A mass spectrometry-cleavable N-terminal tagging strategy for system-level protease activity profiling

Zixiang Fang[1], Maheshika S.K Wanigasekara[1,2], Akop Yepremyan[1,3], Brandon Lam[1], Pawan Thapa[1,2], Frank W. Foss Jr[1], Saiful M. Chowdhury[1]*

1. Department of chemistry and biochemistry, University of Texas at Arlington, Arlington Texas 76019
2. Current address: University of Texas Southwestern Medical Center, Dallas, TX 75390
3. Current Address: McMaster University, ON L8S 4L8, Canada

*Corresponding Author
Saiful M. Chowdhury, PhD
Associate Professor
Chemistry and Biochemistry
University of Texas at Arlington
Arlington, TX 76019
USA
Tel: +1-817-272-5439
Email: schowd@uta.edu

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. This irreversible PTM is regulated and catalyzed by proteases through hydrolysis reaction in the process of protein maturation or degradation. The identification of proteolytic substrates is pivotal in understanding the specificity of proteases and the physiological role of proteolytic process. However, tracking these biological phenomena in native cells is very difficult due to their low abundances. Currently no efficient methods are available to identify proteolytic products from large-scale samples. To facilitate the targeted identification of low-abundant proteolytic products, we devise a strategy incorporating a novel biotinylated reagent PFP (pentafluorophenyl)-Rink-biotin to specifically target, enrich and identify proteolytic N-terminus. Within the PFP-Rink-biotin reagent, a MS-cleavable feature is designed to assist in the unambiguous confirmation of the enriched proteolytic N-termini. This is the first reported study of identifying proteolytic N-terminus by MS-cleavable feature widely adopted in studying low-abundant protein PTMs and cross-linking/MS. The proof-of-concept study was performed with multiple standard proteins whose N-terminus 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 along with a bioinformatics workflow. We believe this method can be very useful to locate proteolytic products in native cellular environment with high confidence.

Introduction

In vivo protein biosynthesis follows a stepwise fashion starting from N-terminal end towards the C-terminal end, usually with the N-terminal amino acid being a methionine (Met)[1]. Among the earliest modifications during or following the manufacturing of most proteins are the removal of initiating Met residue and the acetylation of N-terminal α-amino group, both of which are crucial in protein structure, stability, function and localization[1–3]. As the two most ubiquitous protein N-terminal modifications, around 80% of soluble proteins in human have acetylated N-terminus[4] while all the proteins undergo proteolysis during their maturation or degradation processes[5]. 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[6]. Proteolysis used to be considered predominantly as protein-degrading and disposing procedures, 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[5]. 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[7]. Proteolysis is also intimately involved in significant biological implications and pathological conditions of various cancers, chronic inflammations, neurodegenerative and cardiovascular diseases. It is estimated that 5-10% of all drug targets were designed for anti-proteolytic therapies[8–10]. Because proteolysis occurs co-translationally or post-translationally, the functional N-terminus of proteins cannot be directly inferred from 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[5,6,11].

Each protein has only one protein N-terminal sequence while numerous internal peptides from enzymatic digestion are generated in a regular 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[12]. This challenge will be more pronounced and overwhelming for identifying N-terminus and pertinent proteolytic products of low-abundant proteins in large-scale biological samples of increased complexity and dynamic range. Therefore, certain enrichment strategies of the original protein N-terminus are essential for their confident identifications by mass spectrometric analysis. Enrichment of protein N-terminus is conventionally achieved by either positive or negative selection strategies. Positive selection approaches specifically label the α-amine of the protein N-terminus with an enrichment group while minimizing or preventing the mislabeling of lysine 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)[13–18] or enzymatically-assisted modification (subtiligase has high specificity for N-terminal α-amine)[19,20]. After sample digestion, protein N-terminus can be selectively purified by targeting the enrichment group to be separated from the interfering internal peptides. On the contrary, negative selection approaches, notably the combined fractional diagonal chromatography (COFRADIC)[12,21] and terminal amine isotopic labeling of substrates (TAILS)[22,23], 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-terminus can be retained for LC-MS/MS analysis. Negative selection strategies are advantageous in studying protein N-terminus with native

modifications including N-terminal acetylation. However, enrichment by negative selection tends to result in higher sample complexity from the co-enrichment of free and modified protein N-terminus[24] and requires better enrichment efficiency for removing the more abundant internal peptides[25], which can be inferior in selectively studying proteolytically processed protein N-terminus compared to positive selection[6].

Ideally, a successful N-terminal positional proteomics experiment will only detect a single peptide from each protein. This can be problematic in the confident bioinformatic analysis of the respective protein because the lack of complementary confirmation of associated internal peptides from the same protein, which can in turn compromise the confidence of identifying protein N-terminus. The issue of false N-terminus discovery can be further exacerbated if the enrichment is inefficient or contains non-specific interactions, during the process of targeting the labeled N-terminus in positive selection approach or depleting the internal peptides in negative selection approach. Although these problems can be alleviated by vigorous validations and strict data filtering processes including the utilization of multiple search engines, narrower mass tolerance for identifications, repetitive confirmation of same peptide of different charge states in multiple replicates[22], these criteria may simultaneously filter out low-abundant protein N-terminus.

Here, we design a direct strategy to improve the confidence in the identification of proteolytic products by developing and utilizing a MS-cleavable reagent PFP (pentafluorophenyl)-Rink-biotin (**Figure 1a**) in positive enrichment for investigating proteolytic protein N-terminus of low abundance. The chemical reaction occurs as free protein N-terminus were targeted by PFP group and simultaneously biotinylated for downstream enrichment by removing non-labeled internal peptides. The confident N-terminus identification is enabled by the generation of a fixed

signature ion (SI) and 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 of the enriched protein N-terminus. This strategy incorporates the MS-cleavable feature widely applied in improving the identification capabilities and lowering false-discovery rate of low-abundant peptides in proteomics applications of protein post-translational modification (PTM) study[26] and cross-linking chemical biology[27]. To facilitate the automatic N-terminus confirmation, an in-house Java-based software NTermFinder was developed to validate the identified protein N-terminal peptides by the signature fragments and mass correlations among peptide precursor ion, complementary fragment and the scan numbers during MS acquisition. This strategy aims to achieve confident identification of proteolytic products and protein N-terminus in the native settings of the cells.

