Title

Residue selective crosslinking of proteins through photoactivatable or proximity-enabled reactivity.

Authors

Liu, Jun
Yang, Bing
Wang, Lei

Publication Date

2023-06-01

DOI

10.1016/j.cbpa.2023.102285

Peer reviewed

HHS Public Access

Author manuscript

Curr Opin Chem Biol. Author manuscript; available in PMC 2024 June 01.

Published in final edited form as:

Curr Opin Chem Biol. 2023 June ; 74: 102285. doi:10.1016/j.cbpa.2023.102285.

Residue Selective Crosslinking of Proteins through Photoactivatable or Proximity-Enabled Reactivity

Jun Liu[#],

Bing Yang[#],

Lei Wang[*]

Department of Pharmaceutical Chemistry, the Cardiovascular Research Institute, and Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA, 94158, U.S.A.

Abstract

Photo- and chemical crosslinking of proteins have offered various avenues for studying protein structure and protein interactions with biomolecules. Conventional photoactivatable groups generally lack reaction selectivity toward amino acid residues. New photoactivatable groups reacting with selected residues have emerged recently, increasing crosslinking efficiency and facilitating crosslink identification. Traditional chemical crosslinking usually employs highly reactive functional groups, while recent advance has developed latent reactive groups with reactivity triggered by proximity, which reduce spurious crosslinks and improve biocompatibility. The employment of these residue selective chemical functional groups, activated by light or proximity, in small molecule crosslinkers and in genetically encoded unnatural amino acids is summaried. Together with new software development in identifying protein crosslinks, residue selective crosslinking has enhanced the research of elusive protein-protein interactions in vitro, in cell lysate, and in live cells. Residue selective crosslinking is expected to expand to other methods for the investigation of various protein-biomolecule interactions.

Introduction

Photo-crosslinking of proteins via small molecule crosslinkers, genetically encoded unnatural amino acids (Uaas), or photoaffinity ligand labeling has contributed significantly to our understanding of protein interactions with proteins and various biomolecules [1,2]. Conventional photo-crosslinking mainly uses photoactivatable groups such as aryl azide, benzophenone (BPA), and diazirine, which react through the radical mechanism and thus lack reaction selectivity toward amino acid residues [3]. Such inclusive reactivity, although advantageous for mapping unknown interactions, increases background crosslinking and drastically complicates crosslink analysis via mass spectrometry (MS). The radical intermediates can be readily quenched or convert to less reactive species, often resulting in short half-life (ns to μs) and low crosslinking efficiency. In recent years, a new type of photoactivatable cross-linking groups is emerging, which photo-generates non-radical

[*]:Corresponding author: Lei Wang (Lei.Wang2@ucsf.edu).
[#]:Equal contribution

intermediates with longer half-lives and reaction specificity toward a single or multiple amino acid residues.

Chemical crosslinking does not depend on an external trigger and is often residue selective. Small molecule chemical crosslinkers have been widely used in chemical crosslinking mass spectrometry for studying protein structure and protein-biomolecule interactions [4,5]. Exemplified by *N*-hydroxysuccinimide (NHS) ester reacting with amine, these reactions are generally highly reactive, affording high crosslinking efficiency. However, the high reactivity often leads to spurious cross-links between residues brought into proximity by rare thermal fluctuations, leading to inaccurate distance information. High reactivity of crosslinkers also prevents their usage in living systems. In recent years, a new type of chemical crosslinkers has been developed to have latent reactivity, which is activated only when the functional group is brought into close proximity to its target [6]. Such proximity-enabled reactivity reduces spurious crosslinks, expands the repertoire of targetable amino acid residues, and shows better biocompatibility. In addition, latent bioreactive Uaas have been designed and genetically encoded in proteins, which crosslink various natural amino acid residues via proximity-enabled reactivity [7]. These latent bioreactive Uaas allow genetically encoded chemical crosslinking (GECX) both in vitro and in live cells, opening novel avenues for studying protein interactions.

In this review, we briefly summarize residue selective chemical crosslinking, with reactivity activated by light or by proximity. We focus on small molecule crosslinkers and genetically encoded latent bioreactive Uaas that bear such reactivities. Recent progress in software development for the identification of chemical crosslinks from MS data is also described.

Photoactivatable small molecule chemical crosslinkers

A new type of photoactivatable chemical groups has emerged, which, upon UV light irradiation, releases a non-radical functional group that react with certain amino acid side chains selectively. The released functional group generally has longer half-life than radicals, which can increase the efficiency for protein crosslinking or photoaffinity labeling (Figure 1a-b). One photoreactive group is tetrazole (Figure 1c). Photolysis of tetrazole releases a highly reactive nitrile immine, which undergoes nucleophilic reaction with nearby nuclephiles with preference for the carboxyl group [8,9]. Li et al. showed that aryl tetrazole-derivatized staurosporines photo-crosslink with PKA through the photogenerated nitrile imine reacting with the side chain of Glu [8]. Similarly, Herner et al. demonstrated that 2-aryl-5-carboxytetrazole (ACT) photo-releases a carboxynitrile imine. ACT-attached small molecule drugs crosslink with their target proteins through carboxynitrile imine reacting with a Glu near the target active site [9]. Another photoreactive group is 2-nitrobenzyl alcohol (Figure 1d), which photogenerates an aldehyde to couple with amines. Wang et al. demonstrated that 2-nitrobenzyl alcohol derived D-mannose photoaffinity labels concanavalin A at Lys39 [10]. A heterobifunctional crosslinker consisting of 2-nitrobenzyl alcohol and *N*-hydroxyl succinimide (NHS) photocrosslinks proteins with efficiency higher than the NHS-aryl azide and NHS-BPA [10].

Expanding the number of amino acid residues targetable by the chemical crosslinker would increase the abundance of crosslinks for MS identification and the success rate of photoaffinity labeling when target proteins are unknown. Quinone methide (QM) is an efficient Michael acceptor for various nucleophiles in chemical synthesis [11]. We recently developed a series of photocaged QM-based chemical crosslinkers (Figure 1e), which photogenerates QM capable of crosslinking the side chains of nine amino acid residues (Asp, Glu, Lys, Ser, Thr, Tyr, Gln, Arg, and Asn) [12]. The heterobifunctional NHQM (consisting of NHS and photocaged QM) crosslinker can be used to crosslink proteins *in vitro*, in *E. coli* cell lysate, and on mammalian cell surface, and to crosslink protein with its bound DNA *in vitro*. In addition, the homobifunctional crosslinker HoQM (with photocaged QM at both ends) is able to crosslink dimeric proteins inside live *E. coli* and mammalian cells. Subsequently, Huang et al. reported the use of azidobenzyl-caged QM for proximal labeling of proteins with biotin through $Ir(ppy)_2bpy$-mediated photocatalysis [13].

