# Protein digestion using a cysteine-specific backbone cleavage reagent

Yu-De Chuang[1§], Tsung-Jung Yang[1§], Yu-Chi Chih[1], Po-Cheng Kuo[1], Cheng-Chih Hsu[1], Shao-Lun Chiou[1*], John Chu[12*]

[1]Department of Chemistry, National Taiwan University, Taipei 106319, Taiwan. [2]Center for Emerging Material and Advanced Devices, National Taiwan University, Taipei 10617, Taiwan

§ These authors contributed equally to this work

* Correspondence: johnchu@ntu.edu.tw and d11223104@ntu.edu.tw

**Table of Contents**

**Reagents, instruments, and consumables**

All chemicals are of ACS grade (or higher) and purchased from Acros, Sigma-Aldrich, Nova-Matls, Morrchem, Fluorochem or TCI Chemicals. Reagents and media for cloning and protein expression were purchased from Bioman (Taiwan); thioesterase was produced in-house. All NMR spectra were obtained using a Bruker AVIII 400 MHz instrument; HRMS spectra were obtained by Bruker Daltonics Autoflex Speed MALDI-TOF mass spectrometry.; LC-MS/MS analysis was performed with a Dionex UltiMate 3000 UHPLC system (Thermo Fisher Scientific) coupled to a Q Exactive Plus hybrid quadrupole-orbitrap mass spectrometer (Thermo Fisher Scientific).

**Synthesis of 1**

A mixture of 1,3-indanedione (146 mg, 1.0 mmol) and 60% (w/v) NaH in mineral oil (60 mg, 1.5 mmol) was prepared in 5 mL of dry DMF, followed by the addition of $CS_2$ (60 µL, 1.0 mmol). The reaction was stirred for 1 h at room temperature. Methyliodide (MeI, 134 µL, 2.0 mmol) was then added and stirred at room temperature for 16 h. The reaction mixture was extracted with $CH_2Cl_2$ and $H_2O$. The organic phase was dried over anhydrous $MgSO_4$, filtered, concentrated under reduced pressure, and purified by silica gel chromatography (hexane/EtOAc = 5:1) to give compound **1** as a yellow solid (120 mg, 48% yield). $^1$H NMR (400 MHz, $CDCl_3$) δ 7.84–7.81 (m, 2H), 7.66–7.64 (m, 2H), 2.62 (s, 6H). $^{13}$C NMR (101 MHz, $D_2O$) δ 187.0, 181.4, 140.5, 134.1, 122.7, 121.9, 21.7. HRMS (ES+) m/z: calcd. for $C_{12}H_{10}O_2S_2$ $[M+H]^+$ 251.0194, found 251.0190.

**Synthesis of 6**

Compound **1** (25 mg, 0.10 mmol) was dissolved in a mixture of acetonitrile and 4-(Cyclohexylamino)-1-butanesulfonic acid (CABS) buffer (1:9, 10 mL, pH ≈ 11.0). GSH (61.4 mg, 0.20 mmol) was added into this solution and stirred at room temperature for 24 h and then quenched by adding 200 µL formic acid, resulting in product precipitation. The solid was collected by vacuum filtration and washed sequentially with water and dichloromethane to afford **6** as a pale yellow solid (15 mg, 46% yield). $^{1}$H NMR (400 MHz, $D_2O$): δ 7.59 (d, *J* = 16 Hz, 4H), 4.98-4.93 (m, 1H), 3.81 (s, 2H), 3.78-3.74 (m, 1H), 3.55 (dd, J = 11.1, 5.6 Hz, 1H). $^{13}$C NMR (101 MHz, $D_2O$) δ 190.7, 176.0, 171.1, 169.8, 137.5, 133.5, 121.0, 100.7, 62.6, 43.4, 32.3. HRMS (ES+) m/z: calcd. for $C_{15}H_{12}N_2O_5S$ $[M+H]^+$: 333.0540, found: 333.0522.

**mTE plasmid construction**

The gene for macolacin TE (mTE) was amplified via polymerase chain reaction (PCR) from the genomic DNA of *Paenibacillus xylanexedens* (BCRC 80422) with the following primers:

Mac TE_Fw: 5'- GGACAGCAAGTGCTACCCGCGCCAGAGGAAG-3'

Mac TE_Rv: 5'- GCGACCCATTCAGCTCCCATGCAGGAC-3'

The PCR product was purified via the Nautia Gene PCR Purification Kit and assembled into pET-28a (+) plasmid using NEBuilder HiFi DNA Assembly Tools. The primer sequences were designed for constructing a recombinant mTE with an N-termianl His tag by Gibson assembly. Amplified Mac TE and linearized pET-28a (+) were combined in a 4:1 ratio and incubated for 60 min at 50°C. The recombinant plasmid was transformed into *Escherichia coli* DH5α, grown on an LB-agar plate containing kanamycin (50 µg/mL), and then screened by colony PCR. Plasmids from select colonies were extracted and confirmed by Sanger sequencing. The sequencing results matched the expected construct (Figures S6–S7).

**mTE expression and purification**

The recombinant mTE plasmid was transformed into BL21 (DE3) for expression. An overnight culture of *E. coli* was diluted 50 times in Terrific Broth containing kanamycin (50 μg/mL) and 1% (w/v) glucose. Bacterial culture was incubated at 37°C until OD600 reached 0.6, and then induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM. The culture was grown for 20 h at 16 °C while shaking at 170 rpm. The bacteria were collected by centrifugation at 8,000 rpm at 4°C for 15 min. The cell pellet was resuspended in the lysis buffer (50 mM Tris-HCl, 100 mM NaCl, 0.1% Triton X-100, pH 8.0) and lysed by sonication. The lysate was centrifuged at 12,000 rpm at 4°C for 30 min. The supernatant was mixed with Ni-NTA resins and incubated at 4°C for 1 h. The slurry was first rinsed by wash buffer (50 mM Tris-HCl, 100 mM NaCl, pH 8.0), followed by elution using the same buffer containing increasing concentrations of imidazole (10, 20, 30, 50, 75, 100, 200, and 500 mM). All collected fractions were examined via SDS-PAGE electrophoresis. The purified mTE protein was concentrated by Ultra Centrifugal Filters (Merck Millipore, US), and the concentration was determined using the Bradford assay.

**Reaction monitoring by high performance liquid chromatography (HPLC)**

Analytical reversed-phase HPLC was performed on a Waters instrument (996 UV detector, 600 pump and controller) equipped with a Hypersil GOLD C18 column (5 μm, 4.6 × 250 mm). A two solvent system was used as the mobile phases at a flow rate of 0.8 mL/min. Water containing 0.1% (v/v) of formic acid (FA) is solvent A, and methanol containing 0.1% (v/v) FA is solvent B. The column temperature was fixed at 40°C. The gradient was programmed as follows:

Isocratic elution 1% B for 6 min;

Gradient elution from 10% B to 70% B over 25 min;

Isocratic elution at 70% B for 12 min.

The HPLC trace was monitored by 290 nm and the reported conversions (%) were determined by comparing the integrated peak areas. The reported yields were determined by comparing to a standard calibration curve established using the authentic hydrolysis product, as well as the use of caffeine as the internal standard.

**Optimization of reaction pH**

Cleavage of GSH (2.0 mM) by **1** (1.0 mM) was performed in various buffer solutions from pH 6.0 to 11.0 (10 mM MES (pH 6.0), 10 mM Tris-HCl (pH 7.0 and 8.0), 10 mM ammonium acetate (pH 9.0), 10 mM CABS (pH 10.0 and 11.0)), in presence of 5% (v/v) DMSO to fully dissolve **1**. After incubation at room temperature for 24 h, a 20 μL aliquot of the reaction mixture was withdrawn and quenched by adding FA to a final concentration of 2% (v/v). Caffeine was co-injected as an internal standard (final concentration 0.25 mM) for HPLC analysis.

**Kinetic studies of GSH hydrolysis**

GSH (2.0 mM) cleavage by **1** (1.0 mM) in the CABS buffer (pH 10.0, with 5% (v/v) DMSO) at room temperature was monitored in detail for kinetic analysis. The reaction was quenched by adding FA to a final concentration of 2% (v/v) at various time points ($t$ = 1 min, 30 min, 2 h, 6 h, 15 h, and 24 h). Caffeine was co-injected as an internal standard (final concentration 0.25 mM) for HPLC analysis. The same reaction was performed at pH 11.0 at a different set of time points ($t$ = 1 min, 15 min, 30 min, 50 min, 2 h, and 24 h).

**Kinetic fitting**

Livermore Solver for Ordinary Differential Equations with Automatic method switching (LSODA) was used in Google Colab to fit datasets from HPLC results. Variables including k1, k2, and reaction orders of $OH^-$, GSH, MeSH as catalysts were tested in the initial assessment of the rate equations (Figure S12 to S14). Separated or shared k1 and k2 values for both pH 10 and

pH 11 conditions were tested in the fitting, with the reaction order of $OH^-$ set as variables (Figure S15). Finally, the reaction order of $OH^-$ was determined as 0.5 and 1 for the first reaction and the second reaction, and the k1 and k2 values were fitted again separately for pH 10 and pH 11 datasets to generate the final result (Figure S16 and S17).

**SDS-PAGE analysis**

SDS-PAGE at 12% acrylamide running were performed following standard lab procedures. A 5% stacking gel was used and a broad-range MW marker (10-180 kDa, Prestained Protein Ladder, BIOMAN SCIENTIFIC) was used to estimate protein weights. 24 µL protein samples were mixed with 6 µL loading buffer (composition for 5 × SDS: 10% SDS, 30% glycerol, 250 mM Tris-HCl buffer pH = 6.8, 500 mM DTT, 0.05% bromophenol blue), heated at 95°C for 5 min. 30 µL of each sample was subsequently loaded into the wells. The gel was run at constant 120V for 80 min in SDS running buffer (1X). The gel was further stained in staining solution (Imperial protein Stain, Thermo Fisher Scientific) for 1 h and further destained overnight, followed by visualized inspection.

**pH-Dependent mTE cleavage assays**

The pH-dependency of mTE cleavage induced by the reagent **1** was evaluated across a pH range of 6.0-11.0. To ensure the protein was in a denatured state, mTE (10 $\mu$M) was pre-incubated in various buffers containing 8 M urea. The reagent **1**, prepared as a 20 mM stock solution in DMSO, was added to the mixture to achieve a final concentration of 200 $\mu$M. Reactions were conducted in the following buffers, all containing 150 mM NaCl: 10 mM MES (pH 6.0), 10 mM Tris-HCl (pH 7.0 and 8.0), 10 mM ammonium acetate (pH 9.0), 10 mM CABS (pH 10.0 and 11.0). The reaction mixtures were incubated at 25°C for 24 h. Following incubation, 24 $\mu$L aliquots were quenched with 6 $\mu$L of SDS-PAGE loading buffer (containing 500 mM DTT). After thermal denaturation at 95°C for 10 min, the cleavage profiles were

resolved using 12% SDS-PAGE.

**mTE cleavage under different denaturation conditions**

To investigate the influence of protein conformation on cleavage efficiency, mTE (10 $\mu$M) was subjected to four distinct denaturation treatments at pH 11.0 (10 mM CABS buffer, 150 mM NaCl) prior to the addition of the reagent **1**:

1. Native/control: Pre-incubation in buffer at 25°C.
2. Thermal denaturation: Pre-incubation in buffer at 95°C for 10 min.
3. Chaotropic denaturation: Pre-incubation in buffer containing 8 M urea.
4. Surfactant denaturation: Pre-incubation in buffer containing 10% SDS.

Following the respective pre-treatments, the reagent **1** (20 mM stock in DMSO) was added to each mixture to reach a final concentration of 200 $\mu$M. The reaction mixtures were then incubated at 25°C for 24 h. A 24 $\mu$L aliquot from each condition was quenched with 6 $\mu$L of SDS-PAGE loading buffer (containing 500 mM DTT). After thermal denaturation at 95°C for 10 min, the cleavage profiles were resolved using 12% SDS-PAGE.

