The hitchhiker´s guide to chemoselectivity in chemical protein synthesis

Can Araman

Gorlaeus Laboratories, Einsteinweg 55, 2333 CC Leiden, the Netherlands email: m.c.araman@lic.leidenuniv.nl

Posttranslational modifications (PTMs) of proteins are nature's way of introducing diversity. Even though there are only 22 proteinogenic amino acids (including pyrrolysine and selenocysteine)[1] and thus only limited amount of variety for the ribosomal synthesis of proteins, PTMs provide proteins with distinct functions rendering them useful for different physiological tasks such as homeostasis, apoptosis and cellular signaling. Hence, it is of great interest to elucidate the role of PTMs on proteins in physiological and pathophysiological processes. However, molecular biological methods such as recombinant protein expression either does not yield in proteins with PTMs (*E. coli*) or give rise to heterogeneous mixture thereof (mammalian cells). To provide homogeneity, different methods are applied since 1990s. As discussed in an earlier Peptide Highlights contribution (April 2017), chemoselective ligation methods such as native chemical ligation (NCL)[2] asserted themselves as suitable techniques to install PTMs in a site-specific manner to proteins.

In this commentary, I will discuss exciting new chemoselective synthetic methods to introduce two distinct PTMs (lipidation and SUMOylation), into proteins site-selectively. Hereby, not only NCL, (Dardashti et al. *Chem. Eur. J.* **2020**)[3], but also Ser/Thr-Ligation (STL, Huang et al. *Angew. Chem. Int. Ed. Engl.* **2019**)[4] will be of interest.

S-Palmitoylation

Site-selective chemical lipidation of proteins is a difficult task. Various methods to introduce lipids to proteins have been developed[5], for which solubility of proteins was the major limiting factor. Among all lipidated proteins generated, native S-Palmitoylated variants could not yet be obtained through chemical protein synthesis. S-Palmitoylation is the transfer of palmitoyl-acyl onto the sulfhydryl sidechain of cysteines. This PTM has been associated with protein trafficking and intracellular protein stability.[6] Dysfunction of enzymes for palmitoylation and depalmitoylation are linked to several diseases, e.g. cancer.[6] S-Palmitoylation does not have a consensus sequence and thus the methods to detect palmitoylated proteins are challenging. Moreover, palmitoylated proteins and their interaction partners cannot be isolated from complex mixtures (serum, cell lysates etc.) in a homogeneous manner, which makes the study of inhibitory agents on depalmitoylation even more difficult. In light of these considerations, homogeneously palmitoylated proteins represent valuable tools to study drug candidates against depalmitoylation.

To this end, the research group of Zheng designed and performed the first chemical syntheses of such homogeneously S-palmitoylated membrane proteins sarcolipin (SLN) and matrix 2 ion channel from Influenza virus A (M2). The authors used a γ-aminobutyric acid (GABA)-based reversible backbone modification (RBM^{GABA}) strategy. Hereby, RBM^{GABA} consists of a solubilizing tag (6 lysines, Figure 1) and the hydroxy moiety is capped with GABA (Figure 1, highlighted in blue), which can be cleaved under mild acidic conditions through autocyclization leaving the unprotected peptide unaffected. NCL based methods are not accessible for S-palmitoylated proteins, since those have a S-palmitoylthioester (Figure 1, purple) and can be attacked by N-terminal cysteines. Therefore, non-NCL based chemoselective methods giving rise to native peptide bonds were considered. Serine-Threonine- Ligation (STL, Figure 1) is such a method describing the chemoselective linkage of an N-terminal Ser/Thr peptide and a C-terminal ester, salicyl aldehyde ester SAL (Figure 1, red), peptide with subsequent O → N acyl transfer.[7] So formed N,O-benzylidene acetal linked peptide can then undergo acidolysis to yield in the formation of a nascent peptide bond.