Results and Discussions

Development of MS-cleavable reagent PFP-Rink-biotin for N-terminus labeling

MS-cleavable reagents are increasingly gaining popularities in the research of low-abundant proteins and protein PTMs by producing unique signatures such as signature ion or neutral losses upon tandem MS fragmentation, which assist in the unambiguous identifications[26,28] and more precise quantifications[29] of these otherwise overshadowed targets. This unique feature of MS-cleavable reagent is an attractive strategy for identifying the protein N-terminus which are dramatically outnumbered by internal peptides. Hence, we are designing to improve the confidence in characterizing protein N-terminus by incorporating a CID-cleavable Rink moiety, which has been formerly utilized in solid phase peptide synthesis[30] and novel cross-linker

developments[31,32], in our design of new generation enrichment tag targeting primary amines. N-hydroxysuccinimide (NHS) esters are conventionally used as reactive group towards primary amines[33]. However, PFP esters were reported as less subjective to hydrolysis and having improved chemoselectivity for reaction with amines compared to NHS esters[34]. Combined with the enrichment tag biotin, PFP-Rink-biotin reagent is constructed and synthesized as shown in **Supplemental 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 is preferentially cleaved under lower energy compared to the cleavage of peptide bonds, which readily gives rise to a high-abundant SI peak (m/z 569) from the labeled peptides (**Supplemental Figure 1)**. This feature helps to alleviate the common issue of false-positive identification due to the binding and coelution of non-biotinylated peptides along with the biotinylated peptides. 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-terminus facilitated by specific bioinformatic strategy (**Figure 1b**).

Evaluation of PFP-Rink-biotin in the enrichment and identification of protein N-terminus

The MS-cleavable feature of PFP-Rink-biotin in identifying the biotin-tagged protein N-terminus was investigated using three standard proteins: ubiquitin, β-lactoglobulin and lysozyme. For the model proteins we study, lysozyme[35] and β-lactoglobulin[36] were discovered containing signal peptides that are selectively removed during protein maturation. According to the N-terminal sequences of these model proteins, trypsin (for digesting ubiquitin and β-lactoglobulin) and GluC

(for digesting lysozyme) were selected to obtain adequate lengths of the protein N-terminus for MS analysis against the known N-terminal sequence in the established protein database.

The enrichment of lysozyme N-terminus was first evaluated. 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 quenching of PFP reactive group deactivates its reactivity towards internal peptides to prevent the mislabeling and coelution of internal peptides during enrichment. The removal of excess PFP-Rink-biotin is as 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 affecting their subsequent enrichment. The reagent removal step was compared and evaluated by two different purification techniques: commercial protein concentrator molecular weight cut-off (MWCO) purification and methanol-chloroform precipitation of proteins. In the total ion chromatogram (TIC) of MWCO purification (**Supplemental Figure 2a**), two dominant peaks (arrowed) were observed and identified to be from the unreacted and hydrolyzed PFP-Rink-biotin reagent, whereas two N-terminal peptide precursors (M^{2+} m/z 802.80 and M^{3+} m/z 535.50) were not present in the extracted ion chromatogram (XIC) and the lysozyme N-terminus cannot be identified in the MS data (**Supplemental Figure 3a and 3b**). The same protein sample was alternatively precipitated by methanol-chloroform purification after reacting with the reagent, followed by the same digestion procedures. In this TIC, the two dominant reagent peaks existed in the MWCO treated sample were significantly reduced by methanol-chloroform purification (**Supplemental Figure 2b**). In the XIC, the two precursors (M^{2+} m/z 802.80 and M^{3+} m/z 535.50) emerged in identical retention time (**Supplemental**

Figure 3c and 3d) and identified to be the labeled lysozyme N-terminus by the tandem MS analysis (see **Supplemental Table 1** for the detailed calculations). This experiment clearly shows 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[37].

Two corresponding molecular ions 802.97 (M^{2+}) and 536.18 (M^{3+}) were confirmed by matching against theoretical fragments of the lysozyme N-terminal sequence K_GVFGRC_CE (C_C represents carbamidomethylated cysteine and K_G is guanidinated lysine) by tandem MS analysis (**Figure 2a**). In the MS^2 spectrum of 802.97 (M^{2+}) precursor, three dominant fragments were generated, one of which was the SI (m/z 569.22). The other fragment (labeled as $[M\text{-}SI+H]^+$ in the MS^2 spectrum) was identified to be the complementary fragment (m/z 1036.42) to the signature ion, which was confirmed to match the m/z of [M-SI] in reduced charge state. The lowest intensity of the three being the neutral loss of ammonia from the complementary fragment. The MS^2 spectrum of 536.18 (M^{3+}) also generated highly intense complimentary fragment of 518.88 and the SI peak. While the relative intensity of SI was around 25% relative intensity of the base peak, which was lower compared to around 90% relative intensity from M^{2+} of the same peptide upon fragmentation, showing charge-dependent fragmentation pattern in producing the SI. Further MS^3 spectrum of the complimentary 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. Although different relative intensities of SI and complimentary fragments were observed, this study shows that the same

peptide with different charge states helps to further confirm the protein N-terminus. In this particular case, the higher abundance of SI from M^{2+} helps to identify the enriched peptide utilizing the signature feature of PFP-Rink-biotin, while the higher abundance of complimentary fragment from M^{3+} results in better sequence annotation, both of which are complemented and contributed to the identification of lysozyme N-terminus.

