Supporting Information for

A Genetically Encodable and Fluorogenic Photo-crosslinker via Photo-induced Defluorination Acyl Fluoride Exchange

Jielin Fu, Sitong Li, Lijun Deng, Xiaohu Zhao and Zhipeng Yu*

Key Laboratory of Green Chemistry and Technology of Ministry of Education,

College of Chemistry, Sichuan University, 29 Wangjiang Road,

Chengdu 610064, P. R. China

Email: *zhipeng@scu.edu.cn*

Table of Contents

General Information

Unless otherwise indicated, all solvents and starting materials were purchased from commercial sources and used directly without further purification. Flash column chromatography was performed by using 200-300 mesh silica gel. Anhydrous solvents, were purchased from Acros Organics (THF), and commercially available chemicals were obtained from Adamas, Acros Organics, Aldrich Chemical Co., Alfa, Aesar and TCI. Exact ESI mass spectra were scanned by a Thermo Q-Exactive™ mass spectrometry instrument. LC/ESI-MS data were obtained on a Thermo LTQ-XL mass spectrometer. The ^{1}H, ^{13}C and ^{19}F NMR spectra were recorded on a Brüker Avance 400 spectrometer (^{1}H: 400 or 600 MHz, ^{13}C: 101 or 151 MHz, ^{19}F: 376 MHz). Chemical shifts (δ) for ^{1}H and ^{13}C NMR spectra are given in ppm relative to TMS. The residue solvent signals were used as references for ^{1}H and ^{13}C NMR spectra and the chemical shifts converted to the TMS scale (Chloroform-*d*, 7.26 ppm for ^{1}H NMR and 77.16 ppm for ^{13}C NMR; Dimethyl sulfoxide-d_6, 2.50 ppm for ^{1}H NMR and 39.5 ppm for ^{13}C NMR; Deuterium Oxide, 4.79 ppm for ^{1}H-NMR). Shifts multiplicity was reported as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, brs. = broad. The photo-irradiation power density of 311 nm UV lamp (2.9 mW cm^{-2} or 5.9 mW cm^{-2}, single wavelength output after an optical filter) was measured via an optical power meter purchased from Thorlabs.

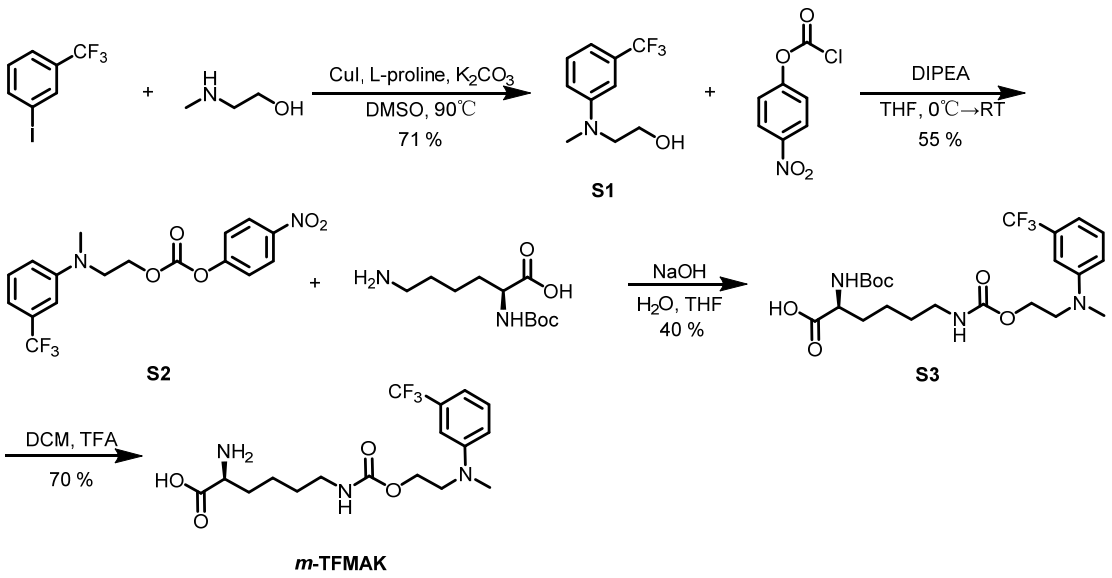
Ultraviolet and visible (UV-Vis) absorption spectra were recorded by using 1.0 cm quartz cuvettes on Thermo NANODROP 2000C Spectrophotometer. Fluorescence emission spectra were recorded by using 1.0 cm quartz cuvettes on Horiba® Fluoromax-4 Spectrofluorometer Detector. Horiba® Duetta fluorescence and absorbance spectrometer and Applied Photo physics® RX2000 stopped-flow device were used for kinetic studies. Exact ESI mass spectra were recorded on a SHIMADZU LCMS-IT-TOF, flow rate = 1.0 mL /min. LC-ESI-MS were obtained on a Thermo LTQ-XL mass spectrometer.

AccuPrime™ *pfx* DNA Polymerase was obtained from Invitrogen. T4 DNA ligase and restriction endonucleases were purchased from Thermo Scientific or NEB®. E.Z.N.A.® Plasmid Mini Kit I (from Omega Bio-tek®) was used for purifying restriction digested DNA and gel-purified DNA according to the manufacturer protocol. Ni-NTA resin was obtained from Thermo

Fisher for purification of protein containing His_6-tag. *E. coli* DH5α competent cells and *E. coli* BL21 (DE3) competent cells (TransGen Biotech) were obtained from commercial source for chemical transformation. Optical density (OD) was measured by using Thermo NANODROP 2000C Spectrophotometer. Images for In-gel fluorescence, Coomassie brilliant blue staining and western blot analysis were captured on a ChampChemi™ 610 Fluorescence/Chemiluminescence Gel Imager. Living cell fluorescence imaging experiments were carried out on an Olympus IX83 epifluorescence microscope. The synthesis of oligonucleotide primers, gene fragments and DNA sequencing were serviced and conducted by Tsingke Biotechnology. Antibodies for Western-Blot analysis were purchased from either Beyotime Biotechnology or ZENBIOSCIENCE. An ECL chemiluminescence kit for WB analysis was also purchased from Beyotime Biotechnology.

The sequences for all the primers used for plasmids construction or site-directed mutation (SDM) are listed in **Table S1**. All the amino acid sequences of proteins mentioned in this work are presented in **Supplementary Notes**.

Synthesis of noncanonical amino acids (*m*-TFMAK)



A mixture of 3-Iodobenzotrifluoride (1.0 eq), 2-methylaminoethanol (1.5 eq), K_2CO_3 (2.0 eq), CuI (0.1 eq) and L-proline (0.2 eq) in DMSO was stirred over night at 90 °C under argon. The reaction was quenched by water and was extracted with EtOAc for 3 times. The organic layer was dried over anhydrous sodium sulfate, concentrated under vacuo and purified by silica chromatography (PE/EA = 2/1) to get **S1** as yellow oil (71%). ^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.33-7.27 (m, 1H), 6.98-6.94 (m, 2H), 6.90 (dd, *J* = 8.4, 2.6 Hz, 1H), 3.78 (t, *J* = 5.6 Hz, 2H), 3.48 (t, *J* = 5.7 Hz, 2H), 2.98 (s, 3H). HRMS (ESI) calcd. for $C_{10}H_{13}F_3NO^+$ $[M+H]^+$ m/z 220.0944, found 220.0944.

To a stirred solution of **S1** (1 eq), 4-nitrophenyl chloroformate (1.6 eq) in THF was added DIPEA (3.0 eq) dropwise at 0 °C. The mixture was warmed to room temperature and stirred overnight. The mixture was concentrated under vacuo and purified by silica chromatography (PE/EA = 6/1) to gain **S2** yellow solid (55%). ^{1}H NMR (400 MHz, Chloroform-*d*) δ 8.24 (d, *J* = 2.2 Hz, 1H), 8.23 (d, *J* = 2.2 Hz, 1H), 7.33 (t, *J* = 7.9 Hz, 1H), 7.24 (d, *J* = 2.2 Hz, 1H), 7.22 (d, *J* = 2.2 Hz, 1H), 6.98 (d, *J* = 7.7 Hz, 1H), 6.94 (s, 1H), 6.93 (d, *J* = 2.3 Hz, 1H), 4.48 (t, *J* = 5.7 Hz, 2H), 3.76 (t, *J* = 5.7 Hz, 2H), 3.06 (s, 3H). ^{13}C NMR (101 MHz, Chloroform-*d*) δ 155.45, 152.61, 149.01, 145.59, 129.88, 131.75 (d, *J* = 31.6 Hz), 125.40, 123.63 (d, *J* = 356.6 Hz), 121.86, 115.20, 113.47, 108.52, 66.26, 50.92, 38.68. ^{19}F NMR (376 MHz, Chloroform-*d*) δ -62.73. HRMS (ESI) calcd. for

$C_{17}H_{16}F_3N_2O_5^+$ $[M+H]^+$ m/z 385.1006 , found 385.1007.

The product **S2** (1.0 eq) was dissolved in THF and added dropwise to a stirring solution of Boc-Lys-OH (1.5 eq) in NaOH (1.0 M) at 0 °C. And the resultant mixture was stirred at 0 °C and detected by TLC analysis. The mixture was regulated to pH = 4-5 by saturated sodium citrate solution and was extracted by EtOAc for 3 times. The organic layer was dried over anhydrous sodium sulfate, concentrated under vacuo and purified by silica chromatography (from PE/EA = 6/1 to DCM/MeOH = 10/1) to get **S3** as yellow oil (40%). ^{1}H NMR (400 MHz, Dimethyl sulfoxide-d_6) δ 7.34 (t, J = 8.3 Hz, 1H), 7.09 (t, J = 5.7 Hz, 1H), 6.98 (dd, J = 8.5, 2.5 Hz, 2H), 6.89 (d, J = 6.5 Hz, 2H), 4.09 (t, J = 5.9 Hz, 2H), 3.88-3.77 (m, 1H), 3.59 (t, J = 5.9 Hz, 2H), 2.96 (s, 3H), 2.94-2.87 (m, 2H), 1.72-1.45 (m, 1H), 1.37 (s, 8H), 1.35-1.29 (m, 1H), 1.29-1.08 (m, 1H). ^{13}C NMR (101 MHz, Dimethyl sulfoxide-d_6) δ 174.73, 163.40, 156.53, 156.08, 149.53, 130.52 (q, J = 30.8 Hz),130.29, 125.02 (q, J = 272.4 Hz),115.72, 112.09 (q, J = 3.9 Hz), 107.70 (q, J = 3.9 Hz), 79.59, 78.37, 61.05, 53.91, 51.09, 40.54, 40.38, 40.33, 40.12, 39.91, 39.70, 39.49, 39.29, 38.74, 30.88, 29.41, 28.61, 23.33. ^{19}F NMR (376 MHz, Dimethyl sulfoxide-d_6) δ -61.31. HRMS (ESI) calcd. for $C_{22}H_{33}F_3N_3O_6^+$ 492.2316 [M+H$^+$], found 492.2316.

The product **S3** was dissolved in DCM, and then TFA (DCM/TFA = 1/1) was added to the solution. The mixture was stirred for 30 min and tracked by LC-MS, and then blew the resulting reaction mixture with nitrogen flow to get rid of the TFA and the solvent. The crude product was then dissolved with DMSO and purified by reverse preparative C18 chromatography with 5% water and 95% acetonitrile to obtain the desired ncAA as a pale-yellow solid (75%). ^{1}H NMR (600 MHz, Deuterium Oxide) δ 7.14 (t, J = 8.1 Hz, 1H), 6.83 (s, 1H), 6.79 (d, J = 7.5 Hz, 1H), 6.77 (d, J = 8.2 Hz, 1H), 4.00 (d, J = 5.7 Hz, 2H), 3.37 (d, J = 5.3 Hz, 2H), 2.78 (t, J = 7.1 Hz, 2H), 2.69 (s, 3H), 1.46-1.36 (m, 1H), 1.35-1.28 (m, 1H), 1.22-1.13 (m, 2H), 1.10-1.02 (m, 2H). ^{13}C NMR (151 MHz, Deuterium Oxide) δ 183.48, 157.91, 149.45, 130.67 (q, J = 30.6 Hz), 129.70, 124.45 (q, J = 272.2 Hz), 116.07, 112.98, 108.72, 61.97, 55.86, 50.97, 40.13, 37.42, 34.37, 28.75, 22.26. ^{19}F NMR (376 MHz, Deuterium Oxide) δ -61.12. HRMS (ESI) calcd. for $C_{17}H_{25}F_3N_3O_4^+$ $[M+H]^+$ m/z 392.1792, found 392.1792.

Materials and Methods

HPLC analysis of competitive test among N^{α}-Boc protected L-Cystine, L-lysine and L-Histidine

N^{α}-Boc protected *m*-TFMAK was diluted with ACN to make a stock solution of 10 mM, N^{α}-Boc-Cystine, Histidine or Lysine was diluted with ACN to make a stock solution of 50 mM, respectively. N^{α}-Boc protected *m*-TFMAK was diluted in ACN/PBS = 1/1 to 100 μM, and N^{α}-Boc protected L-lysine, N^{α}-Boc protected L-Cystine and N^{α}-Boc protected L-Histidine were diluted in ACN/PBS = 1/1 to 500 μM, respectively. N^{α}-Boc protected *m*-TFMAK (100 μM, 1.0 eq) and N^{α}-Boc protected natural amino acid (500 μM, 5.0 eq) were mixed in ACN/PBS = 1/1 with 311 nm activation for 3 min, then the resulting mixtures were analyzed by HPLC-MS to quantify the conversion and yield. To respectively calculate the ratio of the products after photo-DAFEx in mixed solution, we separated the product of hydrolyzation or conjugation with N^{α}-Boc protected L-lysine and presumed that the photo-DAFEx reactions with N^{α}-Boc protected L-Cystine or L-Histidine were completely transformed into corresponding conjugation products or the hydrolyzation. Using the isolated products of hydrolyzation or conjugation with N^{α}-Boc protected L-lysine as an external standard, the results of competitive test were quantified separately and plotted by Origin.

Kinetic study of the photo-DAFEx reaction between corresponding benzoyl fluoride and N^{α}-Boc-lysine

N^{α}-Boc protected *m*-TFMAK was diluted with ACN to make a stock solution of 10 mM, and N^{α}-Boc-lysine was diluted with ACN/PBS = 1/1 to make a stock solution of 50 mM, respectively. The stock solution of N^{α}-Boc protected *m*-TFMAK was diluted by ACN/PBS = 1/1 to 200 μM for the stopped-flow analysis. And the N^{α}-Boc protected lysine stock solution was also diluted by ACN/PBS = 1/1 to the final concentration of 500, 1000, 1500, and 2000 μM in the mixture for the stopped-flow analysis.

For kinetic study, the stopped-flow measurements were performed using an Applied Photo physis® RX2000 rapid kinetics spectrometer accessory connected to Horiba® Duetta fluorescence spectrometer. The N^{α}-Boc protected *m*-TFMAK and N^{α}-Boc-lysine were loaded in two parallel syringes on the stopped-flow device. After the N^{α}-Boc protected *m*-TFMAK solution was irradiated with 311 nm for 3 min to convert to corresponding N^{α}-Boc protected benzoyl fluoride, both the solutions were quickly injected into an optical cell with synchronous fluorescence tracking. Then the time-dependent fluorescence signals were read out by

monitoring the formation of the final benzamide product, at λex = 332 nm; λem = 436 nm. Each data set was collected in triplicate for various concentrations of N^{α}-Boc-lysine, which conformed the *pseudo*-first-order conditions, and the OriginPro software was used for analyzing the resulting data to derive the kinetic results.

Plasmid construction

All primer sequences used for plasmids construction or site-directed mutation (SDM) are listed in **Table S1**. All amino acid sequences of proteins mentioned in this work are presented in **Supplementary Notes**. Plasmid pET-sfGFP-Q204TAG-His_6 and pET-sfGFP-N149TAG-His_6 are reported in our previous work.[1]

The *sj*GST gene was synthesized by Sangon biotech, and then subcloned into pET-21a vector, by using BamHI-EcoRI restriction sites, to construct a plasmid, pET-*sj*GST-His_6. For the construction of *sj*GST mutant expression plasmids, pET-*sj*GST-His_6 was set as the template and the site-directed mutation was used according to the manufacturer's protocol.

Protein expression and purification

For protein expression and purification, pEvol-*m*-TFMAKRS and pET-POI-XTAG-His_6 (with a C-terminal hexa-histidine tagged POI, including sfGFP or *sj*GST gene harbouring an amber codon at X position) were co-transformed into competent *E. coli* BL21 (DE3) cells. Colonies were grown on LB agar containing tetracycline (34 μg·mL^{-1}) and ampicillin (100 μg·mL^{-1}). The grown single colony was picked and inoculated in 5.0 mL LB growth medium supplemented with corresponding antibiotics for overnight culture. Then the overnight culture was diluted into fresh LB medium containing antibiotics at ratio of 1:50. When the OD600 reached 0.4-0.5, 2.0 mM *m*-TFMAK was added to the medium, and protein expression was induced by 1.0 mM IPTG and 0.2% arabinose when OD600 reached 0.6. After 8 h induction at 37 °C, 250 rpm, cells were harvested by centrifugation (3000 rpm, 20 min, 4 °C) and bathed on ice. The harvested cells were resuspended in 2.0 mL lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH = 8.0) and lysed by sonication in an ice/water bath. Then the cell lysates were centrifuged (14000 rpm, 40 min, 4 °C) to get rid of insoluble debris. The supernatant was further

purified by Ni-NTA affinity column (Thermo Scientific). The protein-bound resins were washed with lysis buffer and washing buffer (50 mM Na_2HPO_4, 300 mM NaCl, 50 mM imidazole, pH = 8.0). The protein was finally eluted from the resin by using the elution buffer (50 mM Na_2HPO_4, 300 mM NaCl, 250 mM imidazole, pH = 8.0). The eluted fractions were collected, concentrated and subjected to buffer exchange to PBS buffer (pH = 7.4) using an Amicon Ultra-15 Centrifugal Filter (Millipore) and stored at −80 °C. The identities of purified proteins were confirmed by SDS-PAGE and mass spectrometry (LC-MS).

