ACS Partner Journal

Journal of the American Society for Mass Spectrometry

pubs.acs.org/jasms

Research Article

Mass Spectrometry-Cleavable Protein N-Terminal Tagging Strategy for System-Level Protease Activity Profiling

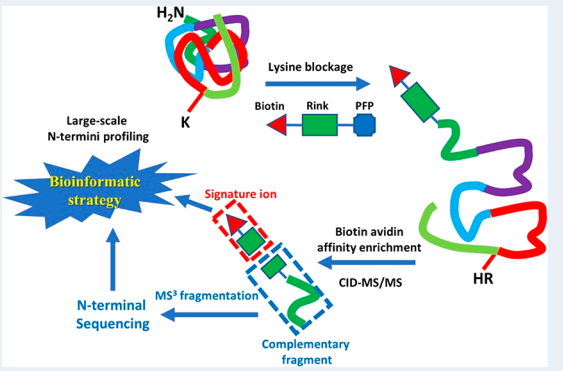
Zixiang Fang, Maheshika S. K. Wanigasekara, Akop Yepremyan, Brandon Lam, Pawan Thapa, Frank W. Foss, Jr., and Saiful M. Chowdhury*

Cite This: *J. Am. Soc. Mass Spectrom.* 2022, 33, 189−197

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Proteolysis is one of the most important protein post-translational modifications (PTMs) that influences the functions, activities, and structures of nearly all proteins during their lifetime. To facilitate the targeted identification of low-abundant proteolytic products, we devised a strategy incorporating a novel biotinylated reagent PFP (pentafluorophenyl)-Rink-biotin to specifically target, enrich and identify proteolytic N-termini. Within the PFP-Rink-biotin reagent, a mass spectrometry (MS)-cleavable feature was designed to assist in the unambiguous confirmation of the enriched proteolytic N-termini. The proof-of-concept study was performed with multiple standard proteins whose N-termini were successfully modified, enriched and identified by a signature ion (SI) in the MS/MS fragmentation, along with the determination of N-terminal peptide sequences by multistage tandem MS of the complementary fragment generated after the cleavage of MS-cleavable bond. For large-scale application, the enrichment and identification of protein N-termini from *Escherichia coli* cells were demonstrated, facilitated by an in-house developed NTermFinder bioinformatics workflow. We believe this approach will be beneficial in improving the confidence of identifying proteolytic substrates in a native cellular environment.



KEYWORDS: *N-terminal tagging, mass spectrometry, enrichment, proteolytic products, post-translational modifications*

Downloaded via UNIV OF TEXAS AT ARLINGTON on April 7, 2022 at 22:36:21 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

INTRODUCTION

Proteolysis is an irreversible and a highly selective hydrolytic reaction of peptide bond catalyzed by more than 560 proteases encoded by over 2% of human genome.[1] Proteolysis was once considered a predominantly protein-degrading and disposing procedure, but gradually it has been discovered to play pivotal roles of precisely adjusting and fine-tuning the protein sequences to construct protein structures, regulate protein functions and control protein cellular localizations in a myriad of biological events.[2] The intracellular proteolytic network of caspases regulates and activates the transmission of inflammatory and apoptotic signals to control many important pathways, one of such examples is the blood coagulation cascade.[3] Proteolysis is also intimately involved in significant biological implications and pathological conditions of various cancers, chronic inflammations, and neurodegenerative and cardiovascular diseases. It is estimated that 5−10% of all drug targets were designed for antiproteolytic therapies.[4] Because proteolysis occurs co-translationally or post-translationally, the functional N-termini of proteins cannot be directly inferred from the genome sequence or transcriptome sequence. To reveal the pathways and mechanisms of proteolysis, proteolytic substrates need to be determined, and the specifically cleaved positions by the proteases need to be identified through the study of protein degradomics.[1,2]

The detection of N-terminal sequences is obscured by numerous internal peptides generated after enzymatic digestion in a typical bottom-up proteomics study. *In silico* endopeptidase Arg-C digestion of human proteome estimates an average of 17.5 internal peptides per N-terminal peptide.[5] This challenge will be more pronounced and overwhelming for identifying N-termini and pertinent proteolytic products of low-abundant proteins in large-scale biological samples. Therefore, certain enrichment strategies of the original protein N-termini are essential for their confident identifications by mass spectrometric analysis. Enrichment of protein N-termini is conventionally achieved by either positive or negative selection strategies. Positive selection approaches specifically label the α-amine of the protein N-termini with an enrichment group while minimizing or preventing the mislabeling of lysine

Received: November 24, 2021
Accepted: December 2, 2021
Published: December 20, 2021

ACS Publications

© 2021 American Society for Mass Spectrometry. Published by American Chemical Society. All rights reserved.

https://doi.org/10.1021/jasms.1c00350
J. Am. Soc. Mass Spectrom. 2022, 33, 189−197

ε-amines, whose abundance is often of higher orders of magnitude. This is achieved either through chemical modifications (for example, by guanidination to block lysine ε-amine)[6] or enzymatically assisted modification (subtiligase has high specificity for N-terminal α-amine).[7] After sample digestion, protein N-termini can be selectively purified by targeting the enrichment group to be separated from the interfering internal peptides. On the contrary, negative selection approaches such as the combined fractional diagonal chromatography (COFRADIC)[5] and terminal amine isotopic labeling of substrates (TAILS)[8] instead target the internal peptides for depletion after the blockage of all primary amines in the proteins. Therefore, the N-terminal sequences of original proteins including the naturally blocked protein N-termini can be retained for LC−MS/MS analysis. Negative selection strategies are advantageous in studying protein N-termini with native modifications including N-terminal acetylation. However, enrichment by negative selection tends to result in higher sample complexity from the coenrichment of free and modified protein N-termini[9] and requires better enrichment efficiency for removing the more abundant internal peptides,[10] which can be inferior in selectively studying proteolytically processed protein N-termini compared to positive selection.[1]

Here, we designed a direct strategy to improve the confidence in the identification of proteolytic products by developing and utilizing an MS-cleavable reagent PFP (pentafluorophenyl)-Rink-biotin in positive selection enrichment approach for investigating proteolytic protein N-termini from large-scale samples. In the experimental workflow, lysine ε-amines were first blocked by a guanidination reaction. The N-terminal labeling reaction occurred as free α-amines are targeted by PFP ester and simultaneously biotinylated for downstream enrichment. The confident N-terminus identification was enabled by the generation of a fixed signature ion (SI) and a high-intensity complementary fragment upon the CID-MS^2 fragmentation, followed by further MS^3 fragmentation on the complementary fragment to reveal the sequence information on the enriched protein N-terminus. This strategy incorporated the MS-cleavable feature widely applied in improving the identification capabilities and lowering false-discovery rate of low-abundant peptides in protein post-translational modification (PTM)[11] and cross-linking proteomics study.[12−14]

To facilitate the automatic confirmation of N-termini, an in-house Java-based software NTermFinder was developed. The software validates the identified protein N-terminal peptides by the presence of the signature ion (SI) in the mass spectra, locating confident peptide precursor ions, MS^3 of the complementary fragments, and connecting scan numbers during MS and tandem MS acquisitions. This strategy aims to achieve confident identification of proteolytic products and protein N-termini in the native cellular settings.