After the optimization and evaluation of lysozyme N-terminus, the tandem MS fragmentation study was also performed with ubiquitin and β-lactoglobulin N-terminus. Although the guanidination step modified lysine to homoarginine within the proteins, tryptic digestion was previously proven to be effective in cleaving the C-terminal of homoarginine[38]. 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^{2+}) in the full MS spectrum. The CID spectrum of M^{2+} peptide precursor produced SI and complimentary fragment ion (m/z 906.54), both in high relative intensities. The complimentary fragment ion was subjected to MS^3 fragmentation to identify the ubiquitin N-terminus 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[39]. 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**).

To summarize, the MS analysis of the labeled N-terminus of standard proteins shows 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 is confirmed to be suitable as a common feature

from PFP-Rink-biotin labeling for identifying the N-terminal peptides. Complementary fragment coexists 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 complimentary fragment contains 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-terminus enrichment and analysis.

Bioinformatic strategy for the application of PFP-Rink-biotin in large-scale N-terminus identification

Strategy for protein N-terminus identification in large-scale by PFP-Rink-biotin was implemented as shown in **Figure 1b**. The proteins from cell lysis were first subjected to denaturation and guanidination for blocking the ε-amines of lysine. Quantitative labeling of lysine guanidination (> 99%) was previously achieved[18] and was adapted in our experiment. The N-terminus that contain free α-amines were then labeled by PFP-Rink-biotin reagent. After the labeling of protein N-terminus, excess reagent was quenched and removed by methanol-chloroform precipitation before enzymatic digestion to prevent mislabeling of internal peptides. The biotinylated N-terminal peptides can be selectively targeted and isolated by biotin-avidin affinity enrichment, and the sample complexity will be significantly reduced before subsequent tandem MS and bioinformatic analysis.

To facilitate the bioinformatic analysis and automatic validation in large-scale N-terminus 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 (569.2 ± 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 Raw file was converted to mzML format and then searched in MS-GF+, where the protein N-terminal peptides were identified, and the corresponding scan numbers were reported as $MS^3Scan\#$. Furthermore, a 3-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 $MS^3Scan\# \leq MS^2Scan\# + N$, N is the number of precursors of each MS^2 scan subjected to MS^3 fragmentation in the LCMS method; (2) Precursor of each MS^3 N-terminal PSM should be present as one of the fragments in the corresponding MS^2 spectrum; (3) 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-terminus), due to the protonation of biotin group within the SI in the same spectrum. This reduce-charge effect is incorporated as a criterion in selecting the positive results of protein N-terminus from potential interferences. The final results after the validation were reported as confident protein N-terminus identifications.

Application of PFP-Rink-biotin to *E. coli* cell lysate

Proof-of-concept large-scale N-terminome study was carried out for the N-terminus enrichment of *E. coli* cell lysate. During this study, the binding of labeled N-terminus peptides with

streptavidin and elution by HFIP was found to be more efficient compared to binding with monomeric avidin and elution by ACN[40]. 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 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 method section for details). The resultant MS data files were searched according to the bioinformatic strategy above and the total number of N-terminus 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**). **Figure 4** shows the MS^n spectra identified to be purine nucleoside phosphorylase N-terminus from *E. coli* confidently confirm 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-terminus[41]. In this catalytic event, the Met being removed is at 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[42]. The efficiency of Met removal is mostly depending on the size of side-chain in P1' amino acid based on the *in vitro* study with synthetic peptides: amino acid of larger size-chain generally has lower MetAP catalytic efficiency, therefore protein N-terminus with small amino acid residue as P1' such as Ala, Gly, Pro, Ser and Cys have near complete Met excision, whereas 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[43,44]. The identified N-terminal sequence motifs[45] were analyzed and summarized in **Figure 5**. Overall, the result of identified *E. coli* N-terminus from the enrichment experiment matches 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. Nearly half of all N-terminal Met identified were cleaved in our experiment, which is consistent with previously reported control study[46]. Though the total number of identifications is comparably lower, but this powerful approach has the ability to identify proteolytic products *in vivo*.

Conclusion

In this study, we explored the potentials of implementing MS-cleavable strategy for the identifications of proteolytic N-terminus after enrichment. The design of this new 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-terminus 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 result files of MS-GF+ searches. Tracking a proteolytic product in native *in vivo* experiments is extremely challenging due to the presence of low number of new N-termini and high sample complexity. Enriching regular N-terminal peptides has some success in this area using reverse purification strategy. We demonstrated an innovative workflow using a mass spectrometry cleavable reagent for the first time in this research 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 our large-scale results in *E. coli* clearly demonstrate 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. Further large-scale application with

this strategy is still 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.

Acknowledgements: Partially support from the grant UA5GM113216-01, NIGMS, NIH is acknowledged. We acknowledge support from UT Arlington chemistry and biochemistry department. We also acknowledge mass spectrometry support from Shimadzu center for advance analytical chemistry at UTA.

