Trypsiligase-Catalyzed Labeling of Proteins on Living Cells

Sandra Liebscher[+],[a] Sebastian Mathea[+],[a] Tobias Aumüller,[b] Andreas Pech,[a] and Frank Bordusa*[a]

Fluorescent fusion proteins are powerful tools for studying biological processes in living cells, but universal application is limited due to the voluminous size of those tags, which might have an impact on the folding, localization or even the biological function of the target protein. The designed biocatalyst trypsiligase enables site-directed linkage of small-sized fluorescence dyes on the N terminus of integral target proteins located in the outer membrane of living cells through a stable native peptide bond. The function of the approach was tested by using the examples of covalent derivatization of the transmembrane proteins CD147 as well as the EGF receptor, both presented on human HeLa cells. Specific trypsiligase recognition of the site of linkage was mediated by the dipeptide sequence Arg-His added to the proteins' native N termini, pointing outside the cell membrane. The labeling procedure takes only about 5 minutes, as demonstrated for couplings of the fluorescence dye tetramethyl rhodamine and the affinity label biotin as well.

Fluorescence microscopy has revolutionized the process of studying biological processes in living cells.[1] The attachment of fluorescent probes to a biomolecule of interest, for example, a desired target protein, enables impressive images providing time-resolved insights into the protein's localization, interaction or half-life and degradation, respectively. The choice of an appropriate fluorophore and its attachment to the molecule of interest is however not in all cases trivial. The fusion of fluorescent proteins at the genetic level, such as the most popular green fluorescent protein (GFP), is a common and elegant strategy for the regiospecific installation of fluorophores on a protein of interest (POI). However, the high molecular weight of ~25 kD means that GFP is not tolerated by all POIs.[2] It should be kept in mind that folding, localization or even the function of the target protein can be affected by the GFP fusion.[3] In contrast, a broad range of commercially available chemoselective and low-molecular-weight reagents exists (e.g., maleimides or succinimides);[4] unfortunately they lack any regiospecificity. As a result, the natural abundance of numerous reactive nucleophilic moieties (e.g., amino groups) in living cells leads to heterogeneous product mixtures and substantial background signal for fluorescence microscopy. Due to their inherent stereo- and regiospecificity and mild reaction conditions, enzymes fulfill all the requirements to mediate site-specific protein derivatization under physiological conditions. A whole range of naturally occurring and engineered enzymes have been established for the gentle modification of proteins in living cells. Phoshopantheinyl transferases,[5] *Escherichia coli* Biotin Ligase BirA,[6] Sortase A from *Staphylococcus aureus*[7] as well as Butelase1 from *Clitoria ternatea*[8] are prominent examples. These enzymes are capable of covalently attaching functional probes to cell-surface-presented proteins if a suitable substrate and the required recognition sequence on the target protein are provided.

We previously established the trypsiligase conjugation technology for enzymatic modification of proteins.[9] Trypsiligase is an engineered trypsin variant (i.e., trypsin K60E/N143H/E151H/D189K) characterized by its pronounced acyl transfer and exceptional low proteolytic activity. Based on the substrate mimetics concept trypsiligase-catalyzed N-terminal modification of proteins with the N-terminal dipeptide sequence Arg-His occurs under native physiological conditions with quantitative product yields in *in vitro* reactions.[9] Such substrate mimetics as peptidyl-4-guanidinophenyl esters (OGp),[10] whose ester leaving groups mimic trypsin-specific arginyl substrate side chains, allow for coupling of diverse and even nonpeptidic acyl moieties to any target bearing the desired dipeptide recognition sequence.

The OGp ester is easy to handle and simplified protocols for its synthesis were recently published.[11] On the other hand, it is not yet clear whether this approach also allows the modification of proteins under *in vivo* conditions. One has to keep in mind that trypsiligase originates from the proteolytically highly active trypsin, which is generally toxic to living systems. Furthermore, the biological stability of the OGp ester would also be a prerequisite for the successfull *in vivo* application of this approach. In this contribution, we present the function of the substrate mimetics-based trypsiligase approach for the site-specific N-terminal labeling of membrane-bound proteins on