**LC-MS/MS analysis**

mTE was excised from SDS-PAGE gel, and washed by pH 8.0 20 mM triethylammonium bicarbonate (TEABC) buffer in supplement of 50% ACN. Then, the gel was dehydrated by ACN, and then dried by vaccum. 20 mL 20 $\mu$g/mL of trypsin in 20 mM pH 8.0 TEABC buffer in supplement of 0.5 mM DTT was added into the gel pieces, which was then incubated at 37°C for 16 hours to digest the protein and reduce it at the same time. 20 mM **1** was added to the gel piece to a final concentration of 2 mM, and then 20 mL pH 11 50 mM CABS buffer in 1:20 DMSO/$H_2O$ was added to adjust the system to pH 11 to initiate the reaction. The mixture was incubated at room temperature for 16 hours and quenched by adding 100 mL 5% formic acid in 1:1 (v/v) $H_2O$/ACN. The gel pieces were extracted by 5% formic acid in 1:1 (v/v) $H_2O$/ACN twice, and the fractions were combined, desalted and then freeze-dried. Samples were stored

at -80ºC fridge and reconstituted by $H_2O$ just before MS analysis. LC-MS/MS analysis was performed with a Dionex UltiMate 3000 UHPLC system (Thermo Fisher Scientific) coupled to a Q Exactive Plus hybrid quadrupole-orbitrap mass spectrometer (Thermo Fisher Scientific). The peptides were separated using ACQUITY UPLC BEH C18 column (130 Å, 1.7 $\mu$m, 2.1 x 100 mm, Waters), with solvent A (0.1% formic acid in water) and solvent B (0.1% formic acid in acetonitrile) as the mobile phases at 45℃ and a flow rate of 300 $\mu$L/min. A 60-min gradient was applied for elution as follows: 0-0.5 min, 2% B; 0.5-35 min, 2-40% B (linear); 35-45 min, 40-60% B (linear); 45-47.5 min, 60-99% B (linear); 47.5-53.5 min, 99% B; 53.5-54 min, 99-2% B (linear); and 54-60 min, 2% B. A heated electrospray ionization (HESI) probe was used as the ion source, and the spray voltage was set at 3.5 kV. The capillary temperature was set at 320℃, and the auxiliary gas temperature was set at 350℃. The spectra of full MS scan (m/z 375-2000) were acquired in the Orbitrap mass analyzer at 70,000 resolution for a maximum injection time of 100 ms with an automatic gain control (AGC) target value of 3e6 in positive ionization mode, followed by 10 data-dependent acquisition (DDA) MS/MS scans at resolution 17,500 for a maximum injection time of 50 ms with an AGC target value of 1e5. The top 10 most abundant precursors were selected for MS2 analysis with an isolation window of 2.0 *m/z* and fragmented by higher-energy collisional dissociation (HCD) using a normalized collision energy of 28%. The dynamic exclusion time was set to 30 s, and all mass spectra were recorded in profile mode.

**Mass spectrometry database search**

All MS raw files were processed with MaxQuant v2.7.5.0[1] against the MTE protein sequence for protein analysis. The tolerance for spectra search allowed 10 ppm mass tolerance for precursor, 0.1 Da for product-ions, and digestion specificity (trypsin and **1**) with up to 4 missed cleavages for a more stringent search. The search engine considered dynamic CA reaction (+154.005 Da) on Cys residues, oxidation (+15.995 Da) on Met residues, and deamidation

(+0.984 Da) on Asn and Gln residues with no fixed modification set. The minimal peptide length was set as 7 residues. The false discovery rate (FDR) of peptide and protein were all set as 1%.

**Table S1**. Cleavage yield of GSH in different buffers.

| buffers | Cleavage efficiency (%) | **1** remaining (%) | **4** remaining (%) |
|---|---|---|---|
| pH 6.0 | 0 | 53 | 0 |
| pH 7.0 | 0 | 29 | 0 |
| pH 8.0 | 20 | 26 | 19 |
| pH 9.0 | 26 | 12 | 8 |
| pH 10.0 | 71 | 9 | 19 |
| pH 11.0 | >99 | 0 | 0 |

**Table S2.** Cleavage yield of GSH at pH 10.0 over time.

| Time (h) | **6** (mM) | **1** (mM) | **4** (mM) |
|---|---|---|---|
| 0.02 | 0.000 | 0.970 | 0.000 |
| 0.5 | 0.080 | 0.702 | 0.060 |
| 2 | 0.200 | 0.618 | 0.112 |
| 6 | 0.300 | 0.410 | 0.228 |
| 15 | 0.503 | 0.252 | 0.201 |
| 24 | 0.745 | 0.097 | 0.195 |

**Table S3.** Cleavage yield of GSH at pH 11.0 over time.

| Time (h) | **6** (mM) | **1** (mM) | **4** (mM) |
|---|---|---|---|
| 0.02 | 0.034 | 0.698 | 0.111 |
| 0.25 | 0.322 | 0.447 | 0.150 |
| 0.5 | 0.538 | 0.304 | 0.138 |
| 0.83 | 0.715 | 0.146 | 0.073 |
| 2 | 0.970 | 0.021 | 0.019 |
| 24 | 0.990 | 0.000 | 0.000 |

**Table S4.** Chromatography peaks area ratio of **6** and caffeine at 290 nm.

| **6** | caffeine | **6** / caffeine |
|---|---|---|
| 1.2 mM | 1.00 mM | 3.690 |
| 0.6 mM | 1.00 mM | 1.899 |
| 0.12 mM | 1.00 mM | 0.635 |
| 0.06 mM | 1.00 mM | 0.237 |
| 0.012 mM | 1.00 mM | 0.067 |
| 0 mM | 1.00 mM | 0 |

**Table S5.** Table of found fragments of mTE in protein sequencing analysis.

| Sequence | **1** (C) | **1** (N-term) | Deamidation (N) | Ox (M) | Missed cleavages | MS/MS m/z | Charge | Mass |
|---|---|---|---|---|---|---|---|---|
| ATLTYEAQLINSGTITAR | 0 | 0 | 0 | 0 | 0 | 962.5155 | 2 | 1922.0109 |
| ATTQAYADYR | 0 | 0 | 0 | 0 | 0 | 580.2734 | 2 | 1158.5306 |
| ATTQAYADYRLEGAHEELLELAR | 0 | 0 | 0 | 0 | 1 | 874.4417 | 3 | 2619.2929 |
| CFGGNLTFEVAK | 1 | 0 | 0 | 0 | 0 | 720.3192 | 2 | 1438.6227 |
| CFPPGSGFGIGYR | 1 | 0 | 0 | 0 | 0 | 1511.6425 | 1 | 1510.6340 |
| CFPPGSGFGIGYRELAGR | 1 | 0 | 0 | 0 | 1 | 1019.9705 | 2 | 2036.9203 |
| DILVQIK | 0 | 0 | 0 | 0 | 0 | 828.5203 | 1 | 827.5117 |
| DNFFELGGNSLSLIR | 0 | 0 | 0 | 0 | 0 | 561.2902 | 3 | 1680.8471 |
| DSEAFRLEHQLPSWQR | 0 | 0 | 0 | 0 | 1 | 666.9988 | 3 | 1997.9708 |
| DTLTPYETSEK | 0 | 0 | 0 | 0 | 0 | 642.3047 | 2 | 1282.5929 |
| DTLTPYETSEKELEEMLADFDEEEK | 0 | 0 | 0 | 0 | 1 | 998.1132 | 3 | 2990.3114 |
| DTLTPYETSEKELEEMLADFDEEEKELMSNPLVR | 0 | 0 | 0 | 0 | 2 | 1344.9648 | 3 | 4029.8599 |
| EGDLNVF | 0 | 0 | 0 | 0 | 0 | 793.3727 | 1 | 792.3654 |
| ELEEMLADFDEEEK | 0 | 0 | 0 | 0 | 0 | 863.8735 | 2 | 1725.7291 |
| ELEEMLADFDEEEKELMSNPLVR | 0 | 0 | 0 | 0 | 1 | 923.1030 | 3 | 2765.2776 |
| ELMSNPLVR | 0 | 0 | 0 | 0 | 0 | 529.7878 | 2 | 1057.5590 |
| GSHMASMTGGQQVLPAPEDAATGGTEYVAPR | 0 | 0 | 0 | 0 | 0 | 1029.8177 | 3 | 3085.4233 |
| IYELIAK | 0 | 0 | 0 | 0 | 0 | 425.2576 | 2 | 848.5008 |
| LEGAHEELLELAR | 0 | 0 | 0 | 0 | 0 | 493.9327 | 3 | 1478.7729 |
| LEGGFVLYGIDFIDDATDYDAMLNR | 0 | 0 | 0 | 0 | 0 | 1412.6649 | 2 | 2822.3109 |
| LEHQLPSWQR | 0 | 0 | 0 | 0 | 0 | 431.8953 | 3 | 1292.6626 |
| LVQAVYDETGIEIPLNR | 0 | 0 | 0 | 0 | 0 | 966.0204 | 2 | 1929.0207 |
| QQIEAEAGVLHGS | 0 | 0 | 0 | 0 | 0 | 1338.6648 | 1 | 1337.6575 |
| QVGVKDNFFELGGNSLSLIR | 0 | 0 | 0 | 0 | 1 | 732.0618 | 3 | 2192.1590 |
| VDETAVVIR | 0 | 0 | 0 | 0 | 0 | 501.2856 | 2 | 1000.5553 |
| VDETAVVIRDILVQIK | 0 | 0 | 0 | 0 | 1 | 605.0385 | 3 | 1810.0564 |
| VWQDTLGVR | 0 | 0 | 0 | 0 | 0 | 537.2910 | 2 | 1072.5665 |
| YVDEIVR | 0 | 0 | 0 | 0 | 0 | 447.2408 | 2 | 892.4654 |

**Figure S1.** Chemical reagents known to cleave proteins at Cys. **a)** This work. References for reagents shown in **b)** and **c)** are cited in the main text.

