ChemComm
Chemical Communications

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: Y. Song, F. Xiong, J. Peng, Y. M. E. Fung, Y. Huang and X. Li, *Chem. Commun.*, 2020, DOI: 10.1039/D0CC01982H.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/chemcomm

View Article Online
DOI: 10.1039/D0CC01982H

COMMUNICATION

Introducing Aldehyde Functionality to Proteins Using Ligand-directed Affinity Labeling

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Yinan Song,[‡] Feng Xiong,[‡] Jianzhao Peng, Yi Man Eva Fung, Yiran Huang, and Xiaoyu Li [*]

Aldehyde is a versatile chemical handle for protein modification. Although many methods have been developed to label proteins with aldehyde, target-specific ones amendable to endogenous proteins are limited. We report a simple affinity probe strategy to install aldehyde on native proteins. Notably, the probe contains a latent aldehyde functionality that is only exposed upon target binding, thereby enabling a one-pot labeling procedure.

Modified proteins are widely used as powerful tools in many research fields including chemical biology, biotech, and medicine.[1-3] In the past decades, a plethora of protein labeling methods have been developed. Among them, site-specific incorporation of functional groups has been a major focus to generate homogeneous protein constructs. Aldehyde is a functional group not typically found in native proteins. Since the natural amino acid side chains are mostly nucleophilic, the electrophilicity of aldehyde makes it "bio-orthogonal" and could be exploited by various nucleophiles.[4, 5] Previously, many methods have been developed to label protein with aldehyde,[5] such as oxidative cleavage,[6-9] biomimetic transamination,[10, 11] and linker exchange.[12, 13] Enzyme-recognizing peptides could be fused with the target for site-specific aldehyde labeling using various enzymes.[14-18] The powerful unnatural amino acid mutagenesis and metabolic labeling methods have also been used to incorporate aldehyde (or ketone) into proteins.[19-21] However, chemical methods are mostly limited to purified proteins, while enzymatic and mutagenesis methods require genetic manipulation and may not be suitable for native proteins. Ligand-directed affinity labeling is an effective approach to achieve target specificity: the ligand guides the probe to bind the target and the proximity effect facilitates target-specific labeling. For example, the Hamachi group pioneered a series of powerful ligand-directed labeling methods leading to diverse applications in cell imaging, biosensing, and inhibitor discovery.[1, 22] The Ball group expanded the scope of modifiable side chains of affinity probes with metal catalysis.[23] However, introducing aldehyde to proteins with affinity probes has been largely underexplored. The studies by the groups of Hamachi[9, 12] and Gothelf[8] remain the only examples. Target-specific labeling of native proteins with aldehyde in complex biological environment, such as live cells, has not been reported. Here, we report a simple ligand-directed labeling strategy to introduce aldehyde to native proteins. As a unique feature, the aldehyde functionality is latent and is only exposed after target conjugation, thus enabling a convenient one-pot labeling procedure. Furthermore, this method also realises the labeling of membrane proteins with aldehyde on live cells.

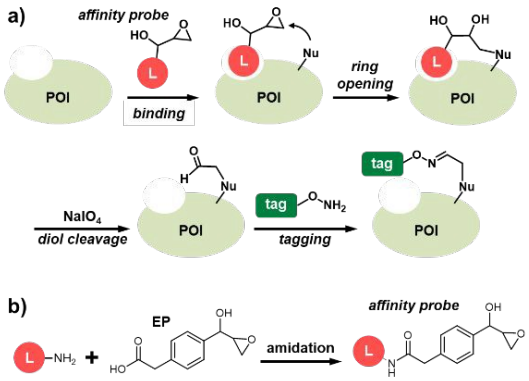


Fig. 1: a) Scheme for protein labeling with the epoxy-alcohol-based affinity probe. b) The precursor **EP** is coupled with the ligand (**L**) to form the affinity probe. POI: protein of interest; Nu: a nucleophilic side chain close to the ligand-binding site.

