A Genetically Encoded Fluorosulfonyloxybenzoyl-L-lysine for Expansive Covalent Bonding of Proteins via SuFEx Chemistry

Jun Liu[1‡], Li Cao[1‡], Paul C. Klauser[1], Rujin Cheng[2], Viktoriya Y. Berdan[1], Wei Sun[1], Nanxi Wang[1], Farid Ghelichkhani[2], Bingchen Yu[1], Sharon Rozovsky[2], and Lei Wang[1*]

1. University of California San Francisco, Department of Pharmaceutical Chemistry and the Cardiovascular Research Institute, 555 Mission Bay Blvd. South, San Francisco, California 94158, United States
2. University of Delaware, Department of Chemistry and Biochemistry, Newark, Delaware 19716, United States

* E-mail: Lei.wang2@ucsf.edu
‡ These authors contributed equally to this work.

Table of Contents

Experimental Procedures:

Supplementary table 1: Primers used for cloning

Primer	Oligonucleotide Sequence (5'→3')
*Ma*PylRS NdeI	GTTGTTCATATGACCGTGAAGTACACCGACGCCCAG
126/129-Rev	CAGGTTTGGCGCCAGCATCGGGCG
126/129-For	CGCCCGATGCTGGCGCCAAACCTGNNKAGCGTGNNKCGTGACCTGCGCGATCACACCGAC
166-Rev	CAGCATGGTGAACTCCTCCAG
166/168-For	CTGGAGGAGTTCACCATGCTGNNKCTGNNKGATATGGGTCCGCGCGGCGAC
227-Rev	GACCCACAGCCGCGCTGCACACTTCCTGACCGTTAATTTCCACGTCGATG
227/228/229-For	GTGTGCAGCGCGGCTGTGGGTCCANNKNNKNNKGATGCGGCCCATGACGTGCATGAGCCG
*Ma*PylRS PstI	GGTGGtCTGCAGTTAATTGATTTTGGCACCATTCAGGTAG
sfGFP2TAG-For	GGTGGTCATATGTAGAAGGGCGAGGAGCTGTTCAC
sfGFP2TAG-Rev	GTTGTTAAGCTTTTAGTGATGGTGATGGTGATGGCTGCCCTTGTACAGCTCGTCCATG
HRpNEUmaRS-For	GAGACCCAAGCTGGCTAGCGCCACCATGACCGTGAAGTACACCGACGC
HRpNEUmaRS_Rev	GGCTGATCAGCGGGTTTAAAGCGGCCGCTTAATTGATTTTGGCACCATTC
ecGST NdeI	GTTGTTCATATGAAATTGTTCTACAAACCGGGTGCCTGC
ecGST HindIII	GTTGTTAAGCTTTTAATGGTGATGGTGATGGTGCTTTAAGCCTTCCGCTGACAG
ecGST86TAG-For	GACCGCCAGTTGCTGGCACCGTAGAACAGTATTTCCCGCTATAAAAC
ecGST86TAG-Rev	GTTTTATAGCGGGAAATACTGTTCTACGGTGCCAGCAACTGGCGGTC
ecGST86TAG92A-For	GCACCGTAGAACAGTATTTCCCGCGCTAAAACCATCGAATGGCT
ecGST86TAG92A-Rev	GCGCGGGAAATACTGTTCTACGGTGCCAGCAACTGGCGGTCGGG
ecGST86TAG92A72A-For	TGACGGAAGGCGTAGCGATTATGCAGGCTCTTGCCGACAGCGTCCCCGA
ecGST86TAG92A72A-For	GCCTGCATAATCGCTACGCCTTCCGTCAGCAAAGTACCGTCATCCA
ecGST65TAG-For	CTGCTGGATGACGGTACTTTGCTGACGTAGGGCGTAGCGATTATGCAGTAT
ecGST65TAG-Rev	CTACGTCAGCAAAGTACCGTCATCCAGCAGCAATGCAGGCACCTGCCCCTT
HR-sjGST NdeI	CTTTAAGAAGGAGATATACATATGACTAGTATGTCCCCTATACTAG
HR-sjGST HindIII	CCGCCAAAACAGCCAAGCTTTTAGTGATGGTGATGGTGATGGGATCCACGCGGAACCAGATCCG
sjGST97TAG-For	GATTTCAATGCTTGAAGGAtagGTTTTGGATATTAGATACGGTG
sjGST97TAG-Rev	CACCGTATCTAATATCCAAAACctaTCCTTCAAGCATTGAAATC
sjGST44A-For	GTGATAAATGGCGAAACAAAgcgTTTGAATTGGGTTTGGAGTTTC
sjGST44A-Rev	GAAACTCCAAACCCAATTCAAAcgcTTTGTTTCGCCATTTATCAC
ecGST HindIII-pCDNA	GTTGTTAAGCTTGCCACCATGAAATTGTTCTACAAACCGGGTG
ecGST XhoI-pCDNA	GGTGGTCTCGAGTTAATGGTGATGGTGATGGTGCTTTAAG
sjGST44S-For	TGAAGGTGATAAATGGCGAAACAAATCGTTTGAATTGGGTTTGGAG
sjGST44S-Rev	GATTTGTTTCGCCATTTATCACCTTCATCGCGCTCATACAAATGC
sjGST44T-For	TGAAGGTGATAAATGGCGAAACAAAACGTTTGAATTGGGTTTGGAG
sjGST44T-Rev	GTTTTGTTTCGCCATTTATCACCTTCATCGCGCTCATACAAATGC
sjGST44Y-For	GATGAAGGTGATAAATGGCGAAACAAATATTTTGAATTGGGTTTGGAGTT
sjGST44Y-Rev	ATATTTGTTTCGCCATTTATCACCTTCATCGCGCTCATACAAATGCTC
sjGST44H-For	GATGAAGGTGATAAATGGCGAAACAAACACTTTGAATTGGGTTTGGAGTT
sjGST44H-Rev	GTGTTTGTTTCGCCATTTATCACCTTCATCGCGCTCATACAAATGCTC
HRpEVOL-For	CTAACAGGAGGAATTACTAGTATGACCGTGAAGTACACCGAC
HRpEVOL-Rev	CAATGATGATGATGATGATGGTCGACTTAATTGATTTTGGCACCATTC
7D12 NdeI	TAAGAAGGAGATATACATATGAAATACCTGCTGCCGACCG
7D12 HindIII	TCCGCCAAAACAGCCAAGCTTTTAGTGGTGGTGGTGATGATGGGACGAGA
7D12 30TAG-For	TGACCTGTGCTGCGAGCGGCCGTACCAGCTAGAGCTACGGCATGGGTTGGTTT
7D12 30TAG-Rev	CTAGCTGGTACGGCCGCTCGCAGCACAGGTCAAACGGAGACTGCCACCGGTT
7D12 31TAG-For	CCTGTGCTGCGAGCGGCCGTACCAGCCGCTAGTACGGCATGGGTTGGTTTCGT
7D12 31TAG-Rev	CTAGCGGCTGGTACGGCCGCTCGCAGCACAGGTCAAACGGAGACTGCCACCG
7D12 109TAG-For	TTCTGCCTGGTATGGCACTCTGTAGGAATACGACTACTGGGG
7D12 109 TAG-Rev	CTACAGAGTGCCATACCAGGCAGAACCCGCCGCCGCGGCGCA

Reagents and molecular biology
Primers were synthesized and purified by Integrated DNA Technologies (IDT), and plasmids were sequenced by GENEWIZ. All molecular biology reagents were either obtained from New England Biolabs or Vazyme. His-HRP antibody, GFP monoclonal antibodies, GAPDH-HRP antibody were obtained from ProteinTech Group. pBAD-ubiquitin (6TAG) and pBAD-ecGST WT and ecGST mutants were used as previously described[1]. ecGST HindIII-pCDNA and ecGST XhoI-pCDNA primers were used to clone ecGST WT and ecGST (86TAG), ecGST (86TAG/92A), ecGST (86TAG/92A/72A) into pCDNA 3.1.

FSKRS amino acid sequence

MTVKYTDAQIQRLREYGNGTYEQKVFEDLASRDAAFSKEMSVASTDNEKKIKGMIANPSRHGL
TQLMNDIADALVAEGFIEVRTPIFISKDALARMTITEDKPLFKQVFWIDEKRALRPMLAPNLGSVA
RDLRDHTDGPVKIFEMGSCFRKESHSGMHLEEFTMLNLFDMGPRGDATEVLKNYISVVMKAA
GLPDYDLVQEESDVYKETIDVEINGQEVCSAAVGPTPIDAAHDVHEPWSGAGFGLERLLTIREK
YSTVKKGGASISYLNGAKIN
Red: mutated residues.

sfGFP (2TAG)
Primers sfGFP2TAG For and sfGFP2TAG Rev were used to construct pBAD-sfGFP (2TAG)

MXKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKG
IDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDG
PVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSHHHHHH
Red: amber codon TAG at 2nd position.

ecGST (86TAG)
Primers ecGST NdeI to GST86TAG-Rev were used to construct pBAD-ecGST (86TAG) by overlap PCR.

