A Journal of the Gesellschaft Deutscher Chemiker

Angewandte International Edition Chemie

GDCh

www.angewandte.org

## Accepted Article

**Title:** Quantitative and Comparative Profiling of Protease Substrates through a Genetically Encoded Multifunctional Photocrosslinker

**Authors:** Dan HE, Xiao Xie, Fan Yang, Heng Zhang, Haomiao Su, Yun Ge, Haiping Song, and Peng Chen



**To be cited as:** *Angew. Chem. Int. Ed.* 10.1002/anie.201708151
*Angew. Chem.* 10.1002/ange.201708151

**Link to VoR:** http://dx.doi.org/10.1002/anie.201708151
http://dx.doi.org/10.1002/ange.201708151

WILEY-VCH

10.1002/anie.201708151

COMMUNICATION

WILEY-VCH

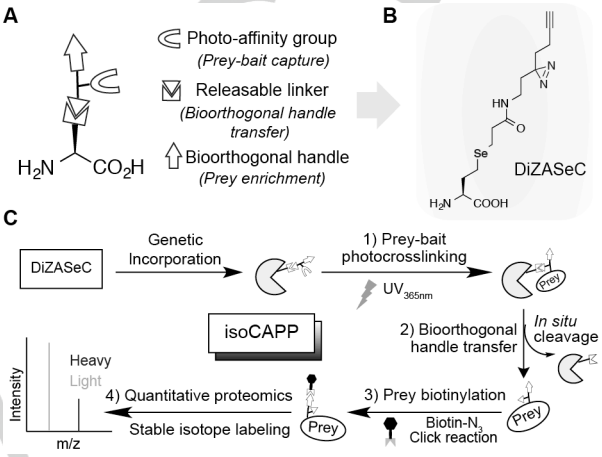
# Quantitative and Comparative Profiling of Protease Substrates through a Genetically Encoded Multifunctional Photocrosslinker

Dan He[a]§, Xiao Xie[a]§, Fan Yang[a], Heng Zhang[a], Haomiao Su[b], Yun Ge[a], Haiping Song[a], and Peng R. Chen*[a,c]

**Abstract:** We developed a genetically encoded, multifunctional photocrosslinker for quantitative and comparative proteomics. By bearing a bioorthogonal handle and a releasable linker in addition to its photo-affinity warhead, this probe enables the enrichment of transient and low-abundance prey proteins after intracellular photocrosslinking and prey-bait separation, which can be subject to stable isotope dimethyl labeling and mass spectrometry analysis. This quantitative strategy (termed isoCAPP) allowed us to adopt a comparative proteomic approach to identify the proteolytic substrates of an *E. coli* protease-chaperone dual machinery DegP. Two newly identified substrates were subsequently confirmed by proteolysis experiments.

Transient protein-protein interactions play essential roles in diverse biological processes such as signal transduction and proteolysis.[1] Genetically encoded photocrosslinking allows direct capture of these non-covalent interactions in live cells.[2] Typically, prey proteins are identified by mass spectrometry (MS) after affinity purification of the photocrosslinked prey-bait complexes. However, a significant technique challenge remains on deconvoluting the *bona fide* prey proteins of a given bait protein because many non-specific contaminants would be co-purified and this issue becomes more problematic for low-abundance interacting proteins.[3] Therefore, efficient enrichment of these photo-captured prey proteins is highly desired, which is essential for further quantitative MS analysis.

We have previously developed two genetically encoded releasable photocrosslinkers that were able to eliminate the non-specific contaminants from overexpressed bait proteins.[4] However, the *in situ* generated chemical handles on prey proteins after prey-bait separation are not fully bioorthogonal,[4a, 5] rendering it difficult to enrich the transient and low-abundance prey proteins such as protease's substrates from a cellular context (Fig. S1-S3). Herein, we report a genetically encoded, multifunctional photocrosslinker DiZASeC that bears a bioorthogonal **a**lkyne moiety in addition to the **diaz**irine photo-affinity group and the **Se**-based releasable linker (Scheme 1A, B). This probe ensured that only the photo-captured prey proteins can be tagged with alkyne group after photolysis and release, and thus allowed us to specifically enrich low-abundance interacting proteins and develop a quantitative proteomic strategy (termed isoCAPP, Scheme 1C, Fig. S4) for profiling protease substrates.