Our approach is shown in Fig. 1. The affinity probe has a very simple design: a ligand and an epoxy alcohol. Epoxide is a moderate electrophile and its reactivity could be improved by the proximity effect upon target binding. Previously, epoxides have been used in affinity probes, activity-based protein profiling, and covalent inhibitors.[24-28] Here, we propose to exploit the structure of the epoxide opening product. After the probe binds to the protein of interest (POI), a nearby nucleo-

Department of Chemistry and State Key Laboratory of Synthetic Chemistry, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China.
Email: xiaoyuli@hku.hk.

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

View Article Online
DOI: 10.1039/D0CC01982H

philic side chain opens the epoxide and forms a vicinal diol, which can be cleaved using periodate to generate the aldehyde. The ligand is released from the pocket and the aldehyde could be further modified with various tags. First, hiding the aldehyde functionality in the crosslinker simplifies the probe synthesis, since the probe is only "bi-functional" structurally but has a latent 3rd functionality. Second, the aldehyde will be exposed only after protein conjugation, while the free unbound probes remain "inactive", thus realizing a one-pot labeling procedure. The periodate cleavage condition is mild, rapid, and could be controlled to have minimal impact on protein target.[6-8, 29]

First, we prepared an epoxy-carboxylic acid precursor **EP** through a simple 3-step synthesis (Fig. S1a), which could be conjugated with an amine-bearing ligand (**L**) to furnish the probe (Fig. 1b). To verify the regioselectivity of epoxide opening, we tested the reaction of an epoxy alcohol with 2-mercaptoethanol and observed the formation of the desired vicinal diol product, where the nucleophilic attack took place at the less hindered carbon (Fig. S2). Next, we tested the method with a model protein, carbonic anhydrase II (CA-II). **EP** was coupled with 4-carboxybenzenesulfonamide (**CBS**), a known CA-II inhibitor (K_d: ~1 μM), to form the probe **CBS-EP** (Fig. 2a). To examine the stability of the epoxy alcohol towards nucleophile under physiological condition, **CBS-EP** was mixed with 5 mM glutathione in PBS (phosphate-buffered saline; pH 7.4); 4% and 10% epoxide opening were observed after 16 and 24 hours, respectively (Fig. S3). The probe also showed similar stability at pH 5.8, which is closer to the physiological pH around cell surface and in the cytosol of cancer cells (Fig. S4). **CBS-EP** also exhibited excellent stability under the typical oxidative cleavage condition (Fig. S5)[30] and showed similar CA-II inhibition activity as **CBS** (Fig. S6). Next, CA-II was incubated with **CBS-EP** and subjected to diol cleavage. Because of the probe's "turn-on" feature, the entire labeling procedure was performed in one pot; without removing the excess probes, the aldehyde-labeled CA-II was tagged with a fluorescent and a biotinylated aminooxy reagent, respectively (**FAM-AO** and **bio-AO**; Fig. 2a). As shown in Fig. 2b, in-gel fluorescence and Western blot clearly showed the specific labeling of CA-II and the labeling efficiencies were determined to be 48% (**FAM-AO**) and 69% (**bio-AO**), respectively (Fig. S7). All negative controls (free **CBS**, no $NaIO_4$, no **CBS-EP**) did not give significant labeling. The labeled CA-II was characterized with mass spectrometry (Fig. 2c), and it showed a ~50% labeling yield (**ii**), largely consistent with the gel analysis. The aldehyde formation was also confirmed in (**iii**); notably, the spectra showed a complete oxidative cleavage (**ii** to **iii**), suggesting that the protein-templated epoxide opening had the same regioselectivity as the intermolecular reaction (Fig. S2). Furthermore, the labeled CA-II was trypsinized, and tandem mass spectrometry analysis of the resulting peptides identified His64 as the modification site (Fig. 2d and S7b). His64 is located near the **CBS**-binding pocket and is the same modification site in a previous study also using an epoxide probe.[24] In addition, we tested the method with another protein: FKBP12 (12-kDa FK506-binding protein). A known ligand, **SLF**, was conjugated with **EP** to form the probe **SLF-EP** and used to