MKLFYKPGACSLASHITLRESGKDFTLVSVDLMKKRLENGDDYFAVNPKGQVPALLLDDGTLLT
EGVAIMQYLADSVPDRQLLAPXNSISRYKTIEWLNYIATELHKGFTPLFRPDTPEEYKPTVRAQL
EKKLQYVNEALKDEHWICGQRFTIADAYLFTVLRWAYAVKLNLEGLEHIAAFMQRMAERPEVQ
DALSAEGLKHHHHHH
Red: amber codon TAG at 86th position.

ecGST (65TAG)
pBAD-ecGST (65TAG) was constructed by site-directed mutagenesis with primers ecGST65TAG-For and ecGST65TAG-Rev

MKLFYKPGACSLASHITLRESGKDFTLVSVDLMKKRLENGDDYFAVNPKGQVPALLLDDGTLLT
XGVAIMQYLADSVPDRQLLAPVNSISRYKTIEWLNYIATELHKGFTPLFRPDTPEEYKPTVRAQL
EKKLQYVNEALKDEHWICGQRFTIADAYLFTVLRWAYAVKLNLEGLEHIAAFMQRMAERPEVQ
DALSAEGLKHHHHHH
Red: amber codon TAG at 65th position.

ecGST (86TAG/92A)
pBAD-ecGST (86TAG/92A) was constructed by site-directed mutagenesis with primers ecGST86TAG92A-For and ecGST86TAG92A-Rev

MKLFYKPGACSLASHITLRESGKDFTLVSVDLMKKRLENGDDYFAVNPKGQVPALLLDDGTLLT
EGVAIMQYLADSVPDRQLLAPXNSISRAKTIEWLNYIATELHKGFTPLFRPDTPEEYKPTVRAQL
EKKLQYVNEALKDEHWICGQRFTIADAYLFTVLRWAYAVKLNLEGLEHIAAFMQRMAERPEVQ
DALSAEGLKHHHHHH
Red: amber codon TAG at 86th position. Blue: 92 A.

ecGST (86TAG/92A/72A)
pBAD-ecGST (86TAG/92A/72A) was constructed by site-directed mutagenesis with primers ecGST86TAG92A72A-For and ecGST86TAG92A72A-Rev

MKLFYKPGACSLASHITLRESGKDFTLVSVDLMKKRLENGDDYFAVNPKGQVPALLLDDGTLLT
EGVAIMQALADSVPDRQLLAPXNSISRAKTIEWLNYIATELHKGFTPLFRPDTPEEYKPTVRAQL
EKKLQYVNEALKDEHWICGQRFTIADAYLFTVLRWAYAVKLNLEGLEHIAAFMQRMAERPEVQ
DALSAEGLKHHHHHH
Red: amber codon TAG at 86th position. Blue: 72/92 A.

sjGST WT
pBAD-sjGST WT was cloned with primers HR-sjGST NdeI and HR-sjGST HindIII.

MTSMSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDG
DVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLP
EMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDK
YLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRGSHHHHHH

sjGST (97TAG) and sjGST (97TAG/44 mutants)
pBAD-sjGST (97TAG) and pBAD-sjGST (97TAG/44A) were constructed by primers HR-sjGST NdeI, sjGST sjGST97TAG-For, sjGST97TAG-Rev, HR-sjGST HindIII rev, sjGST44A-For, and sjGST44A-Rev. And primers set 44S-For, 44S-Rev, 44T-For, 44T-Rev, 44Y-For, 44Y-Rev, 44H-For, 44H-Rev were used to prepare pBAD-sjGST (97TAG/44S), pBAD-sjGST (97TAG/44T), pBAD-sjGST (97TAG/44Y) and pBAD-sjGST (97TAG/44H).

MTSMSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDG
DVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGXVLDIRYGVSRIAYSKDFETLKVDFLSKLP
EMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDK
YLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRGSHHHHHH
Red: amber codon TAG at 97th position. Blue: Paired Lys 44 and its mutation to A, S, T, H, Y.

7D12 WT
Primers 7D12 NdeI and 7D12 HindIII were used to clone 7D12 WT to pBAD plasmid.

MGQVKLEESGGGSVQTGGSLRLTCAASGRTSRSYGMGWFRQAPGKEREFVSGISWRGD
STGYADSVKGRFTISRDNAKNTVDLQMNSLKPEDTAIYYCAAAAGSAWYGTLYEYDYWGQ
GTQVTVSSHHHHHH

7D12 (30TAG)
pBAD-7D12 (30TAG) was constructed by site-directed mutagenesis with primers 7D12 30TAG-For and 7D12 30TAG-Rev

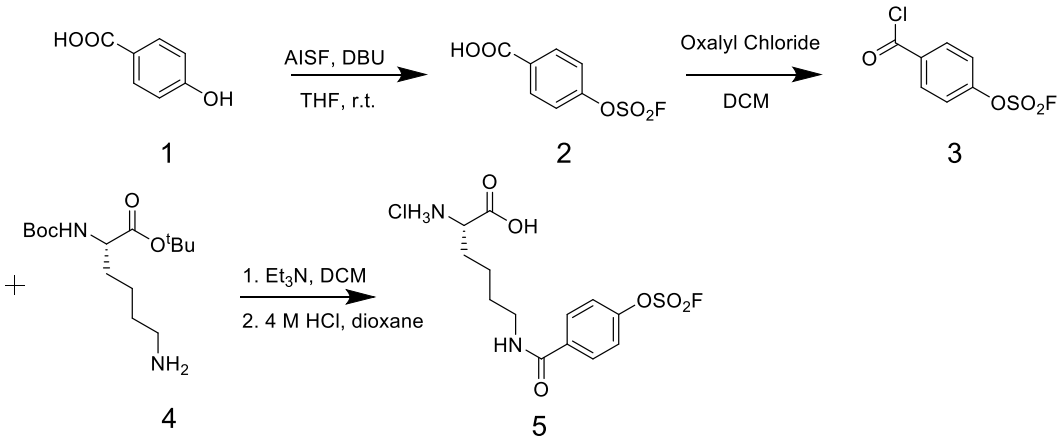
MGQVKLEESGGGSVQTGGSLRLTCAASGRTSXSYGMGWFRQAPGKEREFVSGISWRGD
STGYADSVKGRFTISRDNAKNTVDLQMNSLKPEDTAIYYCAAAAGSAWYGTLYEYDYWGQ
GTQVTVSSHHHHHH
Red: amber codon TAG at 30th position.

7D12 (31TAG)
pBAD-7D12 (31TAG) was constructed by site-directed mutagenesis with primers 7D12 31TAG-For and 7D12 31TAG-Rev

MGQVKLEESGGGSVQTGGSLRLTCAASGRTSRXYGMGWFRQAPGKEREFVSGISWRGD
STGYADSVKGRFTISRDNAKNTVDLQMNSLKPEDTAIYYCAAAAGSAWYGTLYEYDYWGQ
GTQVTVSSHHHHHH
Red: amber codon TAG at 31st position.

Chemical Synthesis of FSK

Synthesis of aryl fluorosulfates was based on recent methods to synthesize sulfur (IV) fluorides using [4-(acetylamino)phenyl]imidodisulfuryl difluoride (AISF) reagent.[2]



*Synthesis of 4-((fluorosulfonyl)oxy)benzoic acid (**2**).* To a 200 mL round-bottom flask were added 4-hydroxybenzoic acid (1, 1.38 g, 10 mmol) and [4-(acetylamino)phenyl]imidodisulfuryl difluoride (AISF) reagent (3.78 g, 12 mmol, 1.2 equiv.). The mixture was dissolved in 50 mL anhydrous tetrahydrofuran and 1,8-diazabicyclo[5.4.0]undec-7-ene (3.35 mL, 22 mmol, 2.2 equiv.) was added dropwise while stirring. The solution was then stirred at r.t. for 20 minutes. The reaction was then diluted with 50 mL ethyl acetate and washed with 1 M HCl (100 mL x 2) and brine (100 mL x 1). The organic fraction was dried with anhydrous sodium sulfate and concentrated under vacuum. The crude product was then purified by column chromatography using MeOH:CH_2Cl_2 (1:100). The product, 4-((fluorosulfonyl)oxy)benzoic acid, was isolated as a white solid (2, 1.72 g, 7.8 mmol, 78%).

*Synthesis of fluorosulfonyloxybenzoyl-L-lysine (**5**, FSK).* To a stirred solution of 4-((fluorosulfonyl)oxy)benzoic acid (2, 0.22 g, 1 mmol) in dry CH_2Cl_2 (15 mL) was added oxalyl chloride (0.21 ml, 2.5 mmol, 2.5 equiv.) dropwise under argon at 0°C. Dimethylformamide (0.1 mL) was then added as catalyst. The reaction mixture was then stirred at r.t. for 5 hours. The

solution was then concentrated under vacuum resulting in a yellow oil. The crude 4-(chlorocarbonyl)phenyl sulfofluoridate (3, ~1 mmol) was redissolved in dry CH_2Cl_2 (10 mL) and cooled to 0°C. *N*-Boc-Lys-^{t}Bu (4, 0.34 g, 1 mmol, 1 equiv.) was then added, after which Et_3N (0.15 mL, 1.1 mmol, 1.1 equiv.) was added dropwise. The reaction mixture was stirred at r.t. overnight. The reaction was quenched with 20 mL of H_2O and washed with 1 M HCl (20 mL x 2). The aqueous phase was combined and extracted with ethyl acetate (20 mL x 2). The organic fractions were combined and dried over anhydrous sodium sulfate and concentrated under vacuum. The crude product was then purified by column chromatography using MeOH:CH_2Cl_2(1:100). The product, N-Boc-FSK-^{t}Bu, was isolated as a yellow oil (0.25 g, 0.50 mmol, 50%).

