A Protein Cleavage Platform Based on Selective Formylation at Cysteine Residues

Naoki Zenmyo[†], Yuya Matsumoto[†], Akihiro Yasuda[†], Shohei Uchinomiya[†], Naoya Shindo[†], Kaori Sasaki-Tabata[†], Tomonori Tamura[‡], Itaru Hamachi[‡], Akio Ojida[†]*

[†]Graduate School of Pharmaceutical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, JAPAN

[‡]Graduate School of Engineering, Kyoto University, Katsura, Nishikyo-ku, Kyoto 615-8510, JAPAN.

***Corresponding author: ojida@phar.kyushu-u.ac.jp (A.O.)

Abstracts:

Site-selective cleavage of the peptide backbone in proteins is an important class of post-translational modification (PTM) in nature. However, the organic chemistry for such site-selective peptide bond cleavages has yet to be fully explored. Herein, we report cysteine *S*-formylation as a means of selective protein backbone cleavage. We developed *N*-formyl sulfonylanilide as a cysteine-selective formylation reagent for proteins. Upon *S*-formylation with the reagent, the amide bond adjacent to the *S*-formylated cysteine is cleaved by hydrolysis under neutral aqueous conditions. Formylation probes bearing a protein ligand enabled the affinity-based selective cleavage of the target proteins not only in the test tube, but also under biorelevant conditions such as in crude cell lysate and on the cell surface. These results demonstrate the high biocompatibility of this protein cleavage technology. A proof-of-concept study of the cleavage-induced protein activation further demonstrates its utility as a platform for the functional regulation of proteins by artificial PTM.

Posttranslational modifications (PTM) are used ubiquitously in nature to modulate protein structure and function. A representative class of PTM is the enzyme-catalyzed modification of amino acid side chains with various chemical groups such as phosphate, acyl, and sugar units[1]. This type of PTM greatly expands the proteome diversity, which cannot be covered by genome encoding and gene transcription processes. Another important class of PTM is the covalent cleavage at specific peptide backbone sites in proteins[1-3]. This controlled proteolysis is widely used in protein maturation, which is necessary to regulate the activity and cellular localization of proteins. In addition to these naturally occurring PTMs, artificial chemical modification is another possible approach to manipulating protein structure and function. This type of artificial PTM allows us to selectively introduce artificial molecules into target proteins and thereby regulate their function. Recent advances in a new class of chemical reactions for chemoselective protein modifications, [4-9] which can be used for localization imaging, activity switching, and drug-transportation of proteins, demonstrate the importance of this approach. In contrast to these extensive studies, little progress has been made for decades in the development of artificial PTMs that allow site-selective cleavage of the peptide backbone in proteins.

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

To date, chemical reactions available for the cleavage of native proteins remain scarce. The conventional approach is based on the cyanation of cysteine residues using reagents such as 2-nitro-5-thiocyanatobenzoic acid (NTCB)[10,11] and 1-cyano-4-dimethylaminopyridinium tetrafluoroborate (CDAP)[12], which induce the amide bond cleavage adjacent to the cyanated cysteine (Figure 1a). However, these reagents require harsh pH conditions (pH 9.0 for NTCB and pH 3.0 for CDAP) and a high concentration of amine (~ 1000 mM) as an additive to facilitate the peptide bond cleavage, making it difficult to avoid protein denaturation. The intrinsic high thiol reactivities of these reagents (reaction half-life with glutathione $t_{1/2}$ < 1 min at 25 °C, Figure S1) also preclude the site-selective cleavage of proteins. As a complement to the cyanation-based cleavage, Kanai *et al.* have developed the stoichiometric copper-mediated oxidative cleavage of proteins at serine residues under denaturing, acidic conditions (CH_3CN/H_2O/AcOH = 9:9:2) (Figure 1b).[13] DeGrado has also developed a sequence-selective nickel-assisted hydrolysis of amide bonds in proteins[14] under conditions with high nickel ion concentrations (~ 1 mM). However, due to their biologically incompatible reaction conditions, the applications of these metal-catalyzed protein cleavages are limited to test tube experiments (Figure 1b). Very recently, Davis *et al.* have reported a reductive, radical-mediated protein cleavage using a diboron reagent (Figure 1c).[15] Although this work marks great progress in chemical proteolysis research, this approach requires the pre-formation of dehydroalanine before the site-selective cleavage. Other reported methods for chemical protein cleavage require harsh conditions such a inducible protein denaturation[6-18] and/or the incorporation of a noncanonical amino acid to invoke the site-selective reaction. [9,20] Furthermore, to our knowledge, there are no reports of ligand-directed site-selective cleavage of proteins using a reactive affinity probe. Given the limited progress in this research area, we aimed to develop a new chemical PTM with high biocompatibility which allows the site-selective and single-step cleavage of protein backbones in a biological setting.

Here, we introduce cysteine *S*-formylation as a new artificial PTM that induces the peptide bond cleavage of proteins (Figure 1d). We found that *N*-formyl sulfonylanilide serves as a cysteine-selective formylation reagent. *S*-formylation by this reagent induced hydrolysis of the peptide bond adjacent to the formylated cysteine to afford the cleaved protein under neutral aqueous conditions. We successfully performed the site- and target-selective cleavage of proteins using liganded *S*-formylation probes not only in a buffer solution, but also in crude cell lysate and on the cell surface. We also applied this chemical proteolysis to activate protein ligand binding, demonstrating the utility of this cleavage platform for the functional regulation of proteins.

Peptide cleavage by cysteine formylation

We initially designed acylating agents with a sulfonylanilide scaffold (Figure 2a) based on the molecular structures of *N*-acylation reagents.[21-23] and evaluated their peptide cleavage activities towards the cysteine-containing peptide-**a** by fluorescence analysis (Figure 2b). This peptide was designed to recover the fluorescence of 7-hydroxycoumarin upon cleavage and disassociation of the 2,4-dinitrophenyl quencher group. In the time-dependent fluorescence analysis, formylation reagent **1** induces the fluorescence recovery of the peptide over 12 hr under neutral aqueous conditions (100 mM phosphate buffer containing 30% CH_3CN, pH

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

7.4, 37 °C, Figure 2c and 2d). LC-MS analysis revealed rapid formation of a new peak at 20 min, which corresponds to the *S*-formylated peptide-**a** ($[M+H]^+$ = 972.33) (Figure 2e). Substitution of the sulfhydryl group with *N*-ethylmaleimide completely blocked the cleavage reaction, suggesting that the formylation occurs at the cysteine residue (Figure S2). *S*-formylation of peptide-**a** was followed by the gradual formation of the two peaks at 31 and 40 min, which were assigned as the hydrolyzed N-fragment ($[M+H]^+$ = 491.18) and the formylated C-fragment ($[M+H]^+$ = 500.16) peptides, respectively. The yield of the cleavage reaction was estimated to be 30% at 12 h based on the quantified peak intensities of the N- and C-fragments. On the other hand, when using the acetylation and butyrylation reagents **2** and **3**, the cleavage reaction proceeded with the much lower rates than with **1**. This finding prompted us to further investigate the chemical properties of the *S*-formylation induced-peptide bond cleavage. The aqueous stability of the *S*-acylated cysteines was evaluated by ^{1}H-NMR under neutral aqueous conditions (100 mM phosphate buffer containing 25% CD_3CN, pD 7.4, Figure 2f, S3 and S4). The data showed that the *S*-formylated cysteine **4** was gradually hydrolyzed with a half-life ($t_{1/2}$) of 6.2 hr to provide the free-thiol species, indicating instability of the *S*-formyl species in neutral aqueous conditions. In contrast, the corresponding *S*-acetylated cysteine **5** was stable under the same conditions ($t_{1/2}$ > 24 h). Combining these data analyses, we propose the reaction mechanism for the *S*-formylation-induced peptide bond cleavage shown in Figure 2g. In the step of the amide bond cleavage, we speculate the formation of the *N*-formyl imide as the reaction intermediate that undergoes the peptide bond cleavage, although we did not indentify this species in the LC-MS analysis.

We next evaluated the sequence dependency of the peptide bond cleavage induced by *S*-formylation using reagent **1**. The initial reaction rates (ΔF/min) of the series of peptides are summarized in Figure 3a. The results showed that the peptide-**b** and -**c**, possessing Asp-Cys and Asn-Cys sequences, respectively, displayed much higher reaction rates than peptide-**a** which has the Gly-Cys sequence (Figure 3b). The cleavage reactions of peptide-**b** and -**c** proceeded almost quantitatively after 3 hr based on LC-MS analyses, which suggest the formation of the corresponding N- and C-fragments from peptide-**b** and -**c** (Figure 3c, S5). In the case of peptide-**c** (Xaa = Asn), we identified the succinimide derivative as the main product of the *N*-fragment (Figure S5), suggesting that the intramolecular attack of the asparagine residue to the amide carbonyl group facilitates the cleavage reaction. In the case of peptide-**b** (Xaa = Asp), the formation of the succinate was not identified by LC-MS, probably due to its rapid hydrolysis. Other tested peptides with different amino acids adjacent to the cysteine did not give significant changes in the reaction rate compared to peptide-**a** (Figure 3a and S6).