**Scheme 1.** The development of a genetically encoded trifunctional photocrosslinker DiZASeC and the quantitative isoCAPP strategy. A. Schematic view of the designed trifunctional photocrosslinker. B. Molecular structure of DiZASeC that contains a diazirine photo-affinity group, an alkyne handle and a C-Se releasable linker. C. Schematic overview of the isoCAPP strategy.

The DiZASeC photocrosslinker was designed to contain a minimalist terminal alkyne-containing diazirine moiety[6] and a Se-based releasable linker,[4b] which bears all the desired functionalities while remains as a structural analogue of our previously reported pyrrolysine-derived photocrosslinker DiZPK[7] to ensure high incorporation efficiency (Fig. S2). To our knowledge, DiZASeC is the first genetically encoded photocrosslinker that carries such an alkyne-diazirine dual-functional minimalist tag. The synthesis route and procedure for DiZASeC was described in Supporting Information (Fig. S5).

We first investigated the amber suppression efficiency and fidelity of DiZASeC by using an *E. coli* acid stress chaperone HdeA as the model protein. To our delight, a pyrrolysine aminoacylesynthetase (PylRS) variant and its $tRNA^{pyl}$ pair was able to recognize and site-specifically incorporate DiZASeC into HdeA (Fig. S6). The molecular weight (MW) of the DiZASeC-bearing protein was analyzed by ESI-MS, which verified the incorporation fidelity (Fig. 1A). Further, DiZASeC was found to possess similar photocrosslinking efficiency as DiZPK for capturing HdeA's client proteins, which suggests that the C-Se bond and alkyne group do not perturb the function of its photo-affinity diazirine moiety (Fig. 1B). Next, we examined whether the alkyne group on DiZASeC could be transferred to prey proteins after subjecting to the CAPP (cleavage after protein photocrosslinking) strategy.[4a, 8] Because HdeA forms a dimer at neutral pH,[9] DiZASeC incorporated into its dimer interface (e.g. at residue F35) was able to capture the other interacting

[a] D. H., X. X, F. Y., H. Z., Y. G., H. S., Prof. P. R. C.
Synthetic and Functional Biomolecules Center, Beijing National Laboratory for Molecular Sciences, Department of Chemical Biology, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China.
E-mail: pengchen@pku.edu.cn
[b] H. S.
College of Chemistry and Molecular Sciences, Wuhan University, Wuhan 430072, China.
[c] Prof. P. R. C.
Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China.
§ These authors contributed equally.

Supporting information for this article is given via a link at the end of the document

monomer upon photolysis. *E. coli* cells expressing HdeA-F35DiZASeC at pH 7 were photo-irradiated (365 nm) for 10 min and the generated covalently cross-linked HdeA dimer was subject to $H_2O_2$-mediated cleavage. Bioorthogonal labeling through the CuAAC reaction was subsequently performed by adding Cy5-$N_3$ and Cu-BTTAA to react with the alkyne handle. The reaction solution was subject to SDS-PAGE analysis and fluorescent imaging. Interestingly, the $H_2O_2$-treated samples without UV irradiation had undetectable signal on the fluorescence gel (Fig. 1C, lane 7), whereas UV-irradiated samples generated fluorescent signal that corresponds to HdeA monomer after $H_2O_2$ treatment (Fig. 1C, lane 3). These results demonstrate that only the successfully generated HdeA monomer bearing a transferred alkyne handle (represents the released prey protein) can be specifically labeled, confirming the efficient tagging of prey proteins after our CAPP strategy.