EXPERIMENTAL SECTION

Materials and Reagents. Ubiquitin of bovine erythrocytes, lysozyme of chicken, and β-lactoglobulin of bovine milk, ammonium bicarbonate, sodium bicarbonate, dimethyl sulfoxide (DMSO), trifluoroacetic acid (TFA), chloroform, and guanidine hydrochloride were purchased from Sigma-Aldrich (MO). *O*-Methylisourea sulfate was obtained from TCI America (OR). Iodoacetamide (IAM), methanol, and acetonitrile (ACN) were obtained from VWR (PA). Formic acid (FA), 3K MWCO protein concentrators, monomeric avidin and streptavidin agarose resins, and C18 desalting tips (100 μL bed) were obtained from Thermo Scientific (IL). Dithiothreitol (DTT) was acquired BioRad (CA), trypsin was from Promega (WI), and endoproteinase GluC was from New England Biolabs (MA). 1,1,1,3,3,3-Hexafluoroisopropanol (HFIP) was obtained from Oakwood Chemical (SC). The bacterial cell lysis kit was acquired from Goldbio (MO). An Aries Filterworks (NJ) water system supplies all of the high-purity water used for preparing aqueous solutions.

Enrichment and Identification of Standard Protein N-Termini. A preliminary experiment for labeling and enriching protein N-termini was performed to identify the N-termini of ubiquitin, lysozyme, and β-lactoglobulin proteins. First, the guanidination procedure was performed as previously described to block the ε-amines of these proteins.[15] Briefly, *O*-methylisourea sulfate was dissolved in water and mixed with 1 mM, 5 μL of standard proteins. The pH of the reaction solution was adjusted to between 10 and 11 with the addition of NaOH and then incubated at 65 °C for 30 min. After guanidination, pH of the protein samples was adjusted to around 7 by adding HCl. PFP-rink-biotin was dissolved in DMSO and added to the guanidinated proteins to a final concentration of 2.5 mM. The α-amine-labeling reaction was allowed to proceed for 2 h at 37 °C. The excess PFP-Rink-biotin was quenched by adding Tris-HCl buffer to a final concentration of 20 mM and removed by either 3K MWCO or methanol−chloroform precipitation, followed by the disulfide cleavages with DTT and alkylation with IAM in the dark. After labeling, the sample was digested either with trypsin or GluC (1:50 w/w) overnight in 50 mM ammonium bicarbonate. The sample was reconcentrated by removing the solvent and reconstituted in pH 7.4 PBS buffer, followed by mixing with monomeric avidin resins for 1 h under room temperature. After peptide binding, the resins were washed extensively with PBS and ultrapure water sequentially to remove the non-binding peptides. Once the washing steps were complete, the beads were incubated with the elution buffer ACN/H_2O/TFA (50/50/0.4, v/v/v) for 1 h at room temperature. The upernatant was separated from the beads by centrifugation and was subsequently evaporated, desalted by C18 tips, and reconstituted in 0.1% FA for LC-MS^n analysis.

Enrichment and Identification of Proteolytic Products from *E. coli* Cell Lysate. The synthesis of PFP-Rink-biotin was prepared as previously reported[16] and is shown in Scheme S1. *Escherichia. coli (E. coli)* top 10 cell was a gift from Dr. Shawn Christensen's lab (UT Arlington). Cell lysis was performed with bacterial cell lysis kit (GoldBio cat. no. GB-177 and GB-176) according to the manufacturer's protocol, and the protein concentration was measured with bicinchoninic acid assay using bovine serum albumin (BSA) standards. Aliquots (0.1 mg) of *E. coli* cell lysate were dissolved with 6 M guanidine hydrochloride containing 50 mM sodium bicarbonate followed by reaction with 4 mM dithiothreitol (DTT) and then alkylation with 12 mM iodoacetamide (IAM) in the dark. Alkylation reaction was quenched by 4 mM DTT solution followed by guanidination reaction at 4 °C for 24 h as previously reported (after adjusting to pH 10−11).[17] Proteins were then purified by a methanol−chloroform method, reconstituted in PBS (pH 7.4), and reacted with 4 mM PFP-Rink-biotin for 2 h at 37 °C. After the reaction, excess PFP-Rink-biotin was quenched by adding Tris-HCl buffer and removed by methanol−chloroform purification. The samples were subjected to digestion with trypsin (1:50, w/w)

https://doi.org/10.1021/jasms.1c00350
J. Am. Soc. Mass Spectrom. 2022, 33, 189−197

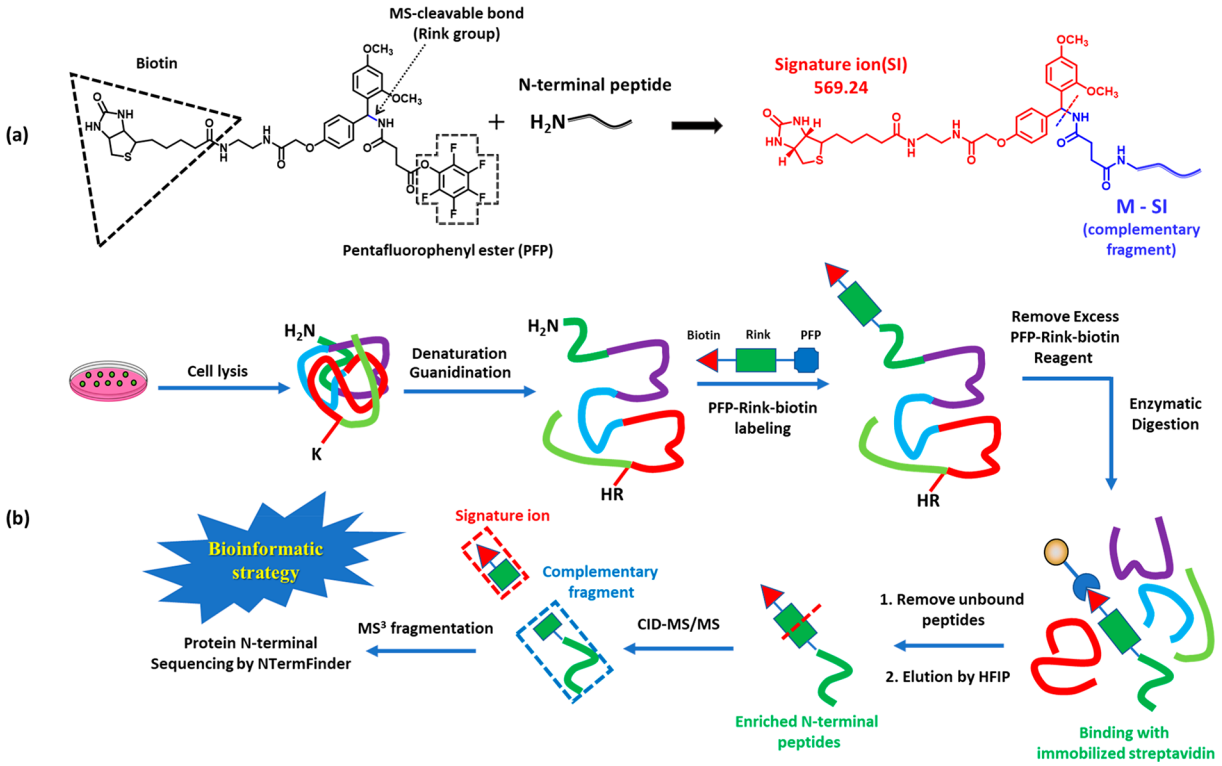


Figure 1. (a) Fragmentation of PFP-Rink-biotin-labeled N-terminal peptide: Under CID fragmentation, the PFP-Rink-biotin-coupled N-terminal peptide produces the signature ion (theoretical m/z 569.24) and complementary fragment in high abundance in the MS^2 spectrum. (b) Workflow of PFP-Rink-biotin-based enrichment and identification of protein N-termini: Following cell lysis, proteins were denatured and guanidinated to block primary amines of lysine. N-Terminal primary amines were then labeled by PFP-Rink-biotin followed by the removal of excess reagent and enzymatic digestion. N-Terminal peptides were selectively enriched by biotin−avidin affinity interaction and analyzed by LC−MS^n for N-termini identification using a bioinformatic approach.

overnight. The samples were then reconstituted in PBS buffer and then with mixing with phosphate-buffered saline (PBS) prewashed streptavidin agarose beads for 1 h at room temperature. After peptide binding, the filtrate was removed, and the beads were washed extensively and sequentially with PBS and ultrapure water to remove the nonbinding peptides. Once done with the washing steps, the beads were incubated with the elution buffer 1,1,1,3,3,3-hexafluoroisopropanol (HFIP) for 5 min at rt.[18] The filtrate was retained, evaporated, desalted by C18 tips, and reconstituted in 0.1% FA for LC-MS^n analysis.

