

Published in final edited form as:

*Bioorg Med Chem Lett.* 2008 November 15; 18(22): 6000–6003. doi:10.1016/j.bmcl.2008.08.044.

## Tags for Labeling Protein N-termini with Subtiligase for Proteomics

Hikari A. I. Yoshihara, Sami Mahrus, and James A. Wells\*

Departments of Pharmaceutical Chemistry and Cellular & Molecular Pharmacology, University of California, San Francisco, CA 94158-2330, USA

### Abstract

The peptide ligase subtiligase, derived from subtilisin, has been employed in the identification of protein N-termini in complex mixtures. Here, the peptide ester substrates for the ligation reaction were optimized with respect to solubility, resulting in greater incorporation of the N-terminal tags. Additionally, the quantitation of the incorporated tags was explored, and a “click” chemistry-based derivatization provided the ability to quantitate the tag to low nanomolar concentrations by sandwich ELISA. These new tags should expand the utility of subtiligase for the proteomic study of N-termini.

The study of the proteome is aided by tools that can separate specific subsets of proteins. For example, in studying post-translational modifications (PTM), methods for enriching phosphoproteins and phosphopeptides have been developed.<sup>1</sup> One common PTM in biology is the production of new N-termini by proteolysis; another is the modification of N-termini through acetylation<sup>3</sup> or myristoylation.<sup>4</sup> Several labs have described indirect means for isolation or enrichment of N-terminal peptides. These techniques include strong cation exchange chromatography of tryptic peptides,<sup>5</sup> and the chemical labeling of amine-containing functionality in the proteome<sup>6</sup> followed by either diagonal chromatography<sup>7</sup> or affinity purification.<sup>8</sup> Alternatively, our laboratory has developed a direct enzymatic method to affinity tag and enrich for the subset of proteins with free N-termini in order to characterize cellular substrates of proteolysis, or, more generally, characterize the ensemble of native N-termini. A recent study using this method identified nearly three hundred caspase substrates and their precise cleavage sites from apoptotic Jurkat cells.<sup>9</sup>

This method utilizes subtiligase, a variant of the protease subtilisin BPN' that contains a cysteine residue in place of the active site serine-221 and an alanine residue replacing proline-225.<sup>10</sup> Unlike the wild-type enzyme, the resulting double mutant has negligible protease activity, yet it can efficiently acylate free N-termini of peptides or proteins using ester substrates. The workflow for the application of subtiligase to the study of the proteome is outlined in Scheme 1. A complex mixture of proteins is N-terminally labeled with subtiligase using a peptide ester containing a biotin affinity tag and a TEV protease<sup>11,12</sup> cleavage site. Tagged proteins are separated from the unincorporated tag and digested with trypsin. Tagged N-terminal peptides are captured on avidin beads, cleaved off the beads by TEV protease treatment and analyzed by LC-MS/MS. While approximately 80% of N-termini are blocked in mammalian cells, mostly due to acetylation,<sup>13,14</sup> certain populations of proteins are believed to have a greater fraction of free N-termini. These include signal-sequence-dependent proteins transported across membranes, such as nuclear-encoded mitochondrial proteins.<sup>15,8</sup>

\*Corresponding Author; tel: +1.415.514.4498; fax: +1.415.514.4507; email: jim.wells@ucsf.edu

While this approach is a powerful and sensitive tool for identifying free protein N-termini and the precise sites of proteolysis, it has limitations. Subtiligase will hydrolyze the ester substrate and ligation yields vary depending on the accessibility and, to some extent, on the nature of the N-terminal residues.<sup>16</sup> In principle, increasing amount of peptide ester will increase the extent of the ligation reaction by providing more ester for ligation before it is consumed by hydrolysis. TEVest2 (**1**, Table 1), the prototypical peptide ester, is somewhat insoluble in aqueous solution and cannot be used at concentrations higher than 1 mM.