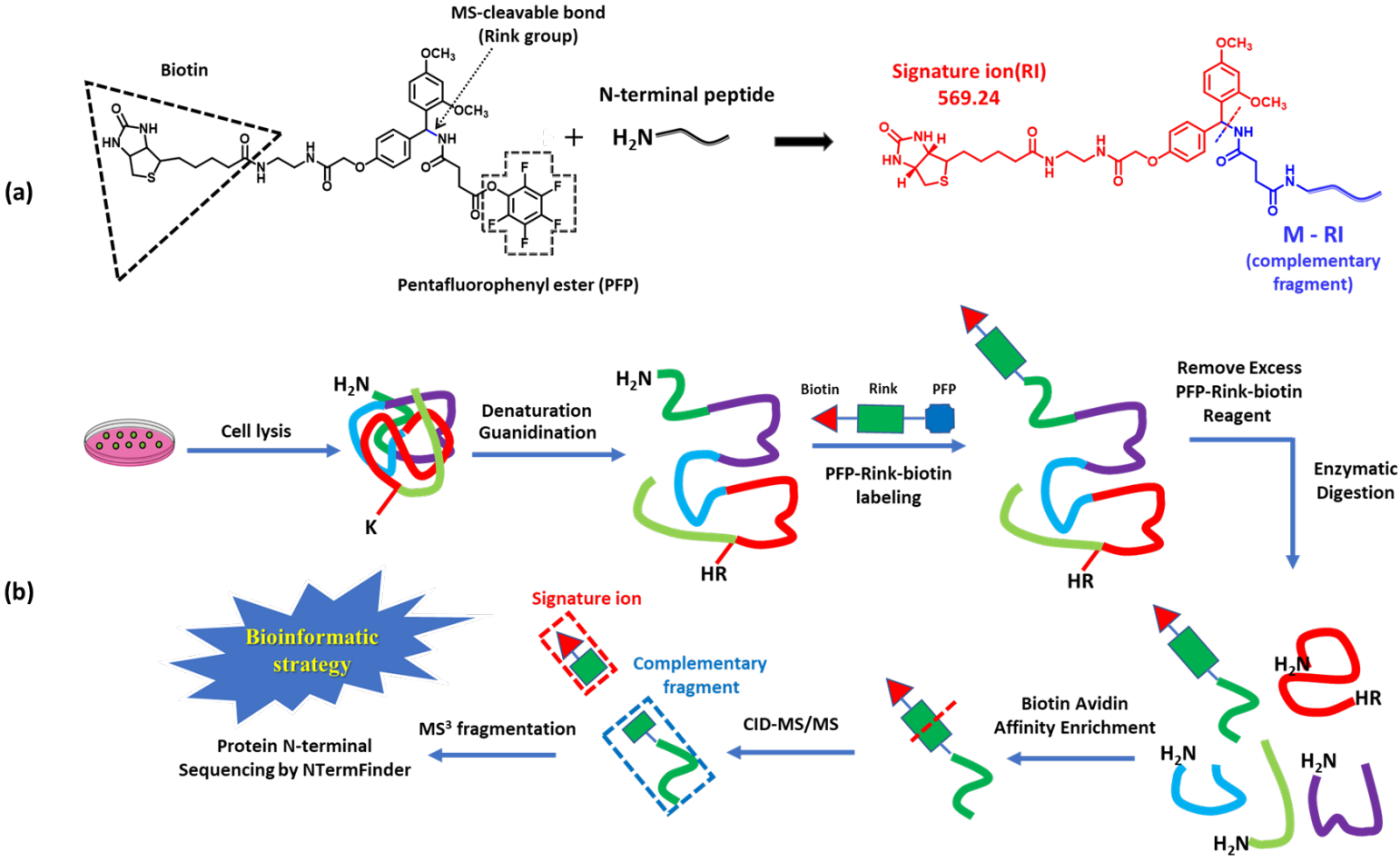


Figure 1. (a) Fragmentation of PFP-Rink-biotin-labeled N-terminal peptide: Under CID fragmentation, the PFP-Rink-biotin-coupled N-terminal peptide produces signature ion (m/z 569.24) and complementary fragment in high abundance in MS^2 spectrum; (b) Workflow of PFP-Rink-biotin-based enrichment and identification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