Verification assay for *m*-TFMAK GCE incorporation in *E. coli*

To investigate the incorporation efficiency of *m*-TFMAK in *E. coli*, pEvol-*m*-TFMAKRS and pET-sfGFP-N149TAG were co-transformed into competent BL21 (DE3) cells. After protein expression, cells were collected and washed with PBS buffer (pH = 7.4) for three times. Then the cells expressed sfGFP were diluted to a density of OD600 = 0.4-0.8 with PBS buffer for OD600 measurement by the NanoDrop 2000c spectrophotometer, while the sfGFP fluorescence of the same sample was measured by Horiba® Fluoromax-4 spectrofluorometer. The intensity of sfGFP fluorescence measurement was normalized by the OD600, and the final amber suppression efficiency was represented by the fluorescence intensity/OD600.

Molecular docking

The crystal structure of *Mm*OmeRS (PDB: 3QTC) with co-crystalized *O*-methyl-L-tyrosine and ANP was obtained from the Protein Data Bank (PDB). The crystal structure of *Mm*PylRS (PDB: 2ZIM) was obtained in the same way. Then, we reconstructed the binding domain by replacing the active site with the mutational amino acid (Y306A). Before docking, water molecular, solvent molecular and original ncAA ligands were removed from the crystal structure. The target *m*-TFMAK ligand was positioned roughly as guided by the original ncAA in the crystal structure and optimized to a minimized ligand energy by applying a CHARMm force field. The effective binding sites were defined within a 5.0 Å spherical space covering all of the active residues to furnish the *m*-TFMAK or *m*-TFMAK-ANP complex. Then, the CDOCKER algorithm was employed to search for a series of molecular docking results with the binding energy

sorted. All the computational experiments and conformational analysis for the *m*-TFMAK-*m*-TFMAKRS complexes were carried out via the Discovery Studio. The results of molecular docking were evaluated depending on the CDOCKER score and CDOCKER interaction energy of each complex against the related targets. As a demonstration, all the docking results were visualized by the Pymol software.

Mammalian cell culture

Human embryonic kidney 293T (HEK 293T) cells were grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) (Adams®) and 1% Pen-Strep (Penicillin-Streptomycin Solution 100×) (Adama®) at 37 °C and 5% CO_2 in a humidified incubator. For cell passage, the cells were detached with 0.25% trypsin-EDTA (Gibco) after incubation for 3.0 min at 37 °C and the suspended cells were treated with complete DMEM (10% FBS and 1% Pen-Strep) to neutralize the trypsin. Then the suspended cells were collected by gentle centrifuging (1000 rpm, 5 min) to get rid of the supernatant. The collected cells were reseeded at a density of about 2.0×10^6 cells in a new batch of 9.0 cm culture dishes supplied with DMEM with 10% FBS and 1% Pen-Strep. The culture medium was refreshed every 3 days after a subculture and the cells were regularly passaged until 10-20 generations.

Evaluation of the mammalian cell viability after treatment of *m*-TFMAK

HEK 293T cells were seeded in 96-well plates with a density of 30000-50000 per well and were cultured for 8-10 h before the compound treatment. Afterward, the cells were cultured with various concentrations of *m*-TFMAK in DMEM supplemented with 10% FBS and 1% Pen-Strep for 30 h. Then, the medium was replaced with fresh DMEM with 10% CCK-8 reagent, and then the cells were further incubated for 1 h prior to the viability testing. The cell viability was determined by monitoring the absorbance at 450 nm in comparison untreated groups by using a microplate reader (Bio-Rad).

Transfection procedure for HEK 293T cells

HEK 293T cells were seeded and cultured to 70-90% confluency in 35-mm glass-bottom tissue-culture dishes and grown in an antibiotic-free DMEM medium supplemented with 10% FBS under the atmosphere containing 5% CO_2 at 37 °C for 12 h. Then the pEM14 plasmid containing *m*-TFMAKRS and EGFP-N149TAG-HA was transiently transfected into the cells via incubation with the Lipofectamine™ (Thermo Fisher) in the opti-MEM medium according to manufacturers' protocol. After transfection for 5-6 h, the medium was replaced with fresh opti-MEM medium with or without 2.0 mM *m*-TFMAK. After 30 h of incubation, the cells were imaged with a Zeiss LSM780 Microscope (40× objective) by monitoring the EGFP fluorescence channel.

Semi-quantitative western blot analysis for mammalian cell lysates

HEK 293T cells were seeded in 6-well plates, and transfected as the previous procedures. The transfected and over-expressing HEK 293T cells were washed with chilled PBS buffer (pH = 7.4), and lysed on ice with RIPA lysis buffer (Sangon biotech). The lysates were collected by centrifuge (10 min, 12000 g, 4 °C) to get rid of insoluble debris. Total cellular protein was determined by bicinchoninic acid (BCA) kit (Beyotime Biotechnology) to ensure consistent loading. After resolving by SDS-PAGE, the separated proteins were transferred to polyvinylidene difluoride (PVDF) membrane via a semi-dry electroblotting. Followed by blocking and washing, anti-HA-tag mouse monoclonal antibody (1:500) was used to recognize and binding to the expressed target protein. The PVDF membrane was then washed for three times before incubating with HRP-labeled goat anti-mouse IgG (H+L) secondary antibody (1:1000). The final quantification of HRP activity was performed by using an ECL substrate (Beyotime Biotechnology). Anti-β-actin mouse monoclonal antibody (1:1000) was used as a reference for the housekeeping protein. Chemiluminescent signals were determined by a gel imager (ChampChemi™ 610Plus, SAGECREATION) and analyzed by ImageJ software. The EGFP-HA band signal was normalized by the internal loading control, β-actin. The suppression efficiency in mammalian cells was semi-quantified by normalized data compared with wild-type EGFP-HA as a reference.

Photo-DAFEx reaction in living cells

After transfecting the plasmid pEM14, the HEK293T cells expressed intact EGFP-N150TAG-HA (with adding 2.0 mM *m*-TFMAK), truncated EGFP-N150TAG-HA (without adding *m*-TFMAK) or wt-EGFP-HA were washed with PBS (pH = 7.4) for three times. The cells were exposed to 311 nm for 5.0 min or dark as control. Then, the cells were lysed as described previously, and resolved by SDS-PAGE, in-gel fluorescence and Western blot with anti-HA-tag mouse monoclonal primary antibody (1:500), HRP-labeled goat anti-mouse IgG (H+L) secondary antibody (1:1000), followed by ECL substrate (Beyotime Biotechnology) for imaging.

Procedures for in-gel digestion and subsequent LC-MS/MS analysis and data mining

The protein bands of interest were excised from the SDS-PAGE gels and cut into cubes (1.0 mm^3). The gel cubes were discolored and dehydrated by shaking in acetonitrile in an Eppendorf centrifuge tube. The dehydrated gels were incubated with were treated with 10 mM DTT (in 25 mM ABC solution) at 56 °C for 1 h. Then, the supernatants were removed, and the gel samples were alkylated with 55 mM IAA (in 25 mM ABC solution) in dark for 45 min at RT. After centrifugation, the supernatants were removed, and the gel samples were washed again with ACN. The in-gel protein samples were digested by MS grade trypsin (1.0 μg, trypsin/protein = 1/50, *w/w*, in 50 mM ABC solution) at 37 °C overnight. After removing supernatants, the gels were washed and dehydrated by ACN. The resulting peptides were gradient extracted by the extraction buffer (from 50%ACN/5%TFA, 75%ACN/0.1%TFA to 100% ACN), then using vacuum-centrifuged to dryness. After desalination via Ziptip® pipette tip (Millipore) and dryness, the digested peptides were subjected to fragmentation and b/y ion signal collection in an Orbitrap Exploris 480 (Thermo Fisher) mass spectrometer for LC-MS/MS analysis. The mass spectroscopic raw data of the peptide fragments were analyzed and spliced by searching the fragments containing the desired residues by using the Thermo proteome discoverer software. Besides, the pLink2 software kit was also used for the analysis of the crosslinked peptides.[2]

In vitro photo-crosslinking reactions for *sj*GST-E51*m*-TFMAK and other protein mutants

Purified *sj*GST-E51*m*-TFMAK-137C-139K, *sj*GST-E51*m*-TFMAK-C137A-K139A, *sj*GST-E51*m*-TFMAK-137C-K139A, *sj*GST-E51*m*-TFMAK-C137A-139K were dulited to 2 μM by PBS buffer (pH = 7.4) and irradiated with a 311 nm UV lamp (21.2 mW cm^{-2}) for 5.0 min. Then, the resulting protein mixture were diluted with 5× SDS-PAGE loading buffer and heated to 95 °C for 5-6 min, and the denatured proteins were resolved and analyzed by SDS-PAGE. Control experiments were performed under the same conditions without irradiation and subjected to the same SDS-PAGE resolving procedures.

Photo-crosslinking of *sj*GST-E51*m*-TFMAK in living *E. coli* and quantification

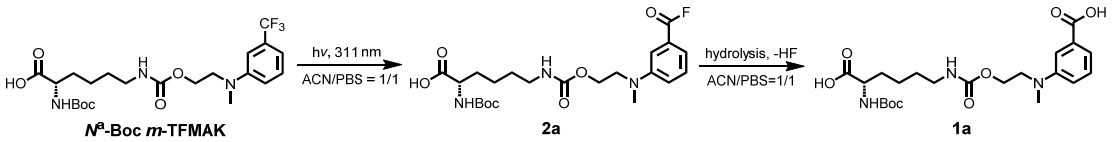
To explore whether photo-crosslinking could occur in living cells, the over-expressing *E. coli* cells containing *sj*GST-E51*m*-TFMAK were harvested and washed for three times by using PBS buffer (pH = 7.4). Cells were then resuspended in PBS buffer, and adjusted to OD600 = 1.0. The living *E. coli* were irradiated with a 311 nm UV lamp for 5.0 min, before being diluted with 2× SDS-PAGE loading buffer and heated to 95 °C for 5-6 min. After resolved by SDS-PAGE, the separated proteins were analyzed by western blot by using the anti-his_6 tag mouse monoclonal antibody (1:500) and anti-GST mouse monoclonal antibody (1:500). Horseradish peroxidase conjugate (HRP) labelled goat anti-mouse IgG (H+L) (1:1000) used as the secondary antibody in the WB to quantify the attached target proteins on the PVDF membrane via catalyzing the oxidation of the ECL substrate (Beyotime Biotechnology) to generate the chemiluminescence. Therefore, the imaging procedures were performed to detect the chemiluminescent bands of the interested protein for further quantification. The crosslinking efficiency was measured by ImageJ.

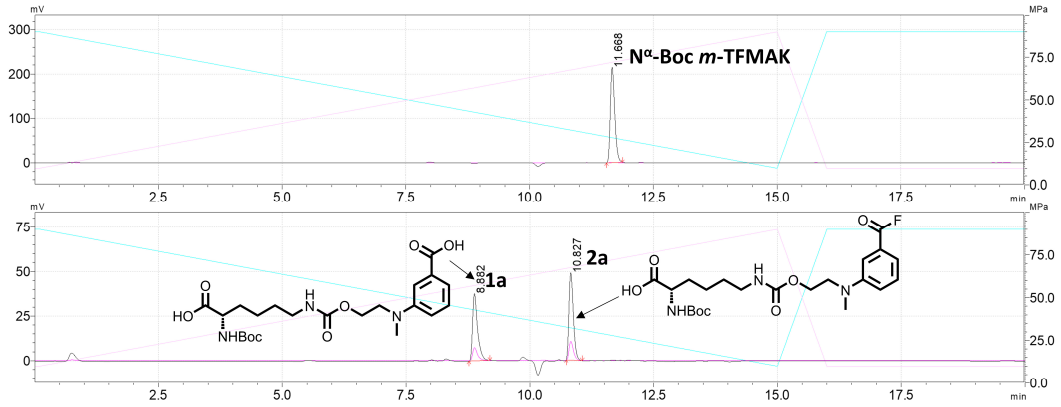
Assessment of phototoxicity of *E. coli* under 311 nm irradiation

The living *E. coli* cells over-expressing *sj*GST-E51*m*-TFMAK were collected, washed with PBS buffer (pH = 7.4) for 3 times and resuspended in PBS buffer to give a cell suspension with OD600 = 1.0. Then, the cells were irradiated with a 311 nm lamp (21.2 mW cm^{-2}) at room

temperature for 5.0 min. The irradiated cells (10 µL) were diluted with LB growth medium, and then plated on LB agar plates were supplemented with tetracycline (34 $\mu g \cdot mL^{-1}$) and ampicillin (100 $\mu g \cdot mL^{-1}$). Afterwards, the smeared LB agar petri dish was placed in incubators at 37 °C for 12 hours to allow the formation of colonies. After imaging under white light to identify the colonies on the agar plates, we were able to manually count the colonies on the plate in order to assess the viability of the cells after the 311 nm irradiation in comparison with unexposed dishes.

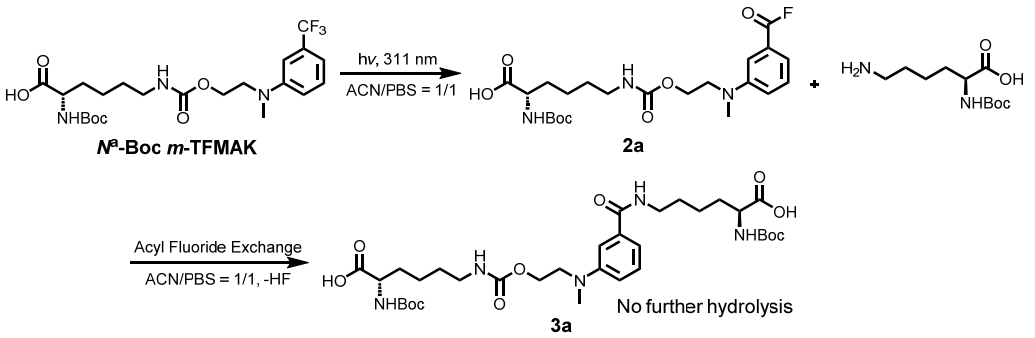
Supplemental Figures

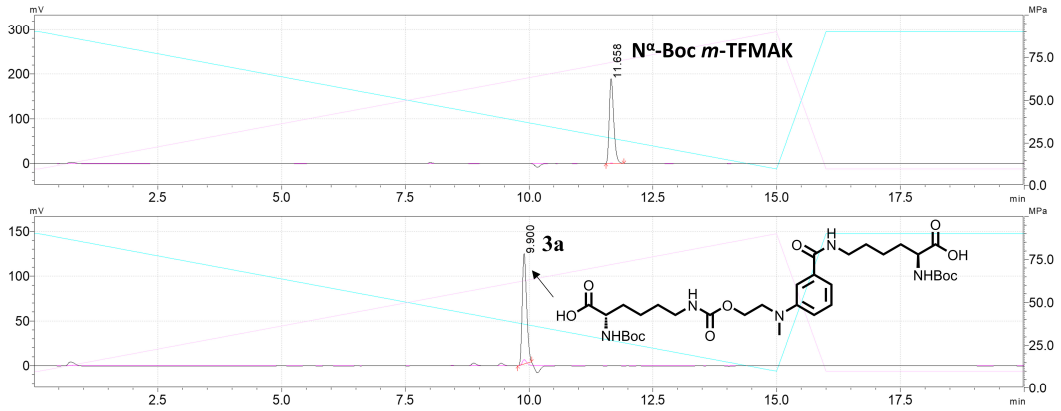




Item	Peak No.	Time (min)	Area (in 254 nm line)
Before irradiation	***N*$^{\alpha}$-Boc-*m*-TFMAK**	11.668	1336260
311 nm for 3 min	**1a**	8.882	263487
	2a	10.827	303075

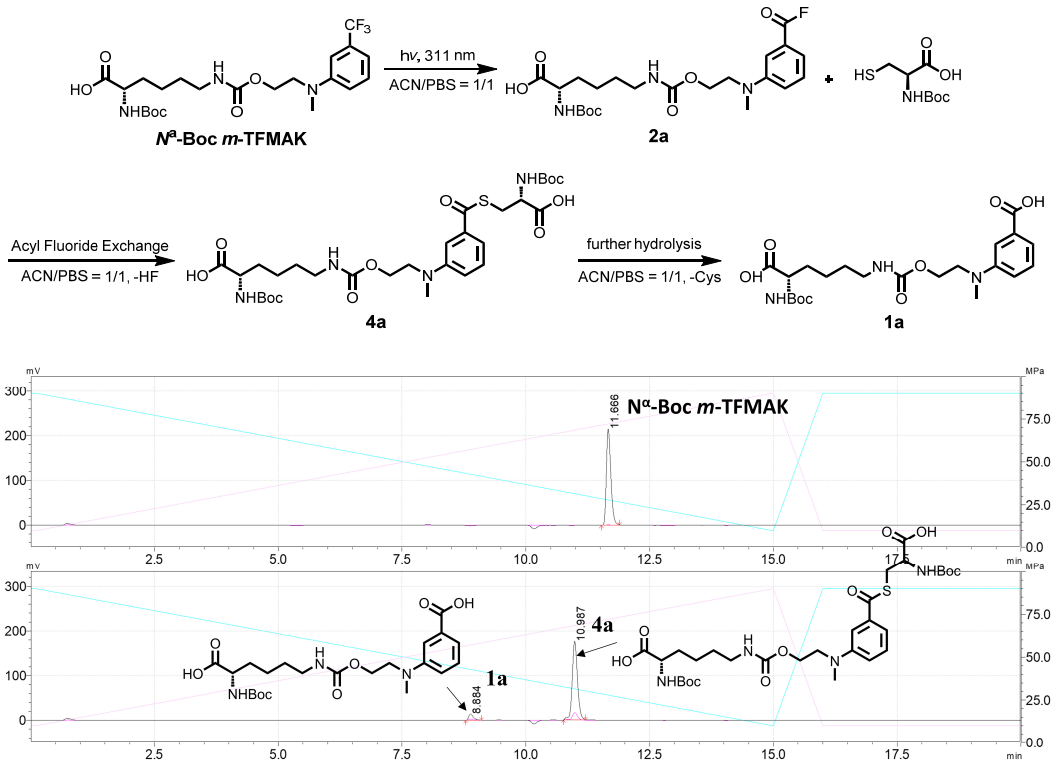
The photo-defluorination reaction of ***N*$^{\alpha}$-Boc-*m*-TFMAK** produces the benzoyl fluoride intermediate (**2a**). Without the presence of any nucleophilic natural amino acid, the benzoyl fluoride intermediate is then hydrolyzed into benzoic acid product (**1a**), which is observed in the HPLC traces at the same time.