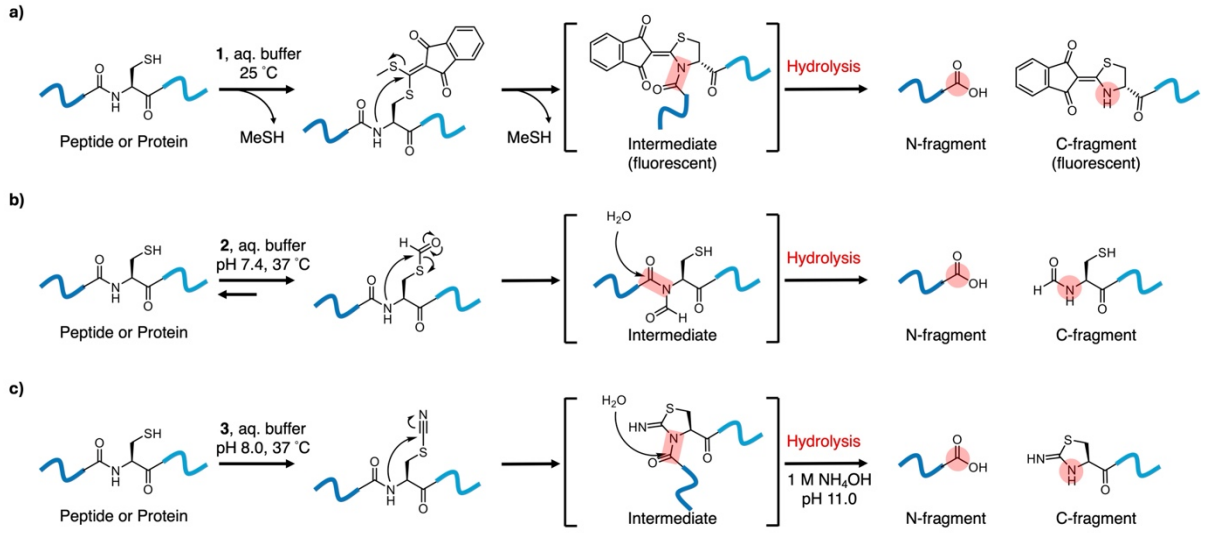


**Figure S2**. $^{1}H$ NMR for **1** in $D_2O$.

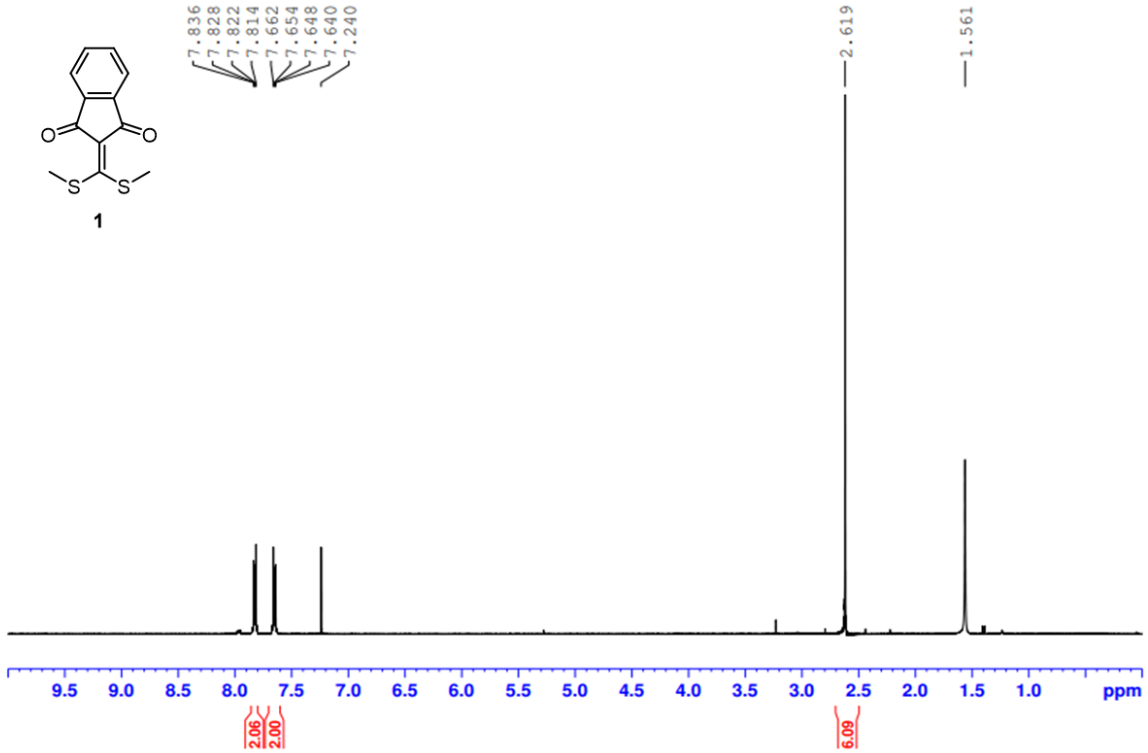



**Figure S3.** $^{13}C$ NMR for **1** in $D_2O$.

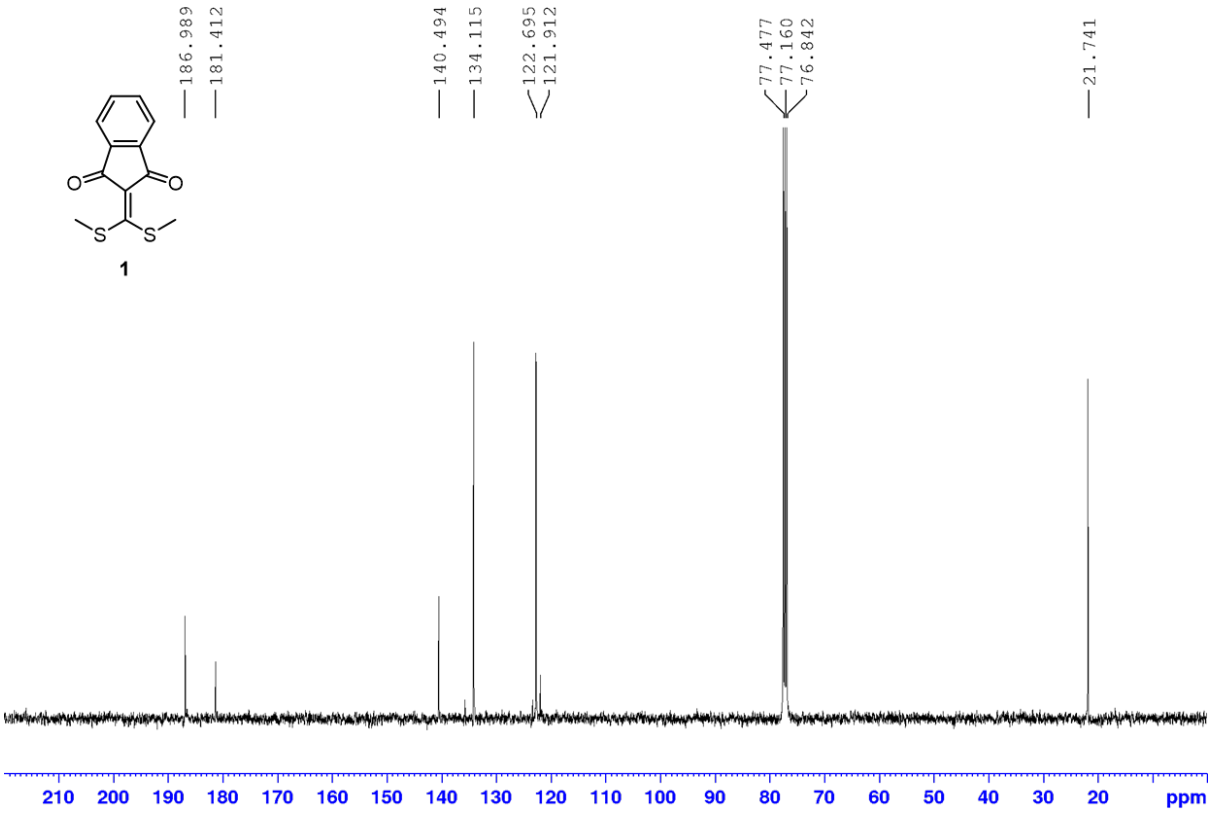


**Figure S4.** $^{1}H$ NMR for **6** in $D_2O$.

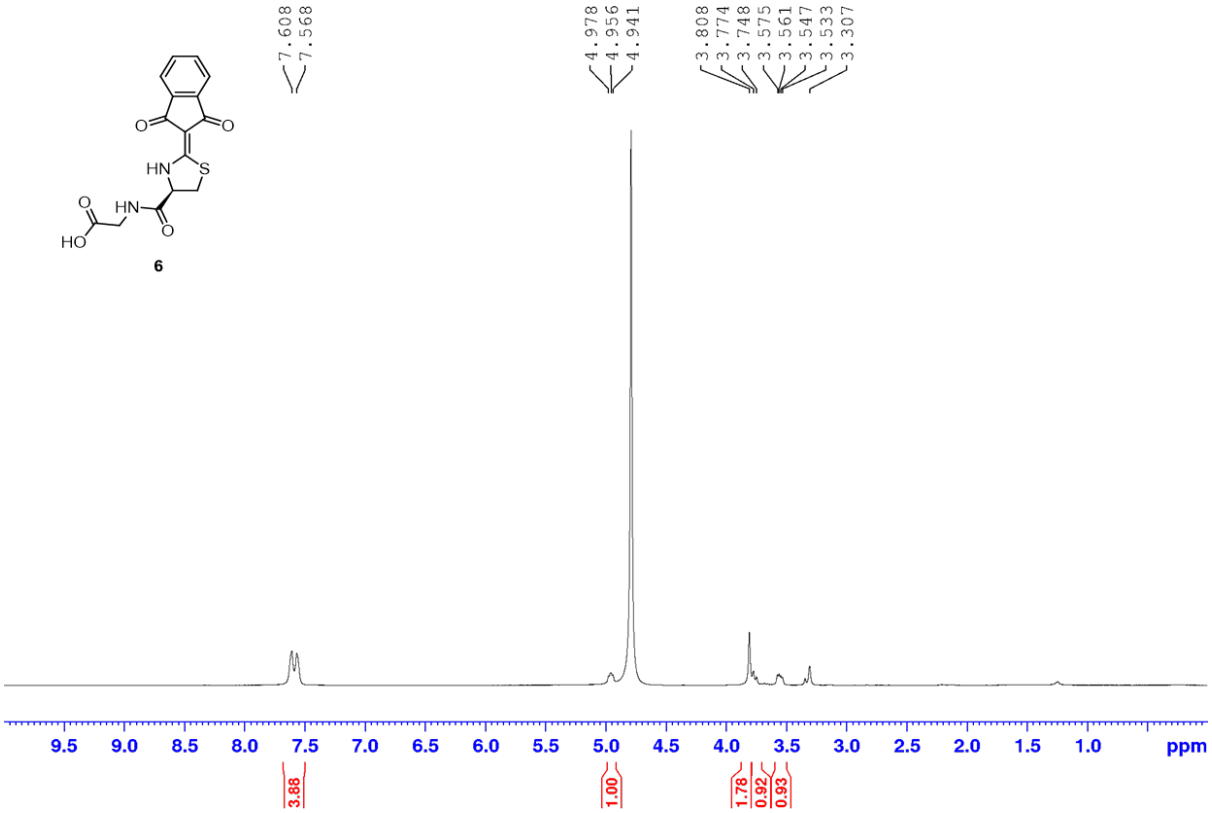



**Figure S5**. $^{13}C$ NMR for **6** in $D_2O$.

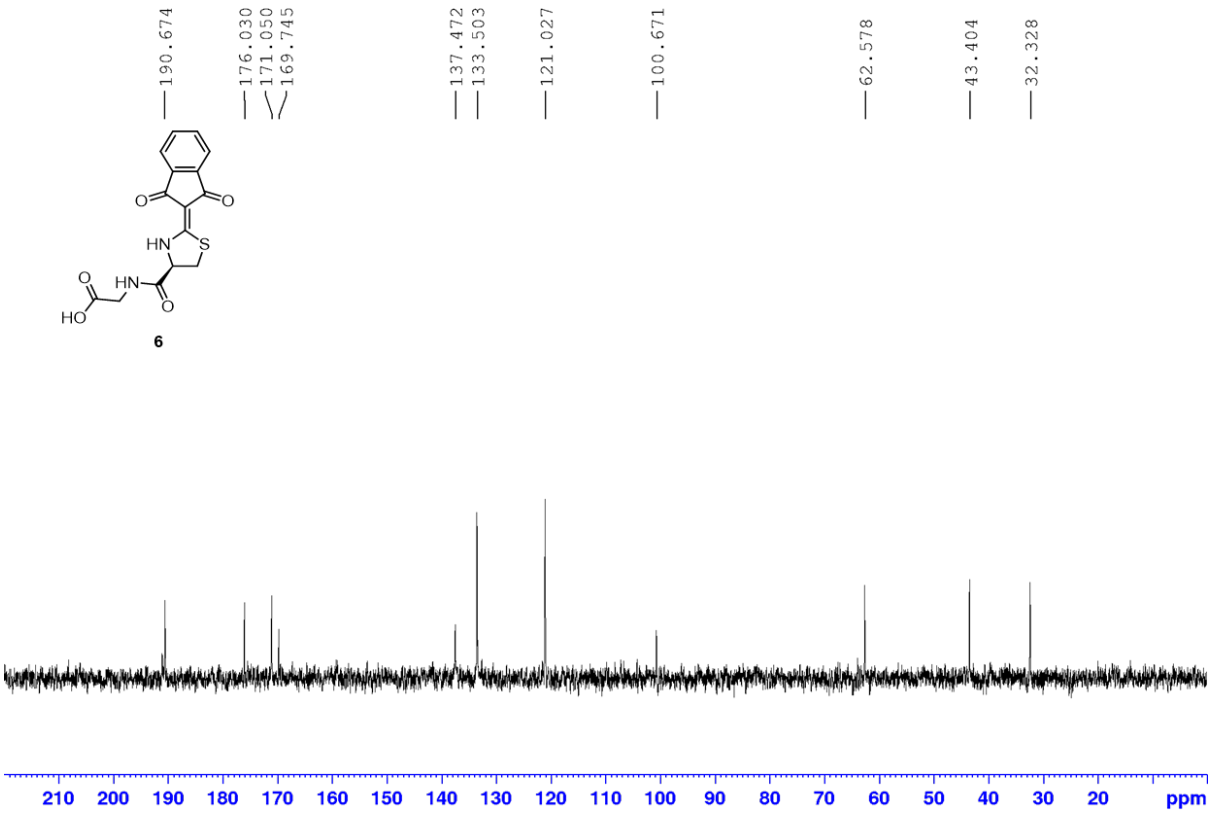


**Figure S6.** Construction and sequence validation of the recombinant pET-28a(+)-mTE plasmid. **a)** Plasmid map of the pET-28a(+)-mTE expression vector. **b)** Sanger sequencing chromatogram of the 5' vector-insert junction obtained using the T7 promoter primer. **c)** Sanger sequencing chromatogram of the 3' vector-insert junction obtained using the T7 terminator primer.