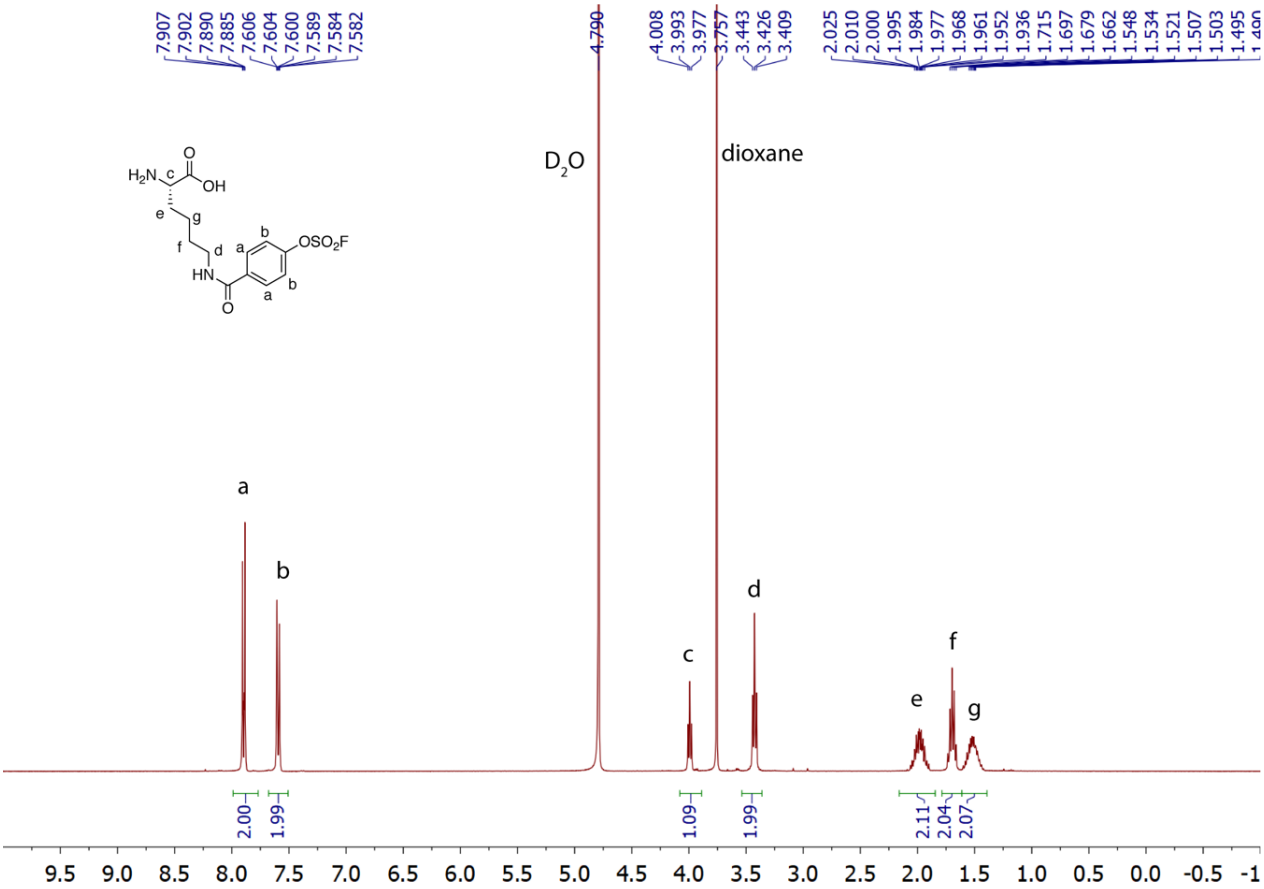
N-Boc-FSK-^{t}Bu (0.25 g , 0.50 mmol) was added to a scintillation vial and dissolved in 4 M HCl in dioxane (10 mL). The reaction was stirred overnight. The resultant solid was filtered off and washed with cool ether (10 mL x 2) affording the product FSK-HCl as a white solid (5, 158 mg, 0.41 mmol, 81%)

^{1}H NMR (400 MHz, D_2O): δ (ppm) 7.89 (d, *J* = 8.8 Hz, 2H), 7.59 (d, *J* = 8.8 Hz, 2H), 3.99 (t, *J* = 6.0, 1H), 3.43 (t, *J* = 6.8 Hz, 2H), 2.03-1.94 (m, 2H), 1.72-1.66 (m, 2H), 1.55-1.49 (m, 2H)

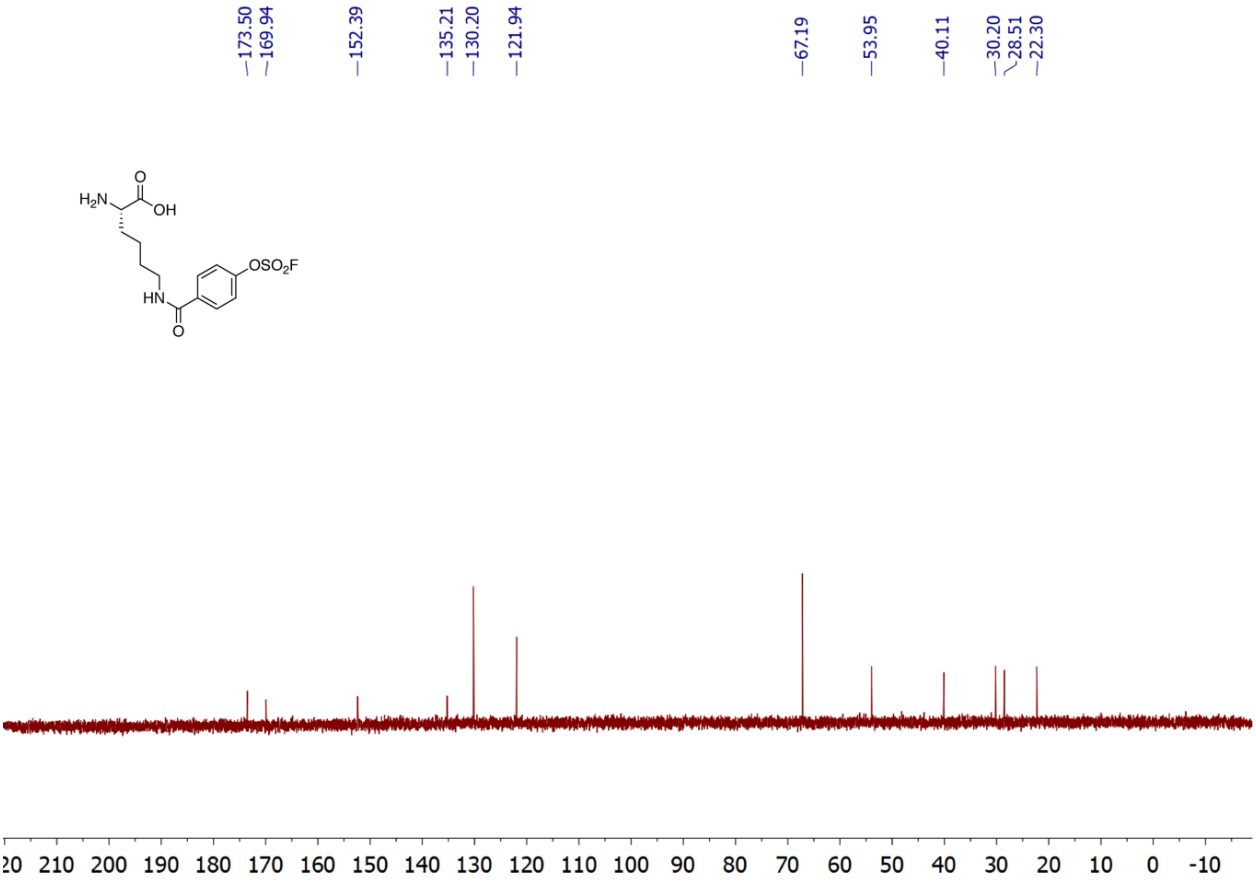
^{13}C NMR (100 MHz, D_2O): δ (ppm) 173.5, 169.9, 152.4, 135.2, 130.2, 121.9, 53.9, 40.1, 30.2, 28.5, 22.3

HR-ESI (+) *m/z*: calculated for $C_{13}H_{17}FN_2NaO_6S$ $[M+Na]^+$, 371.0684; found 371.0690.

^{1}H NMR



^{13}C NMR


173.50
169.94
152.39
135.21
130.20
121.94
67.19
53.95
40.11
30.20
28.51
22.30
H2N
O
OH
HN
O
OSO2F
220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10

Library construction and FSKRS mutant selection

To screen an efficient synthetase for the incorporation of FSK, the primers *Ma*PylRS NdeI to *Ma*PylRS PstI were used to randomize the active site of PylRS and create the library for FSK screening. The selection of an orthogonal synthetase for FSK incorporation was carried out by following the procedures described previously.[3-4] Candidate hits were recloned to pEVOL plasmid with primers HRpEVOL-For and HRpEVOL-Rev followed by investigating the incorporation efficiency into pBAD-EGFP (182TAG). The incorporation efficiency for the hits were compared by measuring the green fluorescence intensity (excitation at 485 nm, emission at 528 nm) normalized to OD at 600 nm. Four candidate hits were identified with the mutations shown below.

Number	AA mutations
1	Y126G/M129A/V168F/H227T/Y228P/L229I
2	Y126G/M129A/V168F/H227S/Y228P/L229V
3	Y126G/M129A/V168F/H227I/Y228P
4	Y126G/M129A/V168F/H227S/Y228P/L229I

Incorporation of FSK into EGFP (182TAG), sfGFP (151TAG), sfGFP (2TAG)

pBAD-sfGFP (2TAG), pBAD-sfGFP (151TAG) or pBAD-EGFP (182TAG) was co-transformed with pEVOL-FSKRS into DH10b, and plated on LB argar plate supplemented with 50 μg/mL kanamycin and 34 μg/mL chloramphenicol. A single colony was picked and inoculated into 1 mL 2xYT (5 g/L NaCl, 16 g/L Tryptone, 10 g/L Yeast extract). The cells were left grown 37 °C, 220 rpm to an OD 0.5, with good aeration for overnight. Next morning, the cells were diluted 10 times in fresh 2xYT supplemented with relevant antibiotics, 0.2% arabinose with or without 1 mM FSK. The cells were then induced at either 30 °C for 6 hr or 18 °C for overnight. The fluorescence was checked by a plate reader as described above.