Optimization of electrophilic reactivity of the formylation reagents

We next optimized the electrophilic reactivity of the formylation reagent that allows selective formylation of a target cysteine residue. Formylation reagent **1** was highly reactive toward a cysteine derivative such as glutathione (GSH) ($t_{1/2}$ < 0.1 hr, Table 1) and unstable by hydrolysis under the neutral aqueous conditions (pH 7.4, 37 °C) ($t_{1/2}$ = 1.0 hr). We found that the high reactivity of **1** could be tuned by the introduction of substituents on the sulfonylanilide moiety. Among the derivatives (Table 1) synthesized, the formylation reagents **9** and **10**

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

exhibited moderate reactivities toward GSH ($t_{1/2}$ = 0.15 and 1.47 hr, respectively) while maintaining sufficient aqueous stability under neutral aqueous conditions ($t_{1/2}$ = 6.6 and 37.2 hr, respectively). The reactivities of **9** and **10** toward other amino acids such as *N*-acetyl lysine, *N*-acetyl tyrosine, and *N*-acetyl histidine, and glycine amide were further evaluated (Table 1, Table S1). The data revealed that **9** and **10** did not show noticeable reactivity toward excess amounts of these amino acids except for N-acetyl histidine, which slightly shortened the half-life of the probes. Based on these results, we designed formylation regents **12** and **13** as the more water-soluble analogs of **9** and **10**, respectively (Figure 4a). The selectivity of these formylation reagents was evaluated using a small ubiquitin protein (8.5 kDa), which possesses seven lysines, one tyrosine and one histidine but lacks cysteine on the surface. The MALDI-TOF mass analysis revealed that the treatment of ubiquitin (2.5 μM) with a high concentration of **13** (1 mM) resulted in scarcely any formylation even after 6 hr, while **12** provided a mono-formylated ubiquitin after 6 hr (Figure S7). These data suggested the high cysteine selectivity of formylation reagents **12** and **13** on the protein surface. In the peptide cleavage experiment, the reactivities of **12** and **13** were not significantly decreased compared to those of probe **1** (Figure S8).

The rather slow cleavage rate of peptide-**a** using the formylation reagents prompted us to find additives that could accelerate the reaction. Based on the proposed reaction mechanism (Figure 2g), we expected that the addition of nucleophilic species would assist the amide bond cleavage. We found that a variety of thiol species effectively accelerated the cleavage reaction of peptide-**a** by formylation regent **12** (Figure 4b), whereas amine derivatives such as glycine amide and propargyl amine did not show noticeable effects on the reaction rate (Figure S9). Among the thiol species, glutathione (GSH) was most effective for reaction rate acceleration. The yield of the cleavage reaction was evaluated to be 77% after 3 hr based on fluorescence intensity analysis. Time-dependent LC-MS analysis revealed the formation of the alkynylated N-fragment (34 min) as a main product in the presence of the alkynylated cysteine (Cys-alkyne) (Figure 4c). Based on this finding, we propose a reaction mechanism of the thiol-assisted amide bond cleavage, which involves a nucleophilic thioester formation with Cys-alkyne and a subsequent intramolecular S to N acyl transfer reaction to give stable alkynylated N-fragment (Figure 4d).

Protein cleavage induced by cysteine formylation

We next applied the cysteine formylation peptide bond cleavage to proteins. As an initial attempt, we employed maltose binding proteins (MBPs) possessing a single cysteine in the Xaa-Cys-Asp8 peptide tag (Xaa = Asp, Asn, and Gly) at their C-termini (Figure 5a). When MBPs were treated with the formylation regents **1**, **12**, and **13** (1 or 5 mM, 2 hr) in the presence of GSH under neutral aqueous conditions (100 mM phosphate buffer, pH 7.4, 37 °C), the proteins truncated at the tag site were observed by SDS-PAGE analysis (Figure 5b). In the absence of GSH, the cleavage reaction proceeded less efficiently with MBP-GC compared to MBP-NC and MBP-DC, as we observed in the peptide cleavage experiments (Figure S10). We next attempted the cleavage of point mutated MBPs possessing a single cysteine (T32C, A207C and A269C) at the internal loop regions. SDS-PAGE analysis showed the partial cleavage of the MBPs at the cysteine site upon treatment with

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

the formylation probes **1**, **12** and **13** (5 mM) in 30% acetonitrile-phosphate buffer solution (pH 7.4) (Figure 5c). The cleavage efficiency of the reagents was found to be **1** > **12** > **13**, which is consistent with the order of their reactivities to cysteine. Treatment of the triple cysteine mutant MBP (T32C/A207C/A269C) with the formylation reagents resulted in the complete disappearance of the parent MBP and provided multiple fragments of MBP. It should be noted that these formylation reagents exhibited higher cleavage efficiencies than the conventional cyanation reagents CDAP and NTCB.[10-12] Interestingly, by harnessing the thiol-assisted cleavage mechanism as shown in Figure 4d, we were able to introduce fluorescent rhodamine or biotin onto the C-terminus of MBP-GC upon treatment with the corresponding cysteine derivatives (*i.e.*, Cys-TAMRA and Cys-biotin) (Figure 5d). The labeling of MBP with Cys-TAMRA and Cys-biotin was confirmed by in-gel fluorescence and western blot analyses, respectively (Figure 5e).

Affinity-based cleavage of proteins by cysteine formylation

To perform site-selective cleavage of proteins by cysteine *S*-formylation, we employed the *N*-terminal domain of human double minute 2 (HDM2)[23] bearing a GST (glutathione S-transferase) (GST-HDM2) (Figure 6a, Figure S11). For the ligand-directed affinity-based cleavage of GST-HDM2, we designed formylation probe **14** possessing a Nutlin-3 (Figure 6b), which is known as a ligand for HDM2. In GST-HDM2, Cys2 is closely positioned at the binding site of Nutlin-3 and was selected as the *S*-formylation target cleavage site. When GST-HDM2 (2 μM) was incubated with probe **14** (100 μM) under neutral aqueous conditions (100 mM phosphate buffer, pH 7.4, 37 °C), the *N*-terminal GST domain gradually increased in the CBB stain (Figure 6c and 6d). The yield of the cleavage reaction was estimated to be 79% after 6 hr based on western blot analysis using anti-His tag antibody. MALDI-TOF mass analysis showed the appearance of a new peak corresponding to the N-terminal GST fragment (Figure S12), suggesting the site-selective cleavage of GST-HDM2 at Cys2. The cleavage reaction by probe **14** was effectively suppressed in the presence of Nutlin-3 (500 μM). Incubation of GST-HDM2 with high concentrations of formylation regent **13** (1 mM) induced the cleavage reaction only slightly. These results exhibit the ligand-directed cleavage of GST-HDM2 by *S*-formylation with **14**.

We next attempted the functional regulation of HDM2 by affinity-based cleavage. For this proof-of-concept experiment, we designed p53-His10-HDM2, which possesses p53 peptide analogue (L26V)[24, 25] and a His10 tag at the *N*-terminus of HDM2 (Figure 6e, Figure S13). Treatment of p53-His10-HDM2 (2 μM) with the binuclear Ni(II)-NTA probe **15**-2Ni(II) (20 μM) (Figure 6f)[26] induced the time-dependent cleavage of the *N*-terminal region of HDM2 in HEPES buffer (pH 7.4) (Figure 6g). The yield of the cleavage reaction after 3 hr was estimated to be 65% based on the CBB stain band intensity. The cleavage reaction did not proceed when using probe **15**, which lacks nickel ions essential for binding to His10 tag. A fluorescence anisotropy experiment revealed that the binding of fluorescent HDM2 probe **17** (0.4 μM) to p53-His10-HDM2 (1.5 μM) was greatly suppressed, mainly due to the intramolecular interaction between the p53 peptide (L26V) and the HDM2 Nutlin-binding domain. We found that the fluorescence anisotropy value gradually increased from 0.06 to 0.13 upon treatment with **15**-2Ni(II) for 3 hr (Figure 6h). This time course corresponds well with the progress of the

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

cleavage reaction (Figure S14a). The analysis based on the calibration curve (Figure S14b, c) suggested that the binding activity of HDM2 towards **17** was recovered from 13% to 67% of that of intact HDM2. The anisotropy value did not change upon treatment of HDM2 with **16**-2Ni(II), which lacks a reactive formyl group. These data demonstrate the utility of the formylation-induced cleavage for activation of proteins.