SLN is a single domain membrane protein, which is S-palmitoylated. It is involved in Ca^{2+}-homeostasis and has recently been identified as a drug target. The researchers aimed to homogeneously palmitoylate this protein via different approaches. First, they tried direct synthesis via Fmoc-SPPS and orthogonal Cys protecting strategy. However, palmitoylated SLN was not to analyze or dissolve in conventional peptide solvents. Thus, the strategy was changed involving RBM^{GABA} and STL (Figure 1). The peptide is divided into two segments a) SLN1-12 with an RBM between Glu2 and Arg3 as well as the palmitoyl on Cys9 and b) SLN13-31 with an RBM between Leu16 and Ile17. Both peptide fragments were ligated in 5 h quantitatively, purified over RP-HPLC and the N,O-benzylidine as well as the RBM groups were cleaved off via acidolysis, giving rise to a final yield of 55%.

The second synthetic target M2 plays a role in the virus entry and replication of influenza A. Its function is governed by numerous PTMs such as phosphorylation and S-palmitoylation. M2 comprises 97 residues and is homo-tetrameric, rendering its synthesis potentially difficult due to high hydrophobicity and spontaneous aggregation. The RBM strategy with the solubilizing lysine tag helps to overcome above problem. To synthesize homogeneously palmitoylated M2, the researchers divided the protein into two segments, M2 comprising amino acids 1-30 with a SAL ester on the C-terminus (M2 1-30) and a RBM containing (Leu36), S-palmitoylated (Cys50) peptide spanning amino acids 31-97 (M2 31-97).

The ligation product with N,O- benzylidene and RBM was obtained in a good yield and purity. There is no report on the yield of the final compound, which is expected given the high hydrophobic nature of the tetramer and handling problems stemming thereof. Nevertheless, the researchers thoroughly characterized M2 via CD, ESI-MS and SDS-PAGE showing that the protein is indeed formed. To summarize, the authors present a novel and elegant method to synthesize native and homogeneously palmitoylated proteins via STL. The use of a backbone modification and its reversible cleavage helps to solubilize hydrophobic sequences and is orthogonal to the chemoselective methods used herein. The future challenges will be the synthesis of larger lipidated proteins (> 120 residues) and introduction of other PTMs alongside lipidation.

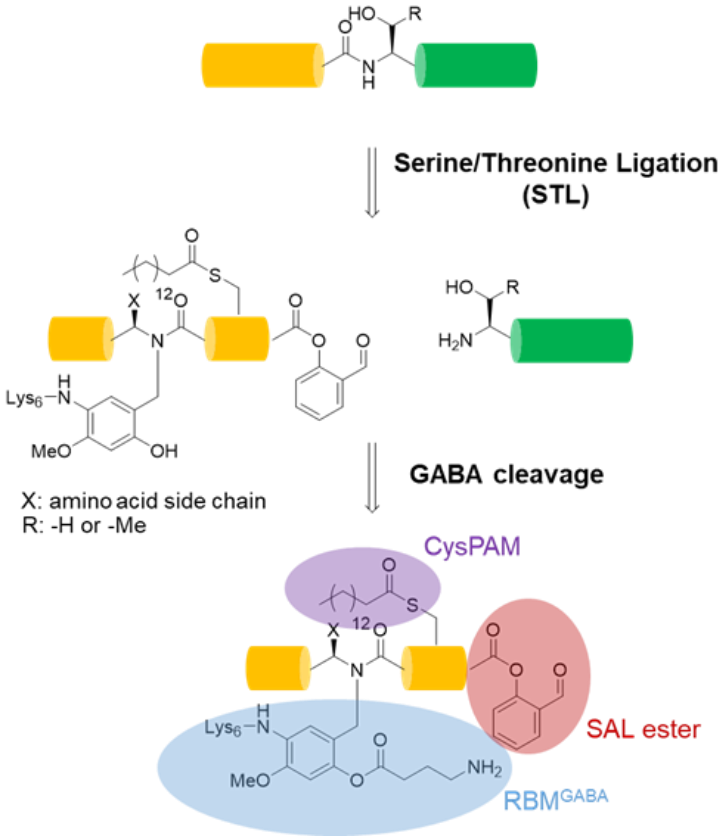

Figure 1. RBMGABA and STL strategy for total synthesis of lipidated proteins

SUMOylation and use of selenolysine as an NCL orthogonal amino acid

Small Ubiquitin like modifier, is a small protein (~ 12 kDa), performing a number of physiological functions, such as nuclear transport, transcriptional regulation, apoptosis as well as cell-cycle regulation to name a few.[8] Similar to ubiquitin (Ub), it is attached to the ε-amino side chain of lysine residues within a consensus sequence of x-K-X-D/E. The enzymatic cascade catalyzing SUMOylation is very similar to that of Ub-attachment.

