

A Journal of the Gesellschaft Deutscher Chemiker

# Angewandte Chemie

GDCh

International Edition

[www.angewandte.org](http://www.angewandte.org)

## Accepted Article

**Title:** Genetic encoding and enzymatic deprotection of a latent thiol side chain to enable new protein bioconjugation applications

**Authors:** Marie Reille-Seroussi, Pascal Meyer-Ahrens, Annika Aust, Anna-Lena Feldberg, and Henning D. Mootz

This manuscript has been accepted after peer review and appears as an Accepted Article online prior to editing, proofing, and formal publication of the final Version of Record (VoR). This work is currently citable by using the Digital Object Identifier (DOI) given below. The VoR will be published online in Early View as soon as possible and may be different to this Accepted Article as a result of editing. Readers should obtain the VoR from the journal website shown below when it is published to ensure accuracy of information. The authors are responsible for the content of this Accepted Article.

**To be cited as:** *Angew. Chem. Int. Ed.* 10.1002/anie.202102343

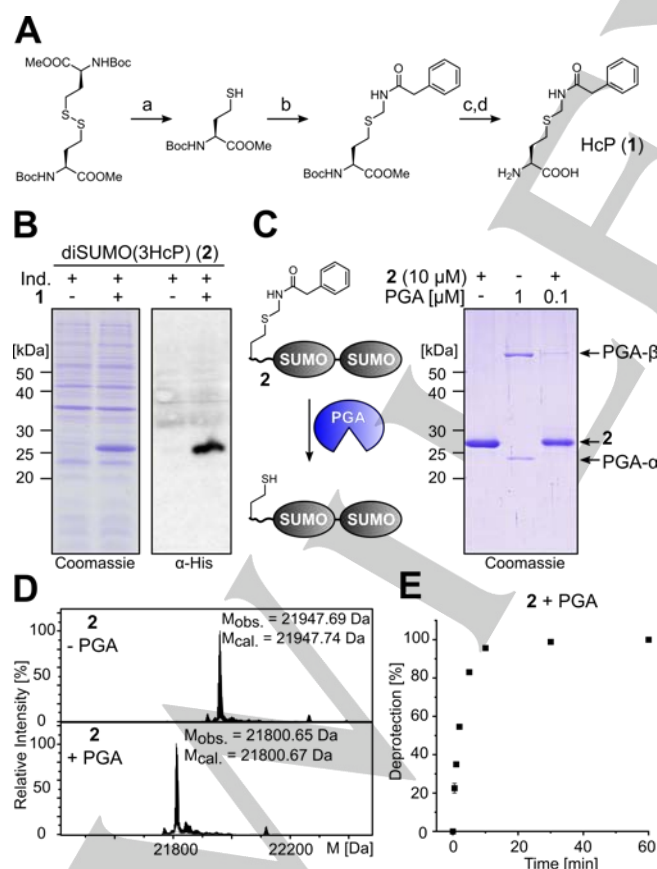
**Link to VoR:** <https://doi.org/10.1002/anie.202102343>



## COMMUNICATION

We chose the Phacm group for protecting a latent thiol because PGA is known to recognize and cleave its phenylacetate (Pac) portion and that the liberated thioaminal further decomposes to reveal the Hcy core structure (Scheme 1). The heterodimeric PGA enzyme is used on an industrial scale to remove the Pac group from penicillin G; a key step in the production of semi-synthetic penicillin variants.<sup>[10]</sup> PGA does not cleave regular peptide bonds.<sup>[10-11]</sup> It is also utilized on short polypeptide substrates in solid phase peptide synthesis (SPPS) to remove these groups from orthogonally protected Lys(Pac) and Cys(Phacm).<sup>[11-12]</sup>

HcP (**1**) was synthesized according to the scheme in Figure 1A and obtained in good yield as a soluble TFA salt. To establish incorporation during protein translation by amber stop codon suppression we found that a mutated *Methanosarcina barkeri* pyrrolysine tRNA synthetase (PylRS\*) evolved for lysine derivatives with bromoalkyl chains<sup>[13]</sup> also accepted **1** as a substrate. *Escherichia coli* BL21(DE3) cells were transformed with two plasmids encoding the PylRS\*/tRNA pair and a H<sub>6</sub>-diSUMO (small ubiquitin related modifier) model protein (**2**) with an amber stop codon at the third position in front of the histidine tag, effectively representing the second position after removal of the start methionine. **1** was incorporated with good efficiency (yielding 13 mg/L expression culture of purified protein) when added at 2 mM to the growth medium (Figure 1B). ESI-MS analysis of the purified diSUMO(3HcP) protein (**2**) confirmed the expected mass with no unprotected species detectable. This result verified the incorporation of **1** and suggested that the Phacm group remained stable during expression and purification.



**Figure 1.** Synthesis, incorporation and deprotection of HcP (**1**). **A**) Scheme of synthesis. **a** = 1.1 eq. tri-*n*-butylphosphine in DCM/H<sub>2</sub>O; **b** = 1.5 eq. *N*-hydroxymethylphenylacetamide, 0.08 eq. TsOH in dioxane; **c** = LiOH in H<sub>2</sub>O/dioxane; **d**

= 1.2 eq. HSiEt<sub>3</sub> in TFA/DCM. Yield = 46%. **B**) Coomassie-stained and α-His western blot SDS-PAGE analysis of incorporation of **1**. Ind. = induced expression. **C**) PGA-His<sub>6</sub>-mediated deprotection of HcP side chain in protein **2**. Left panel: scheme of the reaction; right panel: Coomassie-stained SDS-PAGE of purified proteins and deprotection reaction as indicated. **D**) ESI-MS analysis of the reaction shown in **C** with 0.01 eq. PGA-His<sub>6</sub> (1 h, 25 °C). **E**) Time-course of the analysis shown in **C** & **D**.

