Precise and Controlled Modification of Proteins using Multifunctional Chemical Constructs

Bauke Albada*[a]

Bioconjugation of chemical entities to biologically active proteins has increased our insight in the inner workings of a cell and resulted in novel therapeutic agents. A current challenge is the efficient generation of homogeneous conjugates of native proteins, not only when isolated, but also when still present in their native environment. To do this, various features of protein-modifying enzymes have been combined in artificial constructs. In this concept, the current status of this approach is evaluated, and the interplay between designs and protein modification will be discussed. Particular focus is directed on the protein-binding anchor, the chemistry that is used for the modification, and the linker that connects these two units. Suggestions how to include additional elements such as a trigger-responsive switch that regulated protein modification are also presented.

Introduction

The chemical modification of proteins in their native environment has provided detailed insights in their biological function and even enabled control over their biological role.[1–8] Alternatively, modification of isolated proteins has resulted in superior treatments, as is clear from the recent surge in the developments of antibody-drug conjugates (ADCs).[9] For both avenues, approaches have been developed to incorporate a unique chemically[10,11] and/or enzymatically[12–16] reactive moiety in the protein.[17] In this concept paper, modification strategies for wild-type proteins are surveyed and guiding principles for an integrated toolbox for the modification of a range of wild-type proteins are derived. Considering the high density of functional groups in a protein of interest (POI) (Figure 1a), and the added influence of substantial differences in the micro-environment of different areas of the protein that affect reactivity of local residues (Figure 1b), it is clear that the development of tools for protein modification that is controlled by externally applied triggers is not straightforward. Recently, bifunctional moieties that contain an anchor for the protein-of-interest (POI) and a reactive site that leads to the chemical modification of a specific amino acid residue have recently emerged (Figure 1c). This localizes the modification chemistry (S to P in Figure 1c) or leads to hydrolysis that deactivates the reactive group (S to S* in Figure 11c). As these approaches mimic different features found in protein-modifying enzymes (Table 1), several approaches will now be described and guiding principles for next generation tools are derived. Specifically, the
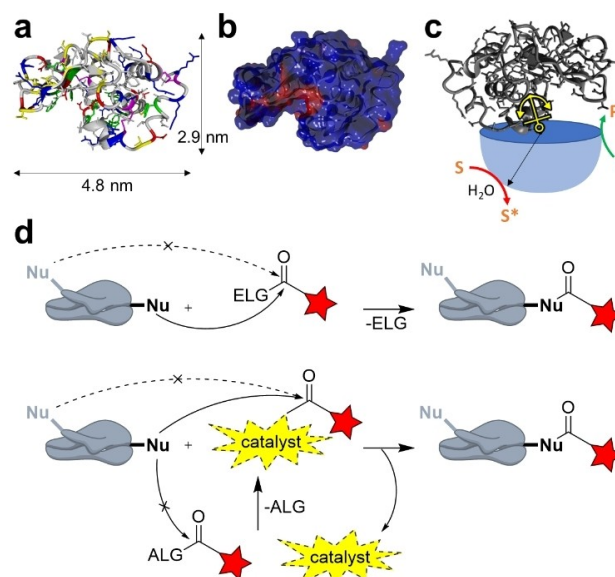
Figure 1. (a) Lysozyme as a typical protein with emphasis on its dimensions and functional amino acid residues (acidic = red [Asp, Glu], basic = blue [Arg, His, Lys], aromatic = green [Phe, Tyr Trp], disulfide = magenta [Cys-Cys]). (b) Differences in electrostatic potential of lysozyme (anionic patches in red, cationic patches in blue). (c) Schematic depiction of the application of a protein-anchor to direct modification to a specific region of the protein. The half dome indicates tha area accessible for the modifying unit. In proximity to the protein, a substrate S is connected to form modified protein P. This desired reaction competes with the undesired deactivation of the substrate S to a hydrolyzed substrate S*, for example. (d) Two types of commonly applied approaches for protein modification as exemplified as an arche-typical nucleophilic acyl substitution reaction: the application of a reactive substrate with an excellent leaving group (ELG) that acylates nucleophilic residues in the protein, and a catalytic acylation of a reactive nucleophilic residue using an otherwise unreactive substrate that contains an average leaving group (ALG).