General incorporation of FSK into proteins for expression and purification

For the incorporation of FSK into ubiquitin (6TAG), ubiquitin (18TAG), 7D12 (30TAG), and 7D12 (31TAG), the procedure of transformation is the same as described above. After transformation, a single colony was picked and left grown at 37 °C, 220 rpm for overnight. Next morning, the cell culture was diluted 100 times and then regrown to an OD 0.5 in 30 to 100 mL scale, with good aeration and the relevant antibiotic selection. Then the medium was added with 0.2% arabinose with or without 1 mM FSK, and the expression were carried out at 18 °C, 220 rpm for 18 hr, 18 °C, or 6 hr at 30 °C. The IMAC chromatography was used for protein purification. And the procedure was done as described elsewhere.[1]

Utilization of FSK and FSY into *ec*GST, *sj*GST and their mutants for protein cross-linking in *E. coli*

For probing ecGST or sjGST and their mutants' cross-linking in living E. coli bacterial cells. pBAD-ecGST WT, pBAD-ecGST (86TAG), pBAD-ecGST (65TAG), pBAD-ecGST (86TAG/92A), pBAD-ecGST (86TAG/92A/72A), or sjGST WT, sjGST (97TAG), sjGST (97TAG/44A, S, T, H, or Y) was was co-transformed with either pEVOL-FSYRS or pEVOL-FSKRS into DH10b cells. FSY or FSK was added with 0.2% arabinose respectively to the cells for induction when the cells were grown to an OD around 0.5. The cells were grown for protein expression at 37 °C for 6 hr, which then were harvested by centrifugation with a benchtop centrifuge and treated with 2xSDS loading dye

containing 100 mM DTT, and boiled for 5 mins at 95 °C. The dimerization of GST due to cross linking was monitored by Western blot using anti-his antibody.

In vitro cross-linking of 7D12 and EGFR

To explore in vitro cross-linking of 7D12 and EGFR, purified 2 μM 7D12 WT, 7D12(30FSK) or 7D12(31FSK) was incubated with 500 nM recombinant human EGFR protein respectively (Abcam, Cat# ab155726) in 15 μL 1xPBS, pH 7.4. After incubation at 37 °C for 16 h, the samples were treated with a final 1xSDS loading dye and boiled for 5 mins at 95 °C. The cross-linking was investigated by running Coomassie blue SDS-PAGE or Western blot with 1:10000 anti-his antibody.

In cellular cross-linking of 7D12 and EGFR

For direct cross-linking of 7D12 to A431 mammalian cells which overexpressed EGFR, A431 cells were seeded in 24-well plate (2×10^5 cells per well) and cultured overnight at 37 °C. The cells were treated with 1 μM 7D12 and 7D12(31TAG) for 1, 2, 4, 8 and 12 h. After digestion with trypsin, the cells were collected by centrifugation at 300 g for 5 min and lysed by adding 100 μL RIPA Buffer with 1x protease inhibitor cocktail. The samples were separated on SDS-PAGE and subjected to Western-blot detection with 1:10000 anti-his antibody. Anti-GAPDH antibody was used as a reference protein.

Genetic incorporation of FSK into Hela GFP (182TAG)

The plasmid pNEU-FSKRS (1 μg) was transfected into Hela-GFP 182(TAG) cells with 3 μL polyethylenimine (PEI) in 2 mL RPMI 1640 media when the cells population reached 80% confluence. A blank Hela-GFP 182(TAG) cell group was used as a negative control. The cells were treated with or without 1 mM FSK 6 hr after transfection and cultured for additional 48 hr. The cells were washed with 1xPBS for one time and subjected for microscope image after which will be harvested and ran Western blot using anti-GFP antibody. Anti-GAPDH antibody was used as a reference protein.

Genetic incorporation of FSK into ecGST mutants in mammalian cells

For probing protein cross-linking in mammalian cells. The plasmid pNEU-FSKRS (1.5 μg) was co-transfected with 1 μg pCDNA 3.1 ecGST WT, 1.5 μg ecGST (86TAG), 1.5 μg ecGST(86TAG/92A), and 1.5 μg ecGST(86TAG/92A/72A) respectively into HEK (293T) cells with 9 μL polyethylenimine (PEI) in 2 mL DMEM media when the cells population reached 80% confluence. The cells were treated with or without 1 mM FSK 6 hr after transfection and cultured for additional 48 hr. The cells were harvested and ran Western blot using anti-His antibody. Anti-GAPDH antibody was used as a reference protein.

Mass spectrometry

Mass spectrometric measurements were performed as previously described.[5] Briefly for electrospray ionization mass spectrometry, mass spectra of intact proteins were obtained using a QTOF Ultima (Waters) mass spectrometer, operating under positive electrospray ionization (+ESI) mode, connected to an LC-20AD (Shimadzu) liquid chromatography unit. Protein samples were separated from small molecules by reverse phase chromatography on a Waters Xbridge BEH C4 column (300 Å, 3.5 μm, 2.1 mm x 50 mm), using an acetonitrile gradient from 30-71.4%, with 0.1% formic acid. Each analysis was 25 min under constant flow rate of 0.2 mL/min at RT. Data were acquired from m/z 350 to 2500, at a rate of 1 sec/scan. Alternatively, spectra were acquired by Xevo G2-S QTOF on a Waters ACQUITY UPLC Protein BEH C4 reverse-phase column (300 Å, 1.7 μm, 2.1 mm x 150 mm). An acetonitrile gradient from 5%-95% was used with 0.1% formic acid, over a run time of 5 min and constant flow rate of 0.5 mL/min at RT. Spectrum were acquired from m/z 350 to 2000, at a rate of 1 sec/scan. The spectra were deconvoluted

using maximum entropy in MassLynx. For tandem mass spectrometry, analysis and sequencing of peptides were carried out using a Q Exactive Orbitrap interfaced with Ultimate 3000 LC system. Data acquisition by Q Exactive Orbitrap was as follows: 10 µL of trypsin-digested protein was loaded on an Ace UltraCore super C18 reverse-phase column (300 Å, 2.5 µm, 75 mm × 2.1 mm) via an autosampler. An acetonitrile gradient from 5%-95% was used with 0.1% formic acid, over a run time of 45 min and constant flow rate of 0.2 mL/min at RT. MS data were acquired using a data-dependent top10 method dynamically choosing the most abundant precursor ions from the survey scan for HCD fragmentation using a stepped normalized collision energy of 28, 30 35 eV. Survey scans were acquired at a resolution of 70,000 at m/z 200 on the Q Exactive. Theoretical patterns of isotopic patterns of peptides were calculated using UCSF MS-ISOTOPE[6] or enviPat Web 2.1.[7] Cross-linked peptides were searched with pLink2 and OpenUaa.[8,9]

