ChemPubSoc Europe

DOI: 10.1002/open.201700180

Open Access ChemistryOPEN Communications

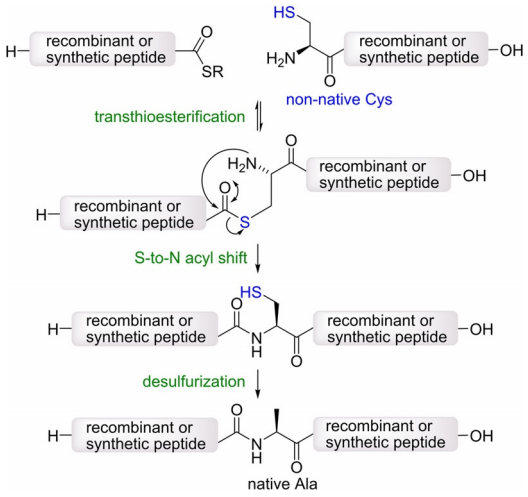
Utility of the Phenacyl Protecting Group in Traceless Protein Semisynthesis through Ligation–Desulfurization Chemistry

Maria Matveenko, Stefanie Hackl, and Christian F. W. Becker*[a]

Semisynthesis of proteins via expressed protein ligation is a widely applicable method, even more so because of the possibility of ligation at non-cysteine sites using β-mercapto amino acids that can be converted to the corresponding native amino acids by desulfurization. A drawback of this ligation–desulfurization approach is the removal of any unprotected native cysteine residues within the ligated protein segments. Here, we show that the phenacyl (PAc) moiety can be successfully used to protect cysteines within recombinantly generated protein segments. As such, this group was selectively appended onto cysteine side chains within bacterially expressed polypeptides following intein cleavage, which reveals a rather sensitive thioester at the C-terminus. The PAc group proved to be compatible with native chemical ligation, radical desulfurization, and reverse-phase HPLC conditions, and was smoothly removed at the end. The utility of the PAc protecting group was then demonstrated by the 'traceless' semisynthesis of two proteins containing one or two native cysteines: human small heat shock protein Hsp27 and murine prion protein.

Expressed protein ligation (EPL) is a versatile method for protein semisynthesis that has tremendously expanded the range of achievable targets and enabled a multitude of discoveries regarding protein structure, function, localization, and other properties.[1] This method relies on the native chemical ligation (NCL)[2] reaction between two protein segments, either of which can be produced recombinantly or synthetically, one containing a C-terminal thioester and the other containing an N-terminal cysteine (Cys), and which react by means of transesterification and rearrange through an S-to-N acyl shift to form a native peptide bond.

The initial requirement for N-terminal cysteine residues, which occur at a low frequency of approximately 1.7% in proteins,[3] has subsequently been abrogated, largely by the development of the ligation–desulfurization strategy, which combines the use of introduced cysteines or non-natural beta-mercapto derivatives of amino acids with desulfurization to clip off the thiol moiety and reveal the native alanine or other proteinogenic amino acids at the ligation site.[4] This strategy, pioneered by Yan and Dawson, who demonstrated the reductive desulfurization of Cys to Ala following NCL (Scheme 1),[4a] was made widely applicable by the development of mild, metal-free radical desulfurization methods,[4b,5] as well as the use of other β- or γ-mercapto amino acids[4f,g,5] and selenocysteine.[6] The ligation–desulfurization chemistry has greatly increased the flexibility of the semi- and total synthesis approaches toward proteins. The potential drawback of most of these methods, besides when using selective desulfurization of β-mercapto aspartate[7] or selenocysteine,[8] is the concomitant desulfurization of any unprotected native cysteines, usually found in the larger recombinant segments of the protein of interest, as those in the synthetic part can be easily protected orthogonally during solid-phase peptide synthesis (SPPS). Both the removal of any native cysteines as well as the incorporation of non-native thiol-containing amino acids is often undesirable, as these residues are capable of forming disulfide bridges that greatly affect the tertiary structure and moreover can themselves be functionally significant.