Item	Peak No.	Time (min)	Area (in 254 nm line)
Before irradiation	***N*$^{\alpha}$-Boc-*m*-TFMAK**	11.658	1189573
311 nm for 3 min	**3a**	9.900	754382

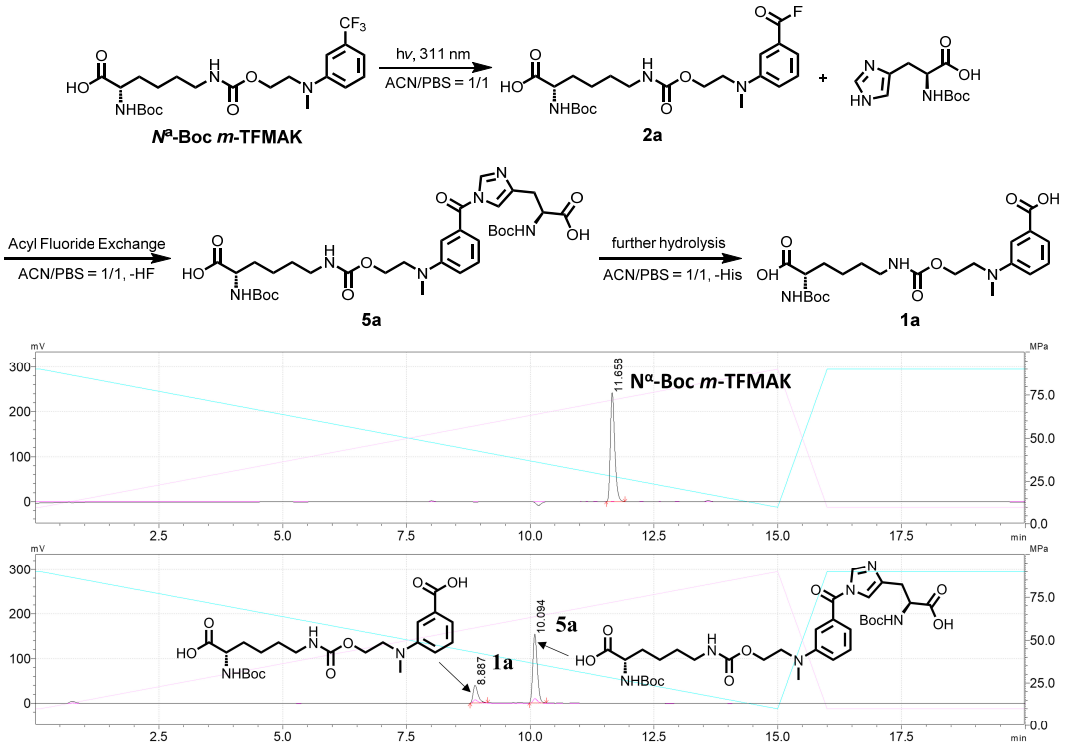
The photo-defluorination reaction of ***N*$^{\alpha}$-Boc-*m*-TFMAK** produces the benzoyl fluoride intermediate (**2a**). In the presence of *N*$^{\alpha}$-Boc-L-lysine, the benzoyl fluoride intermediate is then amidated with the primary amine side chain of lysine to give the ligated product (**3a**). The ligated product (**3a**) is quite stable without any further hydrolysis observed.



Item	Peak No.	Time (min)	Area (in 254 nm line)
Before irradiation	***N*$^{\alpha}$-Boc-*m*-TFMAK**	11.666	1341409
311 nm for 3 min	**1a**	8.884	94273
	4a	10.987	1291893

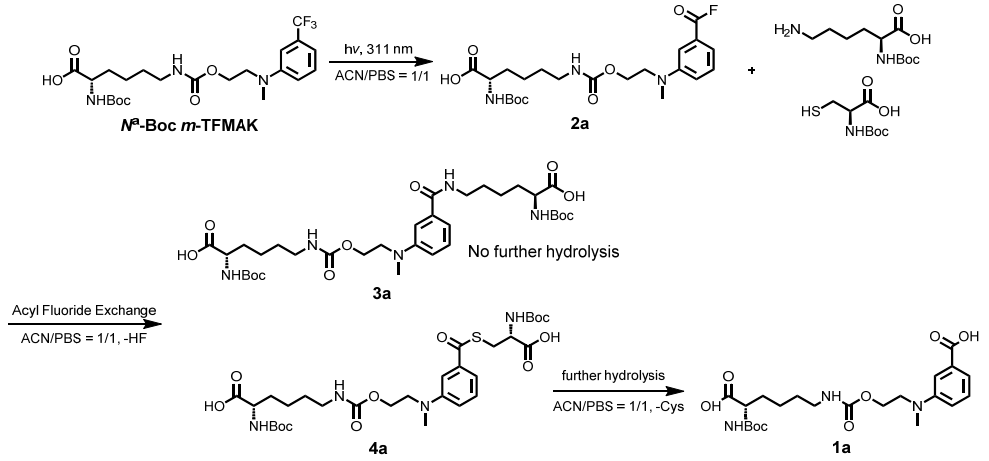
The photo-defluorination reaction of ***N*$^{\alpha}$-Boc-*m*-TFMAK** produces the benzoyl fluoride

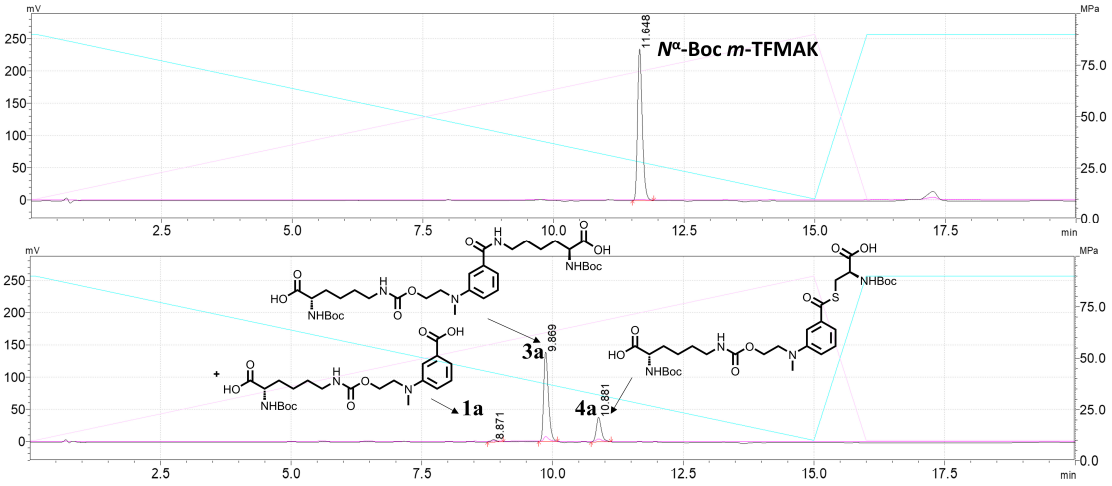
intermediate (**2a**). In the presence of *N*$^{\alpha}$-Boc-L-cysteine, the benzoyl fluoride intermediate is then esterified with the thiol side chain of cysteine to give the ligated product (**4a**). The ligated product (**3a**) is not so stable in water containing solvent, which undergoes further hydrolysis to produce the benzoyl acid (**1a**).



Item	Peak No.	Time (min)	Area (in 254 nm line)
Before irradiation	***N*$^{\alpha}$-Boc-*m*-TFMAK**	11.668	1487153
311 nm for 3 min	**1a**	8.887	269382
	5a	10.094	953336

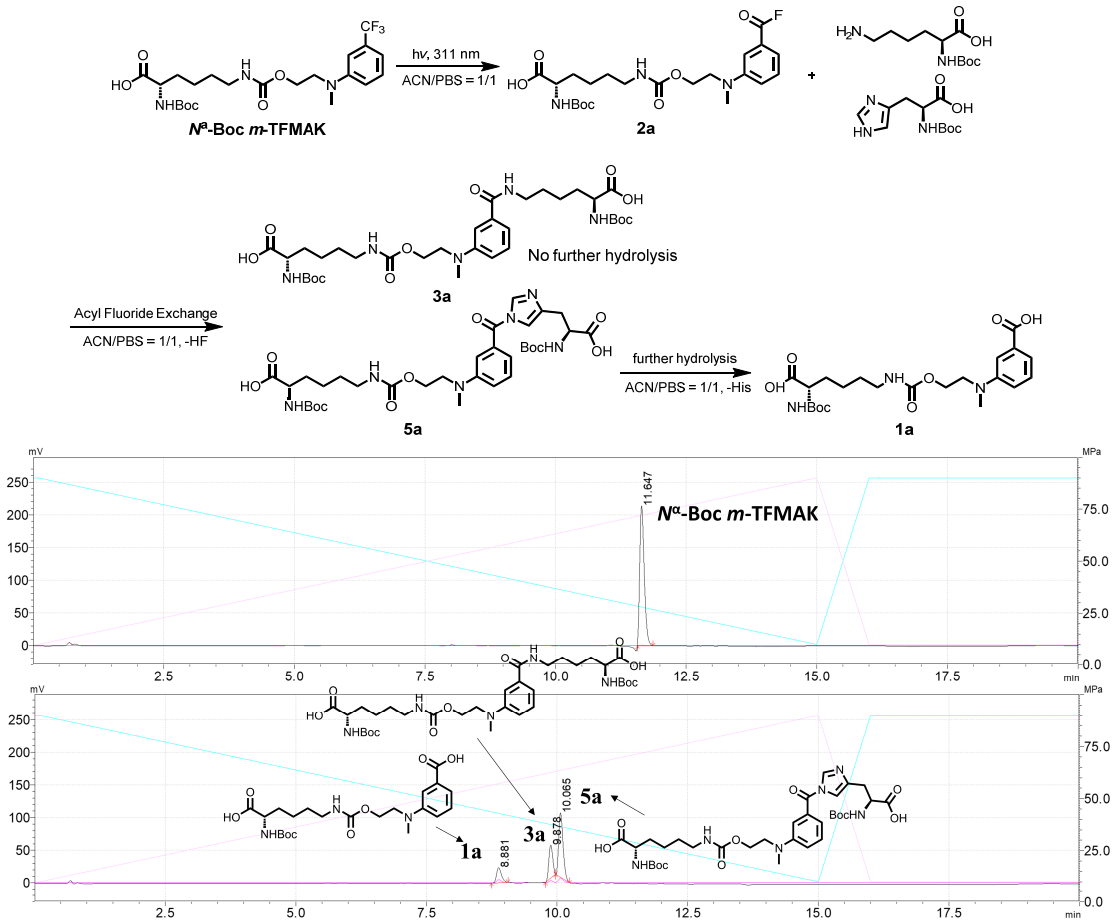
The photo-defluorination reaction of ***N*$^{\alpha}$-Boc-*m*-TFMAK** produces the benzoyl fluoride intermediate (**2a**). In the presence of *N*$^{\alpha}$-Boc-L-histidine, the benzoyl fluoride intermediate is then acylated with the imidazole chain of histidine to give the ligated product (**5a**). The ligated product (**5a**) is not so stable in water containing solvent, which undergoes further hydrolysis to produce the benzoyl acid (**1a**).





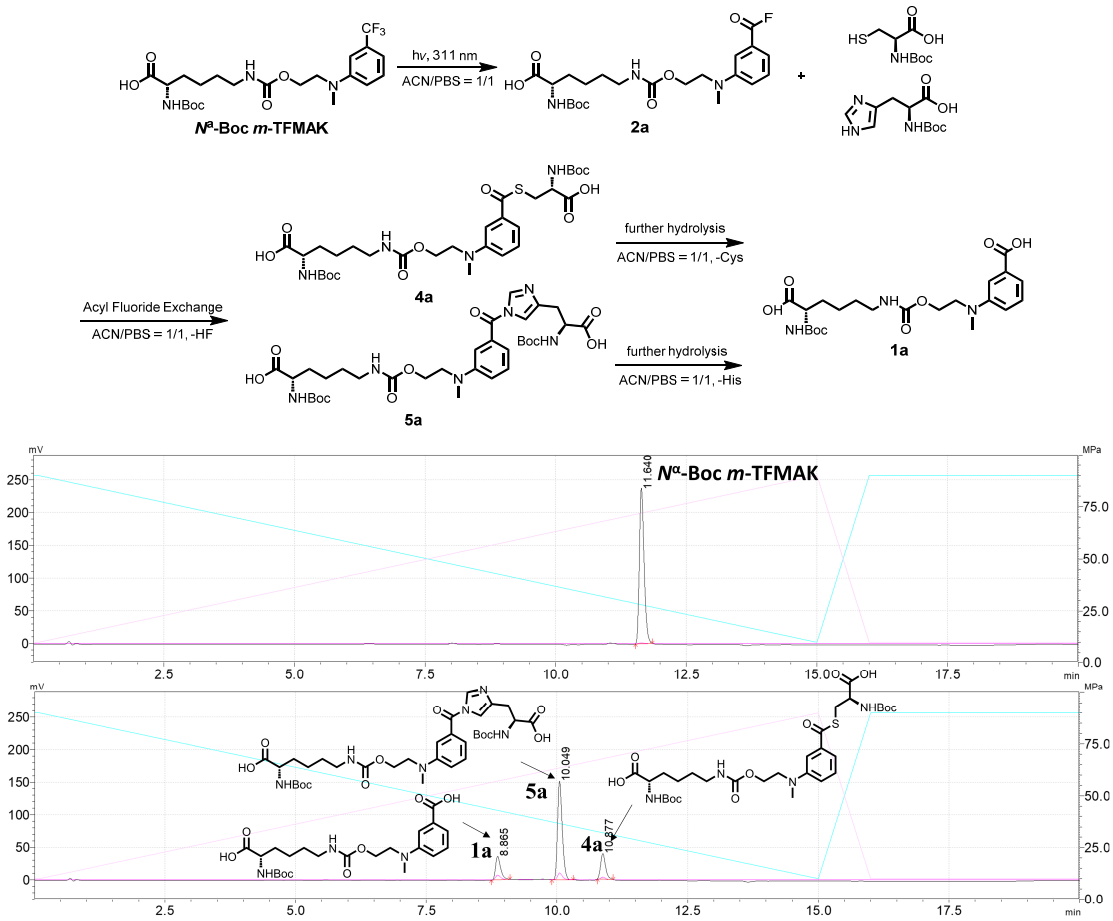
Item	Peak No.	Time (min)	Area (in 254 nm line)
Before irradiation	***N*$^{\alpha}$-Boc-*m*-TFMAK**	11.648	1478886
311 nm for 3 min	**1a**	8.871	24676
	3a	9.869	868795
	4a	10.881	260338

In this competitive reaction, the same concentration of two nucleophilic amino acids (*N*$^{\alpha}$-Boc-L-lysine and *N*$^{\alpha}$-Boc-L-cysteine) exists at the same time, and the photo-generated benzoyl fluoride intermediate (**2a**) reacted with two amino acids to give each ligation product (**3a** and **4a**), respectively. However, due to the hydrolysis of **4a**, the hydrolysate **1a** can be observed.



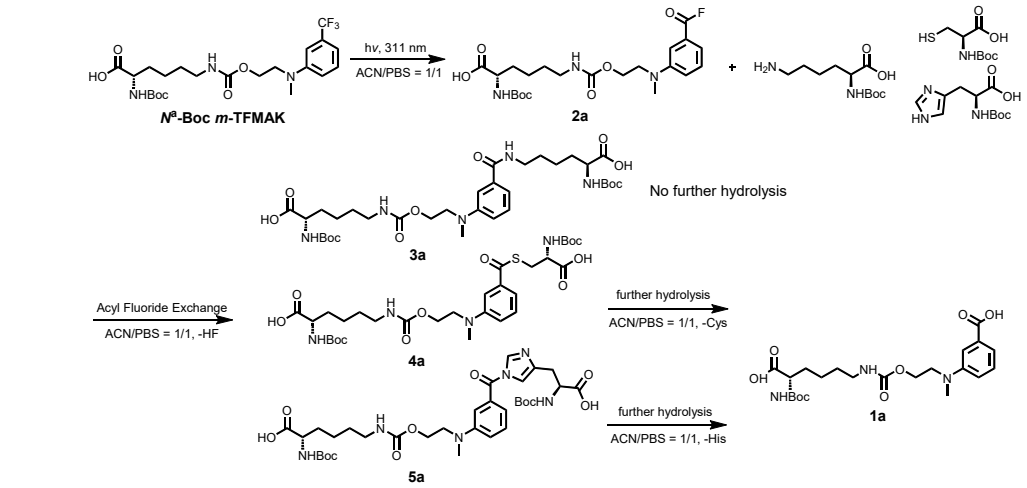
Item	Peak No.	Time (min)	Area (in 254 nm line)
Before irradiation	**N^{α}-Boc-*m*-TFMAK**	11.647	1293291
311 nm for 3 min	**1a**	8.881	153934
	3a	9.878	260987
	5a	10.065	571557

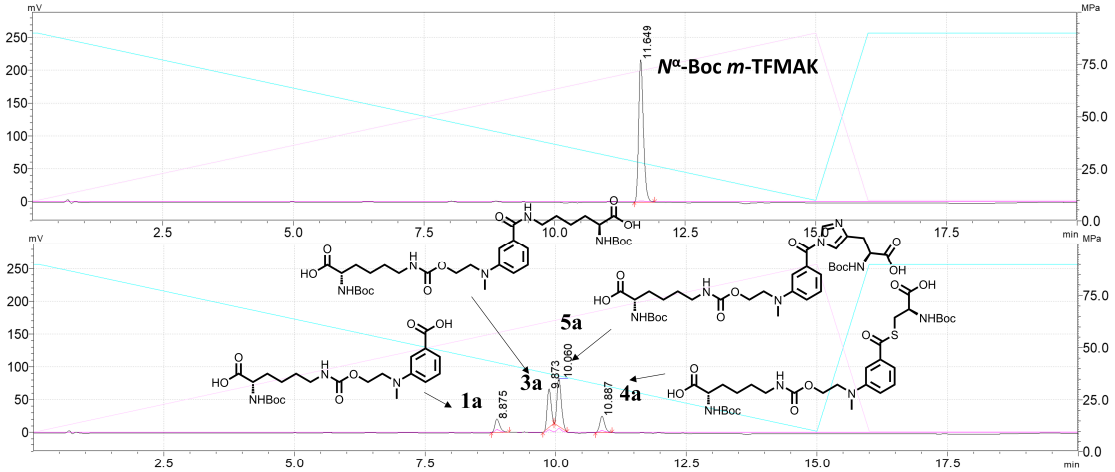
In this competitive reaction, the same concentration of two nucleophilic amino acids (N^{α}-Boc-L-lysine and N^{α}-Boc-L-histidine) exists at the same time, and the photo-generated benzoyl fluoride intermediate (**2a**) reacted with two amino acids to give each ligation product (**3a** and **5a**), respectively. However, due to the hydrolysis of **5a**, the hydrolysate **1a** can be observed.