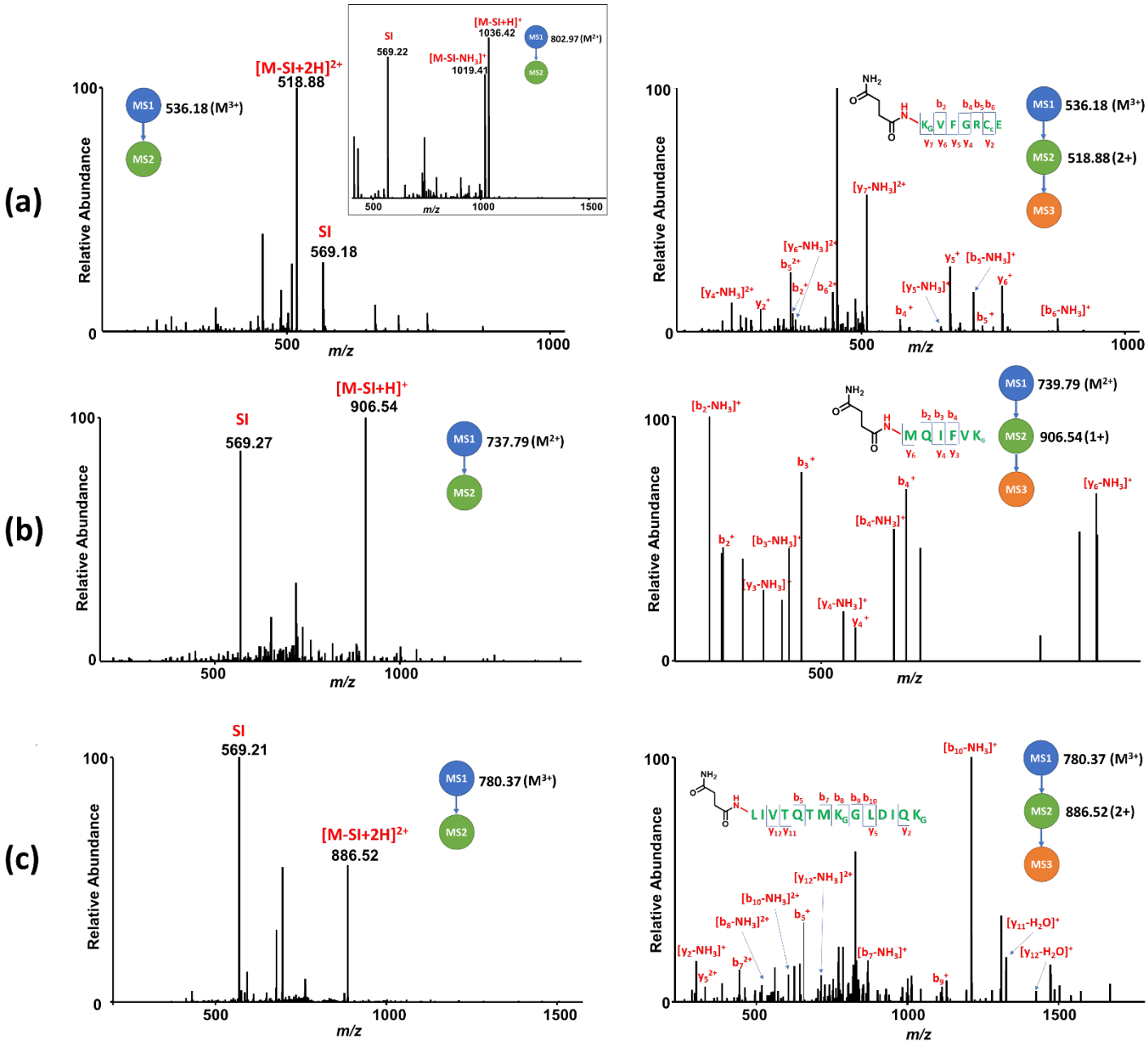


Figure 2. Identification of the labeled standard protein N-terminus by MS^n analysis: (a) MS^2 and MS^3 spectra of lysozyme N-terminus; (b) MS^2 and MS^3 spectra of ubiquitin N-terminus; (c) MS^2 and MS^3 spectra of β-lactoglobulin N-terminus. In all cases, the signature ion (SI) and complementary fragments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

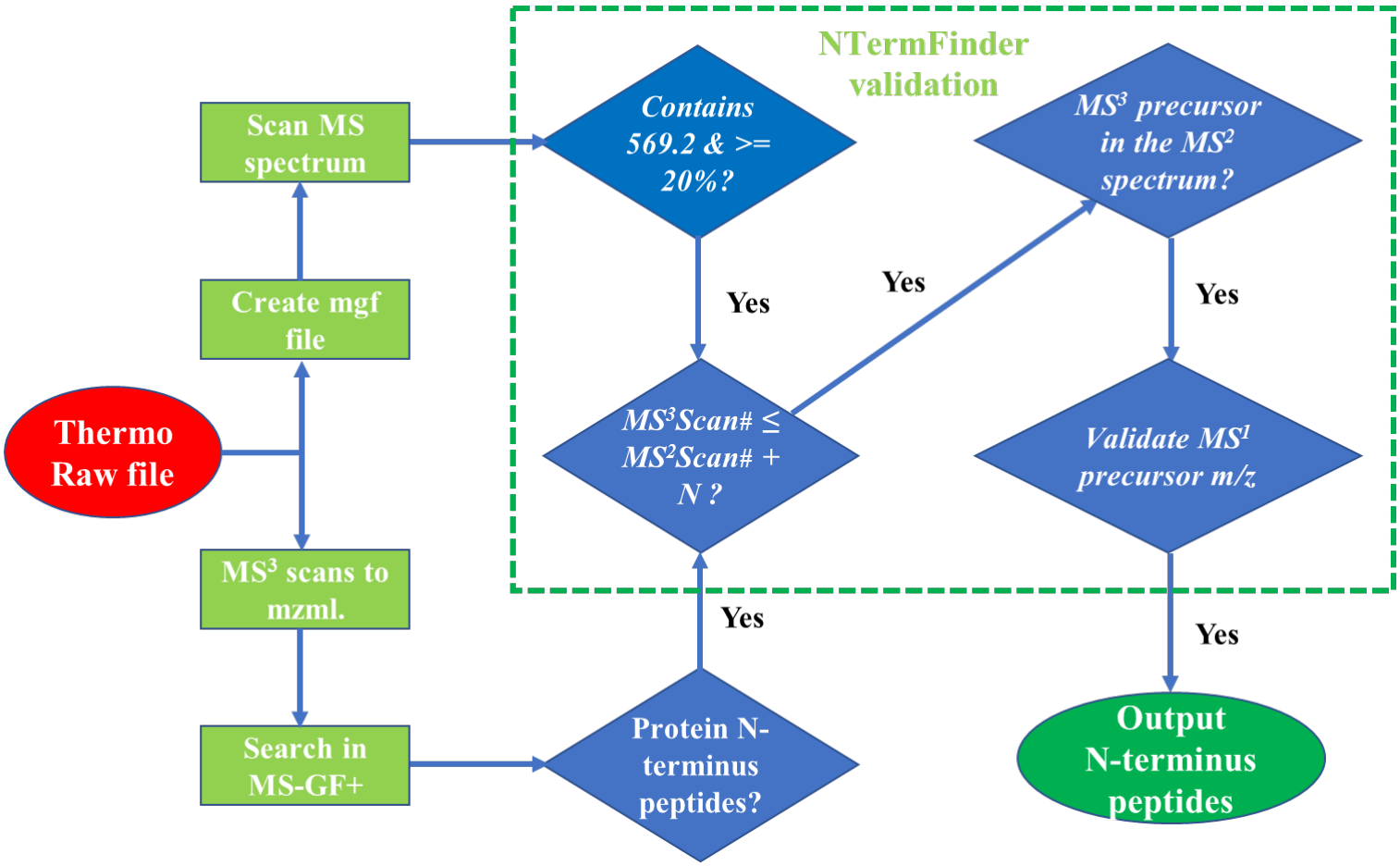


Figure 3. Bioinformatic approach for the validation of PFP-Rink-biotin labeled protein N-terminus. MS^2 spectra are converted to mgf format and MS^3 scans are converted to mzml format and searched in MS-GF+ to identify N-terminal peptides. During the validation of the identified peptides, each MS^2 spectrum containing m/z 569.2 ± 1.0 with higher than 20% relative intensity is matched with 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).*

