

Oxidation-Induced Generation of A Mild Electrophile for A Proximity-Enhanced
Protein-Protein Crosslinking

Electronic Supporting Information

I. General material and method.....	2
II. Experimental procedures.....	3
III. Supplemental table.....	8
IV. Supplemental figures	
1. Figure S1.....	9
2. Figure S2.....	10
3. Figure S3.....	11
4. Figure S4.....	12
5. Figure S5.....	13
6. Figure S6.....	14
7. Figure S7.....	15
8. Figure S8.....	16
9. Figure S9.....	17
10. Figure S10.....	18
V. References.....	19

I. General material and method

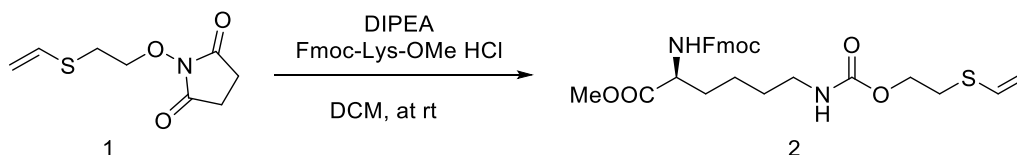
Unless otherwise noted, starting materials, solvents and reagents for chemical synthesis were obtained from commercial suppliers (Acros, Alfa Aesar, Sigma-Aldrich, Chem-impex) and used without further purification. Dry solvents were either freshly distilled by following standard methods or directly purchased from Acros. Deuterated solvents were obtained from Sigma-Aldrich. Flash chromatography (FC) was carried out using SiliaFlash P60 (0.04–0.063 mm, 230–400 mesh) from Silicycle. Thin layer chromatography (TLC) was performed on glass-backed, precoated silica gel plates (Anatech.). NMR spectra were recorded at 25 °C using a Bruker Advance III-HD 400 MHz NMR. Chemical shifts were reported in ppm with deuterated solvents as internal standards (CDCl_3 , H 7.26, C 77.0; D_2O , 4.79). Multiplicity was reported as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, b = broad. Kinetic studies by HPLC was operated on Agilent 1260 Infinity. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed on Bio-Rad mini-PROTEAN electrophoresis system using 15% or 18% homemade SDS-PAGE gels. Bio-Rad Prestained Protein Ladder was applied to at least one lane of each gel for estimation of apparent molecular weights. Protein bands was stained by homemade coomassie brilliant blue staining solution and visualized under Bio-rad Molecular Imager ChemiDoc XRS+ System. For in-gel fluorescence imaging of labeled protein bands, Bio-rad Molecular Imager ChemiDoc XRS+ System was used. Fluorescence intensity and characterization was performed on BioTek Synergy H1 Hybrid Multi-mode Monochromater Fluorescence Microplate reader and Horiba FluoroMax 4 spectrometer. Live cells were imaged on Olympus FV500 inverted (Olympus IX-81) confocal microscope.

II. Experimental procedures

Synthetic procedures

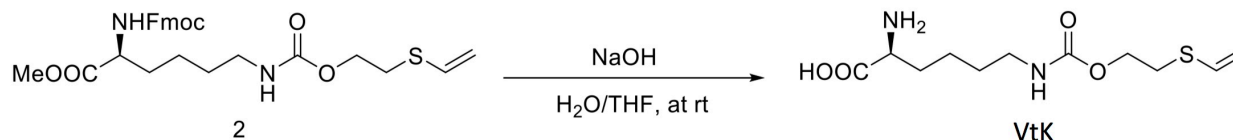
Vinyl sulfide **1**^[1] and *N*^α-Fmoc-L-lysine methyl ester hydrochloride^[2] were obtained by following literature procedures.

N^α-(9-Fluorenylmethoxycarbonyl)-*N*^ε-((2-(vinylthio)ethoxy)carbonyl)-L-lysine methyl ester (**2**)



To the suspension of vinyl sulfide (**1**) (275 mg, 1.12 mmol) and *N*^α-Fmoc-L-lysine methyl ester hydrochloride (612 mg, 1.46 mmol) in 11 mL of dichloromethane (DCM) was added *N,N*-Diisopropylethylamine (DIPEA, 215 μL, 1.24 mmol) with stirring and then the suspension became clear solution. The reaction mixture was continued to be stirred at room temperature. After being stirred for 24 hours, white solid was formed and filtered off. The filtrate was concentrated and further purified by flash chromatography with gradient eluant hexane/ethyl acetate from 5/1 to 2/1. 645 mg (99%) of desired product (**2**) was obtained. ¹H-NMR (300MHz, CDCl₃) δ ppm: 7.77 (d, *J* = 7.5 Hz, 2H), 7.60 (d, *J* = 6.0 Hz, 2H), 7.40 (t, *J* = 7.4 Hz, 2H), 7.32 (t, *J* = 7.3 Hz, 2H), 6.30 (dd, *J* = 10.1, 16.7 Hz, 1H), 5.44 (d, *J* = 8.0 Hz, 1H), 5.19 (m, 2H), 4.82 (br, 1H), 4.42-4.38 (m, 3H), 4.23-4.15 (m, 3H), 3.75 (s, 3H), 3.16 (m, 2H), 2.89 (t, *J* = 6.7 Hz, 2H), 1.90-1.30 (m, 6H); ¹³C-NMR (75 MHz, CDCl₃) δ ppm: 172.87, 156.19, 155.97, 143.82, 143.68, 141.25, 131.24, 127.68, 127.02, 125.03, 119.96, 111.74, 66.96, 62.90, 53.53, 52.44, 47.10, 40.45, 32.11, 30.31, 29.24, 22.18. HRMS (ESI) calcd for C₂₇H₃₂N₂O₆S, [M + H]⁺ 513.2059, found 513.2048.