Here, we explored whether it would be possible to improve the yield of the tagging reaction and increase proteomic coverage by producing more soluble peptide esters. The general strategy in these designs was to add basic or acidic residues such that the net charge of the ester would be strongly positive or negative at neutral pH. Some of the esters conjugated to biotin using a PEG-type linker instead of the two aminohexanoic acid residues in **1**. In addition, to facilitate the routine quantitation of the tagged proteins in complex mixtures by sandwich ELISA, the incorporation of 3-nitrotyrosine and propargylglycine residues into the esters was also investigated. The primary structures of the peptide esters and their solubilities are listed in Table 1.

Ester **2**, a derivative of TEVest2 (**1**) that contained an additional four glutamic acid residues at the N-terminus of the peptide portion, showed improved solubility. However, a much greater increase in solubility was observed with analogues that instead contained additional arginine residues (**3-5**). We chose to incorporate D-arginine in order to preserve the integrity of the tag during the trypsin digest step of the proteomic workflow.

We also wished to produce peptide analogs that could be used to introduce functionality for quantitation and labeling post-TEV protease cleavage. Quantitation would be an important new feature allowing users to analyze how much labeling one has in the sample prior to any further fractionation and LC-MS/MS. We chose to test incorporating 3-nitrotyrosine since this single residue can be used in very quantitative and sensitive ELISA assays.<sup>17</sup> Peptides containing this residue, however, were significantly less soluble. Alternatively, we found that propargylglycine could be incorporated without loss in solubility. The terminal alkyne side chain of propargylglycine provides a “click”-able handle for derivatization with a variety of functional groups post-subtiligase labeling using the copper-catalyzed alkyne-azide cyclization reaction<sup>18,19</sup> with an azide-bearing chemical tag. Ester **4** was thus designed to combine some of the features of a highly soluble peptide ester and contain a propargylglycine residue immediately N-terminal to the ester group. Since non- $\beta$ -branched hydrophobic residues at this position are good substrates for subtilisin<sup>20,21</sup> it was anticipated that the propargyl side chain would be well tolerated. At this position, the propargylglycine residue can function in two roles. First, it can function as a tag for derivatization and quantitation, in principle along the entire workflow and not only prior to the TEV protease cleavage step as with **3**. Second, as a non-natural amino acid, it will produce unique  $a_2$  and  $b_2$  ions when tagged peptides are fragmented during tandem mass spectrometry. Abundant  $a_2$  and  $b_2$  ions, corresponding to the residual serine and tyrosine residues of the tag, are observed with a QqTOF instrument when analyzing proteomic samples labeled using ester **1** and are a hallmark of tagged N-terminal peptides. The incorporation of propargylglycine instead could provide an even more distinct signature.

The performance of the new esters was assessed by small-scale ligation reactions with Jurkat cell lysates, which were then analyzed by SDS-PAGE and blotted with streptavidin to detect the incorporated tag (Figure 1). At 1 mM, all the esters show comparable or modest improvements in labeling compared to **1**. However at higher concentrations the more soluble penta-Arg containing esters (**3**, **4** and **5**) showed significantly higher levels of labeling. Redistribution of some of the Arg residues from the N- to the C-terminal side of the ester

(compare **3** vs. **5**) did not significantly alter their performance. Ester **4**, which introduces a propargylglycine residue directly adjacent to the ligated protein's N-terminus, does not appear to perform as well with respect to labeling as **3** or **5** but is still significantly better than **1**. We note that esters **3**, **4** and **5** caused some protein precipitation in cell lysates. It is known that proteins containing N-terminal polyarginine tags are less thermostable,<sup>22</sup> but we could prevent precipitation by the addition of sodium chloride, urea or guanidinium chloride.

