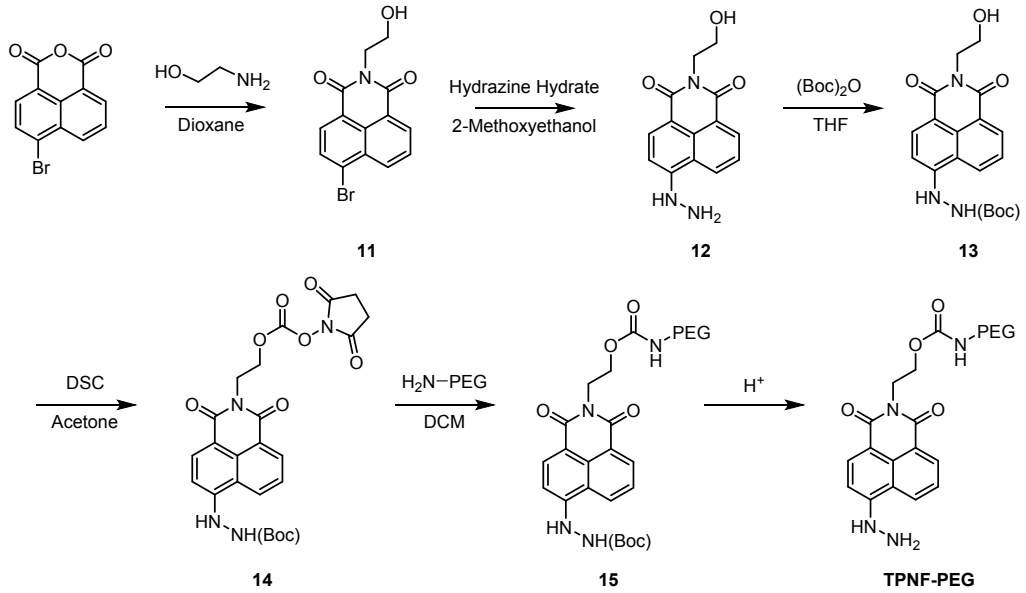
e) Synthesis of **FAM-AO.**

Compound 10

To a solution of 6-carboxyfluorescein (376 mg, 1 mmol) in dry DMF (5 mL), added with EDC (232 mg, 1.5 mmol) and HOBT (202 mg, 1.5 mmol). After stirred at r.t. for half an hour, *N-tert*-butoxycarbonyl-1,6-hexanediamine (237 mg, 1.1 mmol) was added. The solution was stirred at r.t. for another 16 h and then washed with saturated brine solution and extracted with EA. After removing the solvent by evaporation, the residue was purified by flash column chromatography (DCM: MeOH= 10:1) to give compound **10** (244 mg, 42 % yield, which was directly used in the next step of synthesis after being characterized with MS. (m/z, $C_{32}H_{34}N_2O_8$): calculated $[M+Na]^+$= 597.2; observed MS= 597.1.

FAM-AO

Deprotection of boc was carried out by adding 95 % TFA/DCM (5 mL) to compound **10** (45 mg, 0.08 mmol). After stirred at r.t. for 4 hours, the solvent and acid were removed by vacuum. Without further purification, the residue was re-dissolved in dry DMF (1 mL) and added to a well-stirred (boc-aminooxy)acetic Acid (13.77 mg, 0.07 mmol) in DMF (1 mL) with EDC (17 mg, 0.11 mmol) and HOBT (15 mg, 0.11 mmol). The reaction was maintained for 16 hours and then washed with saturated brine solution and extracted with EA. After removing the organic solvents under vacuum, the residue was re-dissolved in 20 % ACN/H_2O and purified by HPLC. After lyophilization, the protecting group was removed by 95 % TFA/DCM (2 mL). The acid and solvent were then removed by vacuum giving final product **FAM-AO** (21 mg, 47 % yield). HR-ESI MS (m/z, $C_{29}H_{29}N_3O_8$): calculated $[M+H]^+$: 548.2027; observed MS: 548.2017. ^{1}H NMR (400 MHz, DMSO-d_6) δ 10.23 (s, 2H), 8.70 (m, 1H), 8.17 (m, 1H), 8.07 (m, 1H), 8.00 (m, 1H), 7.67 (m, 1H), 6.72 (m, 2H), 6.57 (m, 4H), 4.21 (s, 2H), 3.18 (m, 2H), 3.07 (m, 2H), 1.42 (m, 5H), 1.24 (m, 5H).