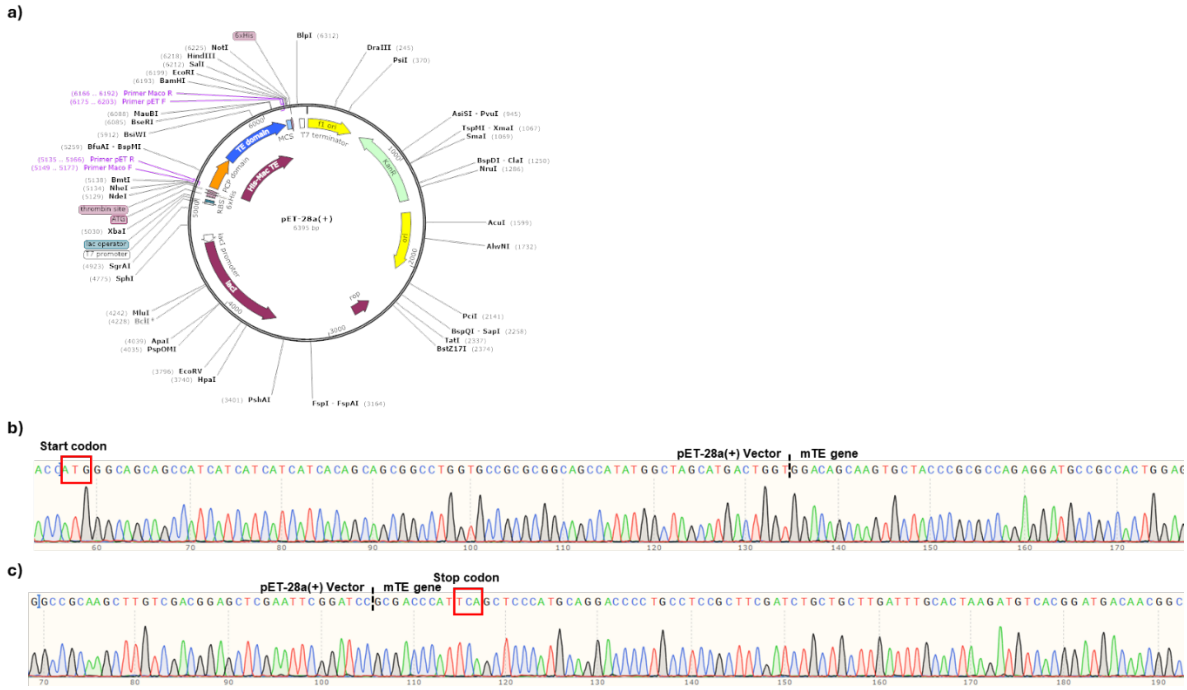


**Figure S7.** Sanger sequencing validation of the mTE gene. Multiple sequence alignment of the expected pET-28a(+)-mTE sequence with forward (M7_T7P) and reverse (M7_T7T) sequencing reads.

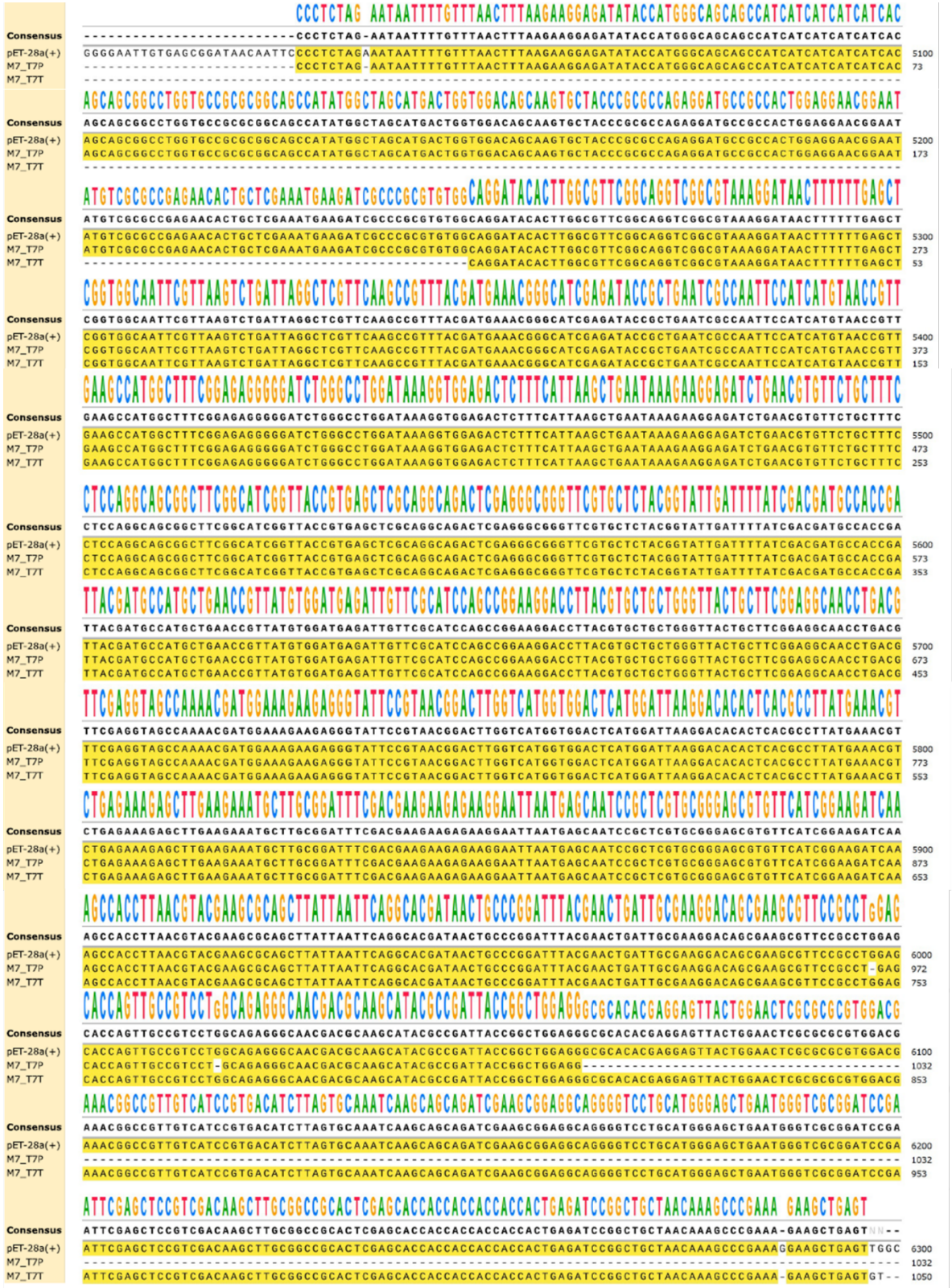


**Figure S8.** HPLC analysis of the cleavage reaction of GSH in different buffers.

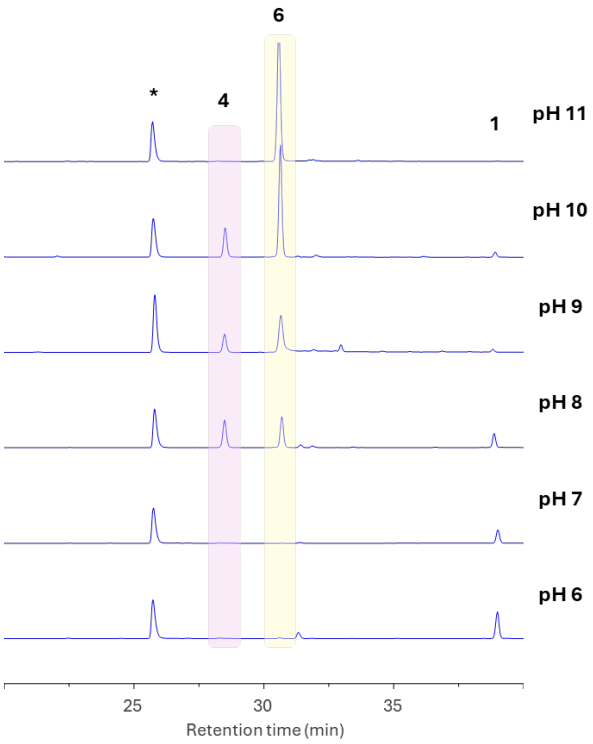



**Figure S9.** HPLC analysis (UV absorption at 290 nm) of the cleavage reaction of GSH under pH 10.0 at different time points from 1 min to 24 h.

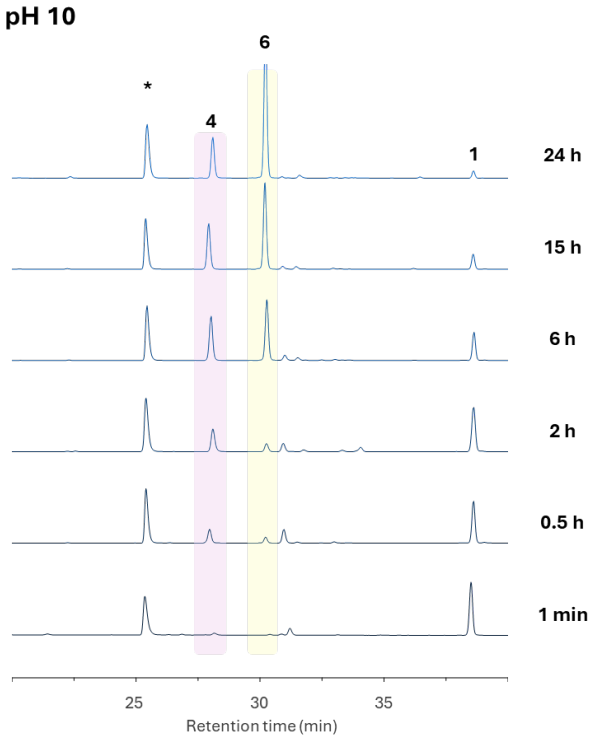


**Figure S10.** HPLC analysis (UV absorption at 290 nm) of the cleavage reaction of GSH under pH 11.0 at different time points from 1 min to 24 h.

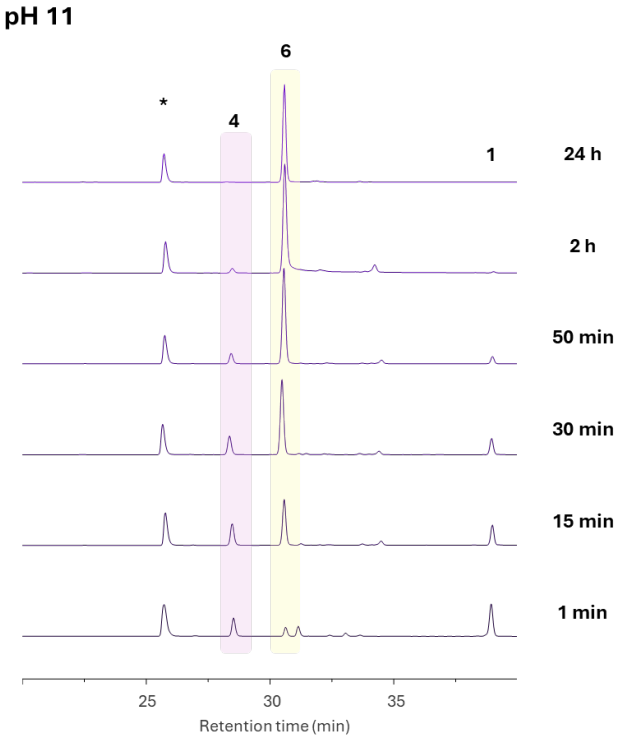



**Figure S11.** Standard calibration curve established using the isolated **6** and caffeine as the internal standard.

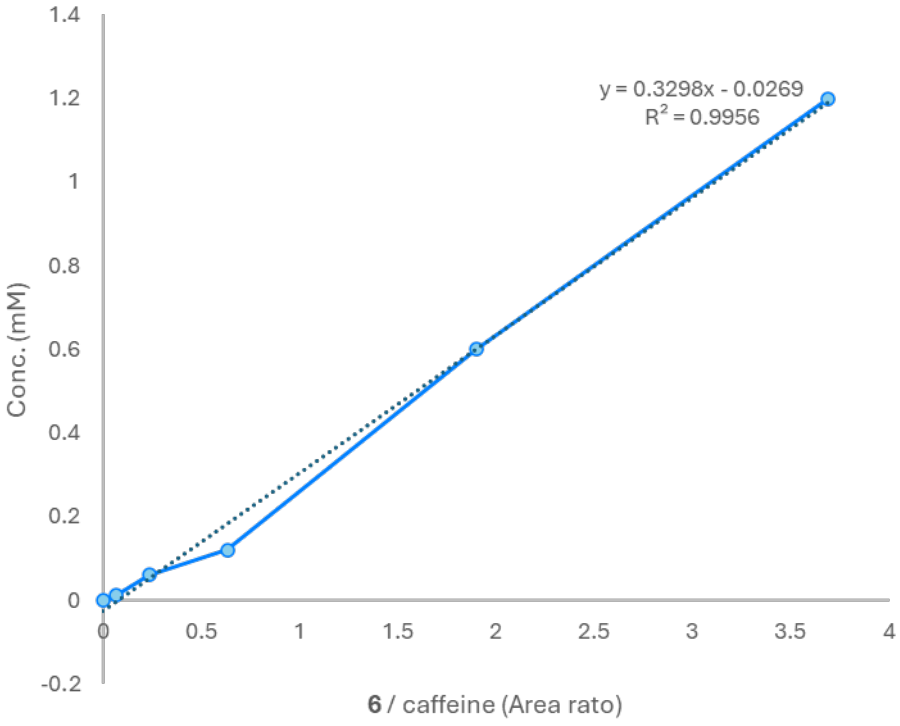


**Figure S12**. Possible reaction pathway of GSH hydrolysis. Compounds were abbreviated as capitalized alphabets A to E and PN and PC. The fitting procedure was stated in following figures (Figure S12 to S16).