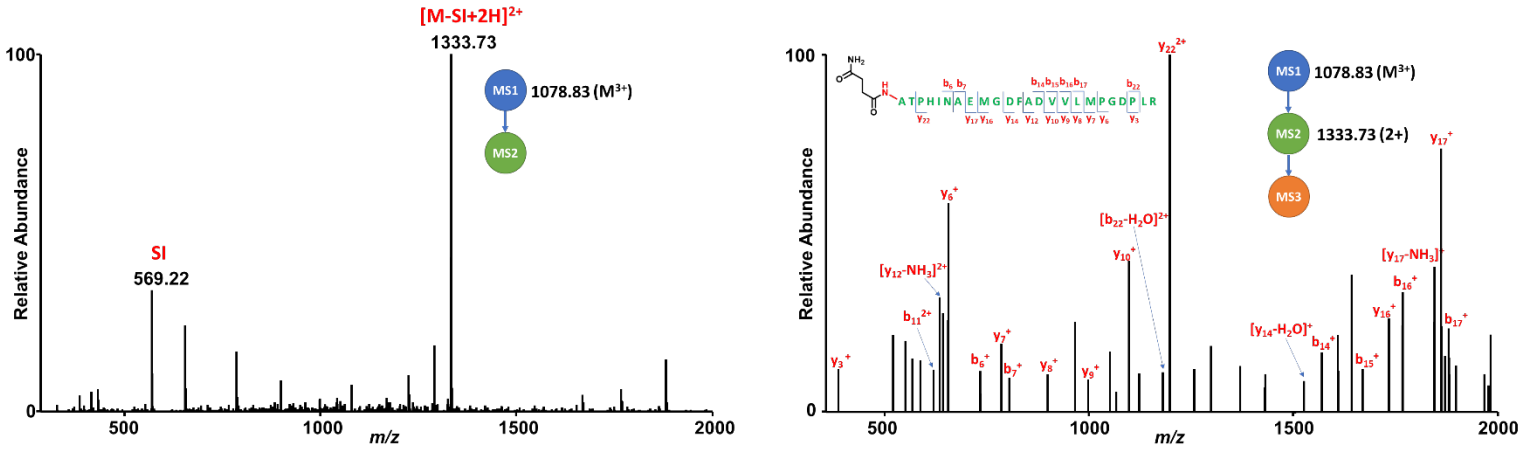
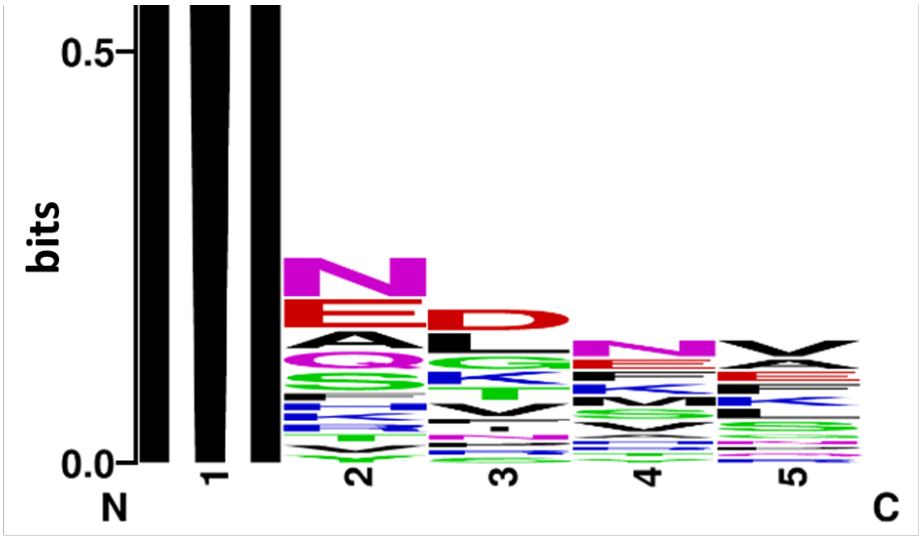
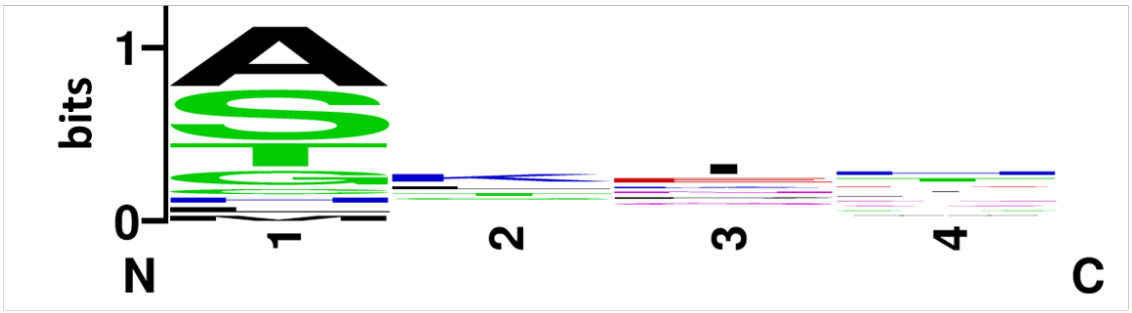


Figure 4. Identification of labeled *E. coli* purine nucleoside phosphorylase DeoD-type protein N-terminus by MS^n analysis. The generation of high-abundance signature ion and complementary fragment is consistent, with the high-confidence identification of the N-terminal peptide sequence by MS^3 fragmentation.