following elements will be addressed: (i) POI anchor, (ii) linker, and (iii) modifying unit. The latter can be either (iiia) an activated functional group that contains an excellent leaving group (ELG) or (iiib) a catalyst that locally activates an otherwise

[a] Dr. B. Albada
Laboratory of Organic Chemistry
Wageningen University & Research
Stippeneng 4, 6708 WE, Wageningen (The Netherlands)
E-mail: bauke.albada@wur.nl

Table 1. Features used by enzymatic and chemical strategies for controlled modification of a protein of interest (POI). YES means present in (almost) all strategies, OR means present in some strategies (potentially combined), RA means rarely included in current strategies, NO means not yet incorporated.
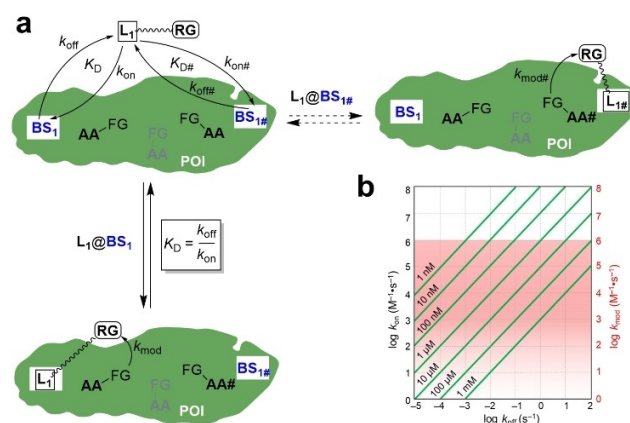
Feature	Enzymatic PTM	Chemical PTM
Reactive Site	YES	YES
POI anchor	YES	OR
Scaffold / Linker	YES	OR
Activity Switch	YES	RA
Reagent Binding Site	YES	NO
Release of Product (POI*)	YES	NO

Figure 2. Schematic simplified depiction of the kinetic parameters that control protein modification in which a ligand is used as anchor for the POI. (a) Two distinct binding sites (BS) are indicated, one is preferred (BS_1) and one is not preferred ($BS_{1\#}$). A ligand that is designed to bind at the preferred site leads to L–POI complex $L_1@BS_1$. However, formation of the undesired L–POI complex $L_1@BS_{1\#}$ can also lead to POI modification, especially when $k_{mod\#} > k_{mod}$. (b) The kinetic map of L–POI interactions, also known as a rate plane with isoaffinity diagonals (RaPID) plot, shows the dependence of dissociation constant (K_D, M) on rate of L-POI formation (k_{on}, s^{-1}) and L-POI dissociation (k_{off}, $M^{-1} \cdot s^{-1}$). For comparison, second order rate constants associated to reactions often used for protein modification are indicated by the red numbers.

unreactive substrate that contains an average leaving group (ALG) (Fig 1d).

Anchoring to the Protein-of-Interest (POI)

The modifying unit can be anchored to the POI by a reversible non-covalent interaction using a ligand, or by a reversible covalent linkage.[18,19] For the application of a ligand as POI anchor, molecules ranging from small to large can be used as long as embedding in a modification construct is tolerated. Whereas small ligands usually have only one site that tolerates modification with a POI-modifying unit, larger POI anchors can facilitate multiple attachment points, which enables its user to target different areas of the POI for modification. Our group showed that thrombin-binding aptamers enriched with catalytic moieties at different positions led to the modification of different sites and regions of the POI.[20] Importantly, protein-specific modification was shown using a mixture of proteins.[7]

This ligand-based approach works when the binding of the ligand is sufficiently strong (a low K_D, typically in the sub micromolar range). Also, the k_{on} for a specific binding site (BS_1) on the POI should be higher than for a competing binding site ($BS_{1\#}$) (see Figure 2a). In case the affinity ($K_D = k_{on}/k_{off}$) of the ligand for BS_1 on the POI is higher than for the modified POI (i.e., POI*) ($K_D > K_{D*}$), turnover numbers > 1 for each construct can be achieved. However, careful consideration of the incubation time is essential as rapid binding of the construct to an alternative binding site ($BS_{1\#}$) may lead to modification in that region, especially when this site contains a hyperreactive residue AA# ($k_{mod\#} > k_{mod}$) (Figure 2a). Depending on the second