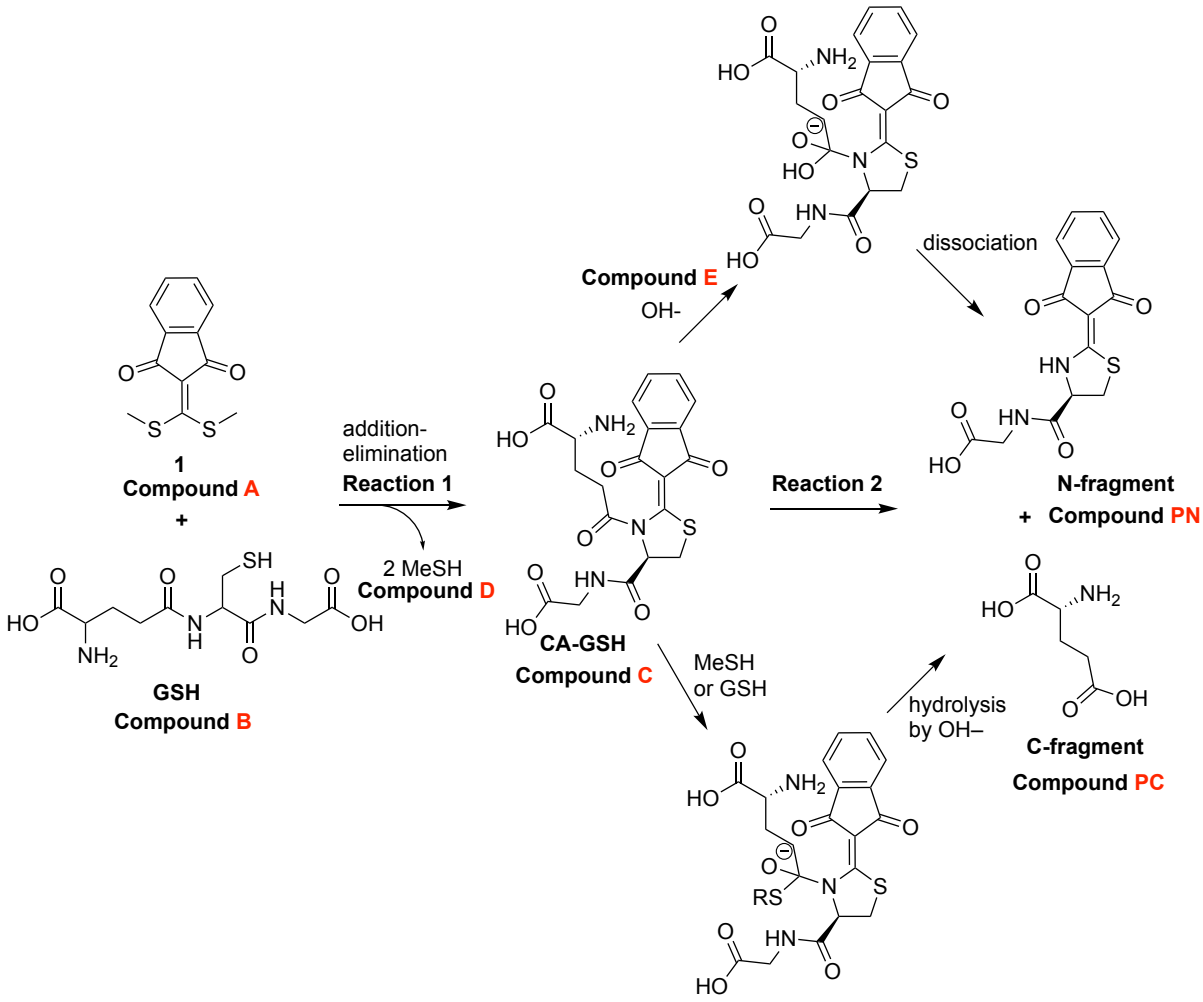
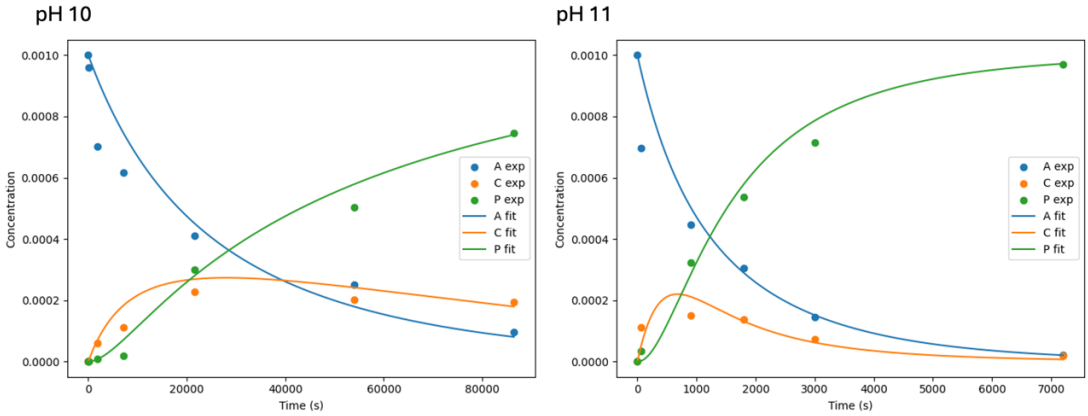


**Figure S13.** Fitting result including GSH (**B**), MeSH (**D**) as catalysts and $OH^-$ (**E**) with variant coefficient for Reaction 2. The fitting was done separately for pH 10 and pH 11. The reaction was written as

$$Reaction\ 1: : A + B \rightarrow C + 2D$$

$$Reaction\ 2: C + alpha\ B + beta\ D + gamma\ E \rightarrow PN + PC + alpha\ B + beta\ D$$

k1 and k2 refer to the reaction constants of Reaction 1 and Reaction 2 respectively. The goal of initial fitting is to obtain similar k values and reaction orders under both pH conditions since the reaction mechanism should be similar. However, the alpha value 2.8 is not chemically reasonable since there's no possible pathway that may consume almost 3 GSH in the Reaction 2. We suggest that including alpha may cause the overfitting even though the curve fits well.




Fitted parameters:
k1 = 2.204e-02
k2 = 3.768e+03
alpha = 2.782 (order in B)
beta = 0.000 (order in D)
gamma = 0.000 (order in E)

Fitted parameters:
k1 = 4.449e-01
k2 = 2.453e+00
alpha = 0.479 (order in B)
beta = 0.189 (order in D)
gamma = 0.379 (order in E)

**Figure S14.** Fitting result including MeSH (**D**) as catalyst and $OH^-$ (**E**) with variable coefficient for Reaction 2 (top). Fitting result including only $OH^-$ (**E**) with variable coefficient for Reaction 2 (bot). Reaction was written as:

$$Reaction\ 1\colon A + B \rightarrow C + 2D$$

$$Reaction\ 2\colon C + beta\ D + gamma\ E \rightarrow PN + PC + beta\ D$$

Comparing the results, MeSH caused worse fitting in pH 11 data, and the fitted reaction order in pH 10 is 0, suggesting MeSH did not play a role in catalysis. Moreover, the boiling point of MeSH is only 5.9°C, so it is also possible that MeSH may dissipate through the time course. The order of $OH^-$ is 1 if no catalysts was taken into calculation (bot), aligning with the mechanistic suggestion of hydrolysis under alkaline. The rate constant k2 is comparable at both pH, while the difference of k1 is large, suggesting missing parameters in Reaction 1.

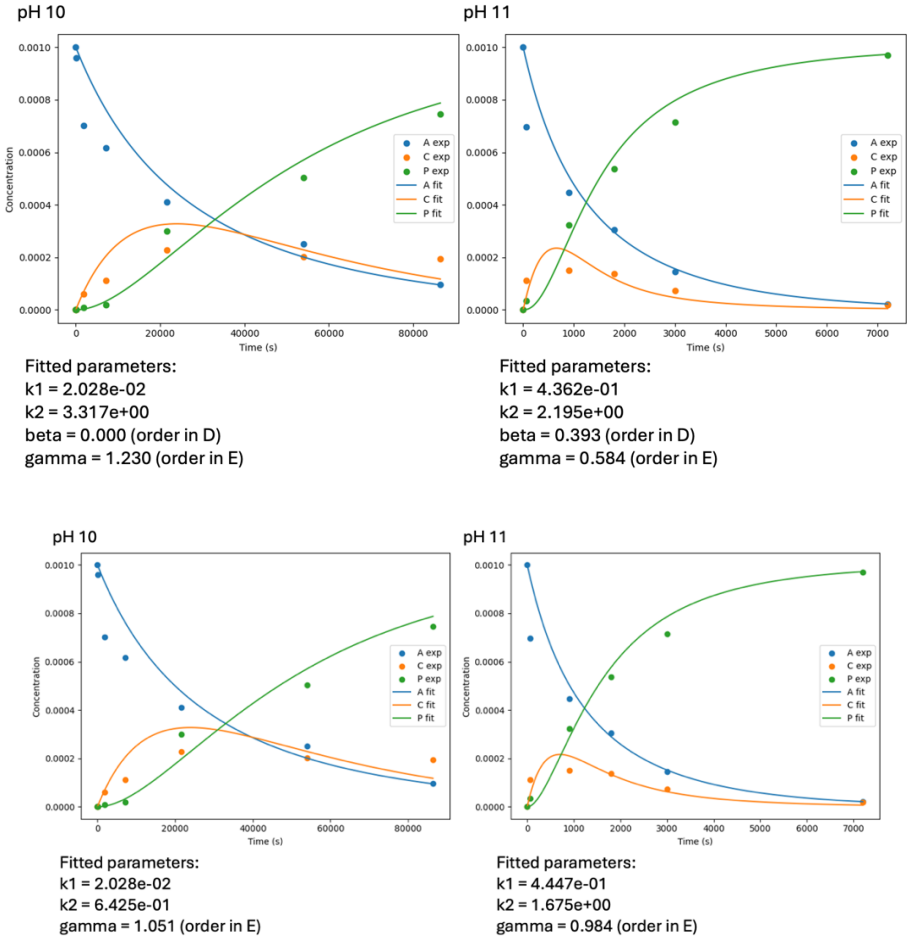


**Figure S15.** Fitting result including $OH^-$ (**E**) for Reaction 1 and 2 (bot), and GSH (**B**), MeSH (**D**) additionally for Reaction 2 (top). The fitting was done separately for pH 10 and pH 11. The reaction was written as

$$Reaction\ 1\colon A + B + gamma1\ E \rightarrow C + 2D + gamma1\ E$$

$$Reaction\ 2\colon C + alpha\ B + beta\ D + gamma2\ E \rightarrow PN + PC + alpha\ B + beta\ D$$

Even though taking $OH^-$ as catalyst for Reaction 1 into account, GSH still caused overfitting, and MeSH may not participate in catalysis. Moreover, the order of $OH^-$ is consistently 1 for Reaction 2 under both pH conditions. Hence, Figure S14 (bot) is the most reasonable fitting result among Figure S12 to S14.