Having established incorporation of **1** via genetic code expansion, we investigated removal of the Phacm group from the model protein using PGA. We prepared purified PGA tagged with a hexahistidine sequence at the C terminus of the β subunit by recombinant expression in *E. coli*<sup>[14]</sup> and tested the enzyme at different concentrations. To our delight, incubation of the protein substrate **2** (10 μM) with just 0.01 eq. PGA-His<sub>6</sub> led to quantitative deprotection, as confirmed by ESI-MS, and was virtually complete already after 10 min (Figure 1D & E). Even with a lower amount of PGA-His<sub>6</sub> (0.001 eq.) we still observed complete deprotection after 2 h (Figure S1). No protein degradation could be detected, even at high PGA-His<sub>6</sub> concentrations (10 μM ± 1 eq., tested up to 24h), consistent with previous reports that PGA does not cleave in the peptide backbone of proteins (Figure S2).<sup>[11]</sup> The ESI-MS analysis also allowed us to observe the thioaminal intermediate, further confirming the mechanism following PGA-mediated removal of the Pac moiety (Figure S3). Of note, this and all subsequent MS analyses with other proteins showed no indication of a potential adduct formation with formaldehyde, suggesting that no such undesired effects are caused by this side product of the Phacm-release mechanism (Scheme 1), likely due to the low concentration in the micromolar range.

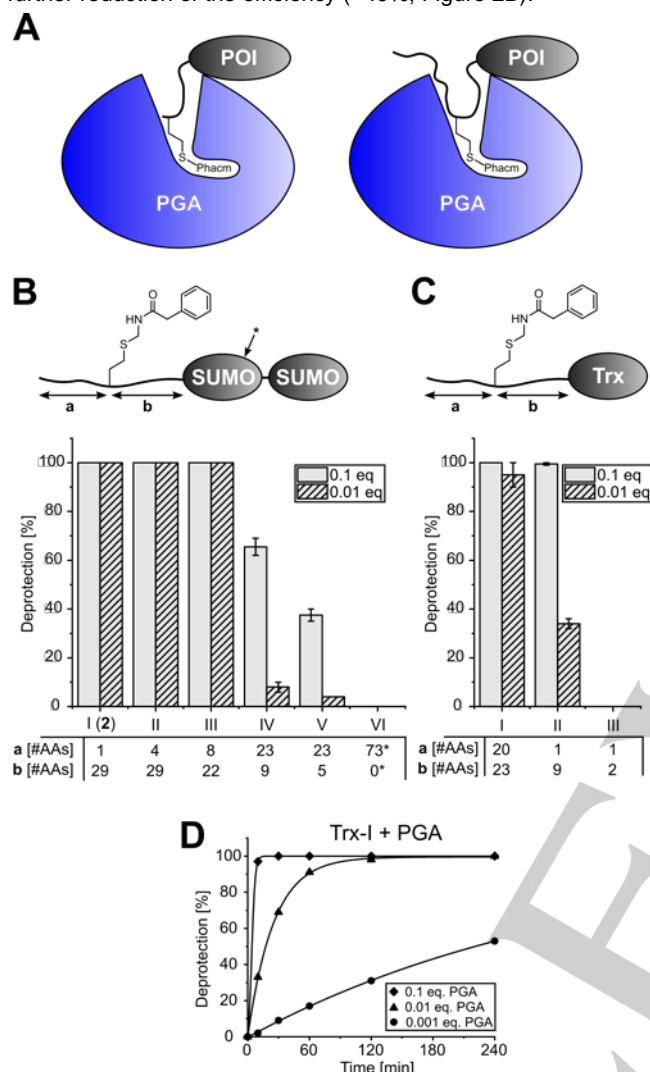
#### Structural requirements for PGA-mediated HcP deprotection

We then studied the structural requirements for such efficient and rapid PGA-mediated deprotection of HcP (**1**) in proteins. PGA's active site is at the bottom of a deep funnel that becomes increasingly constricted and should be easily accessible only by small molecules such as the native substrate penicillin G (Figure 2A).<sup>[15]</sup> Therefore, a clear dependency on the surrounding steric bulk was expected. The accessibility of the side chain of **1** should depend on the globular fold around its position.<sup>[13a]</sup> In our initial deprotection assays (Figure 1), **1** was located close to the N terminus of an unstructured tail region, thus exhibiting nearly maximal accessibility. We systematically tested other positions with varying distance from the N terminus in diSUMO model proteins (Figures 2B and S4-S6) and analyzed time-dependent deprotection by ESI-MS as described above. Constructs with **1** in a distance of 1, 4, or 8 aa from the N terminus and 20-29 aa away from the globular fold of the first SUMO domain were quantitatively deprotected under the chosen conditions (0.1 eq. PGA-His<sub>6</sub>, 4 h; see Figure 2B). These three constructs were even quantitatively deprotected with 0.01 eq. PGA-His<sub>6</sub> in ≤ 1 h (Figures 1E, S3 and S6), suggesting that the position of **1** within an unfolded peptide stretch is largely irrelevant for a rapid and efficient deprotection as long as it is located with enough distance to structured regions.

We then tested deprotection of **1** on the surface of the first SUMO domain (substituting R61), representing the other extreme in terms of structural accessibility. The globular SUMO fold encompasses ~72 aa. In fact, no deprotection could be observed, consistent with the notion that the SUMO domain would not fit deep enough into the funnel to the active site of PGA. Finally, we wondered about the minimally required distance of the **1** position

## COMMUNICATION

relative to a global protein domain for efficient deprotection to occur. With **1** being only nine residues away from the globular SUMO domain, the efficiency was lowered to about 65 % (0.1 eq. PGA-His<sub>6</sub>, 4 h). A shorter distance of only 5 residues led to a further reduction of the efficiency (~40%; Figure 2B).



**Figure 2.** Deprotection efficiency of HcP (**1**) with PGA. A) One-pass (left panel) and two-pass model (right panel) for a flexible polypeptide chain to reach through the funnel into the active site of PGA. B) Deprotection yield of diSUMO constructs with **1** at varying positions ( $t = 4$  h). \* = **1** on surface of globular part of SUMO. C) Deprotection yield of Trx constructs with **1** at varying positions ( $t = 1$  h). D) Deprotection time-course of protein Trx-I with different amounts of PGA-His<sub>6</sub>.

Similar findings were obtained using *E. coli* thioredoxin (Trx; folded globular domain of ~95 aa) with an artificial N-terminal extension as another model substrate protein (Figures 2C and S7). Notably, quantitative deprotection was observed with **1** being flanked by 20-23 aa on both sides as well as with **1** just 9 aa from the globular domain. In contrast, a distance of only 2 aa led to complete impairment (using 0.1 eq. PGA-His<sub>6</sub> for 4h). A further lowering of the PGA-His<sub>6</sub> concentration to 0.01 eq. showed for both model proteins that positions with a distance of only 9 aa to the globular domain are deprotected less rapidly (Figure 2B-D).