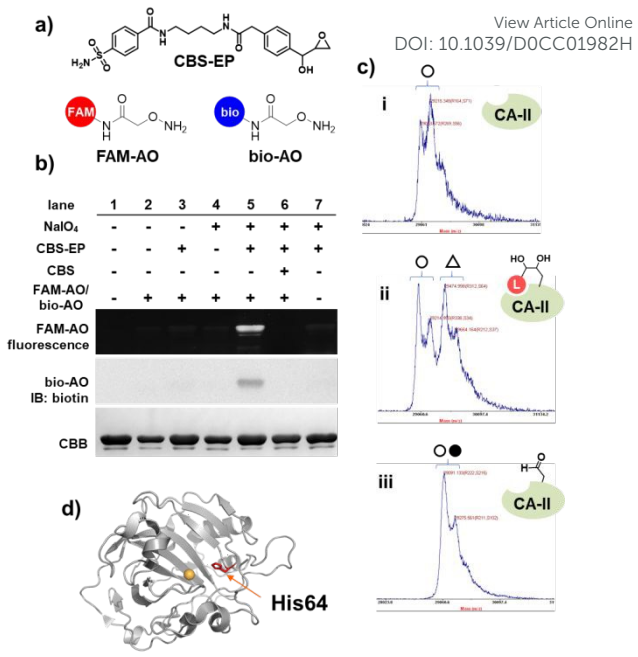


Fig. 2: a) Structure of **CBS-EP**, **FAM-AO**, and **bio-AO**. b) CA-II (10 μM) was labeled with **CBS-EP** (20 μM; 16 h, 37 °C), cleaved with $NaIO_4$ (1 mM, 4 °C, 30 min), and tagged with **FAM-AO** or **bio-AO** (40 μM; r.t. 90 min). The reactions were analyzed with SDS-PAGE and Western blot. c) MALDI-TOF analysis of the labeled CA-II; ○: CA-II (m/z=29025), Δ: **CBS-EP**-labelled CA-II (m/z=29475), ●: aldehyde-labelled CA-II (m/z=29091); L: **CBS**. d) Crystal structure of CA-II (PDB: 1V9E) with the modification site His64 highlighted.

label FKBP12 (Fig. S8). Western blot and MS results confirmed that **SLF-EP** specifically installed an aldehyde on His88 of FKBP12 (Fig. S8-S10), which is the same as in a previous study using a different type of affinity probe.[31]

Next, the labeling method was tested in complex biological contexts. Albeit being mild, periodate also cleaves the terminal sialic acids on glycoproteins in eukaryotic cells.[32, 33] To suppress this background, we used deglycosylases to remove the sialic acids before labeling. First, HeLa cells were lysed and treated with a deglycosylase mix (Deglyco-Mix kit; New England Biolabs); then, CA-II and FKBP12 were spiked in the lysates (2% w/w) and labelled with **CBS-EP** and **SLF-EP**, respectively. After diol cleavage, **bio-AO** was added to tag the aldehyde. There was no need to remove free probes and the **bio-AO** was added to the lysates directly after diol cleavage. As shown in Fig. 3a, both CA-II and FKBP12 could be specifically labeled in lysates. Furthermore, we tested the labeling method on live cells. Although ligand-directed labeling of membrane proteins has been realized,[34] installing aldehyde specifically on an endogenous membrane protein has not been reported. CA-12 is a membrane carbonic anhydrase implicated in malignant cancers. Since **CBS** is also a ligand for CA-12, **CBS-EP** was directly used. First, we verified that **CBS-EP** could specifically label the purified CA-12 in buffer (Fig. S11a); next, A549 cells were cultured in hypoxia to increase CA-12 expression (Fig. S11b) and then incubated with **CBS-EP**. After deglycosylation (Fig. S12) and oxidative cleavage, the cells were tagged with **bio-AO** and lysed;