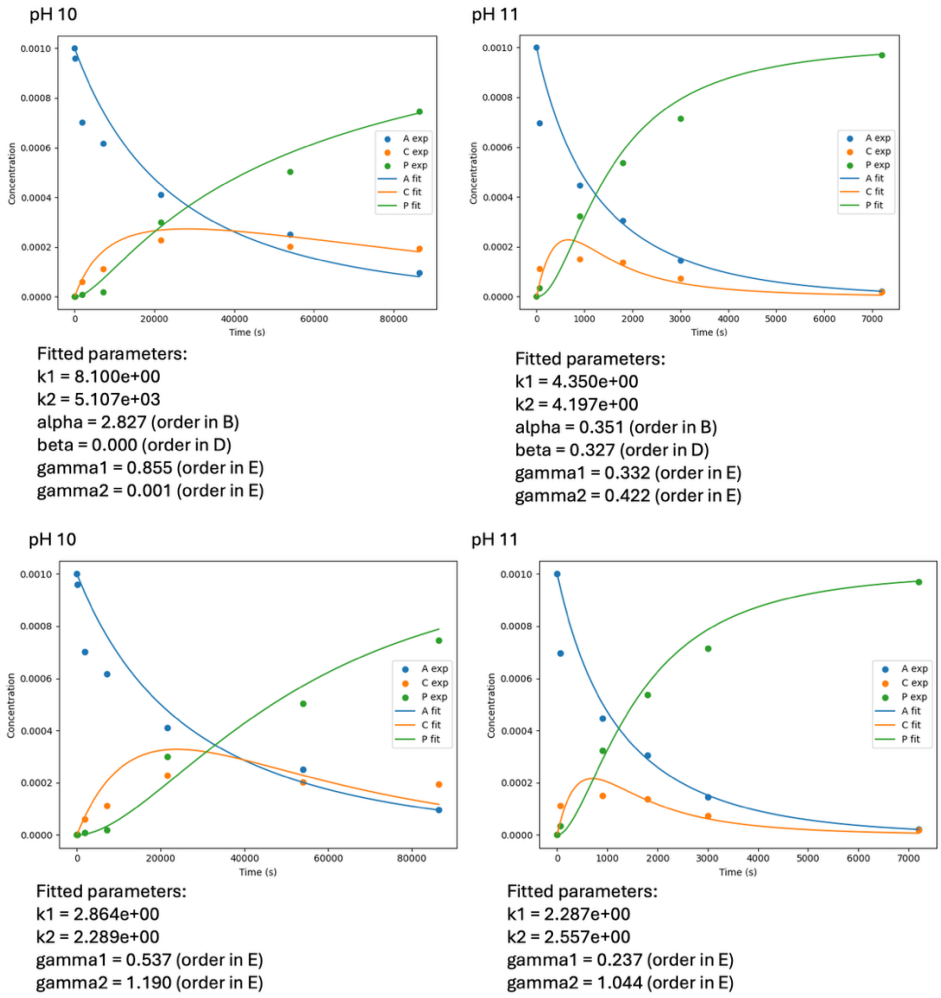
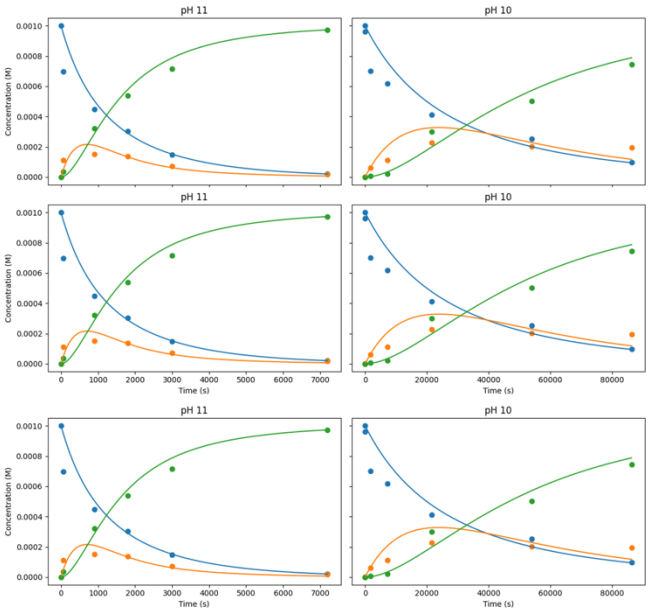


**Figure S16.** Global fitting result of the reactions. As the reaction mechanism should be similar at pH 10 and pH 11, the rate equation should share the same reaction order, and the remaining factors should be combined into rate constants. The reaction was written as

$$A + B + gamma1\ E \rightarrow C + 2D + gamma1\ E$$

$$C + gamma2\ E \rightarrow PN + PC$$

Global fitting was done to the pH 10 and pH 11 datasets. Shared or separated reaction constants k1 and/or k2 were tried in the fitting, and the results suggested that the reaction constants should be separated for both k1 and k2 in each condition (top). The order of OH– is around 0.5 in Reaction 1, suggesting the acid-base equilibrium occurs. The order of $OH^-$ is around 1 in Reaction 2, suggesting $OH^-$ catalyzes the reaction.




Fitted parameters:
$k1_{pH11}$ = 1.177e+01
$k1_{pH10}$= 1.604e+00
$k2_{pH11}$ = 5.873e+00
$k2_{pH10}$ = 1.827e+00
gamma1 = 0.474 (order in E)
gamma2 = 1.165 (order in E)

Separate 1st rxn const.
Fitted parameters:
$k1_{pH11}$ = 1.071e+01
$k1_{pH10}$ = 1.414e+00
k2 = 1.965e+02
gamma1 = 0.461 (order in E)
gamma2 = 1.674 (order in E)

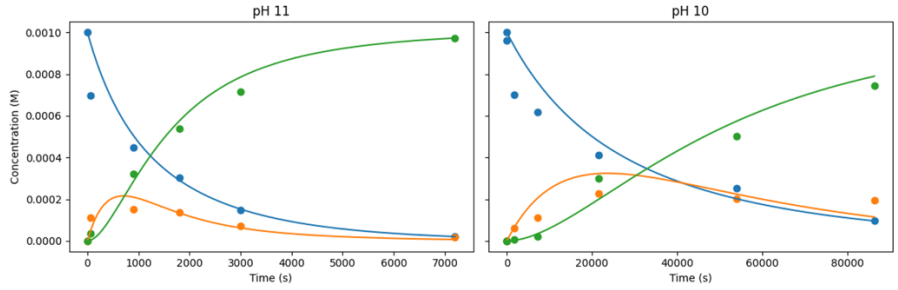
Separate 2nd rxn const.
Fitted parameters:
k1 = 4.652e+03
$k2_{pH11}$= 3.838e+00
$k2_{pH10}$ = 1.031e+00
gamma1 = 1.340 (order in E)
gamma2 = 1.104 (order in E)

**Figure S17.** Final fitting of the GSH hydrolysis reaction (top) and the proposed mechanism (bot). The reaction was written as

$$A + B + 0.5\,E \rightarrow C + 2D + 0.5\,E$$

$$C + E \rightarrow PN + PC$$

The coefficients of $OH^-$ were set as 0.5 and 1 for Reaction 1 and 2 respectively. The rate constants were separately fitted to pH 10 and pH 11 datasets. To check the result, we can compare the k values under both pH conditions. The difference of k2 is 4.5-fold, and the difference of k1 is 7-fold. While considering that the rate-limiting step in the Reaction 1 should be the addition of GSH (conjugate base) to **1**, the difference of the reaction rate constants is 2.5-fold by introducing the term $k_{1,eff} = k_1 \times [OH^-]^{0.5}$. The difference is small enough to validate the deduction, and the proposed mechanism was shown at the bottom.




Fitted parameters:
$k1_{pH11}$ = 1.401e+01
$k1_{pH10}$ = 2.021e+00
$k2_{pH11}$ = 1.864e+00
$k2_{pH10}$ = 4.041e-01

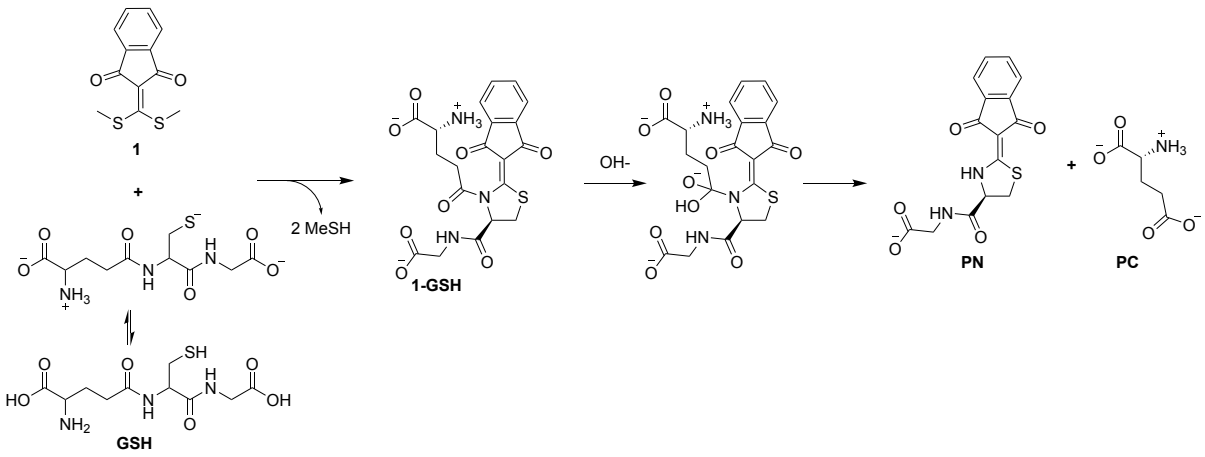


**Figure S18.** MALDI-MS of fragment B of mTE. Asterisk stands for modification by **1**.

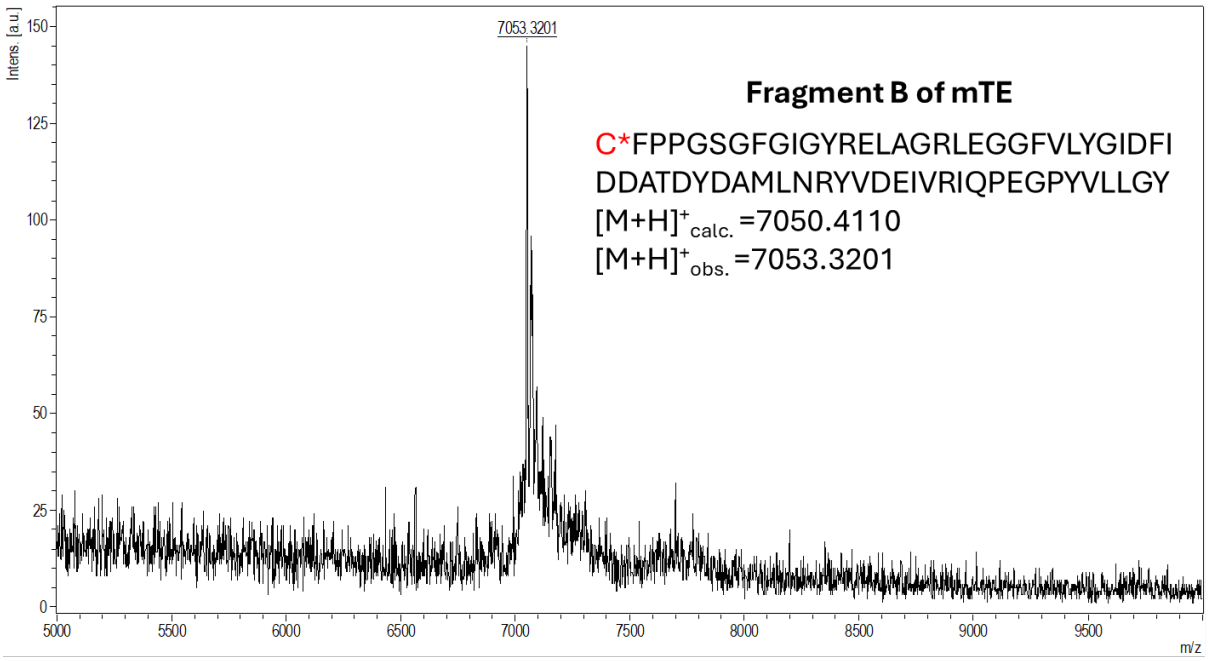


**Figure S19.** Sequence of mTE determined by MS analysis. Found fragments by LC-MS/MS were underlined. Arg and Lys residues were highlighted in blue, and Cys residues were marked in red. Fragments containing Cys that were detected were marked in yellow. CFPPGSGFGIGYR and CFGGNLTFEVAK were both labeled by **1**. EGDLNVF was also found in the search, suggesting complete cleavage at the following N-terminal Cys.