Together, these data are supporting a model that about  $\geq 8$ -10 aa are needed in unstructured or stretched conformation

to reach the PGA active site from the outer rim of its substrate funnel (of 25-30 Å distance) in a one-pass arrangement as shown in Figure 2A (left panel). Even a two-pass arrangement of the polypeptide chain in the funnel is possible and can lead to efficient and rapid deprotection of **1**, suggesting a significant increase of possible protein substrates (Figure 2B, right panel). HcP (**1**) in large globular structures and on flat surfaces is likely inaccessible for PGA but may still be cleaved at a slower rate depending on the architecture when applying longer incubation times and higher PGA amounts.<sup>[13a]</sup>

Of note, the rates and efficiencies observed here for PGA-mediated deprotection reveal an unexpectedly high catalytic potential on protein substrates. They represent a dramatic improvement over previously reported PGA deprotection schemes of Lys(Pac) or Cys(Phacm) on both peptide<sup>[12-13, 16]</sup> and protein<sup>[13a, 17]</sup> substrates, which were often incomplete and required overnight incubation. Next to the structural dependencies, it should be noted that PGA has typically been used as a commercial sample from industrial production batches, often in immobilized form. We recommend not to use such samples for protein applications due to potential problems with protein degradation and precipitation, likely due to their crude nature.<sup>[13a]</sup>

Having established protocols for HcP(**1**) incorporation and enzymatic deprotection to Hcy, we next aimed to develop new protocols in protein chemistry and protein labeling.

### Sequential and regioselective protein dual-labeling using bioconjugation of thiol and latent thiol groups

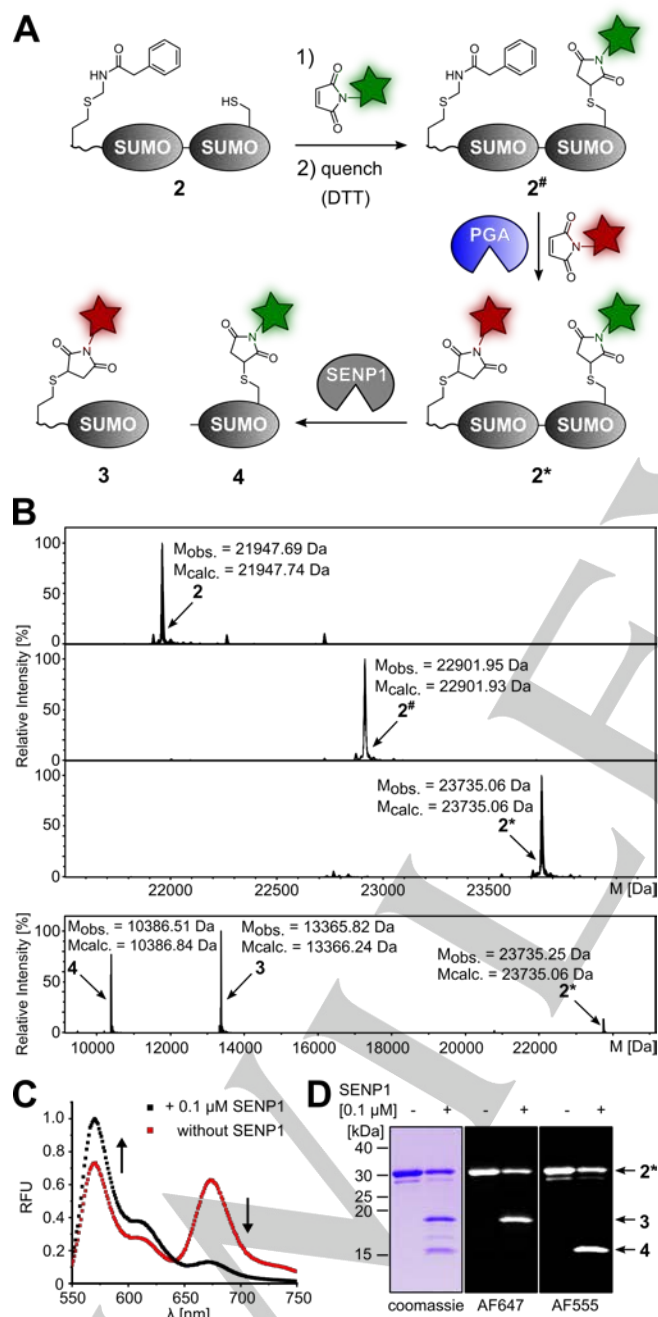
We reasoned that the latent thiol group of **1** would allow for a straight-forward sequential and regioselective dual-labeling strategy of a protein of interest (POI). Following labeling of a single Cys side chain, **1** would be converted to Hcy and the liberated thiol group then be labeled in a second reaction. Assuming no remaining reactive cysteine side chains after the first labeling step, this scheme represents a quasi orthogonal labeling of two thiol groups. Importantly, both labeling steps of this protocol can employ well-established and efficient thiol bioconjugation, which is appealing for its simplicity and availability of a plethora of commercial reagents. Other reported strategies for regioselective dual-labeling involve two different functional moieties, including more specialized types for bioorthogonal conjugation reactions.<sup>[18]</sup>

Dual-labeling of a protein with an acceptor and a donor dye is key to the Förster resonance energy transfer (FRET), a widely used technique to study intramolecular distances and conformational changes. We therefore tested the idea of sequential and regioselective Cys and Hcy labeling with a diSUMO FRET sensor. The design of the sensor was based on the above-described model protein **2** with the latent thiol group of **1** in the N-terminal region of the distal SUMO unit (Figure 3A). This protein also contained a single cysteine (R61C) in the proximal SUMO unit. In the first bioconjugation step, the thiol group of the cysteine was labeled with the donor dye AlexaFluor 555 (AF555) as a maleimide reagent to give **2\*** (Figure 3A). Following quenching with DTT (100 eq.) and subsequent removal of excess quencher and dye by dialysis, we added PGA-SBP and AF647 maleimide as the acceptor dye to simultaneously trigger conversion of HcP to Hcy and labeling of the latter to give dually labeled **2\*** (Figure 3A). PGA-SBP carries a streptavidin-binding peptide instead of the His<sub>6</sub> tag and therefore could easily be



## COMMUNICATION

removed together with the excess AF647 dye by Ni-NTA purification of the His<sub>6</sub>-tagged **2**<sup>\*</sup>. Notably, PGA and PGA-SBP have no cysteines and therefore do not cause any background labeling in such one-pot reactions. The dual-labeling of **2**<sup>\*</sup> without acceptor-acceptor or donor-donor conjugates was confirmed by ESI-MS (Figure 3B). **2**<sup>\*</sup> exhibited intramolecular FRET (Figure 3C), which was lost by enzymatic cleavage into the AF647 and AF555-labeled SUMO monomers **3** and **4**, respectively, with the SUMO protease SENP1 (Figure 3B bottom panel, Figures 3C & D and S8).<sup>[18e]</sup> Together, these data confirmed the intended regioselective conjugation of each SUMO monomer with just one specific dye.