f) Synthesis of **TPNF.**

Compound 11

4-bromo-1,8-naphthalic anhydride (1.358 g, 5 mmol) and 2-aminoethanol (305 mg, 5 mmol) were added to 1,4-dioxane (25 mL) and refluxed for 5 hours. After cooling to 0 ℃, the yellow precipitate was collected by vacuum filtration and washed with Et_2O to give compound **11** (1.052 g, 66 % yield) as a yellow solid. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.68 – 8.63 (m, 1H), 8.58 (m, 1H), 8.41 (d, J = 7.9 Hz, 1H), 8.04 (d, J = 7.9 Hz, 1H), 7.85 (m, 1H), 4.44 (m, 2H), 3.98 (m, 2H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 164.47, 133.60, 132.34, 131.50, 131.16, 130.68, 129.05, 128.12, 122.82, 121.89, 61.59, 42.82.

Compound 12

Compound **11** (560 mg, 1.75 mmol) and hydrazine (1 M, 6 mL, 6 mmol) in 2-methyloxyethanol (20 mL) were refluxed for 3 h. After cooling to r.t., the reaction mixture was concentrated and poured to ice cold water. The orange precipitate was then collected and purified by recrystallization from chlorobenzene to give compound **12** (288 mg, 61 % yield). ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.13 (m, 1H), 8.63 (m, 1H), 8.44 (m, 1H), 8.31 (dd, J = 8.6, 2.5 Hz, 1H), 7.66 (m, 1H), 7.27 (d, J = 8.6 Hz, 1H), 4.86 (t, J = 5.8 Hz, 1H), 4.71 (m, 2H), 4.15 (t, J = 6.7 Hz, 2H), 3.62 (m, 2H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 164.39, 163.51, 153.58, 134.62, 131.00, 129.78, 128.62, 124.54, 122.25, 118.86, 104.58, 104.41, 58.44, 41.81.

Compound 13

Compound **12** (271 mg, 1 mmol) and di-*tert*-butyl decarbonate (327 mg, 1.5 mmol) were added to DMF (4 mL) and heated to 70 ℃ for 8 h. After removing the organic solvent by vacuum, the residue was re-dissolved in

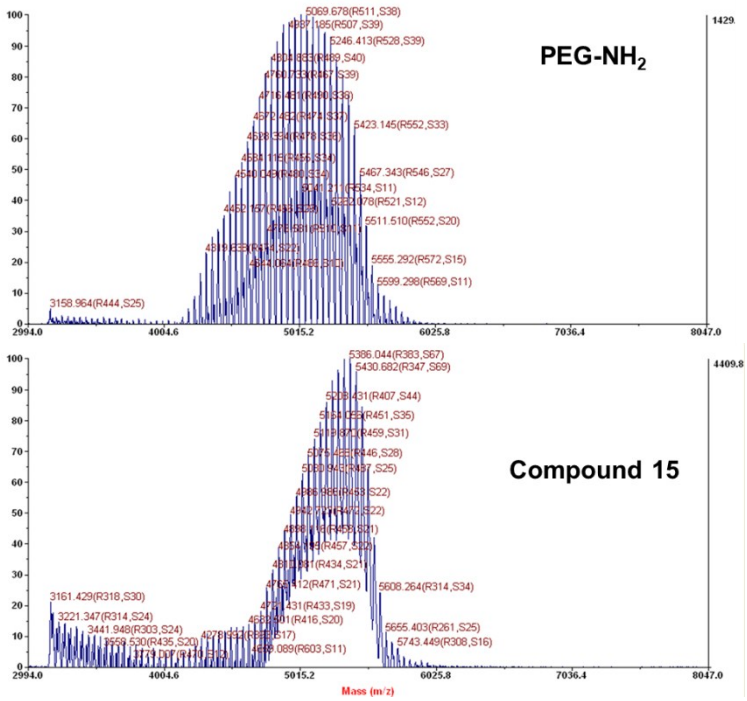
40 % ACN/H_2O and purified by HPLC to give compound **13** (115 mg, 30 % yield). ^{1}H NMR (500 MHz, methanol-*d*4) δ 8.19 (m, 3H), 7.29 (m, 1H), 6.90 (m, 1H), 4.24 (m, 2H), 3.81 (t, *J* = 6.4 Hz, 2H), 1.56 (m, 9H). ^{13}C NMR (125 MHz, Methanol-*d*4) δ 164.58, 164.25, 163.20, 156.92, 150.99, 133.53, 130.37, 127.21, 124.48, 121.53, 111.08, 104.43, 80.86, 58.94, 41.49, 27.24.