N^ε-((2-(vinylthio)ethoxy)carbonyl)-L-lysine (VtK)



To the solution of compound (2) (562 mg, 1.10 mmol) in 5.5 mL of THF was added 5.5 mL of 1 N NaOH aqueous solution. The mixture was stirred at room temperature overnight. The resulting mixture was diluted with 10 mL of water and washed with ethyl acetate (3×5 mL). The aqueous layer was adjusted to pH 6-7 by addition of 1 N HCl to yield white precipitates. The solid (245 mg, 81%) was collected by filtration to yield VtK. ¹H-NMR (400MHz, D₂O, 0.1 N NaOH) δ ppm: 6.37 (dd, *J* = 10.2, 16.9 Hz, 1H), 5.22 (dd, *J* = 10.4, 17.3 Hz, 2H), 4.18 (t, *J* = 6.3 Hz, 2H), 3.15 (t, *J* = 6.6 Hz, 1H), 3.05 (t, *J* = 7.1 Hz, 2H), 2.94 (t, *J* = 6.1 Hz, 2H), 1.60-1.20 (m, 6H); ¹³C-NMR (100 MHz, D₂O, 0.1 N NaOH) δ ppm: 183.77, 158.44, 131.28, 112.59, 63.55, 56.06, 40.39, 34.46, 30.29, 28.92, 22.37. HRMS (ESI) calcd for C₁₁H₂₀N₂O₄S, [M + H]⁺ 277.1222, found 277.1214.

Plasmid construction

pGST-WT: GST-encoding gene was PCR amplified from pGEX vector using primers P1 and P2. The DNA fragment was digested with *Nde*I and *Sac*I, and ligated into pLei-GFP vector that was treated with the same restriction enzymes to afford pGST-WT. Plasmid pGST-WT was confirmed by DNA sequencing.

P1: 5'- CGCCATATGTCCCCTATACTAGGTTA -3'

P2: 5'- ATAGAGCTCTTTTGGAGGATGGTCGCC -3'

pGST mutants: The indicated mutations were introduced by overlapping PCR using pGST-WT plasmid as the template. The digested PCR product was inserted into pLei vector behind a T5 promoter to afford the desired plasmids. Following were primers used in the construction:

P3: 5'-CTTATTATATTGATGGTTAGGTTAAATTAACACAGTC-3'

P4: 5'-GACTGTGTTAATTTAACCTAACCATCAATATAATAAG-3'

P5: 5'-TTATATTGATGGTGATTAGAAATTAACACAGTCTA-3'

P6: 5'-TAGACTGTGTTAATTTCTAATCACCATCAATATAA-3'

P7: 5'-ATAGACTGTGTTAACTGAACCTAACCATCAA-3'

P8: 5'-TTGATGGTTAGGTTTCAGTTAACACAGTCTAT-3'

P9: 5'-TAGACTGTGTTAACTGCTAATCACCATCA-3'

P10: 5'-TGATGGTGATTAGCAGTTAACACAGTCTAT-3'

P11: 5'-TGGACAACCACCCAAC-3'

P12: 5'-GTGGTTGTCCACAGGAGCGTGCAGAGATTTC-3'

P13: 5'-GTTGGGTGGTGCCCCAAAAGAGCGTGCAG-3'

P14: 5'-GCTCTTTTGGGGCACCACCCAACATGTT-3'

PyIRS library construction

A PyIRS (from *Methanosarcina barkeri*) library was created, in which residues Leu270, Tyr271, Leu274, and Cys313 were completely randomized. Overlapping polymerase chain reaction (PCR) was performed with synthetic oligonucleotide primers in which the randomized residues were encoded as NNK (N=A, C, T, or G, K=T or G) to generate a library with a theoretical diversity of 1.05×10^6 . The quality of the library (>99% coverage) was validated by DNA sequencing. The following primers were used in the library construction:

P15: 5'-GGAATTCCATATGGATAAAAAACCATTAGATG-3'

P16: 5'-AGTCGGGGCAAGCATTGGC-3'

P17: 5'-GCCAATGCTTGCCCCGACTNNKNNKAACTATNNKCGAAAACCTCGATAGGATTTTA-3'

P18: 5'-GAAGTTCACCATAGTAAATTC-3'

P19: 5'-GAATTTACTATGGTGAACCTC**NNK**CAGATGGGTTCTGGGATGTAC-3'

P20: 5'-AACTGCAGTTATAGATTGGTTGAAATCCC

Screening of aminoacyl-tRNA synthetase variants for the genetic incorporation of ViP.