(a)



(b)



(c)

	A	E	F	K	N	R	S	H	Q	T	V	Y	G	C	L	M	Total
Met Intact	2	3	1	1	4	1	2	1	2	1	1	1	0	0	0	0	20
Met Removed	7	0	0	0	0	0	6	1	0	3	0	0	2	1	1	1	22
																	42

Figure 5. N-terminal sequence motifs and amino acid frequencies for the identified *E. coli* protein N-terminus: (a) Frequency of five N-terminal amino acids for identified N-terminal sequences with initiating Met intact; (b) Frequency of four N-terminal amino acids for identified N-terminal sequences with initiating Met removed; (c) Summary of the penultimate amino acids in the identified sequences (1st amino acids for Met removed sequences and 2nd amino acids for Met intact sequences)

Table 1. 42 unique protein N-terminus identified from *E. coli* cells

Number	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-terminus identified are 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

Materials and Methods

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 from TCI America (OR). Iodoacetamide (IAM), methanol, acetonitrile (ACN) were obtained from VWR (PA). Formic acid (FA), 3K MWCO protein concentrators, monomeric avidin and streptavidin agarose resins 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 from Oakwood Chemical (SC). Bacterial cell lysis kit was acquired from Goldbio (MO). Synthesis of PFP-Rink-biotin was prepared as we previously reported[47] and shown in **Supplemental Scheme-S1**. An Aries Filterworks (NJ) water system supplies all the high purity water used for preparing aqueous solutions.

Enrichment and identification of standard protein N-terminus

Preliminary experiment of labeling and enriching protein N-terminus was investigated for identifying the N-terminus of ubiquitin, lysozyme and β-lactoglobulin proteins. First, guanidination procedure was performed as previously described to block ε-amines of these proteins[38]. Briefly, O-methylisourea sulfate was dissolved in water and mixed with 1 mM, 5 μL of standard proteins. pH of the reaction solution was adjusted to between 10-11 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. 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 proceeded for 2 hours 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[48] or methanol-chloroform precipitation[37,49], followed by the disulfide cleavages with DTT and alkylation with IAM in dark. After the labeling, sample was digested either with trypsin or GluC (1:50 w/w) overnight in 50 mM ammonium bicarbonate. The sample was re-concentrated 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 done with the washing steps, 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. Supernatant was separated from the beads by centrifugation, and was subsequently evaporated, desalted and reconstituted in 0.1% FA for LC-MS^n analysis.

Enrichment and identification of proteolytic products from *E. coli* cell lysate

The *Escherichia. coli (E. coli)* top 10 cell was a gift from Dr. Shawn Christensen's lab. Cell lysis was performed with bacterial cell lysis kit according to manufacturer's manual and the protein concentration was measured with bicinchoninic acid assay using BSA standards. Aliquots of *E. coli* cell lysate were dissolved with 6 M guanidine hydrochloride containing 50 mM sodium bicarbonate. Followed by reaction with 4 mM DTT and then alkylation with 12 mM IAM in 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)[18]. Proteins were then purified by 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) overnight. The samples were then reconstituted in PBS buffer and then with mixing with PBS pre-washed streptavidin agarose beads for 1 h at room temperature. After peptide binding, the filtrate is removed, and the beads were washed extensively and sequentially with PBS and ultrapure water to remove the non-binding peptides. Once done with the washing steps, the beads were incubated with the elution buffer HFIP for 5 min at RT[40]. The filtrate was retained, evaporated, desalted and reconstituted in 0.1% FA for LC-MS^n analysis.

LC-MS^n analysis

All samples were analyzed in a Dionex UltiMate 3000 nano-UHPLC system hyphened with a nano-ESI-linear ion trap (LIT) Thermo Velos Pro mass spectrometer (Thermo Fisher Scientific, Waltham, MA). For standard protein enriched samples, Acclaim PepMap C18 column (150 mm × 75 μm, 3 μm) was used for the 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: 0.300 μL/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). Source voltage was 2.20 kV and capillary temperature was set to 275 °C. MS data was obtained from 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, activation Q of 0.25, and activation time of 30 ms. For MS^3 data-dependent acquisition, precursor mass range was set according to the mass range of the analyte and MS^3 spectra were collected from the top 2 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.

For *E. coli* enriched samples, Acclaim PepMap C18 column (500 mm × 75 μm, 3 μm) was used for the LC separation with 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). Source voltage was 2.00 kV and capillary temperature was set to 275 °C. MS data was obtained from 400 to 2000 *m*/*z* mass range in rapid scanning mode. Data-dependent MS/MS spectra were collected for the 5 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. Top 2 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% normalized collision energy, activation Q of 0.25, and 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 the MS scans. MS^3 mzML file was analyzed in MS-GF+ against respective databases with the parameter settings as follow[50,51]: 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 protein N-terminus (+99.032027 Da) and methionine-removed protein N-terminus (-32.007458 Da) were added. The MS-GF+ identified 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 discussion.

References

(1) Bradshaw, R. A.; Brickey, W. W.; Walker, K. W. N-Terminal Processing: The Methionine Aminopeptidase and Nα-Acetyl Transferase Families. *Trends Biochem. Sci.* **1998**, *23* (7), 263–267.

(2) Polevoda, B.; Sherman, F. The Diversity of Acetylated Proteins. *Genome Biol.* **2002**, *3* (5), reviews0006.1.

(3) Brown, J. L.; Roberts, W. K. Evidence That Approximately Eighty per Cent of the Soluble Proteins from Ehrlich Ascites Cells Are Nalpha-Acetylated. *J. Biol. Chem.* **1976**, *251* (4), 1009–1014.

(4) Van Damme, P.; Arnesen, T.; Gevaert, K. Protein Alpha-N-Acetylation Studied by N-Terminomics. *FEBS J.* **2011**, *278* (20), 3822–3834.

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

(6) 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.

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

(8) Prudova, A.; Keller, U. auf dem; 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.

(9) Mason, S. D.; Joyce, J. A. Proteolytic Networks in Cancer. *Trends Cell Biol.* **2011**, *21* (4), 228–237.

