Published in final edited form as:

J Am Chem Soc. 2021 July 14; 143(27): 10341–10351. doi:10.1021/jacs.1c04259.

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], **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

[#] These authors contributed equally to this work.

Abstract

Genetically introducing novel chemical bonds into proteins provides innovative avenues for biochemical research, protein engineering, and biotherapeutic applications. Recently, latent bioreactive unnatural amino acids (Uaas) have been incorporated into proteins to covalently target natural residues through proximity-enabled reactivity. Aryl fluorosulfate is particularly attractive due to its exceptional biocompatibility and multi-targeting capability via SuFEx reaction. Thus far, fluorosulfate-L-tyrosine (FSY) is the only aryl fluorosulfate-containing Uaa that has been genetically encoded. FSY has a relatively rigid and short side chain, which restricts the diversity of proteins targetable and the scope of applications. Here we designed and genetically encoded a new latent bioreactive Uaa, fluorosulfonyloxybenzoyl-L-lysine (FSK), in *E. coli* and mammalian cells. Due to its long and flexible aryl fluorosulfate-containing side chain, FSK was particularly useful in covalently linking protein sites that are unreachable with FSY, both intra- and inter-molecularly, *in vitro* and in live cells. In addition, we created covalent nanobodies that irreversibly bound to EGFR receptors on cells, with FSK and FSY targeting distinct positions on EGFR to counter potential mutational resistance. Moreover, we established the use of FSK and FSY for genetically encoded chemical cross-linking to capture elusive enzyme-substrate interactions in live cells, allowing us to target residues aside from Cys and to cross-link at the binding periphery. FSK complements FSY to expand target diversity and versatility. Together, they provide a powerful, genetically encoded, latent bioreactive SuFEx system for creating covalent bonds in diverse proteins *in vitro* and *in vivo*, which will be widely useful for biological research and applications.

Graphical Abstract

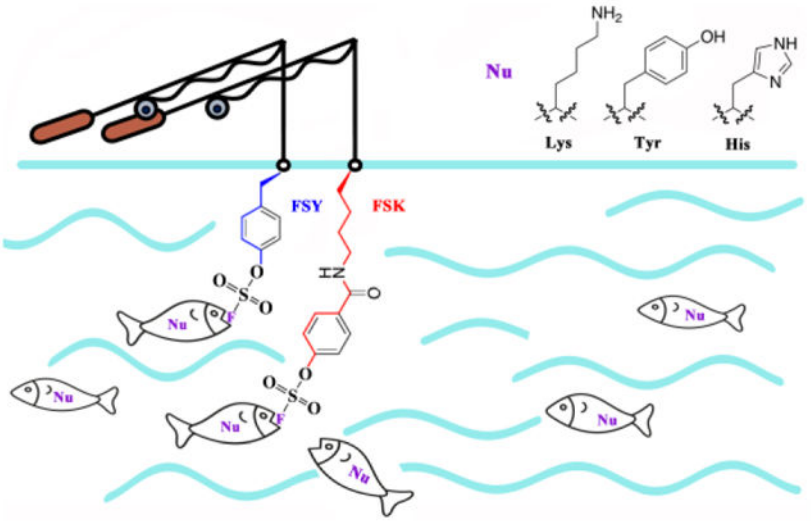
[*] Lei.wang2@ucsf.edu.

Supporting Information

Additional experimental details, materials, and methods; ^{1}H and ^{13}C NMR spectra for FSK; and supplementary figures S1–S13.

Competing Financial Interests Statement

The authors declare no competing financial interests.



INTRODUCTION

Protein side chains can spontaneously form a covalent linkage only via cysteine residues. This natural barrier has been broken in recent years by the addition of genetically encoded, latent bioreactive unnatural amino acids (Uaas) into proteins. Latent bioreactive Uaas can specifically form covalent bonds with nearby natural residues via proximity-enabled reactivity.[1,2] A collection of bioreactive Uaas containing halogen, acrylamide, vinyl sulfone, aryl carbamate, fluorosulfate, or quinone methide moieties have been genetically encoded to target Cys, Lys, His, Tyr, and other nucleophilic residues.[3–8] When engineered to form intramolecular linkages, these new covalent bonds have been used to enhance various protein properties such as fluorescence photon output and thermal stability, as well as to photo-modulate protein structure and function.[1,3,9–11] Covalent linkages can also be formed between proteins, which has been exploited to capture weak and transient protein interactions for identification and analysis.[12,13]

Among the bioreactive functional groups, aryl fluorosulfate is of particular interest due to its exceptional biocompatibility, proximity-dependent reactivity, and multi-targeting capability.[1415161718] It is an excellent latent group which does not indiscriminately react with non-interacting proteins, but reacts efficiently with nucleophilic residues including His, Lys, and Tyr only when they are located in close proximity *in vitro* and in live cells.[7] We recently genetically encoded fluorosulfate-L-tyrosine (FSY) as the first Uaa containing aryl fluorosulfate, and demonstrated its use for creating protein cross-links and generating covalent protein drugs for *in vivo* cancer immunotherapy.[7,19] Nonetheless, as a tyrosine derivative, FSY has a relatively rigid side chain with a length that is similar to canonical amino acids. It has been shown to have a limited reaction radius, which does not allow it to cross-link a target residue located further away in space. Therefore, for a given target protein, the choice of targetable positions near the protein interaction interface is restricted, significantly hindering the general applicability and, when developing a covalent protein drug, the potential to counter mutational resistance with multiple distinct covalent protein binders. In addition, genetically encoded chemical cross-linking (GECX) has great potential

for the identification of unknown weak and transient protein-protein interactions in live cells and *in vivo*. The principle of GECX was initially proven using haloalkane Uaas encoded in the binding site of an enzyme to covalently target Cys residues on substrate proteins.[13] However, free Cys is rare among proteins, and incorporating a Uaa at the binding site may interfere with the protein-protein interaction, limiting the power and scope of GECX. Due to its distinct ability to target multiple nucleophilic residues via the SuFEx reaction, an aryl fluorosulfate group would offer a unique solution to this limitation. Incorporation of an aryl fluorosulfate-containing latent bioreactive Uaa with a long, flexible side chain into the periphery of a binding site would allow for the targeting of a wider range of substrates with different nucleophilic residues without interfering with the binding of the two proteins.

To maximize the capability of using aryl fluorosulfate for generating covalent bonds for various proteins, here we designed and genetically encoded a new latent bioreactive Uaa, fluorosulfonyloxybenzoyl-L-lysine (FSK), which bears an aliphatic side chain extended beyond the length of canonical amino acid side chains, offering greater flexibility and longer reaction distance than FSY. We showed that, for various proteins in *E. coli* and mammalian cells, FSK enabled the formation of covalent linkages both intra- and intermolecularly, whereas FSY was unable to form the same linkages. We generated covalent nanobodies to irreversibly bind with native EGFR receptors on mammalian cell surfaces, and FSK was able to target EGFR positions that were distinct from FSY, creating different covalent nanobody variants. Furthermore, we demonstrated that SuFEx via FSY and FSK incorporation at peripheral sites on an enzyme could be harnessed for GECX in live cells, which directly captured unknown substrate proteins onto the enzyme in *E. coli* for identification, significantly expanding the ability to study elusive protein interactions in their native settings via biocompatible chemistry.