LC-MS^n Analysis of Standard Protein N-Termini. All samples were analyzed in a Dionex UltiMate 3000 nano-UHPLC system joined with a nano-ESI-linear ion trap (LIT) Thermo Velos Pro mass spectrometer (Thermo Fisher Scientific, Waltham, MA). For standard protein-enriched samples, an Acclaim PepMap C18 column (150 mm × 75 μm, 3 μm) was used for LC separation with a 90 min gradient (mobile phase A: 0.1% FA in water; mobile phase B: 0.1% FA in 95% acetonitrile, 5% water; flow rate: 300 nL/min; 90 min gradient: 0−3 min 4.0% B, 3−80 min 4.0−50.0% B, 80−80.1 min 50−95% B, 80.1−85 min 95% B, 85−85.1 min 95−4% B, 85.1−90 min 4% B). The source voltage was 2.20 kV, and the capillary temperature was set to 275 °C. MS data was obtained from the 300 to 2000 m/z mass range. Data-dependent MS/MS spectra were collected for the 10 most abundant precursor ions in each MS scan upon fragmentation (charge state ≥2; isolated width of 2 Da; min signal required: 200) using CID activation with 35.0% normalized collision energy, an activation Q of 0.25, and an activation time of 30 ms. For MS^3 data-dependent acquisition, the precursor mass range was set according to the mass range of the analyte and MS^3 spectra were collected from the top two most abundant precursor ions upon fragmentation (charge state ≥2; isolated width of 2 Da; min signal required: 50) using CID. The activation was set with 45.0% normalized collision energy, activation Q of 0.25, and activation time of 30 ms.

LC-MS^n Analysis of *E. coli* Enriched N-Termini. For *E. coli* enriched samples, an Acclaim PepMap C18 column (500 mm × 75 μm, 3 μm) was used for the LC separation with the same mobile phases as above with a flow rate of 0.200 μL/min in two separate gradient separations (Gradient 1:0−5 min 4.0%B, 5−6 min 4.0−20.0% B, 6−145 min 20−70% B, 145−146 min 70−95% B, 146−150 min 95% B, 150−151 min 95−4% B, 151−170 min 4% B; Gradient 2:0−4 min 4.0% B, 4−120 min 4.0−50.0% B, 120−145 min 50−80% B, 145−145.1 min 80−95% B, 145.1−150 min 95% B, 150−151 min 95−4% B, 151−170 min 4% B). The source voltage was 2.00 kV, and the capillary temperature was set to 275 °C. MS data was obtained from the 400 to 2000 m/z mass range in rapid scanning mode. Data-dependent MS/MS spectra were collected for the five most abundant precursor ions in each MS scan upon fragmentation (charge state ≥2; isolated width of 2 Da; min signal required: 10000) using CID activation with 35.0% normalized collision energy, activation Q of 0.25, and activation time of 10 ms. The top two ions (excluding m/z 569.2) in each MS^2 spectrum were selected for MS^3 data-dependent acquisition (charge state ≥2; isolated width of 2 Da; min signal required: 500) using CID activation of 45.0%

https://doi.org/10.1021/jasms.1c00350
J. Am. Soc. Mass Spectrom. 2022, 33, 189−197

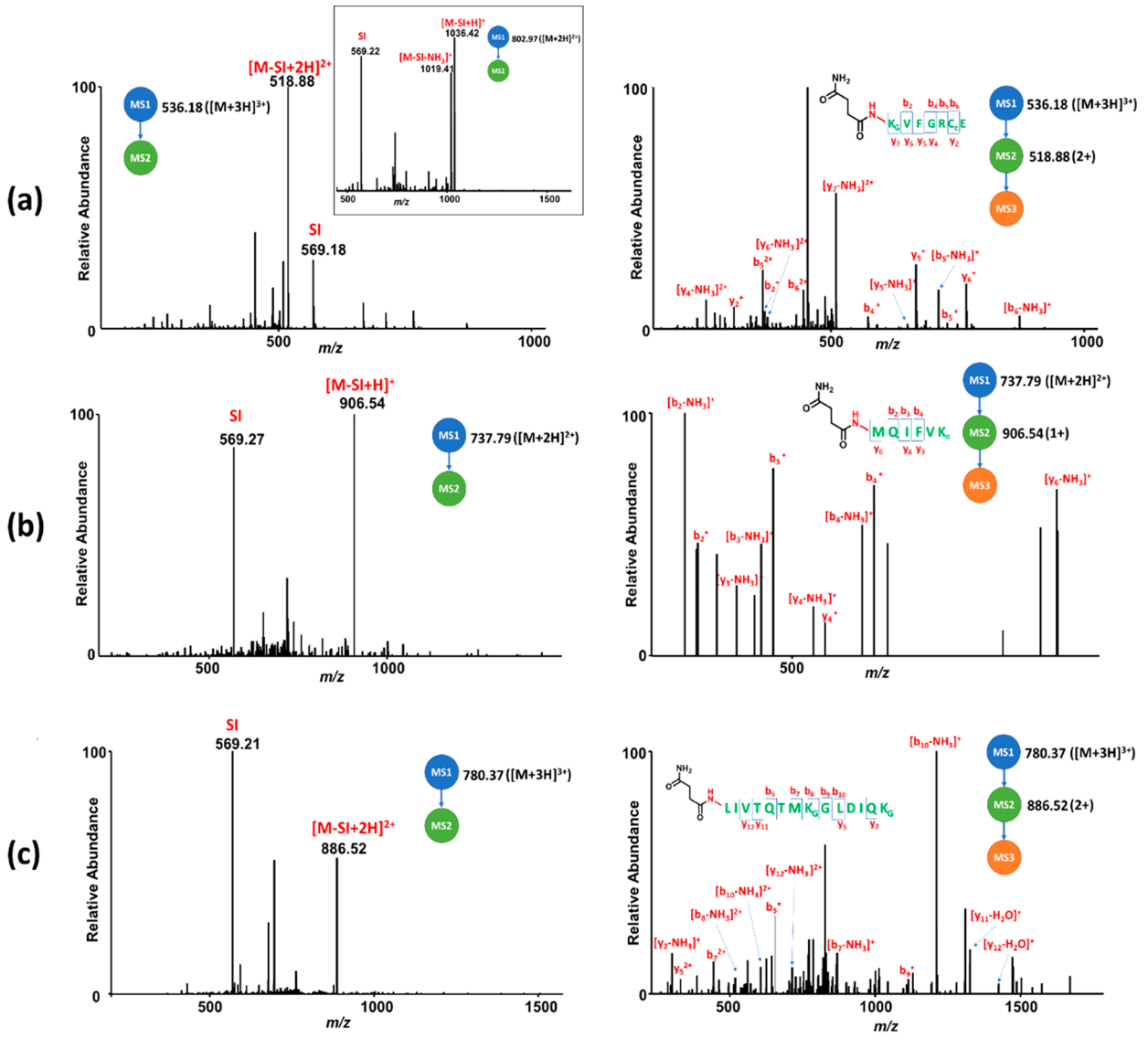


Figure 2. Identification of the labeled standard protein N-termini by MS^n analysis: (a) MS^2 and MS^3 spectra of lysozyme N-terminus; (b) MS^2 and MS^3 spectra of ubiquitin N-terminus; (c) MS^2 and MS^3 spectra of β-lactoglobulin N-terminus. In all cases, the signature ion (SI) and complementary fragments were generated in high relative abundances which facilitated the identification of N-termini labeling and subsequent MS^3 N-terminal sequencing.

normalized collision energy, an activation Q of 0.25, and an activation time of 10 ms.