These photoactivatable chemical crosslinkers, owing to their reaction selectivity, have their crosslinked amino acid residues readily identifiable via tandem MS. With added photo responsivenss, they should be valuable for the investigation of ligand-protein and protein-protein interactions, as well as for structural biology through chemical crosslinking mass spectrometry. Most chemical crosslinkers are currently incompatible for use in living systems; this challenge can be overcome through genetically encoding Uaas.

Genetically encoded photoactivatable chemical crosslinking Uaas

Genetic code expansion allows Uaas harboring diverse functional groups to be incorporated into proteins site-specifically in living systems [14]. Photo-crosslinking Uaas, such as those containing aryl azide, BPA, or diazirine have been genetically encoded and proven useful to capture protein-protein interactions and offer spatiotemporal control for crosslinking [2]. However, the nonspecific reactivity of their radical intermediates produces complex crosslinked products and unpredictable crosslinking sites, which drastically complicate MS data analysis. The short half-life (ns to μs) of the radicals often result in low crosslinking yield, which negatively impacts the capture of weak and transient protein-protein interactions.

Recently, a new class of photoactivatable chemical crosslinking Uaas have been designed and genetically encoded into proteins (Figure 2a). Similar to the mechanisms described in the previous section, these Uaas, after incorporation into proteins and upon UV light activation, release chemical groups that can selectively react with one or multiple amino acid residues in proteins. These Uaas thus enable photo-crosslinking of proteins through selectively targeting certain amino acid side chains, facilitating the identification of the crosslink via MS and often increasing crosslinking efficiency.

On the basis of the ACT photochemistry, Tian et al. designed ACT-containing lysine analogues and incorporated them into proteins in *E. coli* and mammalian cells [15]. One of these Uaas, mPyTK (Figure 2b), allows photo-crosslinking of a GST dimer in vitro and in *E. coli*, exhibiting higher crosslinking efficiency than a diazirine-based photo-crosslinking Uaa. Mutagenesis study suggests that mPyTK crosslinks with Glu92 at the GST dimer

interface. Incorporation of mPyTK in Grb2 also enables the crosslinking with EGFR in mammalian cells. Hu et al. developed an o-nitrobenzyl alcohol containing lysine derivative (o-NBAK) for crosslinking with proximal lysine residues (Figure 2c) [16]. The o-NBAK photogenerates aryl-nitroso intermediates, which have half-lives of ~30 min and react with Lys within GST dimer *in vitro* and in *E. coli* cells. They futher showed that o-NBAK-incorporated SlPatA, a lysine acetyltransferase, captures its substrate protein SeAcs via photoactivated o-NBAK reacting with Lys, as verified by MS.

Amino acid identities at protein-protein interface vary with proteins due to diverse molecular interactions. It is thus valuable to develop photoactivatable Uaas capable of targeting a broader range of natural amino acid residues specifically. Such multi-targeting reactivity is particually useful when identifying unknown protein−protein interactions wherein no information on the target natural residue is available. We recently designed and genetically encoded a photoactivatable *para*-QM generating Uaa, FnbY, into proteins in E. coli and mammalian cells (Figure 2d). Using GST as the model protein, we showed that the incorporated FnbY, in response to UV light, is able to crosslink multiple nucleophilic residues, including Cys, His, Lys, Tyr, Trp, Asn, Gln, Asp, Glu, and Met [17]. The half-life of QM is on the order of seconds, much longer than that of radical intermediates of the conventional photocrosslinking Uaas. Correspondingly, FnbY shows higher crosslinking efficiency than the phenyl azide-based *p*-azido-phenylalanine. We also genetically encoded another photoactivatable Uaa FmnbY [18] (Figure 2d), which photoreleases an *ortho*-QM, showing similar selectivity but higher reactivity than *para*-QM of FnbY. In addition to protein crosslinking, FnbY and FmnbY can both be used for photocontrolled protein conjugation with the widely available amine or thiol reagents, featuring fast kinetics (reaction completes in tens of seconds) [18].

Proximity-enabled small molecule chemical crosslinkers

Warhead reactivity of small molecule crosslinkers is important for protein cross-linking. While poor reactivity may not effectively capture transient protein-protein interactions, excessively high reactivity will cause random non-specific cross-linking, which can lead to high background or inaccurate distance restraints for protein structure modeling. Yang et al. designed a heterobifunctional cross-linker NHSF composed of a highly reactive succinimide ester and a weakly reactive aryl sulfonyl fluoride [6] (Figure 3). Succinimide ester reacts with lysine rapidly, planting the reagent at specific localizations on protein. When the pendant aryl sulfonyl fluoride is cast proximal to the target residues, the increased effective concentration enables aryl sulfonyl fluoride to react with multiple weakly nucleophilic amino acid residues. This proximity-enhanced reaction decreases the nonspecific crosslinks and exhibits a better structural compatibility than the traditional bis-succinimide ester crosslinker. In addition, NHSF is shown to form crosslink between Lys and His, Ser, Thr, Tyr, and Lys sidechains, significantly expanding the diversity of residues targetable by chemical crosslinkers.

This proximity-enabled chemical crosslinking strategy has recently been applied to protein–carbohydrate complex to identify what functional group can react with carbohydrate under biocompatible conditions [19]. After planting various chemical crosslinkers on proteins

via succinimide ester reacting with lysine, the other end of the crosslinker is then tested to crosslink the bound carbodydrate. Aryl sulfonyl fluoride is discovered to react with carbohydrate via proximity-enabled reactivity, offering a novel solution to the low affinity between proteins and carbohydrates.

Proximity-enabled genetically encoded chemical crosslinking (GECX)

GECX via latent bioreactive Uaa enables the capture of weak and transient protein-protein interactions in live cells for identification with MS [20] (Figure 4a). The incorporated latent bioreactive Uaa reacts with a target natural residue only when the two residues are in close proximity. This unique mechanism enables the covalent crosslinking to occur between two interacting proteins specifically, without creating non-specific crosslinks that lead to false positive and complicated MS analysis. As no exogenous trigger is needed, GECX can capture the interacting protein whenever they are in action in live cells and accumulate the crosslinked product for MS identification with high sensitivity. The latent reactivity, genetic encodability, and the spontaneous yet specific crosslinking ability make GECX uniquely suitable for applications in vivo.

A series of latent bioreactive Uaas have been devised and genetically encoded into proteins to covalently target different amino acid residues through proximity-enabled reactivity. These Uaas have been reviewed elsewhere [7,21,22], so we herein focus on their GECX applications in protein complex and protein-biomolecule interactions (Figure 4b).