RESULTS AND DISCUSSION

Design and genetic incorporation of FSK into proteins

To afford flexibility and long reaction radius to aryl fluorosulfate, we designed FSK by attaching the aryl fluorosulfate group, which is critical for the biocompatibility and SuFEx reactivity, to the Lys backbone (Figure 1a). As opposed to FSY, FSK contains a side chain extended beyond the canonical amino acids, making it able to reach target residues at various positions and radii from the protein binding interface. We then evolved an orthogonal tRNA/synthetase pair to incorporate FSK in response to the TAG codon via genetic code expansion,[20] using directed evolution strategies described previously.[21,22] A *Methanomethylophilus alvus* PylRS synthetase mutant library was constructed and subjected to several rounds of selection. Four hits were identified which could efficiently incorporate FSK into the enhanced green fluorescence protein (EGFP), rendering cells green fluorescent. Among them, hit 1 (Y126G/M129A/V168F/H227T/Y228P/L229I) incorporated FSK into EGFP at both 18 °C and 30 °C with highest efficiency, and thus was named as FSKRS (Figure S1, S2). Western blot analysis of EGFP (182TAG) expression also showed that full-length EGFP was produced by FSKRS only when FSK was added to the growth media (Figure S3). We further tested FSK incorporation into the superfolder GFP (sfGFP) at permissive site 2 and site 151. At both positions, strong sfGFP fluorescence was detected

when FSK was added to the growth media, in comparison to control samples without FSK (Figure S4), confirming the specificity of FSKRS for the incorporation of FSK. To evaluate the incorporation fidelity, we incorporated FSK into a small protein, ubiquitin (Ub), at position 6 and measured the intact mass of the purified protein with ESI-MS. The measured molecular weight of 9590.1 Da matched well with the theoretical value (9589.9 Da), and no other peaks corresponding to natural amino acid misincorporation were identified, indicating the high fidelity of FSK incorporation by the evolved FSKRS (Figure 1b, 1c, Figure S5).

FSK enables inter-protein cross-linking at distances unreachable with FSY in cells

We investigated the reaction distance preference for FSY and FSK when they were incorporated into proteins and reacting with nucleophilic natural residues in proximity via SuFEx reaction (Figure 2a). At the energy minimized state, the distance between the Cα and the F atom was 9.0 Å for FSY and 13.8 Å for FSK (Figure 2b). Using these lengths as a guide, we tested their inter-protein cross-linking ability on the *E. coli* glutathione transferase (*ec*GST), a homodimeric protein. FSY or FSK was first incorporated into site 103 of *ec*GST at the dimer interface, near which His106 and Lys107 of the other monomer were potential target residues. Based on the *ec*GST crystal structure,[23] the Cα of residue 103 is 7.8 Å from the Nδ atom of His106 and 6.0 Å from the Nε atom of Lys107. Significant *ec*GST dimeric cross-linking was found when FSY was incorporated, but no apparent dimeric cross-linking was detected when FSK was incorporated (Figure S6), suggesting that FSK was not suitable to target nucleophilic residues located too close in the restricted space of the dimer interface.

We next tested the ability of FSK to cross-link target residues that were too far for FSY to reach. We chose to incorporate FSY or FSK at site Glu65 of *ec*GST, around which multiple nucleophilic residues (Lys93, Tyr100, Lys132, Tyr135) reside with a distance to the Cα of Glu65 spanning from 9.2 to 13.3 Å (Figure 2c). This distance should be favorable for FSK to react but too far for FSY. Indeed, after incorporating FSY or FSK into site 65, we found that FSK induced significant *ec*GST dimeric cross-linking but FSY did not (Figure 2d). We also incorporated FSK or FSY into position 86 of *ec*GST, with residues Tyr92 and Tyr72 located 9.5 Å and 11.3 Å away, respectively (Figure S7a). Similar results were obtained: FSK cross-linked *ec*GST to form the dimer, while FSY incorporated at the same position did not induce apparent cross-linking (Figure S7b).

In addition to cross-linking the homodimeric *ec*GST, we also compared the cross-linking abilities of FSY and FSK in a hetero protein interaction complex. *E. coli* thioredoxin (Trx) interacts with 3'-phosphoadenosine-5'-phosphosulfate (PAPS) reductase. We previously found that FSY incorporated at site 60 of *E. coli* Trx could not efficiently cross-link Trx with PAPS reductase.[7] Examining the structure of Trx in complex with PAPS showed that the nearest possible target residue was His242 of PAPS, which was 12.2 Å away from the Cα of residue 60 of Trx.[24] We therefore tested whether incorporating FSK at the same site would improve the cross-linking efficiency due to the long flexible arm of FSK. As could be expected, while only faint cross-linking was detectable for FSY, FSK enabled a robust cross-linking between Trx and PAPS reductase as visualized via Western blot (Figure S8).

Next, we asked which nucleophilic natural residues FSK could react with. We tested its residue specificity using a residue pair Ala97 and Lys44 of *sf*GST protein, which has a

distance of 11.7 Å between the Cα of Ala97 and the Nε of Lys44.[24] We generated a series of mutants by mutating Lys44 into His, Tyr, Ala, Ser, or Thr, and incorporated FSK into position 97 (Figure 2e). Western blot analysis showed that FSK cross-linked with His, Tyr, and Lys residues, forming a stable *sj*GST dimer (Figure 2f). As previously reported, FSY also forms stable cross-links with His, Tyr, and Lys.[7] The consistent reactivity of FSY and FSK is expected since they both contain the same aryl fluorosufate group. Moreover, we also incorporated FSY at site 97 into these mutants but did not observe any apparent dimeric *sj*GST cross-linking (Figure 2f), further corroborating the difference in reactivity radii between FSK and FSY. Taken together, FSK possessed the same multi-targeting reaction specificity as FSY and enabled protein cross-linking at distances unreachable with FSY.

FSK enables covalent bonding of proteins intramolecularly

Genetically introducing intramolecular cross-linking within a peptide or protein is an innovative way to staple or bridge protein residues and can allow for the engineering of properties such as thermostability and cell permeability.[3,25] Current methods mainly rely on disulfide bond formation between two Cys residues or targeting the thiol group of Cys with a halogen or a Michael receptor installed on a bioreactive Uaa.[2] This greatly limits the number of configurations that can be created for the cross-linked peptides or proteins. Since FSK reacts with multiple nucleophilic residues that are more abundant in proteins than Cys, and has a longer reaction arm that allows it to access different configurations, we reasoned that FSK would expand the diversity of cross-linking patterns for genetically encoded intramolecular cross-links of proteins.

As a proof of concept, we investigated the intramolecular cross-linking ability of FSK using Ub. We incorporated FSK into position 18 of Ub, which is proximal to Lys29 (Figure 3a).[26] Mass spectrometric analysis of the intact protein showed a major peak (75% in intensity) with a mass loss of 20 Da, suggesting the successful formation of an intramolecular cross-link (Figure 3b). To determine whether the cross-linking occurred between FSK and Lys29 as predicted from the crystal structure of Ub, we further trypsin digested the cross-linked Ub (18FSK) and analyzed the digested sample with tandem mass spectrometry in high resolution. The cross-linked peptide was identified, and a series of b and y ions of the cross-linked peptide unambiguously indicated that FSK18 reacted with Lys29 in Ub (Figure 3c). Aside from this cross-linked peptide, we also identified the intact FSK-incorporated peptide (Figure S9), which did not react with other peptides randomly, indicating the high specificity of FSK in generating intramolecular protein cross-links.

FSK and FSY incorporated into nanobodies enable the covalent targeting of native EGFR receptor at different positions

The ability to covalently target native receptors on cells and *in vivo* with various protein binders such as antibodies and nanobodies would afford powerful avenues for imaging, diagnostics, and therapeutics.[25, 19, 27] EGFR is a valuable marker for various cancers, so we aimed to covalently target it with nanobodies. Since FSK and FSY have different reaction distances, we should be able to target different positions of EGFR by incorporating FSK or FSY into the nanobody at different sites. Based on the crystal structure of nanobody 7D12 in complex with EGFR,[28] we predicted that incorporation of FSK into site Ser31 of nanobody

7D12 would potentially enable it to cross-link with His359 of EGFR, as the distance from the Cα of Ser31 to the Nδ atom of His359 is 12.7 Å (Figure 4a). Similarly, incorporation of FSY into site Tyr109 of nanobody 7D12 should target Lys443 of EGFR, which has a distance of 8.1 Å from the Cα of Tyr109 to the Nε atom of Lys443 (Figure 4a). This distance would be too close to work for FSK. The two target sites His359 and Lys443 are located at different positions on EGFR (Figure 4a).