[a] *Dr. S. Liebscher,[+] Dr. S. Mathea,[+] Dr. A. Pech, Prof. Dr. F. Bordusa*
Institute of Biochemistry/Biotechnology, Charles Tanford Protein Centre
Martin Luther University Halle-Wittenberg
Kurt-Mothes-Str. 3a, 06120 Halle (Germany)
E-mail: frank.bordusa@biochemtech.uni-halle.de
[b] *Dr. T. Aumüller*
Max Planck Research Unit for Enzymology of Protein Folding
Weinbergweg 22, 06120 Halle (Germany)
[+] *These authors contributed equally to this work.*

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cbic.202000718

This article is part of a joint Special Collection with the Journal of Peptide Science on SPP 1623: Chemoselective reactions for the synthesis and application of functional proteins. Please see our homepage for more articles in the collection.

living cells (Figure 1a). Orthogonality to other conjugation approaches is achieved by different requirements on the substrate structure, the unique recognition sequence, the differences in the enzyme's mode of action and the linker structure of the final product.

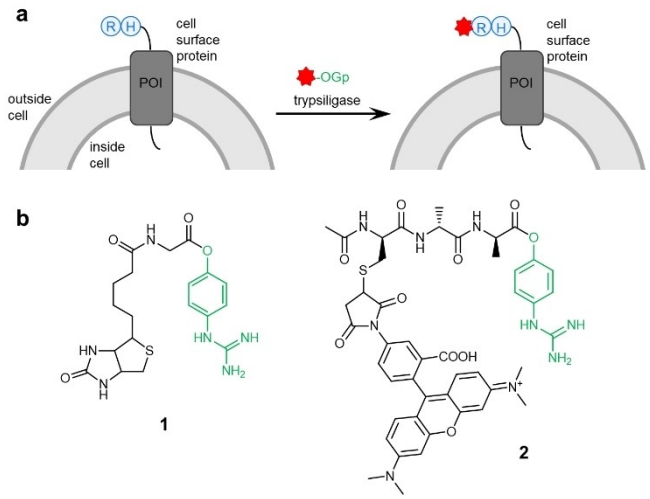

Figure 1. Trypsiligase-catalyzed modification of membrane-bound proteins by the substrate mimetics concept. a) Schematic view of the labeling procedure. The N-terminal Arg-His moiety (blue) of the protein of interest (POI) is presented outside the living cell and covalently modified by acyl transfer from the 4-guanidinophenyl ester (OGp; green) catalyzed by trypsiligase. b) Chemical structures of labeling probes used in this study. Compound **1**: biotinyl-glycyl-4-guanidinophenyl ester (biotinyl-Gly-OGp), compound **2**: acetyl-cysteinyl(tetramethyl rhodamine)-alanyl-alanyl-4-guanidinophenyl ester (ac-Cys(TAMRA)-Ala-Ala-OGp).

The proof of concept was performed with an OGp ester derived from the fluorophore tetramethyl rhodamine (TAMRA; Figure 1b, compound **2**). To demonstrate the flexibility of the approach for coupling functionalities other than fluorophores, an OGp ester containing the affinity tag biotin was also used (Figure 1b, compound **1**). The model proteins were either the epidermal growth factor receptor (EGFR) or CD147 (EMMPRIN or basigin). Both are transmembrane proteins that are involved in a number of tumor diseases. CD147, which is a glycosylated protein that belongs to the immunoglobulin superfamily, exists both in a transmembrane and soluble form.[12] Its extracellular immunoglobulin-like domains are essential for the invasion of *Plasmodium falciparum* into erythrocytes.[13] We transiently expressed full-length CD147, with the residues 22 and 23 mutated to Arg and His, in HeLa cells. After intracellular cleavage of the N-terminal signal peptide 1–21 of CD147, the trypsiligase recognition sequence Arg-His should be N-terminal of CD147 and located outside the living cell (Figure 2a). To specifically monitor CD147 by confocal microscopy, enhanced green fluorescent protein (eGFP) was additionally fused to the intracellular C terminus of the protein. The eGFP fluorescence was also used to demonstrate co-localization of the membrane-bound receptor and the trypsiligase-inserted TAMRA-fluorescence dye. The CD147-eGFP construct was previously