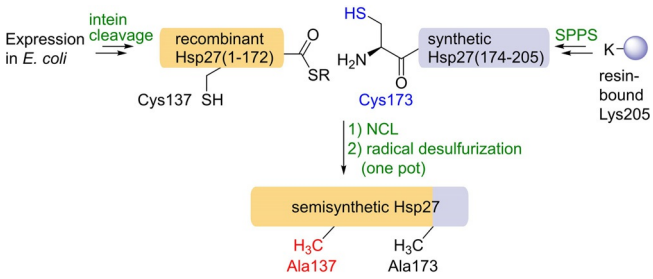
Scheme 1. Overview of the ligation–desulfurization approach.

During our studies on the semisynthesis of human small heat shock protein, Hsp27, through the ligation–desulfurization

[a] *Dr. M. Matveenko, S. Hackl, Prof. Dr. C. F. W. Becker*
Institute of Biological Chemistry, Faculty of Chemistry
University of Vienna
Währinger Str. 38, 1090 Vienna (Austria)
E-mail: christian.becker@univie.ac.at

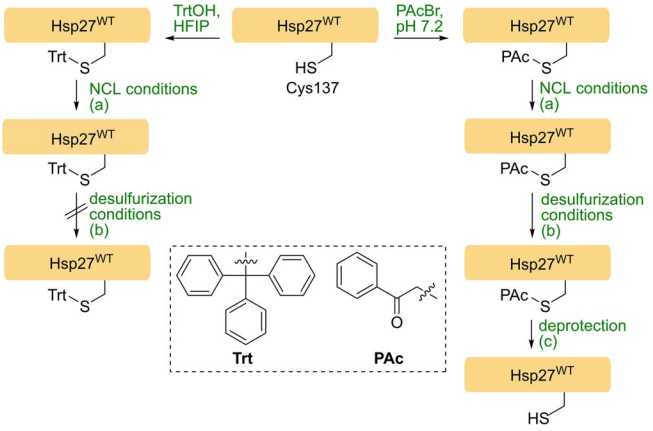
Supporting Information and the ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/open.201700180.

strategy, the unwanted thiol removal of the single native cysteine residue (Cys137) was encountered (Scheme 2). The 205 amino acid protein was generated through NCL between the recombinant Hsp27(1–172) segment carrying a C-terminal thioester and the smaller synthetic fragment Hsp27(173–205), wherein Ala173 had been replaced with a Cys residue, followed by desulfurization. Fortunately, the ensuing Cys137Ala modification had previously been investigated and rendered inconsequential in the context of our functional characterization of the protein.[9] Nevertheless, we decided to solve the undesired desulfurization issue and, thus, to achieve a 'traceless' semisynthesis of Hsp27, keeping Cys137 intact. Two variants of this protein were pursued in accordance with our previous work: one carrying a non-enzymatic posttranslational modification argpyrimidine (Apy) at residue 188, termed Hsp27^{Apy}, and the control, non-modified (NM) variant, referred to as Hsp27^{NM}.



Scheme 2. Overview of the previous ligation–desulfurization approach applied to the semisynthesis of Hsp27, generating a Cys137Ala mutation.[9]