To test these predictions, we incorporated FSK into 7D12 in *E. coli* and isolated 7D12 (31FSK) in high purity (Figure S10). Full-length 7D12 nanobody was obtained only when FSK was added in growth media during expression, and mass spectrometric analysis of the purified nanobody confirmed that only FSK was incorporated in high fidelity (Figure 4b). We next assessed cross-linking *in vitro* by incubating 7D12 (31FSK) with EGFR. SDS-PAGE analysis of the incubated samples showed that the EGFR gel band up-shifted almost completely after incubation with 7D12 (31FSK) but not with 7D12 (WT) (Figure 4c, Figure S11). This up-shifted band was confirmed to be the covalent cross-linking of EGFR with the nanobody 7D12 (31FSK) in Western blot analysis (Figure 4d). In addition, we analyzed the incubation sample with high resolution tandem MS, which unambiguously confirmed that FSK incorporated at site 31 of 7D12 cross-linked with His359 of EGFR as predicted (Figure 4e). These results indicate that nanobody 7D12 (31FSK) cross-linked EGFR in high efficiency. In contrast, at site 109 with a shorter cross-linking distance, EGFR was cross-linked by 7D12 (109FSY) but not by 7D12 (109FSK) in SDS-PAGE and Western blot analysis (Figure 4c, 4d). These cross-linking results were consistent with our predications, suggesting that structural distances in combination with FSY/FSK reaction distances would be valuable guidance for rational engineering of FSY/FSK-mediated covalent linkages.

Furthermore, we also assessed the ability of 7D12 (31FSK) to cross-link native EGFR receptors expressed on cancer cell surface. A431, a human epidermoid carcinoma cell line, was incubated with 7D12 (31FSK) or 7D12 (WT). As expected, Western blot analysis of the cell lysates indicated that 7D12 (WT) could not cross-link with the cells, while 7D12 (31FSK) covalently cross-linked with EGFR robustly, with the cross-linking efficiency increased with incubation time from 1 to 8 h (Figure 4f). Together, these results demonstrate that FSK and FSY can act as complements to each other when designing nanobodies with efficient cross-linking abilities at different reaction distances. The irreversible binding of nanobodies to the EGFR receptor through cross-linking distinct sites of the receptor holds great promise for creating novel protein-based diagnostics and therapeutics that work covalently.

FSK incorporation and protein cross-linking via SuFEx in mammalian cells

To show the potential application of FSK in mammalian cells, we tested FSK incorporation and *in vivo* cross-linking in human HeLa cells. Plasmid pNEU-FSKRS expressing the FSKRS and $tRNA^{Pyl}$ was transfected into the HeLa-EGFP (182TAG) reporter cells.[29] Suppression of the 182TAG codon of the genome-integrated EGFP gene would produce full-length EGFP, rendering cells green fluorescent. Strong EGFP fluorescence was observed from cells using confocal microscopy only when FSK was added to the cell culture (Figure 5a). Western blot analysis of cell lysates using anti-EGFP antibody also showed that full-

length EGFP was produced only in cells fed with FSK (Figure 5b), suggesting successful FSK incorporation into EGFP in HeLa cells.

We next explored the ability of FSK to allow protein cross-linking in a mammalian cellular environment. Plasmid pNEU-FSKRS was co-transfected into HEK293T cells with plasmid pcDNA 3.1 expressing *ec*GST(WT), *ec*GST(86TAG), or *ec*GST(86TAG/92A), and cells were grown in the presence of 1 mM FSK (Figure 5c). Cell lysates were analyzed via Western blot to detect *ec*GST dimeric cross-linking. As shown in Figure 5d, incorporation of FSK into site 86 of *ec*GST successfully led to *ec*GST dimeric cross-linking, which was not observed in the control *ec*GST (WT) and was drastically reduced in *ec*GST (86TAG/92A) wherein one target site Tyr92 was mutated to Ala (Figure S7a). These results indicate that FSK could be incorporated into proteins in a mammalian system using the evolved FSKRS and further used for protein cross-linking in mammalian cells.

Identification of Trx interactome in *E. coli* via FSY or FSK-mediated chemical cross-linking

Previously, haloalkane-containing Uaas were incorporated into the active site of an enzyme to probe its substrate proteins, but were only able to target conserved cysteines on the substrates. Since FSK and FSY can target Lys, His, and Tyr residues, we reasoned that they could be used to capture a broader range of interacting proteins that lack Cys but have Lys, His, or Tyr at the interaction interface. In addition, FSY and FSK can be incorporated at the periphery of the binding interface, rather than in the active site or inside the binding interface to minimize potential interference with the protein interaction under study. To further investigate the reaction distance preference of FSK and FSY in a complex protein environment, we explored their application in identifying unknown substrates of an enzyme in live cells via genetically encoded chemical cross-linking (GECX) (Figure. 6a).

Specifically, we incorporated FSK or FSY into site 59 or site 62 of thioredoxin (Trx) in *E. coli* cells. These two sites are away from the Trx active site and likely located in the periphery of the binding interface of Trx and its substrate proteins (Figure 6b). When incorporated at site 59 of Trx, both FSY and FSK dramatically reduced Trx expression in *E. coli*, suggesting that site 59 did not tolerate mutation well. Consequently, few cross-linking bands were detected (Figure 6c). In contrast, site 62 was permissive for mutation, and both FSK- and FSY-incorporated Trx cross-linked substantially more potential substrate proteins (Fig. 6c). These cross-linked proteins were pulled down, digested with trypsin, and analyzed with tandem mass spectrometry. From analysis of cross-linked proteins, we identified 12 substrate proteins for Trx from both the Trx(FSK) sample and the Trx(FSY) sample (Figure 6d). All identified proteins had either Lys, His, or Tyr cross-linked with the Uaa (Figure S12, S13), showing that they directly bind with Trx. Among these substrate proteins, AHPC, TPX, SDHA, HPTG, CH10 are well known Trx substrates previously reported.[30,31] There are a few overlapped substrates of Trx when using FSK or FSY, such as DNAK, APHC, and TPX. However, for the same substrate protein AHPC and DNAK, FSK and FSY cross-linked to different residues (Figure 6d), a result which is consistent with their different reaction radii. Aside from these overlapping substrate proteins, FSK and FSY, when incorporated into Trx, each captured 9 distinct substrates (Figure 6d). These results demonstrate that both FSY and FSK can be used for GECX in live cells, showing distinct

and complementary abilities in identifying substrate proteins in a complex protein environment.

CONCLUSIONS

We report the development of a new aryl fluorosulfate-containing latent bioreactive Uaa, FSK, for covalent bonding of protein residues via proximity-enabled SuFEx reaction *in vitro* and in cells. Where the previously developed FSY could not generate covalent cross-linking due to its shorter reaction radius and relatively rigid side chain, FSK enabled efficient covalent linkage via its longer and more flexible side chain. This expansion allows a significantly broader range of protein sites to be covalently connected, making this technology generally applicable to various proteins. In addition to inter-protein cross-linking, FSK could also be used for intramolecular cross-linking, which will greatly expand the diversity of cross-linking patterns, such as protein bridges and protein staples,[3,11,25] to facilitate the engineering of novel protein properties.