MS Data Analysis. The MS raw files were converted to MGF format and mzML format (only MS^3 spectra) in Proteome Discoverer 2.1. MGF files were imported to in-house Java software NTermFinder to scan for SI (m/z 569.2) at ±1 Da mass threshold for all of the MS scans. The MS^3 mzML file was analyzed in MS-GF+ against the respective databases with the parameter settings as follows:[19,20] static modification of cysteine carbamidomethylation (+57.02146 Da) and lysine guanidination (+42.021798 Da); variable modification of methionine oxidation (+15.99492 Da). Additional variable modifications from reagent labeling of the protein N-terminus (+99.032027 Da) and methionine-removed protein N-terminus (−32.007458 Da) were added. The MS-GF+ identified peptide-spectrum matches (PSMs) were validated by NTermFinder based on the SI intensity, scan number, the m/z correlations among the SI, complementary fragment, and precursors as described in more details in the Discussion. A user manual to download and use NTermfinder is provided in the Supporting Information.

Data Availability. Mass spectrometric data files of the *E. coli* study are deposited in the ProteomeXchange via PRIDE (PRoteomics IDEntifications Database) repository with identifier PXD023953. The source code of the NTermFinder is available in GitHub (https://github.com/brandonvietlam/NTermFinder).

RESULTS AND DISCUSSION

Development of an MS-Cleavable Reagent PFP-Rink-biotin for N-Termini Labeling. MS-cleavable reagents are increasingly gaining popularity in the search for low-abundant proteins and protein PTMs by producing unique fragment peaks such as signature ions or neutral losses upon tandem MS fragmentation, which further assist in the unambiguous identifications[11,21] and more precise quantifications[22] of these otherwise overshadowed targets. This unique feature of the MS-cleavable reagent is an attractive strategy for identifying

https://doi.org/10.1021/jasms.1c00350
J. Am. Soc. Mass Spectrom. 2022, 33, 189−197

the protein N-termini which are dramatically outnumbered by internal peptides. Thus, we designed to improve the confidence in characterizing protein N-termini by incorporating a CID-cleavable Rink amide moiety in our design of a new generation N-terminal probe targeting primary amines. Rink amide has been formerly utilized in solid-phase peptide synthesis,[23] solid-phase acid-cleavable N-terminal probe, and novel cross-linker developments.[15,24] *N*-Hydroxysuccinimide (NHS) esters were conventionally used as reactive groups toward primary amines. However, PFP esters were reported as being less subjective to hydrolysis and having improved chemoselectivity for reaction with amines compared to NHS esters.[25] Combined with the enrichment tag biotin, PFP-Rink-biotin reagent was constructed and synthesized as shown in Figure 1a and Scheme S1. Within the PFP-Rink-biotin reagent, the key feature for improving the confidence in N-terminus identification is the Rink group. Upon CID activation, the labile Rink group was preferentially cleaved under lower energy compared to the cleavage of peptide bonds, which readily gave rise to a high-abundant SI peak (m/z 569.19, theoretical 569.2434) from the labeled peptides (Figure 1a and Supplemental Figures 1 and 2).

This feature helped to alleviate the common issue of false-positive identification due to the binding and coelution of nonbiotinylated peptides along with the biotinylated peptides. In the experimental workflow, for selectively targeting N-terminal primary amines, at first, lysines side chain ε-primary amines were blocked by guanidination. The guanidination reaction was reported to be very selective to ε-amines by converting lysines to homoarginines[15] (Supplemental Figure 3a). It was also reported that homoarginine-modified lysines are sometimes cleaved by protease trypsin depending on the time, enzyme to protease ratio, and the nearby amino acid sequence.[15] Subsequent labeling of the samples with PFP-Rink-biotin will provide selective tagging of proteins N-termini. Further biotin−avidin affinity chromatography will selectively enrich the N-terminal peptides from the peptide mixtures. Within the same MS^2 spectrum, a complementary fragment that consists of the N-terminal peptide derivatized with the remnant of PFP-Rink-biotin after the loss of SI will be simultaneously generated in high intensity. This reagent-modified fragment can be further subjected to MS^3 data-dependent fragmentation to generate peptide-bond fragmentation for the sequencing of original or proteolytic protein N-termini facilitated by a specific bioinformatic strategy (Figure 1b).

Evaluation of PFP-Rink-biotin for Enrichment and Identification of Protein N-Termini. The MS-cleavable feature of PFP-Rink-biotin in identifying the biotin-tagged protein N-termini was investigated using three standard proteins: ubiquitin, β-lactoglobulin, and lysozyme (Figure 2 and Supplemental Figure 3b). For the model proteins we study, lysozyme[26] and β-lactoglobulin[27] were discovered containing signal peptides that were selectively removed during protein maturation. For the N-terminus characterization of lysozyme, GluC digestion is more favorable over tryptic digestion because of its N-terminal lysine residue. After the labeling reaction, quenching and removal of PFP-Rink-biotin were performed before the digestion of protein for achieving desired enrichment of the lysozyme N-terminal peptide. The removal of excess PFP-Rink-biotin was essential after the deactivation step due to competition for limited avidin binding sites, which can lead to low binding efficiency of the labeled peptides and adversely affect their subsequent enrichment. The reagent removal step was compared and evaluated by two different purification techniques: commercial protein concentrator molecular weight cutoff (MWCO) purification[28] and methanol−chloroform precipitation[29,30] of proteins. In the total ion chromatogram (TIC) of MWCO purification (Supplemental Figure 4a), two dominant peaks (arrowed) were observed and identified to be from the unreacted and hydrolyzed PFP-Rink-biotin reagent, whereas the lysozyme N-terminus could not be identified in the extracted ion chromatogram (XIC) and MS data (Supplemental Figure 5a,b). The same protein sample was alternatively precipitated by methanol−chloroform purification. In this TIC, the two dominant reagent peaks existed in the MWCO treated sample were significantly reduced by methanol−chloroform purification (Supplemental Figure 4b). In the XIC, the two lysozyme N-terminus precursors ($[M + 2H]^{2+}$ m/z 802.80 and $[M + 3H]^{3+}$ m/z 535.50) emerged (Supplemental Figure 5c,d) and were identified labeled in the tandem MS analysis (see Supplemental Table 1 for the detailed calculations). This experiment clearly showed methanol−chloroform purification is more advantageous over the commercial MWCO purification in removing the excess unreacted reagent to help enriching the protein, especially considering the widespread applications of methanol−chloroform in the purification of proteins from large-scale biological samples.[30]

The labeling of lysozyme was confirmed by manually matching MS^2 spectra against theoretical fragments of the lysozyme N-terminal sequence K_GVFGRC_CE (C_C represents carbamidomethylated cysteine and K_G is guanidinated lysine) (Figure 2a). In the MS^2 spectrum of the 802.97 precursor, three dominant fragments were generated, one of which was the SI (m/z 569.22). The other fragment (m/z 1036.42) was identified to be the complementary fragment to the signature ion, which was confirmed to match the m/z of [M-SI] in reduced charge state. The MS^2 spectrum of 536.18 also generated highly intense complementary fragment of m/z 518.88 and the SI peak. Further MS^3 spectrum of this complementary fragment was manually annotated against the N-terminal sequence of lysozyme, which exhibited confident coverage of the sequence confirmation. Note that mass addition of remnant fragment to the peptide was calculated for all the complementary fragments and the resultant fragments in the MS^3 spectrum (Supplemental Table 1).