MGSSHHHHHHSSGLVPRGSHMASMTGGQQVLPAPEDAATGGTEYVAPRTLLEMKIARVW
QDTLGVRQVGVKDNFFELGGNSLSLIRLVQAVYDETGIEIPLNRQFHHVTVEAMAFGEGDL
GLDKGGDSFIKLNKEGDLNVFCFPPGSGFGIGYRELAGRLEGGFVLYGIDFIDDATDYDAML
NRYVDEIVRIQPEGPYVLLGYCFGGNLTFEVAKTMERRGYSVTDLVMVDSWIKDTLTPYET
SEKELEEMLADFDEEEKELMSNPLVRERVHRKIKATLTYEAQLINSGTITARIYELIAKDSEAF
RLEHQLPSWQRATTQAYADYRLEGAHEELLELARVDETAVVIRDILVQIKQQIEAEAGVLHG
S

Sequence coverage: 74%

**Figure S20.** LCMS spectra of fragments C*FPPGSGFGIGYR and C*FGGNLTFEVAK. Asterisk stands for modification by **1**.

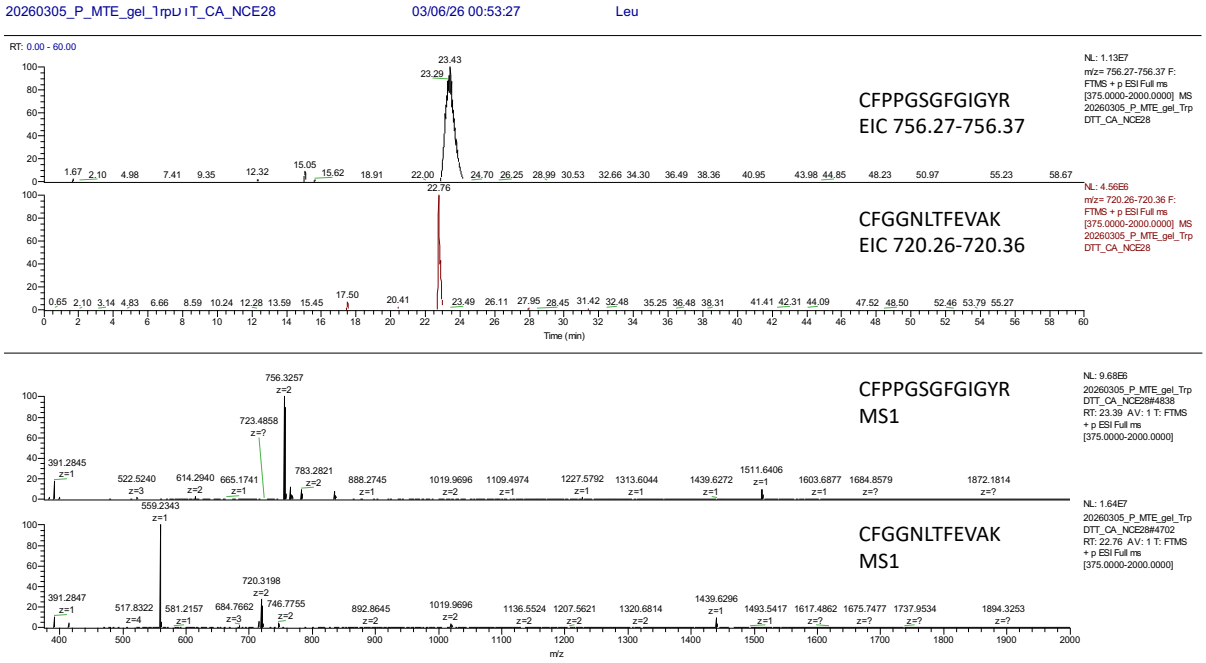


**Figure S21.** MS/MS profile of mTE fragments A) C*FPPGSGFGIGYR and B) C*FGGNLTFEVAK. y-ions and b-ions were noted in blue and red respectively. Asterisk stands for modification by **1**.

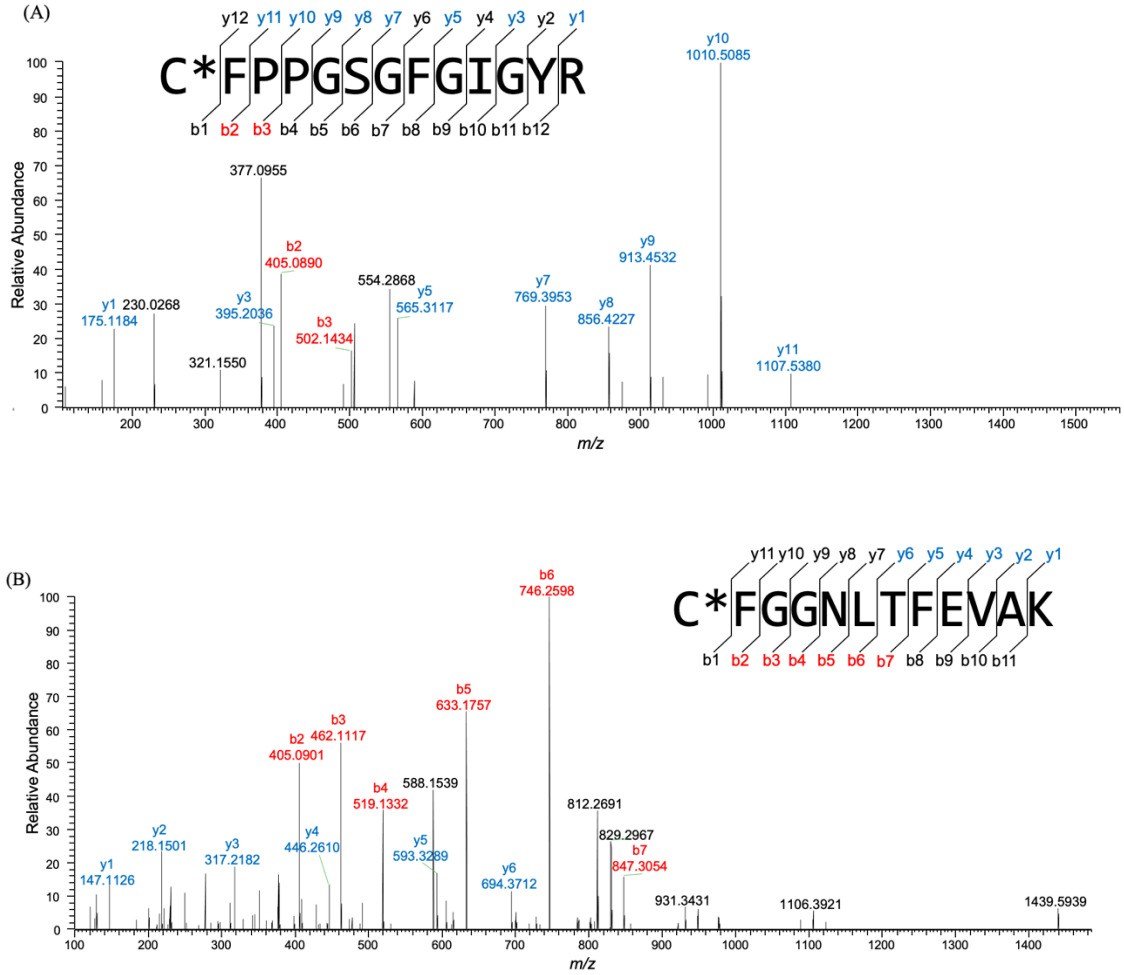


**Figure S22.** Stability test of **1**. **1** (20 mM) was dissolved in an aqueous DMSO-$d_6$ solution (5% v/v) and left at 20 ºC for one week and showed no detectable degradation on $^1$H NMR.

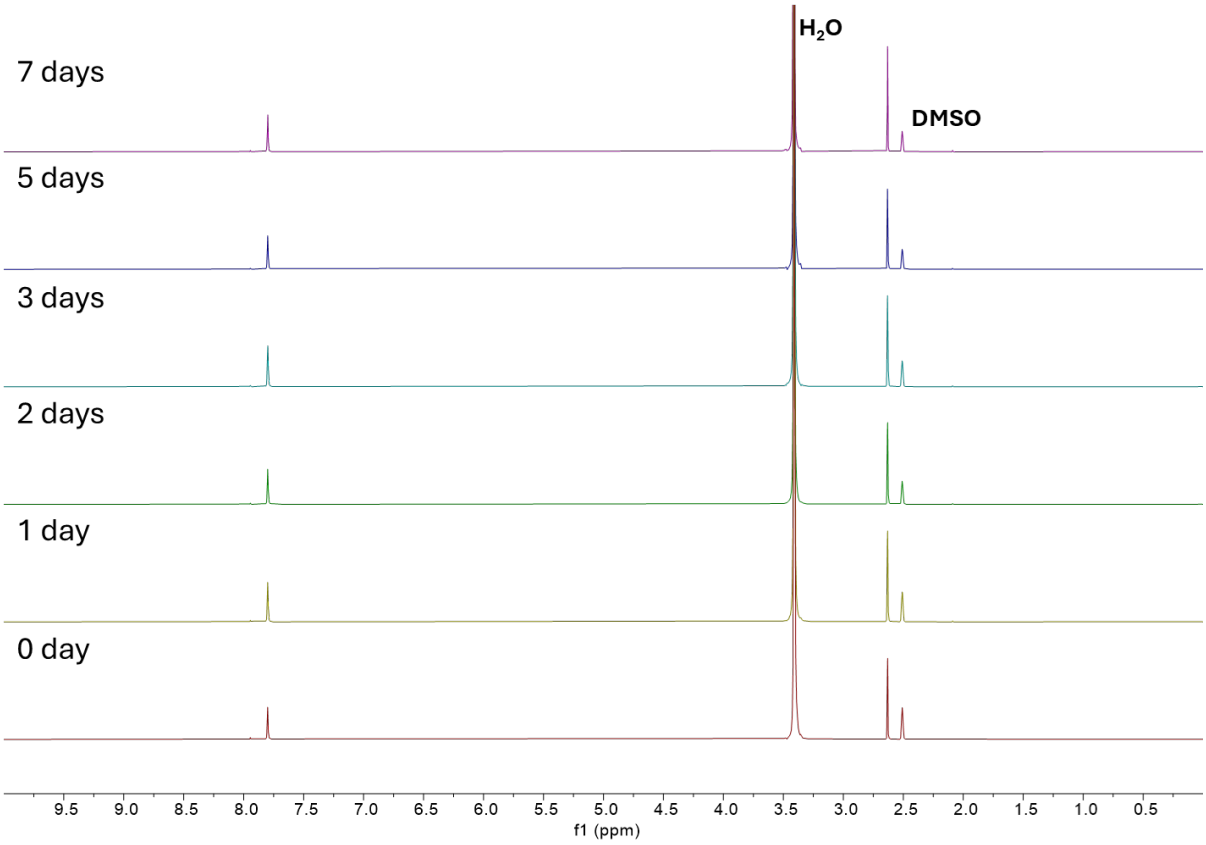
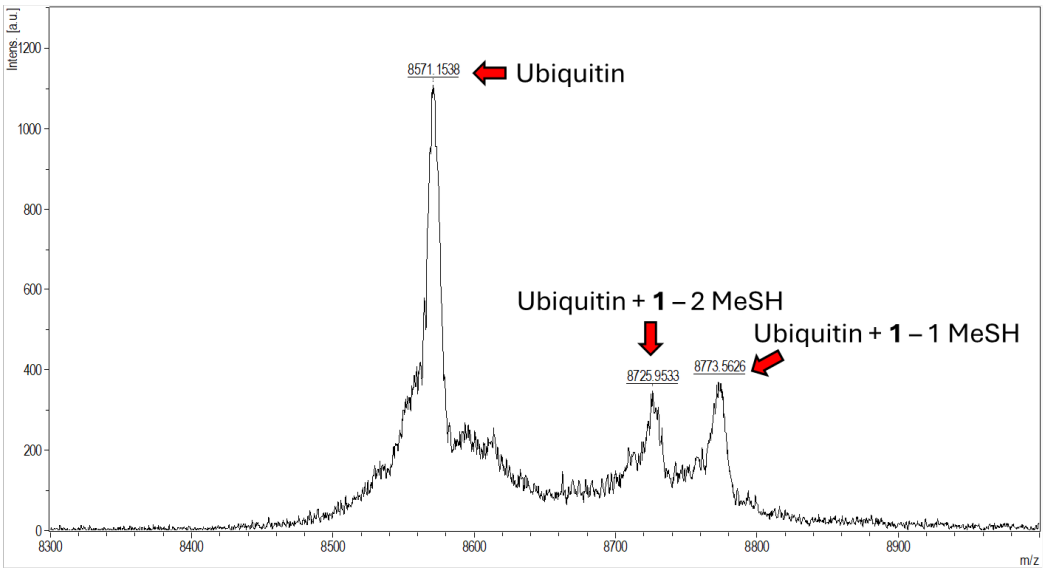


**Figure S23.** Cysteine-Specific test of **1**. **a)** To evaluate the selectivity of reagent **1** toward cysteine, a cysteine-free protein was examined. Ubiquitin (10 $\mu$M) was pre-incubated in CABS buffer (pH 11.0) containing 8 M urea, followed by addition of reagent **1** (100 $\mu$M, from a 10 mM DMSO stock). The reaction was incubated at 25 °C for 24 h prior to MALDI analysis. A mass shift corresponding to reagent **1** addition was observed, indicating off-target reactivity at a specific functional group. **b)** Ubiquitin contains eight nucleophilic amines, including seven lysine side chains and one N-terminal amine, all of which could potentially react with reagent **1**. To determine the reactive site, $H_2N$–Lys–OH and Boc–Lys–OH (10 $\mu$M each) were separately incubated with reagent **1** (10 $\mu$M) in CABS buffer (pH 11.0) containing 8 M urea at 25 °C for 12 h. The reactions were subsequently analyzed by UV–vis spectroscopy. A decrease in absorption at 400 nm was observed only for $H_2N$–Lys–OH, whereas no significant change was detected for Boc–Lys–OH. These results indicate that the reaction occurs selectively at the N-terminal amine rather than at the ε-amino group of lysine.

a)




b)

| Amino acid | # | Amino acid | # |
|---|---|---|---|
| A | 2 | M | 1 (N-term) |
| C | 0 | N | 2 |
| D | 5 | P | 3 |
| E | 6 | Q | 6 |
| F | 2 | R | 4 |
| G | 6 | S | 3 |
| H | 1 | T | 7 |
| I | 7 | V | 4 |
| K | 7 | W | 0 |
| L | 9 | Y | 1 |

Ub sequence:
MQIFVKTLTGKTITLEVEPSDTIENVKAKIQ
DKEGIPPDQQRLIFAGKQLEDGRTLSDYN
IQKESTLHLVLRLRGG

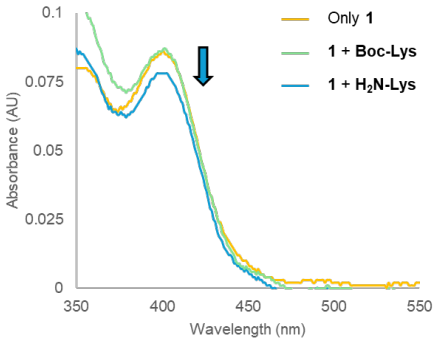


**Figure S24.** Simulation of six proteomic digestion modes on random 3000 human proteins from UniProt. It evaluates fragmentation patterns by combining reagent **1** with trypsin and chymotrypsin. **a)** Fragments are classified into three groups: short peptides (<7 AA, gray), large peptides (>20 AA, blue), and bottom-up standard peptides (7-20 AA, yellow). **b)** The green bars represent the ratio of bottom-up standard peptides. The orange bars represent the residue coverage, indicating the total proportion of MS-accessible residues.