The resulting PyIRS library was transformed into *E. coli* GeneHogs electrocompetent cells subjected to a negative selection to remove PyIRS variants that could charge tRNA^{PyI} with natural amino acid as previously described, followed by a positive selection to identify functional PyIRS variants.^[3] Briefly, the negative selection uses the toxic barnase gene with amber

mutations at permissive sites (Gln2TAG and Asp44TAG) and was carried out in the absence of VtK. The positive selection is based on resistance to chloramphenicol (Cm), which is conferred by the suppression of an amber mutation at a permissive site (Asp112) in the chloramphenicol acetyltransferase-encoding gene in the presence of tRNA^{Pyl}, VtK, and functional PylRS mutants. The surviving PylRS variants were subsequently screened for chloramphenicol resistance level in the presence and absence of VtK. A few clones that survived on 75 µg/mL chloramphenicol in the presence of VtK and did not grow on 34 µg/mL chloramphenicol in the absence of VtK were identified.

Protein expression and purification.

E. coli GeneHogs strain harboring plasmid pBK-VtK and pBAD-Mb-Lys99TAG^[4] was cultured in 50 mL LB media containing Kan (50 mg/L) and Tet (12.5 mg/L) at 37 °C with shaking. The protein expression was induced at OD₆₀₀ of 0.6 by the additions of arabinose (0.01%) and VtK (1 mM). A control contained no VtK was conducted under the same conditions. Following an additional 16 h of cultivation, cells were collected by centrifugation at 5,000g and 4 °C for 15 min. Harvested cells were resuspended in the buffer containing potassium phosphate (20 mM, pH 7.4), NaCl (150 mM), and imidazole (10 mM) and then subsequently lysed by sonication. Cellular debris was removed by centrifugation (21,000g, 30 min, 4 °C). The cell-free lysate was applied to Ni Sepharose 6 Fast Flow resin (GE Healthcare) and the protein purification was carried out by following manufacturer's instructions. Protein concentrations were determined by Bradford assay (Bio-Rad). Purified protein was desalted prior to MS analysis. Similar procedures were applied to the purification of sfGFP-Asp149VtK and GST-VtK variants.

Protein mass spectrometry.

The sample was analyzed by a Q-Exactive HF mass spectrometer under the positive mode.

GST crosslinking using purified proteins

Purified GST variants (1 mg/mL) in 100 mM of sodium phosphate buffer (pH 7.0) were treated with 5 mM H₂O₂ in the presence of Na₂WO₄ for 30 minutes to 90 minutes at 37 °C. TCEP (5 mM) was subsequently added to quench the oxidant. The resulting protein mixture was further

incubated at 37 °C for 3 hours and then quenched by Laemmli sample buffer (contains β -mercaptoethanol) for SDS-PAGE gel analysis.

GST crosslinking using cell lysate

E. coli cells (0.7 mL to 1 mL) expressing either wild-type GST or GST-VtK mutant proteins were harvested by centrifugation (21,000 g, 5 min, 4 °C). Collected cells were washed once in the same volume of sodium phosphate buffer (100 mM, pH 7.0) by vortexing for 5 min and resuspended in 0.1 mL of the phosphate buffer. The cells were subsequently lysed by sonication. The cellular debris was removed by centrifugation (16,000g, 20 min, 4 °C) and the supernatant was treated with 50 mM H₂O₂ in the presence of Na₂WO₄ for 60 minutes. TCEP (50 mM) was added to remove the excess H₂O₂. After additional incubation of the treated cell lysate at 37 °C for 2-4 hours, the cell lysates were analyzed by SDS-PAGE gel and western blotting.

III. Supplemental table

Table S1. The sequence of three hits from the directed evolution.

MbPyIRS	Leu270	Tyr271	Leu274	Cys313
VtKRS-1	Leu270	Tyr271	Ser274	Val313
VtKRS-2	Leu270	Tyr271	Thr274	Thr313
VtKRS-3	Leu270	Tyr271	Phe274	Ala313