We further applied the affinity-based cleavage to the kinase domain of fibroblast growth factor receptor 1 (FGFR1, Met399-Arg822) (Figure 7a, Figure S15). For this purpose, we designed formylation probe **18** based on the structure of futibatinib[27], which has been developed as a covalent inhibitor of FGFR 1–4 (Figure 7b). In the probe design, the acrylamide of futibatinib was replaced with N-formyl sulfonylanilide to target Cys488 for formylation. When the recombinant kinase domain of FGFR1 fused with His6 tag (2 μM) was treated with **18** (50 μM) for 3 hr, the intensity of the FGFR1 band greatly decreased in the SDS-PAGE analysis (Figure 7c) in a concentration-dependent manner (10-50 μM of **18**) (Figure S16). This decrease in the band intensity was not observed with 50 μM of the non-liganded formylation probe **1**, while the high concentration of **1** (1 mM) effectively reduced the band intensity of FGFR1 (50 kDa). The western blot analysis using His tag antibody suggested that *ca.* 85% of FGFR1 was degraded by **18** (Figure 7d). These results demonstrate the application of ligand-directed *S*-formylation cleavage to the kinase domain of FGFR1 with **18**.

Affinity-based cleavage of proteins under biorelevant conditions

We performed the formylation-induced protein cleavage in crude cell lysate to assess its biocompatibility. In this experiment, we employed MBP tagged with Cys-His10 at its C-terminus (Figure S17a, b) and Ni(II)-NTA probe **15**-2Ni(II). We found that the cleavage reaction of the Cys-His10 tag from MBP proceeded upon treatment of MBP-Cys-His10 (0.8 μM) with **15**-2Ni(II) (20 μM) in the crude lysate of A431 cells (Figure 8a and 8b). The yield of the cleavage reaction reached 80% after 6 hr based on western blot analysis using anti-His tag antibody. This yield was comparable to those of the cleavage reaction of MBP-Cys-His10 carried out in phosphate buffer (pH 7.4) (Figure S17c, d). To confirm the protein selectivity of the cleavage reaction, the lysate sample treated with **15**-2Ni(II) (20 μM) was further treated with cysteine-TAMRA to fluorescently label the cleaved proteins (Figure 8c). The in-gel fluorescence analysis showed a single fluorescent band corresponding to the cleaved MBP, suggesting the high target selectivity of the formylation-induced protein cleavage.

Finally, we attempted the peptide bond cleavage of a membrane protein expressed on the cell surface. In this experiment, we designed bradykinin receptor type 2 (B2R) receptor conjugated to a Cys-His10 tag and a SPOT tag on the extracellular N-terminus (Figure 8d, Figure S18). Transient expression of the B2R was confirmed by fluorescence imaging using the anti-SPOT nanobody and B2R antagonist peptide, which were conjugated with Alexa Fluor™ 568 and Cy5 dye, respectively (Figure 8e). When the cells were treated with **15**-2Ni(II) (40 μM) for 3 hr, the fluorescence of the anti-SPOT nanobody decreased on the cell surface, while the fluorescence of the B2R antagonist peptide remained nearly unchanged. The quantitative ratio analysis (anti-

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

SPOT nanobody / B2R antagonist) revealed that the fluorescence of anti-SPOT nanobody decreased by 45% compared to the non-treated cells (Figure 8f). On the other hand, treatments of the cells with probe **15** (without nickel ions) and probe **16**-2Ni(II) (without a formyl group) gave only small changes in the fluorescence of anti-SPOT nanobody. To further confirm this result, we conducted western blot analysis and evaluated the band intensities of B2R detected by anti-SPOT nanobody and anti-B2R antibody (Figure 8g). The data showed that the band detected by the anti-SPOT nanobody becomes weaker upon treatment of the cells with **15**-2Ni(II), while treatments with **15** (without nickel ions) and **16**-2Ni(II) gave only small changes in the band intensity. The quantitative ratio analysis (anti-SPOT nanobody / B2R antibody) disclosed a significant decrease in the band intensity of anti-SPOT nanobody upon treatment with **15**-2Ni(II) (Figure 8h). All these results suggest the cleavage of the SPOT tag at the extracellular N-terminus region from B2R by binding-induced cysteine *S*-formylation. A cell viability assay revealed that **15**-2Ni(II) did not exert significant cytotoxicity, which also suggests good biocompatibility of cysteine *S*-formylation for targeting membrane surface proteins (Figure S19).

Discussion

Formylation of the lysine residue has been known as a protein PTM since it was first discovered in 2007, although its biological function has yet to be elucidated[28, 29]. On the other hand, to our knowledge, there are no reports of cysteine formylation of cellular proteins. This is probably due to the difficulty of detecting the easily hydrolyzed *S*-formylated species and the apparent absence of an enzyme that catalyzes the formylation of proteinaceous cysteines [30]. For these reasons, cysteine formylation has been overlooked as a PTM of proteins. In this manuscript, we show for the first time that cysteine *S*-formylation induces peptide bond cleavage by hydrolysis. Unlike cysteine cyanation using NTCB and CDAP reagents[10,12], the formylation-induced protein cleavage efficiently proceeds under mild neutral conditions. Our study revealed that cysteine *S*-acetylation and *S*-butylation are much less effective for peptide bond cleavage compared to *S*-formylation (Figure 2a, 2d). Degani has reported preliminary research on peptide cleavage by *p*-nitrobezoylation of cysteine, though this reaction requires harsh conditions to yield a cleaved peptide (pH 8.1, 50 °C) [31]. These data suggest the exceptional ability of cysteine formylation to induce peptide bond cleavage. Considering these results, we suggest that, for effective peptide bond cleavage, the cysteine residue must be modified with a small and strong electron withdrawing group such as formyl and cyano groups.

We employed *N*-formyl sulfonylanilides as the formylation reagents for cysteine. While *N*-formylsaccharin has been widely used as a formylation reagent for amines[32], little is known about the utility of less reactive *N*-formyl sulfonylanilide for this purpose. In this study, we revealed that *N*-formyl sulfonylanilides with tuned electrophilic reactivities can work as cysteine-selective formylation reagents under neutral aqueous conditions. Our affinity-based formylation probes bearing ligand units were able to selectively formylate a cysteine residue on the target proteins and induced site-selective cleavage of the peptide bond. To our knowledge, this is the first example of ligand-directed protein cleavage by a chemical reaction. Along with this ligand-directed selectivity, the reversible hydrolytic nature of *S*-formylated cysteine also could contribute to reducing off-target formylation

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

and undesired protein cleavage. We believe that these properties contribute to the target-selective cleavage of the target protein in crude cell lysate containing a multitude of proteins (Figure 8 a-c).

We are currently using this cleavage method for proteins possessing a free cysteine in a highly accessible and conformationally flexible region. We anticipate that further improvement of the formylation probes (for example, higher thiol selectivity and smaller molecular size) will increase the number of proteins that are compatible with this cleavage approach. Global screening of cleavable cysteine sites in cellular proteins will accelerate this advance. Building upon these future studies, we envision that this cleavage platform will facilitate research progress in a variety of ways; targeting disease-related proteins for degradation[33] and mass spectrometry-based shotgun proteomics analysis of cellular proteins[34] are just a few examples of a whole range of possibilities.