order rate constant associated to the modification chemistry that is used (Figure 2b), incubation times prior to activation of the modification chemistry (i.e., only relevant for catalytic modification) should be calculated carefully and monitored precisely. In such cases, ligand-directed approaches that contain stable linkage between POI anchor and functionality that is bound to the POI lead to high second-order rate constants, even approaching that of enzymes. Incorporation of an average or even poor leaving group (ALG or PLG, respectively) is preferred as this leads to desired guided intramolecular modification and low unguided intermolecular background modification ($k_{modify,intra} \gg k_{modify,inter}$).[19,21] If a catalyst is used for the modification, precise control of the timepoint of substrate activation is in place, enabling the ligand-POI complex to reach equilibrium prior to modification. By selecting substrates for which $k_{modify,intra} \gg k_{modify,inter}$ and catalysts for which the modification rate for the bound state exceeds that of the unbound state ($k_{cat,free} < k_{cat,bound}$), unwanted background modifications can be almost completely suppressed.[22] For example, dimethylamine-4-pyridine (DMAP) catalysts that activate thioesters that consist of an average leaving group provide less control than pyridinium oxime (PyOx) catalysts that activate alkylated N-acyl-N-sulfonamide (ANANS) substrates, which contain a poor leaving group. Consequently, the high reactivity of the thioester substrates hindered optimization due to high non-catalyzed modification, whereas the inert nature of the ANANS substrate enabled optimization of the modification via various parameters (e.g., pH, T, reaction time).[20]

My group designed DNA-based constructs containing multiple features found in protein-modifying enzymes. To calibrate

Bauke Albada obtained his PhD-degree from the Utrecht University (the Netherlands), under supervision of prof. Rob Liskamp. After his post-doctoral stays at the labs of prof. Nils Metzler-Nolte (Bochum, Germany) and prof. Itamar Willner (Jerusalem, Israel), he started his independent career in the Laboratory of Organic Chemistry at the Wageningen University. In his group, he develops novel strategies to chemically modify wild type proteins, including the preparation of antibody-drug and antibody-protein conjugates.

the designs of our constructs, we covalently linked our constructs in order to avoid these dynamic interactions.[23] A future ultimate design would require incorporation of reversible POI anchors to achieve >1 turnover numbers. For example, we found that increasing the length of the linker between the POI and construct resulted in a decrease in modification efficiency but did not notably affect the site.[11] Importantly, however, this was not a universal trend, and was strongly dependent on the type of modification chemistry that was applied and the type of POI that was targeted. To be precise, modification of thrombin using a divalent DMAP catalyst and thioester combination *(vide infra)* was insensitive to extension of the linker with two ethylene glycol units (50% *vs* 57% for ethyl versus diethylene glycol, respectively), whereas implementation of a divalent PyOx catalyst and ANANS substrate showed 55% conversion for an ethyl spacer and 34% for the diethylene glycol spacer.[11] Clearly, linker composition and dimensions are crucial to control protein modification in integrated approaches.

Linking the POI anchor to POI modifying unit

As mentioned. specifics of the linker between the anchor and modifying unit greatly affect protein modification. Mathematical models for divalent POI targeting ligand construct[24] suggest that the modification efficiency and precision can be increased by reducing the linker length,[25] as this leads to a shorter reach and higher local concentration of reactive units (Figure 3a). For

example, trimming two units from an ethylene glycol spacer by reducing its length from 10 to 8 ethylene glycol units doubles the local effective concentration (C_{eff}) (Figure 3b). Provided that the targeted residues remain within reach after shortening, this strategy can be considered and has proven its values in recent examples.[25]