We successfully incorporated FSY and FSK into nanobodies and converted them into covalent binders for EGFR, which irreversibly bound to EGFR *in vitro* and on cancer cell surfaces. These covalent nanobodies may provide novel avenues for cancer imaging and therapeutics. In addition, FSK was able to target positions of EGFR different from those of FSY. Expanding the range of targetable positions on a protein of interest would be invaluable to cope with drug resistance through mutation, which often arises in cancer and infectious diseases.

We also established the use of FSY and FSK in GECX to target His, Lys, and Tyr residues of proteins in live cells. Compared with the initial GECX system, which targeted only Cys, FSY and FSK dramatically expand the type and number of proteins amenable to the GECX approach. Moreover, we further demonstrated that FSY and FSK could be incorporated at the binding periphery, rather than inside the binding interface, to capture protein-protein interactions. Incorporation at the periphery will less likely interfere with the protein interactions under study, and thus improve the efficiency and robustness of GECX. Both FSY and FSK could target His, Lys, and Tyr in proximity. The reactivity is dependent on the microenvironment where the residues are located and the relative orientation and distance of the residues. Due to such protein context dependence, so far we have not found a general trend in reactivity in relation to the target residue identity.

In summary, FSK complements FSY with a longer and more flexible side chain while maintaining a similarly exceptional multi-targeting ability and biocompatibility, leading to enhanced targeting diversity and versatility. Together, they serve as a powerful latent bioreactive system for creating covalent bonds via SuFEx chemistry for diverse proteins *in vitro* and in live cells, which will prove widely useful in basic biological studies as well as protein engineering and biotherapeutic applications.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

S.R. acknowledges the support of the NIH (GM121607); L.W. acknowledges the support of the NIH (R01GM118384).

References

(1). Xiang Z; Ren H; Hu YS; Coin I; Wei J; Cang H; Wang L Adding an Unnatural Covalent Bond to Proteins Through Proximity-Enhanced Bioreactivity. Nat. Methods 2013, 10 (9), 885–888. [PubMed: 23913257]

(2). Wang L Genetically Encoding New Bioreactivity. N. Biotechnol 2017, 38 (Pt A), 16–25. [PubMed: 27721014]

(3). Xiang Z; Lacey VK; Ren H; Xu J; Burban DJ; Jennings PA; Wang L Proximity-Enabled Protein Cross-linking Through Genetically Encoding Haloalkane Unnatural Amino Acids. Angew. Chem. Int. Ed. Engl 2014, 53 (8), 2190–2193. [PubMed: 24449339]

(4). Furman JL; Kang M; Choi S; Cao Y; Wold ED; Sun SB; Smider VV; Schultz PG; Kim CH A Genetically Encoded Aza-Michael Acceptor for Covalent Cross-Linking of Protein-Receptor Complexes. J. Am. Chem. Soc 2014, 136 (23), 8411–8417. [PubMed: 24846839]

(5). Kobayashi T; Hoppmann C; Yang B; Wang L Using Protein-Confined Proximity to Determine Chemical Reactivity. J. Am. Chem. Soc 2016, 138 (45), 14832–14835. [PubMed: 27797495]

(6). Xuan W; Shao S; Schultz PG Protein Cross-linking by Genetically Encoded Noncanonical Amino Acids with Reactive Aryl Carbamate Side Chains. Angew. Chem. Int. Ed. Engl 2017, 56 (18), 5096–5100. [PubMed: 28371162]

(7). Wang N; Yang B; Fu C; Zhu H; Zheng F; Kobayashi T; Liu J; Li S; Ma C; Wang PG; Wang Q; Wang L Genetically Encoding Fluorosulfate-L-Tyrosine to React with Lysine, Histidine, and Tyrosine via SuFEx in Proteins in Vivo. J. Am. Chem. Soc 2018, 140 (15), 4995–4999. [PubMed: 29601199]

(8). Liu J; Li S; Aslam NA; Zheng F; Yang B; Cheng R; Wang N; Rozovsky S; Wang PG; Wang Q; Wang L Genetically Encoding Photocaged Quinone Methide to Multitarget Protein Residues Covalently in Vivo. J. Am. Chem. Soc 2019, 141 (24), 9458–9462. [PubMed: 31184146]

(9). Xuan W; Collins D; Koh M; Shao S; Yao A; Xiao H; Garner P; Schultz PG Site-Specific Incorporation of a Thioester Containing Amino Acid Into Proteins. ACS Chem. Biol 2018, 13 (3), 578–581. [PubMed: 29360343]

(10). Hoppmann C; Lacey VK; Louie GV; Wei J; Noel JP; Wang L Genetically Encoding Photoswitchable Click Amino Acids in Escherichia Coli and Mammalian Cells. Angew. Chem. Int. Ed. Engl 2014, 53 (15), 3932–3936. [PubMed: 24615769]

(11). Hoppmann C; Maslennikov I; Choe S; Wang L In Situ Formation of an Azo Bridge on Proteins Controllable by Visible Light. J. Am. Chem. Soc 2015, 137 (35), 11218–11221. [PubMed: 26301538]

(12). Coin I; Katritch V; Sun T; Xiang Z; Siu FY; Beyermann M; Stevens RC; Wang L Genetically Encoded Chemical Probes in Cells Reveal the Binding Path of Urocortin-I to CRF Class B GPCR. Cell 2013, 155 (6), 1258–1269. [PubMed: 24290358]

(13). Yang B; Tang S; Ma C; Li ST; Shao GC; Dang B; DeGrado WF; Dong MQ; Wang PG; Ding S; Wang L Spontaneous and Specific Chemical Cross-Linking in Live Cells to Capture and Identify Protein Interactions. Nat. Commun 2017, 8 (1), 2240. [PubMed: 29269770]

(14). Dong J; Krasnova L; Finn MG; Sharpless KB Sulfur(VI) Fluoride Exchange (SuFEx): Another Good Reaction for Click Chemistry. Angew. Chem. Int. Ed. Engl 2014, 53 (36), 9430–9448. [PubMed: 25112519]

(15). Chen W; Dong J; Plate L; Mortenson DE; Brighty GJ; Li S; Liu Y; Galmozzi A; Lee PS; Hulce JJ; Cravatt BF; Saez E; Powers ET; Wilson IA; Sharpless KB; Kelly JW Arylfluorosulfates Inactivate Intracellular Lipid Binding Protein(S) Through Chemoselective SuFEx Reaction with a Binding Site Tyr Residue. J. Am. Chem. Soc 2016, 138 (23), 7353–7364. [PubMed: 27191344]

(16). Jones LH Emerging Utility of Fluorosulfate Chemical Probes. ACS Med. Chem. Lett 2018, 9 (7), 584–586. [PubMed: 30034581]

(17). Zheng Q; Woehl JL; Kitamura S; Santos-Martins D; Smedley CJ; Li G; Forli S; Moses JE; Wolan DW; Sharpless KB SuFEx-Enabled, Agnostic Discovery of Covalent Inhibitors of Human Neutrophil Elastase. Proc. Natl. Acad. Sci. U. S. A 2019, 116 (38), 18808–18814. [PubMed: 31484779]

(18). Liu Z; Li J; Li S; Li G; Sharpless KB; Wu P SuFEx Click Chemistry Enabled Late-Stage Drug Functionalization. J. Am. Chem. Soc 2018, 140 (8), 2919–2925. [PubMed: 29451783]

(19). Li Q; Chen Q; Klauser PC; Li M; Zheng F; Wang N; Li X; Zhang Q; Fu X; Wang Q; Xu Y; Wang L Developing Covalent Protein Drugs via Proximity-Enabled Reactive Therapeutics. Cell 2020, 182 (1), 85–97.e16. [PubMed: 32579975]

(20). Wang L; Brock A; Herberich B; Schultz PG Expanding the Genetic Code of Escherichia Coli. Science 2001, 292 (5516), 498–500. [PubMed: 11313494]