Compound 14

N, *N*'-disuccinimidyl carbonate (21 mg, 0.08 mmol) and compound **13** (20 mg, 0.05mmol) were dissolved in acetone (1 mL) with triethylamine (22 μL). After stirred at 55 ℃ for 16 h, the solvent was removed by vacuum. The residue was then re-dissolved in 30 % ACN/H_2O and purified by HPLC (C18) and lyophilized for further reaction to give compound **14** (13 mg, 47 % yield), which was directly used in the next step of synthesis after being characterized with MS (m/z, $C_{24}H_{24}N_4O_9$): calculated $[M+H]^+$ = 513.2; observed MS = 513.1 and directly used in the next step.

Compound 15

Amine-PEG (5 kDa, 11 mg, 0.002 mmol) and compound **14** (2 mg, 0.003 mmol) were dissolved in DCM (100 μL). After shaking 4 h at r.t., the solvent was removed by Speedvac and another portion of compound **14** (2 mg, 0.003 mmol) in DCM (100 μL) was added and reacted for another 2 h. Et_2O (1 mL) was then added to the reaction mixture and the precipitate was collected. The residue was purified by HPLC to afford compound **15** (9 mg), which was analyzed by MALDI-TOF MS (shown below). PEG-NH_2 (highest peak): m/z= 5069.678;



compound **15** (highest peak): m/z= 5386.044.

TPNF (PEG conjugate)

Hydrazine is very toxic and reactive; it can even react with trace of formaldehyde in the atmosphere. As a result, the deprotection was only carried out immediately before the experiment. Compound **15** (5 mg) was dissolved in degassed 3 N HCl (100 μL) and reacted for 1 h under Ar protection. Then Et_2O (1 mL) was added. The precipitate was collected and washed with Et_2O. After removed all the solvent by Speedvac, the **TPNF** PEG conjugate (5 mg) was collected for further experiment.

6. Reference

1. Y. Takaoka, H. Tsutsumi, N. Kasagi, E. Nakata and I. Hamachi, *J. Am. Chem .Soc.*, 2006, **128**, 3273-3280.
2. A. Y. Kornilova, B. Algayer, M. Breslin, G. H. Addona and V. Uebele, *Anal. Biochem.*, 2013, **432**, 59-62.
3. I. Chen, M. Howarth, W. Lin and A. Y. Ting, *Nat. Methods*, 2005, **2**, 99-104.
4. A. Shevchenko, H. Tomas, J. Havli, J. V. Olsen and M. Mann, *Nat. Protocols*, 2007, **1**, 2856-2860.
5. M. Bigatti, A. Dal Corso, S. Vanetti, S. Cazzamalli, U. Rieder, J. Scheuermann, D. Neri and F. Sladojevich, *ChemMedChem*, 2017, **12**, 1748-1752.
6. M. Wichert, N. Krall, W. Decurtins, R. M. Franzini, F. Pretto, P. Schneider, D. Neri and J. Scheuermann, *Nat. Chem.*, 2015, **7**, 241-249.
7. Y. Zhou, C. Li, J. Peng, L. Xie, L. Meng, Q. Li, J. Zhang, X. D. Li, X. Li, X. Huang and X. Li, *J. Am. Chem. Soc.*, 2018, **140**, 15859-15867.
8. T. Keenan, D. R. Yaeger, N. L. Courage, C. T. Rollins, M. E. Pavone, V. M. Rivera, W. Yang, T. Guo, J. F. Amara, T. Clackson, M. Gilman and D. A. Holt, *Bioorg. Med. Chem.*, 1998, **6**, 1309-1335.