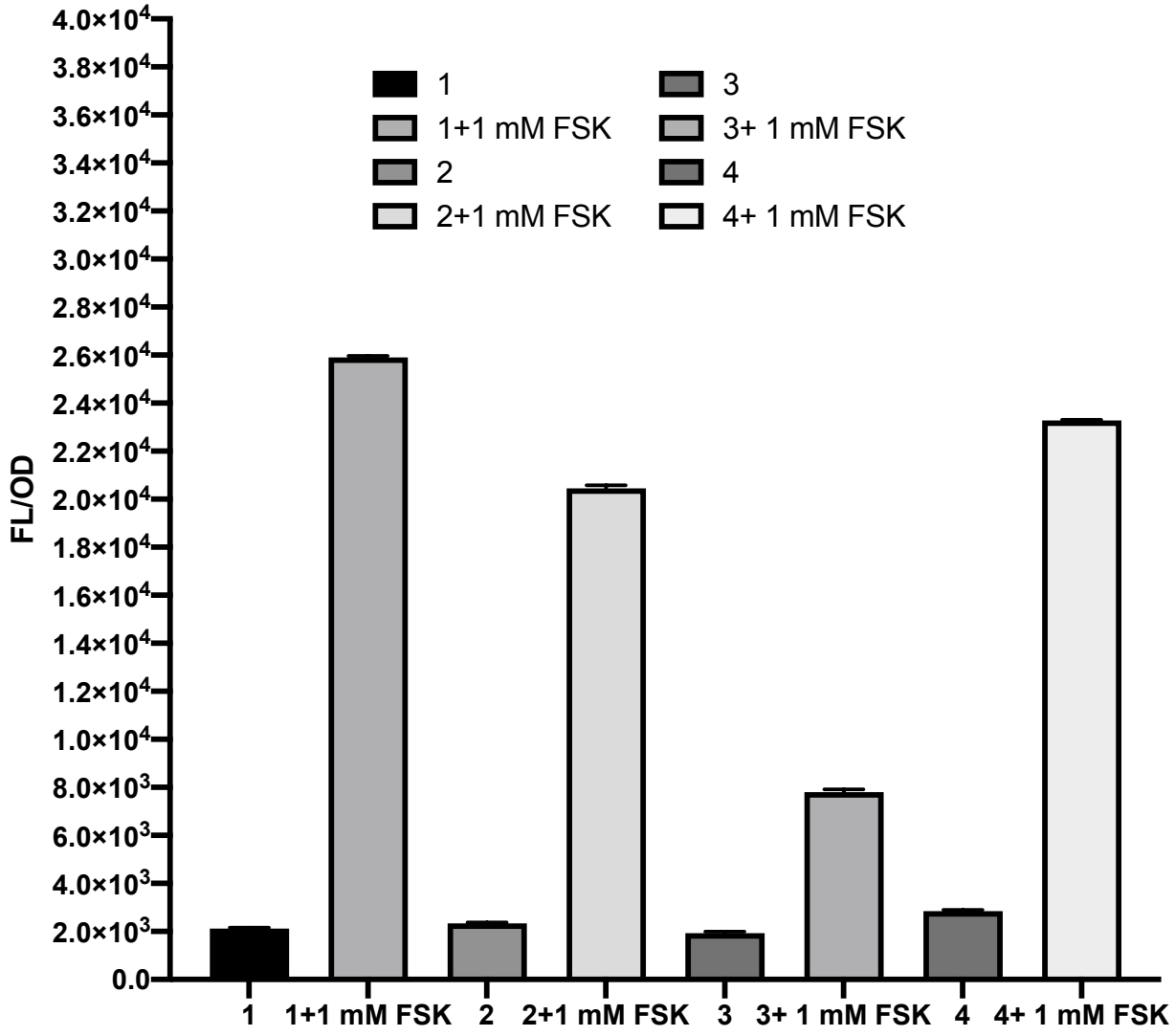


Figure S1. Comparison of FSK incorporation efficiency into EGFP at 18 °C for 24 h. The four identified synthetase hits indicated by numbers were separately used with $tRNA^{Pyl}$ to incorporate FSK into EGFP (182TAG). EGFP expression was induced at 18 °C for 24 h, in the presence or absence of 1 mM FSK in growth media. Cell fluorescence intensity normalized to cell OD_{600} was compared. Error bars represent s.e.m., n = 3 independent experiments.

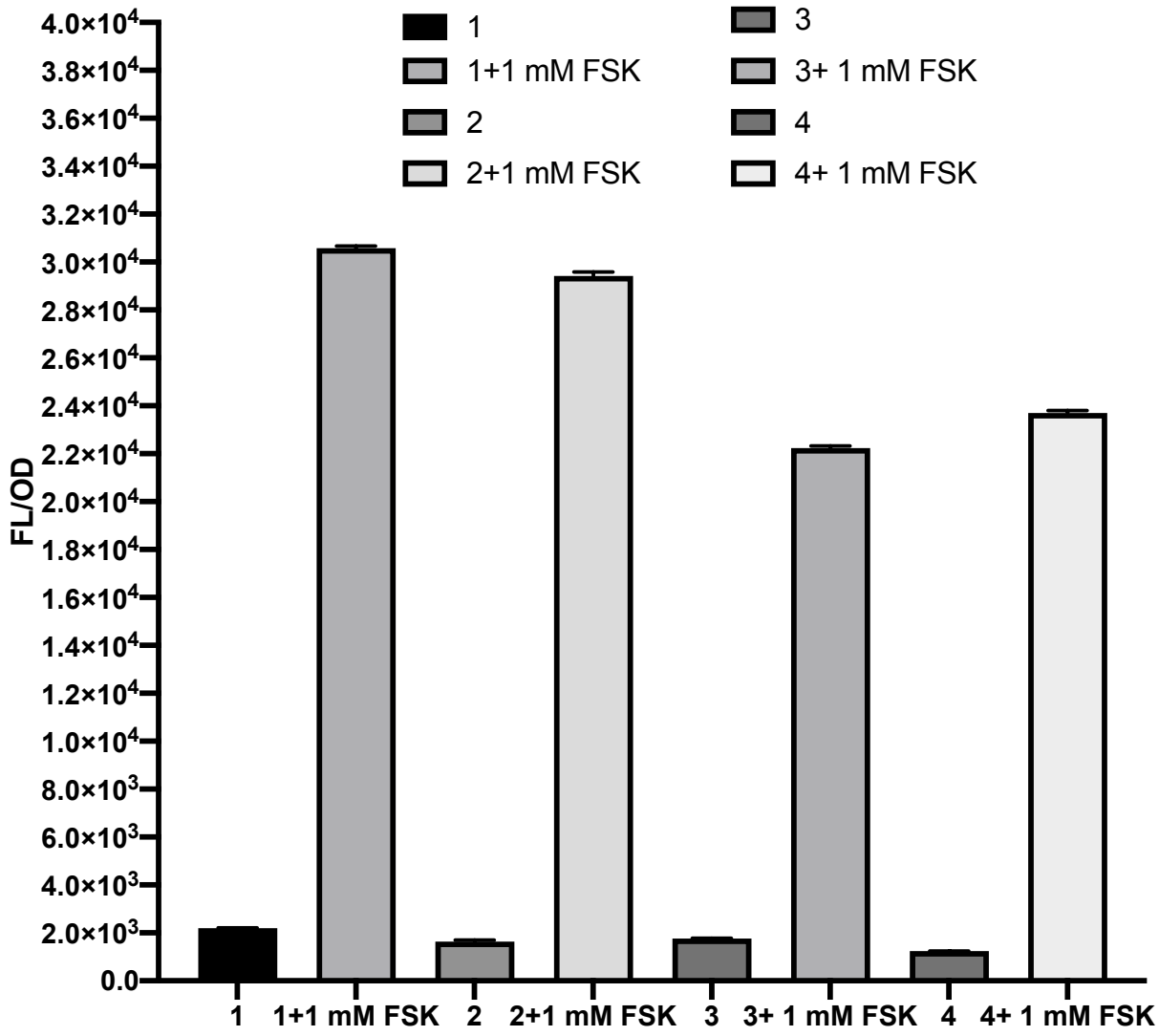


Figure S2. Comparison of FSK incorporation efficiency into EGFP at 30 °C for 6 h. The four identified synthetase hits indicated by numbers were separately used with $tRNA^{Pyl}$ to incorporate FSK into EGFP (182TAG) in *E. coli*. EGFP expression was induced at 30 °C for 6 h, in the presence or absence of 1 mM FSK in growth media. Cell fluorescence intensity normalized to cell OD_{600} was compared. Error bars represent s.e.m., n = 3 independent experiments.

.

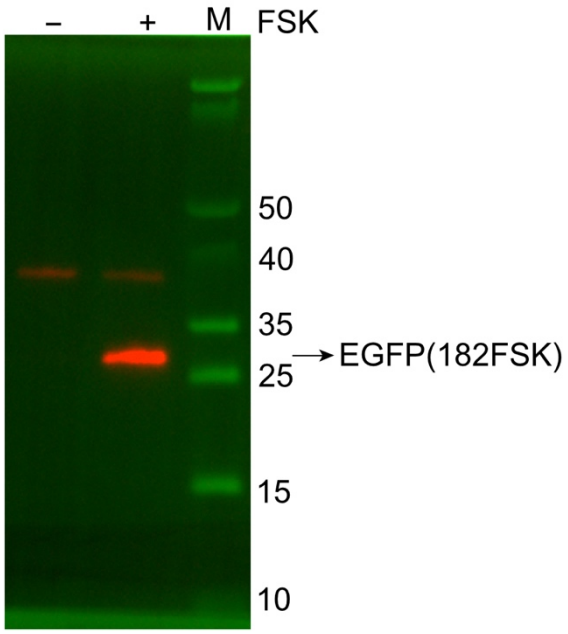


Figure S3. Western blot analysis of incorporation of FSK into EGFP (182TAG). FSKRS/tRNA^{Pyl} were co-transformed with pBAD-EGFP(182TAG) into *E. coli*, and cells were grown with or without 1 mM FSK. Cell lysates were separated, and an anti-His antibody was used to detect the Hisx6 tag appended at the C-terminus of EGFP.