In our own work, we used the programmable and well-established dimensions of double-stranded DNA constructs that were covalently anchored to POIs in order to correlate the position POI-modifying catalysts to the modification efficiency and site of modification.[11] When a catalyst was implemented that produced a soluble reactive agent, no control over the modification efficiency was obtained. To be precise, the catalyst at 2.5 nm ($C_{eff,catalyst}$ 8.4 mM) from the POI yielded the same level of conversion (*i.e.*, 50%) as the catalyst at 9 nm ($C_{eff,catalyst}$ 0.2 mM). As this distance is larger than the diameter of most proteins (Figure 3c), it remains to be determined how useful this such catalytic entities are for precise protein modification, and the role of the linker in controlling this can be negligible. However, when we implemented a catalyst that generated a catalyst-bound reactive species, a gradual decrease in conversion was observed when the catalyst was positioned further from the POI. Specifically, the most active catalyst diPyOx *(vide infra)* positioned at 2.2 nm (C_{eff} = 12.5 mM) from the POI led to 65% conversion, whereas its positioning at 4.6 nm (C_{eff} = 1.3 mM) led to only 16% conversion. (Note: the C_{eff} is actually twice as high for each, due to the bivalent nature of the catalyst.) As expected, LC—MS-MS analysis of tryptic digests of the modified proteins revealed that positioning the catalyst more distant from the POI led to modification of residues beyond the periphery of the attachment point.
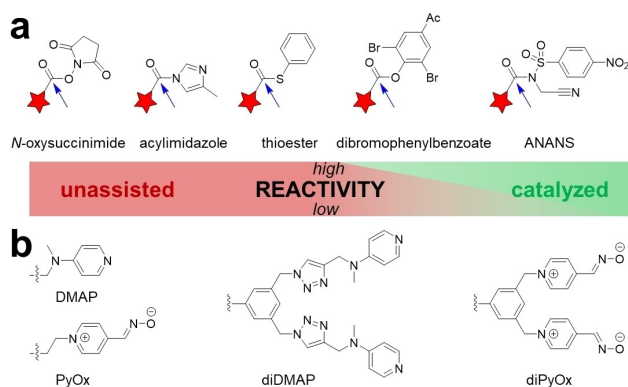
The POI modifying unit

Modification of the POI can be achieved using an activated functional group embedded in the modifying structure, or a catalyst that locally activates an otherwise unreactive soluble substrate (see also Figure 1d). In the first approach, the POI is exposed to a reactive moiety in the construct that is designed to react with a particular amino acid functionality. If a targeted reactive moiety is not within reach, competitive hydrolysis might take place leading to (i) depletion of the reactive construct (see formation of S* in Figure 1c), or (ii) modification of the protein at an undesired site due to the dynamics of the ligand-POI interactions *(vide supra)*. Whereas efficient modification was achieved by the incorporation of acyl moieties that contain excellent leaving group such as *N*-hydroxysuccinimide, acylimidazole and thioester, they are also associated with poor regioselective control and high background modification.[18] Therefore, less reactive constructs such as dibromophenylbenzoate and ANANS proved superior due to cleaner and more controllable modification (Figure 4a).[16] In fact, a study by Hamachi revealed that more stable constructs lead to POI modification that kinetically competes with rapid click reactions.
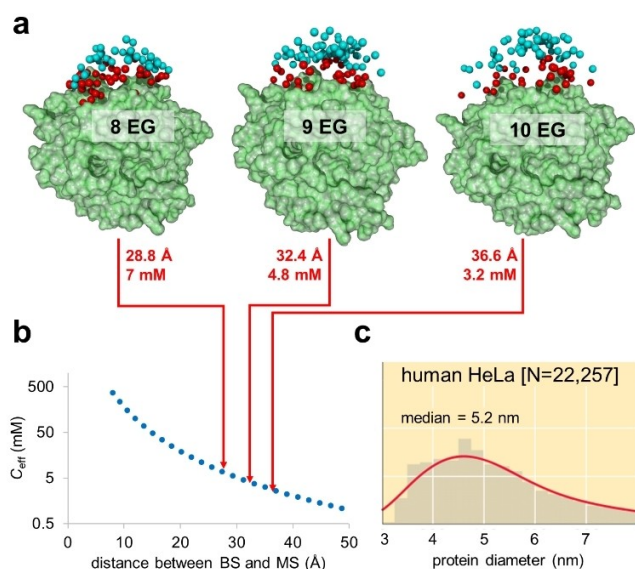
Figure 3. Influence on tether length on the position of the tips with respect to the protein thrombin. (a) Simulation showing the effect of linker length on the position of the ends of the tether where the modifying unit would be positioned. The red balls are within 10 Å from the protein surface, linkers with 8, 9, or 10 ethylene glycol units are shown from left to right, respectively. (b) Dependence of the effective concentration (C_{eff}, mM) on the length of the linker.[26] The positions of the linkers shown above are indicated by the red arrows, and the corresponding numbers on the linker length and effective concentration are given. (c) Distribution of the diameters of the proteins that are part of the human proteome.[27,28]