The first latent bioreactive Uaa, *p*-2'-fluoroacetylphenylalanine (F_{fact}), was designed to use the weak electrophilic fluoromethyl ketone to selectively react with Cys [23]. F_{fact} shows no reactivity to free Cys under physiological conditions, but reacts with the proximal Cys after incorporated into proteins and placed in appropriate orientation. This proximity-enabled F_{fact}-Cys reaction has been used to enable irreversible binding of an affibody with its substrate protein, to intramolecularly attach the chromophore of fluorescent proteins to their β-barrel for enhanced photon output [23], and to pinpoint how the peptide agonist urocortin-1 binds to the class B GPCR CRF-1R on live mammalian cell surface [24]. These work open the door to genetically encoding latent bioreactive Uaas for covalently targeting proteins in vitro and in vivo.

Haloalkane-bearing Uaas with varying aliphatic chain length and halogen atom have been developed to react with Cys, His, or Lys [25,26]. The intra-protein crosslinks increases protein thermostability [26], and the BrC6K-incorporated affibody irreversibly crosslinks native HER2 receptor on mammalian cell surface [25]. Subsequently, bromoalkyl-containing BrCnK Uaas, through reacting with a mutated Cys, were used to stabilize the low-affinity protein complex Rab1b:GDP:DrrA [27]. After crosslinking, unambiguous electron density for GDP became visible in the nucleotide-binding pocket.

Yang et al. firstly established GECX in live cells to capture weak and transient protein interactions including enzyme–substrate binding [20]. Bromoalkyl-containing Uaa BprY and its alkyn derivative EB3 for enrichment via click chemistry were designed to target the Cys residue. GECX via these Uaas enables crosslinking of weak protein binding of

affibody–Z (K_d ~6 μM) directly in *E. coli* cells, exhibiting higher crosslinking efficiency than the conventional photocrosslinker *p*-azido-phenylalanine. In addition, substrate protein PCNA, with EB3 incorporated around the ubiquitination site K164, is able to crosslink the ubiquitin-conjugating enzyme UBE2D3 in live cells. Moreover, incorporation of BprY in the active site of enzyme Trx1 allows the covalent capture of Trx1 substrate proteins in *E. coli* cells. A total of 91 endogenous proteins were identified by MS. Therefore, GECX has the ability to uncover substrate proteins of enzymes and vice versa in live cells.

Tang et al. subsequently incorporated the fluoromethyl ketone-containing FpAcF (a.k.a. F_{fact} [23]) and the bromoalkly-containing Uaa BetY [26] into substrate proteins at their phospho-tyrosine (pY) site to crosslink the conserved active-site Cys of protein tyrosine phosphatases (PTPs) [28]. When incorporated at site pY1023 of HER2, the known PTP SHP2 was crosslinked in HEK293 cells as detected by Western blot. Interestingly, when BetY was incorporated at pY1221 of HER2 in HEK293 cells, 116 proteins were identified by MS, among which only one PTB, PTP1B, was found.

Shu et al. uncovered a broader reactivity of the bromoalkyl-containing BprY [29]. In addition to Cys, His, and Lys, BprY is found also reacting with Tyr, Ser, Thr, Asp, and Glu in proximity although in lower efficiency. They incorporated BprY at residues adjacent to the SIM-binding groove of SUMO2 and identified 264 SUMO2 interaction proteins in 293T cell lysates.

Efficiently targeting noncatalytic and non-Cys residues would expand the scope of GECX. A series of Uaas capable of proximity-enabled sulfur-fluoride exchange (SuFEx) reaction have been designed and genetically encoded in *E. coli* and mammalian cells. Wang et al. initially demonstrated that fluorosulfate-L-tyrosine (FSY) is able to react with proximal His, Lys, and Tyr, forming stable intra- and inter-protein linkages in vitro and in live cells [30]. These residues are located at protein surface and noncatalytic. FSY was found to also react with Ser and Thr, yet converting the latter into dehydroalanine and dehydrobutyrine, respectively [31]. Klauser et al. genetically encoded meta-fluorosulfate-L-tyrosine (mFSY), which has fluorosulfate at the *meta* position vs. the *para* position in FSY [32]. mFSY shows comparable multi-targeting reactivity and can reach certain target orientations unreachable by FSY. Yu et al. genetically encoded a fluorine-substituted fluorosulfate-L-tyrosine (FFY), which accelerates the SuFEx crosslinking rate 2.4-fold over FSY for generating covalent nanobodies that efficiently neutralize SARS-CoV-2 and variants [33]. Liu et al. developed fluorosulfonyloxybenzoyl-L-lysine (FSK) to install fluorosulfate on a long and flexible side chain, extending the reaction radius and enabling crosslinking at distance unreachable with FSY [34]. Moreover, they incorporated FSY or FSK in Trx to covalently capture Trx substrate proteins in *E. coli* cells (Figure 4c). In contrast to BprY and EB3 which had to be incorporated at Trx active site to target the the conserved Cys on the substrate proteins, FSY and FSK can be incorporated at the periphery of the Trx binding site since they can target non-Cys residues His/Lys/Tyr, which are abundant at protein surface. This flexibility allows more substrate proteins with direct crosslinks identified by MS [34], and enables GECX generally applicable to broader enzyme-substrate systems that lack a targetable Cys residue.

Xu et al. genetically incorporated N^2-carboxy-4-aryl-1,2,3-triazole-lysines (CATKs) and showed that CATK-1 crosslinks Lys and Tyr in high efficiency while exhibiting low crosslinking efficiency toward Cys and His [35]. The authors concluded that CATK-1 possesses higher crosslinking reactivity than FPheK (which uses the same acyl transfer mechanism [36]) and FSY. Nonetheless, proximity-enabled crosslinking reactivity is both distance and orientation dependent, as shown in the comparison among FSY, FSK, and mFSY [32,34]. Due to the differences in side chain length and flexibility, a site optimal for CATK-1 crosslinking may not be optimal for FSY and vice versa.

Aside from GECX of proteins with proteins, recently GECX-RNA has been developed for genetically encoded chemical crosslinking of proteins with interacting RNA [37] and GECX-sugar for crosslinking of proteins with interacting carbohydrates [19], expanding the scope of biomolecules covalently targetable via the principle of proximity-enabled reactivity.

Conclusions and perspective

Chemical functional groups that are selectively reactive with amino acid side chains, either activated by UV light or proximity, have increasingly been developed for chemical crosslinking of proteins. They have been installed on small molecule crosslinkers and on genetically encodable Uaas. Combined with mass spectrometry, these residue selective chemical crosslinkers have proven valuable for the identification of elusive protein-protein interactions in vitro, in cell lysate, and in live cells. Compared with conventional photo-crosslinkers reacting through radical mechanisms, these residue selective crosslinkers have longer reaction window to increase crosslinking yield and predictable crosslinks to facilitate MS identification. In addition, enzymatic activation has also been developed to trigger residue specific reactivities for proximity-based protein labeling [38,39]. We thus foresee the increasing application of these residue selective crosslinkers over the traditional non-specific photo-crosslinkers.