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Reference

1. C. T. Walsh, S. Garneau-Tsodikova, G. J. Gatto, Jr, Protein posttranslational modifications: The chemistry of proteome diversifications, *Angew. Chem. Int. Ed.*, **44**, 7342-7372 (2005).
2. H. Paulus, Protein splicing and related forms of protein autoprocessing, *Ann. Rev. Biochem*., **69**, 447-496 (2000).
3. Kwon, Y. T., Ciechanover, A., The ubiquitin code in the ubiquitin-proteasome system and autophagy, *Trend. Biochem. Sci.*, **42**, 873-886 (2017).
4. Emily A. Hoyt, et al. Contemporary approaches to site-selective protein modification. *Nat. Rev. Chem.,* **3**, 147–171 (2019).
5. B. Josephson, B. G. Davis, et al., Light-driven post-translational installation of reactive protein side chains, *Nature*, **585**, 530-537 (2020).
6. M. T. Taylor, et al., A protein functionalization platform based on selective reactions at methionine residues, *Nature*, **562**, 563-568 (2018).
7. S. Lin, et al., Redox-based reagents for chemoselective methionine bioconjugation, *Science*, **355**, 597-602 (2017).
8. C. B. Rosen, M. B. Francis, Targeting the N terminus for site-selective protein modification, *Nat. Chem. Biol.*, **13**, 697-705 (2017).
9. J. Li, P. R. Chen. Development and application of bond cleavage reactions in bioorthogonal chemistry. *Nat. Chem. Biol.* **12**, 129-137 (2016).
10. H.Y.Tang, D. W. Speicher. Identification of alternative products and optimization of 2-nitro-5-thiocyanatobenzoic acid cyanylation and cleavage at cysteine residues. *Anal. Biochem.* **334**, 48 (2004).
11. Y. Qiao, W. R. Liu, et al. Expressed Protein Ligation without Intein. *J. Am. Chem. Soc.*, **142**, 7047 (2020).
12. J. Wu, J. T. Watson. Optimization of the cleavage reaction for cyanated cysteinyl proteins for efficient and simplified mass mapping. *Anal. Bochem.* **258**, 268-276 (1998).
13. Y. Seki, M. Kanai, et al., Serine-Selective Aerobic Cleavage of Peptides and a Protein Using a Water-Soluble Copper–Organoradical Conjugate. *Angew. Chem. Int. Ed.*, **53**, 6501-6505 (2014)
14. B. Dang, W. F. DeGrado, et al., SNAC-tag for sequence-specific chemical protein cleavage. *Nature Methods,* **16**, 319 (2019).
15. T. A. Mollner, B. G. Davis, et al., Reductive site-selective atypical C,Z-type/N2-C2 cleavage allows C-terminal protein amidation. *Sci. Adv.*, **8**, eabl8675 (2022).
16. Denslow et al, Chemical Cleavage of Proteins in Solution, *Current Protocol Pro. Sci.*, 41, 1-11 (2005).
17. R. Okamoto, Y. Kajihara, et al., A synthetic approach to a peptide alpha-thioester from unprotected peptide through cleavage and activation of a specific peptide bond by N-acetylguanidine, *Angew. Chem. Int. Ed.* **51**, 191-196 (2012).
18. K. Pane, E. Notomista et al., Chemical Cleavage of an Asp-Cys Sequence Allows Efficient Production of Recombinant Peptides with an N-Terminal Cysteine Residue, *Bioconjugate Chem.*, **29**, 1373 (2018).

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

19. F. B. Peters, P. G. Schultz et al., Photocleavage of the Polypeptide Backbone by 2-Nitrophenylalanine, *Chem. Biol.*, **16**, 148-152 (2009).
20. Y. Qiu, J. P. Tam et al., Selective Bi-directional Amide Bond Cleavage of N-Methylcysteinyl Peptide. *Eur. J. Org. Chem.*, **20**, 4370-4380 (2014)
21. T. Tamura, I. Hamachi et al., Rapid labelling and covalent inhibition of intracellular native proteins using ligand-directed N-acyl-N-alkyl sulfonamide, *Nat. Commun.*, **9**, No. 1870 (2018).
22. M. Kawano, I, Hamachi et al., Lysine-Reactive N-Acyl-N-aryl Sulfonamide Warheads: Improved Reaction Properties and Application in the Covalent Inhibition of an Ibrutinib-Resistant BTK Mutant, *J. Am. Chem. Soc.*, **145**, 26202-26212 (2023).
23. T. Ueda, T. Tamura, I. Hamachi. et al., Enhanced Suppression of a Protein-Protein Interaction in Cells Using Small-Molecule Covalent Inhibitors Based on an N-Acyl-N-alkyl Sulfonamide Warhead. *J. Am. Chem. Soc.* **143**, 4766-4774(2021).
24. B. Musielak, T. Holak, et. al., Competition NMR for Detection of hit/lead Inhibitors of protein-protein interactions. *Molecules*, **25**, 3017 (2020).
25. J. P . Plante A. J. Wilson. et al. Oligobenzamide proteomimetic inhibitors of the p53–hDM2 protein–protein interaction. *Chem. Commun.* **34**, 5091-5093 (2009).
26. N. Zenmyo, A. Ojida. et al. Optimized reaction pair of the CysHis tag and Ni(II)-NTA probe for highly selective chemical labeling of membrane proteins. *Bull. Chem. Soc. Jpn.,* **92**, 995–1000 (2019).
27. H. Sootome, H. Hirai et. al. Futibatinib Is a Novel Irreversible FGFR 1-4 Inhibitor That Shows Selective Antitumor Activity against FGFR-Deregulated Tumors, *Cancer Res*, **80**, 4986-4997 (2020).
28. T. Jiang, P. C. Dedon, et al. N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. *Proc. Natl. Acad. Sci.* **104**, 60-65 (2007).
29. H. Lin, X. Su, B. He, Protein lysine acylation and cysteine succination by intermediates of energy metabolism, *ACS Chem. Biol.*, 7, 947-960 (2012).
30. N. Harms, J. R. W. Reijnders, R. J. M. Spanning, A. H. Stouthamer, S-formylglutathione hydrolase of *Paracoccs denitrificans* is hololologues to hums esterase D: a universal pathway for formaldehyde detoxification?, *J. Bacter.*, 6296-6299 (1996).
31. Y. Degani, A. Patchornik, J. A. Maclaren, Specific Cleavage of Peptides at Cysteinyl Residues, *J. Am. Chem. Soc.*, **88**, 3460-3461 (1966).
32. T. Cochet, J. Cossy, et al. N-Formylsaccharin: A New Formylating Agent. *Synlett.* **13**, 1920-1922 (2011).
33. M. Bekes, D. R. Langley, C. M. Crews, PROTAC targeted protein degraders: the past is prologue, *Nat. Rev. Drug Discov.*, **21**, 181-200 (2022).
34. M. Iwasaki, T. Masuda, M. Tomita, Y. Ishihama, Chemical Cleavage-Assisted Tryptic Digestion for Membrane Proteome Analysis, *J. Proteom. Res.*, **8**, 3169-3175 (2009).

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

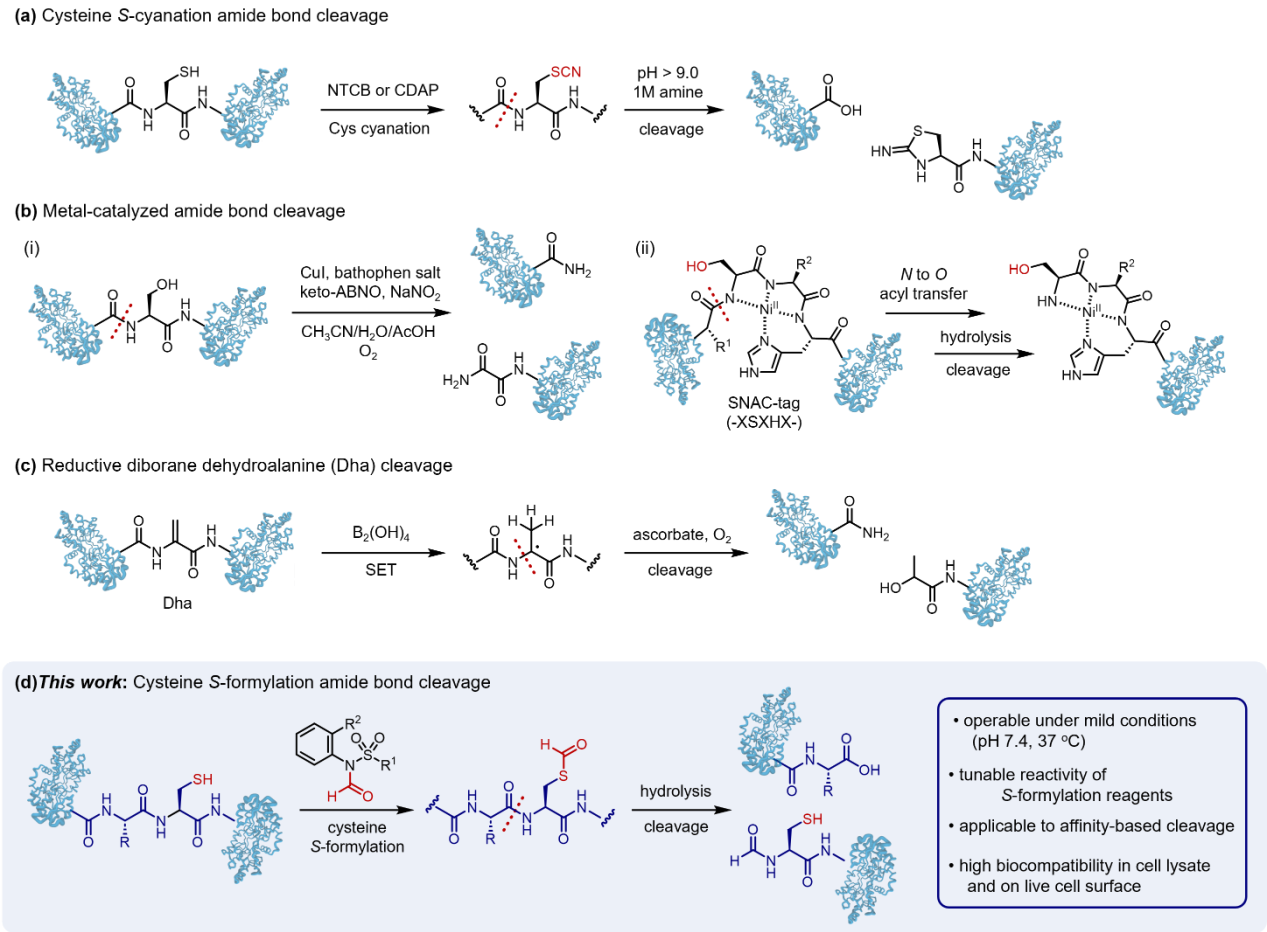


Figure 1. Chemical cleavage of peptide backbone in proteins.

a. Protein cleavage reaction through *S*-Cyanation, under harsh conditions such as pH 9.0 or a large excess of amine (1 M). **b.** Metal-catalyzed amide bond cleavage, i) Aerobic cleavage with Cu/keto-ABNO, ii) Ni(II)-complex facilitated sequence-selective cleavage of SNAC-tag (-XSXHXs-). **c.** Reductivey initiated cleavage via dehydroalanine (Dha) formation. **d. This work**: amide bond cleavage induced cysteine formylation under mild conditions (pH7.4), that could be useful chemical biology tools.

