

**Proteomic studies on protein N-terminus and  
peptide ion mobility by nano-scale liquid  
chromatography/tandem mass spectrometry**

ナノスケール液体クロマトグラフィー/タンデム質量分  
析によるタンパク質 N 末端およびペプチドイオン  
モビリティに関するプロテオミクス研究

**2020**

**Chih-Hsiang Chang**

# CONTENT

ABBREVIATION .....	2
--------------------	---

## CHAPTER 1

### **Isolation of acetylated and unmodified protein N-terminal peptides by strong cation exchange chromatographic separation of TrypN-digested peptides**

INTRODUCTION.....	3
-------------------	---

RESULTS AND DISCUSSION .....	5
------------------------------	---

Retention behavior of TrypN-digested peptides in SCX chromatography .....	5
---	---

SCX HPLC separation of TrypN-digested HEK293T peptides .....	8
--	---

Optimization of SCX separation using TrypN-digested <i>E. coli</i> peptides.....	12
--	----

HEK293T protein N-terminal peptide enrichment by TrypN-SCX approach .....	14
---	----

CONCLUSION .....	17
------------------	----

EXPERIMENTAL SECTION .....	18
----------------------------	----

SUPPLEMENTAL TABLES.....	22
--------------------------	----

REFERENCES.....	23
-----------------	----

## **ABBREVIATION**

AA, ammonium acetate

ACN, acetonitrile

BCA, bicinchoninic acid

BSA, bovine serum albumin

CAA, 2-chloroacetamide

ChaFRADIC, charge-based fractional diagonal chromatography

COFRADIC, combined fractional diagonal chromatography

ESI, electrospray ionization

FDR, false discovery rate

HPG-ALD, highly-branched polyglycerol polymer

HPLC, high performance liquid chromatography

HYTANE, hydrophobic tagging-assisted N-termini enrichment

LC, liquid chromatography

MS, mass spectrometry

MW, molecular weight

PTM, post-translational modification

PTS, phase-transfer surfactants

RP, reversed-phase

SCX, strong cation exchange chromatography

SDC, sodium deoxycholate

SDB-XC, styrene divinylbenzene

SLS, sodium N-lauroylsarcosinate

StageTip, stop and go extraction tip

TAILS, terminal amine isotopic labeling of substrates

TCEP, tris(2-carboxyethyl)phosphine

TFA, trifluoroacetic acid

Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride

UniProtKB, universal protein resource knowledgebase

# CHAPTER 1

## Isolation of acetylated and unmodified protein N-terminal peptides by strong cation exchange chromatographic separation of TrypN-digested peptides

### INTRODUCTION

Characterizing protein N-termini is essential to understand how the entire proteome is generated through biological processes such as translational initiation,<sup>10-12</sup> post-translational modifications<sup>13,14</sup> and proteolytic cleavages.<sup>15,16</sup> In order to perform N-terminomics using mass spectrometry (MS), peptides derived from protein N-termini must be selectively enriched, and many methods have been developed for this purpose.<sup>17,18</sup> Some of them use “positive selection” approaches in which chemically labeled protein N-terminal peptides are enriched by affinity purification.<sup>15,19</sup> However, these approaches are not applicable to proteins with *in vivo* N-terminal modifications. In contrast, “negative selection” approaches to isolate protein N-terminal peptides by depleting internal peptides have been used to comprehensively identify protein N-terminal peptides, including N-terminal modifications such as methylation, acetylation, and lipidation.<sup>20,21</sup> Gevaert et al. pioneered combined fractional diagonal chromatography (COFRADIC),<sup>22</sup> and this was followed by other negative selection approaches such as terminal amine isotopic labeling of substrates (TAILS),<sup>23</sup> the variant of COFRADIC called charge-based fractional diagonal chromatography (ChaFRADIC),<sup>24</sup> and hydrophobic tagging-assisted N-termini enrichment (HYTANE).<sup>25</sup> All of them require blocking of the primary amines at the protein level and depletion of digested internal peptides by means of chemical tagging-based separation. Thus, relatively large amounts of samples (~5 to 10 mg) are generally required to increase the identification number of protein N-terminal peptides. This limits the usefulness of these approaches in the case of hard-to-obtain biological samples.<sup>26,27</sup> Furthermore, limitations in the efficiency and specificity of the chemical derivatizations compromise the confidence of peptide identification. Therefore, a simple and sensitive approach to enrich protein N-terminal peptides is still needed for MS-based proteomics.

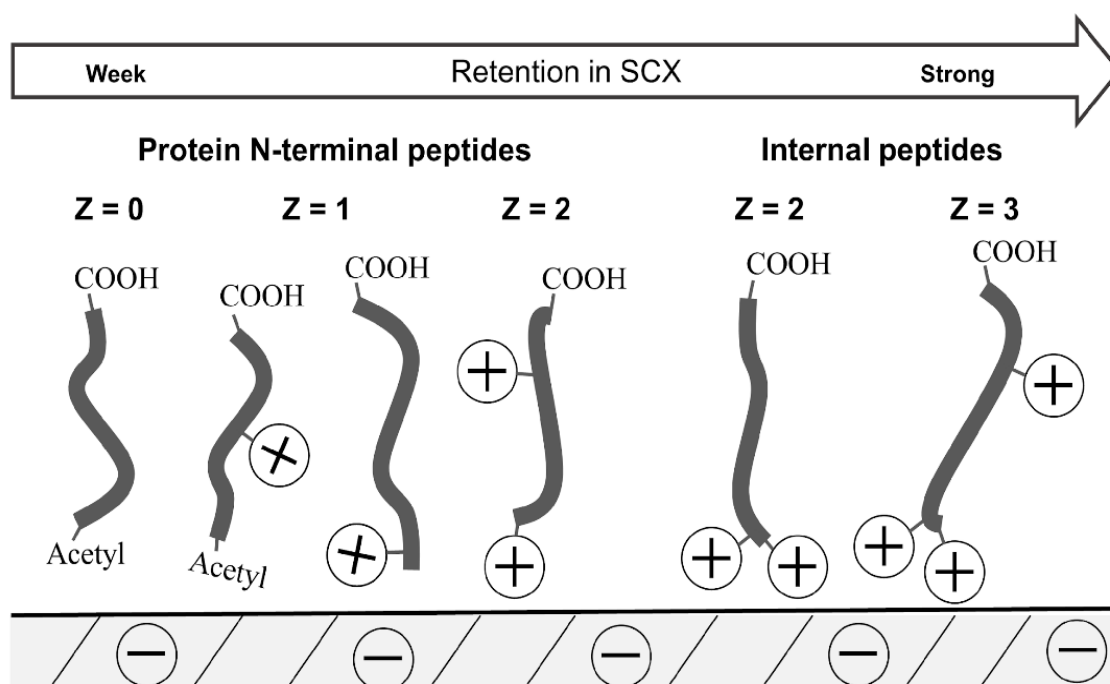
Strong cation exchange (SCX) chromatography, employing Coulombic interactions to separate peptides based on their charge at acidic pH, has been widely applied for deep proteome profiling.<sup>28,29</sup> Alpert et al. reported that the peptide retention in SCX is affected by charge and orientation.<sup>30</sup> In SCX separation of tryptic peptides, acetylated protein N-terminal peptides and protein C-terminal peptides are eluted first. Monophosphorylated peptides with +1 charge are then eluted, followed by peptides with +2 or more charge, such as unmodified protein N-terminal peptides, internal peptides and peptides containing missed cleavages.<sup>31,32</sup> Thus, it is impossible to isolate protein N-terminal peptides from tryptic peptides by SCX chromatography. This is also the case where LysN was used with SCX chromatography, based on the charge/orientation retention model.<sup>30</sup> To overcome this issue, we focused on TrypN, also known as LysargiNase, a metalloprotease that cleaves peptide chains mainly at the N-terminal side of Lys/Arg even in the case of Pro-Lys and Pro-Arg bonds, generating peptides with N-terminal Lys/Arg and yielding protein N-terminal peptides that do not contain Lys/Arg.<sup>33</sup> Unlike other kinds of LysargiNase such as ulilysin<sup>34,35</sup> and mirolysin,<sup>36</sup> which preferentially cleave the N-terminal side of either Lys or Arg, TrypN cleaves the N-terminal side of Lys and Arg equally at pH 6~8. Moreover, the peptide identification performance for N-terminal Lys/Arg peptides is comparable to that for tryptic peptides.<sup>37</sup>

In this chapter, we developed a new method to enrich protein N-terminal peptides without the need for chemical derivatization or complex procedures, taking advantage of the combination of proteinase TrypN-mediated protein cleavage and SCX separation of N-terminal peptides based on the extended charge/orientation retention model. We show that this rapid and simple approach to enrich protein N-terminal peptides enables comprehensive, high-throughput analysis of the human and bacteria N-terminomes.