Analysis of crosslinked protein complex with MS facilitates the discovery of unknown protein interactions [40]. Identification of crosslinked peptides offers evidence for direct (rather than indicrect) interaction and further helps to map the interaction interface. Mapping crosslinked peptides generated by conventional photo-crosslinkers has been challenging due to the relative low crosslinking efficiency and complicated crosslinks resultant from non-specific reactivity. Recently, software SpotLink was developed to identify non-specific crosslinks of peptides at the proteome scale, identifying over 3000 cross-links from human proteome database [41]. We expect that Spotlink has the potential to identify protein crosslinks generated by conventional photo-crosslinking Uaas.

To identify peptides crosslinked by residue selective chemical crosslinking Uaas, regular database searching software such as Mascot and Maxquant can be theoretically used [42,43], but the entire or partial fragment ions of Uaa-incorporated peptides will be missed, which decreases the identification sensitivity. Small molecular cross-linking identification software can also analyze Uaa mediated cross-linking. However, different from small molecule crosslinking, one chain of Uaa mediated cross-linked peptides is known and fixed before database searching [44]. Liu et al. developed a new database search engine, OpenUaa, to

analyze Uaa–mediated crosslinking at the proteomic scale [45]. Compared with traditional search engines which consider Uaa-incorporated peptide as a large modification and miss the Uaa fragment ions, OpenUaa preserves the information of all fragments. As the Uaa-incorporated peptide is known before database search, OpenUaa first identifies the Uaa-incorporated peptides via in silico digestion, which reduces the redundancy of candidate peptide generation; these spectra are then searched against a global protein database to identify the second peptides. This inclusive and open search capability drastically improves identification sensitivity and coverage. For instance, integrating GECX with OpenUaa, the direct interactome of Trx in *E. coli* cells is identified, yielding 289 crosslinked peptides and corresponding to 205 direct binding protein of Trx [45]. Improvement of search engine is thus invaluable to effectively identify peptide crosslinks from MS data of complex and sophisticated protein crosslinking samples.

Protein cross-linking mediated by small molecular chemical cross-linkers tends to generate more intra-protein crosslinks than inter-protein crosslinks, with the former competitively suppressing the latter. To study direct protein interactions and protein interacting interfaces, inter-protein cross-linked peptides are more informative. New cross-linking strategies and reagents are thus desired to enhance inter-protein cross-linking efficiency. Genetically encoded latent bioreactive Uaas can also form both intra- and inter-protein crosslinks, depending on whether a target residue is in proximity or not. Selection of incorporation site and/or mutation of interfering residues can help generate one type of crosslinks over the other.

With further optimization we expect that small molecule chemical crosslinkers will be more compatible for use in live cells. As proximity-enabled reactivity allows for rational engineering of covalent linkages in proteins, we expect that GECX will increasingly be used for engineering novel protein properties that require strong, stable, and irreversible bonding. The in vivo compatibility, as demonstrated by FSY-incorporated protein drugs [46], will foster further exploration of covalent therapeutics and diagnostics. Lastly, we expect that these chemical functional groups to be introduced into biomolecules through other methods such as enzyme-mediated tagging [47] and metabolic incorporation [48], which will further expand their applications.

Acknowledgements

L.W. acknowledges the support of NIH (R01GM118384 and R01CA258300).

References

1. Burton NR, Kim P, Backus KM: Photoaffinity labelling strategies for mapping the small molecule–protein interactome. Org Biomol Chem 2021, 19:7792–7809. [PubMed: 34549230]
2. Pham ND, Parker RB, Kohler JJ: Photocrosslinking approaches to interactome mapping. Curr Opin Chem Biol 2013, 17:90–101. [PubMed: 23149092]
3. Hassan MM, Olaoye OO: Recent Advances in Chemical Biology Using Benzophenones and Diazirines as Radical Precursors. Molecules 2020, 25:2285. [PubMed: 32414020]
4. Piersimoni L, Kastritis PL, Arlt C, Sinz A: Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein–Protein Interactions—A Method for All Seasons. Chem Rev 2022, 122:7500–7531. [PubMed: 34797068]

5. Yu C, Huang L: Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. Anal Chem 2017, 90:144–165. [PubMed: 29160693]

6. Yang B, Wu H, Schnier PD, Liu Y, Liu J, Wang N, DeGrado WF, Wang L: Proximity-enhanced SuFEx chemical cross-linker for specific and multitargeting cross-linking mass spectrometry. Proc Natl Acad Sci USA 2018, 115:11162–11167. [PubMed: 30322930] ** The crosslinking group sulfonyl fluoride with low but proximity-enhanced reactivity was used in small molecule crosslinker, which reduces spurious crosslinks, exhibits a better structural compatibility, and expands the diversity of residues targetable for chemical cross-linking mass spectrometry of proteins.

7. Cao L, Wang L: New covalent bonding ability for proteins. Protein Sci 2022, 31:312–322. [PubMed: 34761448] * An overview of genetically encoded latent bioreactive unnatural amino acids, the development history, and associated various applications.

8. Li Z, Qian L, Li L, Bernhammer JC, Huynh HV, Lee J-S, Yao SQ: Tetrazole Photoclick Chemistry: Reinvestigating Its Suitability as a Bioorthogonal Reaction and Potential Applications. Angew Chem Int Ed 2016, 55:2002–2006.

9. Herner A, Marjanovic J, Lewandowski TM, Marin V, Patterson M, Miesbauer L, Ready D, Williams J, Vasudevan A, Lin Q: 2-Aryl-5-carboxytetrazole as a New Photoaffinity Label for Drug Target Identification. J Am Chem Soc 2016, 138:14609–14615. [PubMed: 27740749]

10. Wang C, Liu Y, Bao C, Xue Y, Zhou Y, Zhang D, Lin Q, Zhu L: Phototriggered labeling and crosslinking by 2-nitrobenzyl alcohol derivatives with amine selectivity. Chem Commun 2020, 56:2264–2267.

11. Toteva MM, Richard JP: The Generation and Reactions of Quinone Methides. Advances in physical organic chemistry 2011, 45:39–91. [PubMed: 24511169]

12. Liu J, Cai L, Sun W, Cheng R, Wang N, Jin L, Rozovsky S, Seiple IB, Wang L: Photocaged Quinone Methide Crosslinkers for Light-Controlled Chemical Crosslinking of Protein–Protein and Protein–DNA Complexes. Angew Chem Intl Ed 2019, 58:18839–18843.