After the evaluation of the lysozyme N-terminus, the tandem MS fragmentation study was also performed with ubiquitin and β-lactoglobulin N-termini. MS^2 and MS^3 spectra of the enriched ubiquitin protein N-terminal peptide precursors were extracted (Figure 2b). Ubiquitin N-terminus $MQIFVK_G$ modified by PFP-Rink-biotin was generated upon the cleavage after the first guanidinated lysine (K_G), with the m/z of 737.79 ($[M + 2H]^{2+}$) in the full MS spectrum. The CID spectrum of the $[M + 2H]^{2+}$ peptide precursor produced SI and the complementary fragment ion (m/z 906.54), both in high relative intensities. The complementary fragment ion was subjected to MS^3 fragmentation to identify the ubiquitin N-terminal sequence. This result matched with the previously Edman-degradation-determined complete amino acid sequence where the N-terminal methionine is not removed in the mature form of ubiquitin.[31] Similarly, identification of the β-lactoglobulin N-terminal sequence after tryptic digestion was achieved by labeling with PFP-Rink-biotin reagent and MS analysis with multistage activation (Figure 2c).

https://doi.org/10.1021/jasms.1c00350
J. Am. Soc. Mass Spectrom. 2022, 33, 189−197

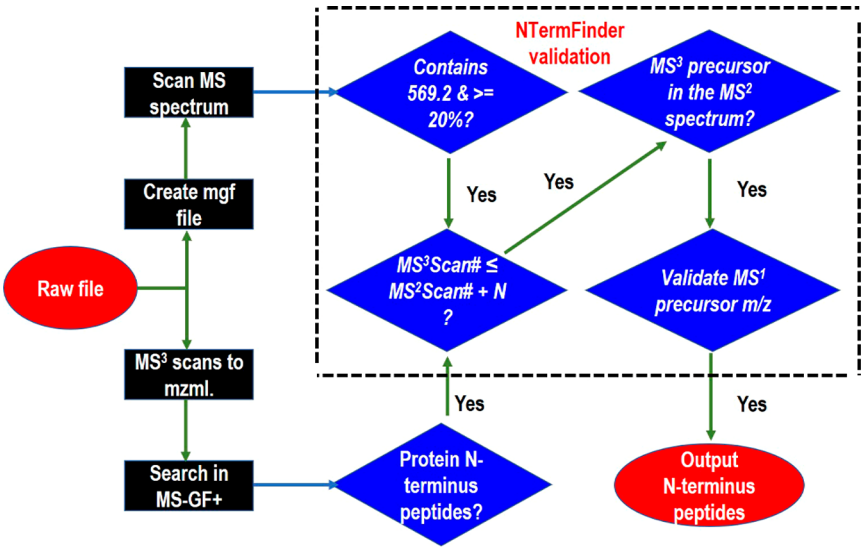


Figure 3. Bioinformatic approach for the validation of PFP-Rink-biotin labeled protein N-termini. MS^2 spectra were converted to MGF format and MS^3 scans were converted to mzML format and searched in MS-GF+ to identify N-terminal peptides. During the validation of the identified peptides, each MS^2 spectrum containing m/z 569.2 ± 1.0 with higher than 20% relative intensity was matched with the corresponding MS^3 scan of the identified peptide for validating the scan number and ensuring the MS^3 precursor ion is in the MS^2 spectrum, followed by the validation of precursor in the MS^1 to be within ±1 Da of the calculated mass of (Precursor × Charge + 569.2 + 1.0078)/(Charge + 1).

To summarize, the MS analysis of the labeled N-termini of standard proteins showed that the SI can be consistently and predominantly produced across different modified N-terminal peptides, usually one of the highest fragments observed, exhibiting low energy pathway of cleavage than most of peptide bond fragmentations. This SI was confirmed to be suitable as a common feature from PFP-Rink-biotin labeling for identifying the N-terminal peptides. Complementary fragment coexisted as one of the most intense fragments among the CID-MS^2 spectra to be paired with the SI as identifiers of the derivatized N-terminal peptides. Furthermore, this complementary fragment contained the complete N-terminal sequence, which can be further subjected to multistage data-dependent MS^3 fragmentation for peptide sequencing. The enrichment strategy was confirmed to be efficient after optimizations and proceeded for large-scale N-termini enrichment and analysis.

Bioinformatic Strategy for the Identification of Large-Scale Proteolytic Products. Strategy for protein N-termini identification in large-scale by PFP-Rink-biotin was implemented as shown in Figure 1b. To facilitate the bioinformatic analysis and automatic validation in large-scale N-termini identification, an in-house Java script NTermFinder was developed and incorporated to streamline the filtering and selection of N-terminal peptides from our MS analysis (Figure 3). Raw MS files of enriched samples were converted to MGF (Mascot Generic File) format which were subsequently imported into NTermFinder to select spectra containing m/z of SI (569.24 ± 1 Da) with at least 20% relative abundance within the spectra. The scan numbers of these spectra were denoted as $MS^2Scan\#$. On the other hand, the MS^3 spectra from the same file were converted to mzML format and then searched in MS-GF+,[19,20] where the protein N-terminal peptides were identified, and the corresponding scan numbers were reported as $MS^3Scan\#$. Furthermore, a three-step validation of the identified N-terminal PSMs from MS-GF+ was performed in the NTermFinder: (1) Scan number validation: scan numbers of MS^2 and MS^3 spectra have to match $MS^3Scan\# \leq MS^2Scan\# + N$, where N is the number of precursors of each MS^2 scan subjected to MS^3 fragmentation in the LCMS method; (2) the precursor of each MS^3 N-terminal PSM should be present as one of the fragments in the corresponding MS^2 spectrum; (3) and the MS^1 precursor of the identified MS^3 PSM should be within ±1 Da of the theoretical result calculated by (Precursor × charge + 569.2 + 1.0078) /(charge + 1), where Precursor is the MS^3 precursor and Charge is the charge state of the identified peptide. Both Precursor and Charge were extracted from the MS-GF+ output results. The rationale behind this validation step is that the complementary fragments were always found to be one charge less than the respective charge states of their precursors in the full MS spectra (as confirmed in the analysis of standard proteins N-termini), due to the protonation of biotin group within the SI in the same spectrum. This reduce-charge effect was incorporated as a criterion in selecting the positive results of protein N-termini from potential interferences. The final results after the validation were reported as confident protein N-termini identifications.