IV. Supplemental figures

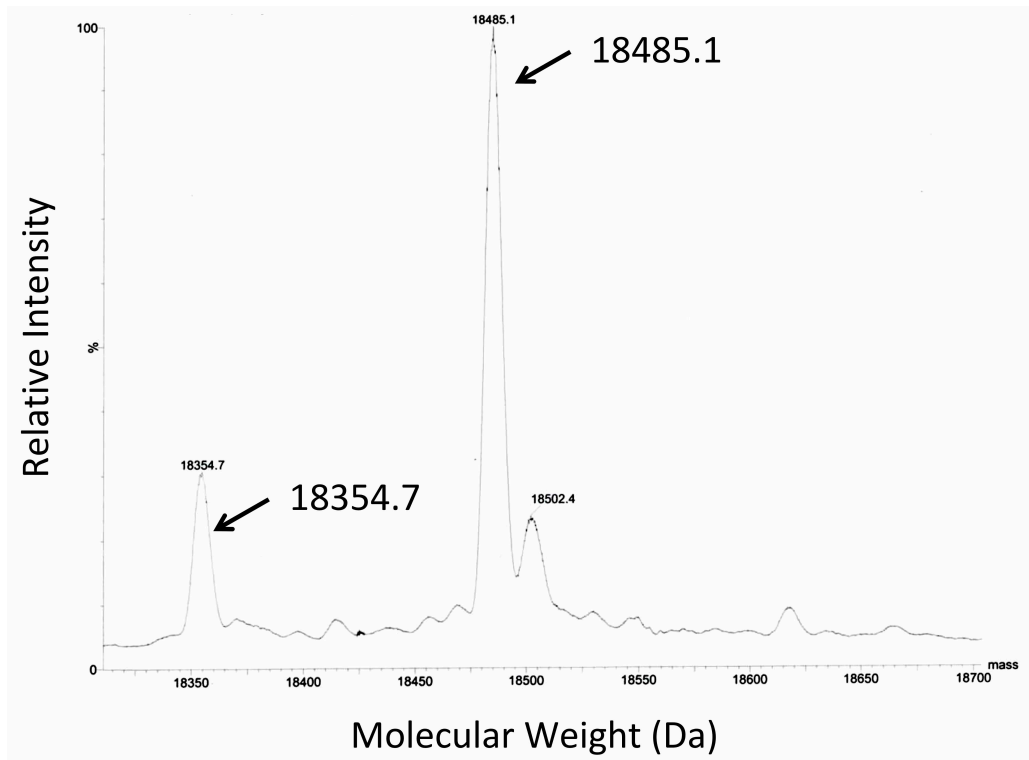


Fig. S1 Deconvoluted ESI-MS spectra of the Mb-Lys99VtK mutant. Expected masses: 18485.1 Da (with the N-terminal Met) and 18353.9 (without the N-terminal Met); observed masses: 18485.1 Da (with the N-terminal Met) and 18354.7 Da (without the N-terminal Met). The other signals do not correspond to sfGFP mutant with any proteinogenic amino acids at position Lys99.

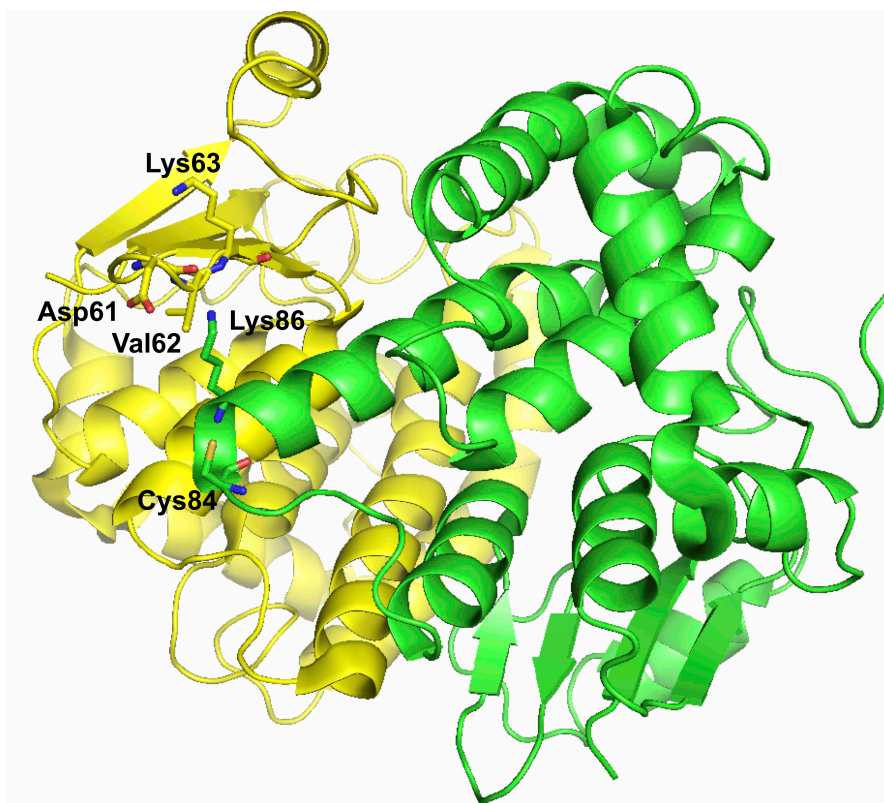


Fig. S2 Crystal structure of GST dimer (PDB: 1Y6E).