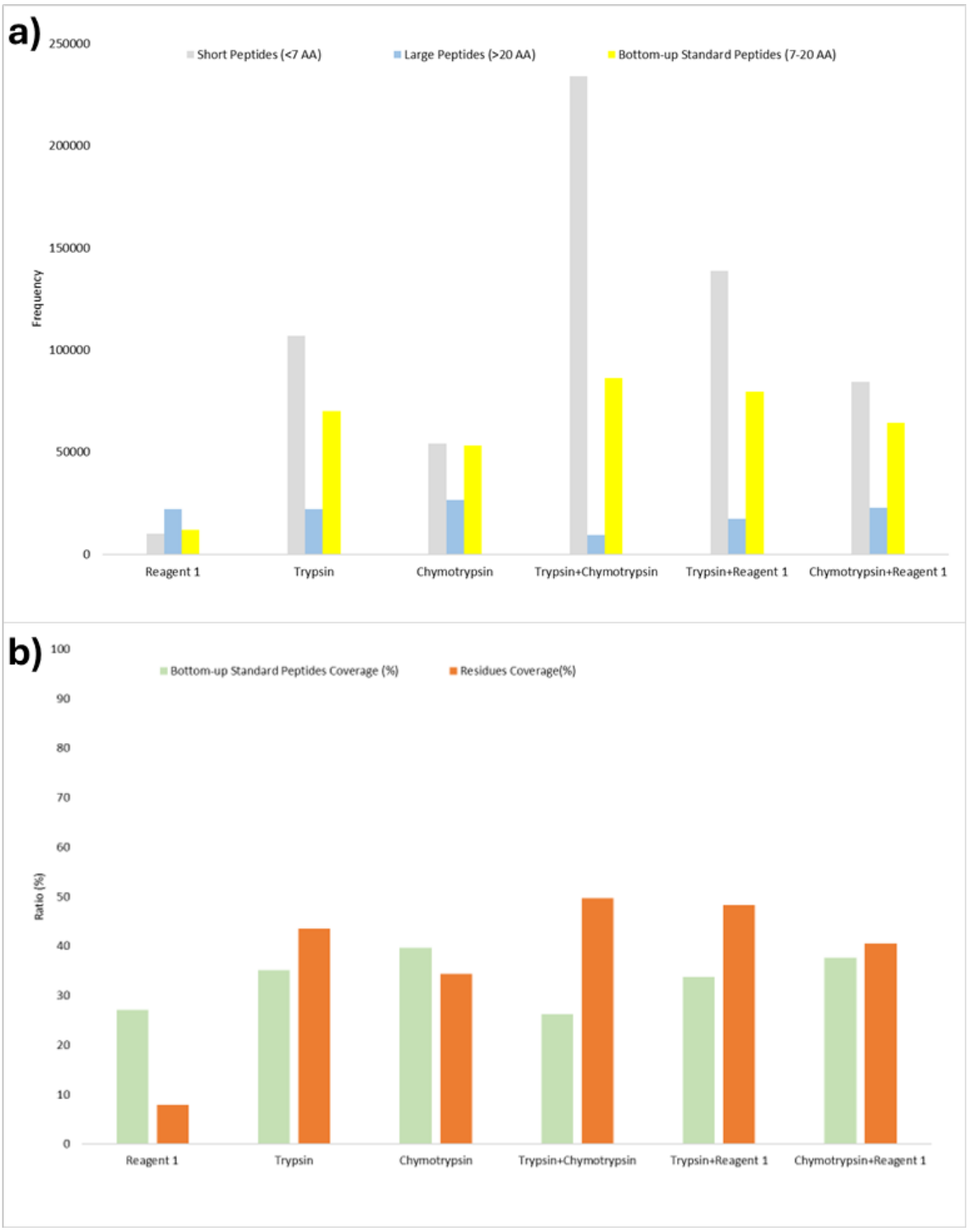
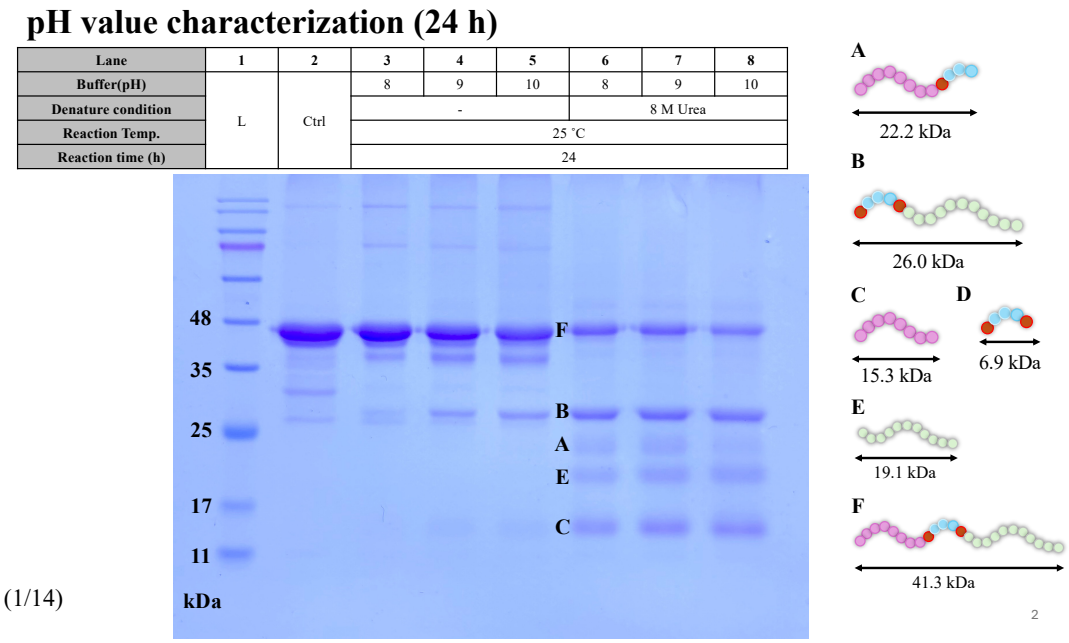


**Figure S25.** Original gel images corresponding to Figure 3B



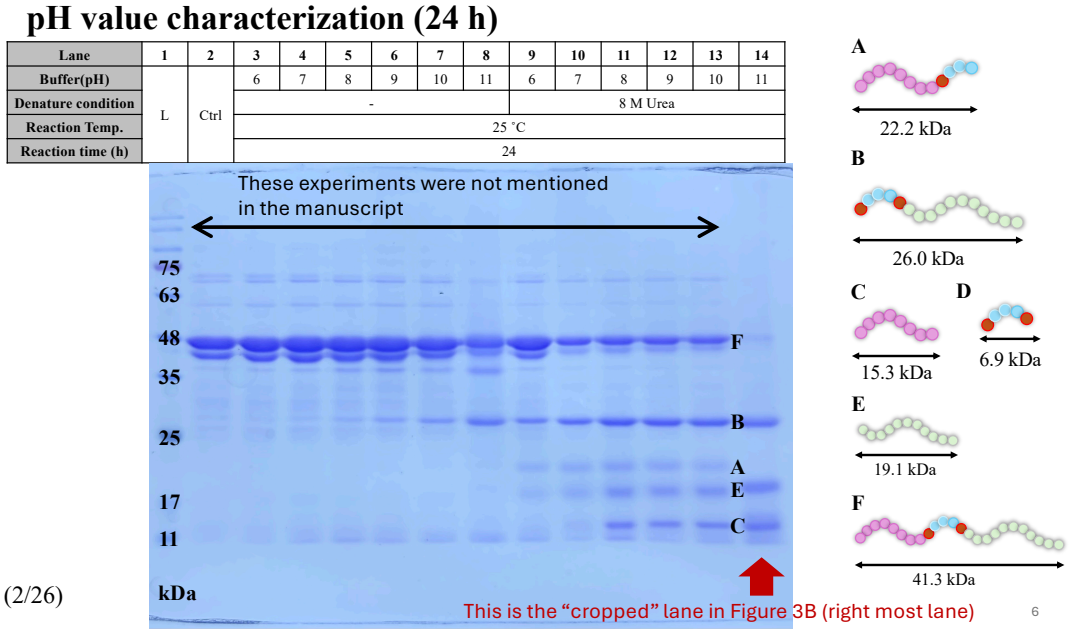
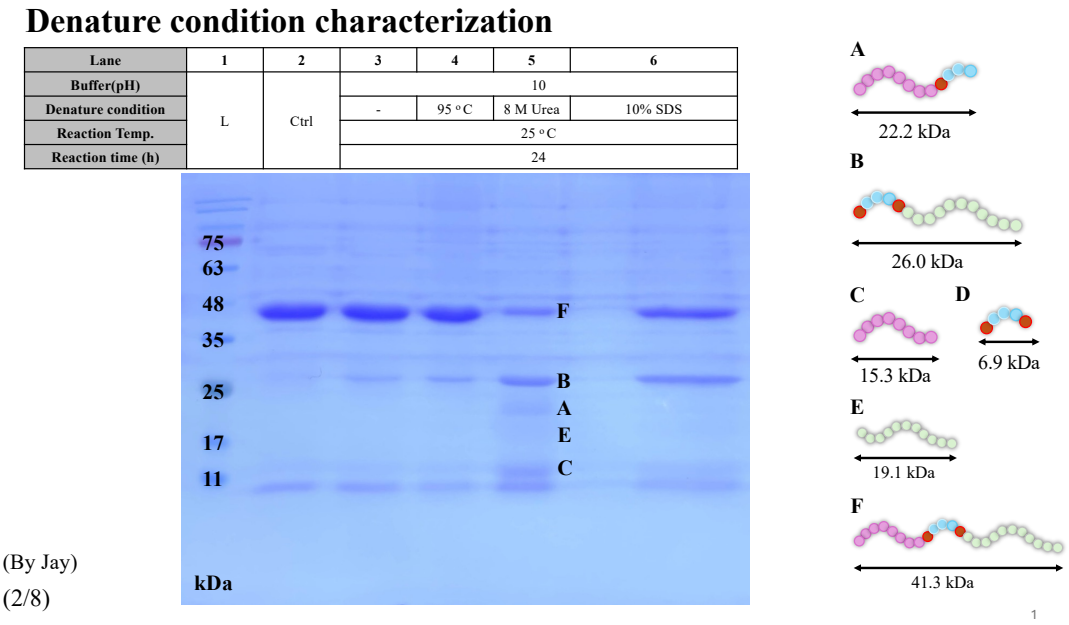



**Figure S26.** Original gel image corresponding to Figure 3C



**Reference**

1. Sinitcyn, P.; Hamzeiy, H.; Salinas Soto, F.; Itzhak, D.; McCarthy, F.; Wichmann, C.; Steger, M.; Ohmayer, U.; Distler, U.; Kaspar-Schoenefeld, S.; et al. MaxDIA enables library-based and library-free data-independent acquisition proteomics. *Nat. Biotechnol.* **2021**, *39*, 1563–1573.