# RESULTS AND DISCUSSION

## Retention behavior of TrypN-digested peptides in SCX chromatography

Proteolysis with TrypN yields peptides with at least a +2 charge with Lys or Arg and an  $\alpha$ -amino group at the peptide N-terminus at the acidic pH. On the other hand, peptides derived from protein N-termini have neither Lys nor Arg and are often acetylated at the protein N-terminus, so that most of them have a 0 or +1 charge, and only His-containing peptides with an unmodified protein N-terminus have a +2 charge (Figure 1-1). In this study, we focused on the fact that SCX chromatography under the acidic conditions might be able to separate peptides based on the number of positive charges as well as the localization of the charges according to the charge/orientation retention model,<sup>30</sup> and we attempted to separate protein N-terminal peptides from internal peptides among TrypN-digested peptides.

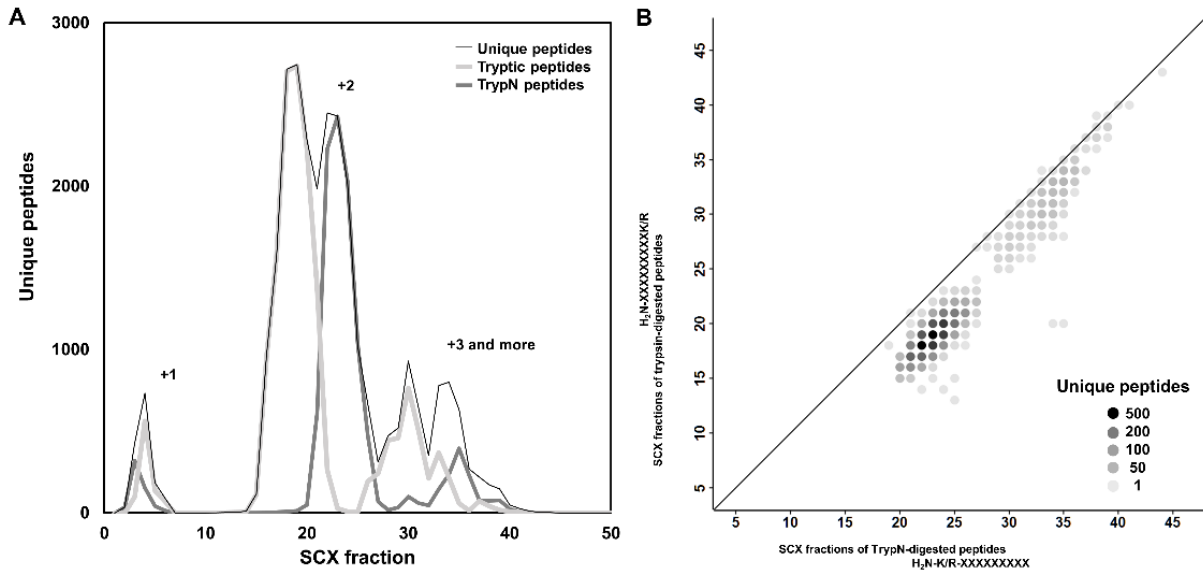


**Figure 1-1. Charge/orientation-based SCX retention model for TrypN-digested peptides.** Six types of TrypN-digested peptides are shown. Z is the charge number at acidic pH, which is based on the number of basic residues per peptide, such as unmodified N-terminus, Lys, Arg and His.

We first examined the number of missed cleavages in TrypN digestion. When digestion was performed in 0.1% RapiGest according to the manufacturer's protocol, the missed cleavage rate (the content of peptides with two or more missed cleavage sites) was 14%, almost equal to the value in the condition without addition of RapiGest (16%). On the other hand, when 1%

SDC was added instead of RapiGest, the missed cleavage rate was dramatically reduced to 5.8%. Thus, TrypN digestion was performed according to the PTS protocol in this study.<sup>38</sup>

Next, keeping in mind the need to separate protein N-terminal-derived peptides with both His residues and unmodified N-termini from TrypN-digested internal peptides, we investigated whether the peptides could be separated based on the position of the positive charge, in addition to the number of positive charges, by SCX chromatography. Studies with proteases that cleave either Lys/Arg, such as LysC and LysN or trypsin and TrypN, have indicated that the position of the positive charge affects the outcome in shotgun proteomics.<sup>35,39</sup> For example, it has been reported that peptides with N-terminal Lys or Lys/Arg are more strongly retained than peptides with a C-terminal Lys or Lys/Arg in reversed-phase liquid chromatography (RP-LC).<sup>37</sup> To determine how the Lys/Arg position of peptides affects their retention behavior in SCX chromatography at acidic pH, we examined a mixture of TrypN- and trypsin-digested peptides using the SCX HPLC system, followed by nanoLC/MS/MS. The 19,853 unique tryptic peptides generally showed weaker retention than the 11,334 unique TrypN peptides with the same charge states (Figure 1-2A, Supplemental Table 1). To characterize the SCX elution profiles in more detail, we compared the retention time in SCX HPLC for approximately 4,000 peptide pairs having sequences that differ only in the position of terminal Lys/Arg (Figure 1-2B). As expected, TrypN-digested peptides exhibit stronger SCX retention compared to Trypsin analogs. This would be because the TrypN peptides carry two positively charged groups at the N-terminus, due to the  $\alpha$ -amino group of the N-terminal Lys/Arg and the side chain  $\epsilon$ -amino or guanidino group, whereas the positive charge of the C-terminal Lys/Arg of trypsin peptides was partially neutralized by the  $\alpha$ -carboxy group under acidic condition (Figure 1-1). Alpert et al. and Gauci et al. reported that LysN-digested phosphopeptides with two basic moieties in close proximity tend to be more strongly retained on an SCX column than tryptic phosphopeptides.<sup>30,32</sup> Gussakovskiy et al. reported a retention model for predicting the retention times in SCX chromatography of tryptic peptides, in which the position-dependent coefficient of basic amino acids is higher near the N-terminus.<sup>40</sup> We also found that the TrypN-digested peptides eluted in a narrower SCX fraction range than the tryptic peptides (Figure 1-2A). This may be due to the fact that the distance between N-terminal positive charge in the tryptic peptides differs depending upon the length of peptide, whereas the TrypN-digested peptides have the N-terminal Lys/Arg that minimizes the distance between the two positive charges. To our knowledge, the present work is the first to validate the peptide charge/orientation retention model in SCX using thousands of identical sequence pairs.

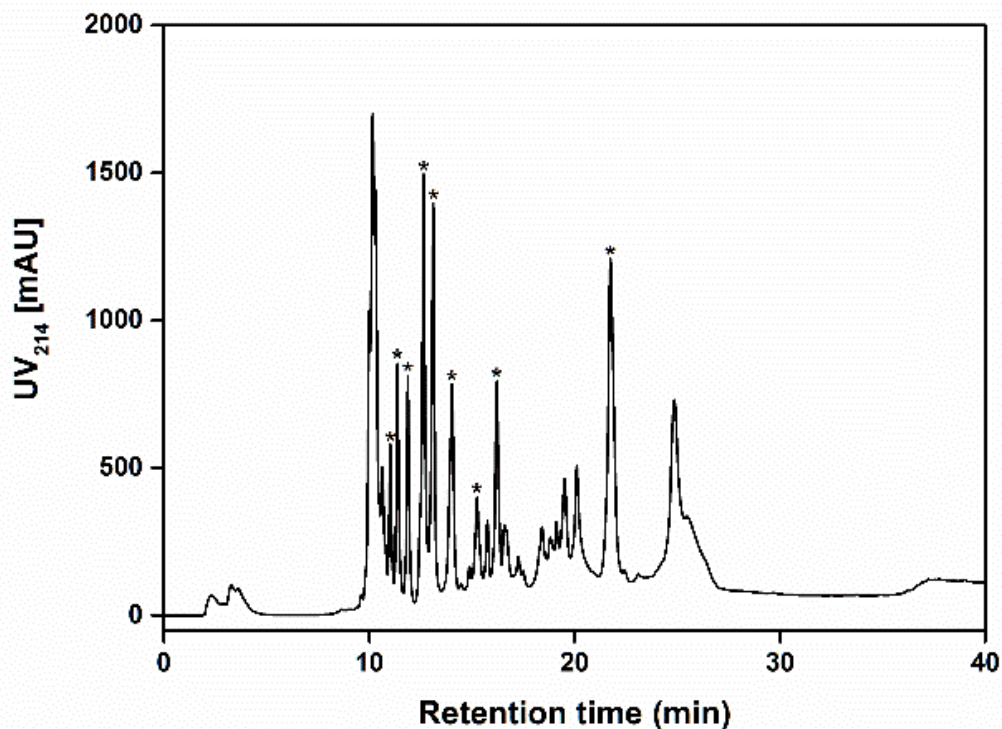


**Figure 1-2. SCX separation of TrypN-digested and trypsin-digested peptides.** (A) SCX separation of different types of peptides in digested HEK293T cell lysate. The peptide sample was prepared by mixing TrypN- and trypsin-digested HEK293T peptides with 1:1. The black, light gray and dark gray curves represent unique peptides, trypsin-digested peptides and TrypN-digested peptides, respectively. The charge number at acidic pH was labeled. (B) Comparison of SCX elution profiles of TrypN- and trypsin-digested peptides. Peptides with the same sequence except for the termini were selected (K/R-XXXXXXXX and XXXXXX-K/R for TrypN- and trypsin-digested peptides, respectively). The shade of the circle color indicates the number of peptides.