(10) Quesada, V.; Ordonez, G. R.; Sanchez, L. M.; Puente, X. S.; Lopez-Otin, C. The Degradome Database: Mammalian Proteases and Diseases of Proteolysis. *Nucleic Acids Res.* **2009**, *37* (Database), D239–D243.

(11) McQuibban, G. A. Inflammation Dampened by Gelatinase A Cleavage of Monocyte Chemoattractant Protein-3. *Science* **2000**, *289* (5482), 1202–1206.

(12) 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.

(13) 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.

(14) Xu, G.; Shin, S. B. Y.; Jaffrey, S. R. Global Profiling of Protease Cleavage Sites by Chemoselective Labeling of Protein N-Termini. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (46), 19310–19315.

(15) Yamaguchi, M.; Nakazawa, T.; Kuyama, H.; Obama, T.; Ando, E.; Okamura, T.; Ueyama, N.; Norioka, S. High-Throughput Method for N-Terminal Sequencing of Proteins by MALDI Mass Spectrometry. *Anal. Chem.* **2005**, *77* (2), 645–651.

(16) Ji, C.; Guo, N.; Li, L. Differential Dimethyl Labeling of N-Termini of Peptides after Guanidination for Proteome Analysis. *J. Proteome Res.* **2005**, *4* (6), 2099–2108.

(17) Enoksson, M.; Li, J.; Ivancic, M. M.; Timmer, J. C.; Wildfang, E.; Eroshkin, A.; Salvesen, G. S.; Tao, W. A. Identification of Proteolytic Cleavage Sites by Quantitative Proteomics. *J. Proteome Res.* **2007**, *6* (7), 2850–2858.

(18) 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.

(19) 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.

(20) Wildes, D.; Wells, J. A. Sampling the N-Terminal Proteome of Human Blood. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (10), 4561–4566.

(21) Staes, A.; Impens, F.; Van Damme, P.; Ruttens, B.; Goethals, M.; Demol, H.; Timmerman, E.; Vandekerckhove, J.; Gevaert, K. Selecting Protein N-Terminal Peptides

by Combined Fractional Diagonal Chromatography. *Nat. Protoc.* **2011**, *6* (8), 1130–1141.

(22) 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.

(23) Kleifeld, O.; Doucet, A.; Prudova, A.; auf dem Keller, U.; Gioia, M.; Kizhakkedathu, J. N.; Overall, C. M. Identifying and Quantifying Proteolytic Events and the Natural N Terminome by Terminal Amine Isotopic Labeling of Substrates. *Nat. Protoc.* **2011**, *6* (10), 1578–1611.

(24) 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.

(25) 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.

(26) 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.

(27) Sinz, A. Divide and Conquer: Cleavable Cross-Linkers to Study Protein Conformation and Protein–Protein Interactions. *Anal. Bioanal. Chem.* **2017**, *409* (1), 33–44.

(28) 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.

(29) Stadlmeier, M.; Bogena, J.; Wallner, M.; Wühr, M.; Carell, T. A Sulfoxide-Based Isobaric Labelling Reagent for Accurate Quantitative Mass Spectrometry. *Angew. Chemie Int. Ed.* **2018**, *57* (11), 2958–2962.

(30) Rink, H. Solid-Phase Synthesis of Protected Peptide Fragments Using a Trialkoxy-Diphenyl-Methylester Resin. *Tetrahedron Lett.* **1987**, *28* (33), 3787–3790.

(31) Tang, X.; Munske, G. R.; Siems, W. F.; Bruce, J. E. Mass Spectrometry Identifiable Cross-Linking Strategy for Studying Protein−Protein Interactions. *Anal. Chem.* **2005**, *77* (1), 311–318.

(32) Tang, X.; Bruce, J. E. A New Cross-Linking Strategy: Protein Interaction Reporter (PIR) Technology for Protein-Protein Interaction Studies. *Mol. Biosyst.* **2010**, *6* (6), 939–947.

(33) Kalkhof, S.; Sinz, A. Chances and Pitfalls of Chemical Cross-Linking with Amine-Reactive N-Hydroxysuccinimide Esters. *Anal. Bioanal. Chem.* **2008**, *392* (1–2), 305–312.

(34) 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.

(35) Canfield, R. E. The Amino Acid Sequence of Egg White Lysozyme. *J. Biol. Chem.* **1963**, *238*, 2698–2707.

(36) Braunitzer, G.; Chen, R.; Schrank, B.; Stangl, A. [The Sequence of Beta-Lactoglobulin (Author's Transl)]. *Hoppe. Seylers. Z. Physiol. Chem.* **1973**, *354* (8), 867–878.

(37) 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**, jssc.201901273.

(38) 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.

(39) 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.

(40) WO2016120247A1 - Methods for a quantitative release of biotinylated peptides and proteins from streptavidin complexes - Google Patents https://patents.google.com/patent/WO2016120247A1/ja (accessed Mar 26, 2020).

(41) 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.

(42) Schechter, I.; Berger, A. On the Size of the Active Site in Proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27* (2), 157–162.

(43) 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.

(44) Frottin, F.; Martinez, A.; Peynot, P.; Mitra, S.; Holz, R. C.; Giglione, C.; Meinnel, T. The Proteomics of N-Terminal Methionine Cleavage. *Mol. Cell. Proteomics* **2006**, *5* (12), 2336–2349.

(45) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14* (6), 1188–1190.

(46) 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.

(47) Yepremyan, A. Syntheses of Novel Peptide Linkers. M.S. Thesis, The University of Texas at Arlington, Arlington, TX, **2015**.

(48) 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.

(49) 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.

(50) 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.

(51) Kim, S.; Pevzner, P. A. MS-GF+ Makes Progress towards a Universal Database Search Tool for Proteomics. *Nat. Commun.* **2014**, *5*.