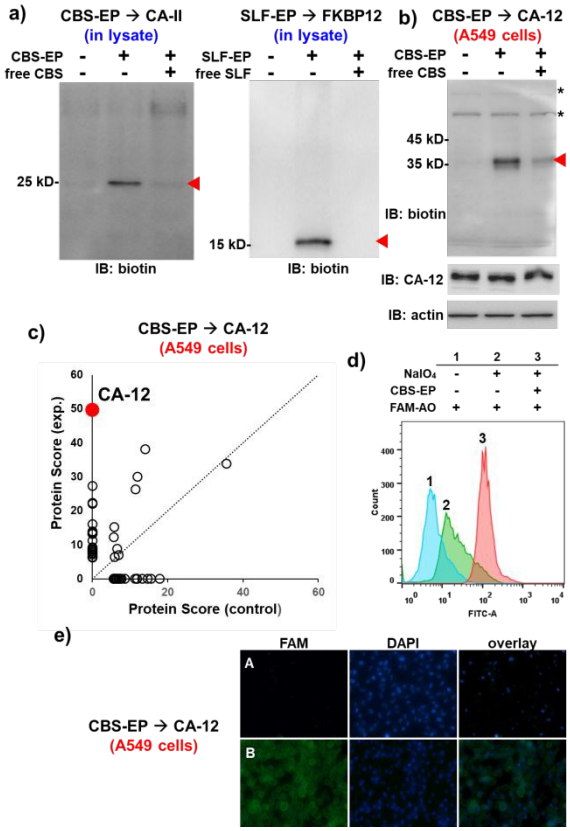


Fig. 3: a) Western blot analysis of CA-II and FKBP12 labeling in HeLa lysates with **CBS-EP** and **SLF-EP**. b) Labeling of the endogenous CA-12 on A549 cells using **CBS-EP**. Conditions are the same as in Fig. 2. Loading control: Fig. S13; *: endogenously biotinylated protein; IB: immunoblotting. c) CA-12 on A549 cells were labeled with **CBS-EP**, tagged with **bio-AO**, affinity-purified, and analyzed with tandem MS. *y*-axis: with **CBS-EP**; *x*-axis: without **CBS-EP** (negative control). d)-e) Labeled CA-12 on A549 cells were tagged with **FAM-AO** and analyzed with flow cytometry and confocal imaging. **CBS-EP**: 4 μM, 37 °C, 4 h; $NaIO_4$: 1 mM, 4 °C, 5 min; tagging: **bio-AO**/**FAM-AO** (40 μM), aniline (5 mM), 4 °C, 90 min. In e) A: without **CBS-EP**; B: with **CBS-EP**. See Fig. S14a for flow cytometry histograms.

Western blot analysis showed specific CA-12 labeling (Fig. 3b). Next, the biotinylated proteins were isolated with streptavidin beads and characterized with MS, which confirmed the specific capture of CA-12 from the cell surface (Fig. 3c). In addition, the CA-12 labeled cells were also tagged with **FAM-AO**, and both fluorescence imaging and flow cytometry corroborated the specific labeling (Fig. 3d-3e). Although **CBS** is known to bind multiple CA isoforms, **CBS-EP** appeared to have only labeled CA-12. Since $NaIO_4$ is not cell permeable, in principle, only membrane CAs may be labeled, which includes CA-9 and CA-12.[35] Although both CA-9 and CA-12 are highly expressed on A549 cells,[36] only CA-12 could be efficiently labeled by **CBS-EP** in the purified form (Fig. S11c). We reason the lack of CA-9 labeling in Fig 3b and 3c may be due to that CA-9 might bind **CBS** weakly on the cell surface or it lacks a properly positioned nucleophile for probe crosslinking. In addition, the Hamachi group observed the similar phenomenon with a **CBS**-based affinity probe on A549 cells.[36]

Besides the aldehyde on the protein, the diol cleavage generates another aldehyde on the original probe that may also be tagged; although there was no extensive washing, the cell