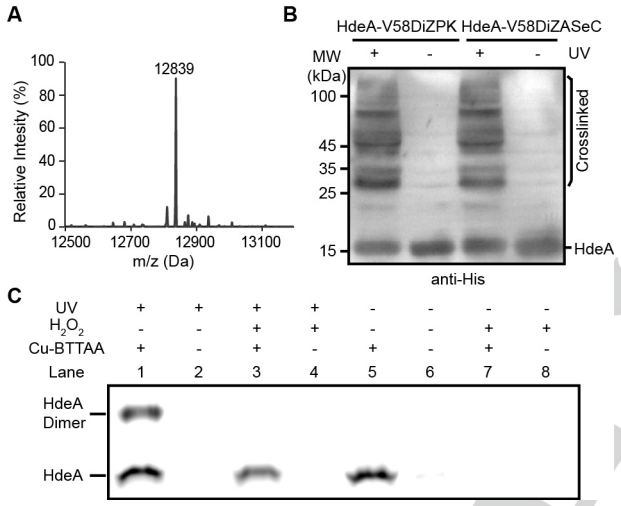



**Figure 1.** Verifying the multi-functions of DiZASeC. A. Successful incorporation of DiZASeC into a model protein verified by ESI-MS. The measured MW is 12839 Da (calculated 12840 Da). B. DiZASeC and DiZPK showing similar photocrosslinking efficiency. C. Fluorescent imaging gel showing specific tagging of prey proteins after our CAPP strategy. After photo-capture and oxidative cleavage, only cross-linked HdeA monomer was fluorescent labeled.

Next, we set to utilize a comparative proteomic strategy to study the proteolytic substrates of a protease. Despite playing fundamental roles in virtually all living organisms,[10] the proteolytic substrates for many proteases remain poorly understood.[11] Moreover, some proteases possess moonlighting functions such as chaperoning activity, which further complicated the identification of their high-fidelity proteolytic substrates.[12] For example, DegP is a key protein quality control (PQC) factor during outer membrane protein (OMP) biogenesis within the periplasmic space of Gram-negative bacteria.[13] It is a member of the high temperature requirement A (HTRA) family of serine proteases that belongs to the core set of protease found in cells and is widely conserved from single to multicellular organisms.[14] Additionally, it possesses a chaperone-protease dual function that is essential for protecting bacterial cell envelope from harsh environmental conditions such as heat shock or oxidative stress.[13, 15] Because identifying the direct proteolytic substrates of DegP remains difficult via traditional methods that rely on proteomic comparisons between cells harbouring active and inactive protease variants, we decided to directly photo-capture and enrich DegP-substrate interactions that can be subject to quantitative MS studies. To this end, we coupled our DiZASeC probe and isoCAPP strategy to investigate DegP's proteolytic substrates under heat shock conditions.

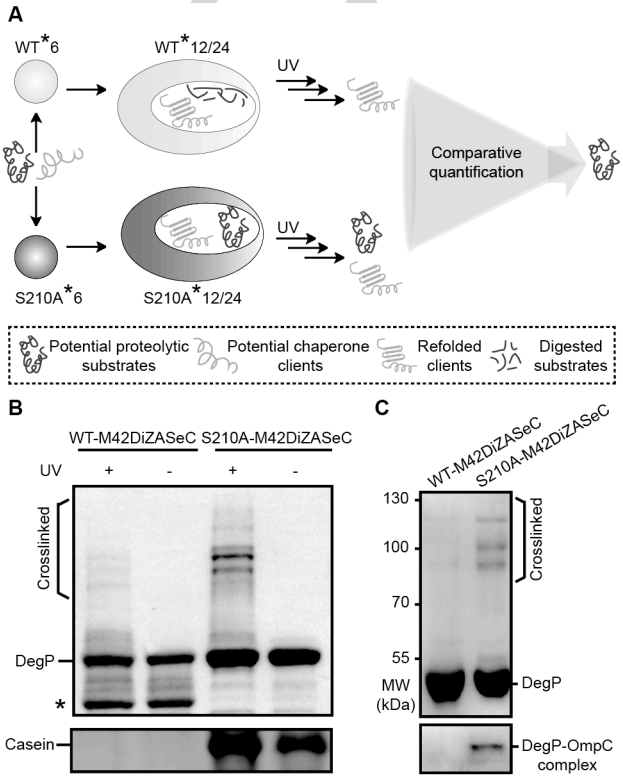



**Figure 2.** A comparative proteomic approach for profiling the proteolytic substrates of DegP by DiZASeC. A. Schematic model of discrepant cross-linked complexes of proteolytically active (WT) and inactive (S210A) DegP. The differences could be subsequently analyzed by comparative proteomics. B. Western blotting showing overwhelming cross-linked complexes of β-casein with S210A over those with WT detected by anti-His antibody (top panel). Coomassie blue staining gel showing cleaved β-casein by WT (bottom panel). Asterisk indicates the self-cleaved DegP. C. Western blotting results showing differences in crosslinking proteins of WT and S210A at 42 °C in living *E. coli* cells. OmpC is a known substrate of DegP and only cross-linked with S210A. Samples were analyzed with anti-His antibody (top panel) and anti-OmpC antibody (bottom panel).