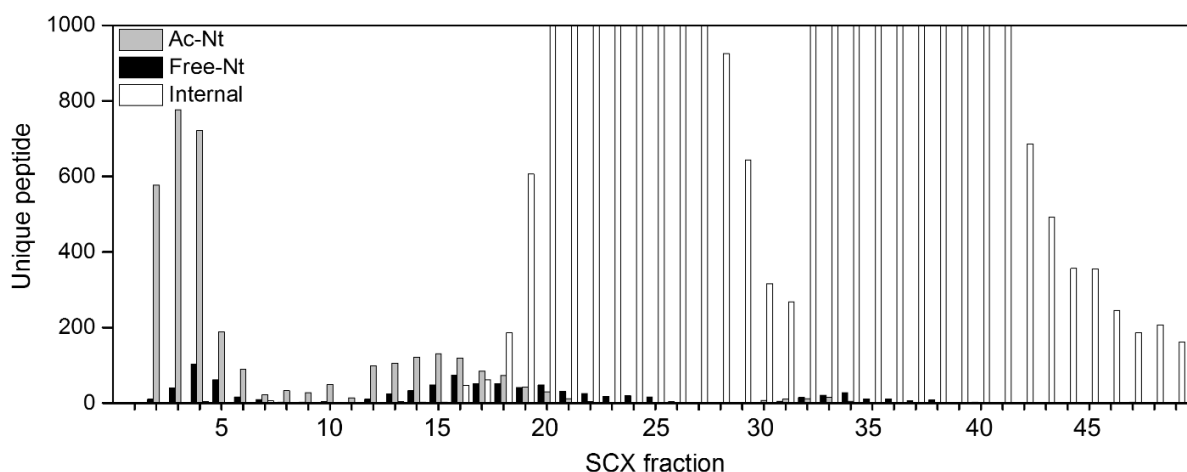


## SCX HPLC separation of TrypN-digested HEK293T peptides

The HPLC system used in this study was equipped with a nonporous hydrophilic SCX column having a separation efficiency equivalent to that of a typical reversed-phase column (the peak width at half height was  $12.4 \pm 4.2$  seconds and the peak capacity was 122; see Figure 1-3). As already shown in Figure 1-2A, this SCX HPLC system was able to separate TrypN-digested peptides with +1 and +2 charges from each other. Comprehensive SCX fractionation of TrypN-digested peptides derived from HEK293T cells was performed with a KCl salt gradient elution at pH 2.6, and peptide identification for each fraction was performed by nanoLC/MS/MS (Supplemental Table 2). As shown in Figure 1-4, nearly all of the protein N-terminal-derived peptides were clearly separated from the internal peptides, regardless of whether their N-termini were acetylated or unmodified. The fractions from 2 to 11 min contained mainly 0 and +1 peptides, including 2,207 acetylated protein N-terminal peptides, 345 His-containing acetylated N-terminal peptides, and 262 unmodified N-terminal peptides. The 12-18 min fractions contained +2 peptides, i.e., unmodified protein N-terminal peptides containing one His, Lys or Arg and acetylated protein N-terminal peptides containing two basic amino acids. The next fractions from 19 to 30 min also contained +2 peptides, but most of them were internal peptides based on the orientation effect, i.e., retention was stronger due to the high density of positive charge at the N-terminus of the peptides (Figure 1-2B). Thus, the protein N-terminal peptides can be easily isolated. Peptides with a charge greater than 2+ were sequentially eluted in the fractions after 31 min. These included protein N-terminal peptides containing missed cleavage sites, but their number was small due to the high efficiency of TrypN digestion by the PTS method. Up to 90% of non-redundant protein N-terminal peptides could be recovered in fractions up to 18 min by this approach (Figure 1-5), demonstrating that the combination of TrypN digestion with SCX HPLC enables simple and rapid protein N-terminal peptide enrichment. In addition, unlike trypsin, which is unable to cleave Lys-Pro and Arg-Pro bonds, TrypN can cleave Pro-Lys and Pro-Arg bonds, generating protein N-terminal peptides with Pro at the C-termini, and thus improving the coverage in the N-terminome analysis.



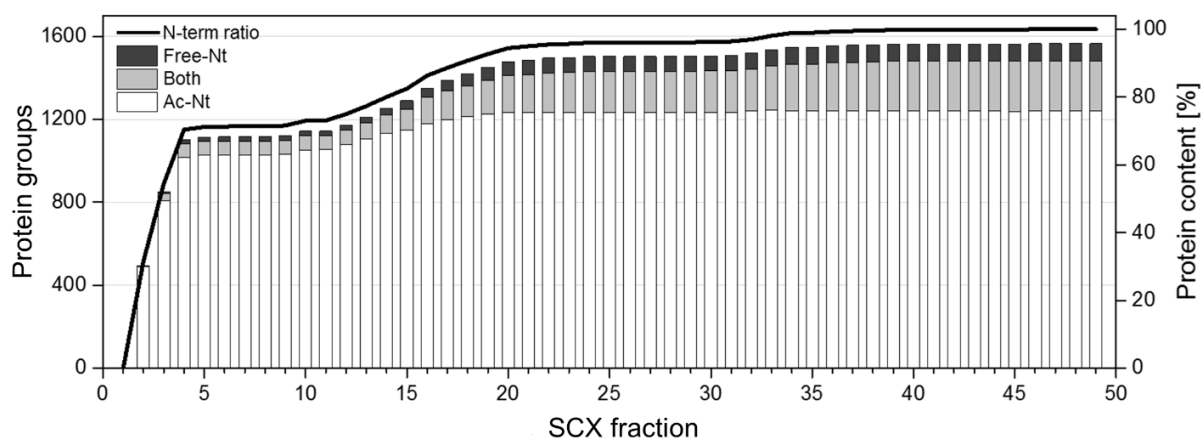
**Figure 1-3. Typical chromatogram of high resolution SCX chromatography employed in this study.** Tryptic HSA peptides (500  $\mu$ g) were analyzed by using the Shimadzu Prominence HPLC system with Agilent BioEX SCX column (250 mm  $\times$  4.6 mm inner diameter, 5  $\mu$ m, non-porous beads made of poly(styrene-divinylbenzene) modified with sulfonate groups). A mixture of 7 mM sodium phosphate (pH 2.6) and ACN (7:3) was used for SCX buffer A, and 1 M NaCl was added to buffer A for SCX buffer B. The two-step linear gradient was performed as follows: 0% B for 5 min, 0-40% in 30 min, 40-99% in 5 min and 100% B for 10 min. Peak capacity was calculated using the following equation:  $PC = 1 + tg/W0.5$ . ( $W0.5 = 12.4 \pm 4.2$  sec,  $PC = 122$ .) Peaks with asterisk were selected for measuring the half peak width ( $W0.5$ ) to calculate peak capacity.



**Figure 1-4. SCX elution profiles of TrypN-digested peptides.** SCX HPLC fractionation of different types of peptides in TrypN-digested HEK293T cell lysates using KCl salt gradient elution. In fractions 2-6, protein N-terminal peptides with  $Z = 0$  and 1 were observed, such as acetylated protein N-terminal peptides with or without one basic amino acid and unmodified protein N-terminal peptides without basic amino acid.  $Z$  is the charge number at acidic pH, which is based on the number of basic residues per peptide, such as unmodified N-terminus, Lys, Arg and His.