(21). Takimoto JK; Dellas N; Noel JP; Wang L Stereochemical Basis for Engineered Pyrrolysyl-tRNA Synthetase and the Efficient in Vivo Incorporation of Structurally Divergent Non-Native Amino Acids. ACS Chem. Biol 2011, 6 (7), 733–743. [PubMed: 21545173]

(22). 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. Am. Chem. Soc 2018, 140 (28), 8807–8816. [PubMed: 29984990]

(23). Nishida M; Harada S; Noguchi S; Satow Y; Inoue H; Takahashi K Three-Dimensional Structure of Escherichia Coli Glutathione S-Transferase Complexed with Glutathione Sulfonate: Catalytic Roles of Cys10 and His106. J. Mol. Biol 1998, 281 (1), 135–147. [PubMed: 9680481]

(24). Chartron J; Shiau C; Stout CD; Carroll KS 3′-Phosphoadenosine-5′-Phosphosulfate Reductase in Complex with Thioredoxin: a Structural Snapshot in the Catalytic Cycle. Biochemistry 2007, 46, 3942–3951. [PubMed: 17352498]

(25). Chen XH; Xiang Z; Hu YS; Lacey VK; Cang H; Wang L Genetically Encoding an Electrophilic Amino Acid for Protein Stapling and Covalent Binding to Native Receptors. ACS Chem. Biol 2014, 9 (9), 1956–1961. [PubMed: 25010185]

(26). Cook WJ; Jeffrey LC; Carson M; Chen Z; Pickart CM Structure of a Diubiquitin Conjugate and a Model for Interaction with Ubiquitin Conjugating Enzyme (E2). J. Biol. Chem 1992, 267 (23), 16467–16471. [PubMed: 1322903]

(27). Berdan VY; Klauser PC; Wang L Covalent Peptides and Proteins for Therapeutics. Bioorg. Med. Chem 2021, 29, 115896. [PubMed: 33285408]

(28). Schmitz KR; Bagchi A; Roovers RC; van Bergen en Henegouwen PMP; Ferguson KM Structural Evaluation of EGFR Inhibition Mechanisms for Nanobodies/VHH Domains. Structure 2013, 21 (7), 1214–1224. [PubMed: 23791944]

(29). Wang W; Takimoto JK; Louie GV; Baiga TJ; Noel JP; Lee K-F; Slesinger PA; Wang L Genetically Encoding Unnatural Amino Acids for Cellular and Neuronal Studies. Nat. Neurosci 2007, 10 (8), 1063–1072. [PubMed: 17603477]

(30). Kumar JK; Tabor S; Richardson CC Proteomic Analysis of Thioredoxin-Targeted Proteins in Escherichia Coli. Proc. Natl. Acad. Sci. U. S. A 2004, 101 (11), 3759–3764. [PubMed: 15004283]

(31). Lu J; Holmgren A The Thioredoxin Antioxidant System. Free Radic. Biol. Med 2014, 66, 75–87. [PubMed: 23899494]

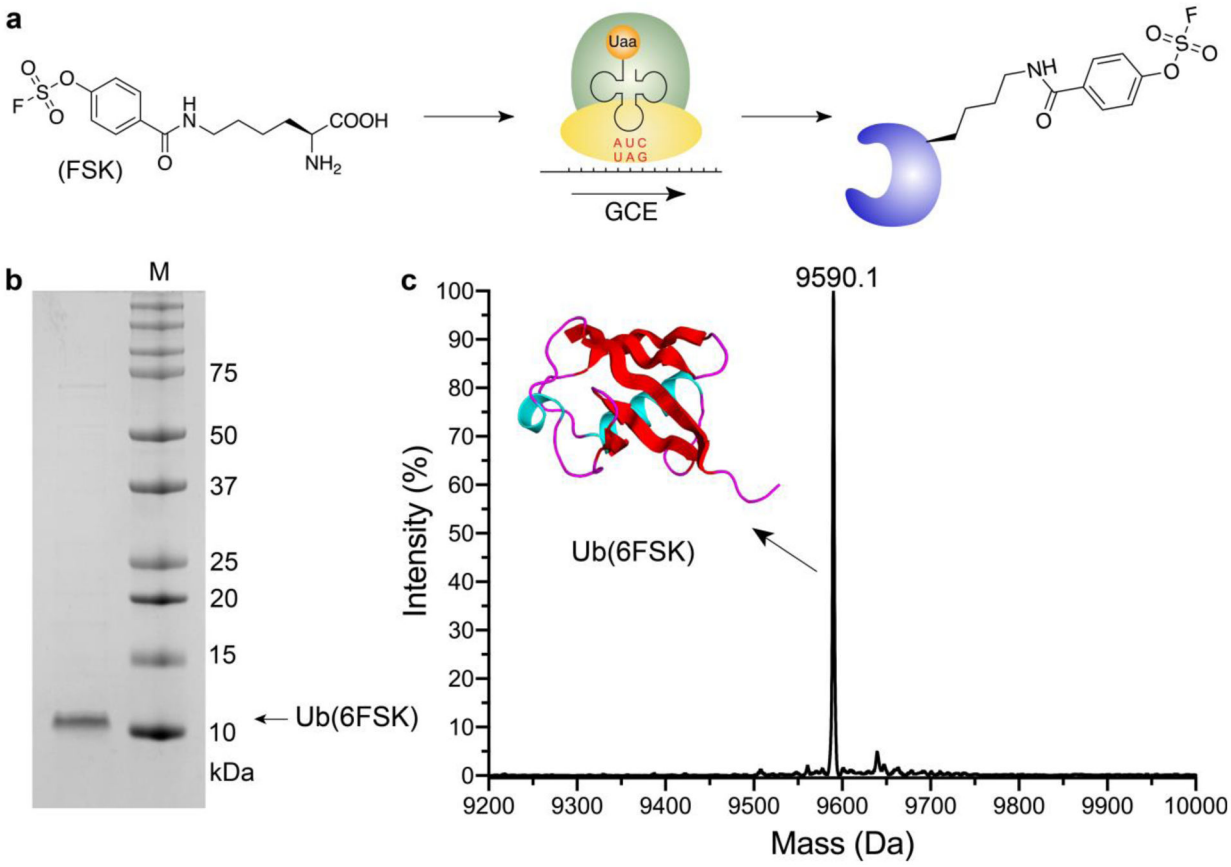


Figure 1. Site-specific incorporation of FSK into proteins via genetic code expansion.

(a) Schematic illustration of FSK structure and incorporation via genetic code expansion. **(b)** SDS-PAGE of purified ubiquitin (6FSK). **(c)** Mass spectrum of the intact ubiquitin (6FSK). Theoretical molecular weight: 9589.9 Da; observed: 9590.1 Da.