We reasoned that wild-type DegP (DegP-WT) would digest its proteolytic substrates and only formed cross-linked complexes with its chaperone clients, whereas the protease deficient DegP-S210A variant would crosslink to both proteolytic substrates and chaperone clients upon photo-irradiation. Therefore, this comparative proteomic analysis may allow the identification of proteolytic substrates of this chaperone-protease dual function machinery. Notably, DegP changes its oligomeric state from the inactive hexamer to the active 12- and 24-mers to gain its proteolytic activity.[16] This possesses a formidable

challenge for adopting our comparative proteomic analysis by using non-cleavable photocrosslinkers because the altered oligomeric states will generate highly non-uniformed cross-linked complexes. Moreover, the hydrophobic cavity in DegP will render the biotin-based labeling probe inaccessible even under denatured conditions due to the steric hindrance (Fig. S7). In contrast, the cleavable feature of our DiZASeC probe makes the analysis unaffected by the change of oligomeric states, while the exposed alkyne handle can significantly facilitate the subsequent prey protein labeling and enrichment (Fig. S7). Together, we envisioned that our DiZASeC probe was particularly suitable for analyzing the proteolytic substrates of DegP (Fig. 2A).

To this end, we incorporated DiZASeC into both DegP-WT and DegP-S210A variants (Fig. S8). Residue M42 located on the regulating LA loop in DegP's protease domain was chosen as the incorporation site, which is involved in substrate recognition (Fig. S9).[16a, 17] The incorporation did not perturb proteolysis function of DegP as verified by detecting its proteolytic activity with resorufin-labeled casein (Fig. S10). The resulting proteins DegP-WT-M42DiZASeC and DegP-S210A-M42DiZASeC were incubated with β-casein at 37 °C for 30 min followed by photolysis at 365 nm for 10 min (Fig. 2B). As a common substrate for proteases, β-casein was efficiently digested by DegP-WT as visualized on the coomassie blue staining gel. The western blotting results showed that more cross-linked bands were formed with DegP-S210A than with DegP-WT. We then conducted photocrosslinking on live *E. coli* cells harboring DegP-WT or DegP-S210A. The *degP* mutant cells expressing DegP-WT-M42DiZASeC or DegP-S210A-M42DiZASeC were cultured at 42 °C (Fig. 2C) and 30 °C (Fig. S11), respectively, as the activity of DegP is temperature-dependent.[15c] Immunoblotting analysis demonstrated that the proteolytic activity of DegP in live cells at 42 °C was higher than that at 30 °C (Fig. S12). Accordingly, when cultured at 42 °C, DegP-S210A formed more cross-linked bands than DegP-WT, which is consistent with the *in vitro* results. Indeed, by using an antibody against OmpC, a previously known DegP substrate,[18] we verified the efficient crosslinking between OmpC and DegP-S210A (Fig. 2C). In contrast, both DegP-S210A and DegP-WT formed similar crosslinked bands when the bacterial cells were cultured at 30 °C (Fig. S11). Together, these results demonstrate the feasibility for revealing the proteolytic substrates of DegP by differential analysis of the photo-captured protein pools between DegP-WT and DegP-S210A.