Since the charge number of the protein N-terminal peptides at acidic pH is generally smaller than that of internal peptides, we examined whether the identification efficiency of protein N-terminal peptides is affected by the low positive charge number. Since peptide identification is influenced by several steps, including ionization, ion transmission from MS1 to MS2, and fragmentation, four parameters such as the UV absorbance at 214 nm in SCX, the total ion currents in MS and MS/MS scans, and the Mascot peptide score distribution were measured for SCX 1-18 min fraction (protein N-terminal peptides were enriched) and 19-50 min fraction (internal peptides were enriched), respectively (

Table 1-1). Considering that the UV absorbance ratio of the 1-18 min fraction to the 19-50 min fraction was smaller than the ratio of the average total ion current per MS scan, the ionization efficiency of protein N-terminal peptides was better than that of the internal peptides due to the lower sample complexity of the 1-18 min fraction. For ion transmission efficiency from MS1 to MS2, we did not see any difference between protein N-terminal peptides and internal peptides. As for fragmentation, the profiles of charge number distribution at acidic pH were significantly different between the protein N-terminal and internal peptides, but the obtained profiles of the score distribution were almost identical. This could be due to the similar distribution profiles of the charge states of the precursor ions. These results indicate that there is no clear disadvantage of using TrypN for the identification of the protein N-terminal peptides.



**Figure 1-5. Accumulation of non-redundant protein groups based on protein N-terminal peptides during SCX elution.** Sample: TrypN-digested HEK293T cell lysate. Free-Nt (black), both (gray) and Ac-Nt (white) squares in the bar graph represent unmodified, partially acetylated and acetylated protein N-termini, respectively. The curve represents the accumulated content of protein groups.

**Table 1-1. Evaluation of the identification efficiency of TrypN-digested HEK293T peptides used in Figure 1-4.**

	Total peak area in SCX (UV 214 nm)	Average total ion current per MS1 scan	Average total ion current per MS2 scan	Mascot peptide score (average $\pm$ SD)	Precursor ion charge state (1+/2+/3+>3+)	Charge number at acidic pH (0/+1/+2/+3/>+3)
Protein N-terminal peptides (fr1-18)	5.8E+04	1.9E+08	3.0E+06	51 $\pm$ 26	1/86/11/2	21/41/37/1/0
Internal peptides (fr19-50)	3.5E+05	9.4E+08	1.5E+07	52 $\pm$ 24	0/62/33/4	0/0/52/28/20
Protein N-term/Internal ratio	0.17	0.20	0.20	0.98	-	-



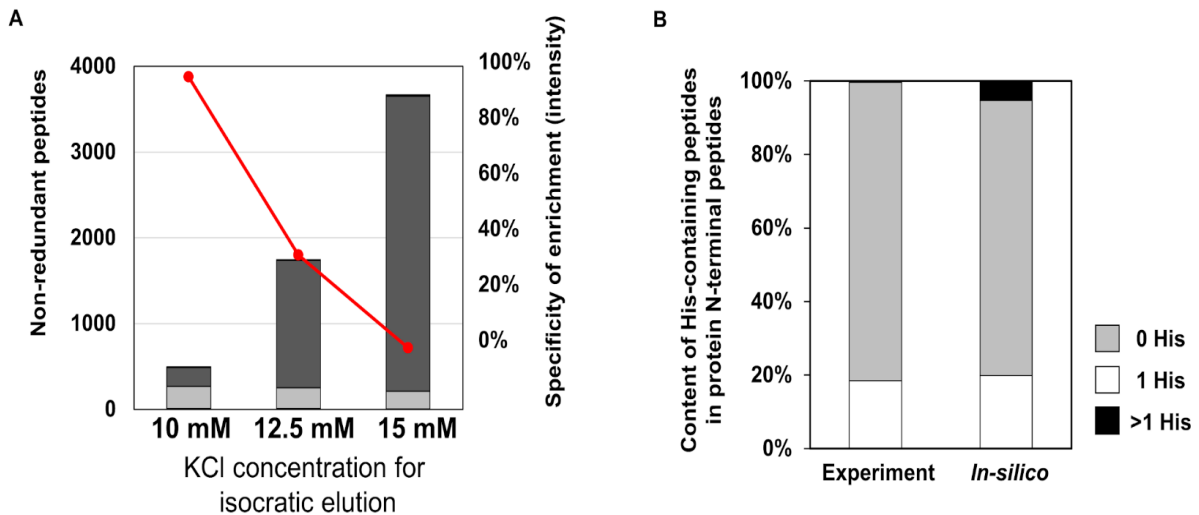
## Optimization of SCX separation using TrypN-digested *E. coli* peptides

To optimize the elution conditions for isolation of protein N-terminal peptides, we employed *E. coli* TrypN-digested peptides. Because bacterial proteins have less N-terminal modification than mammalian proteins, the bacterial sample was considered preferable to optimize the conditions for separating the protein N-terminal peptides with +2 charge (peptides with an unmodified N-terminus and one His residue) from the internal peptides (Figure 1-1). Three SCX buffers with different KCl concentrations were used for isocratic elution for 30 min, and the enrichment efficiencies for protein N-terminal peptides were compared (Supplemental Table 3). An enrichment specificity of more than 97% was obtained with 10 mM KCl (Table 1-2). When buffers with higher KCl concentrations were used, more internal peptides were identified (Figure 1-6A). In the case of 10 mM KCl buffer, we identified 53 His-containing protein N-terminal peptides out of 270 non-redundant protein N-terminal peptides without missed cleavage from 20 µg of *E. coli* lysate (19.6%, Figure 1-6B). Among *in silico* TrypN-digested peptides from the *E. coli* proteome, 20% of the protein N-terminal peptides contain one His, suggesting that our enrichment conditions have no bias in identifying His-containing protein N-terminal peptides. In other words, this SCX chromatography was able to isolate the protein N-terminal peptides from TrypN-digested *E. coli* peptides even in the most difficult cases where the unmodified protein N-terminal peptides contain an additional basic amino acid such as His, Lys or Arg near the N terminus (Figure 1-7). Although this SCX separation can be explained by the charge/orientation model, it is the first report to apply the retention model to N-unmodified protein N-terminal peptides.

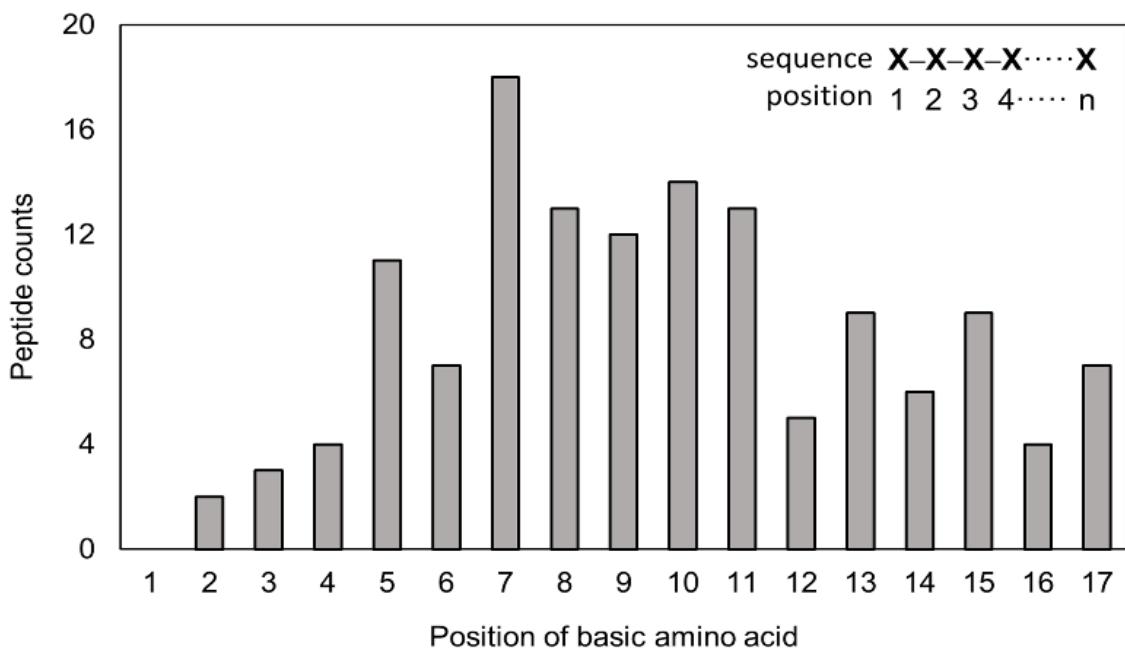
**Table 1-2. Enrichment of *E. coli* protein N-terminal peptides by SCX HPLC with isocratic elution at three different KCl concentrations.**

Salt concentration	10 mM	12.5 mM	15 mM
Unique peptides	432	1,669	3,416
Unmodified protein N-terminal peptides	326	444	387
Acetylated protein N-terminal peptides	31	25	22
Protein N-terminal peptides (%) based on peptide counts	82.6	28.1	12.0
Protein N-terminal peptides (%) based on peak area	98.2	49.2	18.5

The enrichment specificity of protein N-terminal peptides is obtained by calculating the number and LC/MS peak area of protein N-terminal peptides among all identified peptides.