To this end, we searched for a suitable protecting group that would temporarily mask the native Cys137 of Hsp27 and satisfy several criteria. Firstly, this protecting group has to be selective for cysteine residues, as it needs to be appended following thiol-mediated intein cleavage of the long, recombinantly produced protein segment containing many nucleophilic side chains. In addition, the conditions for this protection need to be mild, so as not to compromise the unstable C-terminal thioester moiety. Protection prior to intein cleavage and concomitant thioester formation is not possible, owing to the requirement for the free thiol group of the critical cysteine residue in intein to be next to the cleavage site. Secondly, the resulting protected thiol has to be stable enough to withstand the NCL reaction followed by desulfurization of the ligation-site cysteine, thus excluding any groups that are sensitive to reducing, basic, and radical conditions. Finally, it should be straightforward to remove such a protecting group by using reasonably mild reagents. Based on these criteria, two candidates were selected, trityl (Trt)[10] and phenacyl (PAc)[11] groups, both of which were attractive as they had previously been successfully used for similar purposes, although neither has been used to protect a thioester-containing peptide.[10,11c] Specifically, Trt was utilized by Mochizuki et al. to protect two internal Cys sulfhydryls within a 32-amino-acid-long synthetic peptide, which was then deprotected at the N-terminal cysteine and ligated to a C-terminal thioester fragment, followed by desulfurization, Trt-deprotection, oxidation, and purification, to obtain the desired product in 42% overall yield from the Trt-free peptide.[10] Kawakami and co-workers, on the other hand, used the PAc moiety, first applied in the protection of amino and thiol functions by Tang et al.,[11a] to mask two cysteines within a recombinant C-terminal fragment of histone H3.1. After successful introduction of the PAc groups, the orthogonally protected N-terminal cysteine was revealed and the resulting fragment underwent NCL, desulfurization, and PAc-deprotection to afford semisynthetic H3.1 in 18% yield over six steps from the protecting-group-free recombinant fragment.[11c] With these two promising candidates in mind, we decided to use wild-type (WT) recombinant, HPLC-purified Hsp27 (Hsp27^{WT}) as a model system for the comparison of the protecting groups (Scheme 3).