As a model to test the performance of a quantitation strategy involving derivatization of a propargylglycine residue (Figure 2a), tripeptides consisting of (Ac)GlyTyr(NO<sub>2</sub>)Ala, a PEG linker and an azide group were prepared. The derivatization of the propargylglycine residue of **3** pre-bound to neutravidin in 96-well plates proceeded efficiently under typical conditions (Figure 2b). The large excess (250 equivalents or greater) of the azide-bearing nitrotyrosine peptide appears to ensure the complete reaction of the avidin-immobilized **3**, as standard curves using peptides containing nitrotyrosine residues and pre-derivatized peptides showed similar sensitivity compared to standard curves prepared by on-plate derivatization. Using this assay, **3** can be reliably detected at 30 nM, a level of sensitivity that allows quantitation of ligated N-termini to an estimated lower limit of 0.2% of free N-termini in a eukaryotic cell lysate. In practice, the material analyzed would be a complex mixture of proteins, with some fraction N-terminally labeled and any residual unreacted tag removed by gel filtration.

In summary, we have developed new peptide esters as subtiligase substrates for use in the proteomic labeling of free N-termini. By incorporating several additional arginine residues, the solubility of the peptides is markedly improved, and the ligation reaction can be driven further by the addition of more ester. Adding propargylglycine residues into the peptide esters allows the quantitation of the labeled N-termini by sandwich ELISA following derivatization with a nitrotyrosine and azide containing peptide. These features should to expand the utility of subtiligase for the study of proteolysis and other PTMs of protein N-termini.

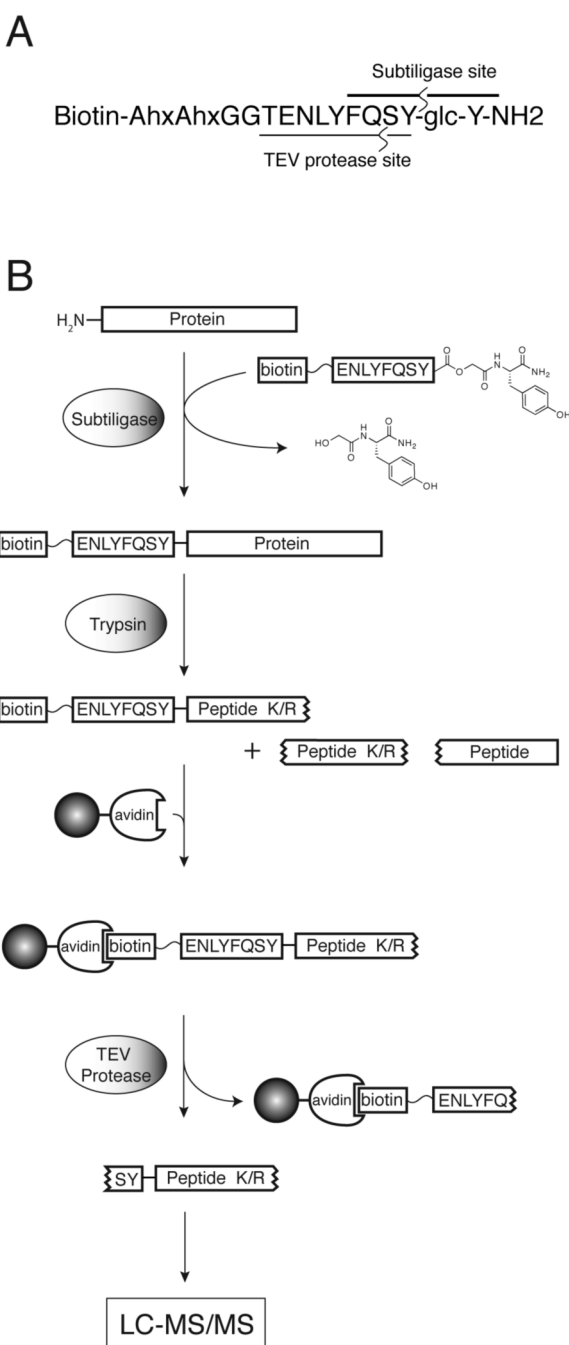
## Acknowledgements

We wish to thank Pete Wildes, Nick Agard and Emily Crawford for helpful discussions. Financial support was provided by the Sandler Foundation, NIH (RO1 GM081051), and the Hartwell Foundation. HAIY was supported in part by a Stewart Trust Cancer Research Award from the UCSF Cancer Center.