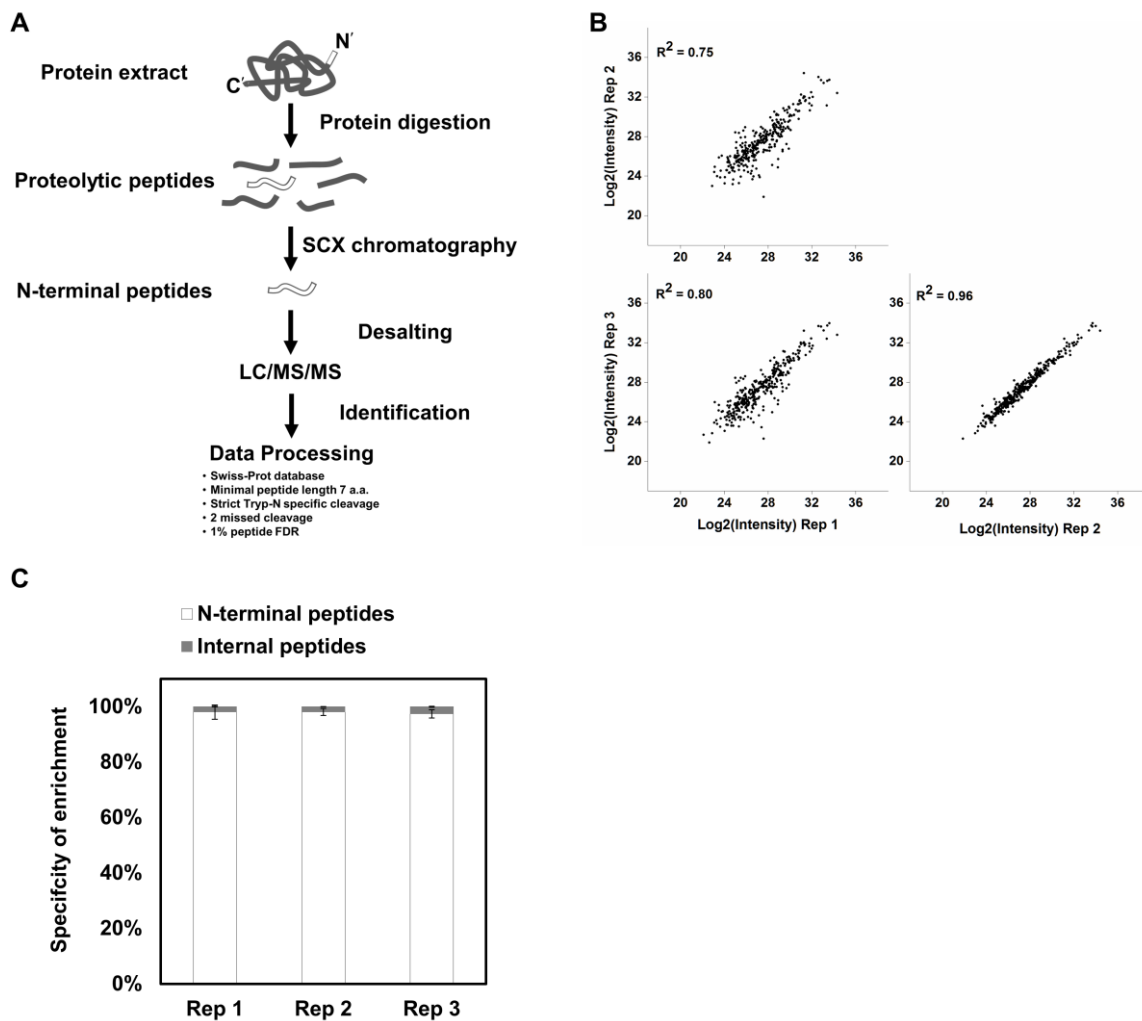
**Figure 1-6. SCX elution profiles of TrypN-digested peptides.** (A) SCX HPLC separation of TrypN-digested *E. coli* peptides under isocratic conditions with different KCl concentrations. The specificity in enriching protein N-terminal peptides and the numbers of peptides with different Z values are shown as a red curve. (B) The content of His-containing peptides in protein N-terminal peptides. The experimental results were those obtained with 10 mM KCl isocratic elution, and the *in-silico* results were calculated from the *E. coli* K-12 MG1665 protein sequence database. Details of *in-silico* digestion are described in the experimental section.



**Figure 1-7. Dependence of the localization of basic amino acids (K, R, or H) on isolating protein N-terminal peptides by SCX with 10 mM KCl isocratic elution.** Unmodified protein N-terminal peptides with +2 charge (137 peptides) isolated from TrypN-digested *E. coli* cell lysate were plotted.

## HEK293T protein N-terminal peptide enrichment by TrypN-SCX approach

The N-terminal peptides of His-containing proteins could be successfully separated from the internal peptides of human and bacterial samples by SCX HPLC under optimized elution conditions. To validate the applicability of this method to large-scale N-terminal proteomics, we performed triplicate analyses using HEK293T cells, which have been widely used in N-terminal proteomics.<sup>26,27</sup> Triplicate SCX HPLC fractionations using 10 mM KCl isocratic elution were done for TrypN-digested HEK293T peptides (80 µg each time), and we subjected one-fourth of the isolated peptides to nanoLC/MS/MS in triplicate (9 runs in total). Default parameters, such as the Swiss-Prot human protein sequence database, specific TrypN cleavage, and minimum peptide length of 7 amino acids, were applied for peptide identification by database search (Figure 1-8A). The results are shown in Figure 1-8B, Table 1-3 and Supplemental Table 4. High correlations of peak areas of identified peptides were observed for intra- and interday preparation samples, ( $R^2 = 0.96$  and  $R^2 = 0.75, 0.80$ , respectively). On average, we identified 1,550 unique acetylated and 200 unmodified protein N-terminal peptides from 20 µg of TrypN-digested HEK293T peptides in a single LC/MS/MS analysis. Contamination by internal peptides amounted to only 3% and 9% in peptide peak area and peptide number, respectively (Figure 1-8C, Table 1-3). Protein N-terminal peptides with missed cleavage were also enriched in the same elution, and 850 (~50%) miscleaved unique N-terminal peptides were identified on average, improving the coverage of the N-terminome. We identified 1,640 acetylated, 106 partially acetylated and 167 unmodified non-redundant protein N-termini. Note that 1,600 additional neo-N-terminal peptides were identified when semi-specific cleavage at the N-terminus was allowed in the data processing, although our purpose in this study was not to find novel proteoforms, but to establish a novel approach for N-terminomics. Furthermore, to compare our results with two published N-terminome datasets for HEK293T human cells,<sup>17,18</sup> we re-analyzed those datasets under the same conditions without the use of their original customized database or non-specific cleavage. In terms of the contents of acetylated and unmodified protein N-terminal peptides, all three datasets provided identical results, whereas the content of internal peptides as well as the number of unique peptides varied depending on the approach and the sample amount (Figure 1-9).