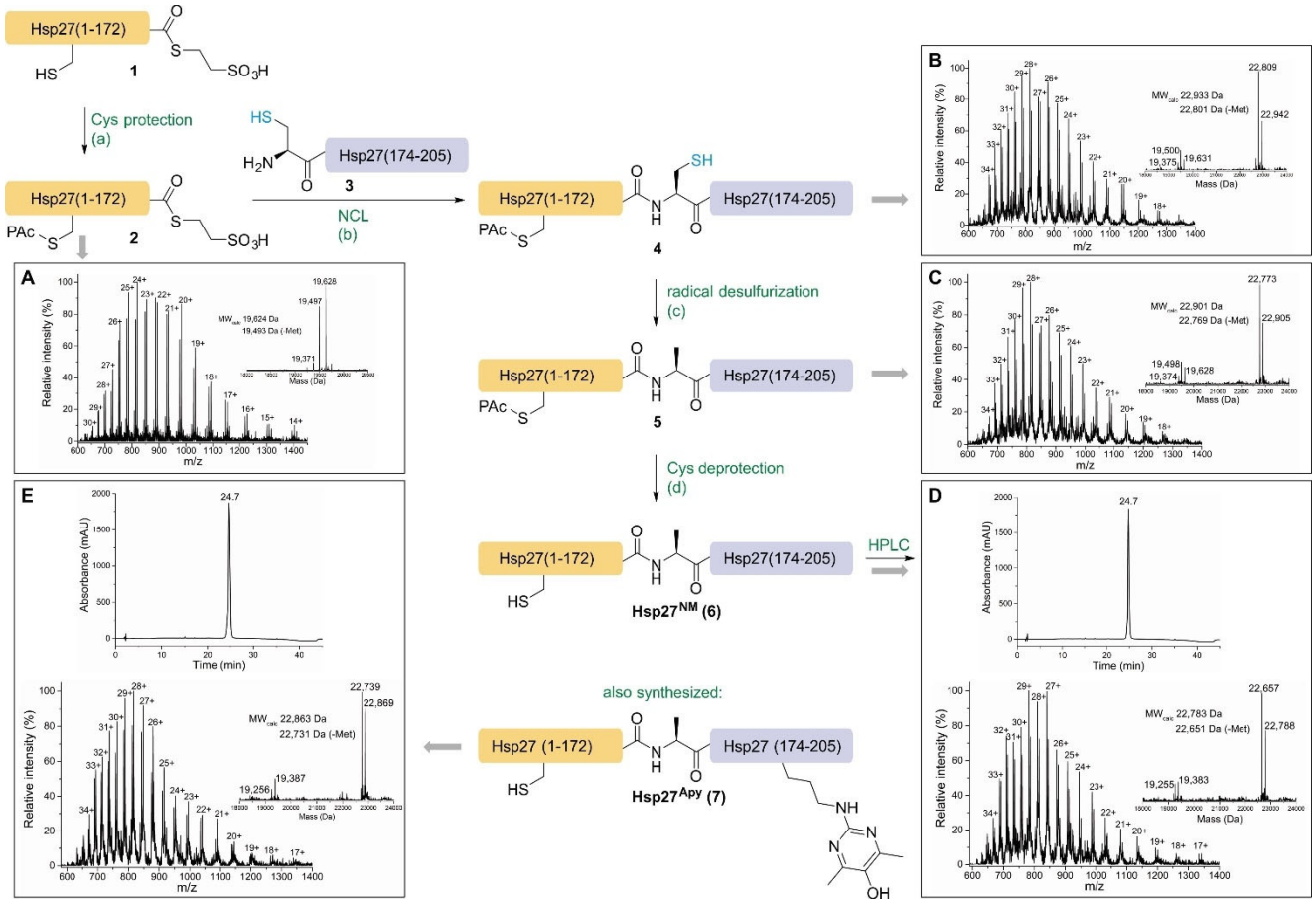


Scheme 3. Overview of the process to find a suitable cysteine protecting group to enable the 'traceless' semisynthesis of Hsp27. Reagents and conditions: a) 6 M Gdn·HCl, 0.2 M NaPi, 0.25 M MESNA, 50 mM TCEP, pH 7.8, 37 °C, 48 h; b) 6 M Gdn·HCl, 0.2 M NaPi, 0.2 M TCEP, 50 mM MESNA, 16 mM V-50, pH 6.7, 40 °C, 16 h; c) Zn, MPA (15% v/v in 6 M Gdn·HCl).

Gratifyingly, both Trt and PAc protecting groups could be introduced easily onto Cys137 within the model system Hsp27^{WT} by alkylation with either Trt cation, generated from TrtOH in the presence of a moderately acidic hexafluoroisopropanol (HFIP), or PAcBr used in slight excess at a carefully controlled pH (Scheme 3) under the published reaction conditions.[10,11c] The ESI-MS spectra of the ensuing proteins (Figures S1 and S2) indicated a clean conversion, and thus the respective buffers were exchanged, without isolation of proteins, with that required for an NCL reaction. In previous work, we were able to conduct the ligation and the desulfurization steps in one pot by using sodium 2-mercaptoethanesulfonate (MESNA) as the mediator for NCL, as this reagent is compatible with the subsequent radical-mediated step.[9,12] Thus, the S-protected proteins were submitted to these optimized conditions, consisting of 6 M Gdn·HCl, 50 mM tris(2-carboxyethyl)phosphine (TCEP) and 250 mM MESNA at pH 7.8. Both protecting groups performed well under these conditions, and the corresponding protein samples remained unaffected even after 48 h (the typical reaction time from previous studies), as judged by their LC-MS spectra (Figures S3 and S4). The more critical test was the subsequent radical desulfurization process, for which the reaction