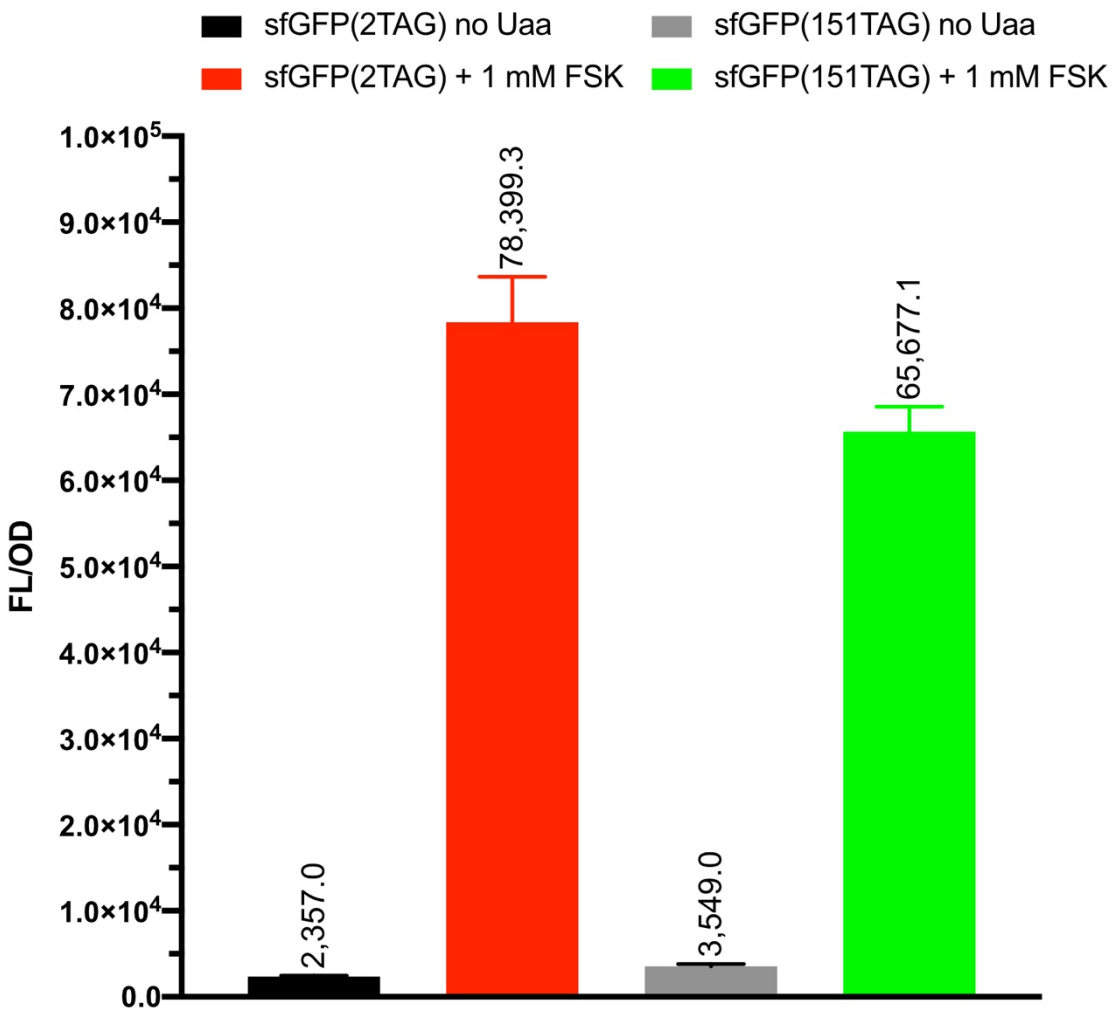


Figure S4. Incorporation of FSK into sfGFP (2TAG) and sfGFP(151TAG). pEVOL-FSKRS was co-transformed with pBAD-sfGFP(2TAG) or pBAD-sfGFP(151TAG) into *E. coli* DH10b cells, respectively. Cells were grown with or without 1 mM FSK. sfGFP expression was detected using a plate reader (485 nm excitation wavelength, 528 nm emission wavelength). The plot represented the sfGFP fluorescence intensity after normalization to bacterial growth at optical density 600 nm. Error bars represent s.e.m., n = 3 independent experiments.

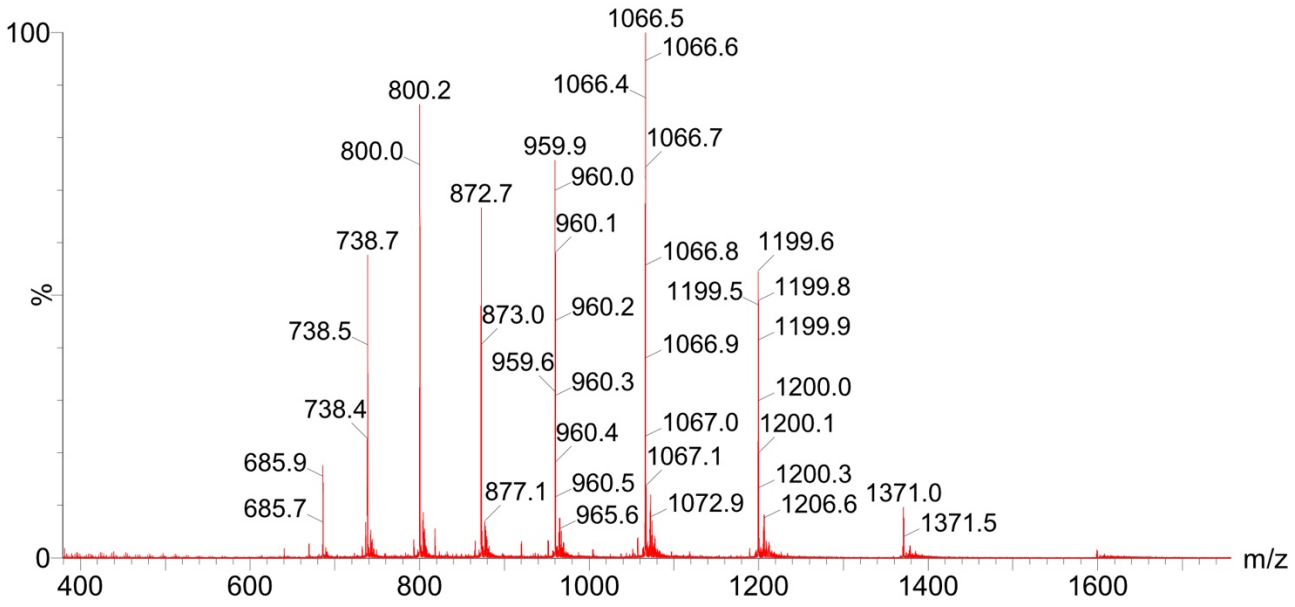


Figure S5. M/Z mass spectrum of Ub (6FSK).

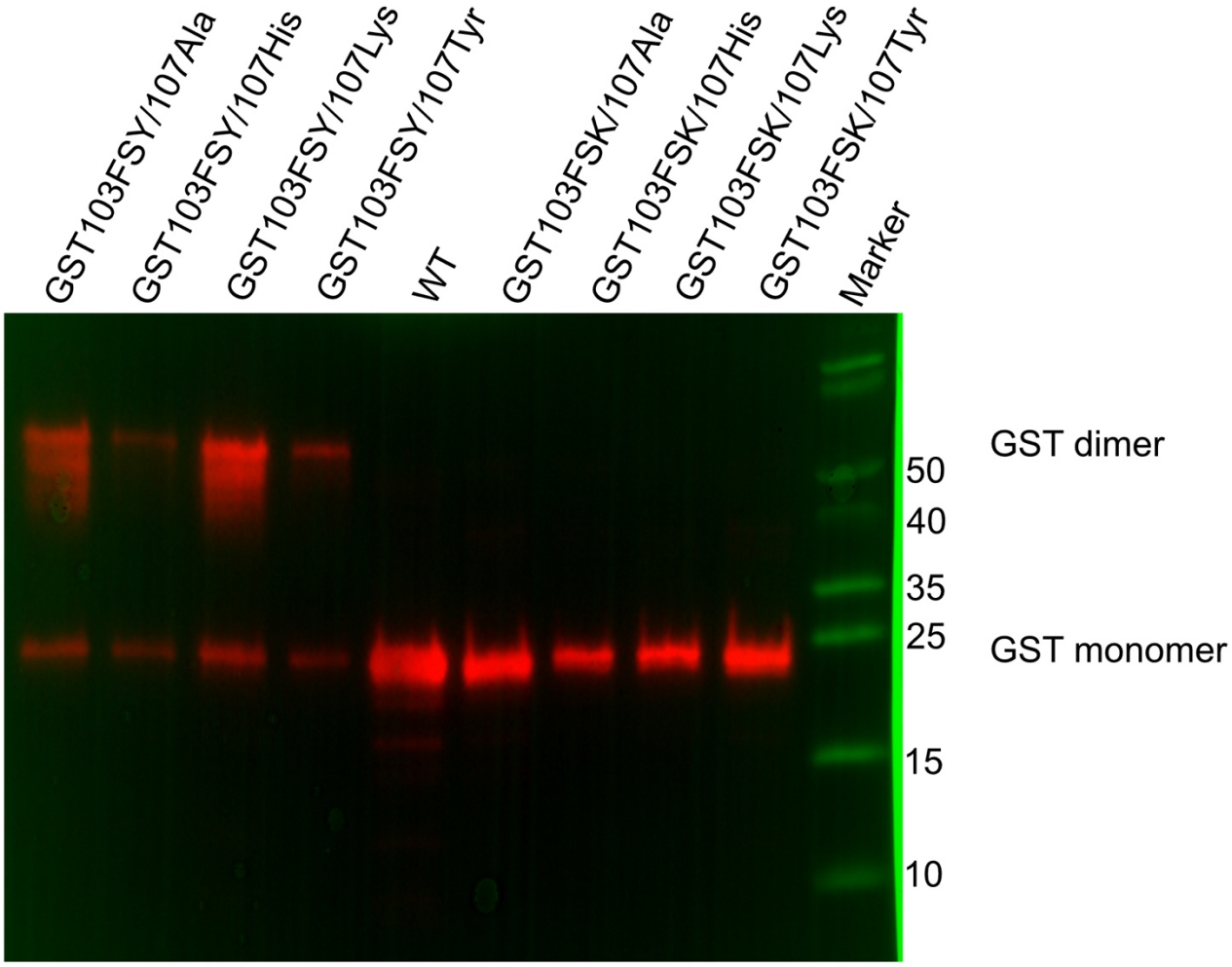


Figure S6. Comparison of the FSY- and FSK-mediated GST cross-linking in short distance proximity. The pEVOL-FSYRS or pEVOL-FSKRS was co-expressed with *ec*GST (103TAG/107Ala), *ec*GST (103TAG/107His), *ec*GST (103TAG/107Lys), *ec*GST (103TAG/107Tyr), respectively in *E. coli* in the presence of 1 mM FSK or FSY at 37 °C for 6 h. The WT GST was used as a negative control. The GST dimeric cross-linking was detected through Western blot of *E. coli* cell lysate by using anti-His antibody to detect the Hisx6 tag appended at the C-terminus of GST. *ec*GST has His at site 106, which could also be cross-linked by FSY103 in the first four lanes.