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

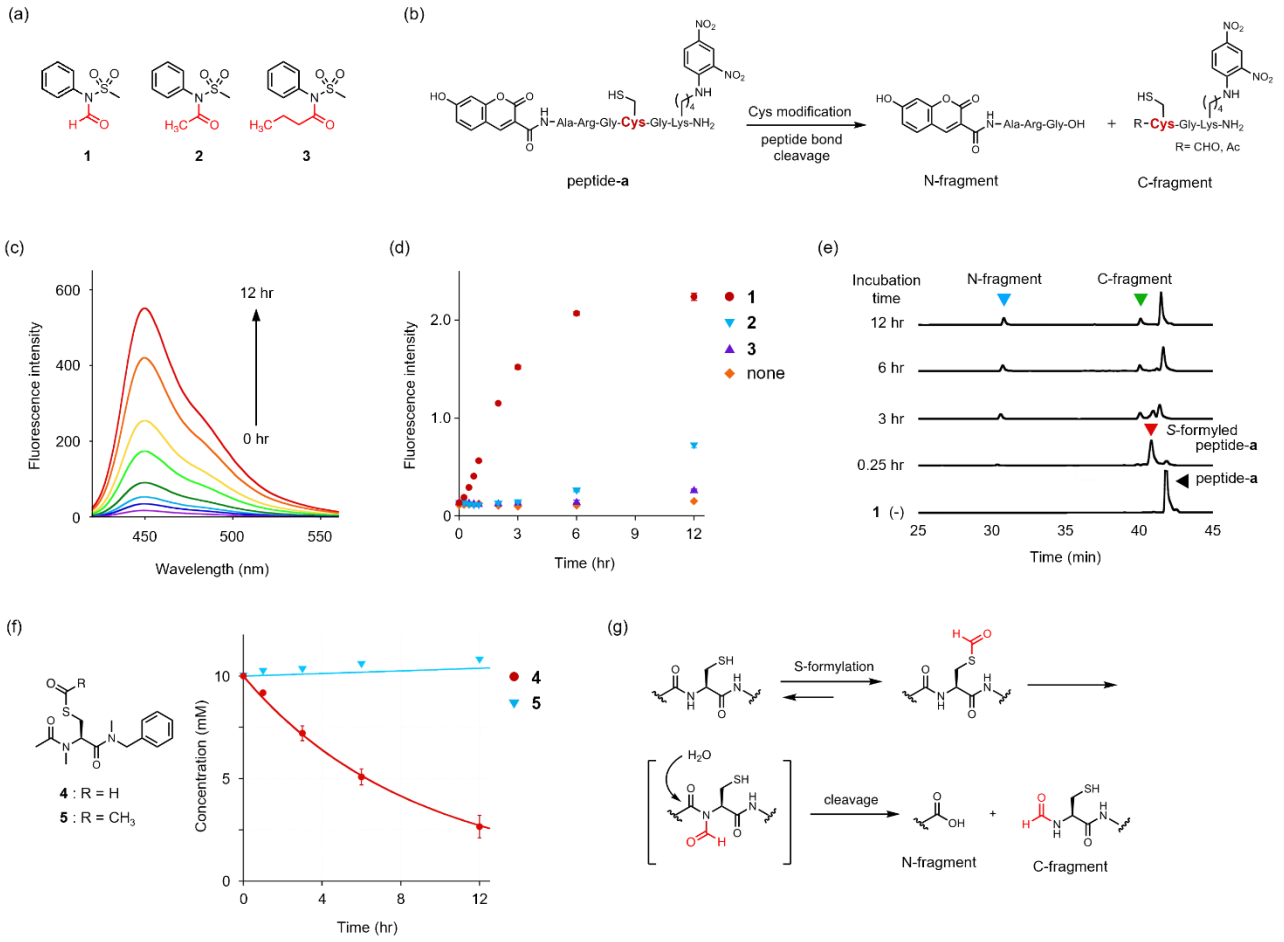


Figure 2. Peptide bond cleavage induced by cysteine *S*-formylation.

a, Structures of the sulfonamide-based acylation reagents. **b**, Peptide bond cleavage of peptide-**a** through cysteine modification. **c**, Fluorescence spectral change of peptide-**a** (15 μM) upon addition of formylation reagent **1** (1 mM). Conditions: 100 mM sodium phosphate buffer (pH 7.4), 30% MeCN, 37 °C, λ_{ex} = 400 nm. **d**, Time trace plot of the fluorescence change of peptide-**a** (15 μM) upon treatment with acylation reagent **1-3** (1 mM). Conditions: 100 mM sodium phosphate buffer (pH 7.4), 30% MeCN, 37 °C, $\lambda_{ex}/\lambda_{em}$ = 400/450 nm, mean ± s.d.: three independent experiments. **e**, HPLC analysis of the cleavage reaction of peptide-**a** thorough *S*-formylation. Peptide-**a** (30 μM) was treated with **1** (3 mM) in 100 mM sodium phosphate buffer (pH 7.4) containing 10% DMF in the presence of 0.3 mM TCEP at 37 °C. The peaks were detected by UV absorbance at 370 nm. The peptides were identified by ESI-TOF-MS analysis. **f**, Time plot of the hydrolysis of *S*-formylated cysteine **4** and *S*-acetyled cysteine **5** (1 mM) analyzed by ^{1}H-NMR. Conditions: 100 mM sodium deuterium phosphate buffer, 25% CD_3CN, pD 7.4, 37 °C, mean ± s.d.: three independent experiments. **g**, Proposed reaction mechanism of the peptide bond cleavage induced by cysteine *S*-formylation.

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

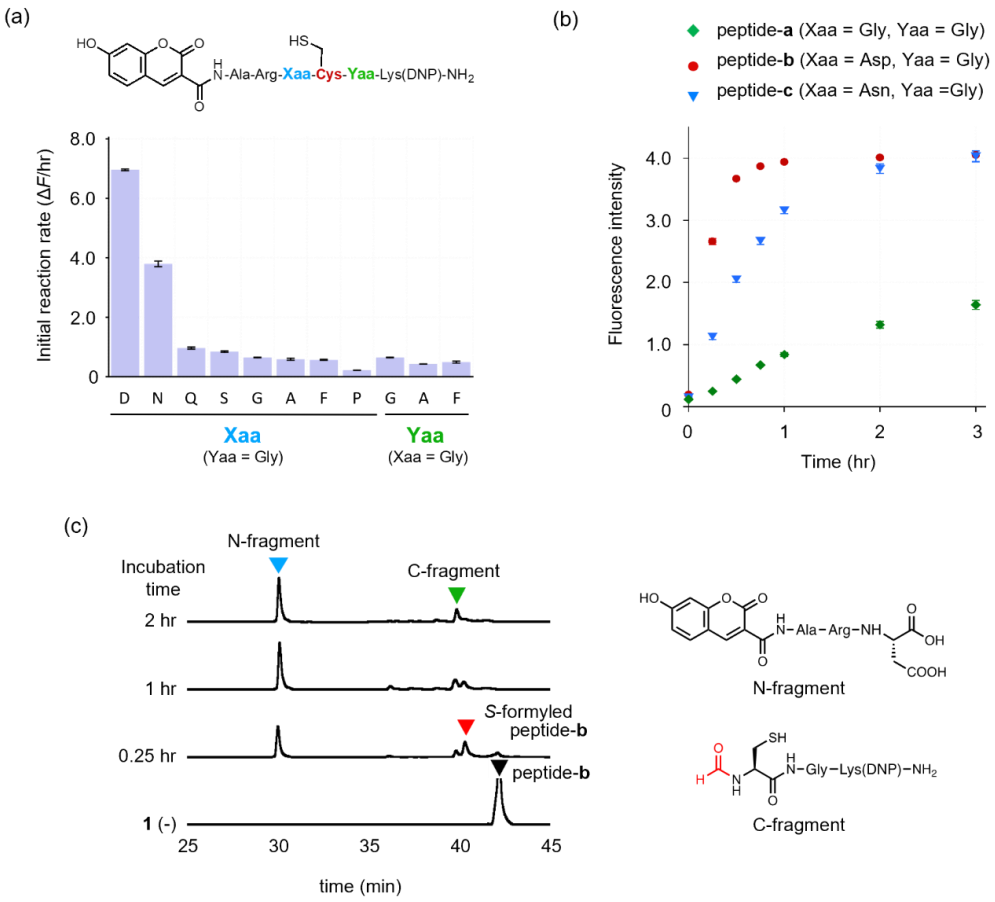
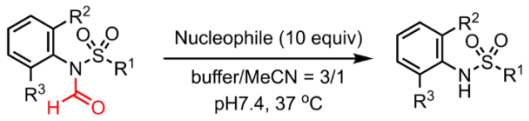