ChemBioChem

Concept
doi.org/10.1002/cbic.202300187

**Chemistry
Europe**
European Chemical
Societies Publishing

Figure 4. Chemical structures of nucleophilic acyl substitution derivatives that have been explored in ligand-meditated protein modification. The blue arrow indicates where the nucleophilic residue of the POI attacks the construct. Whereas the excellent leaving groups on the left are too reactive to provide the desired control over modification, the average or poor leaving groups to the right benefit from local activation by a catalyst (see also Figure 1d). (b) Acylation catalyst that have aided the local activation of the thioester (for DMAP) and ANANS (for PyOx) substrates from panel a by forming transiently activated acylating agents (as exemplified in Figure 1d). The wiggly line indicate where the catalyst is bound to the POI anchor.

To suppress unwanted modification even further, a catalyst that activates a non-reactive substrate can be implemented (see Figure 1d).[29—34] In this approach, however, distinction should be made between catalysts that generate a covalently bound reactive species and those that generate soluble reactive species. Those that locally generate a soluble reactive species display a wider modification range while also being associated with inferior control[23,35] than the ones that generate a catalyst-bound reactive unit.[23,35—37] If a catalyst that operates via a covalently-bound intermediate is not applicable, a catalyst can be used that activates a POI-bound residue (e.g., Tyr), which then reacts with an otherwise unreactive soluble component.[38] A higher level of control over the site of modification is obtained when a catalyst is incorporated that generates a covalently-bound reactive intermediate from an otherwise unreactive substrate. In such cases, unassisted modification is almost completely suppressed. In addition, constructs for which the modification rate for the bound state exceeds that of the unbound state ($k_{cat,free} < k_{cat,bound}$) are beneficial for clean modification. At the moment, various catalysts have emerged that strike this desired balance between low background modification and high local modification.[18—20] Furthermore, multivalent catalysts can be used to improve labeling efficiency by increasing the effective concentration (C_{eff}) of the modifying unit (Figure 4b).[22] For our aptamer-based modification approach, an increase in conversion from 28% to >90% was observed when PyOx was replaced with diPyOx.[20] For DMAP, a clear positive effect of implementing a divalent catalyst was not discernable.

Integration and Switching

During the exposition of the three different elements that are commonly incorporated in the latest chemical constructs that have been designed for protein modification (Table 1), it became clear that each aspect is not free of complications. Therefore, the incorporation of multiple elements in one construct has been challenging. Despite these challenges, chemical methods advanced to a level that enable the synthesis of densely-functionalized small molecules (Figure 5a) as well as late-stage functionalization of commercially available biomolecules with clickable catalysts that can be used for POI modification (Figure 5b). As such, it becomes feasible to incorporate additional elements such as externally addressable triggers in order to regulate the protein modification chemistry on demand. As this switchable element has to navigate the entire construct around the pitfalls mentioned for the different components that have been integrated, robust synthesis and activity screening and analytical techniques have to be used in an integrated approach so that successful designs can be identified and further refined.