Item	Peak No.	Time (min)	Area (in 254 nm line)
Before irradiation	**N^{α}-Boc-*m*-TFMAK**	11.640	1492213
311 nm for 3 min	**1a**	8.865	242104
	5a	10.049	950161
	4a	10.877	254020

In this competitive reaction, the same concentration of two nucleophilic amino acids (N^{α}-Boc-L-cysteine and N^{α}-Boc-L-histidine) exists at the same time, and the photo-generated benzoyl fluoride intermediate (**2a**) reacted with two amino acids to give each ligation product (**4a** and **5a**), respectively. However, due to the hydrolysis of both **4a** and **5a**, the hydrolysate **1a** can be observed.





Item	Peak No.	Time (min)	Area (in 254 nm line)
Before irradiation	***N*$^{\alpha}$-Boc-*m*-TFMAK**	11.649	1373582
311 nm for 3 min	**1a**	8.875	138153
	3a	9.873	305185
	5a	10.060	419819
	4a	10.887	167131

In this competitive reaction, the same concentration of three nucleophilic amino acids (*N*$^{\alpha}$-Boc-L-lysine, *N*$^{\alpha}$-Boc-L-cysteine and *N*$^{\alpha}$-Boc-L-histidine) exists at the same time, and the photo-generated benzoyl fluoride intermediate (**2a**) reacted with three amino acids to give each ligation product (**3a**, **4a** and **5a**: 41:11:31), respectively. However, due to the hydrolysis of both **4a** and **5a**, the hydrolysate **1a** (17%) can be observed.

Figure S1. *N*$^{\alpha}$-Boc protected *m*-TFMAK reacted with *N*$^{\alpha}$-Boc protected L-lysine, *N*$^{\alpha}$-Boc protected L-Cystine and *N*$^{\alpha}$-Boc protected L-Histidine. *N*$^{\alpha}$-Boc protected *m*-TFMAK was diluted in ACN/PBS = 1/1 to 100 μM, and *N*$^{\alpha}$-Boc protected L-lysine, *N*$^{\alpha}$-Boc protected L-Cystine and *N*$^{\alpha}$-Boc protected L-Histidine were diluted in ACN/PBS = 1/1 to 500 μM, respectively. *N*$^{\alpha}$-Boc protected *m*-TFMAK (100 μM, 1.0 eq) and *N*$^{\alpha}$-Boc protected natural amino acid (500 μM, 5.0 eq) were mixed in ACN/PBS = 1/1 with 311 nm activation for 3 min, then the resulting mixtures were analyzed by HPLC-MS to quantify the conversion and yield.

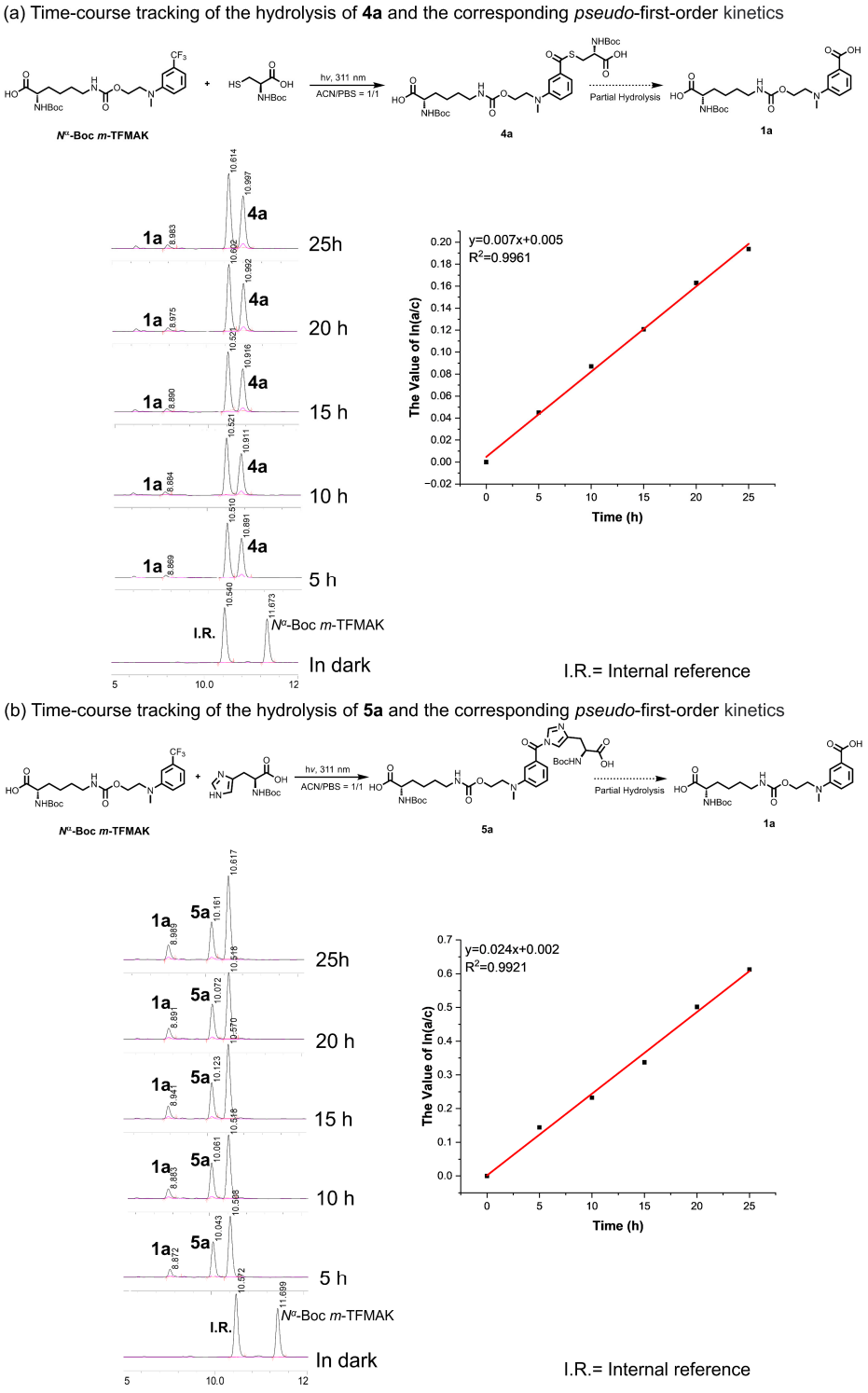


Figure S2. The hydrolysis of conjugation product of N^{α}-Boc protected L-Cystine (**4a**) or N^{α}-Boc protected L-Histidine (**5a**) in ACN/PBS (v/v = 1:1, pH = 7.4) can be regarded as a pseudo-first-order reaction because the concentration of water is almost constant in the buffer and in great excess. (a) HPLC traces of **4a** after incubation for a specific time and a linear fitting curve for time-course tracking the hydrolysis, and the half-life of hydrolysis for the thioester was 99 hours. (b) HPLC traces of **5a** after incubation for a specific time and a linear fitting curve for time-course tracking the hydrolysis, and the half-life of hydrolysis for the benzoyl imidazole was 28 hours.

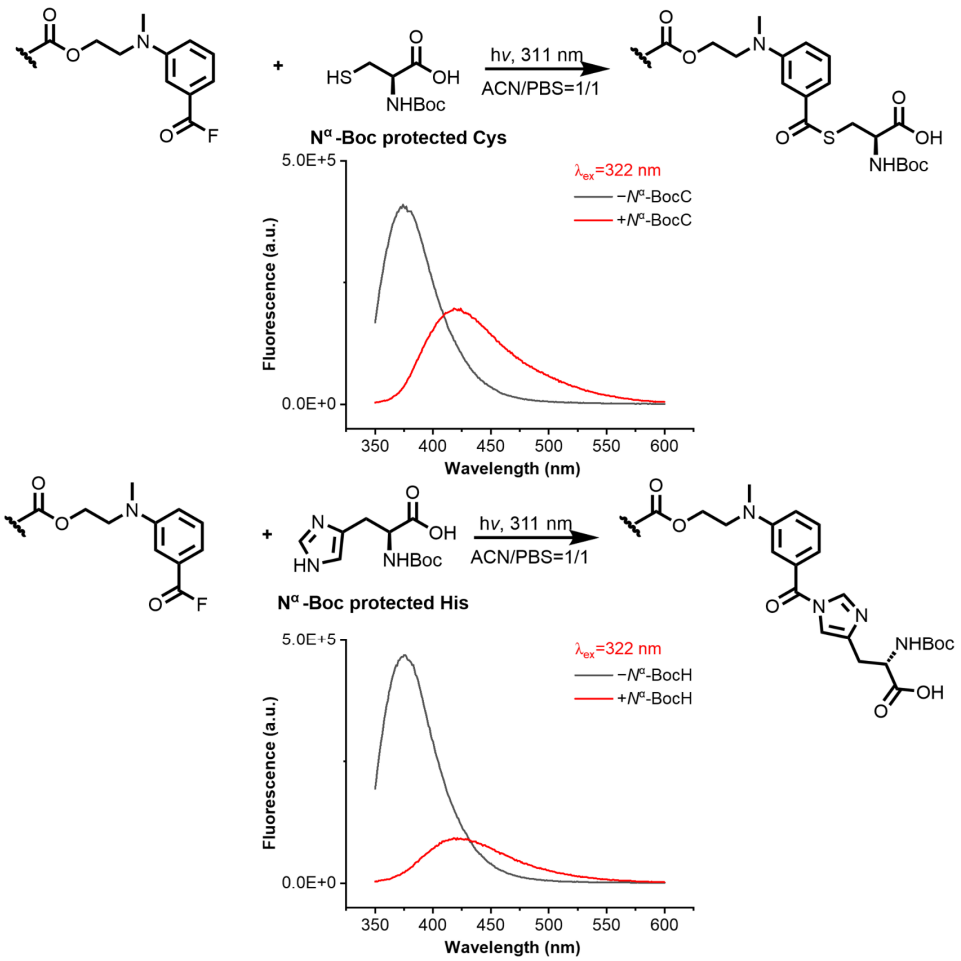


Figure S3. Spectral characters in the fluorescence emission intensity for *in situ* monitoring the corresponding acyl fluoride (100 μM) reacting with *N*$^{\alpha}$-Boc protected L-cysteine or *N*$^{\alpha}$-Boc protected L-histidine (500 μM) for 60 s. To ensure *N*$^{\alpha}$-Boc protected *m*-TFMAK (100 μM) complete conversion into corresponding acyl fluoride, the solution was irradiated for 3.0 min under 311 nm lamp (5.9 mW cm^{-2}, single wavelength output after an optical filter) before mixing with *N*$^{\alpha}$-Boc protected L-cysteine or *N*$^{\alpha}$-Boc protected L-histidine solution.

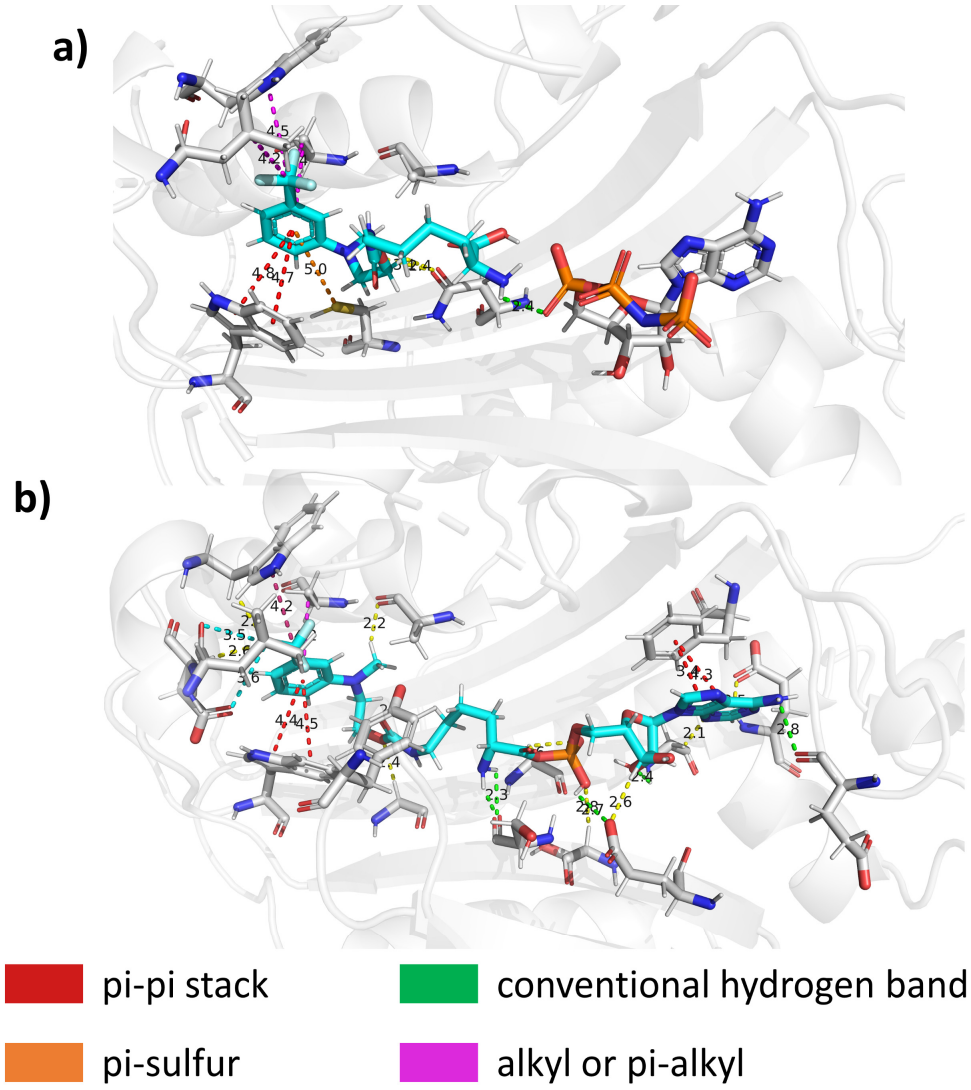


Figure S4. Molecular docking results of *m*-TFMAK (a) and *m*-TFMAK-ANP complex (b) to the binding sites of *Mm*PylRS-Y306A. *m*-TFMAK or *m*-TFMAK-ANP complex was shown in the stick model, and *Mm*PylRS-Y306A was shown in cartoon model. (a) 3D display of the docking posture of *m*-TFMAK in *Mm*PylRS-Y306A with an optimal binding energy (−50.4 kcal·mol^{-1}). Pi-pi stack interaction (red-colour): W417, pi-sulfur (orange-colour): C348, hydrogen-bonding (green-colour): ANP, alkyl or pi-alkyl interactions (purple-colour): A306, L407, w411. (b) 3D display of the docking posture of *m*-TFMAK-ANP complex in *Mm*PylRS-Y306A with an optimal binding energy (−78.7 kcal·mol^{-1}). Pi-pi stack interaction (red-colour): W417, F342, hydrogen-bonding (green-colour): E337, E396, S399, Y384, G423, alkyl or pi-alkyl interactions (purple-colour): A306, L339, I413.

Table S1. The docking results with scoring and interaction energies of *m*-TFMAK to *Mm*PylRS-Y306A.

Pose number	CDOCKER energy (kcal·mol^{-1})	CDOCKER interaction energy (kcal·mol^{-1})	remarks
1	**−50.2**	**−50.4**	**The most favorable pose and docking site.**
2	−44.6	−47.7	
3	−44.0	−49.1	
4	−43.0	−46.9	
5	−42.7	−45.6	
6	−42.2	−46.0	
7	−41.0	−46.6	

Table S2. The docking results with scoring and interaction energies of *m*-TFMAK-ANP to *Mm*PylRS-Y306A.

Pose number	CDOCKER energy (kcal·mol^{-1})	CDOCKER interaction energy (kcal·mol^{-1})	remarks
1	**−26.0**	**−78.7**	**The most favorable pose and docking site.**
2	−25.3	−76.9	
3	−24.9	−71.3	

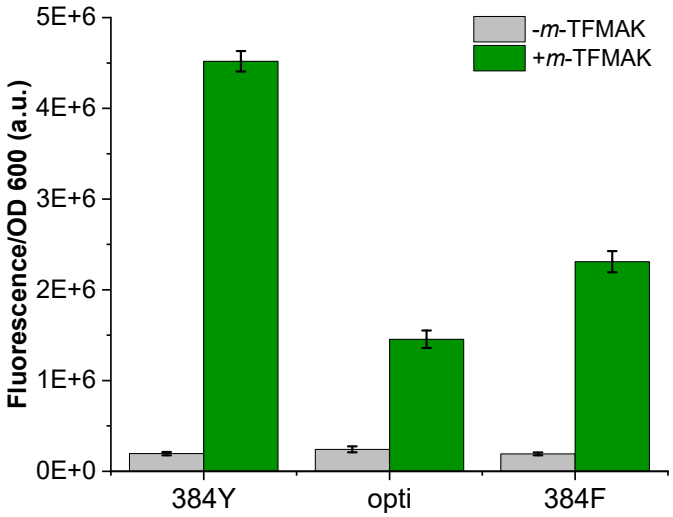


Figure S5. Evaluation of amber codon suppression efficiency of *m*-TFMAK (2.0 mM) in *E. coli*. via fluorescence/OD 600 screening. Error bars represent three independent expressions. 384F mutant means Y306A/Y384; opti-mutant means R19S/H29R/T122S/Y306A; 384F mutant means Y306A/Y384F.