13. Huang Z, Liu Z, Xie X, Zeng R, Chen Z, Kong L, Fan X, Chen PR: Bioorthogonal Photocatalytic Decaging-Enabled Mitochondrial Proteomics. J Am Chem Soc 2021, 143:18714–18720. [PubMed: 34709827]

14. Wang L, Brock A, Herberich B, Schultz PG: Expanding the Genetic Code of *Escherichia coli*. Science 2001, 292:498–500. [PubMed: 11313494]

15. Tian Y, Jacinto MP, Zeng Y, Yu Z, Qu J, Liu WR, Lin Q: Genetically Encoded 2-Aryl-5-carboxytetrazoles for Site-Selective Protein Photo-Cross-Linking. J Am Chem Soc 2017, 139:6078–6081. [PubMed: 28422494] * This work describes the genetic incorporation of a photoactivatable chemical crosslinking Uaa that selectively crosslinks with Glu.

16. Hu W, Yuan Y, Wang C-H, Tian H-T, Guo A-D, Nie H-J, Hu H, Tan M, Tang Z, Chen X-H: Genetically Encoded Residue-Selective Photo-Crosslinker to Capture Protein-Protein Interactions in Living Cells. Chem 2019, 5:2955–2968. * A photoactivatable chemical crosslinking Uaa selectively reacting with Lys is developed. Genetic incorporation of this Uaa in an enzyme to capture its substrate protein is also demonstrated.

17. Liu J, Li S, Aslam NA, Zheng F, Yang B, Cheng R, Wang N, Rozovsky S, Wang PG, Wang Q, et al. : Genetically Encoding Photocaged Quinone Methide to Multitarget Protein Residues Covalently in Vivo. J Am Chem Soc 2019, 141:9458–9462. [PubMed: 31184146] * Two photoactivatable chemical crosslinking Uaas are developed, which exhibit cross-linking selectively toward nine nucleophilic amino acid residues, significanly expanding the number of residues targetable.

18. Liu J, Cheng R, Van Eps N, Wang N, Morizumi T, Ou W-L, Klauser PC, Rozovsky S, Ernst OP, Wang L: Genetically Encoded Quinone Methides Enabling Rapid, Site-Specific, and Photocontrolled Protein Modification with Amine Reagents. J Am Chem Soc 2020, 142:17057–17068. [PubMed: 32915556]

19. Li S, Wang N, Yu B, Sun W, Wang L: Genetically encoded chemical crosslinking of carbohydrate. Nat Chem 2022, doi:10.1038/s41557-022-01059-z.

20. Yang B, Tang S, Ma C, Li S-T, Shao G-C, Dang B, DeGrado WF, Dong M-Q, Wang PG, Ding S, et al. : Spontaneous and specific chemical cross-linking in live cells to capture and identify protein interactions. Nat Commun 2017, 8:2240. [PubMed: 29269770] ** First report of GECX in live

cells, establishing the principle and feasibility of capturing weak and transient protein interactions through genetically encoded chemical crosslinking. The ability of GECX to capture the substrates of an enzyme or the enzyme of a substrate protein is demonstrated.

21. Wang N, Wang L: Genetically encoding latent bioreactive amino acids and the development of covalent protein drugs. Curr Opin Chem Biol 2022, 66:102106.
22. Wang L: Genetically encoding new bioreactivity. N Biotech 2017, 38:16–25.
23. 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, 10:885–888. [PubMed: 23913257] ** This report firstly establishes the proximity-enabled reactivity concept and genetically encodes latent bioreactive Uaa to covalently target a natural amino acid residue for generating covalent linkages within and between proteins.
24. 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, 155:1258–1269. [PubMed: 24290358]
25. Chen X-H, 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, 9:1956–1961. [PubMed: 25010185]
26. Xiang Z, Lacey VK, Ren H, Xu J, Burban DJ, Jennings PA, Wang L: Proximity-Enabled Protein Crosslinking through Genetically Encoding Haloalkane Unnatural Amino Acids. Angew Chem Int Ed 2014, 53:2190–2193.
27. Cigler M, Müller TG, Horn-Ghetko D, von Wrisberg M-K, Fottner M, Goody RS, Itzen A, Müller MP, Lang K: Proximity-Triggered Covalent Stabilization of Low-Affinity Protein Complexes In Vitro and In Vivo. Angew Chem Intl Ed 2017, 56:15737–15741.
28. Tang H, Dai Z, Qin X, Cai W, Hu L, Huang Y, Cao W, Yang F, Wang C, Liu T: Proteomic Identification of Protein Tyrosine Phosphatase and Substrate Interactions in Living Mammalian Cells by Genetic Encoding of Irreversible Enzyme Inhibitors. J Am Chem Soc 2018, 140:13253–13259. [PubMed: 30247891]
29. Shu X, Asghar S, Yang F, Li S-T, Wu H, Yang B: Uncover New Reactivity of Genetically Encoded Alkyl Bromide Non-Canonical Amino Acids. Front Chem 2022, 10:815991.
30. Wang N, Yang B, Fu C, Zhu H, Zheng F, Kobayashi T, Liu J, Li S, Ma C, Wang PG, et al. : Genetically Encoding Fluorosulfate- l -tyrosine To React with Lysine, Histidine, and Tyrosine via SuFEx in Proteins *in Vivo*. J Am Chem Soc 2018, 140:4995–4999. [PubMed: 29601199] ** FSY was designd and genetically encoded, which possesses exceptional biocompatibility and multi-targeting ability, reacting with Lys/His/Tyr through click chemistry SuFEx directly in live cells. FSY has been successfully used in a variety of applications including the development of covalent protein drugs for cancer immunotherapy in mice.
31. Yang B, Wang N, Schnier PD, Zheng F, Zhu H, Polizzi NF, Ittuveetil A, Saikam V, DeGrado WF, Wang Q, et al. : Genetically Introducing Biochemically Reactive Amino Acids Dehydroalanine and Dehydrobutyrine in Proteins. J Am Chem Soc 2019, 141:7698–7703. [PubMed: 31038942]
32. Klauser PC, Berdan VY, Cao L, Wang L: Encoding latent SuFEx reactive meta-fluorosulfate tyrosine to expand covalent bonding of proteins. Chem Commun 2022, 58:6861–6864.
33. Yu B, Li S, Tabata T, Wang N, Cao L, Kumar GR, Sun W, Liu J, Ott M, Wang L: Accelerating PERx reaction enables covalent nanobodies for potent neutralization of SARS-CoV-2 and variants. Chem 2022, 8:2766–2783. [PubMed: 35874165]
34. Liu J, Cao L, Klauser PC, Cheng R, Berdan VY, Sun W, Wang N, Ghelichkhani F, Yu B, Rozovsky S, et al. : A Genetically Encoded Fluorosulfonyloxybenzoyl- l -lysine for Expansive Covalent Bonding of Proteins via SuFEx Chemistry. J Am Chem Soc 2021, 143:10341–10351. [PubMed: 34213894] * A lysine deriavtive containing a long and flexible aryl fluorosulfate moiety was genetically encoded to complement FSY to expand target diversity and versatility. FSY and FSK were used in GECX to capture the substrate proteins of Trx in E. coli cells, demonstrating that GECX can target non-active site and non-Cys residues.
35. Xu Y, Rahim A, Lin Q: Spontaneous Orthogonal Protein Crosslinking via a Genetically Encoded 2-Carboxy-4-Aryl-1,2,3-Triazole. Angew Chem Int Ed 2022, 61:e202202657.