Summary and Outlook

In this concept paper, approaches that integrate elements used by protein-modifying enzymes into artificial constructs have been evaluated. It appears that such a combination of elements is a promising avenue leading to controlled modification of proteins, even in their native environment. Despite the fact that
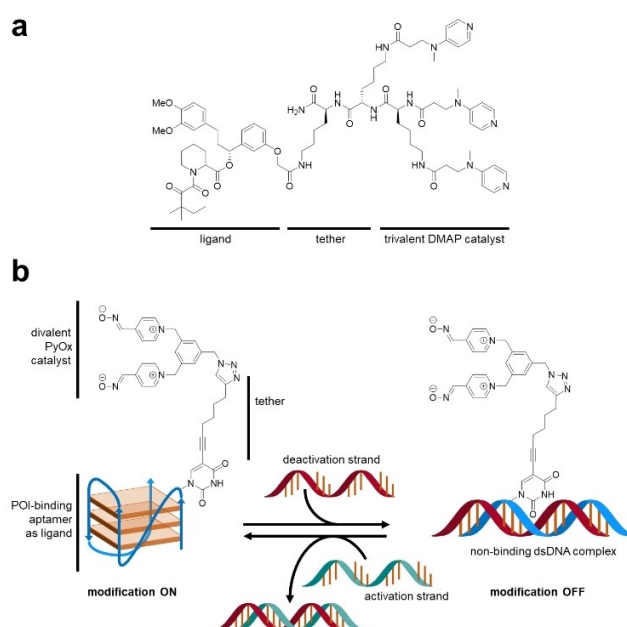
Figure 5. Two examples of chemical constructs that have been designed for protein modification. (a) A small organic molecules that combines a ligand, tether, and trivalent acylation catalyst.[39] (b) A late-stage catalyst functionalized protein-binding DNA construct that contains a switchable element that enables external regulation of the modification activity.

ChemBioChem

Concept
doi.org/10.1002/cbic.202300187

Chemistry Europe
European Chemical
Societies Publishing

the performance of artificial constructs for POI modification are affected by a multidimensional set of mostly interconnected parameters, patterns emerge that enable incorporation of elements that allow higher levels of control. Examples of constructs that contain a POI anchor, POI modifying catalyst (including appropriate substrate), programmable scaffold, and activity switch have emerged.[20] It remains challenging to identify the optimal constructs, as a large physical space can be sampled by an enormous set of chemical constructs. In addition, implementation of such constructs in more complex settings such as cell lysates, let alone in living cells brings additional challenges. Nevertheless, generating a set of tools that enables the efficient and precision modification of target proteins in their native environment in a trigger-dependent manner is appealing for many applications in life sciences, and therefore is a worthwhile but challenging task to pursue.

Acknowledgements

Conflict of Interests

There is not conflict of interest.