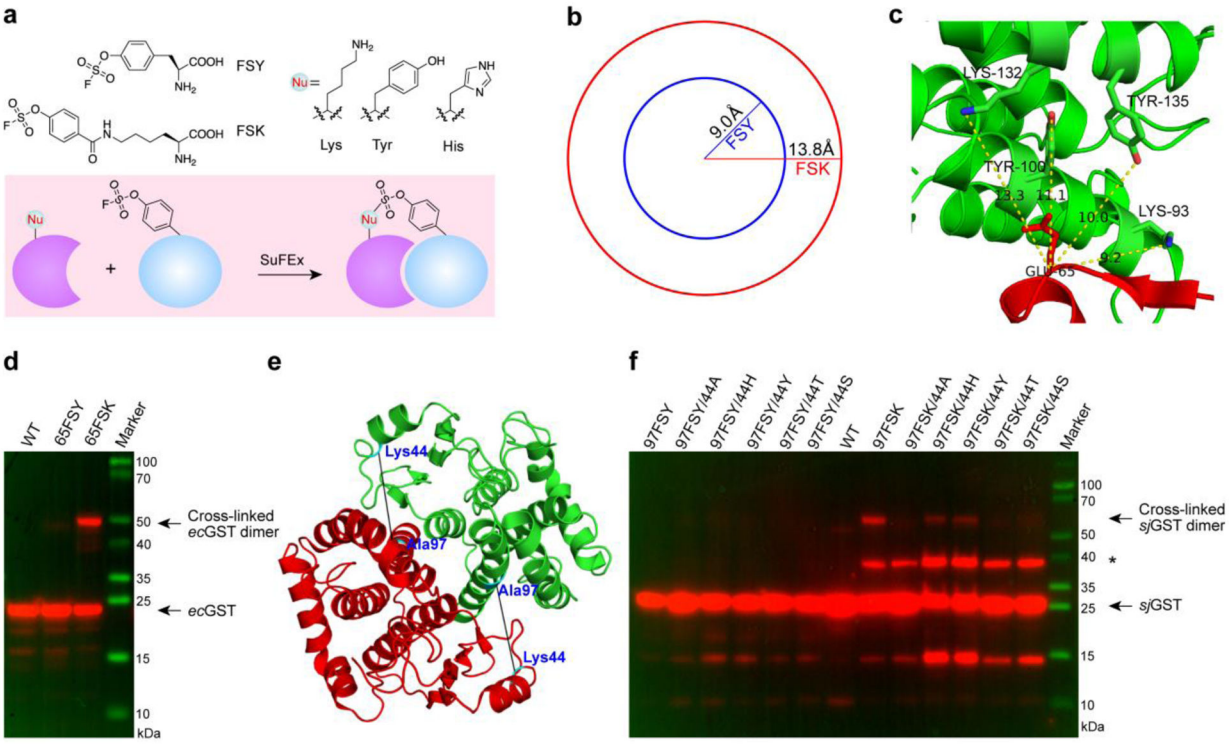


Figure 2. Genetically encoded FSK enables protein cross-linking at longer distances unreachable with FSY.

(a) Chemical structures of FSY and FSK, and schematic illustration of aryl fluorosulfate reacting with nucleophilic residues in proximity via SuFEx chemistry. **(b)** Reaction distances of FSY and FSK measured from the Cα to the F atom at their energy minimized states. **(c)** Crystal structure of *ec*GST (PDB code: 1A0F) showing the distances between Glu65 and the adjacent nucleophilic residues in yellow dotted lines. **(d)** Western blot analysis of *ec*GST dimeric cross-linking induced by FSY or FSK incorporated at site 65 of *ec*GST. **(e)** Crystal structure of *sj*GST (PDB code: 1Y6E) showing the distance between the Cα atoms of Lys 44 and Ala97). **(f)** Western blot analysis of *sj*GST dimeric cross-linking mediated by FSY or FSK. * indicates other proteins interacting with *sj*GST in *E. coli*.

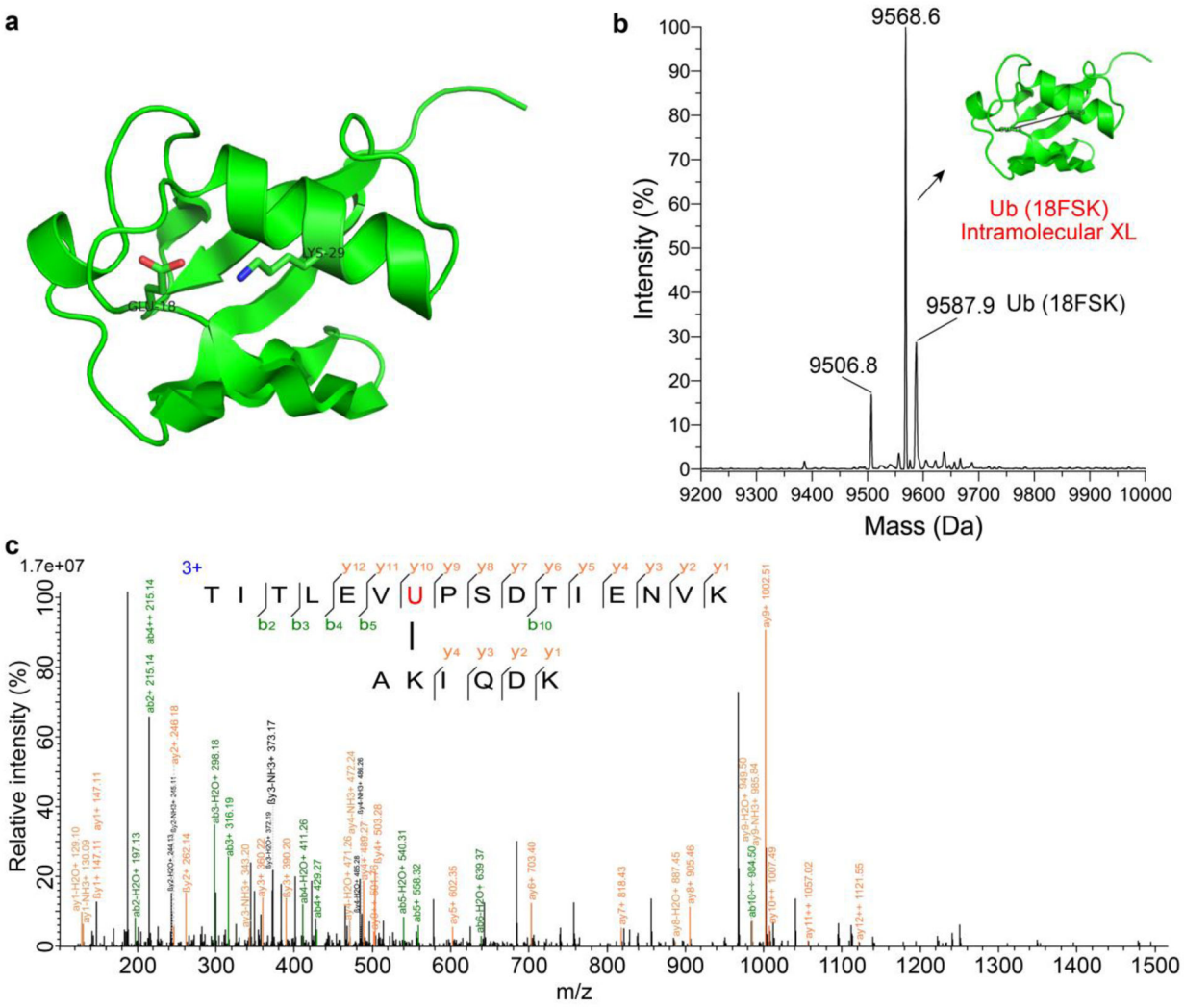


Figure 3. FSK-mediated intramolecular cross-linking in ubiquitin (Ub).

(a) Structure of Ub (PDB code: 1AAR) showing Glu18 for FSK incorporation to target Lys29. **(b)** ESI-MS of Ub (18FSK). The peak of 9587.9 Da corresponds to the intact Ub(18FSK) (calculated MW: 9587.9 Da). The peak of 9568.6 Da corresponds to the intramolecularly cross-linked Ub via FSK18 reacting with Lys29 and losing HF (calculated MW: 9567.9). The peak 9506.8 Da corresponds to Ub (18FSK) losing SO_2F (calculated MW: 9506.9). **(c)** Tandem mass spectrum of the cross-linked peptide identified from the trypsin-digested Ub (18FSK), showing clearly that FSK (indicated with red U) reacted with Lys29 as designed.