36. Xuan W, Shao S, Schultz PG: Protein Crosslinking by Genetically Encoded Noncanonical Amino Acids with Reactive Aryl Carbamate Side Chains. Angew Chem Intl Ed 2017, 56:5096–5100.

37. Sun W, Wang N, Liu H, Yu B, Jin L, Ren X, Shen Y, Wang L: Genetically encoded chemical crosslinking of RNA in vivo. Nat Chem 2022, doi:10.1038/s41557-022-01038-4.

38. Ke M, Yuan X, He A, Yu P, Chen W, Shi Y, Hunter T, Zou P, Tian R: Spatiotemporal profiling of cytosolic signaling complexes in living cells by selective proximity proteomics. Nat Commun 2021, 12:71. [PubMed: 33397984]

39. Dubovetskyi A, Cherukuri KP, Ashani Y, Meshcheriakova A, Reuveny E, Ben-Nissan G, Sharon M, Fumagalli L, Tawfik DS: Quinone Methide-Based Organophosphate Hydrolases Inhibitors: *Trans* Proximity Labelers versus *Cis* Labeling Activity-Based Probes. ChemBioChem 2021, 22:894–903. [PubMed: 33105515]

40. Zheng J, Chen X, Yang Y, Tan CSH, Tian R: Mass Spectrometry-Based Protein Complex Profiling in Time and Space. Anal Chem 2021, 93:598–619. [PubMed: 33146515]

41. Zhang W, Gong P, Shan Y, Zhao L, Hu H, Wei Q, Liang Z, Liu C, Zhang L, Zhang Y: SpotLink enables sensitive and precise identification of site nonspecific cross-links at the proteome scale. Brief Bioinformatics 2022, 23:bbac316.

42. Cox J, Mann M: MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 2008, 26:1367–1372. [PubMed: 19029910]

43. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS: Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 1999, 20:3551–3567. [PubMed: 10612281]

44. Yang B, Wu Y-J, Zhu M, Fan S-B, Lin J, Zhang K, Li S, Chi H, Li Y-X, Chen H-F, et al. : Identification of cross-linked peptides from complex samples. Nat Methods 2012, 9:904–906. [PubMed: 22772728]

45. Liu C, Wu T, Shu X, Li S, Wang DR, Wang N, Zhou R, Yang H, Jiang H, Hendriks IA, et al. : Identification of Protein Direct Interactome with Genetic Code Expansion and Search Engine OpenUaa. Adv Biology 2021, 5:2000308. ** Software OpenUaa is the first software dedicated to searching Uaa-mediated protein crosslinks. OpenUaa significantly improves the sensitivity and coverage in identifying cross-linked proteins out of complex biological samples. In addition, OpenUaa is also able to identify endogenous protein crosslinks in same high quality.

46. Li Q, Chen Q, Klauser PC, Li M, Zheng F, Wang N, Li X, Zhang Q, Fu X, Wang Q, et al. : Developing Covalent Protein Drugs via Proximity-Enabled Reactive Therapeutics. Cell 2020, 182:85–97.e16. [PubMed: 32579975]

47. Milczek EM: Commercial Applications for Enzyme-Mediated Protein Conjugation: New Developments in Enzymatic Processes to Deliver Functionalized Proteins on the Commercial Scale. Chem Rev 2018, 118:119–141. [PubMed: 28627171]

48. Pedowitz NJ, Pratt MR: Design and synthesis of metabolic chemical reporters for the visualization and identification of glycoproteins. RSC Chem Biol 2021, 2:306–321. [PubMed: 34337414]