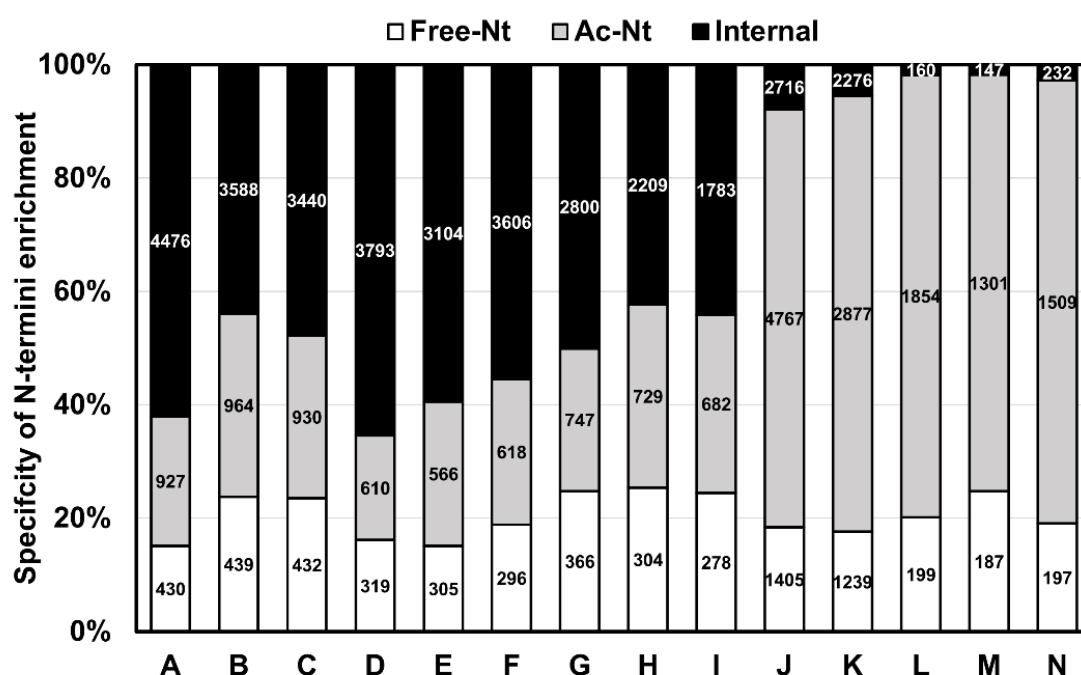
**Figure 1-8. N-terminal proteomics using SCX chromatography.** (A) Workflow of TrypN/SCX-based N-terminal proteomics. For details, see Materials and Methods. Three technical replicates were conducted on Day 1 (Rep 1) and Day 2 (Rep 2 and Rep 3) to evaluate inter- and intraday reproducibility. (B) Reproducibility in quantifying peptide peak areas between three technical replicates. Rep 1-Rep 2 and Rep 1-Rep 3 correlation represent interday reproducibility whereas Rep 2-Rep 3 correlation shows intraday reproducibility. (C) Enrichment specificity for protein N-terminal peptides in three technical replicates.



**Table 1-3. Identification of protein N-terminal peptides from TrypN-digested HEK293T cells.**

	Replicate 1	Replicate 2	Replicate 3	Total
Unmodified protein N-terminal peptides	199 ( $\pm 3$ )	187 ( $\pm 5$ )	197 ( $\pm 3$ )	352
Acetylated protein N-terminal peptides	1,854 ( $\pm 13$ )	1,301 ( $\pm 18$ )	1,509 ( $\pm 15$ )	2,666
Internal peptides	160 ( $\pm 8$ )	147 ( $\pm 3$ )	232 ( $\pm 6$ )	433
N-term ratio (% , peptide counts)	92.8 ( $\pm 0.3$ )	91.0 ( $\pm 0.2$ )	88.0 ( $\pm 0.3$ )	
N-term ratio (% , peptide area)	97.4 ( $\pm 0.3$ )	98.0 ( $\pm 0.5$ )	97.4 ( $\pm 0.5$ )	
Unmodified protein groups	115 ( $\pm 3$ )	100 ( $\pm 4$ )	116 ( $\pm 3$ )	167
Partially acetylated protein groups	36 ( $\pm 2$ )	60 ( $\pm 3$ )	50 ( $\pm 2$ )	106
Acetylated protein groups	1,223 ( $\pm 7$ )	1,000 ( $\pm 9$ )	1,187 ( $\pm 16$ )	1,640

Samples were prepared in triplicate (Replicates 1-3) and nanoLC/MS/MS of each sample was conducted in triplicate. Each number in the table is the average of triplicate measurements, and the total number is calculated after merging all results (n=9) and removing redundancy. The enrichment specificity of protein N-terminal peptides is obtained by calculating the number and peak area of protein N-terminal peptides among all identified peptides.



**Figure 1-9. Comparison of the enrichment efficiency for protein N-terminal peptides from HEK293T cell lysate between published datasets and this study.** The published raw files were downloaded from PRIDE, reanalysis by Mascot 2.6.1 with Swiss-Prot database (version 2017\_04, 20,199), and peptides were considered identified if the false discovery rate is set to be less than 1% at peptide level. Specificity was calculated by the peptides area, and the numbers in the bar chart show the number of non-redundant peptides. Entry A-I depleted internal peptides by NHS-beads.<sup>27</sup> Entry J and K depleted internal peptides by HPG-ALD.<sup>26</sup> Entry L, M and N are the triplicate results in this article. Ac-Nt, Free-Nt and Internal represent acetylated protein N-terminal peptides, unmodified protein N-terminal peptides and internal peptides, respectively.

## CONCLUSION

We have succeeded in developing a new N-terminomics method that does not require chemical reactions. This simple and rapid approach is suitable for high-throughput screening with minimal sample amounts. Our TrypN-SCX N-terminomics can enrich protein N-terminal peptides without bias, including peptides containing basic amino acids, with or without N-terminal modifications. This TrypN-SCX approach has great potential for expanding N-terminomics.

# EXPERIMENTAL SECTION

## Materials

Ammonium bicarbonate, tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), sodium deoxycholate (SDC), sodium N-lauroylsarcosinate (SLS), ammonium bicarbonate, tris(2-carboxyethyl)phosphine (TCEP), 2-chloroacetamide (CAA), calcium chloride, ethyl acetate, acetonitrile (ACN), acetic acid, trifluoroacetic acid (TFA) and other chemicals were purchased from Fujifilm Wako (Osaka, Japan). RapiGest was purchased from Waters (Milford, MA). Modified trypsin was from Promega (Madison, MA). TrypN was from Protifi (Huntington, NY). Styrene divinylbenzene (SDB-XC) Empore<sup>TM</sup> disk was purchased from GL Sciences (Tokyo, Japan). Water was purified by a Millipore Milli-Q system (Bedford, MA).

## Cell culture and protein extraction

HEK293T (human embryonic kidney) cells were cultured to 80% confluence in 10-cm diameter dishes. *Escherichia coli* K-12 BW25113 cells were grown to mid-log phase in LB broth with vigorous shaking at 37°C. These cells were collected by centrifugation and resuspended in the PTS lysis buffer containing protease inhibitors (Sigma), 12 mM SDC, 12 mM SLS, 10 mM TCEP, 40 mM CAA in 100 mM Tris buffer (pH 8.5).<sup>38,41</sup> The lysate was vortexed and sonicated on ice for 20 min. The concentration of protein crude extract was determined by means of bicinchoninic acid (BCA) protein assay (ThermoFisher Scientific, Rockford, IL).

## Protein Digestion

For optimization of TrypN digestion conditions, protein pellets were prepared by methanol/chloroform precipitation as described previously,<sup>32</sup> and were dissolved with 0.1% RapiGest in the buffer consisting of 25 mM trimethylammonium acetate, 2 mM CaCl<sub>2</sub> and 0.1 mM MnCl<sub>2</sub> at pH 7.4, followed by TrypN digestion overnight at 55°C according to the manufacturer's protocol. The PTS buffer<sup>38</sup> or the urea buffer consisting of 1 M urea, 25 mM trimethylammonium acetate, 2 mM CaCl<sub>2</sub> and 0.1 mM MnCl<sub>2</sub> at pH 7.4 instead of the RapiGest buffer was also used for the TrypN digestion.

For TrypN digestion after optimization, the cell lysate in the PTS buffer was diluted 10-fold with 10 mM CaCl<sub>2</sub> and digested with TrypN (1: 50 w/w) overnight at 37 °C. Note that TrypN can be replaced with LysargiNase (Merck Millipore, Darmstadt, Germany). In the case of tryptic digestion, the protein solution was digested with LysC (1:50 w/w) for 3 h at 37 °C,

followed by 5-fold dilution with 50 mM ammonium bicarbonate and trypsin digestion (1:50 w/w) overnight at 37 °C. After enzymatic digestion, an equal volume of ethyl acetate was added to each sample solution, and the mixture was acidified with 0.5% trifluoroacetic acid (TFA, final concentration) according to the PTS protocol reported previously.<sup>38</sup> The resulting mixture was shaken for 1 min and centrifuged at 15,700 g for 2 min to separate the ethyl acetate layer. The aqueous layer was collected and desalted by using StageTips with SDB-XC disk membranes (SDB-StageTip).<sup>42</sup> The peptides were quantified by LC-UV at 214 nm using BSA digest as a standard and kept in 80% ACN and 0.5% TFA at -20 °C until use.

### **Peptide fractionation by SCX HPLC**

SCX chromatography was performed using a Prominence HPLC system (Shimadzu, Kyoto, Japan) with a BioIEX SCX column (250 mm × 4.6 mm inner diameter, 5 μm non-porous beads made of poly(styrene-divinylbenzene) modified with sulfonate groups (Agilent, Santa Clara, CA).