Figure 3. Sequence dependency of the peptide cleavage reaction induced by cysteine *S*-formylation.

a, Initial reaction rates of the peptide cleavage reaction (Δ*F*/min) upon treatment with **1**. Peptide (15 μM) was treated with **1** (1 mM) in 100 mM sodium phosphate buffer (pH 7.4) containing 30% MeCN at 37 °C. $\lambda_{ex}/\lambda_{em}$ = 400/450 nm. mean ± s.d.; three independent experiments. **b**, Time trace plot of cleavage reactions of peptide-**a**–**c** (Xaa-Cys-Gly, Xaa = Gly, Asp and Asn) by *S*-formylation with **1**. Peptide (15 μM) was treated with **1** (1 mM) in 100 mM sodium phosphate buffer (pH 7.4) containing 30% MeCN at 37 °C. $\lambda_{ex}/\lambda_{em}$ = 400/450 nm, mean ± s.d.: three independent experiments. **c**, HPLC analysis of the cleavage reaction of peptide-**b** by *S*-formylation. Peptide-**b** (30 μM) was treated with **1** (3 mM) in 100 mM sodium phosphate (pH 7.4) containing 10% DMF in the presence of 0.3 mM TCEP at 37 °C. The peaks were detected by UV absorbance at 370 nm. The peptides were identified by ESI-TOF-MS analysis.

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

364 **Table 1**. Summary of the aqueous stability and electrophilic reactivity of N-formyl sulfonylanilides.[a]



	Substituent			Nucleophile: $t_{1/2}$ (hr)		
compound	R^1	R^2	R^3	GSH	*N*Ac-Lys	none
1	Me	H	H	< 0.10	0.99	1.02
6	Me	Me	H	0.31	1.70	1.68
7	Me	Me	Me	2.98	16.3	16.1
8	^{i}Pr	H	H	0.16	3.89	4.01
9	^{t}Bu	H	H	0.15	6.15	6.63
10	^{t}Bu	Me	H	1.47	37.3	37.2
11	^{t}Bu	Me	Me	20.9	> 48	> 48

365

366 [a]Data represent the mean ± standard deviation of repeated experiments (n = 3).

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

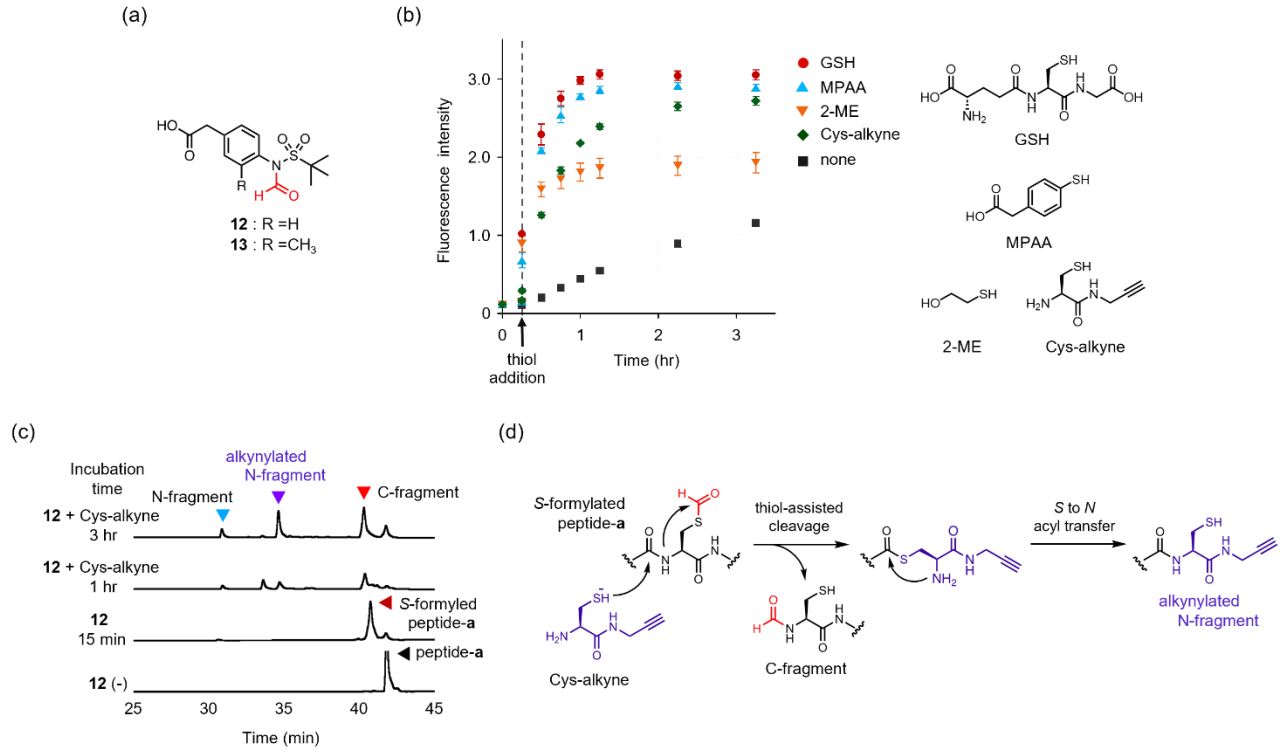


Figure 4. Thiol-assisted cleavage of cysteine *S*-formylated peptide.

a, Structures of the water-soluble *S*-formylation reagents **12** and **13**. **b**, Fluorescence change of peptide-**a** (15 μM) upon treatment with **12** (1 mM) in the absence and presence of various thiol species (1 mM). Conditions: 100 mM sodium phosphate (pH 7.4), 30% CH_3CN, 37 °C. mean ± s.d.: three independent experiments. **c**, HPLC analysis of the cleavage reaction of peptide-**a** (30 μM) upon treatment with **12** (3 mM) in the presence of Cys-alkyne (3 mM). Conditions: 100 mM sodium phosphate (pH 7.4), 10% DMF, 3 mM TCEP, 37 °C. The peaks were detected by UV absorbance at 370 nm. **d**, Proposed reaction mechanism of the thiol-assisted cleavage of *S*-formyl peptide and subsequent *S* to *N* acyl transfer.