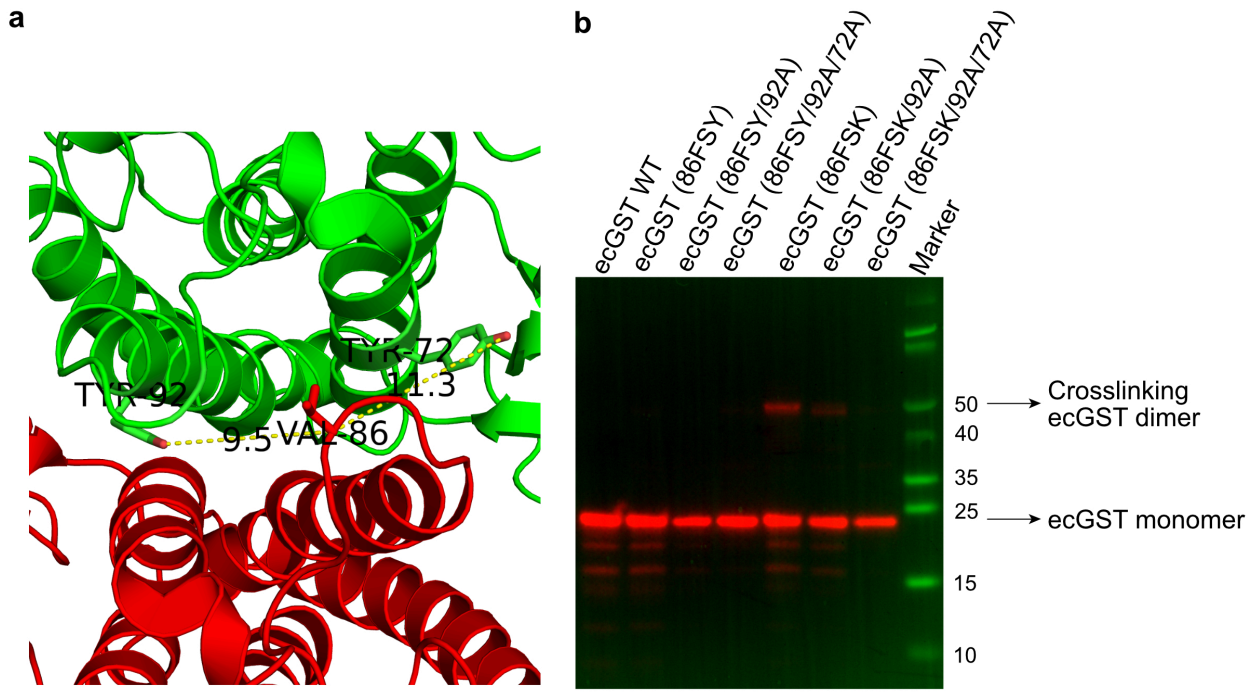


Figure S7. Comparison of the FSY- and FSK- mediated *E. coli* GST cross-linking at site 86. The pEVOL-FSYRS or pEVOL-FSKRS was co-expressed with *ec*GST WT or pBAD-GST (86TAG) in the presence of 1 mM FSK or FSY at 37 °C for 6 h. The WT GST was used as a negative control. The GST dimer cross-linking was detected with Western blot by using an anti-His antibody.

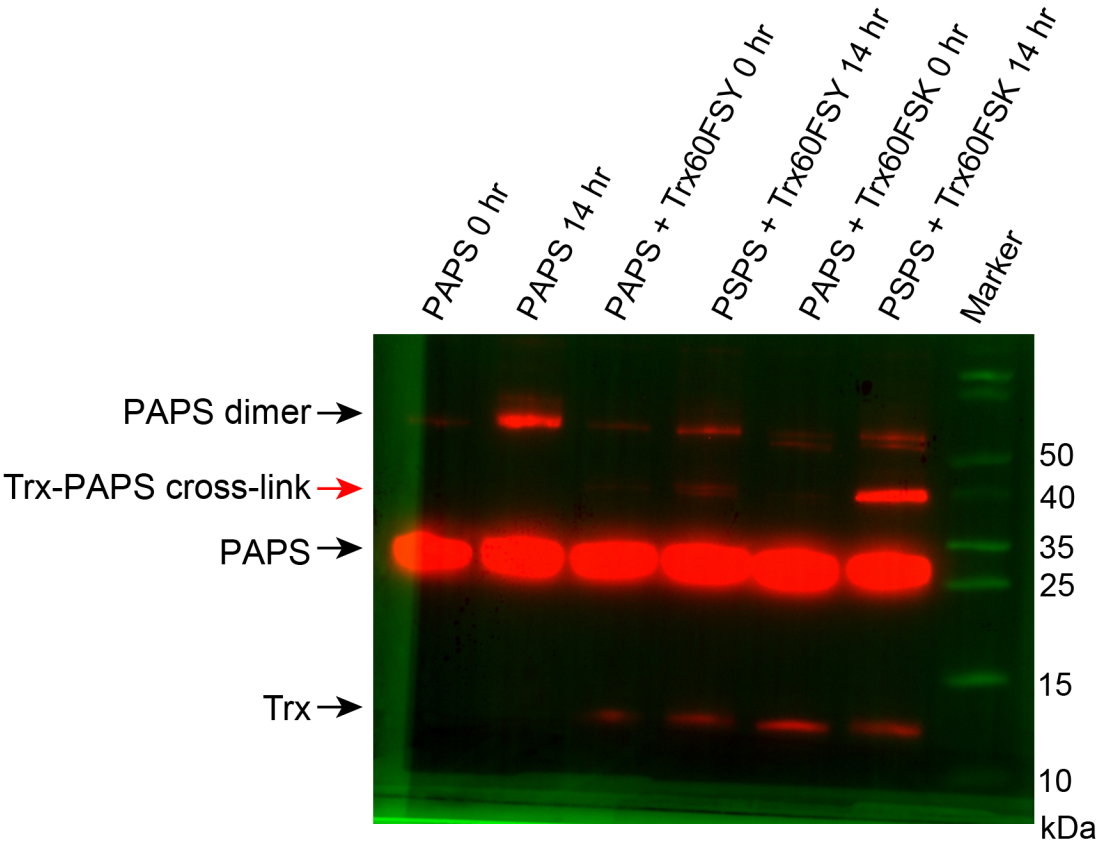


Figure S8. Comparison of the cross-linking efficiency of FSK and FSY in mediating Trx and PAPS cross-linking. An anti-His antibody was used to detect both Trx and PAPS in the Western blot.

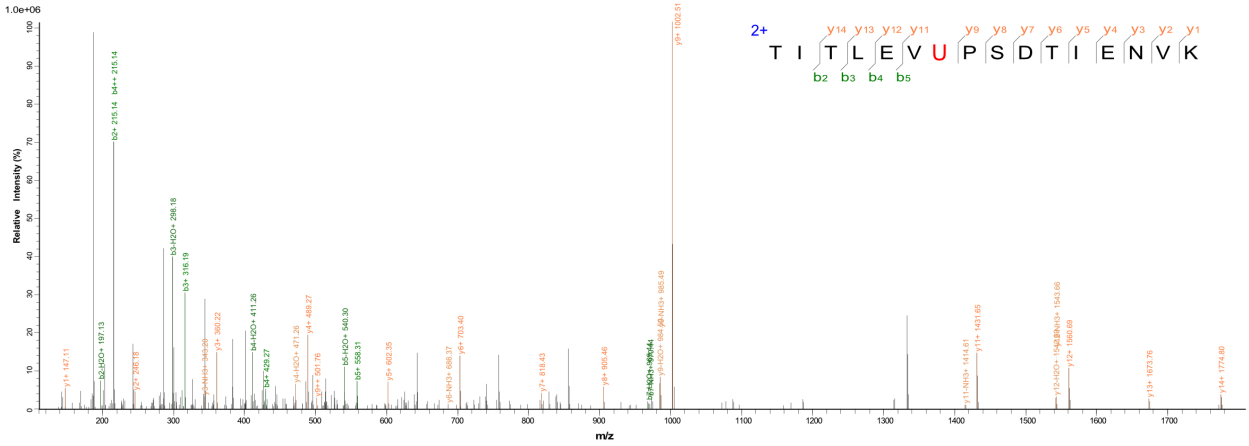


Figure S9. Tandem mass spectrum of the intact FSK-incorporated peptide of ubiquitin. U represents FSK incorporated at site 18 of ubiquitin.

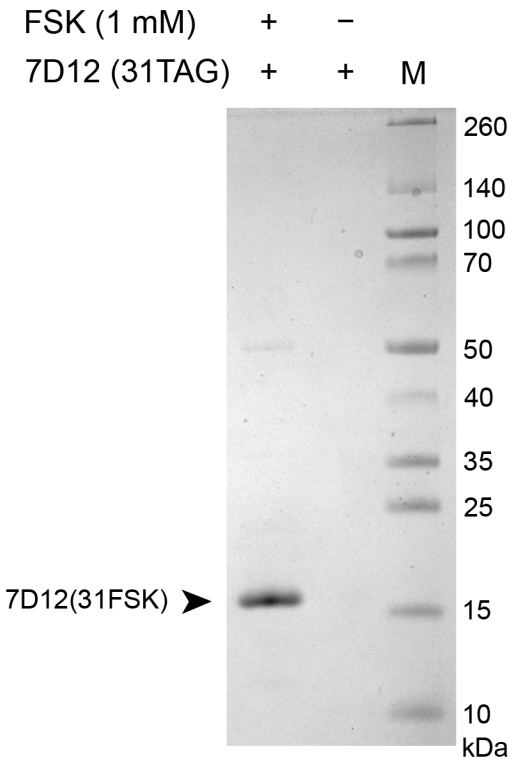


Figure S10. SDS-PAGE analysis of purified nanobody 7D12 (31FSK). In the absence of 1 mM FSK during expression, no full-length nanobody was produced.