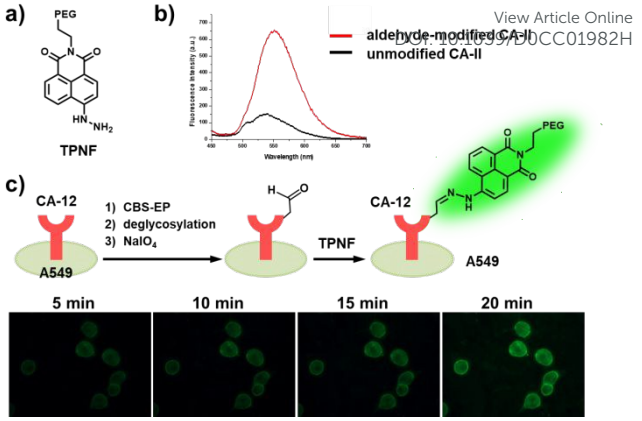


Fig. 4: a) Structure of **TPNF**. b) Fluorescence spectra of **TPNF** (100 μM) with CA-II (40 μM); red: labeled with aldehyde; black: unlabeled. Condition: 1x PBS (pH 6.7), 5 mM aniline, 30 min, r.t., excitation: 430 nm. b) Scheme for the labeling of CA-12 on A549 cells and the real-time cell images. After installing the aldehyde, the cells were treated with **TPNF** (10 μM), aniline (5 mM) in PBS buffer (pH 6.7) at room temperature.

labeling involved buffer exchanges to provide the optimal conditions for each step (see the SI for details), which may have washed off some of the non-covalently bound probe. Thus, we also performed the CA-12 labeling with a completely "no-wash" procedure and used free **CBS** to displace the probe after diol cleavage. The results showed similar cell fluorescence was obtained and only a slight signal decrease upon **CBS** competition (Fig. S14b), suggesting that the signal was mostly from the labeled CA-12. Next, the method was further tested with folate receptor (FR), another membrane protein and also a prime target for anti-cancer drug delivery on HeLa cells. The labeling was performed with **FA-EP**, a probe prepared based on the known FR ligand folic acid (**FA**; Fig. S15a). The results confirmed the specific capture of FR on live HeLa cells (Fig. S15). In addition, after the probe labeling, oxidative cleavage, deglycosylation, and **bio-AO** tagging, we observed little detrimental effect on cell viability, as well as on the ligand-binding affinity of CA-2 and FR (Fig. S16).

Fluorogenic probes, also known as "turn-on" probes, are widely used in imaging applications due to the high signal to noise ratio and the capability of real-time imaging.[37] Elegantly designed fluorogenic probes have been developed to detect formaldehyde *in vivo*.[38] However, fluorogenic imaging of proteins via an aldehyde tag has not been reported. To test this, we used a naphthalimide-based two-photon fluorescent probe (**TPNF**; Fig. 4a), which fluoresces upon hydrazone formation.[38-40] To limit the probe to cell membrane, **TPNF** was modified with a 5 kDa PEG to prevent it from entering the cell. First, we tested **TPNF** with purified CA-II either unlabeled or labeled with an aldehyde. As shown in Fig. 4b, the labeled CA-II exhibited significantly higher fluorescence upon **TPNF** addition, suggesting the aldehyde was able to activate **TPNF**. Next, we labeled CA-12 on A549 cells with an aldehyde using **CBS-EP**. After deglycosylation and cleavage, the cells were incubated with **TPNF**, and the generation of fluorescence on cell surface could be observed in real time (Fig. 4c). In addition, we also used **CBS** competition to eliminate the signal from the non-covalently bound probe. The results showed that the cell fluorescence was

mostly from the aldehyde label on CA-12 (Fig. S17). Finally, flow cytometry also confirmed the specific labeling (Fig. S18).

In conclusion, we have developed a simple method to install aldehyde on proteins. In this method, the latent aldehyde functionality enables a one-pot labeling procedure; the free unbound probes were not "activated" and this method is suitable for real-time imaging of membrane proteins. Like many affinity labeling approaches, the target specificity/isoform selectivity of the ligand has important implications for this method: the labeling outcome is a combination of all the proteins that can bind the ligand. Thus, additional measures and controls would be necessary to deconvolute the labeling outcome and the verification of the labeling specificity is always important. Unfortunately, the use of oxidative cleavage is a major limitation due to protein glycosylation. Deglycosylation is able to suppress the background, but it also alters the native proteome. In addition, cells contain a range of biomolecules having the diol moiety but cannot be removed by deglycosylases, such as the glycosylphosphatidyl- inositol (GTI)-anchored proteins, the glycosaminoglycan (GAG) chain on proteoglycans, glycolipids, and the furanose-containing molecules such as RNA and ATP. These molecules may complicate the labeling selectivity and careful control experiments would be required. For *in vitro* studies, this approach serves as a simple way to incorporate aldehyde into proteins; for native proteins, the method may be limited to certain applications where protein isolation is the main objective, such as target identification.