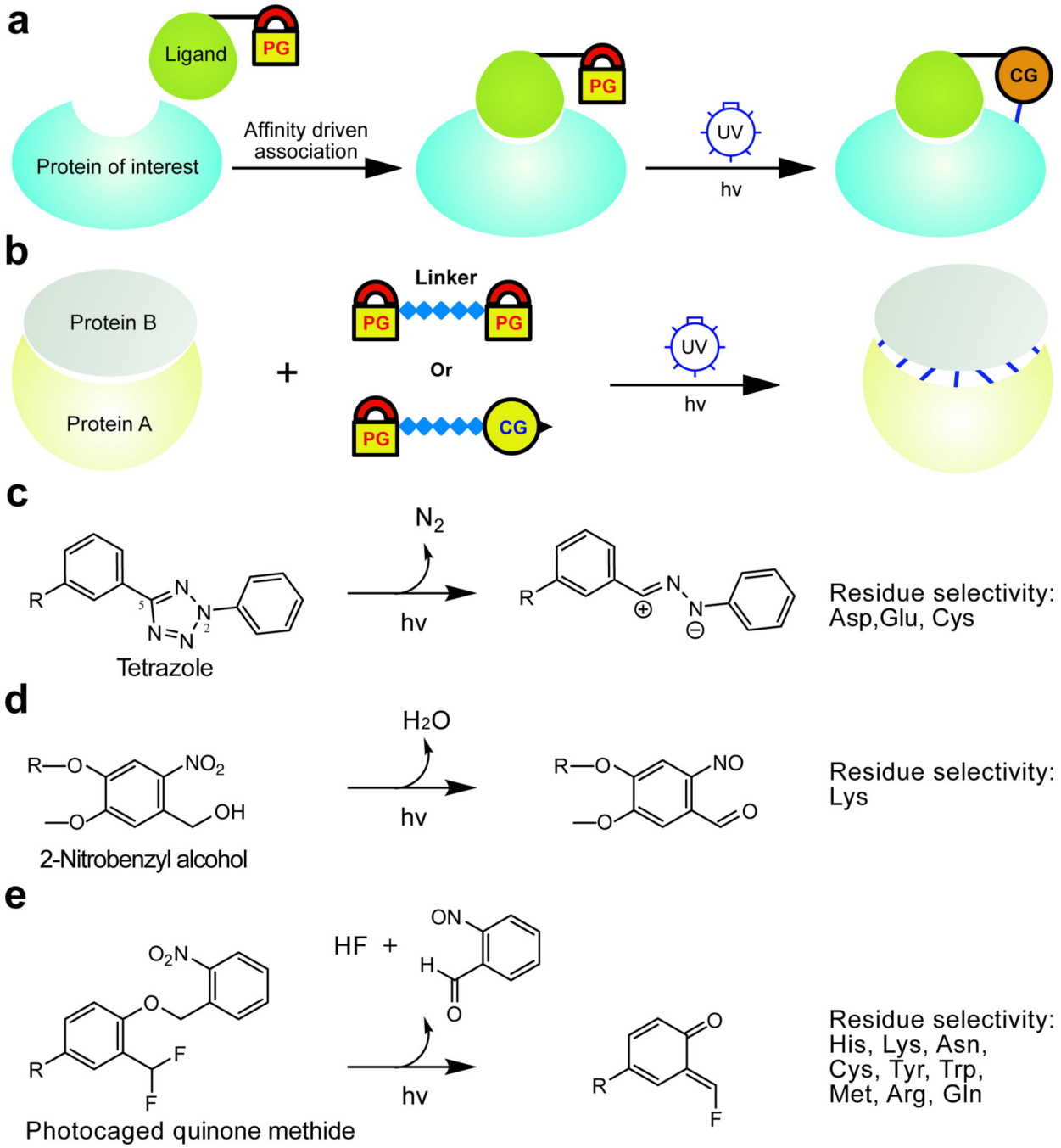

Figure 1.

(**a**) A schematic view of photoaffinity labeling. PG: photoactivatable crosslinking group; CG: crosslinking group. (**b**) A schematic view of protein crosslinking with photoactivatable small molecule crosslinkers. The crosslinker can have one or two PGs. (**c-e**) Photoreactive mechanisms for tetrazole (c), 2-nitrobenzyl alcohol (d), and photocaged quinone methide (e).

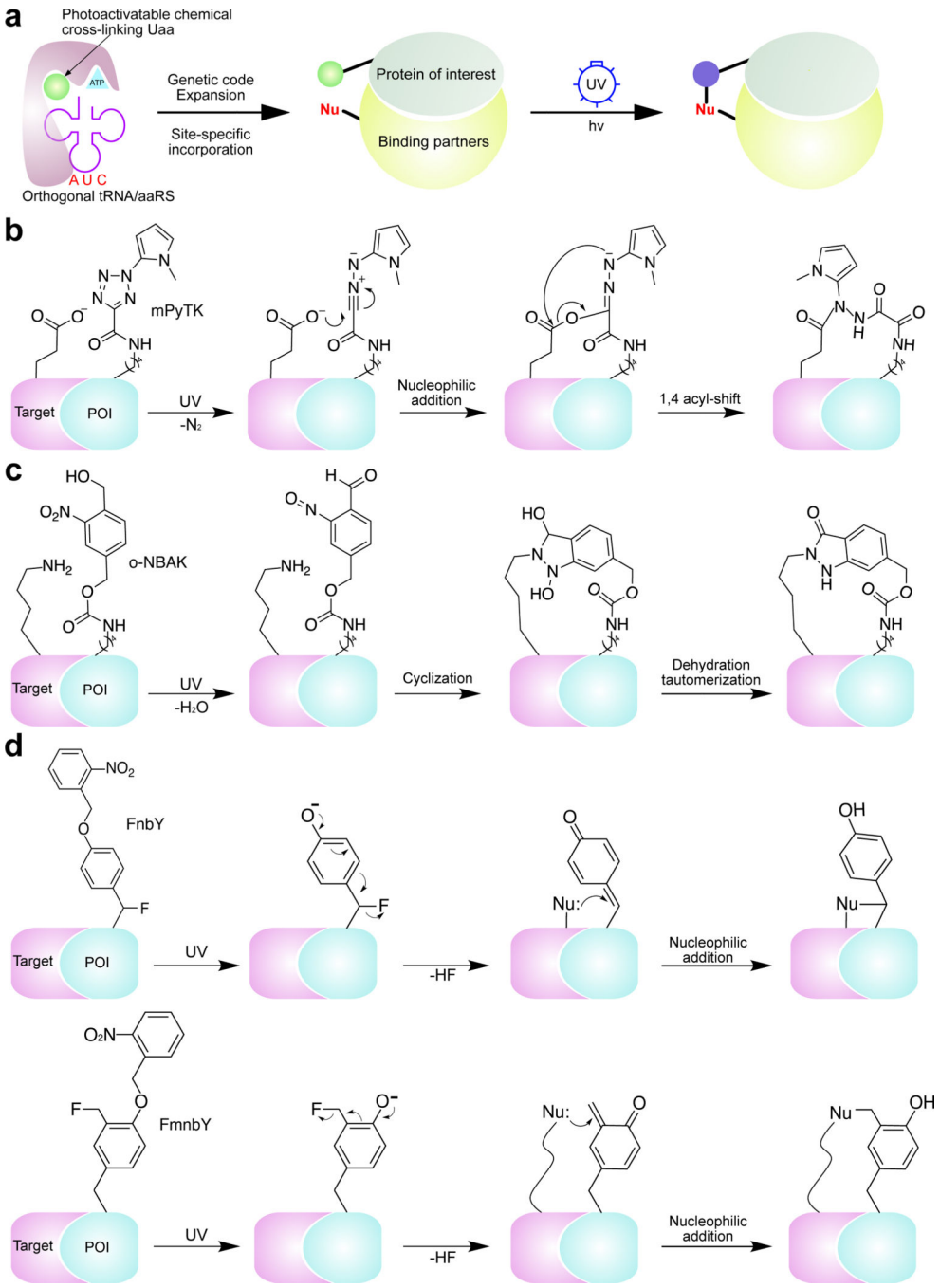

Figure 2.

(**a**) A schematic view of genetically encoding photoactivatable chemical crosslinking Uaa to crosslink proteins upon UV light activation and with residue selectivity. (**b-d**) Photoactivation and subsequent residue-specific reaction mechanism of the genetically encoded mPyTK (b), o-NBAK (c), and FnbY/FmnbY (d).