mixtures were diluted (4×) with a buffer containing additional TCEP to serve as the thiyl radical acceptor, and the water-soluble radical initiator V-50,[4b] whereas the remaining MESNA (ca. 50 mM final concentration) should function as the hydride source. After 16 h at 40 °C,[9] the LC-MS spectrum of the Trt-protected protein (Figure S5) showed a mixture of products, including deprotected but not desulfurized proteins, indicating that the easily formed and stable trityl radical had potentially compromised the stability of the protecting group and the reaction itself. On the other hand, the spectrum of the PAc-protected Hsp27^{WT} did not change, showing only the protected proteins (Figure S6) and indicating that only the PAc group was suitable for our purposes. In the last hurdle, this moiety was then smoothly cleaved by treatment with powdered Zn under acidic conditions [mercaptopropionic acid (MPA) in 6 M Gdn·HCl, pH ≈ 1, unadjusted].[11c] After 30 min, the peak for the substrate was already undetectable in the LC-MS spectrum of the reaction mixture, and the deprotected protein was isolated by HPLC (recovery > 60% over 4 steps; see Figure S7 for characterization). Notably, Katayama and Hojo have demonstrated that such reducing conditions do not affect acetamidomethyl (Acm) or *p*-methoxybenzyl groups,[11b] which are commonly used for Cys protection should such orthogonality be of use.

In this way, the PAc group was selected and was then used to synthesize the non-Apy-modified (Hsp27^{NM}) and Apy-modified (Hsp27^{Apy}) proteins that retain their native Cys137 residues (Scheme 4). PAcBr could then be smoothly reacted with Hsp27(1–172)-thioester **1** to produce the corresponding Cys137-protected variant **2**; the formation of which was verified by LC-MS analysis of the reaction mixture (mass difference owing to PAc addition = +118 Da; Figure S8). The only inhomogeneity in the ensuing mass spectrum (Scheme 4A) arises from the presence of traces of the corresponding C-terminal acid of **2** (mass deviation from desired product = −124 Da), formed by hydrolysis during the preceding MESNA-mediated intein cleavage. On the other hand, the doubling of the main peak is the result of the inconsequential incomplete cleavage of the N-terminal methionine by bacterial peptidases, which often occurs during expression in *E. coli*.[13] Subsequently, buffer exchange to the NCL conditions followed by addition of the N-terminal Cys-carrying peptide **3** in a slight excess generated, after 48 h, ligation product **4**, as indicated by LC-MS analysis of the crude reaction mixture and the ensuing mass spectrum (Scheme 4B), which also shows traces of unreacted **2** as well as the corresponding acid. Buffer exchange and submitting the crude **4** to radical desulfurization resulted in clean