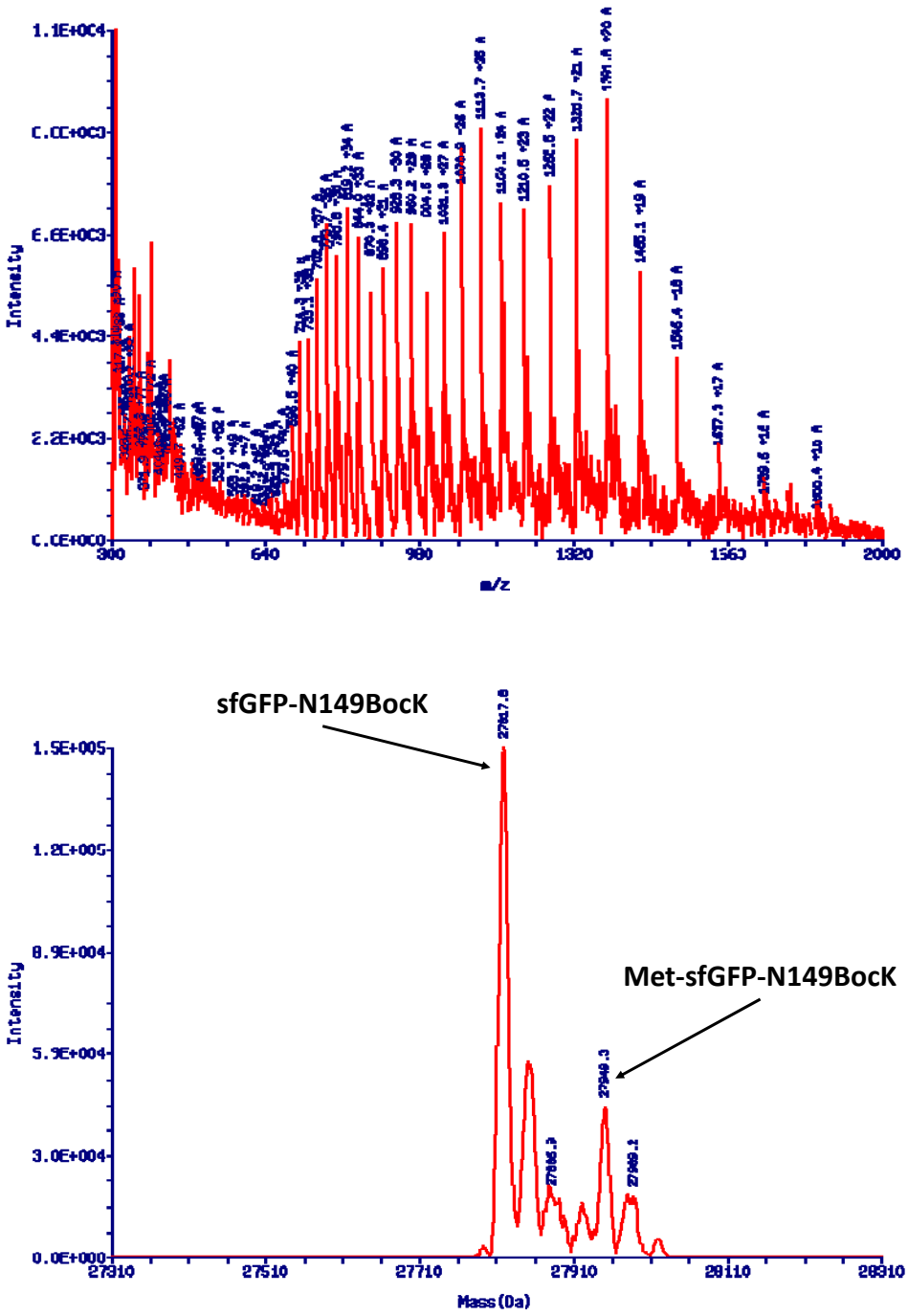


Figure S6. Deconvoluted LC-MS analysis of sfGFP-N149BocK with the charge ladder and the deconvoluted mass spectrum. Found [M-Met+H$^+$]: 27817.8 Da. The peak of Met-sfGFP-N149BocK represents the intact sfGFP protein without losing the first methionine (Met) in its protein sequence during the native PTM in living cells.

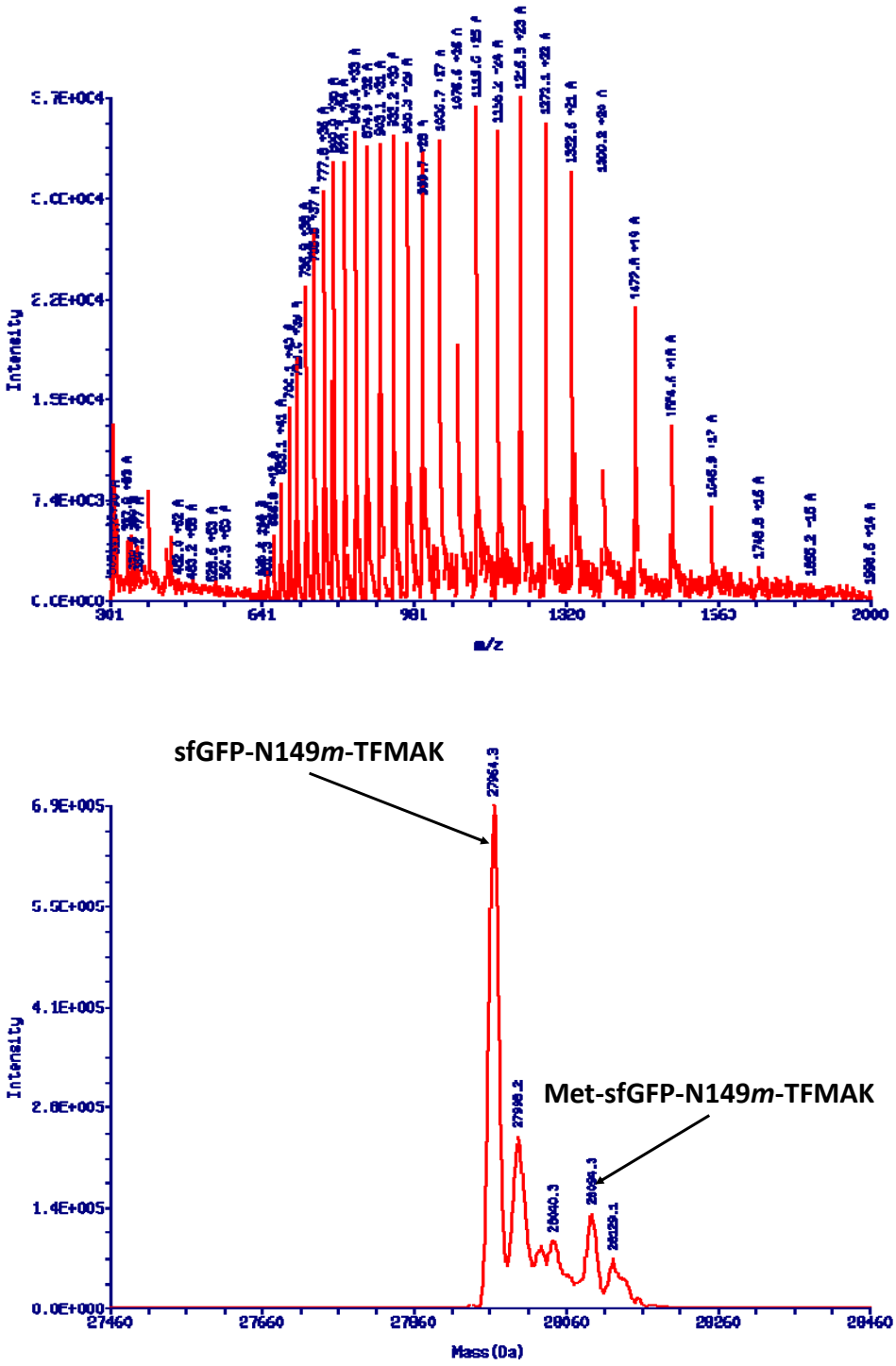


Figure S7. Deconvoluted LC-MS analysis of sfGFP-N149*m*-TFMAK with the charge ladder and the deconvoluted mass spectrum. Found [M-Met+H$^+$]: 27964.3 Da. Calcd. mass 27962.8 Da.

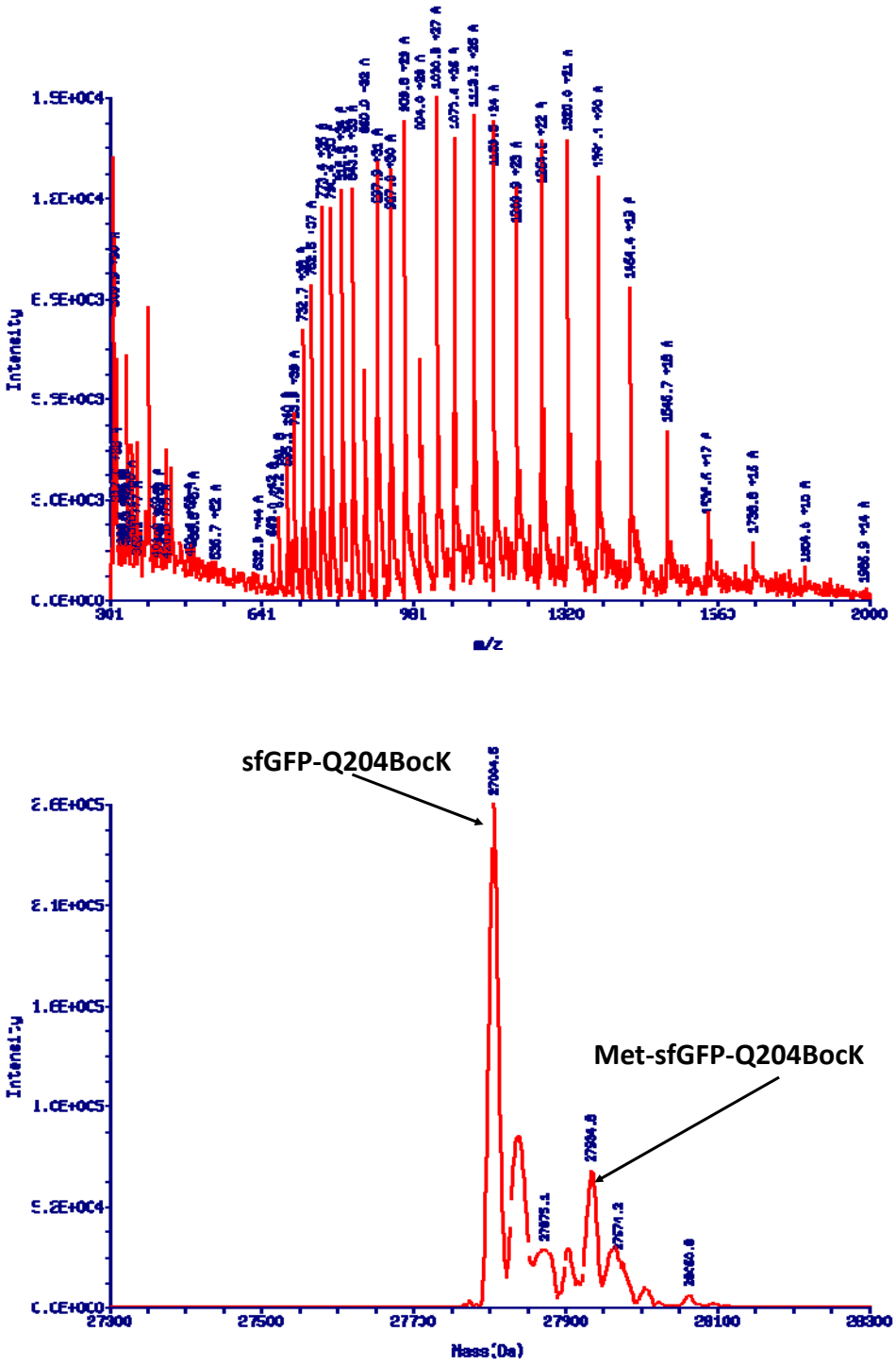


Figure S8. Deconvoluted LC-MS analysis of sfGFP-Q204BocK with the charge ladder and the deconvoluted mass spectrum. Found [M-Met+H$^+$]: 27804.5 Da.

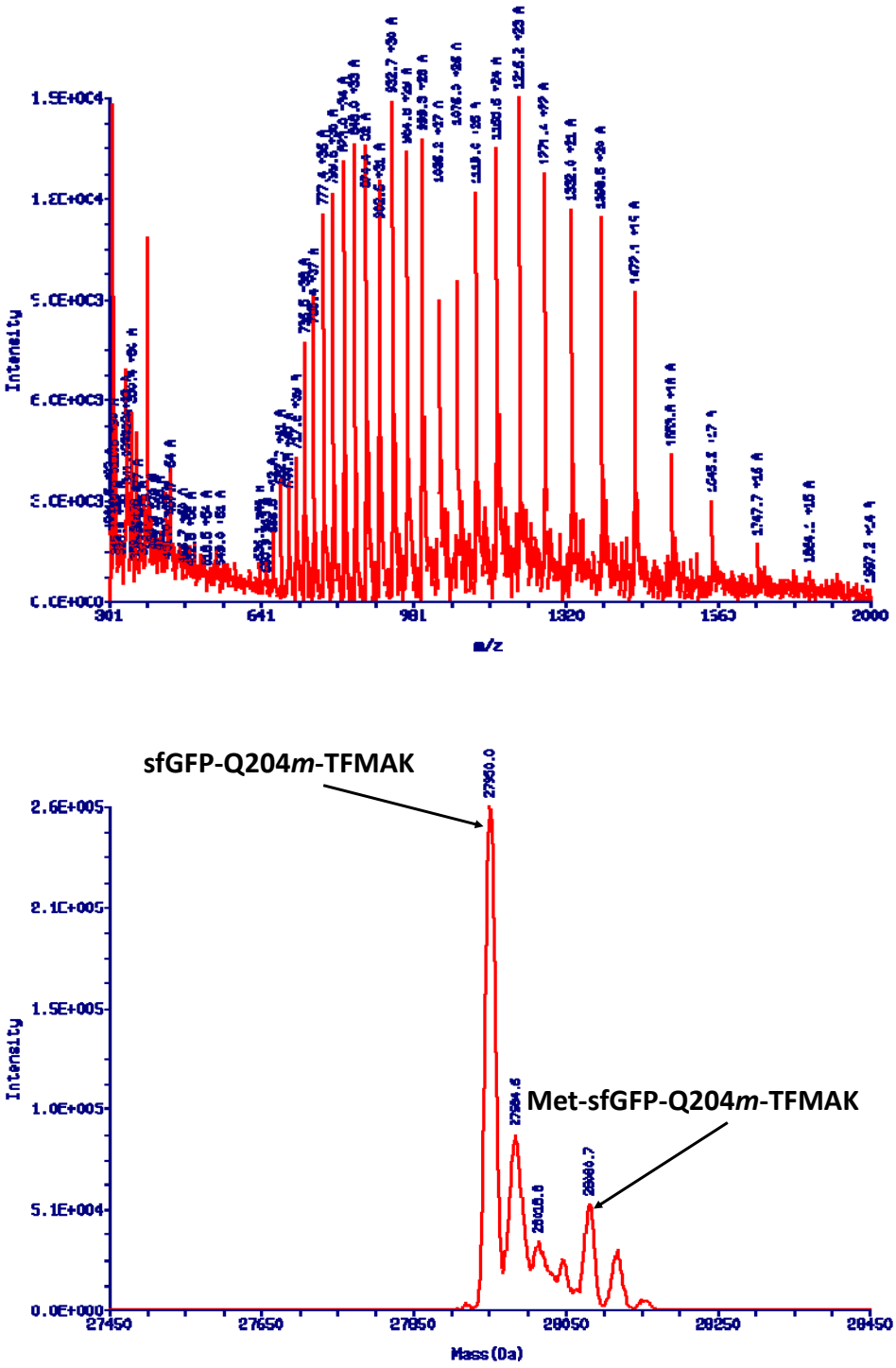


Figure S9. Deconvoluted LC-MS analysis of sfGFP-Q204*m*-TFMAK with the charge ladder and the deconvoluted mass spectrum. Found [M-Met+H$^+$]: 27950.5 Da. Calcd. mass 27949.5 Da.