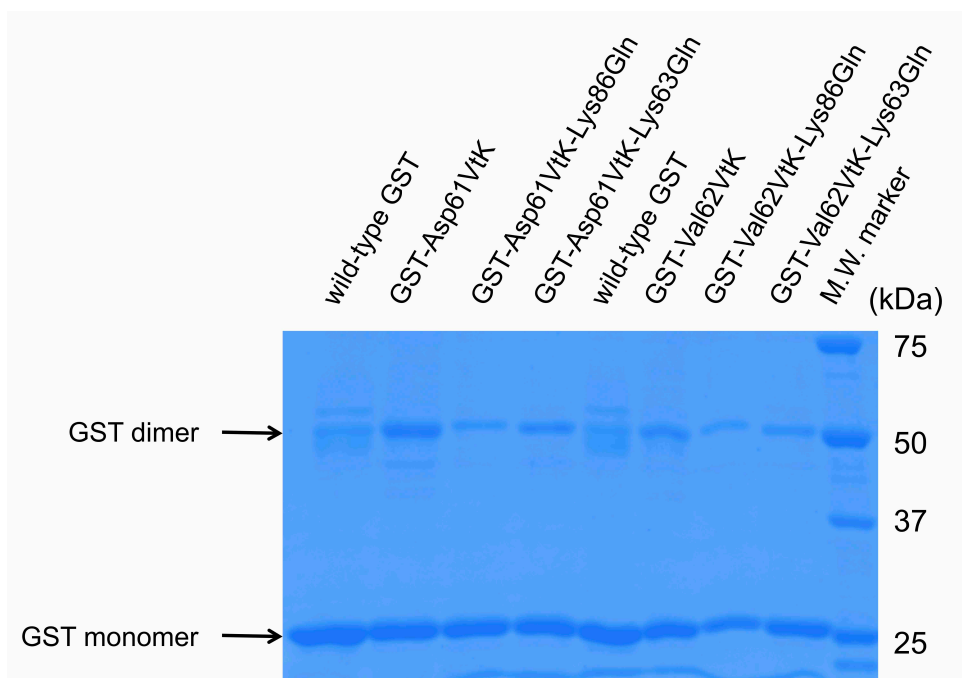


Fig. S3 SDS-PAGE analysis of crosslinking of GST variants. All GST samples were incubated in the presence of 5 mM H_2O_2 and 0.5 mM Na_2WO_4 at 37 °C for 60 min. TCEP (10 mM) was added after the oxidation reaction before SDS-PAGE analysis.

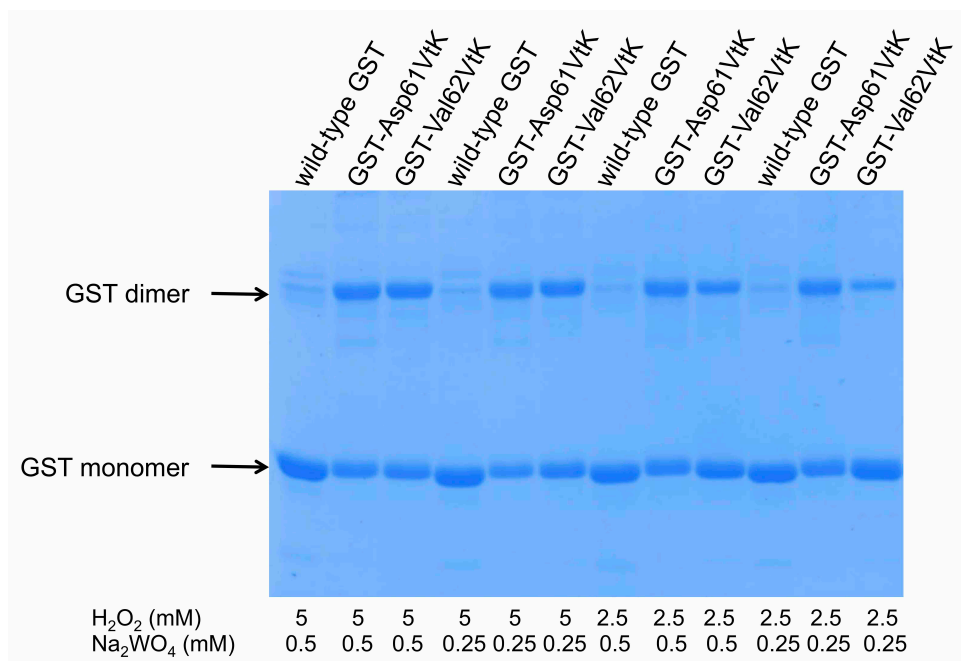


Fig. S4 SDS-PAGE analysis of crosslinking of GST variants in the presence of different concentrations of H₂O₂ and Na₂WO₄.

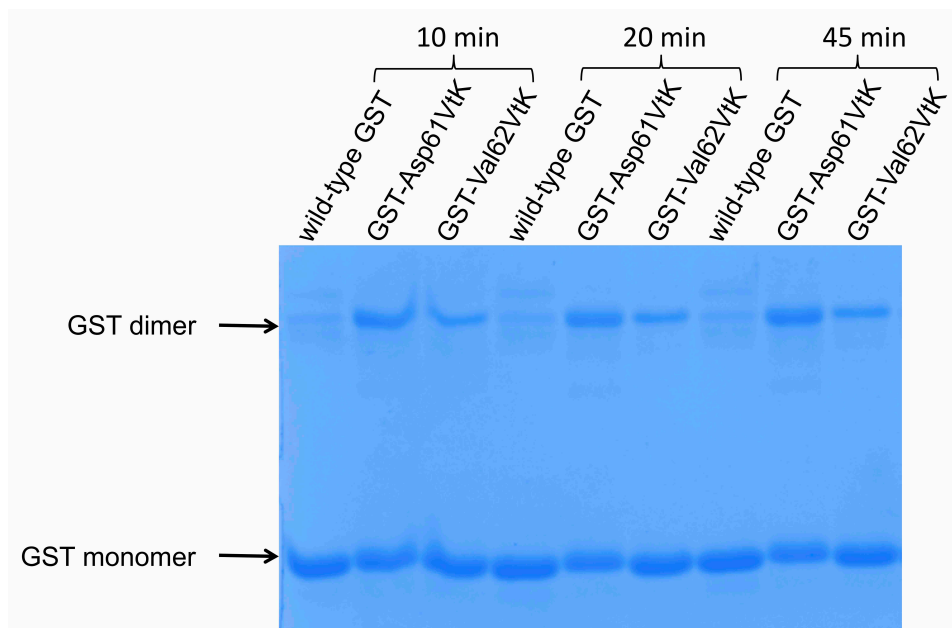


Fig. S5 SDS-PAGE analysis of crosslinking of GST variants with different reaction time. The concentrations of the reagents are: 2.5 mM H₂O₂ and 0.25 mM Na₂WO₄.