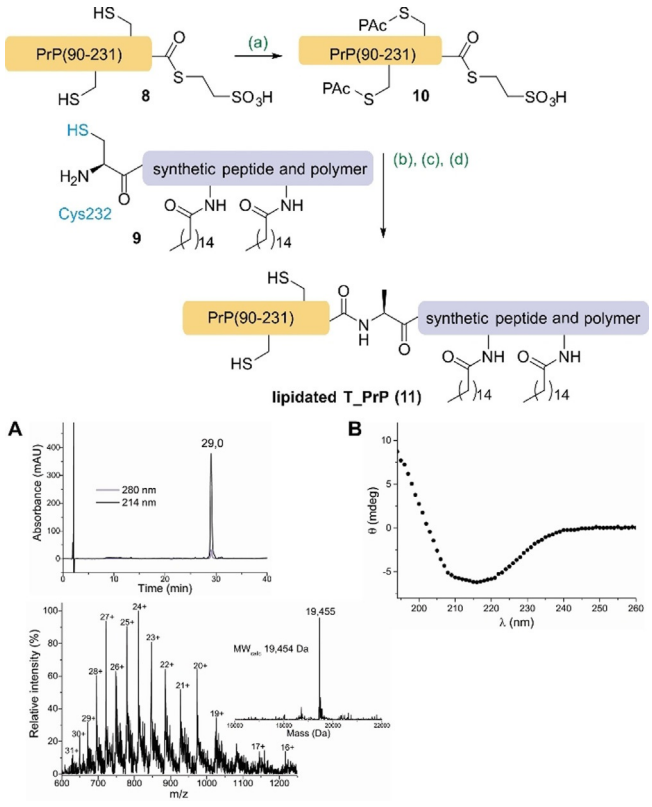


Scheme 4. Semisynthesis of Hsp27 variants through traceless one-pot ligation–desulfurization by using the PAc protecting group. Reagents and conditions: a) [**1**] = 0.5 mM, PAcBr (2.5 equiv), 6 M Gdn·HCl, 0.4 M NaPi, pH 7.15, 25 °C, 1 h; b) [**2**] = 0.3 mM, **3** (3 equiv), 6 M Gdn·HCl, 0.2 M NaPi, 0.25 M MESNA, 50 mM TCEP, pH 7.8, 37 °C, 48 h; c) 6 M Gdn·HCl, 0.2 M NaPi, 0.15 M TCEP, 0.125 M MESNA, 13 mM V-50, pH 6.6, 40 °C, 16 h; d) powdered Zn (37 mg mL^{-1}), MPA (7.5% v/v in 6 M Gdn·HCl), 25 °C, 1 h, 35% over four steps. A–C) Mass spectra obtained by LC-MS analysis of crude reaction mixtures of **2**, **4**, or **5**. D,E) Analytical HPLC traces (λ = 214 nm) and ESI-MS spectra of isolated **6**, **7**.

conversion into compound **5** (Scheme 4C), which was then deprotected as described above and purified by the single application of reverse-phase HPLC to provide Hsp27^{NM} still containing Cys137 (**6**, 35% from thioester **1**, 2.0 mg). The Apy-modified variant **7** was generated by identical means, but using the Apy-derivatized variant of **3** in 38% yield (2.2 mg) over four steps. The proteins were obtained in good purity, as judged by analytical HPLC chromatograms and ESI-MS spectra (Schemes 4D and 4E), although both samples contained small amounts of the unreacted acid Hsp27(1–172)–OH, which also retained Cys137. The two samples were analyzed by far-UV circular dichroism (CD) spectroscopy and analytical size-exclusion chromatography (Figures S11 and S12, respectively). We did not observe any major differences between the properties of the two variants or, moreover, to those obtained previously and containing the Cys127Ala modification (Figures S11 and S12).[9]

To probe the scope of the PAc-protection approach, we chose another recombinant protein segment that had been utilized in the semisynthesis of a lipidated prion protein (PrP) variant in our laboratories. The C-terminal thioester of N-terminally truncated PrP (residues 90–231, T PrP-SR, **8**; Scheme 5) contains two cysteines, Cys178 and Cys213, and was previously ligated to a synthetic C-terminal peptide **9** that features a non-native Cys232 residue at its N-terminus for NCL, with two lysine-bound palmitoyl groups to serve as a membrane anchor and a polymer tag to enhance solubility.[14] The introduced ligation site Cys was left un-desulfurized in the previous syntheses, which could potentially be problematic for the folding of the protein, as it involves the Cys178–Cys213 disulfide formation. In an attempt to achieve a 'traceless' semisynthesis of lipidated T PrP, we obtained the bis-Cys-protected product **10** (Scheme 5 and Figure S13). Ligation of this thioester to peptide **9** (2.5 equiv) did not proceed with MESNA as a mediator, and the more reactive thioester-forming mercaptophenylacetic acid (MPAA) had to be utilized, which resulted in rapid and efficient ligation (Figure S14). Unfortunately, this aromatic thiol was not compatible with the radical desulfurization reaction, owing to its radical-quenching properties and, moreover, it could not be removed sufficiently by using precipitation or buffer-exchange methods. Thus, the ligation product was isolated by reverse-phase HPLC (Figure S15), and then submitted to the optimized desulfurization conditions. The readily formed Cys232Ala product (Figure S16) was then treated with the PAc-deprotection mixture used previously. Unfortunately, the reaction was very sluggish in this case, producing insoluble material and unable to reach completion. Although even worse results were observed with Mg as the metal, using Zn in AcOH (45% v/v in 6 M Gdn·HCl)[11b] allowed the desired product to be formed with full conversion in 60 min, as indicated by LC-MS. After centrifugation and isolation by reverse-phase HPLC, the desired lipidated T PrP variant **11** was isolated in 65–90% yield over two steps and characterized by ESI-MS and analytical reverse-phase HPLC (Scheme 5A). The semisynthetic PrP samples were folded, as described previously,[14a] and analyzed by CD spectroscopy (Scheme 5B), which showed that they adopt a secondary structure similar to that observed for the samples containing the additional Cys232 residue (Figure S17 and Table S1), corroborating the previous results.