Finally, we combined stable isotope labeling with this comparative proteomic approach (termed isoCAPP strategy) to identify the proteolytic substrates of DegP (Fig. 3A). The *degP* mutant *E. coli* strain expressing DegP-WT-M42DiZASeC or DegP-S210A-M42DiZASeC were grown at 42 °C for 6 h and subject to isoCAPP strategy. Briefly, after photo-irradiation, the oxidatively cleaved prey protein pool was enriched by click labeling with biotin-$(PEG)_2$-$N_3$ and pulling down by streptavidin resins. On-bead trypsin digestion was performed overnight after reduction with DTT (DL-Dithiothreitol) and alkylation with IAA (Iodoacetamide). The pools of DegP-WT and DegP-S210A were then subject to stable isotope dimethyl labeling with light and heavy labels, respectively.[19] Finally, the combined pools were subject to LC-MS/MS identification and quantification. Quality assessment of workflow was performed by examining the abundance of DegP proteins to guarantee the fidelity of our results (Fig. S13).

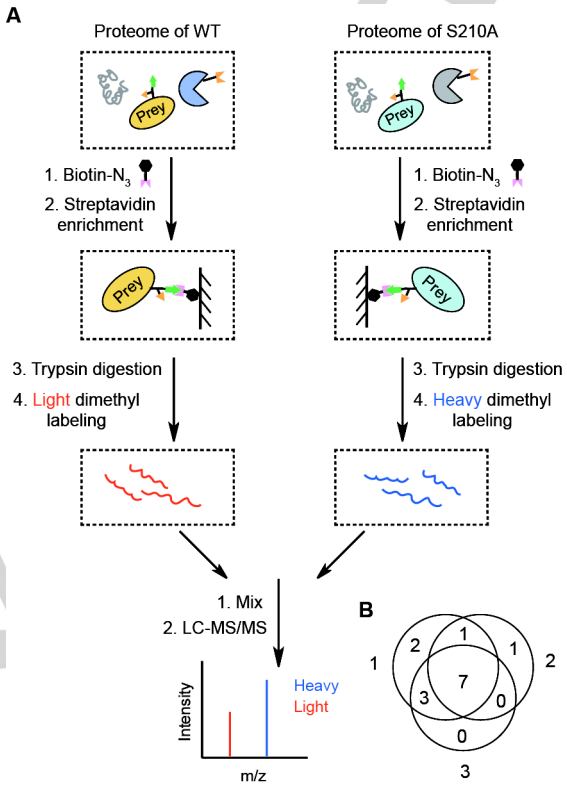



**Figure 3.** Identifying the proteolytic substrates of DegP by isoCAPP. A. Schematic workflow of quantitative proteomics analysis based on stable isotope dimethyl labeling (isoCAPP strategy). After enrichment and on-bead trypsin digestion, the prey proteins of DegP-WT were 'light' labeled while those of DegP-S210A were 'heavy' labeled. B. Venn diagrams illustrating the number of identified proteolytic substrates of DegP in three biological replicates.

An overview of all the protein identified was included in Supplementary Table 1. Proteins with over two-fold heavy/light ratio were assigned as the potential proteolytic substrates and the results from three biological replicates are summarized in Figure 3B. A total of 14 proteins met our criteria and were identified as the high-fidelity proteolytic substrates of DegP (Table 1). All the 8 previously known substrates were covered by this list (shown as grey in Table 1),[20] which demonstrated the robustness of our method. In addition, we also uncovered six new proteins as the DegP's substrates (shown as black in Table 1), which all have lower abundance than the other eight proteins (Supplementary Table 2). These results demonstrate the power of our trifunctional photocrosslinker DiZASeC and isoCAPP strategy for identifying low-abundance and transient protein-protein interactions. The proteolysis test was conducted on OmpN (Outer membrane protein N) and MetQ (D-methione binding lipoprotein), two newly identified DegP's proteolytic substrates from this study. The decreasing amount of OmpN

and MetQ in the presence of DegP-WT (Fig. S14) confirmed that these two proteins were indeed the proteolytic substrates of DegP. This indicates DegP may be involved in other pathways in addition to OMP biogenesis and protection.