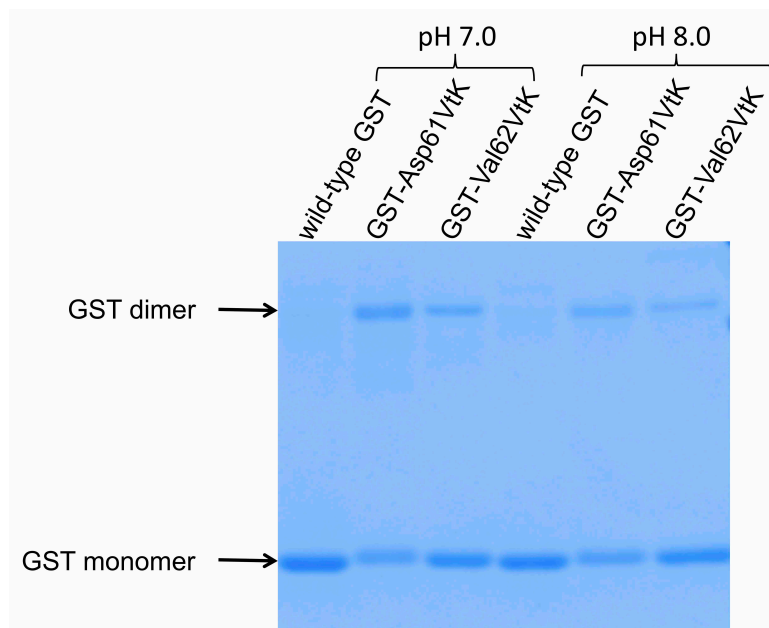


Fig. S6 SDS-PAGE analysis of crosslinking of the GST variants at different pH values. The concentrations of the reagents are: 2.5 mM H_2O_2 and 0.25 mM Na_2WO_4 .

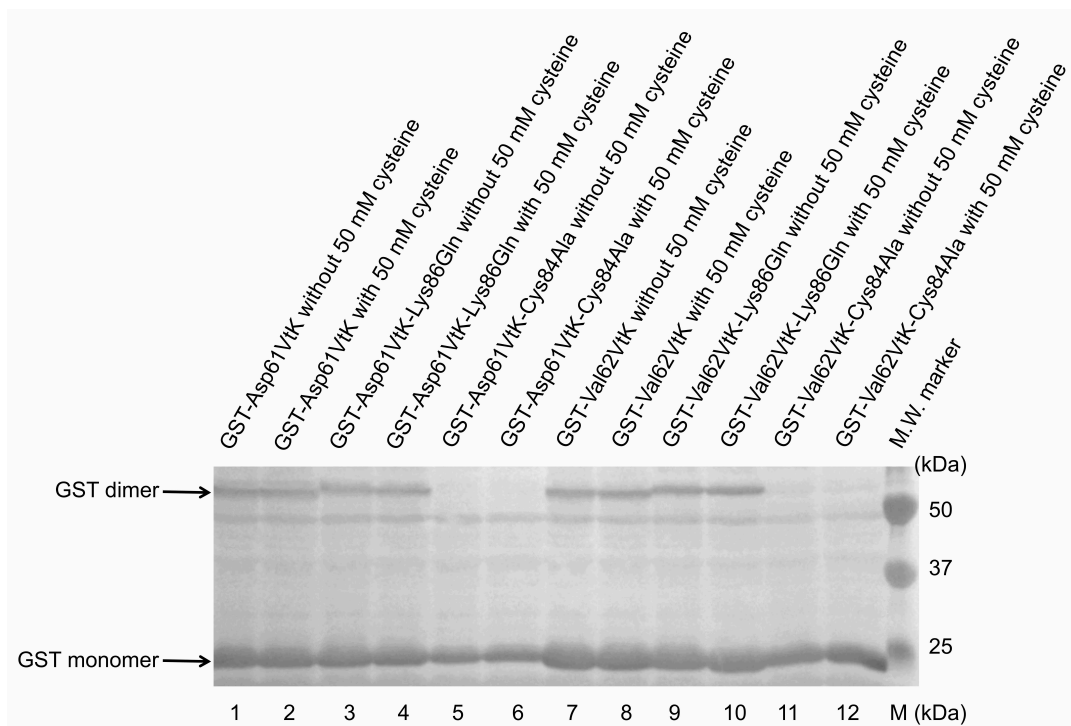


Fig. S7 Crosslinking of GST dimer in cell lysate. His-tagged GST variants were detected by Western blot using anti-His antibody. All GST samples (as cell lysate) were incubated in the presence of 50 mM H_2O_2 and 0.25 mM Na_2WO_4 at 37 °C. TCEP (50 mM) was added after the oxidation reaction.