Overall, the PAc moiety proved to be a useful cysteine protecting group that can be appended onto recombinantly produced protein segments containing a sensitive thioester moiety. The group remains unaffected during the NCL and radical desulfurization reactions as well as reverse-phase HPLC, and can be cleanly removed under reductive, acidic conditions that are orthogonal to the deprotection of common cysteine protecting groups. Therefore, this methodology may help to achieve even more precise semi- and total synthesis approaches for proteins in the future.



Scheme 5. Traceless semisynthesis and characterization of lipidated T PrP variant. Reagents and conditions: a) [**8**] = 0.4 mM, PAcBr (5 equiv), 6 M Gdn·HCl, 0.4 M NaPi, pH 7.15, 20 °C, 1 h; b) [**10**] = 0.6 mM, **9** (2.5 equiv), 6 M Gdn·HCl, 0.2 M NaPi, 30 mM MPAA, 20 mM TCEP, pH 7.2, 37 °C, 6 h, ca. 30% over two steps; c) 6 M Gdn·HCl, 0.2 M NaPi, 0.25 M TCEP, 0.1 M MESNA, 20 mM V-50, pH 6.7, 40 °C, 16 h; d) Zn (180 mg mL^{-1}), AcOH (45% v/v in 6 M Gdn·HCl), 25 °C, 1 h, 65–90% over 2 steps. A) Analytical HPLC trace (baseline corrected) and ESI-MS spectrum of isolated **11**. B) Far-UV CD spectrum of folded **11**.

Experimental Section

Procedures and characterization data can be found in the Supporting Information.

Acknowledgements

We gratefully acknowledge Gilyana Kücükkaplan and Gerhard Niederacher for excellent technical assistance and the Austrian

Academy of Sciences for financial support (APART fellowship to M.M.).

Conflict of Interest

The authors declare no conflict of interest.

Keywords: cysteine · ligation–desulfurization · protecting groups · protein modifications · protein semisynthesis

[1] a) T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705 – 6710; b) T. W. Muir, *Annu. Rev. Biochem.* **2003**, *72*, 249 – 289; c) T. Durek, C. F. W. Becker, *Biomol. Eng.* **2005**, *22*, 153 – 172; d) T. W. Muir, *Biopolymers* **2008**, *90*, 743 – 750; e) P. Siman, A. Brik, *Org. Biomol. Chem.* **2012**, *10*, 5684 – 5697; f) M. Holt, T. Muir, *Annu. Rev. Biochem.* **2015**, *84*, 265 – 290; g) H. M. Burke, L. McSweeney, E. M. Scanlan, *Nat. Commun.* **2017**, *8*, 15655.

[2] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, *266*, 776 – 779.

[3] B. L. Nilsson, M. B. Soellner, R. T. Raines, *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 91 – 118.

[4] a) L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, *123*, 526 – 533; b) Q. Wan, S. J. Danishefsky, *Angew. Chem. Int. Ed.* **2007**, *46*, 9248 – 9252; *Angew. Chem.* **2007**, *119*, 9408 – 9412; c) A. Dirksen, P. E. Dawson, *Curr. Opin. Chem. Biol.* **2008**, *12*, 760 – 766; d) H. Rohde, O. Seitz, *Biopolymers* **2010**, *94*, 551 – 559; e) P. E. Dawson, *Isr. J. Chem.* **2011**, *51*, 862 – 867; f) L. R. Malins, R. J. Payne, *Curr. Opin. Chem. Biol.* **2014**, *22*, 70 – 78; g) L. R. Malins, R. J. Payne, *Aust. J. Chem.* **2015**, *68*, 521; h) L. R. Malins, R. J. Payne in *Protein Ligation and Total Synthesis I* (Ed.: L. Liu), Springer International Publishing, Cham, **2015**, pp. 27 – 87.