This work was supported by grants from the Research Grants Council of the HKSAR (AoE/P-705/16, 17321916, 17302817, 17301118, 17111319), the Laboratory for Synthetic Chemistry and Chemical Biology of Health@InnoHK of ITC, HKSAR, the NSFC of China (21572014, 21877093, 91953119).

Conflicts of interest

There are no conflicts to declare.

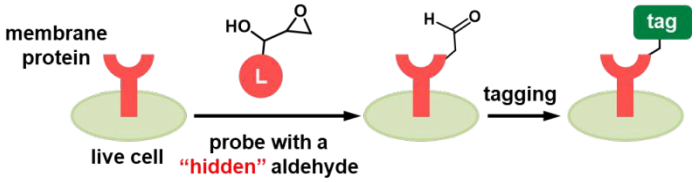
Notes and references

‡ These authors contributed equally to this work.

1. T. Tamura and I. Hamachi, *J. Am. Chem. Soc.*, 2019, **141**, 2782.
2. N. Krall, F. P. da Cruz, O. Boutureira and G. J. Bernardes, *Nat. Chem.*, 2016, **8**, 103-113.
3. L. Xue, I. A. Karpenko, J. Hiblot and K. Johnsson, *Nat. Chem. Biol.*, 2015, **11**, 917-923.
4. D. K. Kolmel and E. T. Kool, *Chem. Rev.*, 2017, **117**, 10358.
5. R. J. Spears and M. A. Fascione, *Org. Biomol. Chem.*, 2016, **14**, 7622.
6. K. F. Geoghegan and J. G. Stroh, *Bioconjug Chem*, 1992, **3**, 138.
7. Y. Zeng, T. N. Ramya, A. Dirksen, P. E. Dawson and J. C. Paulson, *Nat. Methods*, 2009, **6**, 207-209.
8. A. L. Kodal, C. B. Rosen, M. R. Mortensen, T. Torring and K. V. Gothelf, *ChemBioChem*, 2016, **17**, 1338.
9. T. Nagase, S. Shinkai and I. Hamachi, *Chem. Commun.*, 2001, **3**, 229.
10. J. M. Gilmore, R. A. Scheck, A. P. Esser-Kahn, N. S. Joshi and M. B. Francis, *Angew. Chem. Int. Ed.*, 2006, **45**, 5307.
11. J. I. MacDonald, H. K. Munch, T. Moore and M. B. Francis, *Nat. Chem. Biol.*, 2015, **11**, 326.
12. Y. Takaoka, H. Tsutsumi, N. Kasagi, E. Nakata and I. Hamachi, *J. Am. Chem. Soc.*, 2006, **128**, 3273.
13. H. Wakabayashi, M. Miyagawa, Y. Koshi, Y. Takaoka, S. Tsukiji and I. Hamachi, *Chem. Asian J.*, 2008, **3**, 1134.
14. I. S. Carrico, B. L. Carlson and C. R. Bertozzi, *Nat Chem Biol*, 2007, **3**, 321.
15. M. Rashidian, J. M. Song, R. E. Pricer and M. D. Distefano, *J. Am. Chem. Soc.*, 2012, **134**, 8455.
16. J. D. Cohen, P. Zou and A. Y. Ting, *ChemBioChem*, 2012, **13**, 888.
17. D. Schumacher, J. Helma, F. A. Mann, G. Pichler, F. Natale, E. Krause, M. C. Cardoso, C. P. Hackenberger and H. Leonhardt, *Angew. Chem. Int. Ed. Engl.*, 2015, **54**, 13787.
18. I. Chen, M. Howarth, W. Lin and A. Y. Ting, *Nat. Methods*, 2005, **2**, 99.
19. V. W. Cornish, K. M. Hahn and P. G. Schultz, *J. Am. Chem. Soc.*, 1996, **118**, 8150.
20. L. K. Mahal, K. J. Yarema and C. R. Bertozzi, *Science*, 1997, **276**, 1125.