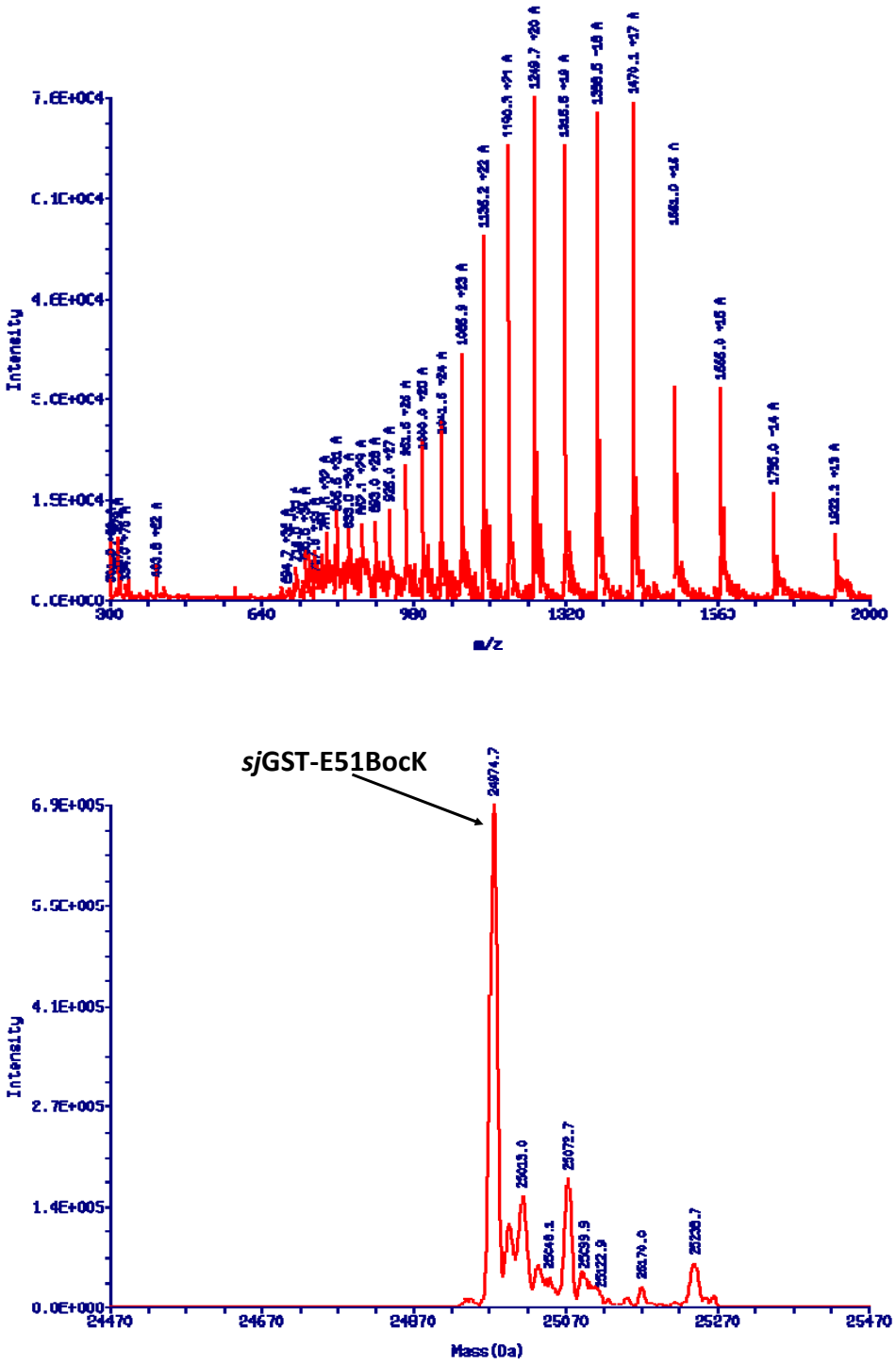


Figure S10. Deconvoluted LC-MS analysis of *sj*GST-E51BocK with the charge ladder and the deconvoluted mass spectrum. Found [M-Met+H$^+$]: 24974.4 Da.

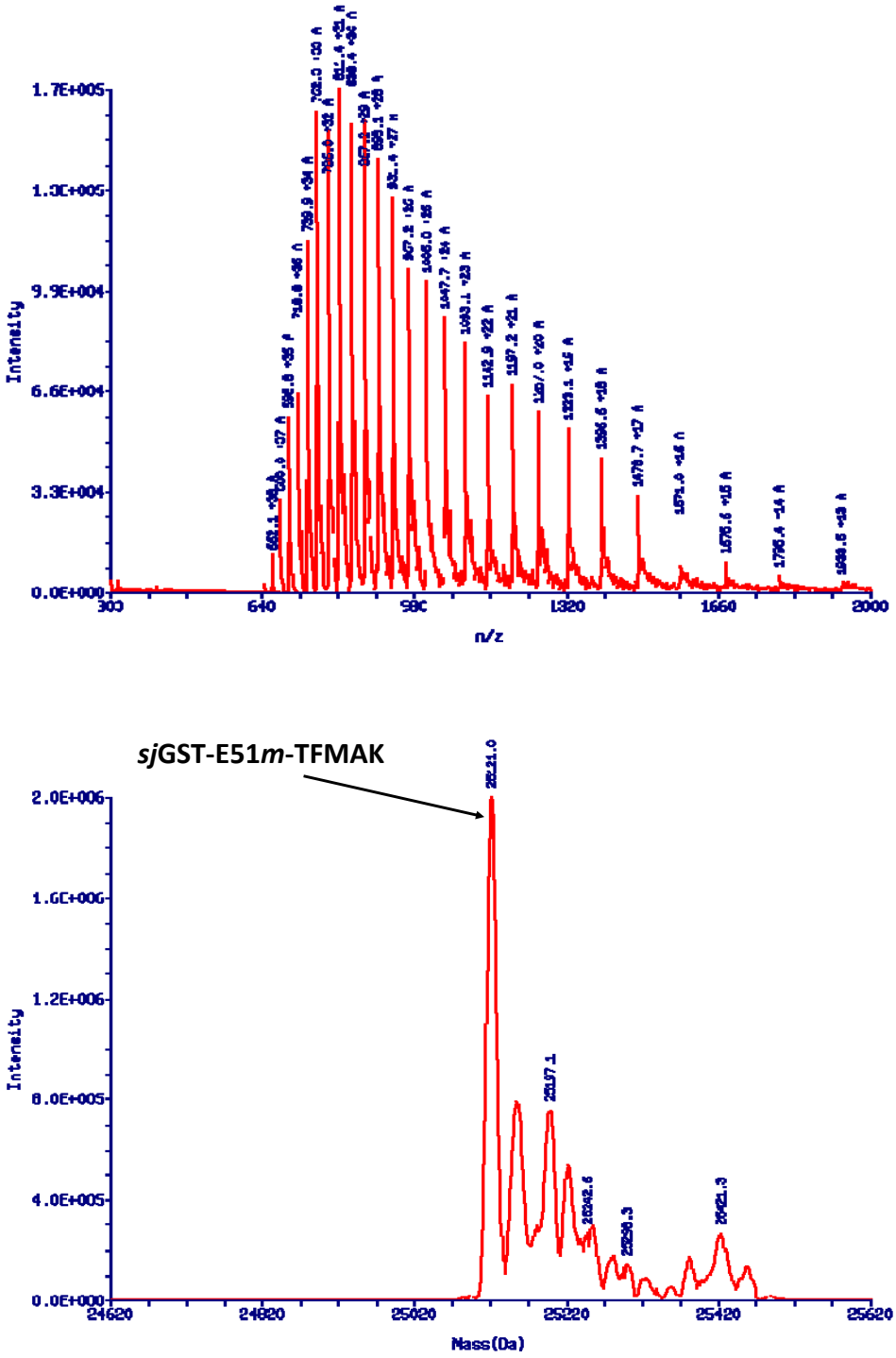


Figure S11. Deconvoluted LC-MS analysis of *sj*GST-E51*m*-TFMAK with the charge ladder and the deconvoluted mass spectrum. Found [M-Met+H$^+$]: 25121.0 Da. Calcd. mass 25119.4 Da.

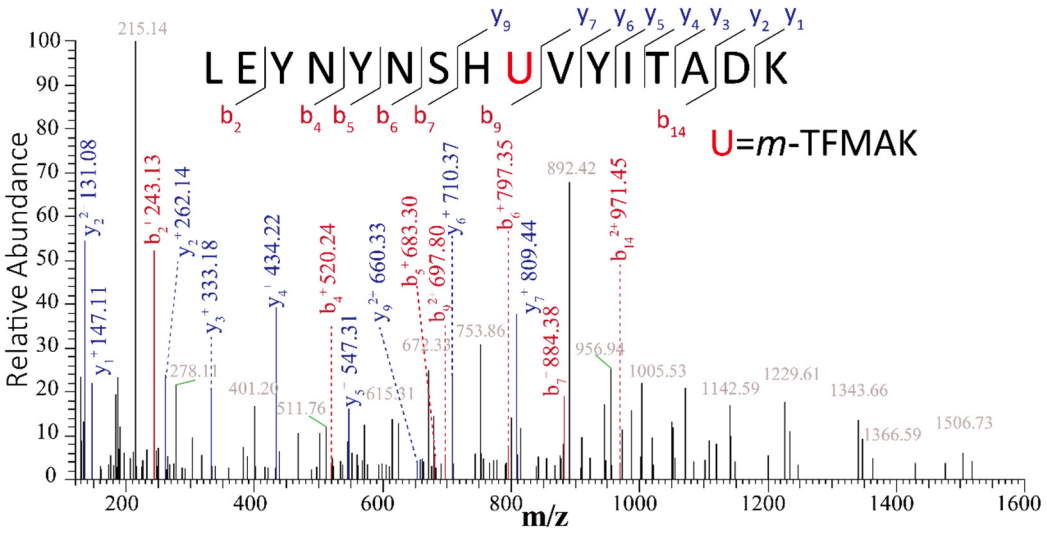


Figure S12. The LC-MS/MS spectrum for spliced sfGFP-N149*m*-TFMAK peptide fragments with sequence displayed. A series of fragmented b/y ions clearly confirmed the *m*-TFMAK incorporation in sfGFP at position 149. U = *m*-TFMAK.

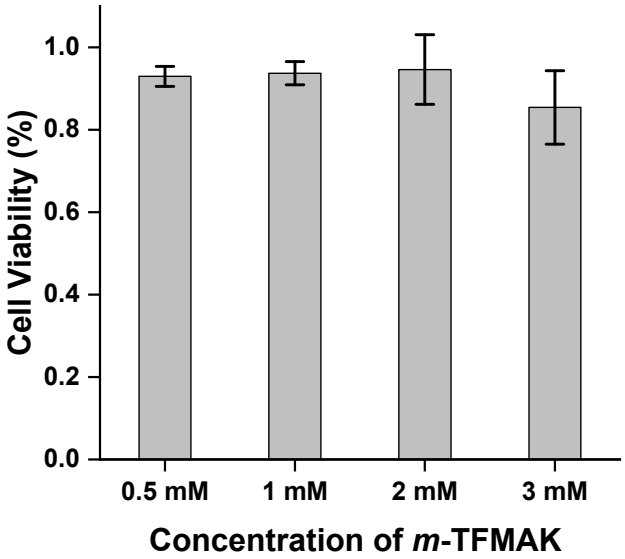


Figure S13. HEK293T cell viability under the treatment of various concentration of *m*-TFMAK for 30 h, which was determined by CCK-8 assay (purchased from Beyotime Biotechnology). Error bar represented three independent experiments.

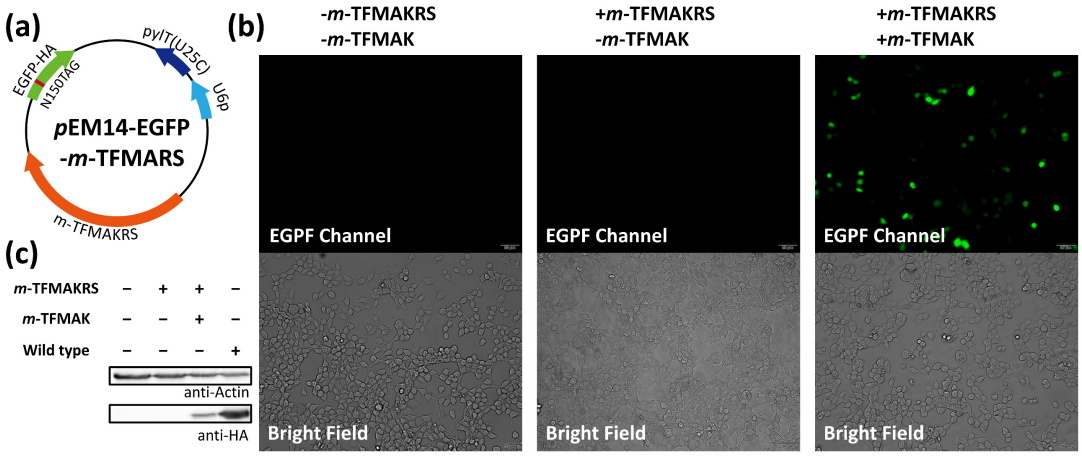


Figure S14. The GCE incorporation system of *m*-TFMAK applied in living HEK293T cells. (a) The plasmid used for transfection into the HEK293T cells. (b) Fluorescence imaging after transfection for 30 h. Conditions: 2.0 mM *m*-TFMAK. (c) Western blot analysis to verify the full-length EGFP expression by anti-HA-tag Mab. The PVDF membrane was incubated with anti-HA-tag mouse monoclonal antibody (1:500) or anti-β-actin mouse monoclonal antibody (1:1000), followed by incubating with HRP-labelled Goat anti-Mouse IgG (H+L) as the secondary antibody (1:1000), which was further incubated with ECL substrate for the chemiluminescence imaging.

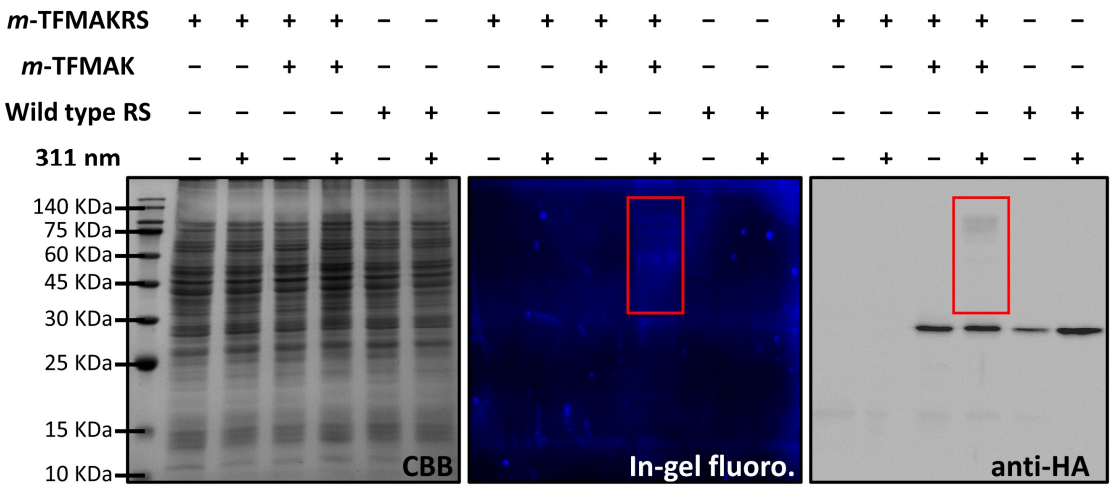


Figure S15. The photo-DAFEx reaction in living HEK293T cells over-expression EGFP-N150*m*-TFMAK. After the transfection and expression, SDS-PAGE, in-gel fluorescence imaging and western blot analysis indicated the crosslinked bands with endogenous protein appeared only when expressing intact EGFP-N150*m*-TFMAK followed by irradiation of 311 nm for 5 min.

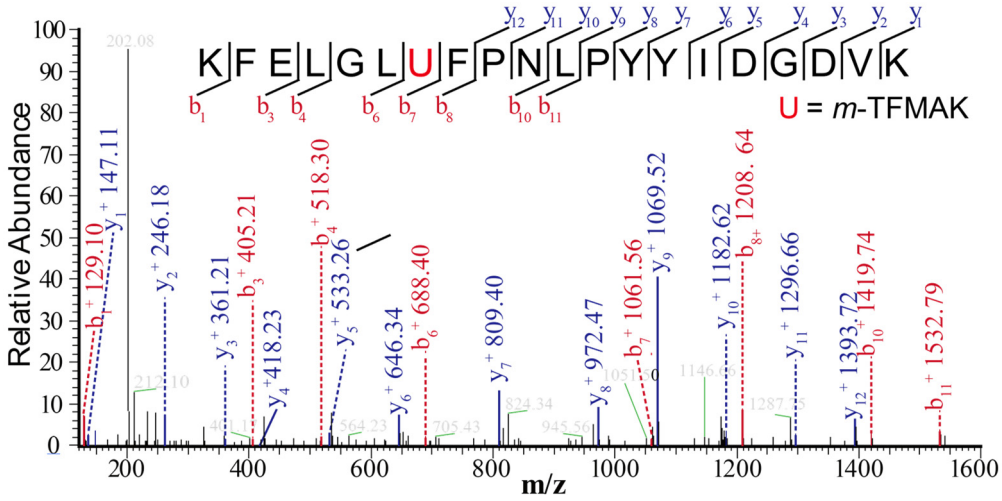


Figure S16. The LC-MS/MS spectrum for the spliced peptide fragments of *sj*GST-E51*m*-TFMAK with the sequence displayed. A series of fragmented ions clearly confirmed the successful incorporation of *m*-TFMAK at position 51 of *sj*GST. U = *m*-TFMAK.

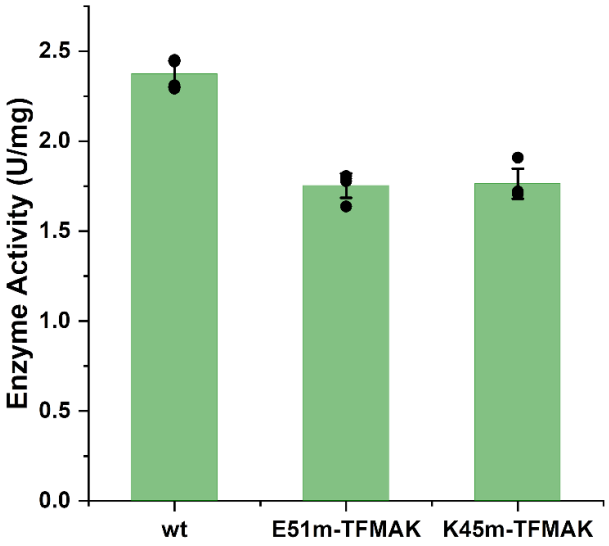


Figure S17. The enzymatic activity of two *sj*GST mutant analyzed by the conjugation of L-glutathione to CDNB via commercial CDNB-based GST assay kit (Sorlabio®). The results showed that the incorporation of *m*-TFMAK affected the enzyme activity by decreasing 26.2% for E51*m*-TFMAK (replacing a negatively charged E) or 25.7% for K45*m*-TFMAK (replacing a positively charged K). Therefore, we have to admit that the side chain of *m*-TFMAK was relatively rigid & non-polar compared to natural amino acids, and there might be a steric hindrance that affected the natural interactions if around the interacting interface. However, the effect of incorporating *m*-TFMAK as a single mutation on native protein interaction was not a substantial impact.