**Table 1.** List of identified proteolytic substrates of DegP using the DiZASeC probe and isoCAPP strategy.

| Name | MW (kDa) | Function |
|---|---|---|
| OmpC | 40 | Outer membrane protein C |
| OmpF | 39 | Outer membrane protein F |
| OmpA | 37 | Outer membrane protein A |
| OmpX | 19 | Outer membrane protein X |
| OmpW | 21 | Outer membrane protein W |
| NmpC | 40 | Putative outer membrane porin protein |
| MalE | 43 | Maltose-binding periplasmic protein |
| FkpA | 29 | FKBP-type peptidyl-prolyl cis-trans isomerase |
| OmpN | 41 | Outer membrane protein N |
| MetQ | 29 | D-methionine-binding lipoprotein |
| CyoA | 35 | Cytochrome bo(3) ubiquinol oxidase subunit 2 |
| CydA | 58 | Cytochrome bd-I ubiquinol oxidase subunit 1 |
| MtlA | 68 | PTS system mannitol-specific EIICBA component |
| FdoG | 112 | Formate dehydrogenase-O major subunit |

In summary, we developed a quantitative and comparative proteomics strategy that relies on a genetically encoded trifunctional photocrosslinker (DiZASeC) bearing a Se-based releasable linker and an alkyne-diazirine minimalist moiety. The bioorthogonal enrichment and stable isotope labeling capabilities allowed us to perform the quantitative isoCAPP strategy for identifying the proteolytic substrates of an essential *E. coli* PQC factor DegP under heat shock conditions. A total of 14 proteolytic substrates were uncovered including 6 newly identified proteins with lower abundance. Noteworthy, the releasable feature of DiZASeC permits the transfer of an alkyne handle to the captured prey proteins upon prey-bait separation, which ensures the efficiency and fidelity of the subsequent bioorthogonal labeling. Meanwhile, the enrichable feature of DiZASeC helps improve the detection limit and the signal-to-noise ratio, which facilitates uncovering of low-abundance substrates. By comparing the photo-captured protein pools between the proteolytically active and inactive variants, our strategy maintains the native proteolysis network and cellular context, and enables quantitative profiling of potential substrates, which is generally applicable to diverse proteases.[21] In addition, our isoCAPP strategy is suitable for quantitative comparison of the intracellular interaction networks of a given protein in response to different stimulus conditions.

## Acknowledgements

We thank Prof. Zengyi Chang for the DegP-WT and MalE plasmids and OmpC antibody; Prof. Chu Wang for valuable discussions; Prof. Xiang Zhou for offering help in compound synthesis. This work was supported by the National Natural Science Foundation of China (21521003 and 21432002) and National Key Research and Development Program of China (2016YFA0501500).

**Keywords:** genetically encoded photocrosslinkers• protein-protein interactions • proteomics

[1] a) S. E. Acuner Ozbabacan, H. B. Engin, A. Gursoy, O. Keskin, *Protein. Eng. Des. Sel.* **2011**, *24*, 635-648; b) A. G. Ngounou Wetie, I. Sokolowska, A. G. Woods, U. Roy, J. A. Loo, C. C. Darie, *Proteomics* **2013**, *13*, 538-557.

[2] a) N. D. Pham, R. B. Parker, J. J. Kohler, *Curr. Opin. Chem. Biol.* **2013**, *17*, 90-101; b) Y. Tanaka, M. R. Bond, J. J. Kohler, *Mol. Biosyst* **2008**, *4* 473-480; c) Y. Yang, H. P. Song, P. R. Chen, *Iubmb Life* **2016**, *68*, 879-886; d) S. Zhang, D. He, Z. Lin, Y. Yang, H. Song, P. R. Chen, *Acc. Chem. Res.* **2017**, *50*, 1184-1192.

[3] W. H. Dunham, M. Mullin, A. C. Gingras, *Proteomics* **2012**, *12*, 1576-1590.

[4] a) S. Lin, D. He, T. Long, S. Zhang, R. Meng, P. R. Chen, *J. Am. Chem. Soc.* **2014**, *136*, 11860-11863; b) Y. Yang, H. Song, D. He, S. Zhang, S Dai, S. Lin, R. Meng, C. Wang, P. R. Chen, *Nat. Commun.* **2016**, *7*, 12299.

[5] Z. Yu, Q. Lin, *J. Am. Chem. Soc.* **2014**, *136*, 4153-4156.

[6] Z. Li, P. Hao, L. Li, C. Y. Tan, X. Cheng, G. Y. Chen, S. K. Sze, H. M. Shen, S. Q. Yao, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 8551-8556.