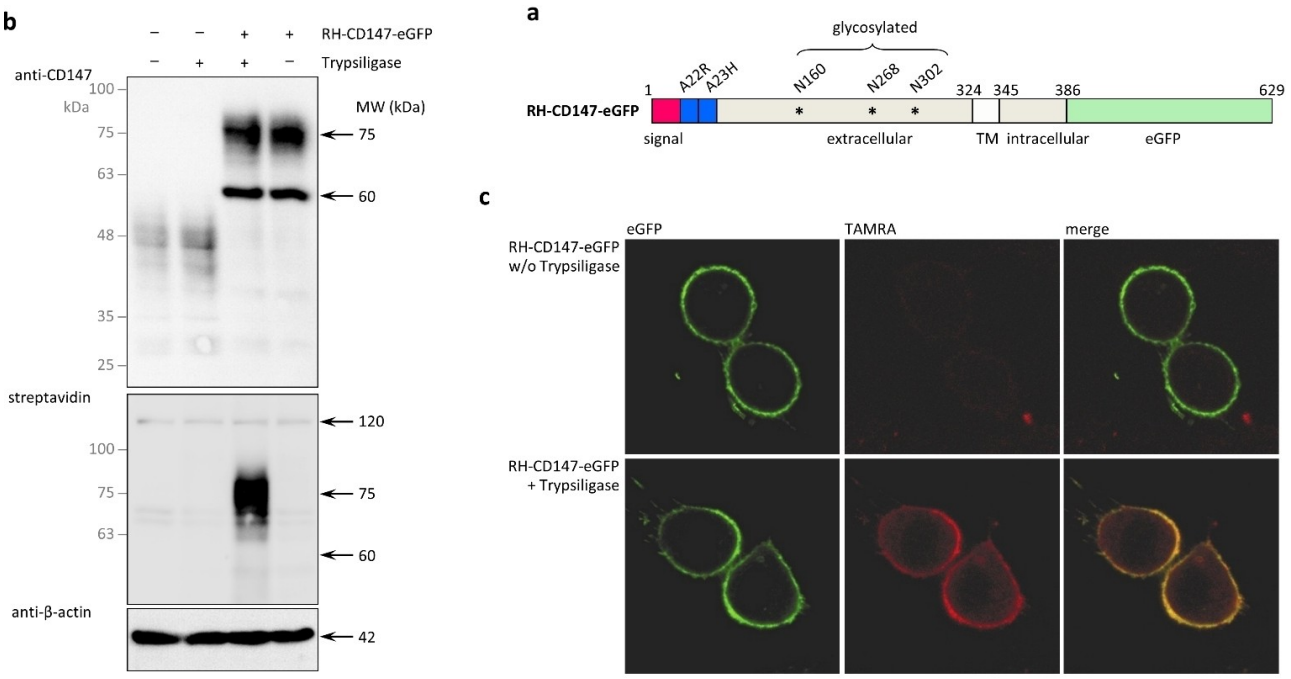

Figure 2. Trypsiligase-catalyzed modification of the receptor protein CD147. a) Schematic view of the primary structure of the target protein. The N-terminal signal sequence is cleaved off during translocalization of the protein into the membrane, generating the N-terminal dipeptide sequence Arg-His. TM: transmembrane segment. b) Western blot analysis of the biotinylation reactions. HeLa cells expressing the appropriate construct were incubated with 5 μM trypsiligase and 50 μM biotinylated OGp ester (**1**) for 5 min. Cells were analyzed by western blotting using anti-CD147 antibody and peroxidase-conjugated streptavidin. c) Confocal microscopy analysis of fluorescence labeling reactions. Cells were incubated with 5 μM trypsiligase and 50 μM TAMRA-conjugated OG ester (**2**) for 5 min. After several washing steps, cells were analyzed by confocal microscopy. eGFP was excited at 488 nm, and the emitted light was detected after passing a 515/30 filter. TAMRA was excited at 561 nm, and the emitted light was detected after having passed a 650LP filter.