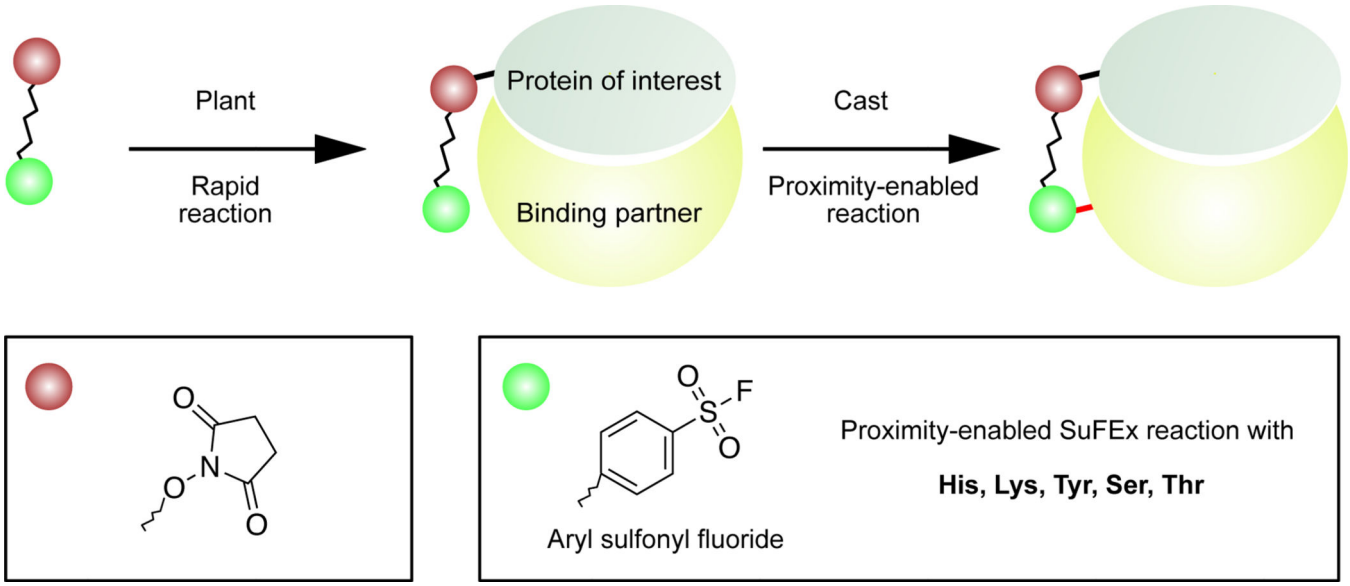

Figure 3.

Crosslinking mechanism of small molecule crosslinker containing one weakly reactive group whose reactivity is enhanced by proximity effect.

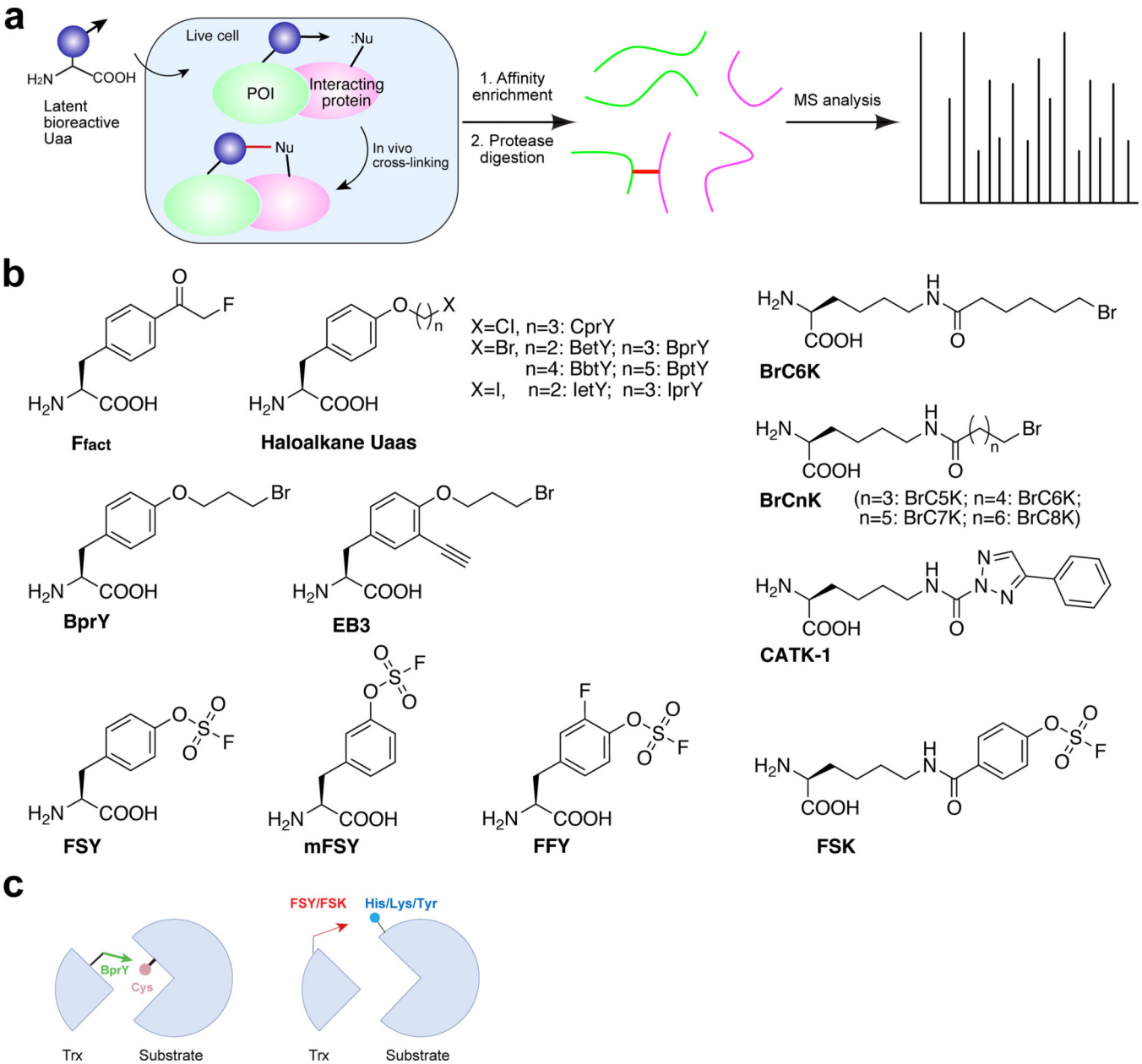

Figure 4.

(**a**) A schematic view of genetically encoded chemical crosslinking (GECX) in live cells to capture protein-protein interactions for MS identification. (**b**) Structures of latent bioreactive Uaas used in GECX as discussed in the text. (**c**) Not limited to the active site, multi-targeting ability of FSY/FSK allows the Uaa to be placed at the binding periphery to target the more abundant Lys/His/Tyr, expanding the diversity of proteins amenable to GECX.