For examination of the SCX separation characteristics, 75 μg each of trypsin- and TrypN-digested HEK293T peptides were mixed and directly loaded onto the SCX column at 0.8 mL/min. A mixture of 5 mM potassium phosphate (pH 3.0) and ACN (7:3) was used as SCX buffer A, and a mixture of 500 mM potassium phosphate (pH 3.0) and ACN (7:3) was used as SCX buffer B. Gradient elution was performed as follows: 0% B for 5 min, 0-10% in 20 min, 10-50% in 10 min, 50-100% in 5 min and 100% B for 4 min. Fractions were manually collected at one min intervals for 45 min. After evaporation of the solvent in a SpeedVac SPD121P (ThermoFisher Scientific), fractionated peptides were resuspended in 50 μL of 0.1% TFA and desalted by using SDB-StageTips. One-fourth of each fraction was analyzed by nanoLC/MS/MS using a TripleTOF 5600 (SCIEX, Foster City, CA) as described below.

For gradient SCX fractionation of TrypN-digested HEK293T peptides, 80 μg of digested peptides were analyzed using the SCX HPLC system described above. A mixture of 7.5 mM potassium phosphate (pH 2.6) and ACN (7:3) was used as SCX buffer A, and 350 mM KCl was added to buffer A for SCX buffer B. Gradient elution was performed as follows: 0.5% B for 15 min, 0.5-1% B in 10 min, 1-4% B in 10 min, 4-10% B in 3 min, 10-100% B in 3 min, and 100% B for 5 min. Fractions were manually collected at one min intervals for 50 min. The fractionated peptides desalted by using SDB-StageTips as described above. One-fourth of each fraction for Fr.1-43 and one-tenth of each fraction for Fr.44-50 were analyzed by nanoLC/MS/MS using an Orbitrap Fusion Lumos mass spectrometer (ThermoFisher Scientific) as described below.

### **Enrichment of protein N-terminal peptides by SCX HPLC with isocratic elution**

Enrichment of protein N-terminal peptides from 30 µg of TrypN-digested *E. coli* peptides was performed using the SCX HPLC system under the following isocratic conditions: SCX buffer A was a mixture of 7.5 mM potassium phosphate solution (pH 2.2) containing 10, 12.5 or 15 mM KCl and ACN (7:3), and buffer B was a mixture of 7.5 mM potassium phosphate solution (pH 2.2) containing 500 mM KCl and ACN (7:3). Isocratic elution was performed with 100% A for 30 min and then the system was washed with 100% B. The collected fractions were lyophilized, re-suspended in 50 µL of 0.1% TFA and desalted using SDB-StageTips. Two-thirds of the enriched peptides were analyzed by nanoLC/MS/MS using the Orbitrap Fusion Lumos.

To isolate protein N-terminal peptides from TrypN-digested HEK293T peptides, the digested peptides (80 µg) were analyzed by the SCX HPLC system under isocratic conditions, eluting with a mixture of 7.5 mM potassium phosphate solution (pH 2.2) containing 10 mM KCl and ACN (7:3) for 30 min to collect the desired fraction and desalted using SDB-StageTips as described above. We analyzed one-fourth of the enriched peptides by nanoLC/MS/MS in triplicate using the Orbitrap Fusion Lumos.

### **NanoLC/MS/MS analysis**

NanoLC/MS/MS analyses were performed on a TripleTOF 5600 mass spectrometer or an Orbitrap Fusion LUMOS mass spectrometer, connected to a Thermo Ultimate 3000 pump and an HTC-PAL autosampler (CTC Analytics, Zwingen, Switzerland). Peptides were separated on self-pulled needle columns (150 mm length × 100 µm ID, 6 µm opening) packed with Reprosil-Pur 120 C18-AQ 3 µm reversed-phase material (Dr. Maisch, Ammerbuch, Germany). The injection volume was 5 µL and the flow rate was 500 nL/min. The mobile phases were (A) 0.5% acetic acid and (B) 0.5% acetic acid and 80% ACN. For TripleTOF 5600 analysis, gradient elution was performed as follows: 12–40% B in 20 min, 45–100% B in 1 min, 100% B for 5 min. For Orbitrap analysis, gradient elution of fractionated samples was performed as follows: 12–40% B in 15 min, 40–100% B in 1 min, 100% B for 5 min. For protein N-terminal peptide-enriched samples, gradient elution was performed as follows: 10–40% B in 100 min, 40–100% B in 10 min, 100% B for 10 min. For proteome samples, gradient elution was performed as follows: 8–35% B in 100 min, 35–50% B in 10 min, 5–100% B in 5 min, 100% B for 10 min. Spray voltages of 2300 V in the TripleTOF 5600 system and 2400 V in the Orbitrap system were applied. The mass scan range of the TripleTOF 5600 system was  $m/z$  300–1500, and the top ten precursor ions were selected in each MS scan for subsequent MS/MS scans. The mass

scan range for the Orbitrap system was  $m/z$  300–1500, with an automatic gain control value of  $1.00e + 06$ , a maximum injection time of 50 ms and detection at a mass resolution of 60,000 at  $m/z$  200 in the orbitrap analyzer. The top ten precursor ions with +2, +3 or +4 charge were selected in each MS scan for subsequent MS/MS scans with an automatic gain control value of  $5.00e + 04$  and a maximum injection time of 300 ms. Dynamic exclusion was set for 25 s with a 10 ppm gate. The normalized HCD was set to be 30, with detection at a mass resolution of 15,000 at  $m/z$  200 in the Orbitrap analyzer. A lock mass (445.1200025) function was used to obtain constant mass accuracy during the gradient.

### **Proteomics Data Processing for HEK293T and *E. coli***

Two peak lists in “.mgf” and “.apl” formats were generated from the MS/MS spectra by MaxQuant 1.5.8.0 33. The peptides and proteins were identified by Mascot v2.6.1 (Matrix Science, London, U.K.) against the Swiss-Prot database (version 2017\_4, 20,199 sequences) or the *E. coli* K-12 MG1665 protein sequence database 34 with a precursor mass tolerance of 20 ppm (TripleTOF 5600) or 10 ppm (Orbitrap), a fragment ion mass tolerance of 0.1 Da (TripleTOF 5600) or 20 ppm (Orbitrap), TrypN/trypsin specificity allowing for up to 4 missed cleavages for TrypN/trypsin mixed proteolytic peptides and strict TrypN specificity allowing for up to 2 missed cleavages for TrypN-digested peptides. Carbamidomethylation of cysteine was set as a fixed modification, and methionine oxidation and protein N-terminal acetylation were allowed as variable modifications. False discovery rates at a peptide level of less than 1% were applied for peptide identification based on a target-decoy approach.

### **Preparation of *in-silico* digested *E. coli* Protein N-terminal peptide list**

The list of protein N-terminal peptides obtained by *in silico* TrypN-digestion of *E. coli* proteome was prepared using *E. coli* K-12 MG1665 protein sequence database (4,316 sequences).<sup>43</sup> First, a strict TrypN-specific cleavage at protein N-terminus was performed and missed cleavage was prohibited. Second, peptides shorter than 7 amino acids in length and the redundant sequences were removed. A total of 1,494 protein N-termini were obtained.

## **SUPPLEMENTAL TABLES**

Supplemental Table 1. Identification list of TrypN- and trypsin-mixed HEK293T peptides used in Figure 1-2.

Supplemental Table 2. Identification list of TrypN-digested HEK293T peptides used in Figure 1-4, 1-5.

Supplemental Table 3. Identification list of TrypN-digested *E. coli* peptides used in Table 1-1, Figure 1-6.

Supplemental Table 4. Identification list of TrypN-digested HEK293T peptides used in Table 1-3 and Figure 1-8.

Supplemental Table 1-4 are available free of charge at ASBMB website (<https://www.mcponline.org/cms/10.1074/mcp.TIR120.002148/attachment/143946ff-5f35-433e-8c14-c4e8069b9ea8/mmc1.xlsx>).