[1] H. Yu, J. Feng, F. Zhong, Y. Wu, *Macromol. Rapid Commun.* **2022**, *43*, 2200195.
[2] K. Shiraiwa, R. Cheng, H. Nonaha, T. Tamura, I. Hamachi, *Cell Chem. Biol.* **2020**, *27*, 970–985.
[3] R. J. Ernst, T. P. Krogager, E. S. Maywood, R. Zanchi, V. Beranek, T. S. Elliott, N. P. Barry, M. H. Hastings, J. W. Chin, *Nat. Chem. Biol.* **2016**, *12*, 776–778.
[4] X. Chen, Y.-W. Wu, *Org. Biomol. Chem.* **2016**, *14*, 5417–5439.
[5] K. Lang, J. W. Chin, *Chem. Rev.* **2014**, *114*, 4764–4806.
[6] C. D. Spicer, B. G. Davis, *Nat. Commun.* **2014**, *5*, 4740.
[7] Y. Takaoka, A. Ojida, I. Hamachi, *Angew. Chem. Int. Ed.* **2013**, *52*, 4088–4106.
[8] J. A. Prescher, C. R. Bertozzi, *Nat. Chem. Biol.* **2005**, *1*, 13–21.
[9] K.-L. Wu, C. Yu, C. Lee, C. Zuo, Z. T. Ball, H. Xiao, *Bioconjugate Chem.* **2021**, *32*, 1947–1959.
[10] B. Albada, J. F. Keijzer, H. Zuilhof, F. L. van Delft, *Chem. Rev.* **2021**, *121*, 7032–7058.
[11] N. L. Kjaersgaard, T. B. Nielsen, K. V. Gothelft, *ChemBioChem* **2022**, *23*, e202200245.
[12] K. J. Jensen, M. B. Thygesen, K. K. Sorensen, S. Wu, T. Treiberg, S. Schoffelen, *ChemBioChem* **2022**, *23*, e202200359.
[13] J. J. Bruins, D. Blanco-Ania, V. van der Doef, F. L. van Delft, B. Albada, *Chem. Commun.* **2018**, *54*, 7338–7341.
[14] J. J. Bruins, C. van de Wouw, K. Wagner, L. Bartels, B. Albada, F. L. van Delft, *ACS Omega* **2019**, *4*, 11801–11807.
[15] J. J. Bruins, A. Westphal, B. Albada, K. Wagner, L. Bartels, H. Spits, W. van Berkel, F. L. van Delft, *Bioconjugate Chem.* **2017**, *28*, 1189–1193.
[16] J. J. Bruins, J. A. M. Damen, M. Wijdeven, I. Hurkmans, F. L. van Delft, *Bioconjugate Chem.* **2021**, *32*, 2167–2172.
[17] O. Boutureira, G. J. L. Bernardes, *Chem. Rev.* **2015**, *115*, 2174–2195.
[18] S. R. Adusumalli, D. G. Rawale, K. Thakur, L. Purushottam, N. C. Reddy, N. Kalra, S. Shukla, V. Rai, *Angew. Chem. Int. Ed.* **2020**, *59*, 10332–10336.
[19] D. G. Rawale, K. Thakur, P. Sreekumar, T. K. Sajeev, A. Ramesh, S. R. Adusumalli, R. M. Mishra, V. Rai, *Chem. Sci.* **2021**, *12*, 6732–6736.
[20] J. F. Keijzer, J. Firet, B. Albada, *Chem. Commun.* **2021**, *57*, 12960–12963.
[21] T. Tamura, T. Ueda, T. Goto, T. Tsukidate, Y. Shapira, Y. Nishikawa, A. Fujisawa, I. Hamachi, *Nat. Commun.* **2018**, *9*, 1870.
[22] T. Tamura, Z. Song, K. Amaike, S. Lee, S. Yin, S. Kiyonaka, I. Hamachi, *J. Am. Chem. Soc.* **2017**, *139*, 14181–14191.
[23] J. F. Keijzer, J. Firet, B. Albada, *Chem. Eur. J.* **2022**, *28*, e202200895.
[24] T. A. Shewmake, F. J. Solis, R. J. Gillies, M. R. Caplan, *Biomacromolecules* **2008**, *9*, 3057–3064.
[25] T. Tamura, S. Tsukiji, I. Hamachi, *J. Am. Chem. Soc.* **2012**, *134*, 2216–2226.
[26] M. Kjaergaard, J. Glavina, L. B. Chemes, *Meth. Enzymol.* **2021**, *647*, 145–171.
[27] T. Geiger, A. Wehner, C. Schaab, J. Cox, M. Mann, *Mol. Cell. Proteomics* **2012**, *11*, M111.014050.
[28] R. Milo, R. Phillips, *Cell Biology by the Numbers*, Taylor & Francis Ltd, New York, USA **2015**.
[29] M. Yang, J. Li, P. R. Chen, *Chem. Soc. Rev.* **2014**, *43*, 6511–6526.
[30] T. Mino, S. Sakamoto, I. Hamachi, *Biosci. Biotechnol. Biochem.* **2021**, *85*, 53–60.
[31] Z. T. Ball, *Acc. Chem. Res.* **2013**, *46*, 560–570.
[32] Z. T. Ball, *Curr. Opin. Chem. Biol.* **2015**, *25*, 98–102.
[33] S. Sato, H. Nakamura, *Angew. Chem. Int. Ed.* **2013**, *52*, 8681–8684.
[34] M. Kunishima, S. Nakanishi, J. Nishida, H. Tanaka, D. Morisaki, K. Hioki, H. Nomoto, *Chem. Commun.* **2009**, *37*, 5597–5599.
[35] J. F. Keijzer, B. Albada, *Bioconjugate Chem.* **2020**, *31*, 2283–2287.
[36] P. G. Isenegger, B. G. Davis, *J. Am. Chem. Soc.* **2019**, *141*, 8005–8013.
[37] J. Ohata, Z. T. Ball, *J. Am. Chem. Soc.* **2017**, *139*, 12617–12622.
[38] H. A. Beard, J. R. Hauser, M. Walko, R. M. George, A. J. Wilson, R. S. Bon, *Nature Commun. Chem.* **2019**, *2*, 133.
[39] H. Wang, Y. Koshi, D. Minato, H. Nonaka, S. Kiyonaka, Y. Mori, S. Tsukiji, I. Hamachi, *J. Am. Chem. Soc.* **2011**, *133*, 12220–12228.