**Figure 3.** Regioselective Cys and Hcy dual labeling of proteins. A) Scheme to prepare diSUMO FRET sensor **2**<sup>\*</sup> and its proteolytic cleavage. B) ESI-MS analysis of the unpurified reaction mixtures to analyze intermediates and products. C) Fluorescence spectra of dually labeled **2**<sup>\*</sup> (2 μM) with and without treatment with SENP1 (0.1 μM, 2 h) using excitation at 520 nm. C) SDS-PAGE analysis of the proteolytic cleavage reaction of **2**<sup>\*</sup> (10 μM) with SENP1 (0.1 μM,

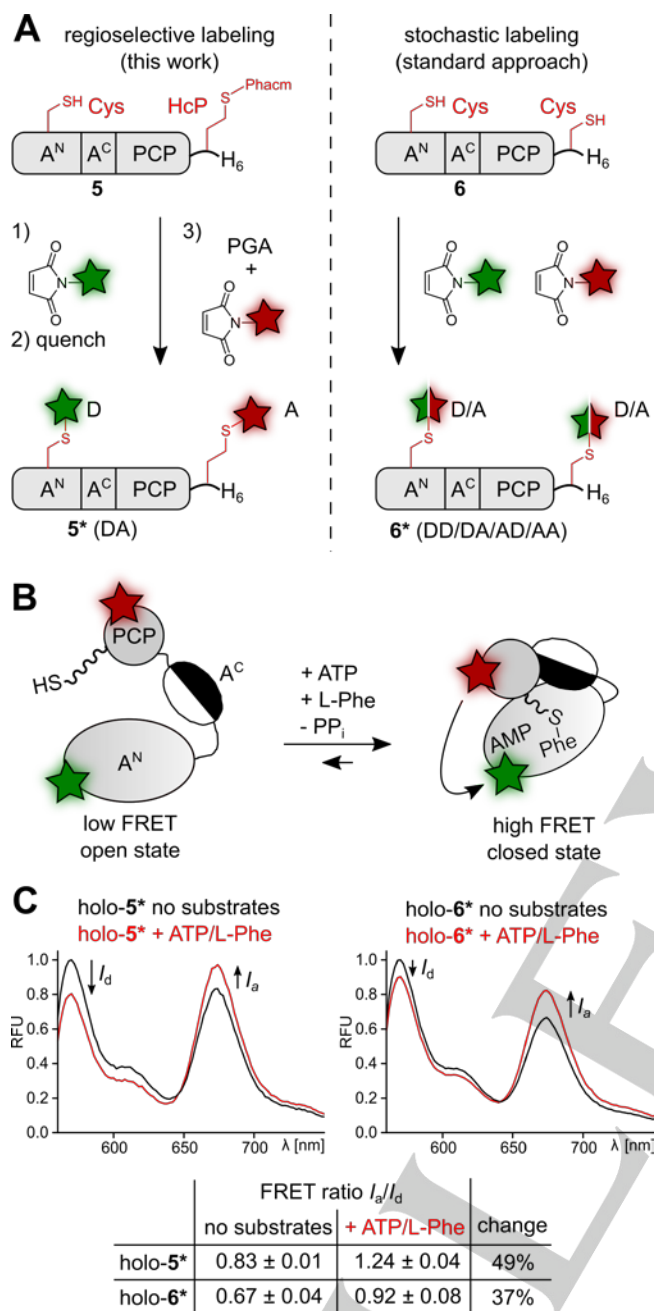
2 h). Excitation of the gel was performed at 609 nm for AF647 and 535 nm for AF555.

We then aimed to demonstrate the advantage of the new regioselective dual-labeling protocol over the typically used stochastic labeling of two cysteines in terms of the FRET response. Whereas the stochastic labeling cannot avoid the partial dual labeling with only acceptor and only donor dyes (Figure 4A, right panel), no such unproductive species are formed in our new approach and hence an increase in dynamic FRET range would be expected (Figure 4A, left panel). To this end, we designed a new nonribosomal peptide synthetase (NRPS) FRET sensor to monitor the interaction of its adenylation (A) and peptidyl-carrier-protein (PCP) domains, which together comprise 71 kDa. The in-solution analysis of the domain interaction is important to understand the timing and dynamics of conformational changes in these multi-domain biosynthetic enzymes.<sup>[19]</sup> The only currently reported FRET sensor design for these large, multidomain proteins utilized EGFP as a large fluorescent protein donor dye, which is not optimal due to its possible impact on domain mobility.<sup>[19]</sup> For a design based on two small synthetic dyes, we introduced a single cysteine at position N152C by site-directed mutagenesis in the large A<sup>N</sup> subunit of the A domain<sup>[19a]</sup> and fused a short tag (AGV(HcP)TEH<sub>6</sub>) containing the new amino acid at the C-terminal end of the PCP domain to give A-PCP(N152C/HcP-tag) (construct **5**). For comparison, cysteine was used at both positions to give the control construct A-PCP(N152C/Cys-tag) (**6**) (Figure 4A). The Cys-HcP construct **5** was labeled with AF555 and AF647 maleimides according to our new sequential protocol involving PGA-SBP to deprotect the latent thiol group. Excess dye and PGA-SBP were separated off by Ni-NTA chromatography. Successful regioselective dual-labeling (91%) was confirmed by SDS-PAGE and ESI-MS (Figures S9A and S10). The Cys-Cys control construct **6** was bioconjugated under the same conditions, however, by adding AF555 and AF647 maleimides simultaneously as a mixture to afford stochastic labeling of the two cysteines (Figure S9B). Subsequently, both labeled proteins **5**<sup>\*</sup> and **6**<sup>\*</sup> were converted into the catalytically active holo-forms by addition of the 4'-phosphopantetheinyl (Ppant) transferase Sfp<sup>[20]</sup> and coenzyme A (Figure S10) to give the FRET sensors holo-**5**<sup>\*</sup> and holo-**6**<sup>\*</sup>.