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

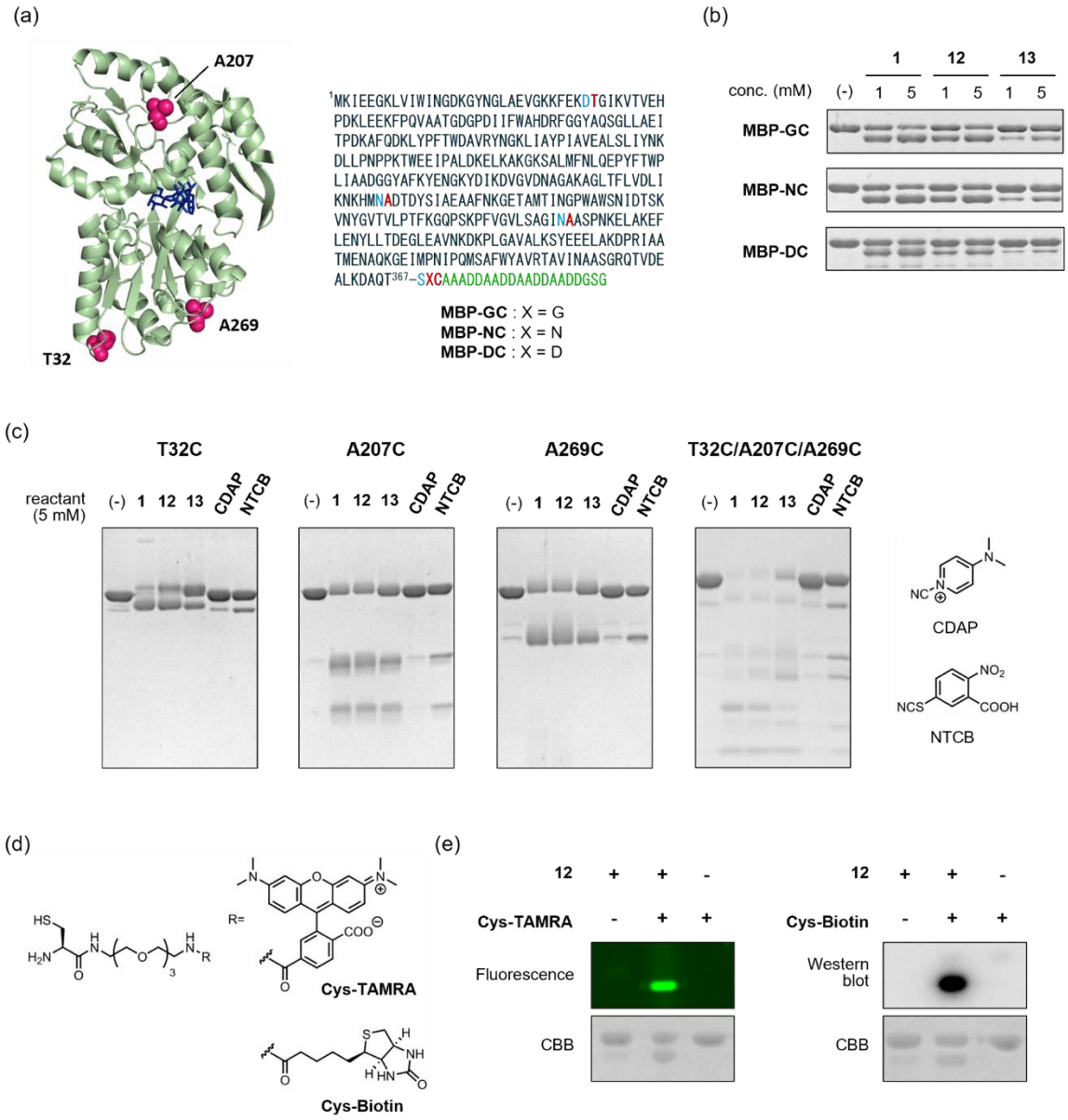


Figure 5. Protein cleavage induced by cysteine *S*-formylation.

a, Crystal structure and amino acid sequence of maltose binding protein (PDB:1ANF, cysteine mutation sites are shown in magenta). The peptide tag containing one cysteine residue (shown in green) is fused to the C-terminus for the cleavage experiments shown in Figure 5b. **b**, Cleavage of tag-fused MBP using formylation reagents **1**, **12**, and **13**. MBP (4 μM) was first treated with the probe (1 or 5 mM) for 1 hr and then treated with GSH (5 mM) for 1 hr in 100 mM sodium phosphate buffer (pH 7.4) containing 40 μM TCEP at 37 °C. **c**, Cleavage of the point mutated MBPs (4 μM) using the formylation reagents (**1**, **12**, or **13**), CDAP, or NTCB (5 mM, 1 hr) in the presence of GSH (5 mM, 1 hr) in 100 mM sodium phosphate containing 30% CH_3CN (pH 7.4, 37 °C). **d**, Structures of Cys-TAMRA and Cys-Biotin. **e**, Cleavage of MBP-GC (30 μM) using **12** (10 mM) and subsequent chemical labeling with Cys-TAMRA and Cys-biotin (1 mM, pH 7.4, 37 °C). left panel: in-gel fluorescence analysis, right panel: western blot analysis using streptavidin-horseradish peroxidase (HRP).

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

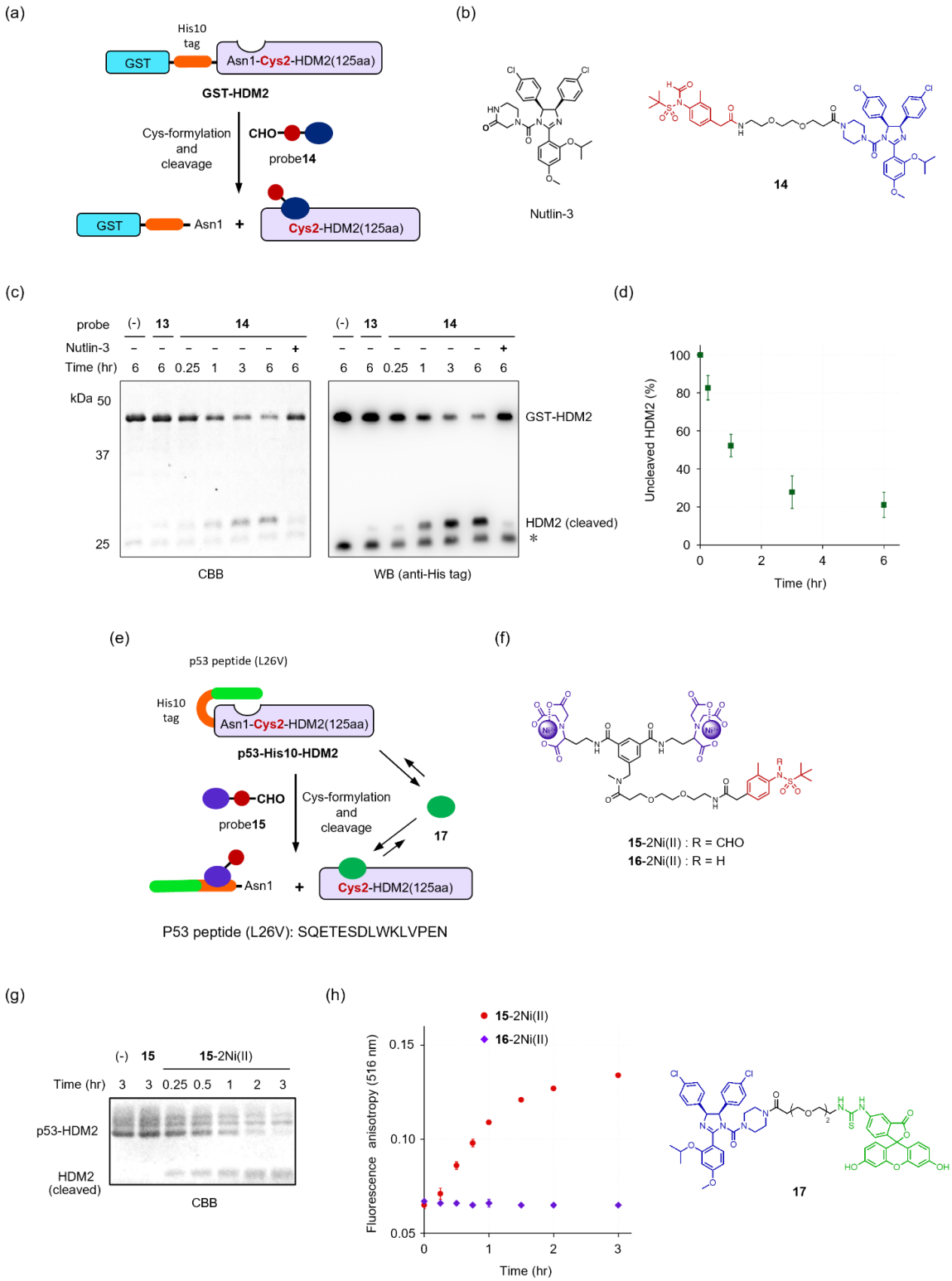


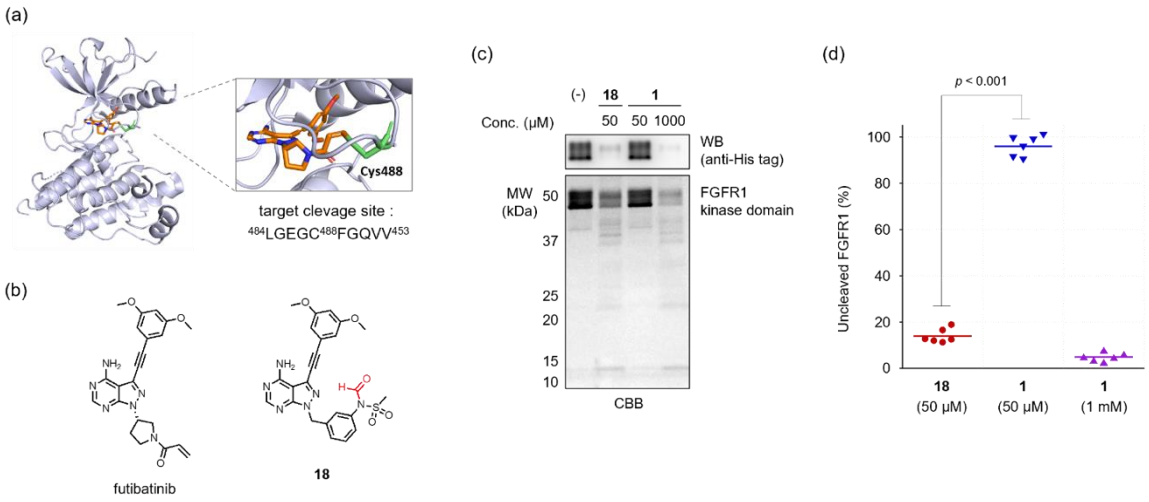
Figure 6. Affinity-based cleavage and functional regulation of HDM2.