[7] M. Zhang, S. Lin, X. Song, J. Liu, Y. Fu, X. Ge, X. Fu, Z. Chang, P. R. Chen, *Nat. Chem. Biol.* **2011**, *7*, 671-677.

[8] S. Zhang, D. He, Y. Yang, S. Lin, M. Zhang, S. Dai, P. R. Chen, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 10872-10877.

[9] F. Yang, K. R. Gustafson, M. R. Boyd, A. Wlodawer, *Nat. Struct. Biol.* **1998**, *5*, 763-764.

[10] X. S. Puente, L. M. Sanchez, C. M. Overall, C. Lopez-Otin, *Nat. Rev. Genet.* **2003**, *4*, 544-558.

[11] O. Julien, M. Zhuang, A. P. Wiita, A. J. O'Donoghue, G. M. Knudsen, C. S. Craik, J. A. Wells, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E2001-2010.

[12] a) V. M. Jarocki, J. L. Tacchi, S. P. Djordjevic, *Proteomics* **2015**, *15*, 1075-1088; b) C. J. Jeffery, *Curr. Opin. Struct. Biol.* **2004**, *14*, 663-668.

[13] T. Clausen, M. Kaiser, R. Huber, M. Ehrmann, *Nat. Rev. Mol. Cell. Biol.* **2011**, *12*, 152-162.

[14] M. J. Page, E. Di Cera, *J. Biol. Chem.* **2008**, *283*, 30010-30014.

[15] a) T. Clausen, C. Southan, M. Ehrmann, *Mol. Cell.* **2002**, *10*, 443-455; b) J. Skorko-Glonek, E. Laskowska, A. Sobiecka-Szkatula, B. Lipinska, *Arch. Biochem. Biophys.* **2007**, *464*, 80-89; c) C. Spiess, A. Beil, M. Ehrmann, *Cell* **1999**, *97*, 339-347.

[16] a) T. Krojer, J. Sawa, E. Schafer, H. R. Saibil, M. Ehrmann, T. Clausen, *Nature* **2008**, 453, 885-890; b) N. J. Thompson, M. Merdanovic, M. Ehrmann, E. van Duijn, A. J. R. Heck, *Structure* **2014**, 22, 281-290.

[17] D. Figaj, A. Gieldon, A. Polit, A. Sobiecka-Szkatula, T. Koper, M. Denkiewicz, B. Banecki, A. Lesner, J. Ciarkowski, B. Lipinska, J. Skorko-Glonek, *J. Biol. Chem.* **2014**, 289, 15880-15893.

[18] M. CastilloKeller, R. Misra, *J. Bacteriol.* **2003**, 185, 148-154.

[19] P. J. Boersema, R. Raijmakers, S. Lemeer, S. Mohammed, A. J. Heck, *Nat. Protoc.* **2009**, 4, 484-494.

[20] X. Ge, R. Wang, J. Ma, Y. Liu, A. N. Ezemaduka, P. R. Chen, X. Fu, Z. Chang, *FEBS. J.* **2014**, 281, 1226-1240.

[21] B. H. van den Berg, A. Tholey, *Proteomics* **2012**, 12, 516-529.

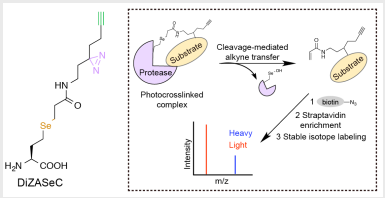
10.1002/anie.201708151


WILEY-VCH

## Entry for the Table of Contents

# COMMUNICATION

**Quantitative and Comparative Photocrosslinking**: A genetically encoded trifunctional photocrosslinker was developed that bears a bioorthogonal handle and a releasable linker in addition to its photo-affinity warhead. This probe enabled the enrichment of transient and low-abundance prey proteins after intracellular photocrosslinking and prey-bait separation. A quantitative and comparative proteomic strategy using this probe revealed the substrates of a protease.




*Dan He, Xiao Xie, Fan Yang, Heng Zhang, Haomiao Su, Yun Ge, Haiping Song, and Peng R. Chen**

***Page No. – Page No.***

**Quantitative and Comparative Profiling of Protease Substrates through a Genetically Encoded Multifunctional Photocrosslinker**