The A-PCP FRET sensors were then biochemically and biophysically characterized. The enzyme catalyzes L-Phe-AMP formation from the substrates L-Phe and ATP as well as the subsequent covalent binding of the amino acid as an L-Phe-thioester on the Ppant prosthetic group. These chemical reactions correlate with a shift of the conformational equilibrium to a preferential binding of the PCP domain to the A domain (Figure 4B).<sup>[19]</sup> Importantly, in the absence of substrates, the regioselectively labeled sensor holo-**5**<sup>\*</sup> showed a higher FRET ratio (acceptor intensity over donor intensity =  $I_a/I_d$ ) compared to holo-**6**<sup>\*</sup>, consistent with the lack of the unproductive donor-donor and acceptor-acceptor combinations (FRET ratios of 0.83 vs. 0.67, respectively; Figure 4C). Both FRET sensors reported on the conformational change induced by the addition of substrates ATP and L-Phe. Again, the regioselective design of holo-**5**<sup>\*</sup> proved to be more sensitive as it led to a greater change in the FRET ratio (49% vs. 37% for holo-**6**<sup>\*</sup>, respectively; Figure 4C), corresponding to a higher dynamic range. Together, these findings clearly demonstrate the advantage of regioselective over stochastic

## COMMUNICATION

labeling in FRET studies, and show the utility of our new labeling protocol for such purposes.

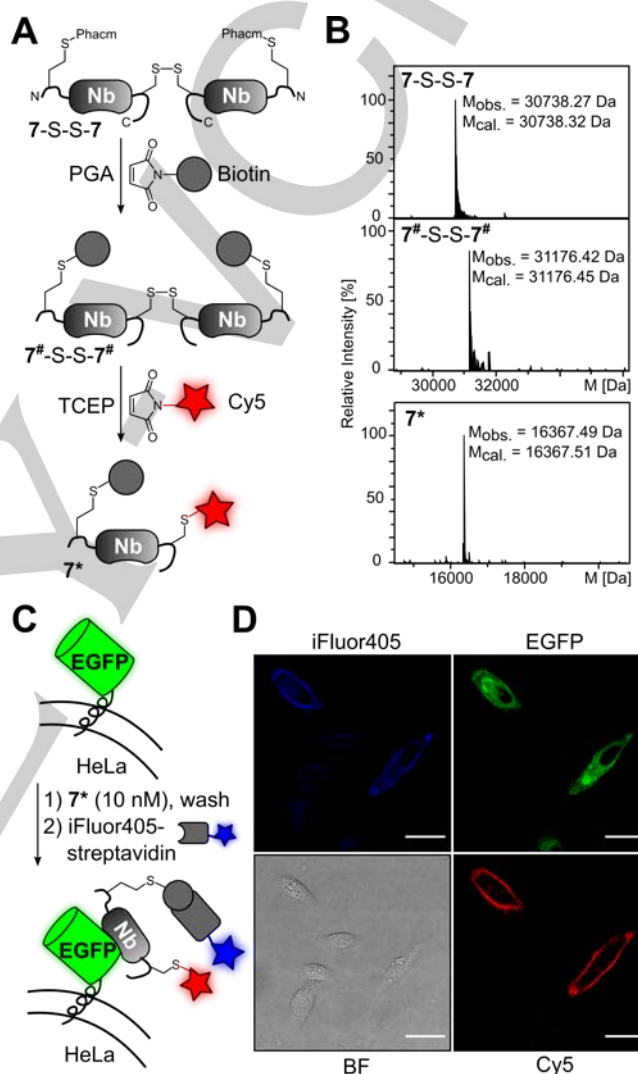


**Figure 4.** Intramolecular multidomain NRPS FRET-sensor improved by regioselective dual labeling. A) Schematic comparison of regioselective and stochastic labeling. B) Scheme of conformational change of the NRPS di-domain A-PCP-FRET sensor following substrate addition. C) Fluorescence spectra of holo-5\* or holo-6\* (0.3  $\mu$ M) to monitor FRET ratios with and without substrates (using excitation at 520 nm). Measurements with substrates were performed 30 min after adding 2 mM ATP and 2 mM L-Phe.

### Latent thiol bioconjugation in presence of intact cysteine disulfide

In another novel strategy for protein modification we aimed to take advantage of the fact that the latent thiol group of **1** does not interfere with existing disulfide bonds in the protein during expression and purification. This point is particularly noteworthy

in conjunction with the nascent nature of the Hcy thiol group following deprotection by PGA, which should allow quantitative bioconjugation without a preceding treatment with reductants like TCEP or DTT, as is typically required for cysteine bioconjugation following protein expression and purification. We hypothesized that protein thiol bioconjugation on the latent thiol group of **1** should be possible with preservation of disulfide bonds and with avoiding otherwise potentially harmful reducing conditions.



**Figure 5.** Latent thiol bioconjugation in presence of disulfide bond. A) Reaction scheme involving a disulfide-linked anti-GFP nanobody (Nb) with subsequent dual labeling after reducing the disulfide bond. B) ESI-MS analysis of unpurified intermediates in the reaction scheme shown in A). C) Scheme of Nb-mediated cell surface labeling of HeLa cells expressing EGFP. D) Confocal microscopy images of transiently transfected HeLa cells to present EGFP as shown in C). Note that non-transfected cells did not bind the nanobody. Scale bar = 25  $\mu$ m.

Nanobodies (Nbs) are single-domain antibody fragments derived from heavy-chain-only antibodies with countless applications in protein biochemistry, diagnostics and therapy.<sup>[21]</sup> We noticed that when nanobodies are expressed into the periplasm of *E. coli*, an extra single cysteine in a short appended tag sequence tends to form a disulfide with a second monomer to give a covalent homodimer. We therefore aimed to test our

## COMMUNICATION

hypothesis by selectively bioconjugating the latent thiol group of **1** in the presence of this disulfide bond between the two nanobodies. In an additional extension, the disulfide could be reduced to give the monomeric nanobody and the unpaired cysteine be used for a second labeling reaction (Figure 5A).