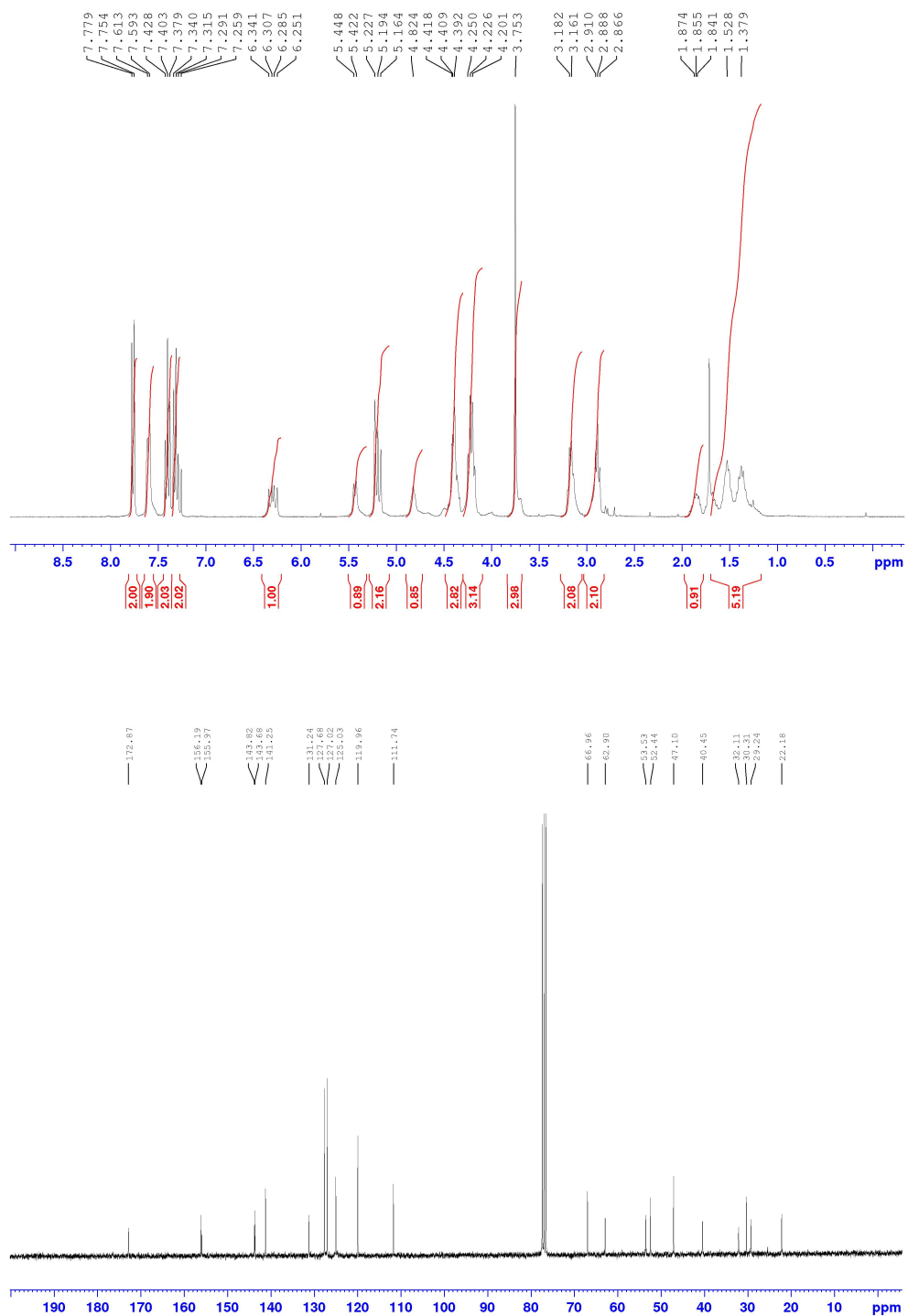


Fig. S8 ¹H and ¹³C NMR spectra of *N*^α-(9-Fluorenylmethoxycarbonyl)-*N*^ε-((2-(vinylthio)ethoxy)carbonyl)-*L*-lysine methyl ester (2).