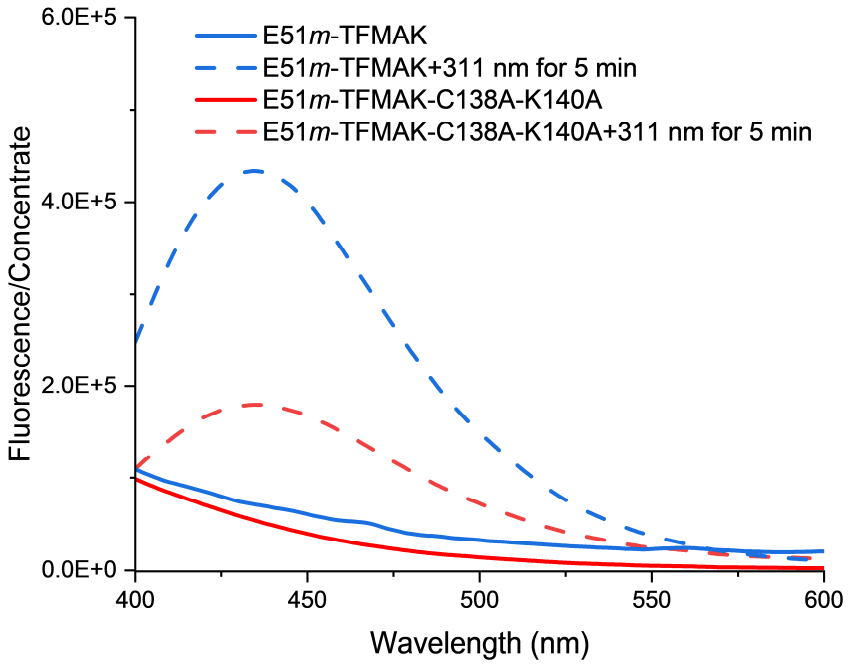


Figure S18. Fluorescence intensity differences of crosslinked sjGST between monomeric and dimeric (mixed with dimer) from. The fluorescence intensity was normalized by total protein concentration (BCA assay, Beyotime), the emission wavelength was the same at 436 nm for either intra-crosslinking or inter-crosslinking sjGST. As a result, the inter-crosslinked (dimeric) fluorescence intensity was stronger than intra-crosslinked (monomeric) one.

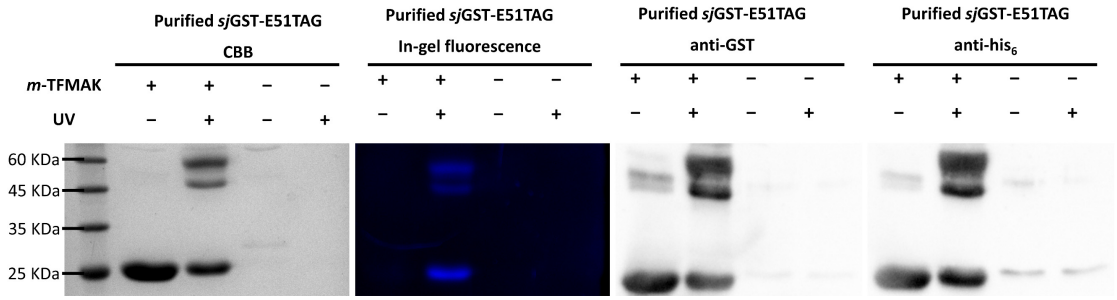


Figure S19. SDS-PAGE and western blot (PVDF membrane) analysis of purified *sj*GST-E51*m*-TFMAK by anti-his_6 or anti-GST antibody to confirm the formation of dimeric *sj*GST. The PVDF membrane was incubated with anti-GST-tag mouse monoclonal antibody (1:500) or anti-his-tag mouse monoclonal antibody (1:500) respectively, followed by incubating with HRP-labelled Goat anti-Mouse IgG (H+L) as the secondary antibody (1:1000), which was further incubated with the ECL substrate for the chemiluminescence imaging.

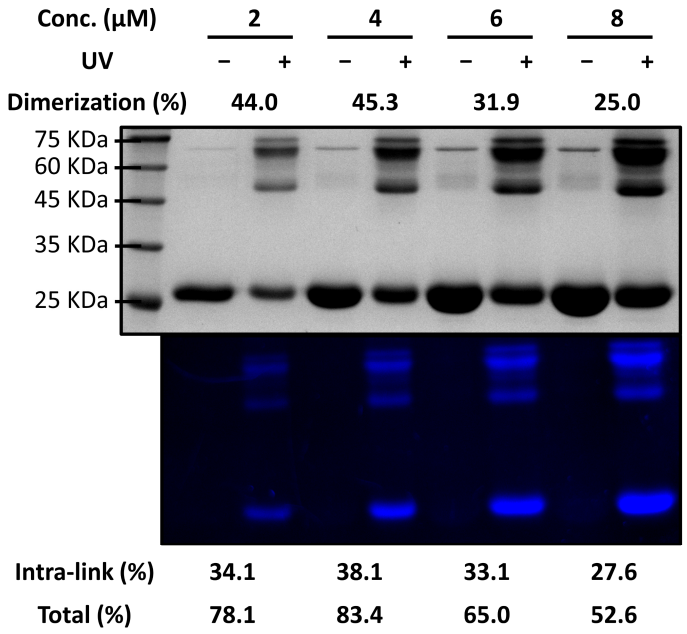


Figure S20. Photo-crosslinking of various concentrations of purified *sj*GST-E51*m*-TFMAK analysis by SDS-PAGE and in-gel fluorescence image.

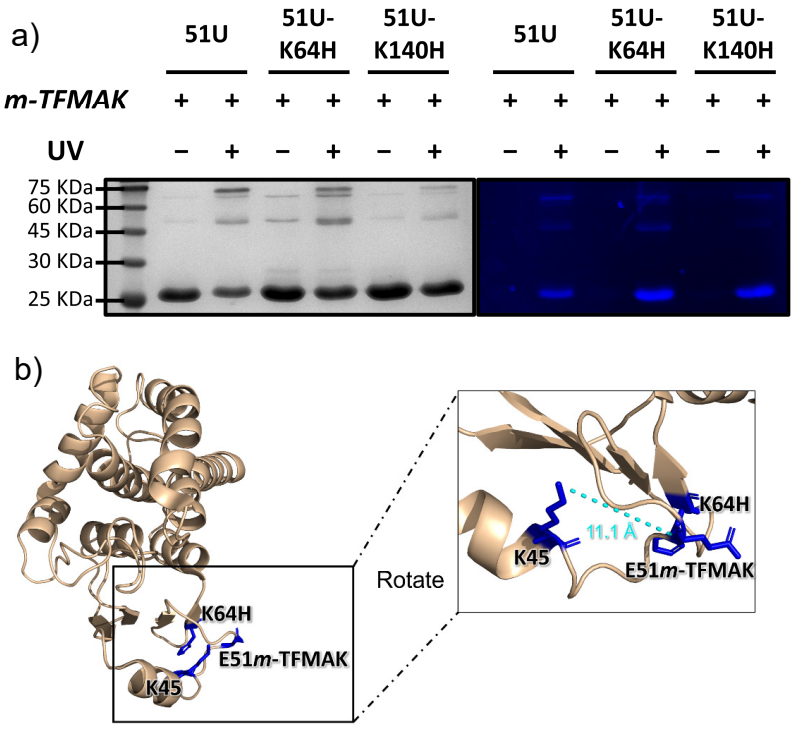
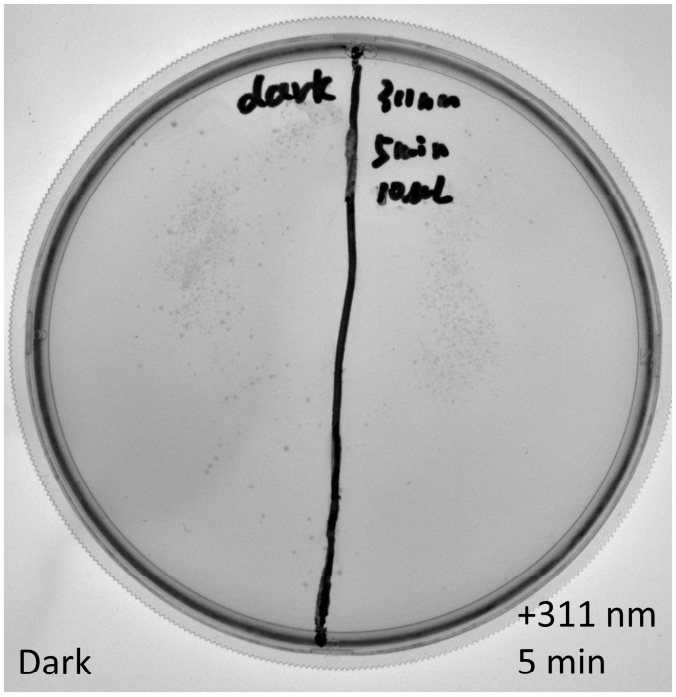


Figure S21. (a) SDS-PAGE and in-gel fluorescence analysis of the photo-DAFEx crosslinking for the purified *sj*GST-E51*m*-TFMAK, *sj*GST-E51*m*-TFMAK-K64H or *sj*GST-E51*m*-TFMAK-K140H. The results showed that a stronger primaeval fluorescence band appeared for the sjGST-E51*m*-TFMAK-K64H mutant, which indicated that the fast-generated *N*-acyl imidazole (K64AcH) might undergo a trans-acylation with other proximal Lys (on longer at the K64 residue) intramolecularly. (b) According to the crystal structure of sjGST, the other proximal Lys site at this time should be at position K45 with a distance of 11.1 Å from C^{α}_{E51} to $N^{\varepsilon\prime}_{K45}$, which is within the maximum length of the crosslinking

radius for a lysine-targeted linkage. The dimer band of sjGST-E51*m*-TFMAK-K140H showed a much weaker fluorescence band with weaker CBB bands. This experimental phenomenon also indirectly reflected the instability of dimeric crosslinking on the His site.



Colonies: 869 (100%) Colonies: 629 (72%)

Figure S22. Assessment of bacterial viability after 311 nm irritation by using over-expressing *sj*GST-E51*m*-TFMAK BL21 (DE3) cells. By imaging the colonies on the LB plate, we were able to perform the manual counting of the number of colonies on agar plates illuminated by transmitted light to evaluate the viability of the bacteria after photo-DAFEx conjugation.

Supplementary Notes and Tables

Note S1: The amino acid sequence of *m*-TFMAKRS

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARALRHHKYRKTCKR

CRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARAPKPLENTEAAQAQPSGSKFSPAIPVST

QESVSVPASVSTSISSISTGATASALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPF

RELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVD

KNFCLRPMLAPNLANYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFTMLNFCQMGSGCTRENLESIIT

DFLNHLGIDFKIVGDSCMVYGDTLDVMHGDLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKVKHDF

KNIKRAARSESYYNGISTNL*

Where * is the “ochre” stop codon, TAA.

Note S2: The amino acid sequence of sfGFP-Q204TAG-His6

MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFS
RYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEY
NYNSHNVYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTXSALSKDPNE
KRDHMVLLEFVTAAGITHGMDELYKEHHHHHH*

Where * is the “ochre” stop codon, TAA; X = “amber” codon, TAG.

Note S3: The amino acid sequence of sfGFP-N149TAG-His6

MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFS
RYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEY
NYNSHXVYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNE
KRDHMVLLEFVTAAGITHGMDELYKELHHHHHH*

Where * is the “ochre” stop codon, TAA; X = “amber” codon, TAG.

Note S4: The amino acid sequence of EGFP-N149TAG-HA

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCF
SRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLE
YNYNSHXVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPN
EKRDHMVLLEFVTAAGITLGMDELYKMYPYDVPDYA*

Where * is the “ochre” stop codon, TAA; X = “amber” codon, TAG. Underline means HA tag.

Note S4: The amino acid sequence of *sj*GST-E51TAG-His_6

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLXFPNLPYYIDGDVKLTQSMAIIR
YIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNG
DHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGD
HPPKPNSSSVDKLAAALEHHHHHH*

Where * is the “ochre” stop codon, TAA; X = “amber” codon, TAG.

Table S3: The primer sequences used for introduction TAG amber stop codon or create mutants

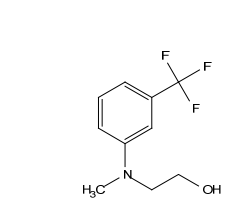
Description	Oligo type	Primer sequence (5’-3’)
*Mm*PylRS-Y384F	Forward	tgcatggtctttggggataccttgatgtaatgc
	Reverse	tatccccaaagaccatgcaggaatcgcctac
*Mm*PylRS-Y306A	Forward	atgcttgctccaaaccttgcgaactacctgcgca
	Reverse	agcttgcgcaggtagttcgcaaggtttggagca
*Mm*PylRS-R19S	Forward	accgggctctggatgtccagcaccggaacaattcat
	Reverse	atgaattgttccggtgctggacatccagagc
*Mm*PylRS-H29R	Forward	tcataaaataaaacaccgcgaagtctctcgaagc
	Reverse	ttcgagagacttcgcggtgttttattttatgaattgttccgg
*Mm*PylRS-T122S	Forward	aaacctcttgagaatagcgaagcggcacag
	Reverse	tgagcctgtgccgcttcgctattctcaagagg
sjGST-E51TAG	Forward	tttgaattgggtttgtagtttcccaatct
	Reverse	taaggaagattgggaaactacaaaccca
sjGST-K139A	Forward	agatcgtttatgtcatgcgacatatttaaat
	Reverse	accatttaaatatgtcgcatgacataaacga
sjGST-C137A	Forward	ttcgaagatcgtttagcgcataaaacatat
	Reverse	tttaaatatgttttatgcgctaaacgatcttcg
sjGST-C137A-K139A	Forward	ttcgaagatcgtttagcgcatgcgacatatttaaat
	Reverse	tttaaatatgtcgcatgcgctaaacgatcttcgaaca

Reference

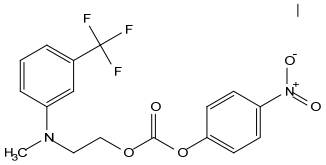
[1] Q. Xiong, T. Zheng, X. Shen, B. Li, J. Fu, X. Zhao, C. Wang and Z. Yu, *Chemical Science*, 2022, **13**, 3571-3581.

[2] Z.-L. Chen, J.-M. Meng, Y. Cao, J.-L. Yin, R.-Q. Fang, S.-B. Fan, C. Liu, W.-F. Zeng, Y.-H. Ding, D. Tan, L. Wu, W.-J. Zhou, H. Chi, R.-X. Sun, M.-Q. Dong and S.-M. He, *Nat. Commun.*, 2019, **10**, 3404.

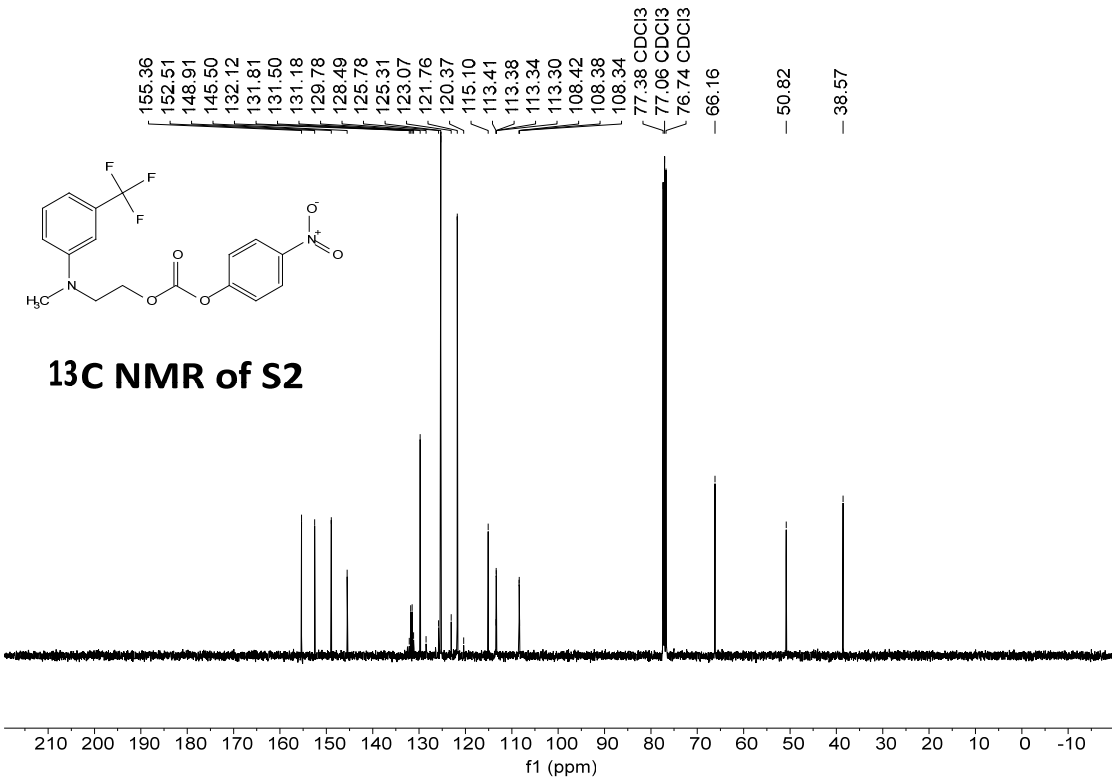
^{1}H NMR, ^{13}C NMR and ^{19}F NMR Spectra