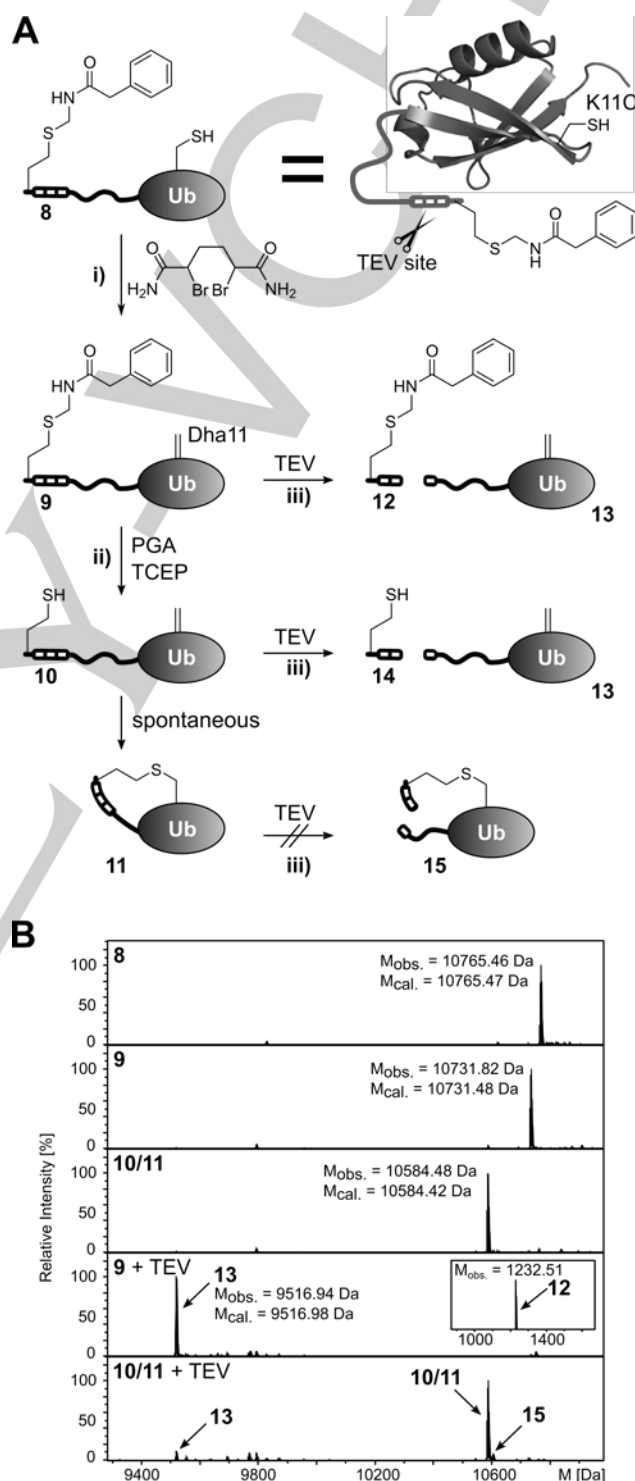
To this end, we prepared a GFP-enhancer nanobody<sup>[22]</sup> with a C-terminally appended cysteine and additionally fused a short N-terminal tag to introduce the latent thiol group of **1**. The resulting protein **7** was purified as a disulfide-bridged dimer **7-S-S-7** (Figure 5B) without any detectable free thiol of a monomeric species (Figure S11). Addition of PGA-His<sub>6</sub> (0.02 eq., 1 h) to convert **1** into Hcy and *in situ* labeling with biotin maleimide afforded the desired conjugate **7<sup>#</sup>-S-S-7<sup>#</sup>** with the intact disulfide bond in quantitative yield, as confirmed by ESI-MS (Figure 5B, top panels). Furthermore, to achieve subsequent dual-labeling, TCEP was added to **7<sup>#</sup>-S-S-7<sup>#</sup>** to reduce the disulfide and give the monomeric species of **7<sup>#</sup>**. Subsequent addition of Cy5 maleimide led to bioconjugation of the unpaired cysteine side chain and furnished the dually labeled nanobody **7\*** (Figure 5B, bottom panel and Figure S12; note that under these conditions the other, internal disulfide bond of the nanobody remained unaffected). Finally, we proved the functional integrity of **7\*** (10 nM) by its specific binding to EGFP presented on HeLa cells<sup>[23]</sup> and by visualizing the biotin moiety using iFluor405 conjugated streptavidin (Figure 5C and D).

#### A latent thiol for enzyme-activated protein crosslinking

Finally, the latent thiol group of **1** inspired us to a novel strategy to create stable protein crosslinks. We reasoned that as long as the latent thiol group is kept protected, the thiol group of a free cysteine could be chemically converted into a suitable electrophile. Deprotection of the latent thiol by PGA then reveals the nucleophilic reaction partner.

To test this idea, we chose to create an intramolecular thioether covalent crosslink on ubiquitin as the model protein (Figure 6A). We incorporated **1** into an unstructured N-terminal peptide extension, that also contained a TEV protease cleavage site, and introduced a single cysteine by site-directed mutagenesis on the surface of Ub's globular fold (K11C) to give protein **8**. The correct masses of this protein and the following modification steps were confirmed by ESI-MS (Figure 6B). Treatment of purified **8** with dibromoadipic amide converted Cys11 into dehydroalanine (Dha11; protein **9**).<sup>[24]</sup> Following removal of excess 2,5-dibromoadipic amide we added PGA-His<sub>6</sub> (0.01 eq.) to quantitatively deprotect the latent thiol of **1** to Hcy (protein **10**). Overnight incubation was necessary to allow for the slow spontaneous addition of the Hcy thiol to the Dha double bond to give cyclized protein **11** with the novel protein crosslink. However, since this reaction does not change the molecular mass of the protein we were unable to distinguish **11** from **10** by ESI-MS. To verify that the cyclization reaction had occurred, we reasoned that the constrained conformation of the cyclized peptide chain in **11** would likely twist the TEV cleavage site from its stretched conformation necessary for recognition by TEV protease and thereby partially or completely block TEV cleavage. In contrast, the uncyclized structure in **9** and **10** should be quantitatively cleaved (Figure 6A; right panels). Indeed, we found **9** to be quantitatively cleaved by TEV protease into **12** and **13**. In contrast, following incubation of the potential mixture of uncyclized **10** and cyclized **11** with TEV we could only detect a weak signal for cleavage product **13** and most of the signal for **10/11** remained unchanged (Figure 6B, bottom panel). This

finding is consistent with our hypothesis that the cyclized protein **11** was resistant against TEV cleavage, and suggests that a near quantitative intramolecular crosslink to **11** had occurred. Analysis of the protein species by SDS-PAGE electrophoretically separated cleaved **13** from **10/11** and corroborated an efficient yield of >90% for the crosslink reaction (Figure S13).