shown to be physiologically similar to the wild-type CD147.[14] We could further demonstrate that the expression and localization of CD147-eGFP was not affected by the introduction of the short dipeptide trypsiligase tag (Figure S1 in the Supporting Information). In the post-mitotic cells the protein is, as expected, clearly localized in the plasma membrane (Figure S1b). For the initial biotinylation reactions of CD147-eGFP, the presenting HeLa cells were incubated with 50 μM compound **1** in the presence of trypsiligase (5 μM) for 5 min (Supporting Methods). Subsequently, the lysed cells were analyzed by western blotting using peroxidase-conjugated streptavidin, which binds specifically to the biotin moiety, and an anti-CD147 antibody that recognizes both biotinylated and non-biotinylated CD147 (Figure 2b). Detection with peroxidase-conjugated streptavidin shows a strong biotinylation level of cell-surface bound CD147 only in the presence of trypsiligase, whereas no biotinylation of the target protein could be detected in reactions without biocatalyst (cf. Figure 2b). Comparative *in vitro* experiments with the heptapeptide H-RHGFVQA-NH_2, which corresponds to the N-terminal sequence of CD147, proved the specific recognition of the RH-CD147 sequence by trypsiligase. In this case quantitative product yields could be obtained (Figure S2). Furthermore, the analysis of biotinylated CD147 with anti-CD147 antibody show no difference in the protein ladders in reactions with and without enzyme.

Thus, there is no evidence for proteolytic site reactions caused by trypsiligase. It should be noted that western blotting revealed two forms of CD147, one at 60 kDa and the other at 75 kDa, which are described to derive from receptor shedding (lanes 2 and 3 in Figure 2b).[15] Lanes 1 and 2 of Figure 2b show both protein species in their wild-type forms without eGFP and RH-mutation prior transfection of the cells. Consistently, only the high molecular weight form (75 kDa) was labeled with biotin, while the 60 kDa species was not modified due to the missing N-terminal Arg-His-moiety (cf. Figure 2b).

Next, we tested the flexibility of the trypsiligase approach versus the POI by exchanging CD147 with the EGF receptor. The latter represents a typical member of the protein kinase superfamily. As such, it dimerizes after ligand binding and is characterized by a ligand dependent tyrosine autophosphorylation activity resulting in cell proliferation.[16] First, the trypsiligase dipeptide recognition sequence was likewise introduced by mutating the extracellular N-terminal amino acids 25 and 26 of the receptor to Arg and His, as already tested for CD147 (Figure S3). It has also been demonstrated that neither the C-terminal fluorescent fusion nor the introduction of the short dipeptide trypsiligase tag affects the expression or localization of the EGF receptor (data not shown). Furthermore, previous studies by Subach et al. showed no influence of the C-terminal fluorescent protein on the physiology of the EGF receptor.[17] The analysis of the modification reactions was again performed by western blotting using anti-EGFR antibody and peroxidase-conjugated streptavidin. In the presence of trypsiligase, a high degree of receptor biotinylation was again detected (Figure S3b). In contrast, no biotinylation was observed in the absence of the biocatalyst.

Finally, we investigated the flexibility of the trypsiligase approach with regard to the functionality to be introduced in the POI. For this purpose, we exchanged the biotin-derived OGp ester **1** with the TAMRA-conjugated counterpart **2**. Arg-His-CD147-eGFP containing HeLa cells were again incubated for five minutes in the labeling buffer according to the reactions performed with compound **1**. In parallel, analog control reactions without trypsiligase were performed. Subsequent analysis by confocal microscopy clearly show a localization of the fluorescence dye on the cell surface only in the presence of the biocatalyst. Furthermore, the overlay of eGFP and TAMRA fluorescence showed a distinct co-localization on the membrane visible by a distinct orange color of the cell membrane (Figure 2c).

Based on the acceptance of a large number of additional functionalities and POI's shown in previous studies,[9,18] these results suggest that trypsiligase in combination with OGp esters represents a powerful tool also for the site-specific modification of proteins on the surface of living cells. The recognition of the target protein by trypsiligase only requires the introduction of a simple dipeptide sequence at its N terminus. The reactions themselves are completely enzyme-specific. Neither with the OGp ester alone nor with the biocatalyst in the absence of the trypsiligase recognition sequence at the target proteins, unwanted background signals based on unspecific reactions were detected. There are also, no site reactions based on the original high proteolysis activity of the parental wt trypsin. Its zymogen-like structure enables the formation of an acyl enzyme intermediate by OGp esters, but not the hydrolysis of originally trypsin-specific peptide bonds making trypsiligase-catalyzed bond formations irreversible.[9] This high reaction specificity could also explain why trypsiligase does not affect the cell viability (cf. Table S1) and could also be an explanation for the high velocity of these labeling reactions. Finally, trypsiligase is a robust enzyme in terms of its high tolerance towards pH, temperature or various additives as it is known for trypsin. Like the parental enzyme it is easy to handle, requires no cofactors or reducing conditions and can be produced in large quantities and high purity. In summary, the approach presented here is a practical and at the same time flexible approach for site-specific N-terminal modification of proteins on the surface of living cells, which is not limited to the examples shown here but also includes other membrane proteins or functionalities such as, click reactants or non-expressible ligands.[18]