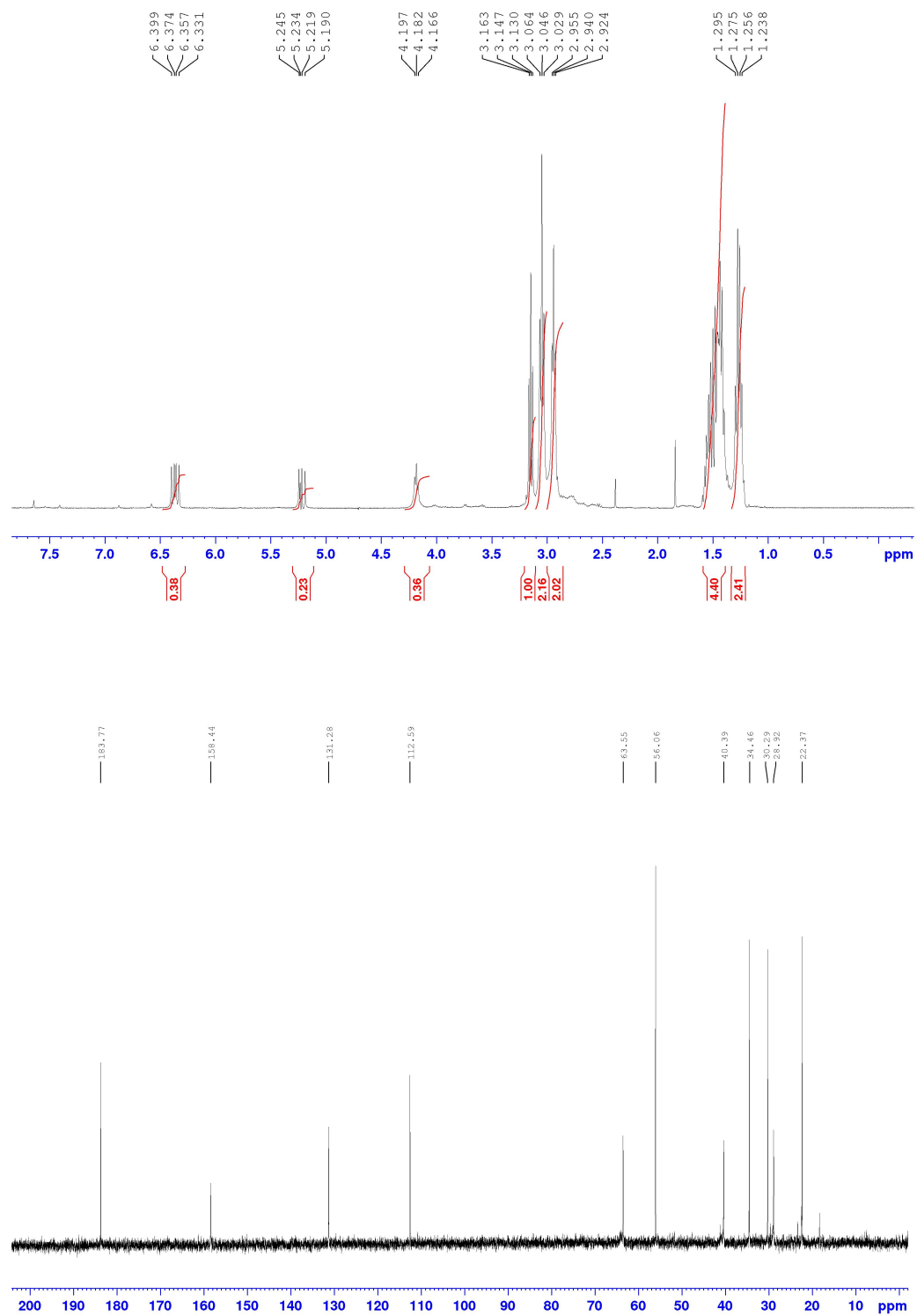


Fig. S9 ¹H and ¹³C NMR spectra of VtK.

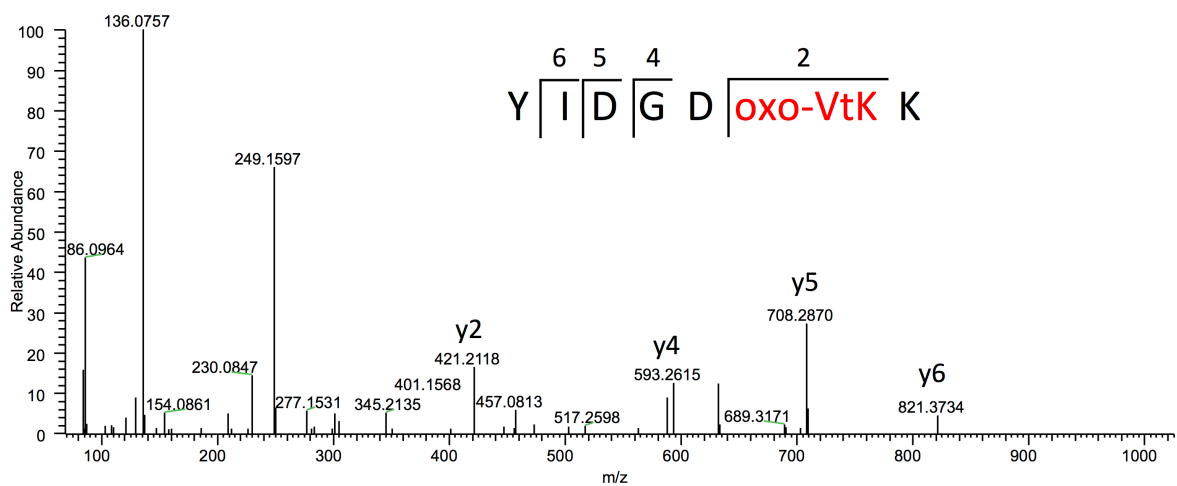


Fig. S10 Mass spectrometry analysis of GST-Val62VtK after a treatment with H₂O₂. The y ions are marked in the spectrum. The amino acid sequence of the peptide fragment, YIDGD-oxoVtK-K, from GST-Val62VtK is shown on top.

References

- [1] J. W. Hojfeldt, P. Blakskjaer, K. V. Gothelf, *The Journal of organic chemistry* **2006**, *71*, 9556-9559.
- [2] H. Pan, P. Kopeckova, J. Yang, J. Kopecek, Google Patents, **2011**.
- [3] K. Liu, B. Enns, B. Evans, N. Wang, X. Shang, W. Sittiwong, P. H. Dussault, J. Guo, *Chem. Commun.* **2017**, *53*, 10604-10607.
- [4] P. R. Chen, D. Groff, J. Guo, W. Ou, S. Cellitti, B. H. Geierstanger, P. G. Schultz, *Angew. Chem., Int. Ed.* **2009**, *48*, 4052-4055.