Application of PFP-Rink-biotin to *E. coli* Cell Lysate. a proof-of-concept large-scale N-terminome study was carried out for the N-termini enrichment of *E. coli* cell lysate. During this study, the binding of labeled N-terminal peptides with streptavidin and elution by HFIP was found to be more efficient compared to binding with monomeric avidin/elution by ACN.[18] In the LC−MS/MS analysis, two LC gradients with high ACN elution were performed to account for the increased hydrophobicity due to the attachment of a bulky Rink-biotin group to the N-terminal peptides. In addition, to reduce the selection of background noises for the fragmentation in MS^2, the minimum signal threshold for MS/MS fragmentation was increased to 10000 (see the Experimental Section for details). The resultant MS data files were searched according to the bioinformatic strategy above, and the total number of N-termini identified from the sample runs was 42 after the in-house bioinformatic approach along with manual removal of repetitive sequences and validation of the modifications identified (Table 1). The number of N-termini identifications

https://doi.org/10.1021/jasms.1c00350
J. Am. Soc. Mass Spectrom. 2022, 33, 189−197

Table 1. Forty-Two Unique Protein N-Termini Identified from *E. coli* Cells

no.	N-terminal peptide[a]	protein name
1	mARYFRRRK$_G$FC$_c$R	30S ribosomal protein S18
2	mELK$_G$K$_G$LmGHISIIPDYR	H repeat-associated putative transposase YdcC
3	mFK$_G$RRYVTLLPLFVLLAAC$_c$SSK$_G$	membrane-bound lytic murein transglycosylase B
4	mK$_G$DK$_G$VYK$_G$	dihydroneopterin triphosphate diphosphatase
5	mNTEATHDQNEALTTGARLR	cytoskeleton protein RodZ
6	mRVNLLITmIIFALIWPVTALR	fimbria adhesin EcpD
7	mSGFFQRLFGK$_G$	uncharacterized protein YjfK
8	MAK$_G$APIRARK$_G$RVRK$_G$	30S ribosomal protein S11
9	MENFK$_G$HLPEPFR	tryptophanase
10	METTQTSTIASK$_G$DSR	serine transporter
11	MHPMLNIAVR	inositol-1-monophosphatase
12	MNDSEFHR	iron−sulfur cluster assembly protein CyaY
13	MNFEGK$_G$IALVTGASR	3-oxoacyl-[acyl-carrier-protein] reductase FabG
14	MNLHEYQAK$_G$	succinate-CoA ligase [ADP-forming] subunit beta
15	MQGSVTEFLK$_G$PR	DNA-directed RNA polymerase subunit alpha
16	MQLNSTEISELIK$_G$QR	ATP synthase subunit alpha
17	MSIVVK$_G$	anaerobic nitric oxide reductase flavorubredoxin
18	MTDMNILDLFLK$_G$	Tol-Pal system protein TolQ
19	MVSNASALGR	succinate dehydrogenase hydrophobic membrane anchor subunit
20	MYVVSTK$_G$	D-tagatose-1,6-bisphosphate aldolase subunit GatY
21	GFTTR	D-allose transport system permease protein AlsC
22	GIFSR	phage shock protein A
23	HSLQR	formyltetrahydrofolate deformylase
24	SK$_G$IVK$_G$	enolase
25	AAK$_G$DVK$_G$	60 kDa chaperonin
26	AK$_G$GQSLQDPFLNALR	RNA-binding protein Hfq
27	ALNLQDK$_G$QAIVAEVSEVAK$_G$	50S ribosomal protein L10
28	AQQTPLYEQHTLC$_c$GAR	aminomethyltransferase
29	ASENMTPQDYIGHHLNNLQLDLR	ATP synthase subunit a
30	ATPHINAEMGDFADVVLMPGDPLR	purine nucleoside phosphorylase DeoD-type
31	ATVSMR	30S ribosomal protein S2
32	C$_c$GIVGAIAQR	glutamine-fructose-6-phosphate aminotransferase [isomerizing]
33	LYIDK$_G$ATILK$_G$FDLEmLK$_G$K$_G$	protein HdeD
34	mVQHLK$_G$RRPLSRYLK$_G$DFK$_G$	regulator of sigma S factor FliZ
35	SK$_G$EHTTEHLR	uncharacterized protein YqjD
36	SK$_G$IFEDNSLTIGHTPLVR	cysteine synthase A
37	SLINTK$_G$	alkyl hydroperoxide reductase C
38	SLSTEATAK$_G$	30S ribosomal protein S15
39	STEIK$_G$TQVVVLGAGPAGYSAAFR	dihydrolipoyl dehydrogenase
40	TDK$_G$LTSLR	transaldolase B
41	TK$_G$PYVR	probable hydrolase YcaC
42	TMNITSK$_G$QMEITPAIR	ribosome-associated inhibitor A

[a]All N-termini identified were labeled by PFP-biotin with a 99.03 mass addition, 'm' represents oxidized methionine, "K$_G$" represents guanidinated lysine, and "C$_c$" represents carbamidomethylated cysteine.

appears to be lower when compared to the previous studies in *E. coli*,[6,32] mainly due to the lower sensitivity in the MS^3 acquisition and additional time required to extract both signature ion (MS^2 level) and sequence information (MS^3 level) from the modified N-termini. A recently reported strategy named chemical enrichment of protease substrates (CHOPS) adapted a similar labeling approach by synthesizing a biotinylated reagent for targeting low-abundant proteolytic substrates.[33] Nevertheless, our approach provides the advantage of incorporating an MS-cleavable functionality for achieving higher confidence and reproducibility. However, we achieved these goals using a low-resolution mass analyzer like ion trap. Figure 4 showed the MS^n spectra identified to be

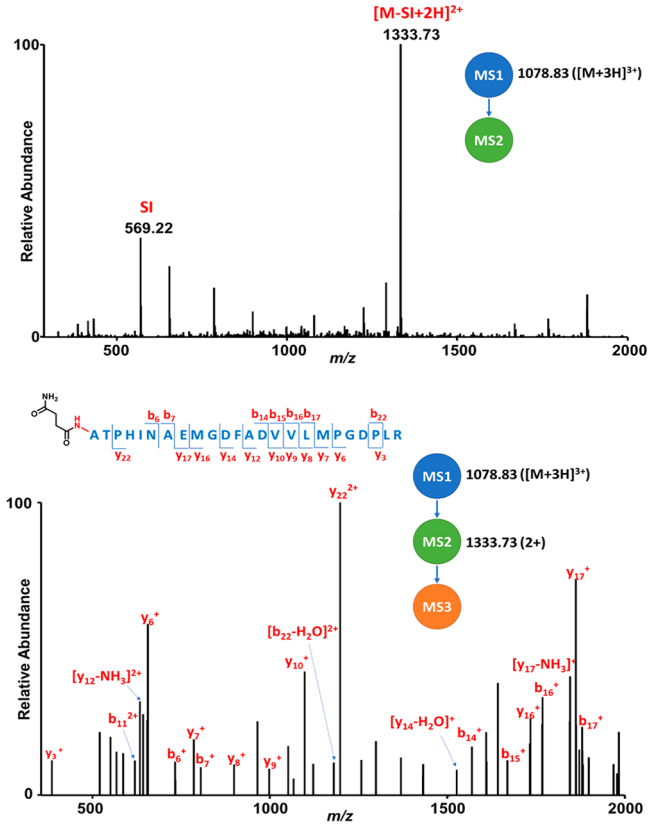


Figure 4. Identification of labeled *E. coli* purine nucleoside phosphorylase DeoD-type protein N-terminus by MS^n analysis. The generation of the high-abundance signature ion and complementary fragment was consistent with the high-confidence identification of the N-terminal peptide sequence by MS^3 fragmentation.

purine nucleoside phosphorylase N-terminus from *E. coli*, confidently confirming the labeling and N-terminal sequence. The predominant proteolytic event in *E. coli* is the excision of Met by methionine aminopeptidase (MetAP) after the deformylation of the N-termini.[34] In this catalytic event, the Met being removed is at the P1 position, whereas the new N-terminal amino acid (penultimate position in the original sequence) is the P1′ position according to Schechter and Berger nomenclature.[35] The efficiency of Met removal is mostly dependent on the size of the side chain in the P1′ amino acid based on the *in vitro* study with synthetic peptides. Amino acids with a larger size chain generally have lower MetAP catalytic efficiency; therefore, protein N-termini with a small amino acid residue as P1′ such as Ala, Gly, Pro, Ser, and

https://doi.org/10.1021/jasms.1c00350
J. Am. Soc. Mass Spectrom. 2022, 33, 189−197

Cys have nearly complete Met excision, whereas the cleavage efficiency of penultimate Val and Thr could depend on the P2′−P4′ amino acids and lower cleavage efficiency was observed for Phe, Arg, Tyr, Trp.[36] Overall, the result of identified *E. coli* N-termini from the enrichment experiment matched well with the expectation of this size-dependent proteolytic event, except for two Met intact peptides with P1′ Ala potentially due to partial Met removal (Table 1). Nearly half of all N-terminal Met identified were cleaved in our experiment, which is consistent with a previously reported control study.[37] Though the total number of identifications is comparably lower, this powerful approach has the ability to identify proteolytic products *in vivo*.