To study the precise physiological contribution of SUMO, site-specifically SUMOylated proteins were generated by different groups using chemical protein synthesis and/or semisynthesis.[9] In most of mentioned publications, γ- and δ-thiolysine was used to generate the iso-peptide bond, which is disfavored in the presence of free cysteines, since those can also be reduced with the methods used to reduce γ- and δ-sulfhydryl groups. This is usually the case in NCL based methodologies and require a complex sulfhydryl protection chemistry leading to an additional HPLC purification step with a decrease in overall yield. Metanis and coworkers now describe an elegant addition to the existing toolbox to generate SUMOylated proteins via semisynthesis. Hereby, the researchers synthesized a γ-Seleno analogue of lysine (γ-Selenolysine, γ-*Se*K, Figure 2) for traceless iso-peptide bond formation. On one hand, the method is fully orthogonal to NCL, on the other hand, γ-*Se*K can be selectively deselenized under the conditions where unprotected cysteine residues are not affected rendering this new amino acid a valuable tool for protection group economical NCL approaches. In only six steps, the researchers synthesized Fmoc-SPPS suitable γ-*Se*K building block with *p*-methoxybenzyl as Se-protecting group (Se-Mob) and Boc for ε-NH_2 side chain protection. However, for the formation of an isopeptide bond, the ε-NH_2 must be protected with an orthogonal group, which should not be freed during acidic cleavage from the resin. Hence, an ε-Alloc-protected derivative of γ-*Se*K in 15 steps with a very good overall yield (~ 87%) was generated (Figure 2).