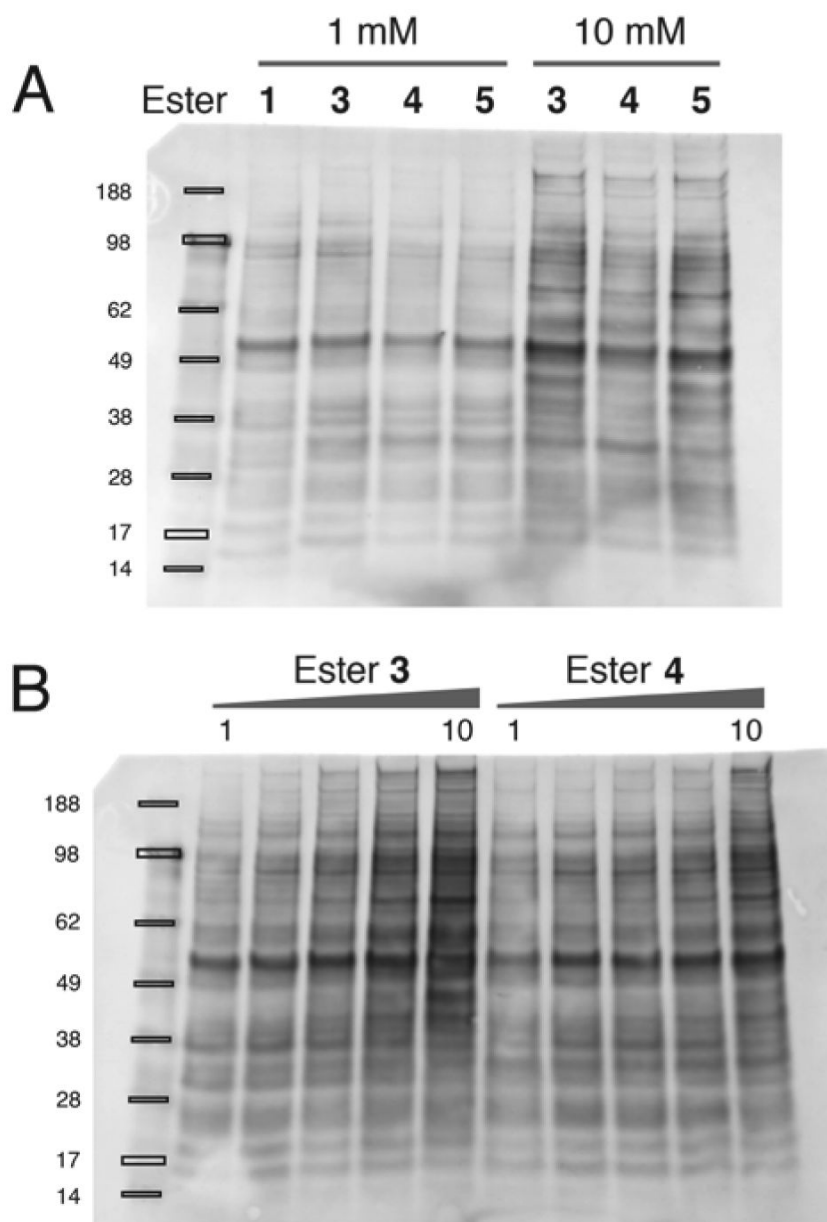
## References

1. Reinders J, Sickmann A. *Proteomics* 2005;5:4052. [PubMed: 16196093]
2. Bradshaw RA, Brickey WW, Walker KW. *Trends Biochem. Sci* 1998;23:263. [PubMed: 9697417]
3. Polevoda B, Sherman F. *J. Biol. Chem* 2000;275:36479. [PubMed: 11013267]
4. Towler DA, Gordon JI, Adams SP, Glaser L. *Annu. Rev. Biochem* 1988;57:69. [PubMed: 3052287]
5. Dormeyer W, Mohammed S, Breukelen B, Krijgsveld J, Heck AJ. *J. Proteome Res* 2007;6:4634. [PubMed: 17927228]
6. Enoksson M, Li J, Ivancic MM, Timmer JC, Wildfang E, Eroshkin A, Salvesen GS, Tao WA. *J. Proteome Res* 2007;6:2850. [PubMed: 17547438]
7. Gevaert K, Goethals M, Martens L, Van Damme J, Staes A, Thomas GR, Vandekerckhove J. *Nat. Biotechnol* 2003;21:566. [PubMed: 12665801]
8. Timmer JC, Enoksson M, Wildfang E, Zhu W, Igarashi Y, Denault JB, Ma Y, Dummitt B, Chang YH, Mast AE, Eroshkin A, Smith JW, Tao WA, Salvesen GS. *Biochem. J* 2007;407:41. [PubMed: 17650073]
9. Mahrus S, Trinidad JC, Barkan DT, Sali A, Burlingame AL, Wells JA. *Cell* 2008;134:866. [PubMed: 18722006]