CONCLUSION

In this study, we explored the potentials of implementing an MS-cleavable signature-ion based strategy for the identifications of proteolytic N-termini after enrichment. The design of this new gas-phase MS-cleavable labeling reagent PFP-Rink-biotin inherently provides a fixed signature SI generated in high intensity for the confirmation of the labeled peptides, with the sequence information being extracted and analyzed by multistage MS fragmentation. Further validation of the N-termini authenticity was facilitated by the correlations among precursors/product ions from MS^1 to MS^3 levels and automated by our custom high-throughput bioinformatic approach catering to the resulting files of MS-GF+ searches. Tracking a proteolytic product in native *in vivo* experiments is extremely challenging due to the presence of a low number of new N-termini and high sample complexity. Enriching regular N-terminal peptides had some success in this area using reverse purification strategy. We demonstrated an innovative workflow using a MS-cleavable reagent for the first time in this research area. Together with other methods in this area, we believe this method can be routinely applicable to decode N-terminal heterogeneity of antibody-based drugs, as well as pinpointing and profiling system-level protease activities. We will not claim that this is fully optimized yet, but large-scale results in *E. coli* clearly demonstrated the power of this approach. Although we utilized common practice, such as guanidination to block lysines, further improvement in this area will reduce false positive identifications significantly. Large-scale application of this strategy is under development with the ongoing focus being on the optimization of enrichment and LC−MS acquisition methods for identifying new proteolytic substrates to improve our understanding of proteolytic mechanisms in normal and diseased biological systems.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jasms.1c00350.

Experimental procedures for enriching and identifying N-termini of standard proteins and *E. coli* cells; NMR characterizations of PFP-Rink-biotin and its synthetic intermediates; MS characterization of PFP-Rink-biotin and the mechanism of generating signature fragment; guanidination reaction and the N-termini sequences of standard proteins; optimization of protein purification; mass calculation examples of identified protein N-termini; link and user manual to download and use NTermFinder (PDF)

AUTHOR INFORMATION

Corresponding Author

Saiful M. Chowdhury − *Department of Chemistry and Biochemistry, University of Texas at Arlington, Arlington, Texas 76019, United States;* orcid.org/0000-0002-2553-985X; Phone: +1-817-272-5439; Email: schowd@uta.edu

Authors

Zixiang Fang − *Department of Chemistry and Biochemistry, University of Texas at Arlington, Arlington, Texas 76019, United States;* Present Address: Washington University School of Medicine, St. Louis, MO 63110

Maheshika S. K. Wanigasekara − *Department of Chemistry and Biochemistry, University of Texas at Arlington, Arlington, Texas 76019, United States;* Present Address: University of Texas Southwestern Medical Center, Dallas, TX 75390.

Akop Yepremyan − *Department of Chemistry and Biochemistry, University of Texas at Arlington, Arlington, Texas 76019, United States;* Present Address: McMaster University, Hamilton, ON L8S 4L8, Canada.

Brandon Lam − *Department of Chemistry and Biochemistry, University of Texas at Arlington, Arlington, Texas 76019, United States*

Pawan Thapa − *Department of Chemistry and Biochemistry, University of Texas at Arlington, Arlington, Texas 76019, United States;* Present Address: University of Texas Southwestern Medical Center, Dallas, TX 75390.

Frank W. Foss, Jr. − *Department of Chemistry and Biochemistry, University of Texas at Arlington, Arlington, Texas 76019, United States;* orcid.org/0000-0003-1940-6580

Complete contact information is available at: https://pubs.acs.org/10.1021/jasms.1c00350

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Partial support from grants IOS 1655735 NSF and UA5GM113216-01, NIGMS, NIH is acknowledged. We acknowledge support from UT Arlington Chemistry and Biochemistry Department. We also acknowledge mass spectrometry support from the Shimadzu Center for Advanced Analytical Chemistry at UTA. NMR experimentation was made possible by partial support from the National Science Foundation under Award CHE-0840509.

REFERENCES

(1) Doucet, A.; Butler, G. S.; Rodríguez, D.; Prudova, A.; Overall, C. M. Metadegradomics: Toward in Vivo Quantitative Degradomics of Proteolytic Post-Translational Modifications of the Cancer Proteome. *Mol. Cell. Proteomics* **2008**, *7* (10), 1925−1951.

(2) López-Otín, C.; Overall, C. M. Protease Degradomics: A New Challenge for Proteomics. *Nat. Rev. Mol. Cell Biol.* **2002**, *3* (7), 509−519.

(3) Salvesen, G. S.; Dixit, V. M. Caspases: Intracellular Signaling by Proteolysis. *Cell* **1997**, *91* (4), 443−446.

(4) Prudova, A.; auf dem Keller, U.; Butler, G. S.; Overall, C. M. Multiplex N-Terminome Analysis of MMP-2 and MMP-9 Substrate Degradomes by ITRAQ-TAILS Quantitative Proteomics. *Mol. Cell. Proteomics* **2010**, *9* (5), 894−911.

(5) Gevaert, K.; Goethals, M.; Martens, L.; Van Damme, J.; Staes, A.; Thomas, G. R.; Vandekerckhove, J. Exploring Proteomes and