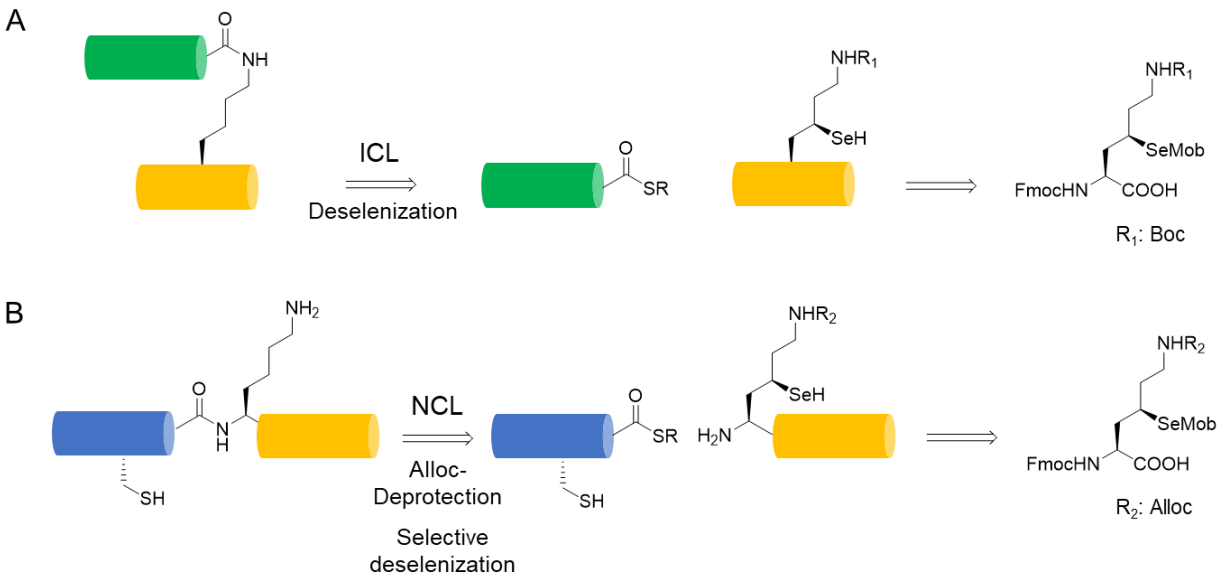

Figure 1. A) γ-Selenolysine assisted iso-peptide ligation (ICL) and B) NCL

Subsequently, a semisynthetic approach was used to link a native SUMO modification to pancreatic Glucokinase (GCK). GCK is SUMOylated at one or more lysine residues, however the impact of this PTM on the function of GCK is under investigated due to its heterogeneity. In a three-segment approach, the researchers aimed to ligate SUMO-1 to the lysine at position 15 of GCK. For this purpose, a latent C-terminal peptide thioester (α-hydrazide) spanning amino acids 1-19 of GCK was synthesized and γ-*Se*K was introduced on position 15. The isopeptide-bond ligation (ICL) between SUMO-1 thioester and γ-*Se*K of GCK1-19 proved to be difficult (long reaction time >24h and low yields ~ 22%). However, the free cysteine of SUMO-1 at position 52 was not affected by deselenization procedure after ICL, highlighting the suitability of this method for a broad spectrum. Note that efforts to obtain full length GCK with homogeneously introduced SUMO-1 were not successful, which might be caused by previously reported problems such as solubility of SUMO in common ligation buffers. Taken together Metanis group reports a very valuable tool for the synthesis of iso-peptide linked PTMs. The orthogonality to NCL and sulfhydryl groups certainly qualifies γ-*Se*K for a broader use. Considering this and other recent contributions, the future of chemoselective seleno-ligation chemistry tends to be bright and certainly more is to come.

REFERENCES

1. Y. Zhang, V. N. Gladyshev, Nuc. Acid. Res. 2007, 35, 4952-4963. PMID: 17626042
2. P. Dawson, T. Muir, I. Clark-Lewis, S. Kent, Science 1994, 266, 776-779. PMID: 7973629
3. R. Notis Dardashti, S. Kumar, S. Sternisha, S. R. Post, B. G. Miller, N. Metanis, Chem. Eur. J., 2020.
4. D.-L. Huang, C. Montigny, Y. Zheng, V. Beswick, Y. Li, X.-X. Cao, T. Barbot, C. Jaxel, J. Liang, M. Xue, C.-L. Tian, N. Jamin, J.-S. Zheng, Angew. Chem. Int. Ed. Engl., 2019.
5. T. Mejuch, H. Waldmann, Bioconjugate Chemistry 2016, 27, 1771-1783. PMID: 27444727
6. H. Jiang, X. Zhang, X. Chen, P. Aramsangtienchai, Z. Tong, H. Lin, Chem. Rev. 2018, 118, 919-988. PMID: 29292991
7. X. Li, H. Y. Lam, Y. Zhang, C. K. Chan, Org. Lett. 2010, 12, 1724-1727. PMID: 20232847
8. R. T. Hay, Molecular Cell 2005, 18, 1-12. PMID: 15808504
9. [a]S. Bondalapati, E. Eid, S. M. Mali, C. Wolberger, A. Brik, Chem. Sci. 2017, 8, 4027-4034. PMID: 28580118; [b]M. P. C. Mulder, R. Merkx, K. F. Witting, D. S. Hameed, D. El Atmioui, L. Lelieveld, F. Liebelt, J. Neefjes, I. Berlin, A. C. O. Vertegaal, H. Ovaa, Angew. Chem. Int. Ed. Engl. 2018, 57, 8958-8962. PMID: 29771001; [c]O. Melnyk, J. Vicogne, Tetrahedron Lett. 2016, 57, 4319-4324; [d]T. G. Wucherpfennig, V. R. Pattabiraman, F. R. P. Limberg, J. Ruiz-Rodríguez, J. W. Bode, Angew. Chem. Int. Ed. Engl. 2014, 53, 12248-12252. PMID: 25244549; [e]E. Boll, H. Drobecq, N. Ollivier, A. Blanpain, L. Raibaut, R. Desmet, J. Vicogne, O. Melnyk, Nat. Prot. 2015, 10, 269-292. PMID: 25591010