10. Abrahmsen L, Tom J, Burnier J, Butcher KA, Kossiakoff A, Wells JA. *Biochemistry* 1991;30:4151. [PubMed: 2021606]

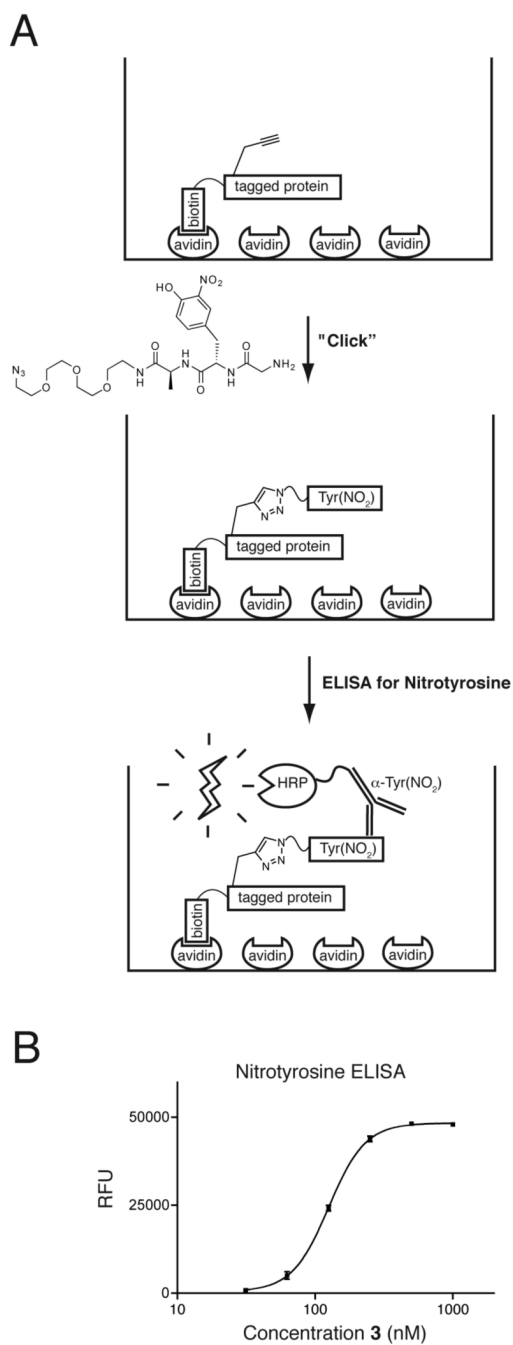
11. Carrington JC, Dougherty WG. Proc. Natl. Acad. Sci. U. S. A 1988;85:3391. [PubMed: 3285343]
12. Speers AE, Cravatt BF. J. Am. Chem. Soc 2005;127:10018. [PubMed: 16011363]
13. Brown JL. Biochim. Biophys. Acta 1970;221:480. [PubMed: 5499431]
14. Brown JL, Roberts WK. J. Biol. Chem 1976;251:1009. [PubMed: 1249063]
15. Neupert W. Annu. Rev. Biochem 1997;66:863. [PubMed: 9242927]
16. Chang TK, Jackson DY, Burnier JP, Wells JA. Proc. Natl. Acad. Sci. U. S. A 1994;91:12544. [PubMed: 7809074]
17. Ye YZ, Strong M, Huang ZQ, Beckman JS. Methods Enzymol 1996;269:201. [PubMed: 8791650]
18. Speers AE, Cravatt BF. Chem Biol 2004;11:535. [PubMed: 15123248]
19. Wang Q, Chan TR, Hilgraf R, Fokin VV, Sharpless KB, Finn MG. J. Am. Chem. Soc 2003;125:3192. [PubMed: 12630856]
20. Gron H, Meldal M, Breddam K. Biochemistry 1992;31:6011. [PubMed: 1627543]
21. Wells JA, Cunningham BC, Graycar TP, Estell DA. Proc. Natl. Acad. Sci. U. S. A 1987;84:5167. [PubMed: 3299378]
22. Fuchs SM, Raines RT. Protein Sci 2005;14:1538. [PubMed: 15930002]