Acknowledgements

The authors are grateful for the professional technical assistance of C. Langer and S. Kaufmann. Financial support by the SPP 1623 "Chemoselctive reactions for the synthesis and application of functional proteins" (project: "Semisynthesis of antibody drug conjugates via biocatalysis") is grateful acknowledged. Open access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: cell-surface engineering · site-specific labeling · substrate mimetics · trypsiligase · trypsin variants

[1] J. W. Lichtman, J. A. Conchello, *Nat. Methods* **2005**, *2*, 910–919.
[2] G.-J. Kremers, S. G. Gilbert, P. J. Cranfill, M. W. Davidson, D. W. Piston, *J. Cell Sci.* **2011**, *124*, 157–160.
[3] C. S. Lisenbee, S. K. Karnik, R. N. Trelease, *Traffic* **2003**, *4*, 491–501.
[4] E. Baslé, N. Joubert, M. Pucheault, *Chem. Biol.* **2010**, *17*, 213–227.
[5] N. George, H. Pick, H. Vogel, N. Johnsson, K. Johnsson, *J. Am. Chem. Soc.* **2004**, *126*, 8896–8897.
[6] I. Chen, M. Howarth, W. Lin, A. Y. Ting, *Nat. Methods* **2005**, *2*, 99–104.
[7] M. W. Popp, J. M. Antos, G. M. Grotenbreg, E. Spooner, H. L. Ploegh, *Nat. Chem. Biol.* **2007**, *3*, 707–708.
[8] X. Bi, J. Yin, G. K. T. Nguyen, C. Rao, N. B. A. Halim, X. Hemu, J. P. Tam, C. F. Liu, *Angew. Chem. Int. Ed.* **2017**, *56*, 7822–7825; *Angew. Chem.* **2017**, *129*, 7930–7933.
[9] S. Liebscher, M. Schöpfel, T. Aumüller, A. Sharkhuukhen, A. Pech, E. Höss, C. Parthier, G. Jahreis, M. T. Stubbs, F. Bordusa, *Angew. Chem. Int. Ed.* **2014**, *53*, 3024–3028; *Angew. Chem.* **2014**, *126*, 3068–3072.
[10] M. Thormann, S. Thust, H. J. Hofmann, F. Bordusa, *Biochemistry* **1999**, *38*, 6056–6062.
[11] S. Liebscher, F. Bordusa in *Bioconjugation: Methods and Protocols* (Eds.: S. Massa, N. Devoogdt), Springer, New York, **2019**, pp. 95–115.
[12] L. Xiong, C. Edwards, L. Zhou, *Int. J. Mol. Sci.* **2014**, *15*, 17411.
[13] K. E. Wright, K. A. Hjerrild, J. Bartlett, A. D. Douglas, J. Jin, R. E. Brown, J. J. Illingworth, R. Ashfield, S. B. Clemmensen, W. A. de Jongh, S. J. Draper, M. K. Higgins, *Nature* **2014**, *515*, 427–430.
[14] S. Zucker, M. Hymowitz, E. E. Rollo, R. Mann, C. E. Conner, J. Cao, H. D. Foda, D. C. Tompkins, B. P. Toole, *Am. J. Pathol.* **2001**, *158*, 1921–1928.
[15] W. Tang, S. B. Chang, M. E. Hemler, *Mol. Biol. Cell* **2004**, *15*, 4043–4050.
[16] X. Zhang, J. Gureasko, K. Shen, P. A. Cole, J. Kuriyan, *Cell* **2006**, *125*, 1137–1149.
[17] F. Subach, L. Zhang, T. W. Gadella, N. G. Gurskaya, K. A. Lukyanov, V. V. Verkhusha, *Chem. Biol.* **2010**, *17*, 745–755.
[18] C. Meyer, S. Liebscher, F. Bordusa, *Bioconjugate Chem.* **2016**, *27*, 47–53.

Manuscript received: October 15, 2020
Revised manuscript received: November 10, 2020
Accepted manuscript online: November 11, 2020
Version of record online: January 15, 2021