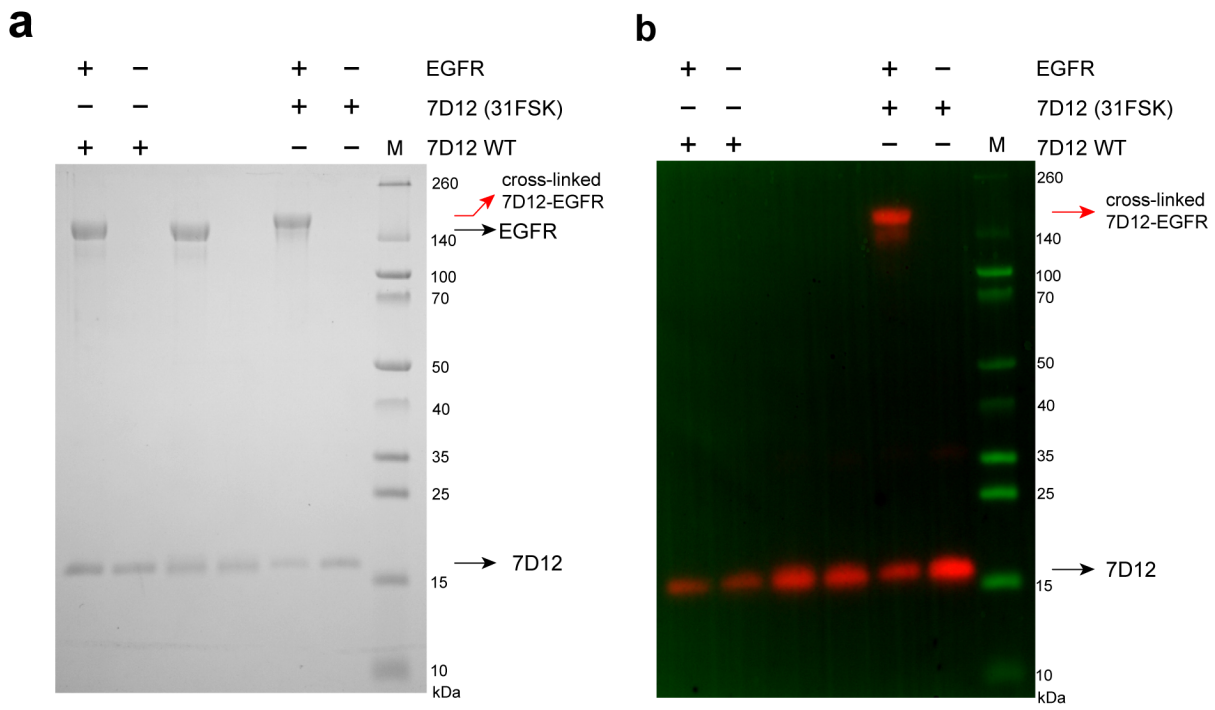


Figure S11. SDS-PAGE and Western blot analysis of covalent cross-linking of nanobody 7D12 (WT) and 7D12 (31FSK) to EGFR *in vitro*. (**a**) SDS-PAGE. (**b**) Western blot using an anti-His6 antibody to detect nanobody 7D12.

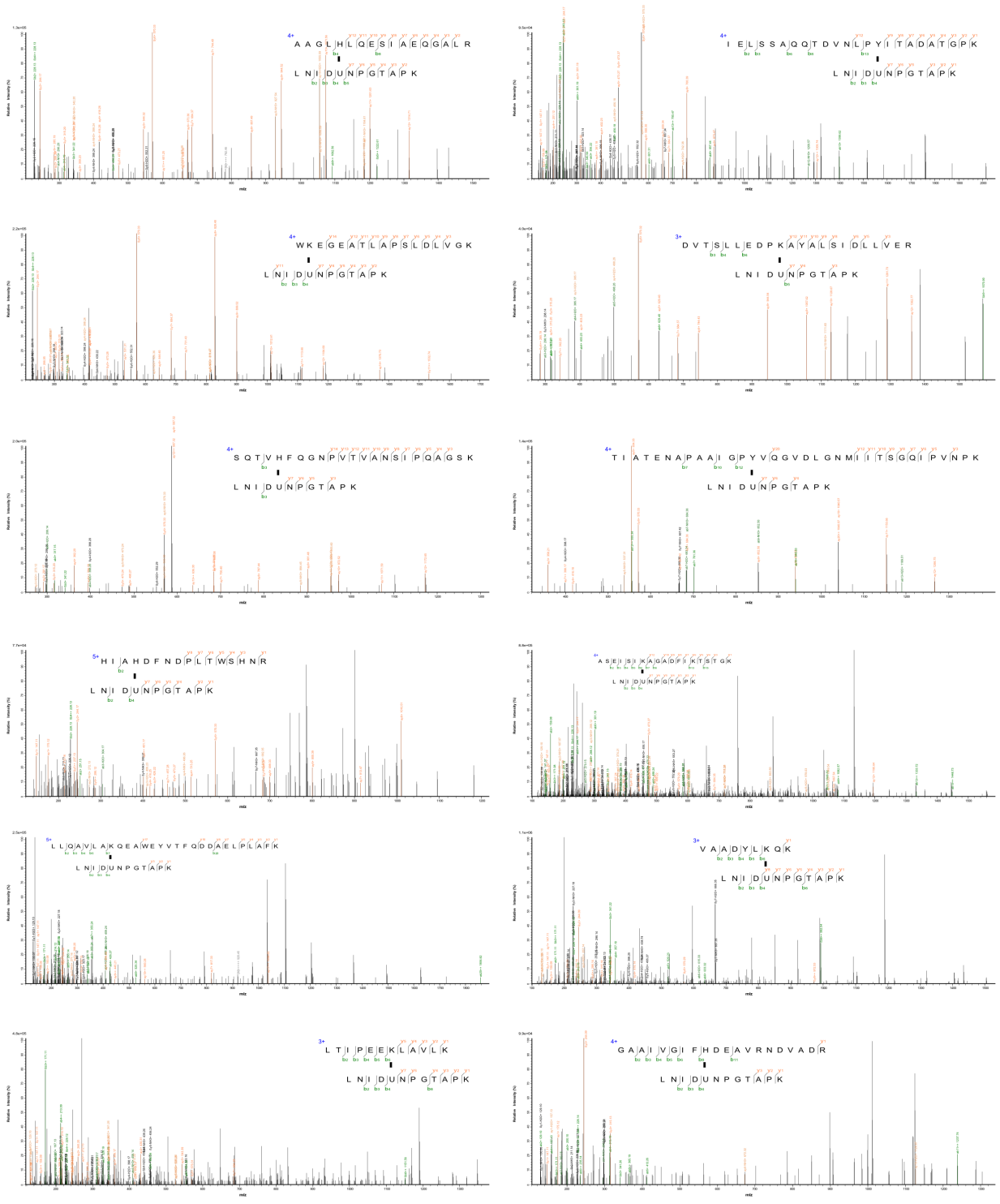


Figure S12. Tandem mass spectra of cross-linked peptides identified from Trx (62FSK).

Figure S13. Tandem mass spectra of cross-linked peptides identified from Trx (62FSY).

Supplementary Reference

1. Liu, J.; Li, S.; Aslam, N. A.; Zheng, F.; Yang, B.; Cheng, R.; Wang, N.; Rozovsky, S.; Wang, P. G.; Wang, Q.; Wang, L., Genetically Encoding Photocaged Quinone Methide to Multitarget Protein Residues Covalently in Vivo. *J. Amer. Chem. Soc.* **2019,** *141* (24), 9458-9462.
2. Zhou, H.; Mukherjee, P.; Liu, R.; Evrard, E.; Wang, D.; Humphrey, J.M.; Butler, T.W.; Hoth, L.R.; Sperry, J.B.; Sakata, S.K.; Helal, C.J.; am Ende, C.W. *Org. Lett.* **2018,** 20(3), 812-815.
3. Liu, J.; Zheng, F.; Cheng, R.; Li, S.; Rozovsky, S.; Wang, Q.; Wang, L., Site-Specific Incorporation of Selenocysteine Using an Expanded Genetic Code and Palladium-Mediated Chemical Deprotection. *J. Amer. Chem. Soc.* **2018,** *140* (28), 8807-8816.
4. Liu, J.; Cheng, R.; Wu, H.; Li, S.; Wang, P. G.; DeGrado, W. F.; Rozovsky, S.; Wang, L., Building and Breaking Bonds via a Compact S-Propargyl-Cysteine to Chemically Control Enzymes and Modify Proteins. *Angew. Chem. Int. Ed. Engl.* **2018,** *57* (39), 12702-12706.
5. Liu, J.; Chen, Q.; Rozovsky, S., Utilizing Selenocysteine for Expressed Protein Ligation and Bioconjugations. *J. Amer. Chem. Soc.* **2017,** *139* (9), 3430-3437.
6. Baker, P. R., Clauster, K.R., http://prospector.ucsf.edu.
7. Loos, M.; Gerber, C.; Corona, F.; Hollender, J.; Singer, H., Accelerated Isotope Fine Structure Calculation Using Pruned Transition Trees. *Anal. Chem.* **2015,** *87* (11), 5738-5744.
8. Chen, Z.; Meng, J.; Cao, Y.; Yin, J.; Fang, R.; Fan, S.; Liu, C.; Zeng, W.; Ding, Y.; Tan, D.; Wu, L.; Zhou, W.; Chi, H.; Sun, R.; Dong, M.; He, S. A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat. Commun.* **2019**, *10*, 3404.
9. Liu, C.; Wu, T.; Shu, X.; Li, S.; Wang, D.R.; Wang, N.; Zhou, R.; Yang, H.; Jiang, H.; Hendriks, I.A.; Gong, P.; Zhang, L.; Nielsen, M.L.; Li, K.; Wang, L.; Yang, B. Identification of Protein Direct Interactome with Genetic Code Expansion and Search Engine OpenUaa. *Adv. Biology* **2021**, *5*, 2000308.