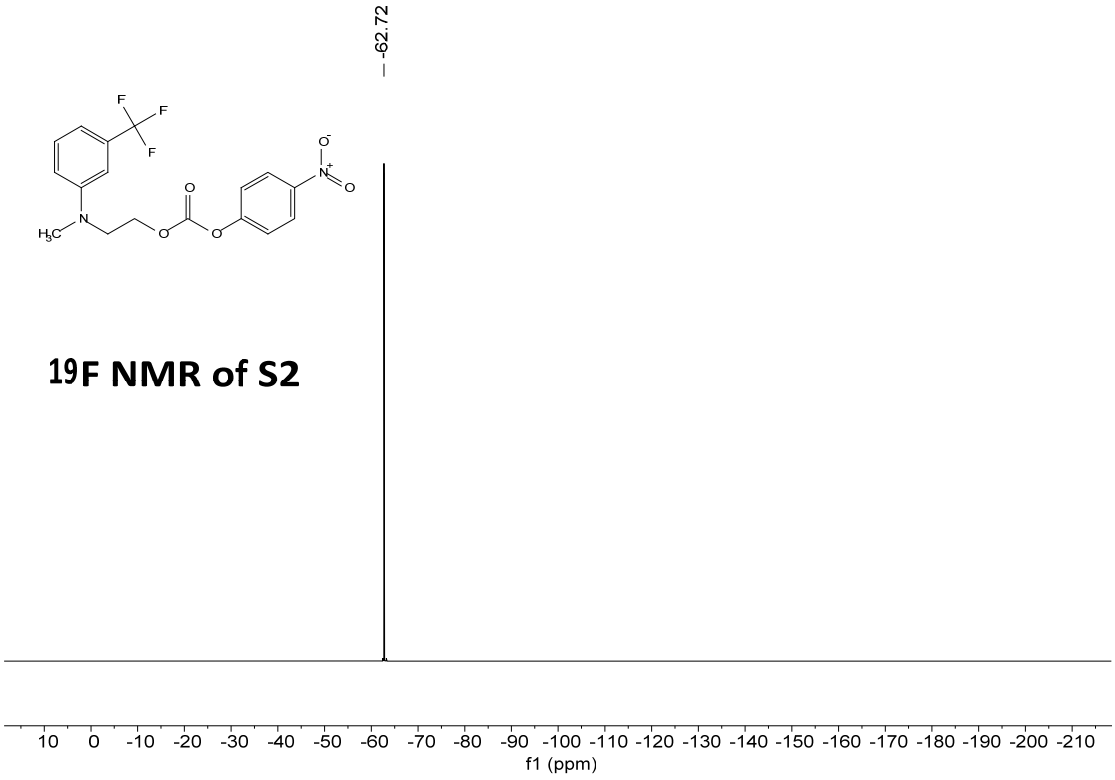
^{1}H NMR of S1



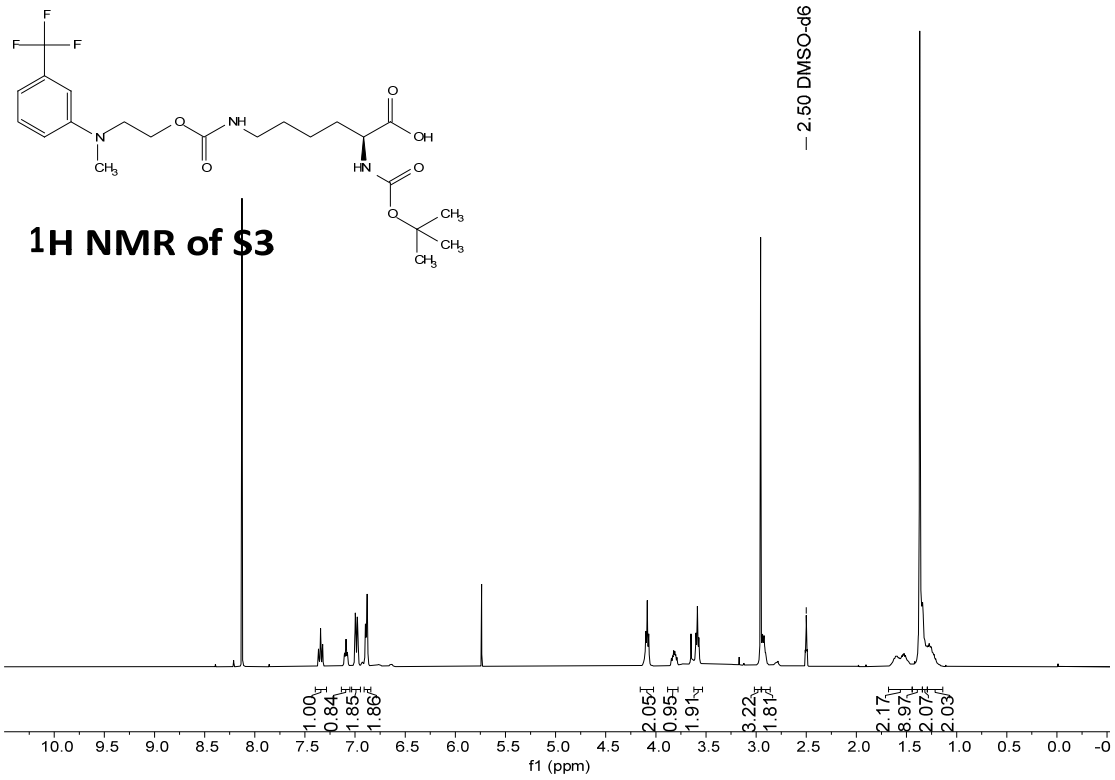
^{1}H NMR of S2



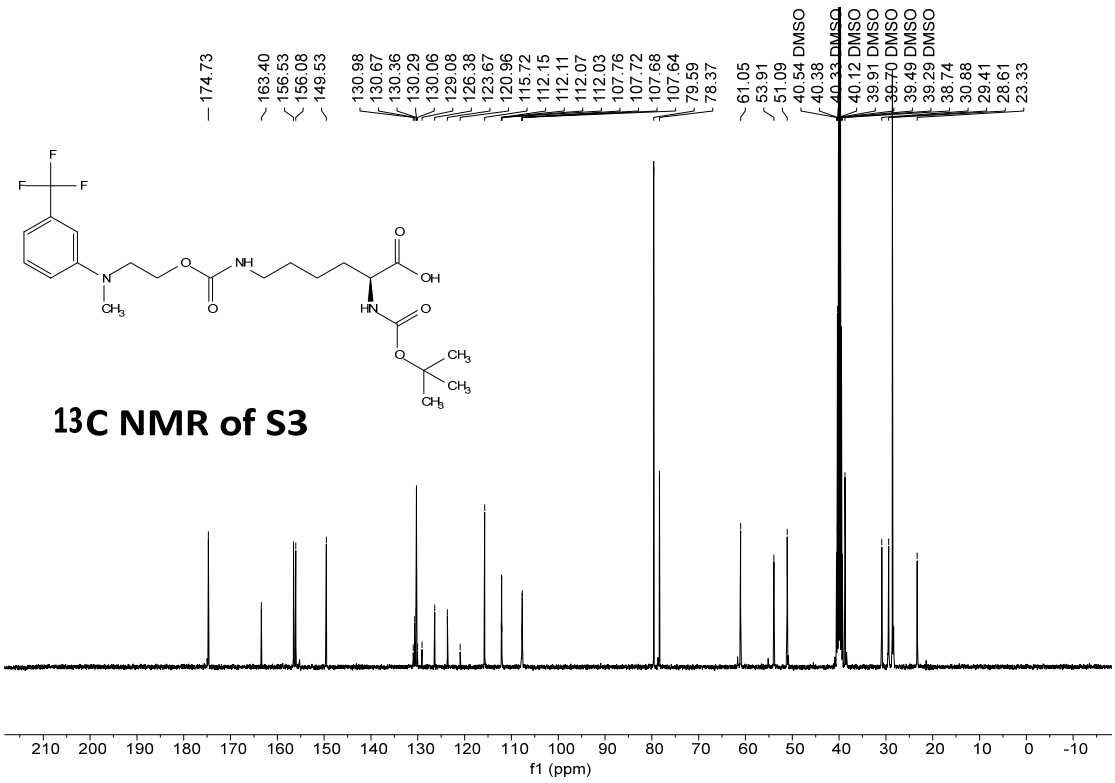
^{13}C NMR of S2



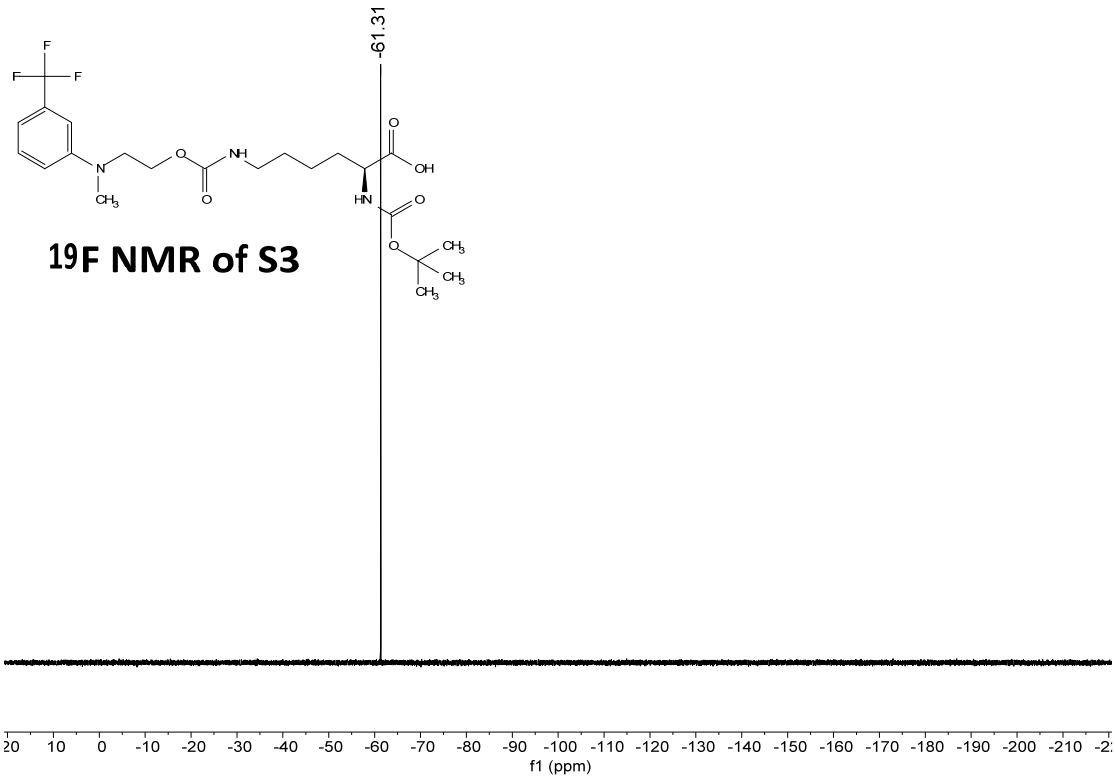
^{19}F NMR of S2



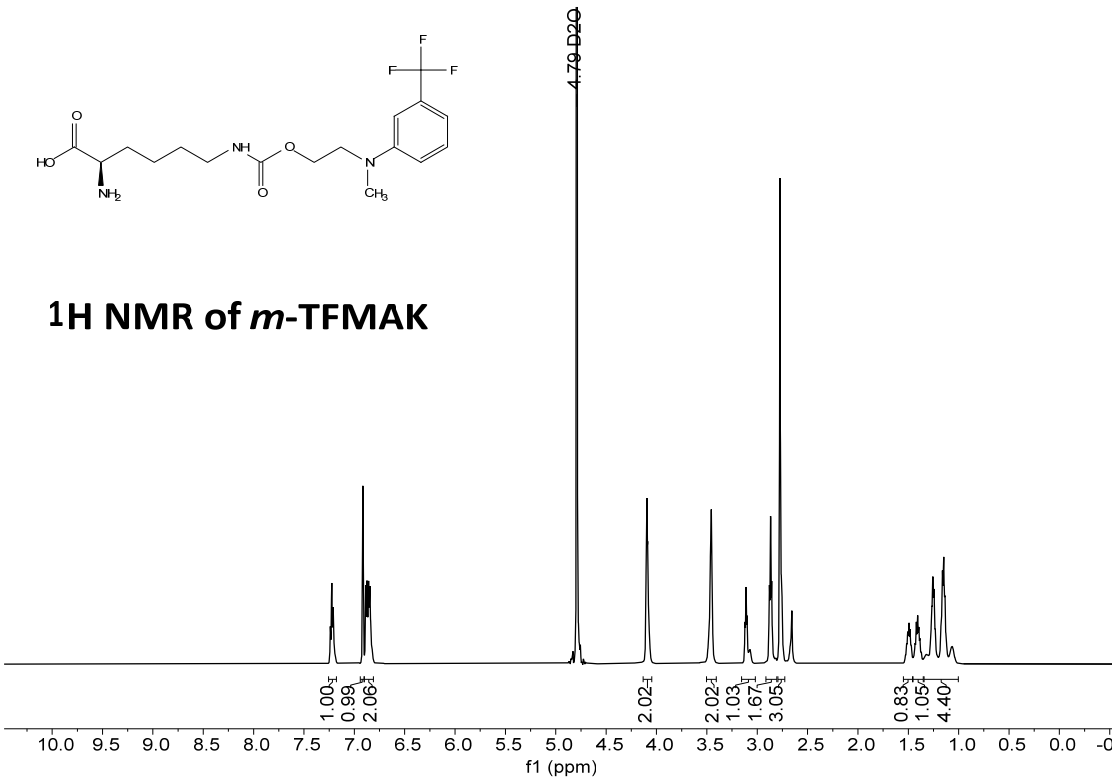
^{1}H NMR of S3



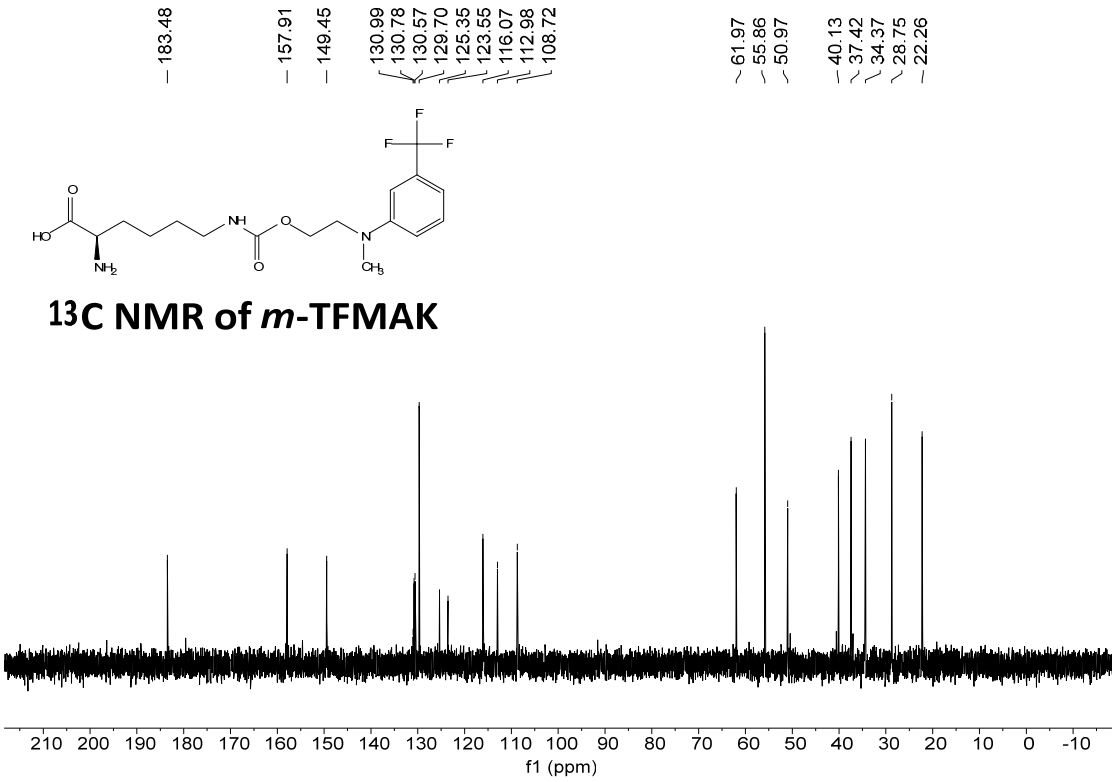
^{13}C NMR of S3



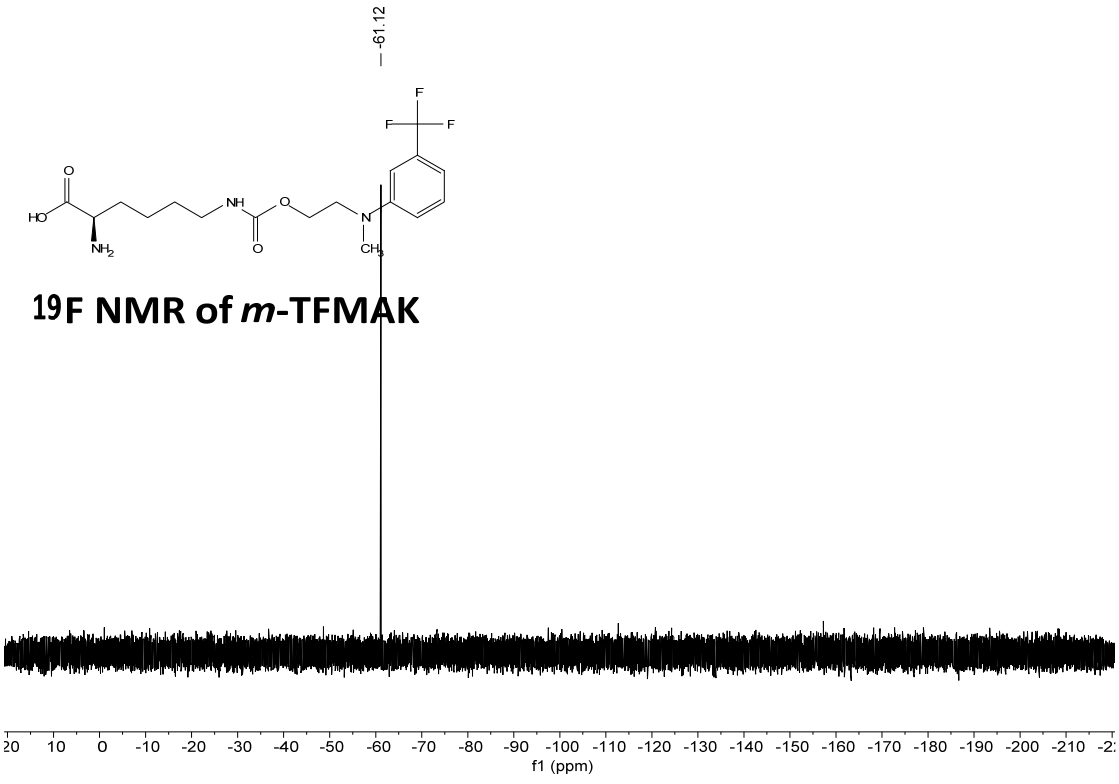
^{19}F NMR of S3



^{1}H NMR of *m*-TFMAK



^{13}C NMR of *m*-TFMAK



^{19}F NMR of *m*-TFMAK