a, Schematic illustration of the affinity-based cleavage of recombinant GST-HDM2 using formylation probe **14**. **b**, Structure of Nutlin-3 and formylation probe **14**. **c**, Cleavage of GST-HDM2 (2 μM) using **13** (100 μM) and **14** (100 μM) in 100 mM phosphate buffer (pH 7.4) at 37 °C. Left panel: CBB staining, Right panel: western blot analysis using anti-His tag antibody. **d**, Time trace plot of GST-HDM2 cleavage using **14** by western blot analysis (mean ± s.d.:

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

three independent experiments). **e**, Schematic illustration of functional regulation of p53-His10-HDM2 through peptide tag cleavage using probe **15**-2Ni(II). **f**, Structure of Ni(II)-NTA probe **15** targeting the oligo-histidine tag and the negative control probe **16** (without a formyl group). **g**, Cleavage of p53-His10-HDM2 (2 μM) using **15**-2Ni(II) (20 μM) or **15** (without $NiCl_2$) in 100 mM phosphate buffer (pH 7.4) at 37 °C. **h**, Time trace plot of the fluorescence anisotropy values of probe **17** upon treatment of p53-His10-HDM2 with **15**-2Ni(II) or **16**-2Ni(II). (mean ± s.d.: three independent experiments). p53-His10-HDM2 (1.5 μM) was treated with **15**-2Ni(II) or **16**-2Ni(II) (20 μM, 37 °C, 3 hr). The fluorescence anisotropy measurement was conducted using **17** (0.4 μM) in 100 mM phosphate buffer (pH 7.4) at 37 °C.

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0



405

406 **Figure 7. Affinity-based cleavage of FGFR1 kinase domain.**

407 **a**, X-ray structure of FGFR1 in complex with futibatinib (PDB:6MZW). **b**, Structures of futibatinib and formylation
408 probe **18**. **c**, Cleavage of FGFR1 kinase domain (1 μM) upon treatment with **18** (50 μM) or **1** (50 μM or 1 mM, pH
409 7.4, 37 °C, 3 hr) in 100 mM phosphate buffer (pH 7.4) containing 150 μM TCEP in the presence of GSH (1 mM, 1
410 hr) at 37 °C. (upper panel: western blot analysis using anti-His tag antibody, lower panel: CBB staining). **d,** Plot of
411 the relative band intensity of FGFR1 kinase domain upon treatment with **18** or **1.** The data were obtained from
412 western blot analysis shown in Figure **6c** (n = 6).

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

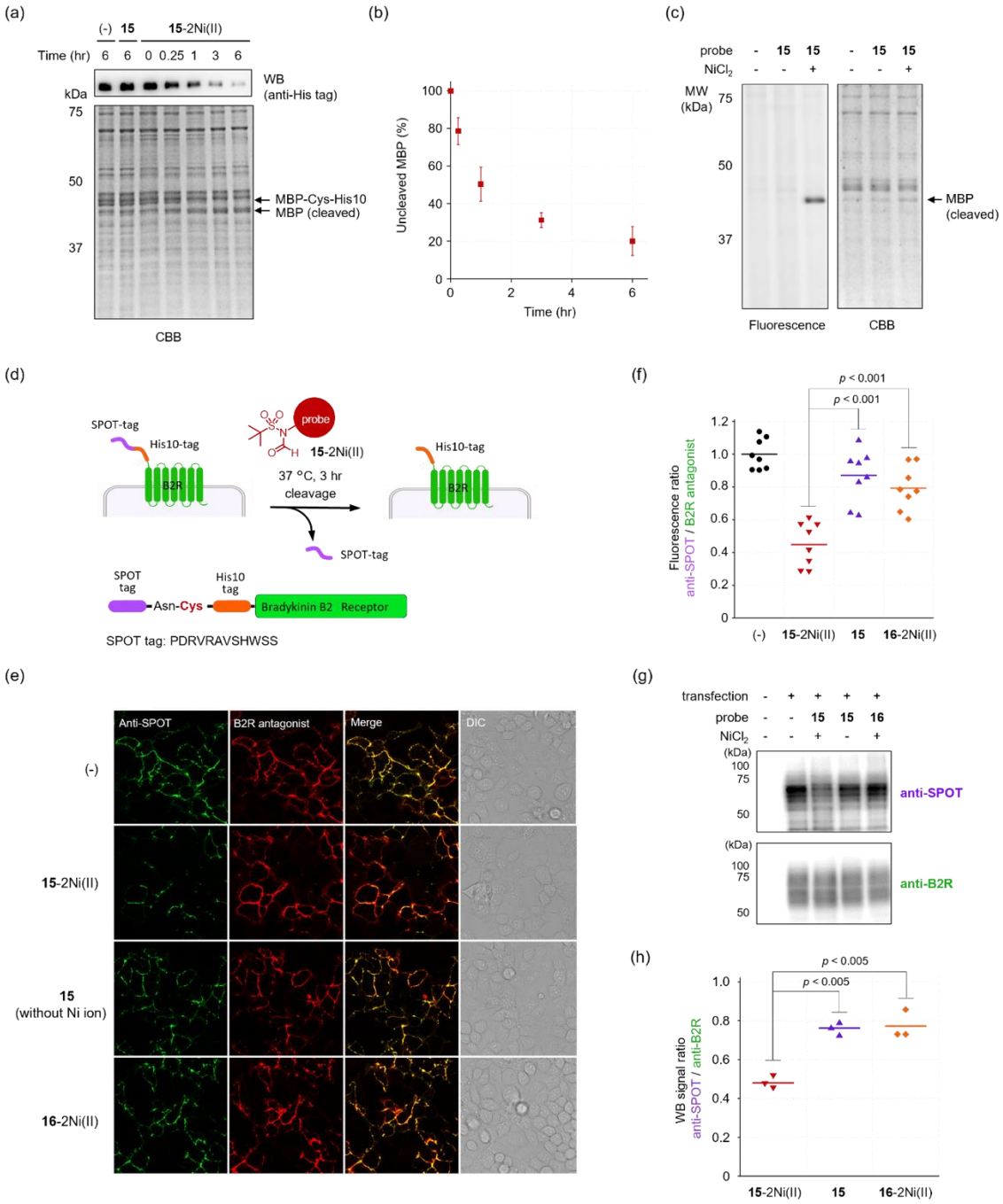


Figure 8. Chemical cleavage of proteins by cysteine *S*-formylation under biorelevant conditions.

a, Chemical cleavage of MBP-Cys-His10 (0.8 μM) using **15**-2Ni(II) (20 μM) in A431 cell lysate (1 mg/mL in 100 mM phosphate buffer, pH 7.4) at 37 °C. upper panel: western blot of MBP-Cys-His10 using anti-His tag antibody, lower panel: CBB staining. **b**, Time trace plot of the cleavage of MBP-Cys-His10 using **15**-2Ni(II) in A431 Lysate by western blot analysis (mean ± s.d.: three independent experiments). **c**, Cleavage of MBP-Cys-His10 (0.8 μM) using **15**-2Ni(II) (20 μM, 1 hr) in A431 lysate and subsequent fluorescence labeling with Cys-TAMRA (200 μM, 1 hr) at 37 °C. left: in-gel fluorescence analysis, $\lambda_{ex}/\lambda_{em}$ = 520/575 nm, right: CBB staining. **d**, Schematic illustration of chemical cleavage of SPOT-His10 tag fused B2R on the cell surface. **e**, Fluorescence imaging of HEK293T cells expressing SPOT-His10-B2R upon treatment with the formylation probes. The cells were treated with **15**-2Ni(II), **15**

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

(without $NiCl_2$) or **16**-2Ni(II) (40 μM, 3 hr) in HEPES-buffered saline (pH 7.4) at 37 °C. **f**, Plot of the fluorescence intensity ratio (Alexa Fluor 568-anti-SPOT nanobody / Cy5-anti-B2R antagonist peptide) of HEK293 cells (n = 8). **g**, Western blot analysis of SPOT-His10-B2R cleavage on the cell surface upon treatment with the formylation probes. **h**, Plot of band intensity ratios (anti-SPOT/anti-B2R) in the western blot analysis (n = 3).

https://doi.org/10.26434/chemrxiv-2024-cnt58 **ORCID:** https://orcid.org/0009-0007-5841-9778 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0