21. A. Tuley, Y. J. Lee, B. Wu, Z. U. Wang and W. R. Liu, *Chem. Commun.*, 2014, **50**, 7424.
22. R. Kubota and I. Hamachi, *Chem. Soc. Rev.*, 2015, **7**, 4454.
23. J. Ohata and Z. T. Ball, *J. Am. Chem. Soc.*, 2017, **139**, 12617.
24. G. Chen, A. Heim, D. Riether, D. Yee, Y. Milgrom, M. A. Gawinowicz and D. Sames, *J. Am. Chem. Soc.*, 2003, **125**, 8130.
25. S. R. Adusumalli, D. G. Rawale, U. Singh, P. Tripathi, R. Paul, N. Kalra, R. K. Mishra, S. Shukla and V. Rai, *J. Am. Chem. Soc.*, 2018, **140**, 15114.
26. M. J. Evans, A. Saghatelian, E. J. Sorensen and B. F. Cravatt, *Nat. Biotechnol.*, 2005, **23**, 1303.
27. F. Yuan, S. H. Verhelst, G. Blum, L. M. Coussens and M. Bogyo, *J. Am. Chem. Soc.*, 2006, **128**, 5616.
28. S. R. Adusumalli, D. G. Rawale, K. Thakur, L. Purushottam, N. C. Reddy, N. Kalra, S. Shukla and V. Rai, *Angew. Chem. Int. Ed.*, 2020, DOI: 10.1002/anie.202000062.
29. H. K. Munch, J. E. Rasmussen, G. Popa, J. B. Christensen and K. J. Jensen, *Chem. Commun.*, 2013, **49**, 1936.
30. A. Sudalai, A. Khenkin and R. Neumann, *Org. Biomol. Chem.*, 2015, **13**, 4374.
31. T. Tamura, S. Tsukiji and I. Hamachi, *J. Am. Chem. Soc.*, 2012, **134**, 2216.
32. K. Zuberbuhler, G. Casi, G. J. Bernardes and D. Neri, *Chem. Commun.*, 2012, **48**, 7100.
33. Q. Zhou, J. E. Stefano, C. Manning, J. Kyazike, B. Chen, D. A. Gianolio, A. Park, M. Busch, J. Bird, X. Zheng, H. Simonds-Mannes, J. Kim, R. C. Gregory, R. J. Miller, W. H. Brondyk, P. K. Dhal and C. Q. Pan, *Bioconjug. Chem.*, 2014, **25**, 510.
34. S. Sakamoto, S. Kiyonaka and I. Hamachi, *Curr. Opin. Chem. Biol.*, 2019, **50**, 10.
35. C. T. Supuran, *Nat. Rev. Drug Discov.*, 2008, **7**, 168.
36. T. Miki, S. H. Fujishima, K. Komatsu, K. Kuwata, S. Kiyonaka and I. Hamachi, *Chem. Biol.*, 2014, **21**, 1013.
37. A. Nadler and C. Schultz, *Angew Chem Int. Ed.*, 2013, **52**, 2408.
38. K. J. Bruemmer, T. F. Brewer and C. J. Chang, *Curr. Opin. Chem. Biol.*, 2017, **39**, 17.
39. Y. H. Lee, Y. Tang, P. Verwilst, W. Lin and J. S. Kim, *Chem. Commun.*, 2016, **52**, 11247.
40. X. Li, B. Zhang, C. Yan, J. Li, S. Wang, X. Wei, X. Jiang, P. Zhou and J. Fang, *Nat. Commun.*, 2019, **10**, 2745.

View Article Online
DOI: 10.1039/D0CC01982H

Table of Content



An affinity probe with a "hidden" aldehyde functionality for protein labeling.