Analyzing Protein Processing by Mass Spectrometric Identification of Sorted N-Terminal Peptides. *Nat. Biotechnol.* **2003**, *21* (5), 566−569.
(6) Timmer, J. C.; Enoksson, M.; Wildfang, E.; Zhu, W.; Igarashi, Y.; Denault, J.-B.; Ma, Y.; Dummitt, B.; Chang, Y.-H.; Mast, A. E.; Eroshkin, A.; Smith, J. W.; Tao, W. A.; Salvesen, G. S. Profiling Constitutive Proteolytic Events in Vivo. *Biochem. J.* **2007**, *407* (1), 41−48.
(7) Mahrus, S.; Trinidad, J. C.; Barkan, D. T.; Sali, A.; Burlingame, A. L.; Wells, J. A. Global Sequencing of Proteolytic Cleavage Sites in Apoptosis by Specific Labeling of Protein N Termini. *Cell* **2008**, *134* (5), 866−876.
(8) Kleifeld, O.; Doucet, A.; auf dem Keller, U.; Prudova, A.; Schilling, O.; Kainthan, R. K.; Starr, A. E.; Foster, L. J.; Kizhakkedathu, J. N.; Overall, C. M. Isotopic Labeling of Terminal Amines in Complex Samples Identifies Protein N-Termini and Protease Cleavage Products. *Nat. Biotechnol.* **2010**, *28* (3), 281−288.
(9) Yeom, J.; Ju, S.; Choi, Y.; Paek, E.; Lee, C. Comprehensive Analysis of Human Protein N-Termini Enables Assessment of Various Protein Forms. *Sci. Rep.* **2017**, *7* (1), 6599.
(10) Chen, L.; Shan, Y.; Weng, Y.; Sui, Z.; Zhang, X.; Liang, Z.; Zhang, L.; Zhang, Y. Hydrophobic Tagging-Assisted N-Termini Enrichment for In-Depth N-Terminome Analysis. *Anal. Chem.* **2016**, *88* (17), 8390−8395.
(11) Bhawal, R. P.; Sadananda, S. C.; Bugarin, A.; Laposa, B.; Chowdhury, S. M. Mass Spectrometry Cleavable Strategy for Identification and Differentiation of Prenylated Peptides. *Anal. Chem.* **2015**, *87* (4), 2178−2186.
(12) Chakrabarty, J. K.; Naik, A. G.; Fessler, M. B.; Munske, G. R.; Chowdhury, S. M. Differential Tandem Mass Spectrometry-Based Cross-Linker: A New Approach for High Confidence in Identifying Protein Cross-Linking. *Anal. Chem.* **2016**, *88* (20), 10215−10222.
(13) Chakrabarty, J. K.; Sadananda, S. C.; Bhat, A.; Naik, A. J.; Ostwal, D. V.; Chowdhury, S. M. High Confidence Identification of Cross-Linked Peptides by an Enrichment-Based Dual Cleavable Cross-Linking Technology and Data Analysis Tool Cleave-XL. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (2), 173−182.
(14) Kamal, A. H. M.; Aloor, J. J.; Fessler, M. B.; Chowdhury, S. M. Cross-Linking Proteomics Indicates Effects of Simvastatin on the TLR2 Interactome and Reveals ACTR1A as a Novel Regulator of the TLR2 Signal Cascade. *Mol. Cell. Proteomics* **2019**, *18* (9), 1732−1744.
(15) Chowdhury, S. M.; Munske, G. R.; Yang, J.; Zhukova, D.; Nguyen, H.; Bruce, J. E. Solid-Phase N-Terminal Peptide Enrichment Study by Optimizing Trypsin Proteolysis on Homoarginine-Modified Proteins by Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2014**, *28* (6), 635−644.
(16) Yepremyan, A. Syntheses of Novel Peptide Linkers. M.S. Thesis, The University of Texas at Arlington, Arlington, TX, 2015.
(17) Kim, J.-S.; Dai, Z.; Aryal, U. K.; Moore, R. J.; Camp, D. G.; Baker, S. E.; Smith, R. D.; Qian, W.-J. Resin-Assisted Enrichment of N-Terminal Peptides for Characterizing Proteolytic Processing. *Anal. Chem.* **2013**, *85* (14), 6826−6832.
(18) Warnken, U.; Schnölzer, M.; Linder, E. Methods for a quantitative release of biotinylated peptides and proteins from streptavidin complexes. WO 2016120247 A1, 2016-08-04.
(19) Kim, S.; Gupta, N.; Pevzner, P. A. Spectral Probabilities and Generating Functions of Tandem Mass Spectra: A Strike against Decoy Databases. *J. Proteome Res.* **2008**, *7* (8), 3354−3363.
(20) Kim, S.; Pevzner, P. A. MS-GF+ Makes Progress towards a Universal Database Search Tool for Proteomics. *Nat. Commun.* **2014**, *5* (1), 5277.
(21) Bhawal, R. P.; Shahinuzzaman, A. D. A.; Chowdhury, S. M. Gas-Phase Fragmentation Behavior of Oxidized Prenyl Peptides by CID and ETD Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (4), 704−707.
(22) Stadlmeier, M.; Bogena, J.; Wallner, M.; Wühr, M.; Carell, T. A Sulfoxide-Based Isobaric Labelling Reagent for Accurate Quantitative Mass Spectrometry. *Angew. Chem., Int. Ed.* **2018**, *57* (11), 2958−2962.
(23) Rink, H. Solid-Phase Synthesis of Protected Peptide Fragments Using a Trialkoxy-Diphenyl-Methylester Resin. *Tetrahedron Lett.* **1987**, *28* (33), 3787−3790.
(24) Chowdhury, S. M.; Munske, G. R.; Tang, X.; Bruce, J. E. Collisionally Activated Dissociation and Electron Capture Dissociation of Several Mass Spectrometry-Identifiable Chemical Cross-Linkers. *Anal. Chem.* **2006**, *78* (24), 8183−8193.
(25) Wang, J.; Zhang, R.-Y.; Wang, Y.-C.; Chen, X.-Z.; Yin, X.-G.; Du, J.-J.; Lei, Z.; Xin, L.-M.; Gao, X.-F.; Liu, Z.; Guo, J. Polyfluorophenyl Ester-Terminated Homobifunctional Cross-Linkers for Protein Conjugation. *Synlett* **2017**, *28* (15), 1934−1938.
(26) Canfield, R. E. The Amino Acid Sequence of Egg White Lysozyme. *J. Biol. Chem.* **1963**, *238*, 2698−2707.
(27) Braunitzer, G.; Chen, R.; Schrank, B.; Stangl, A. [The Sequence of Beta-Lactoglobulin (Author's Transl)]. *Hoppe-Seyler's Z. Physiol. Chem.* **1973**, *354* (8), 867−878.
(28) Fang, Z.; Baghdady, Y. Z.; Schug, K. A.; Chowdhury, S. M. Evaluation of Different Stationary Phases in the Separation of Inter-Cross-Linked Peptides. *J. Proteome Res.* **2019**, *18* (4), 1916−1925.
(29) Wessel, D.; Flügge, U. I. A Method for the Quantitative Recovery of Protein in Dilute Solution in the Presence of Detergents and Lipids. *Anal. Biochem.* **1984**, *138* (1), 141−143.
(30) Shahinuzzaman, A. D. A.; Chakrabarty, J. K.; Fang, Z.; Smith, D.; Kamal, A. H. M.; Chowdhury, S. M. Improved In-solution Trypsin Digestion Method for Methanol-Chloroform Precipitated Cellular Proteomics Sample. *J. Sep. Sci.* **2020**, *43*, 2125.
(31) Schlesinger, D. H.; Goldstein, G.; Niall, H. D. Complete Amino Acid Sequence of Ubiquitin, an Adenylate Cyclase Stimulating Polypeptide Probably Universal in Living Cells. *Biochemistry* **1975**, *14* (10), 2214−2218.
(32) Berry, I. J.; Steele, J. R.; Padula, M. P.; Djordjevic, S. P. The Application of Terminomics for the Identification of Protein Start Sites and Proteoforms in Bacteria. *Proteomics* **2016**, *16* (2), 257−272.
(33) Griswold, A. R.; Cifani, P.; Rao, S. D.; Axelrod, A. J.; Miele, M. M.; Hendrickson, R. C.; Kentsis, A.; Bachovchin, D. A. A Chemical Strategy for Protease Substrate Profiling. *Cell Chem. Biol.* **2019**, *26* (6), 901−907.
(34) Solbiati, J.; Chapman-Smith, A.; Miller, J. L.; Miller, C. G.; Cronan, J. E. Processing of the N Termini of Nascent Polypeptide Chains Requires Deformylation Prior to Methionine Removal. *J. Mol. Biol.* **1999**, *290* (3), 607−614.
(35) Schechter, I.; Berger, A. On the Size of the Active Site in Proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27* (2), 157−162.
(36) Hirel, P. H.; Schmitter, J. M.; Dessen, P.; Fayat, G.; Blanquet, S. Extent of N-Terminal Methionine Excision from Escherichia Coli Proteins Is Governed by the Side-Chain Length of the Penultimate Amino Acid. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86* (21), 8247−8251.
(37) Bienvenut, W. V.; Giglione, C.; Meinnel, T. Proteome-Wide Analysis of the Amino Terminal Status of Escherichia Coli Proteins at the Steady-State and upon Deformylation Inhibition. *Proteomics* **2015**, *15* (14), 2503−2518.

https://doi.org/10.1021/jasms.1c00350
J. Am. Soc. Mass Spectrom. 2022, 33, 189−197