**Figure 6.** Intramolecular protein crosslinking using the latent thiol group. A) Scheme of the strategy involving the conversion of a single cysteine into an electrophile. The structural representation of the Ub model protein is based on pdb-file 1UBQ. Reaction conditions: i) dibromoadipic amide (10 eq., overnight, 37 °C); ii) PGA-His<sub>6</sub> (0.01 eq.), TCEP (5 eq.); overnight; iii) TEV protease (0.1



## COMMUNICATION

eq., 3h). B) ESI-MS analysis of the reaction intermediates and products (**12**:  $M_{\text{cal.}} = 1232.52$  Da).

Together, these results show that the latent thiol group in **1** can be used to efficiently introduce novel crosslinks into proteins. Notably, the formed thioether crosslink is of minimal atom economy, also due to the small size of the Hcy side chain. The latter contrasts to the mostly long and bulky side chains incorporated by the genetic code expansion technology for related purposes.<sup>[25]</sup>

## Conclusions

We have reported the genetic encoding of the novel non-canonical amino acid HcP (**1**) with a latent thiol group and its enzymatic deprotection to Hcy. The PGA-mediated deprotection was found to be surprisingly rapid and efficient at sterically well-accessible positions in unstructured regions, but was poor or impossible on flat surfaces of stably folded globular proteins. Importantly, the deprotection is performed under mild, physiological, non-denaturing and non-destructive conditions to the protein of interest, representing a significant advancement in terms of preparative utility over previously reported chemical or photochemical deprotection schemes of latent thiol or selenol groups. We have demonstrated the potential of the latent thiol group in several novel approaches of selective chemical modification of proteins, thereby significantly expanding the scope of thiol bioconjugation. The variety of these approaches reflect the versatility of thiol group chemistry. Protein thiol bioconjugation is a classical and well-established technique, and is still highly attractive due to its simplicity, chemoselectivity, efficiency and wide-spread use. The synthesis of the required unnatural amino acid **1** is easy to perform. The requirement for bioorthogonal reactions is circumvented in our chemical modification schemes. Bioorthogonal reactions require more sophisticated reagents of more restricted availability, in particular for dual labeling,<sup>[18]</sup> and are of more restricted applicability in many regards, for example as caused by the interference of cyclooctynes with free cysteines.<sup>[26]</sup> Our reported examples include a) two different routes to regioselective and consecutive dual labeling of cysteine and homocysteine (Hcy) as two quasi-orthogonal thiol groups, b) selective thiol conjugation under preservation of disulfide bonds that are otherwise sensitive to reducing conditions, and c) a protein crosslinking strategy to introduce stable thioether bridges. These protocols would not have been possible without the latent thiol group. We believe that they present powerful new tools for the protein chemist and biochemist and that the novel latent thiol group will enable many other new and unique applications.

## Acknowledgements

We thank Kathrin Lang and Sheng Yang for providing plasmids for the PyIRS/tRNA system and PGA expression, respectively. We gratefully acknowledge funding of this work by the Deutsche Forschungsgemeinschaft (DFG; grants MO1073/8-1 and SFB858/B14 to H. D. M.).