**Scheme 1.**

A. Structure of peptide ester **1** with ligation and TEV protease cleavage sites annotated. B. Proteomics workflow for N-terminal labeling using subtiligase.



**Figure 1.** Avidin blots of ligation reactions in Jurkat cell lysates. A. Comparison of peptide esters at 1 mM and 10 mM. The limited solubility of **1** precludes its use at 10 mM. B. Lysates incubated with increasing concentrations (1, 2, 4, 7 & 10 mM) of peptide ester in the presence of 10  $\mu$ M subtiligase.

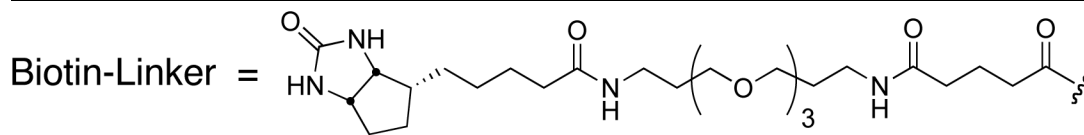


**Figure 2.** Quantitation of tagging by derivatization and ELISA. A. Tagged proteins are captured on an analytical scale in neutravidin-coated 96-well plates. The propargylglycine residue is derivatized using a copper(I)-catalyzed azide-alkyne cyclization reaction and the incorporated 3-nitrotyrosine is quantitated using an antibody conjugated to horseradish peroxidase (HRP). B. ELISA using peptide ester **3**.

**Table 1**

Peptide ester structures and solubilities

Ester	Sequence	Solubility (mM)
1	Biotin-AhxAhxGGTENLYFQSY-glc-Y-NH <sub>2</sub>	~1-2
2	Biotin-AhxAhx- <b>EEEE</b> -GGTENLYFQSY-glc-Y-NH <sub>2</sub>	~10
3	Biotin-Linker- <b>dRdRdRdR</b> -prG-TENLYFQSY-glc-R-NH <sub>2</sub>	>40
4	Biotin-Linker- <b>dRdRdRdR</b> -G-TENLYFQS-prG-glc-R-NH <sub>2</sub>	>50
5	Biotin-Linker- <b>dRdRdR</b> -prG-TENLYFQSY-glc-RR-NH <sub>2</sub>	>40



Ahx = 6-aminohexanoyl

prG = L-propargylglycyl

glc = glycolyl