## REFERENCES

- (1) Kim, M. S.; Pinto, S. M.; Getnet, D.; Nirujogi, R. S.; Manda, S. S.; Chaerkady, R.; Madugundu, A. K.; Kelkar, D. S.; Isserlin, R.; Jain, S.; Thomas, J. K.; Muthusamy, B.; Leal-Rojas, P.; Kumar, P.; Sahasrabudde, N. A.; Balakrishnan, L.; Advani, J.; George, B.; Renuse, S.; Selvan, L. D. N., et al. *Nature* **2014**, *509*, 575-581.
- (2) Wilhelm, M.; Schlegl, J.; Hahne, H.; Gholami, A. M.; Lieberenz, M.; Savitski, M. M.; Ziegler, E.; Butzmann, L.; Gessulat, S.; Marx, H.; Mathieson, T.; Lemeer, S.; Schnatbaum, K.; Reimer, U.; Wenschuh, H.; Mollenhauer, M.; Slotta-Huspenina, J.; Boese, J. H.; Bantscheff, M.; Gerstmair, A., et al. *Nature* **2014**, *509*, 582-587.
- (3) Iwasaki, M.; Sugiyama, N.; Tanaka, N.; Ishihama, Y. *J. Chromatogr. A* **2012**, *1228*, 292-297.
- (4) Wolters, D. A.; Washburn, M. P.; Yates, J. R. *Anal. Chem.* **2001**, *73*, 5683-5690.
- (5) Masuda, T.; Sugiyama, N.; Tomita, M.; Ishihama, Y. *Anal. Chem.* **2011**, *83*, 7698-7703.
- (6) Thingholm, T. E.; Jorgensen, T. J. D.; Jensen, O. N.; Larsen, M. R. *Nat. Protoc.* **2006**, *1*, 1929-1935.
- (7) Zhang, H. M.; Guo, T. N.; Li, X.; Datta, A.; Park, J. E.; Yang, J.; Lim, S. K.; Tam, J. P.; Sze, S. K. *Mol. Cell. Proteomics* **2010**, *9*, 635-647.
- (8) Bouwmeester, R.; Gabriels, R.; Hulstaert, N.; Martens, L.; Degroeve, S. *bioRxiv* **2020**, 2020.2003.2028.013003.
- (9) Gabriels, R.; Martens, L.; Degroeve, S. *Nucleic Acids Res.* **2019**, *47*, W295-W299.
- (10) Nakahigashi, K.; Takai, Y.; Kimura, M.; Abe, N.; Nakayashiki, T.; Shiwa, Y.; Yoshikawa, H.; Wanner, B. L.; Ishihama, Y.; Mori, H. *DNA Res.* **2016**, *23*, 193-201.
- (11) Ingolia, N. T. *Nat. Rev. Genet.* **2014**, *15*, 205-213.
- (12) Van Damme, P.; Gawron, D.; Van Crielinge, W.; Menschaert, G. *Mol. Cell. Proteomics* **2014**, *13*, 1245-1261.



- (13) Hwang, C. S.; Shemorry, A.; Varshavsky, A. *Science* **2010**, *327*, 973-977.
- (14) Starheim, K. K.; Gevaert, K.; Arnesen, T. *Trends Biochem. Sci.* **2012**, *37*, 152-161.
- (15) Mahrus, S.; Trinidad, J. C.; Barkan, D. T.; Sali, A.; Burlingame, A. L.; Wells, J. A. *Cell* **2008**, *134*, 866-876.
- (16) McDonald, L.; Robertson, D. H. L.; Hurst, J. L.; Beynon, R. J. *Nat. Methods* **2005**, *2*, 955-957.
- (17) Leitner, A. *Anal. Chim. Acta* **2018**, *1000*, 2-19.
- (18) Klein, T.; Eckhard, U.; Dufour, A.; Solis, N.; Overall, C. M. *Chem. Rev.* **2018**, *118*, 1137-1168.
- (19) Xu, G.; Shin, S. B.; Jaffrey, S. R. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 19310-19315.
- (20) Varland, S.; Osberg, C.; Arnesen, T. *Proteomics* **2015**, *15*, 2385-2401.
- (21) Lai, Z. W.; Petrera, A.; Schilling, O. *Curr. Opin. Chem. Biol.* **2015**, *24*, 71-79.
- (22) Gevaert, K.; Goethals, M.; Martens, L.; Van Damme, J.; Staes, A.; Thomas, G. R.; Vandekerckhove, J. *Nat. Biotechnol.* **2003**, *21*, 566-569.
- (23) 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. *Nat. Biotechnol.* **2010**, *28*, 281-288.
- (24) Venne, A. S.; Solari, F. A.; Faden, F.; Paretti, T.; Dissmeyer, N.; Zahedi, R. P. *Proteomics* **2015**, *15*, 2458-2469.
- (25) Chen, L. F.; Shan, Y. C.; Weng, Y. J.; Sui, Z. G.; Zhang, X. D.; Liang, Z.; Zhang, L. H.; Zhang, Y. K. *Anal. Chem.* **2016**, *88*, 8390-8395.
- (26) Na, C. H.; Barbhuiya, M. A.; Kim, M. S.; Verbruggen, S.; Eacker, S. M.; Pletnikova, O.; Troncoso, J. C.; Halushka, M. K.; Menschaert, G.; Overall, C. M.; Pandey, A. *Genome Res.* **2018**, *28*, 25-36.
- (27) Yeom, J.; Ju, S.; Choi, Y.; Paek, E.; Lee, C. *Sci. Rep-Uk* **2017**, *7*.

- (28) Adachi, J.; Hashiguchi, K.; Nagano, M.; Sato, M.; Sato, A.; Fukamizu, K.; Ishihama, Y.; Tomonaga, T. *Anal. Chem.* **2016**, *88*, 7899-7903.
- (29) Essader, A. S.; Cargile, B. J.; Bundy, J. L.; Stephenson, J. L., Jr. *Proteomics* **2005**, *5*, 24-34.
- (30) Alpert, A. J.; Petritis, K.; Kangas, L.; Smith, R. D.; Mechtler, K.; Mitulovic, G.; Mohammed, S.; Heck, A. J. R. *Anal. Chem.* **2010**, *82*, 5253-5259.
- (31) Helbig, A. O.; Gauci, S.; Raijmakers, R.; van Breukelen, B.; Slijper, M.; Mohammed, S.; Heck, A. J. R. *Mol. Cell. Proteomics* **2010**, *9*, 928-939.
- (32) Gauci, S.; Helbig, A. O.; Slijper, M.; Krijgsveld, J.; Heck, A. J.; Mohammed, S. *Anal. Chem.* **2009**, *81*, 4493-4501.
- (33) Wilson, J. P.; Ipsaro, J. J.; Del Giudice, S. N.; Turna, N. S.; Gauss, C. M.; Dusenbury, K. H.; Marquart, K.; Rivera, K. D.; Pappin, D. J. *J. Proteome Res.* **2020**, *19*, 1459-1469.
- (34) Tallant, C.; Garcia-Castellanos, R.; Seco, J.; Baumann, U.; Gomis-Ruth, F. X. *J. Biol. Chem.* **2006**, *281*, 17920-17928.
- (35) Huesgen, P. F.; Lange, P. F.; Rogers, L. D.; Solis, N.; Eckhard, U.; Kleifeld, O.; Goulas, T.; Gomis-Ruth, F. X.; Overall, C. M. *Nature Methods* **2015**, *12*, 55-58.
- (36) Koneru, L.; Ksiazek, M.; Waligorska, I.; Straczek, A.; Lukasik, M.; Madej, M.; Thogersen, I. B.; Enghild, J. J.; Potempa, J. *Biol. Chem.* **2017**, *398*, 395-409.
- (37) Tsiatsiani, L.; Giansanti, P.; Scheltema, R. A.; van den Toorn, H.; Overall, C. M.; Altelaar, A. F. M.; Heck, A. J. R. *J. Proteome Res.* **2017**, *16*, 852-861.
- (38) Masuda, T.; Tomita, M.; Ishihama, Y. *J. Proteome Res.* **2008**, *7*, 731-740.
- (39) Raijmakers, R.; Neerincx, P.; Mohammed, S.; Heck, A. J. *Chem Commun (Camb)* **2010**, *46*, 8827-8829.
- (40) Gussakovsky, D.; Neustaeter, H.; Spicer, V.; Krokhin, O. V. *Anal. Chem. (Wash.)* **2017**, *89*, 11795-11802.

(41) Humphrey, S. J.; Azimifar, S. B.; Mann, M. *Nat. Biotechnol.* **2015**, *33*, 990-U142.

(42) Rappsilber, J.; Ishihama, Y.; Mann, M. *Anal. Chem.* **2003**, *75*, 663-670.

(43) Riley, M.; Abe, T.; Arnaud, M. B.; Berlyn, M. K. B.; Blattner, F. R.; Chaudhuri, R. R.; Glasner, J. D.; Horiuchi, T.; Keseler, I. M.; Kosuge, T.; Mori, H.; Perna, N. T.; Plunkett, G.; Rudd, K. E.; Serres, M. H.; Thomas, G. H.; Thomson, N. R.; Wishart, D.; Wanner, B. L. *Nucleic Acids Res.* **2006**, *34*, 1-9.