**Keywords:** nanobody • genetic code expansion • penicillin G acylase • thiol bioconjugation • regioselective labeling

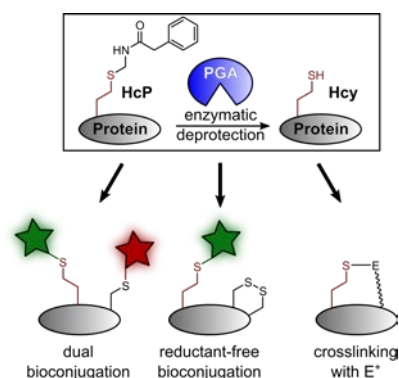
- [1] D. Fass, C. Thorpe, *Chem Rev* **2018**, *118*, 1169–1198.
- [2] S. B. Gunnoo, A. Madder, *ChemBiochem* **2016**, *17*, 529–553.
- [3] S. B. Kent, *Chem Soc Rev* **2009**, *38*, 338–351.
- [4] a) S. Virdee, P. B. Kapadnis, T. Elliott, K. Lang, J. Madrzak, D. P. Nguyen, L. Riechmann, J. W. Chin, *J Am Chem Soc* **2011**, *133*, 10708–10711; b) X. Li, T. Fekner, J. J. Ottesen, M. K. Chan, *Angew Chem Int Ed Engl* **2009**, *48*, 9184–9187; c) J. R. Frost, N. T. Jacob, L. J. Papa, A. E. Owens, R. Fasan, *ACS Chem Biol* **2015**, *10*, 1805–1816; d) E. Brustad, M. L. Bushey, A. Brock, J. Chittuluru, P. G. Schultz, *Bioorg Med Chem Lett* **2008**, *18*, 6004–6006; e) T. Liu, Y. Wang, X. Luo, J. Li, S. A. Reed, H. Xiao, T. S. Young, P. G. Schultz, *Proc Natl Acad Sci U S A* **2016**, *113*, 5910–5915; f) M. Koh, H. Y. Cho, C. Yu, S. Choi, K. B. Lee, P. G. Schultz, *Bioconjug Chem* **2019**, *30*, 2102–2105.
- [5] a) R. Upreti, J. Luo, J. Liu, Y. Naro, S. Samanta, A. Deiters, *ChemBiochem* **2014**, *15*, 1793–1799; b) N. Wu, A. Deiters, T. A. Cropp, D. King, P. G. Schultz, *J Am Chem Soc* **2004**, *126*, 14306–14307; c) D. P. Nguyen, M. Mahesh, S. J. Elsasser, S. M. Hancock, C. Uttamapinant, J. W. Chin, *J Am Chem Soc* **2014**, *136*, 2240–2243; d) D. P. Nguyen, T. Elliott, M. Holt, T. W. Muir, J. W. Chin, *J Am Chem Soc* **2011**, *133*, 11418–11421; e) J. Liu, F. Zheng, R. Cheng, S. Li, S. Rozovsky, Q. Wang, L. Wang, *J Am Chem Soc* **2018**, *140*, 8807–8816; f) W. Ren, A. Ji, H. W. Ai, *J Am Chem Soc* **2015**, *137*, 2155–2158; g) R. Rakauskaitė, G. Urbanaviciute, A. Ruksenaite, Z. Liutkeviciute, R. Juskenas, V. Masevicius, S. Klimasauskas, *Chem Commun (Camb)* **2015**, *51*, 8245–8248; h) K. Yang, G. Li, P. Gong, W. Gui, L. Yuan, Z. Zhuang, *ChemBiochem* **2016**, *17*, 995–998.
- [6] L. Johansson, C. Chen, J. O. Thorell, A. Fredriksson, S. Stone-Elander, G. Gavelin, E. S. Arner, *Nat Methods* **2004**, *1*, 61–66.
- [7] C. E. Hoyle, C. N. Bowman, *Angew Chem Int Ed Engl* **2010**, *49*, 1540–1573.
- [8] J. K. Böcker, W. Dörner, H. D. Mootz, *Chem Commun (Camb)* **2019**, *55*, 1287–1290.
- [9] C. C. Liu, P. G. Schultz, *Annu Rev Biochem* **2010**, *79*, 413–444.
- [10] K. Srirangan, V. Orr, L. Akawi, A. Westbrook, M. Moo-Young, C. P. Chou, *Biotechnol Adv* **2013**, *31*, 1319–1332.
- [11] D. Kadereit, H. Waldmann, *Chem Rev* **2001**, *101*, 3367–3396.
- [12] M. Royo, J. Alsina, E. Giral, U. Slomczynska, F. Albericio, *J. Chem. Soc., Perkin Trans. 1* **1995**, 1095–1102.
- [13] a) M. Reille-Seroussi, S. V. Mayer, W. Dörner, K. Lang, H. D. Mootz, *Chem Commun (Camb)* **2019**, *55*, 4793–4796; b) M. Cigler, T. G. Müller, D. Horn-Ghetko, M. K. von Wrisberg, M. Fottner, R. S. Goody, A. Itzen, M. P. Müller, K. Lang, *Angew Chem Int Ed Engl* **2017**, *56*, 15737–15741.
- [14] T. Cheng, M. Chen, H. Zheng, J. Wang, S. Yang, W. Jiang, *Protein Expr Purif* **2006**, *46*, 107–113.
- [15] H. J. Duggleby, S. P. Tolley, C. P. Hill, E. J. Dodson, G. Dodson, P. C. Moody, *Nature* **1995**, *373*, 264–268.
- [16] J. Tulla-Puche, M. Gongora-Benitez, N. Bayo-Puxan, A. M. Francesch, C. Cuevas, F. Albericio, *Angew Chem Int Ed Engl* **2013**, *52*, 5726–5730.
- [17] a) Q. C. Wang, J. Fei, D. F. Cui, S. G. Zhu, L. G. Xu, *Biopolymers* **1986**, *25 Suppl*, S109–114; b) L. Zakova, D. Zyka, J. Jezek, I. Handlova, M. Sanda, A. M. Brzozowski, J. Jiracek, *J Pept Sci* **2007**, *13*, 334–341.
- [18] a) E. M. Brustad, E. A. Lemke, P. G. Schultz, A. A. Deniz, *J Am Chem Soc* **2008**, *130*, 17664–17665; b) B. Wu, Z. Wang, Y. Huang, W. R. Liu, *ChemBiochem* **2012**, *13*, 1405–1408; c) I. Nikic, T. Plass, O. Schraidt, J. Szymanski, J. A. Briggs, C. Schultz, E. A. Lemke, *Angew Chem Int Ed Engl* **2014**; d) A. Sachdeva, K. Wang, T. Elliott, J. W. Chin, *J Am Chem Soc* **2014**, *136*, 7785–7788; e) L. J. Kost, H. D. Mootz, *ChemBiochem* **2018**, *19*, 177–184.
- [19] a) J. Alfermann, X. Sun, F. Mayerthaler, T. E. Morrell, E. Dehling, G. Volkmann, T. Komatsuzaki, H. Yang, H. D. Mootz, *Nat Chem Biol* **2017**, *13*, 1009–1015; b) F. Mayerthaler, A.-L. Feldberg, J. Alfermann, X. Sun, W. Steinchen, H. Yang, H. D. Mootz, *RSC Chem Biol* **2021**, <https://doi.org/10.1039/D1030CB00220H>.
- [20] R. H. Lambolt, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla, C. T. Walsh, *Chem Biol* **1996**, *3*, 923–936.



## COMMUNICATION

- [21] R. W. Cheloha, T. J. Harmand, C. Wijne, T. U. Schwartz, H. L. Ploegh, *J Biol Chem* **2020**, *295*, 15307-15327.
- [22] A. Kirchhofer, J. Helma, K. Schmidhals, C. Frauer, S. Cui, A. Karcher, M. Pellis, S. Muyldermans, C. S. Casas-Delucchi, M. C. Cardoso, H. Leonhardt, K. P. Hopfner, U. Rothbauer, *Nat Struct Mol Biol* **2010**, *17*, 133-138.
- [23] B. Jedlitzke, Z. Yilmaz, W. Dörner, H. D. Mootz, *Angew Chem Int Ed Engl* **2020**, *59*, 1506-1510.
- [24] J. M. Chalker, S. B. Gunnoo, O. Boutureira, S. C. Gerstberger, M. Fernandez-Gonzalez, G. J. Bernardes, L. Griffin, H. Hailu, C. J. Schofield, B. G. Davis, *Chem Sci* **2011**, *2*, 1666-1676.
- [25] a) V. Y. Berdan, P. C. Klauser, L. Wang, *Bioorg Med Chem* **2020**, *29*, 115896; b) H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat Chem Biol* **2008**, *4*, 232-234; c) T. Kobayashi, C. Hoppmann, B. Yang, L. Wang, *J Am Chem Soc* **2016**, *138*, 14832-14835.
- [26] R. van Geel, G. J. Pruijn, F. L. van Delft, W. C. Boelens, *Bioconjug Chem* **2012**, *23*, 392-398.

## Entry for the Table of Contents



Enzymatic deprotection of a latent thiol under mild and physiological conditions expands the scope of classic thiol bioconjugation. New reaction schemes enable regioselective labeling and dual modifications for a variety of preparative applications and using the established and commercially available plethora of thiol-directed reagents.

Institute and/or researcher Twitter usernames: ((optional))