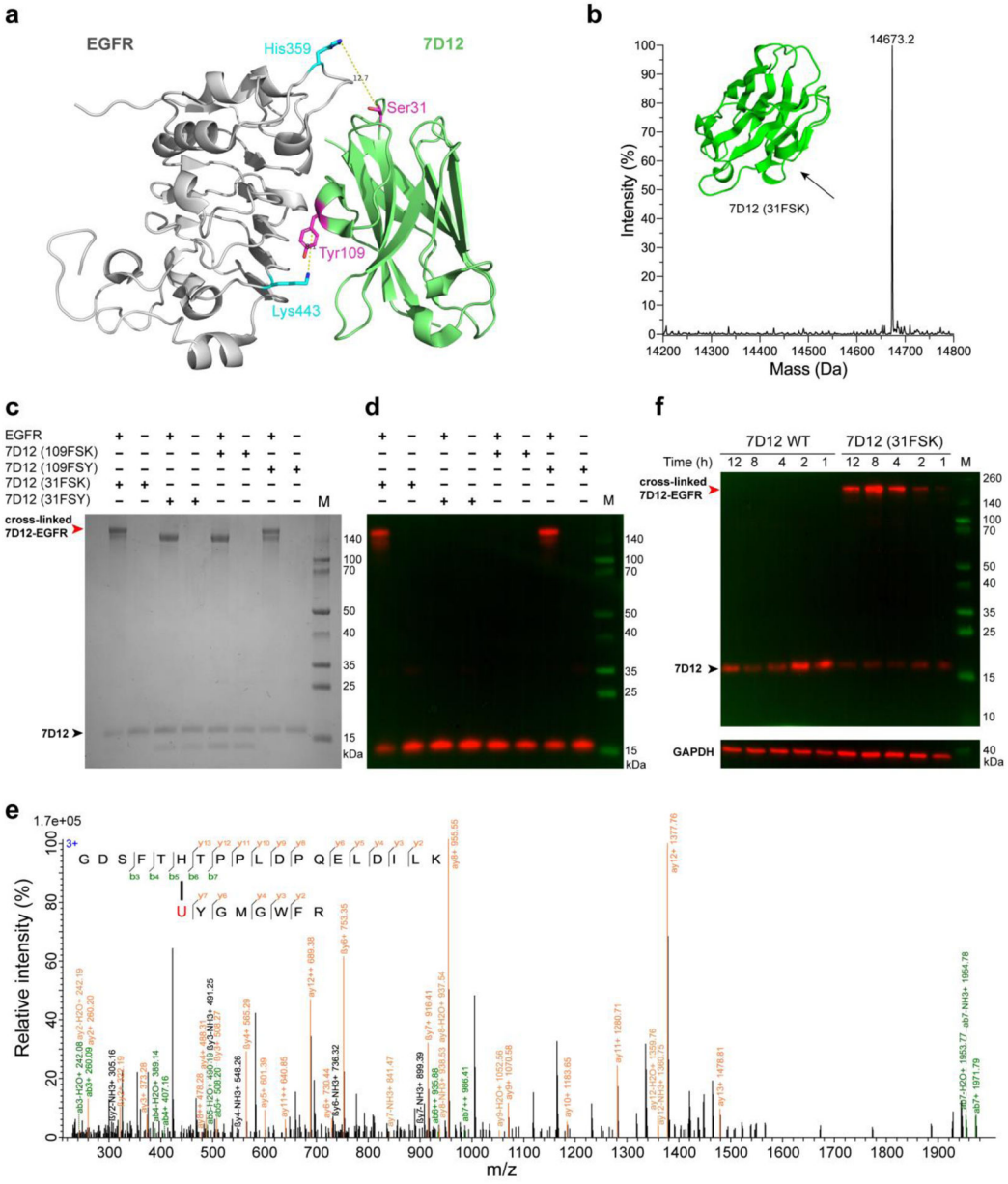


Figure 4. FSK enabled 7D12 nanobody to covalently target the EGFR receptor at positions different from FSY.

(a) Structure of nanobody 7D12 in complex with EGFR (PDB code: 4KRL), showing Ser31 of 7D12 for FSK incorporation to target His359 of EGFR and Tyr109 for FSY incorporation to target Lys443 of EGFR. Distances are labeled between the Cα of the incorporation site to the nucleophilic N atom of the target residue. **(b)** ESI-MS analysis of 7D12 (31FSK) confirmed FSK incorporation in high fidelity. Calculated MW: 14673.1 Da (forming 1 pair of disulfide bond); measured MW: 14673.2 Da. **(c)** SDS-PAGE analysis of covalent cross-linking of nanobody 7D12 with EGFR *in vitro*. **(d)** Western blot analysis of covalent cross-linking of nanobody 7D12 with EGFR *in vitro*. Results for 7D12 (WT) control were shown in Figure S11. (**e**) Tandem mass spectrum of 7D12 (31FSK) incubation with EGFR confirmed that FSK (represented by U) of 7D12 nanobody cross-linked with His359 of

EGFR as designed. **(f)** Western blot analysis of nanobody 7D12 cross-linking with native EGFR expressed on A431 cell surface. Nanobody 7D12 (WT) or 7D12 (31FSK) was incubated with A431 cells for the indicated time periods, and cell lysates were immunoblotted with anti-His antibody to detect the nanobody 7D12.

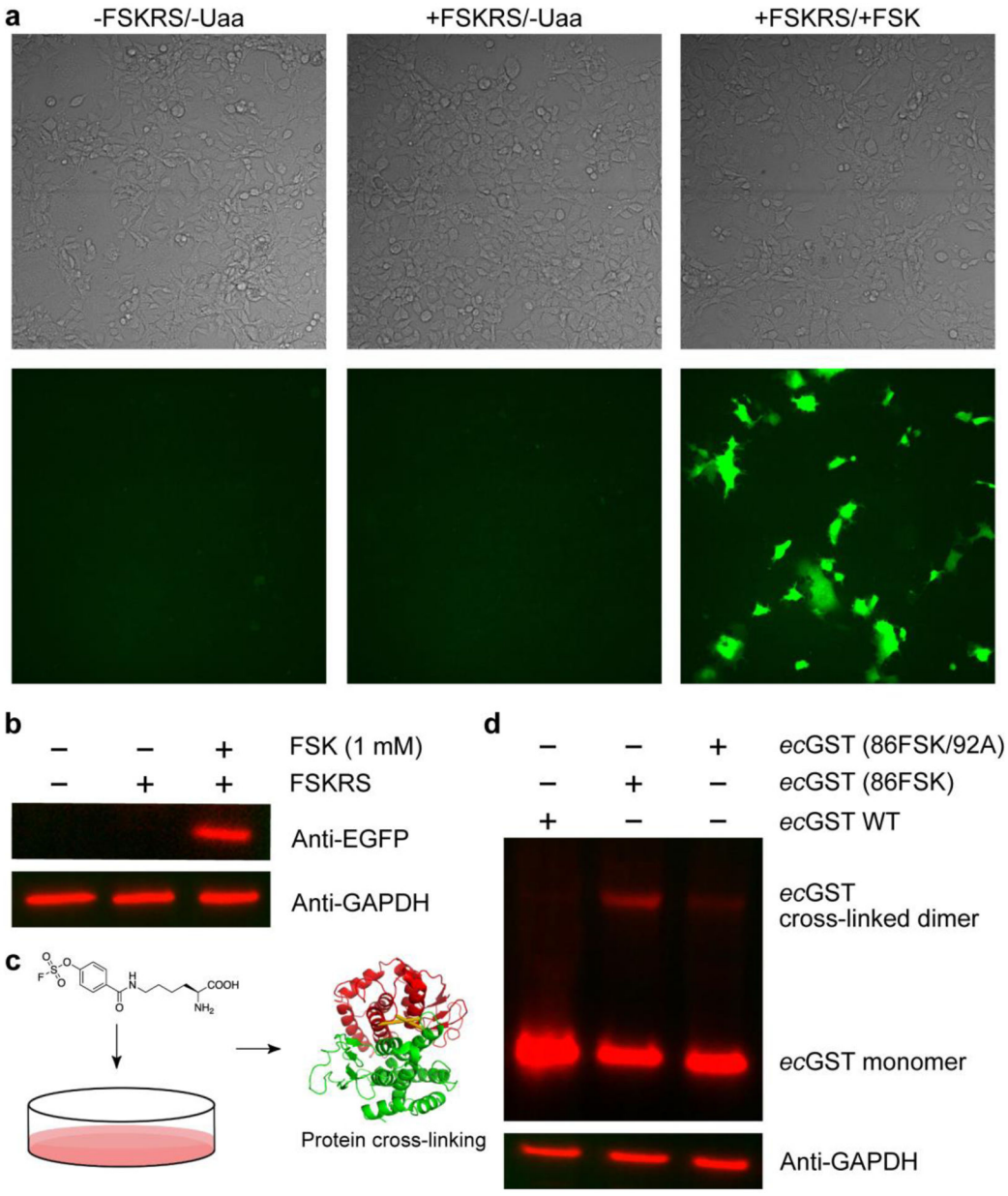
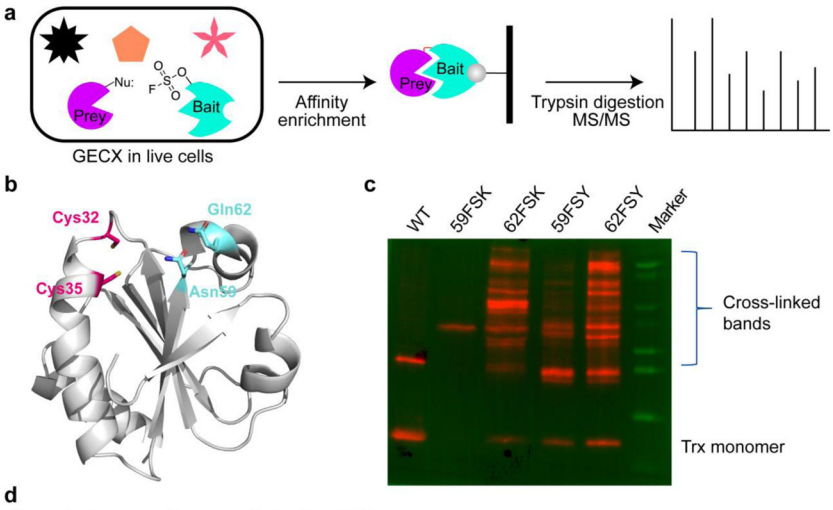