[5] C. Haase, H. Rohde, O. Seitz, *Angew. Chem. Int. Ed.* **2008**, *47*, 6807 – 6810; *Angew. Chem.* **2008**, *120*, 6912 – 6915.

[6] a) N. Metanis, E. Keinan, P. E. Dawson, *Angew. Chem. Int. Ed.* **2010**, *49*, 7049 – 7053; *Angew. Chem.* **2010**, *122*, 7203 – 7207; b) L. R. Malins, N. J. Mitchell, R. J. Payne, *J. Pept. Sci.* **2014**, *20*, 64 – 77.

[7] a) R. E. Thompson, B. Chan, L. Radom, K. A. Jolliffe, R. J. Payne, *Angew. Chem. Int. Ed.* **2013**, *52*, 9723 – 9727; *Angew. Chem.* **2013**, *125*, 9905 – 9909; b) C. Araman, R. E. Thompson, S. Wang, S. Hackl, R. J. Payne, C. F. W. Becker, *Chem. Sci.* **2017**, *8*, 6626 – 6632.

[8] a) S. Dery, P. S. Reddy, L. Dery, R. Mousa, R. N. Dardashti, N. Metanis, *Chem. Sci.* **2015**, *6*, 6207 – 6212; b) L. R. Malins, N. J. Mitchell, S. McGowan, R. J. Payne, *Angew. Chem. Int. Ed.* **2015**, *54*, 12716 – 12721; *Angew. Chem.* **2015**, *127*, 12907 – 12912.

[9] M. Matveenko, E. Cichero, P. Fossa, C. F. Becker, *Angew. Chem. Int. Ed.* **2016**, *55*, 11397 – 11402; *Angew. Chem.* **2016**, *128*, 11569 – 11574.

[10] M. Mochizuki, H. Hibino, Y. Nishiuchi, *Org. Lett.* **2014**, *16*, 5740 – 5743.

[11] a) G. Tang, T. Ji, A.-F. Hu, Y.-F. Zhao, *Synlett* **2008**, *2008*, 1907 – 1909; b) H. Katayama, H. Hojo, *Org. Biomol. Chem.* **2013**, *11*, 4405 – 4413; c) T. Kawakami, R. Yoshikawa, Y. Fujiyoshi, Y. Mishima, H. Hojo, S. Tajima, I. Suetake, *J. Biochem.* **2015**, *158*, 403 – 411.

[12] P. Siman, O. Blatt, T. Moyal, T. Danieli, M. Lebendiker, H. A. Lashuel, A. Friedler, A. Brik, *ChemBioChem* **2011**, *12*, 1097 – 1104.

[13] P. H. Hirel, M. J. Schmitter, P. Dessen, G. Fayat, S. Blanquet, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 8247 – 8251.

[14] a) D. Olschewski, R. Seidel, M. Miesbauer, A. S. Rambold, D. Oesterhelt, K. F. Winklhofer, J. Tatzelt, M. Engelhard, C. F. W. Becker, *Chem. Biol.* **2007**, *14*, 994 – 1006; b) N. K. Chu, C. F. W. Becker, *Methods Enzymol.* **2009**, *462*, 177 – 193; c) N. K. Chu, W. Shabbir, E. Bove-Fenderson, C. Araman, R. Lemmens-Gruber, D. A. Harris, C. F. W. Becker, *J. Biol. Chem.* **2014**, *289*, 30144 – 30160.

Received: November 8, 2017
Revised manuscript received: December 12, 2017