Figure 5. Genetic incorporation of FSK into proteins for protein cross-linking in mammalian cells.

(a) Fluorescence microscopic images of HeLa-EGFP (182TAG) reporter cells under different conditions. Cells were transfected with or without pNEU-FSKRS, and grown with or without 1 mM FSK. Top: bright field; bottom: EGFP fluorescence channel. **(b)** Western blot analysis of FSK incorporation into EGFP in HeLa cells. Samples from (a) were lysed and detected using an anti-EGFP antibody. GAPDH expression level was used as reference. (**c**) Schematic illustration of genetic incorporation of FSK into mammalian cells for protein cross-linking. **(d)** Western blot analysis of FSK-mediated *ec*GST cross-linking in mammalian cells. pNEU-FSKRS was co-transfected with pcDNA3.1-ecGST (WT), *ec*GST (86TAG), or *ec*GST (86TAG/92A) into HEK293T cells. The dimeric cross-linking of *ec*GST was detected using an anti-His antibody. GAPDH was used as a reference.



d

Trx substrate proteins cross-linked by FSK

Protein	Description	Cross-linked peptide	Score
sp|P0AC41|SDHA	Succinate dehydrogenase flavoprotein subunit	AAGLHLQESIAEQGALR	28.58
sp|P0A6Y8|DNAK	Chaperone protein DnaK	IELSSAQQTDVNLPYITADATGPK	26.30
sp|P0AE08|AHPC	Alkyl hydroperoxide reductase C	WKEGEATLAPSLDLVGK	25.84
sp|P69503|APT	Adenine phosphoribosyltransferase	DVTSLLEDPKAYALSIDLLVER	14.42
sp|P0A862|TPX	Thiol peroxidase	SQTVHFQGNPVTVANSIPQAGSK	21.72
sp|P0AF93|RIDA	2-iminobutanoate/2-iminopropanoate deaminase	TIATENAPAAIGPYVQGVDLGNMIITSGQIPVNPK	14.22
sp|P0A6Z3|HTPG	Chaperone protein HtpG	HIAHDFNDPLTWSHNR	13.58
sp|P0A6L0|DEOC	Deoxyribose-phosphate aldolase	ASEISIKAGADFIKTSTGK	10.96
sp|P0A725|LPXC	UDP-3-O-acyl-N-acetylglucosamine deacetylase	LLQAVLAKQEAWEYVTFQDDAELPLAFK	14.22
sp|P33362|YEHZ	Glycine betaine-binding protein YehZ	VAADYLKQK	10.06
sp|P77481|YCJV	Putative uncharacterized ABC transporter ATP-binding protein YcjV	LTIPEEKLAVLK	10.04
sp|P16679|PHNL	Alpha-D-ribose 1-methylphosphonate 5-triphosphate synthase subunit PhnL	GAAIVGIFHDEAVRNDVADR	9.40

Trx substrate proteins cross-linked by FSY

Protein	Description	Cross-linked peptide	Score
sp|P0A6Y8|DNAK	Chaperone protein DnaK	IELSSAQQTDVNLPYITADATGPK	27.53
sp|P0A6Y8|DNAK	Chaperone protein DnaK	AKLESLVEDLVNR	15.51
sp|P0AE08|AHPC	Alkyl hydroperoxide reductase C	AAQYVASHPGEVCPAK	10.91
sp|P0AE08|AHPC	Alkyl hydroperoxide reductase C	EGEATLAPSLDLVGKI	19.83
sp|P0A862|TPX	Thiol peroxidase	SQTVHFQGNPVTVANSIPQAGSK	23.64
sp|P0A799|PGK	Phosphoglycerate kinase	SVNDVKADEQILDIGDASAQELAEILK	22.53
sp|P08312|SYFA	Phenylalanine--tRNA ligase alpha subunit	TMKAQQPPIR	18.95
sp|P39199|PRMB	50S ribosomal protein L3 glutamine methyltransferase	HEPELGLASGTDGLKLTR	15.17
sp|P06720|AGAL	Alpha-galactosidase	TAHIALMDIDPTRLEESHIVVR	14.22
sp|P0A6F9|CH10	10 kDa chaperonin	VGDIVIFNDGYGVK	12.40
sp|P0A6X1|HEM1	Glutamyl-tRNA reductase	EADIIISSTASPLPIIGKGMVER	12.14
sp|P0AAB8|USPD	Universal stress protein D	HNDAHLTLIHIDDGLSELYPGIYFPATEDILQLLK	22.02

Figure 6. FSK and FSY enable GECX in *E. coli* cells to capture and identify Trx substrates.
(a) Schematic illustration of using FSK or FSY to capture Trx substrate proteins through GECX in live cells followed with MS identification. (**b**) Trx structure (PDB 1XOB) showing the two active site Cys residues (Cys 32 and 35) and two selected peripheral sites (Asn59 and Gln62) for FSK/FSY incorporation. (**c**) Western blot analysis of cell lysates of *E. coli* cells expressing the indicated Trx proteins. An anti-Hisx6 antibody was used to detect the Hisx6 tag appended at the C-terminus of Trx. **(d)** Identified substrate proteins of Trx and their peptides cross-linked by FSK or FSY. Red: cross-linked residues; Blue: Cys alkylated by 2-iodoacetamide.