Analyst

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/analyst

Analyst

ARTICLE

Controlling Nonspecific Trypsin Cleavages in LC−MS/MS-Based Shotgun Proteomics Using Optimized Experimental Conditions

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Pan Fang,[a] Mingqi Liu,[a] Yu Xue,[b] Jun Yao,[a] Yang Zhang,[a] Huali Shen,[a,c,*] and Pengyuan Yang[a,b,*]

Trypsin has traditionally been used for enzymatic digestion during sample preparation in shotgun proteomics. The stringent specificity of trypsin is essential for accurate protein identification and quantification. However, nonspecific trypsin cleavages are often observed in LC-MS/MS-based shotgun proteomics. To explore the extent of nonspecific trypsin cleavages, a series of biological systems including standard protein mixtures, saccharomyces cerevisiae, human serum, human cancer cell lines and mouse brain were examined. We found that nonspecific trypsin cleavages commonly occurred in various trypsin digested samples with high frequency. To control these nonspecific trypsin cleavages, we optimized fundamental parameters during sample preparation with mouse brain homogenates. These parameters included denaturing agents and protein storage time, trypsin type, enzyme-to-substrate ratio, as well as protein concentration during digestion. The optimized experimental conditions significantly decreased the ratio of partially tryptic peptides in total identifications from 28.4% to 2.8%. Furthermore, the optimized digestion protocol was applied to the study of N-glycoproteomics, and the proportions of partially tryptic peptides in enriched mixtures were also sharply reduced. Our work demonstrates the importance of controlling nonspecific trypsin cleavages in both shotgun proteomics and glycoproteomics and provides a better understanding and standardization for routine proteomics sample treatment.

Introduction

Trypsin plays an important role in mass spectrometry (MS)-based shotgun proteomics as the single most utilized protease. The preference for trypsin stems from its stability under a wide range of conditions, its high activity and high substrate specificity. Trypsin specifically cleaves at the carboxyl side of lysine and arginine residues, resulting in uniform peptides with a length of 10-12 residues that can be analyzed by mass spectrometry. Furthermore, the resulting C-terminal basic residues not only improve ionization in MS-analysis but also contribute to the generation of an abundant y-ion series in tandem mass spectrometry experiments to facilitate confident identification.

Although trypsin is a very specific protease,[1] a significant number of nonspecific tryptic peptides are often confidently identified.[2-7] Identifying peptides derived from nontryptic pathways may increase sequence coverage, but searching for all potential nonspecific peptides is a challenge for shotgun proteomics data from complex samples.[6] Interpretation of MS/MS spectra using nonspecific search mode is time consuming and may not improve overall peptide identification because of the possible sensitivity reduction from search space expansion. Moreover, the vast majority of identified nontryptic peptides in nonspecific searches are often considered as less confident identifications than fully tryptic peptides or to be incorrect.[8, 9] Generally, it is less favorable to search MS/MS spectra without enzyme constraint. On the other hand, irregular cleavages present in peptide populations can substantially affect the accuracy of protein identification and quantification, reproducibility of mass spectrometry analysis.[10, 11] Therefore, controlling nonspecific trypsin cleavages in LC−MS/MS-based shotgun proteomics is quite important and necessary.

Many of nonspecific tryptic peptides are thought to be signaling peptides or to be formed during sample preparation.[7] These nonspecific cleavage events can be partly explained by several possible reasons. Firstly, the presence of numerous endogenous proteases in biological samples such as plasma or cereal seeds,[12, 13] may result in protein degradation. These pre-existing proteolytic fragments in the original sample are typically non-specific. Sometimes nontryptic peptides are considered as manifestations of proteolytic events, either in vivo or in vitro.[14] In addition, the quality of commercially available trypsins has a great impact on cleavage specificity.[15-17] The cleavage specificity of various trypsin types are often influenced by chymotrypsin contamination,[18] pseudotrypsin activity due to partial autolysis of trypsin,[19] inappropriate storage,[20] other contaminating proteins present in trypsin preparations as well as the origin of the trypsin.[21] Thirdly, digestion conditions, *e.g.* temperature, pH, time, enzyme-to-substrate ratio, detergents and digestion solvents, are directly related with the digestion processes and thereby may have a large effect on digestion efficiency and specificity.[5, 17] Finally, in-source fragmentation (ISF) have been

reported to be a major source of partially tryptic peptides in LC-MS/MS-based proteomics.[12]

In this study, we examined the extent of nonspecific trypsin cleavages within a series of biological systems and revealed the great difference in frequency of nonspecific trypsin cleavages among various samples. To control these nonspecific trypsin cleavages, we optimized major experimental parameters during sample preparation. The optimized experimental conditions were further applied to N-glycoproteomics to control the amounts of nonspecific N-glycopeptides.

Experimental sections

Protein samples

Four standard proteins, including β-Casein from bovine milk, ribonuclease B from bovine pancreas, fetuin from fetal calf serum and bovine serum albumin, were obtained from Sigma-Aldrich (USA). Each standard protein in equal amounts was dissolved in the lysis buffer consisting of 4% SDS, 0.1 M Tris-HCl, pH 7.6 and then mixed for further analysis. Saccharomyces cerevisiae S288c were purchased from Thermo Fisher (USA). The yeast cells were harvested by centrifugation in the exponential growth phase (O.D. was about 1.0 at 600 nm) after growing in YPD overnight. Then they were suspended in the same lysis buffer with standard protein mixture and lysed using the MiniBeadbeater (Biospec) at maximum speed, four cycles of 40 s each, with 2 min pauses between cycles to minimize protein degradation. After centrifugation at 14,000 rpm for 40 min, the supernatants were collected for further analysis. Normal human serum samples were supplied by Zhongshan Hospital and collected with informed consent under an institutional review board approved protocol. Human metastatic cancer cell lines MHCC97L were obtained from the Liver Cancer Institute of Zhongshan Hospital affiliated to Fudan University (Shanghai, China). Cell pellets were suspended in the same lysis buffer after centrifugation and briefly sonicated to reduce viscosity of the lysate (3 × 20 s). Mouse brain tissues dissected from adult C57BL/6J male mice (3-12 months) were purchased from SLRC LABORATORY ANIMAL (Shanghai, China). The lysis of solid tissue samples was facilitated by homogenization using High-throughout Tissue Grinding machine (ONEBIO, China) at 65 Hz for 60 s. All samples were dissolved in the same lysis buffer consisting of 4% SDS, 0.1 M Tris-HCl, pH 7.6. Additionally, mouse brain tissue was also extracted by a lysis buffer containing 8 M urea, 2 M thiourea, 1 mM phenylmethanesulfonyl fluoride (PMSF) and a mixture of protease inhibitor (Roche, Switzerland). The protein concentration was determined by BCA (Pierce, Rockford, IL) or Bradford method depending on the lysis buffer. Extracted proteins were stored in −80 ℃ until used.

Protein digestion

Proteins were reduced in 10 mM DTT at 37 ℃ for 60 min, and alkylated in dark by 20 mM iodoacetamide (IAA) at room temperature for 30 min. For proteins dissolved in the lysis buffer containing SDS, 6 volumes of acetone was added to precipitate proteins at −20 ℃ for at least 3 h. Afterwards, the precipitates were resuspended in denaturing buffers (8 M urea in 50 mM ammonium bicarbonate) followed by a tenfold dilution with 50 mM ammonium bicarbonate. For proteins extracted by urea, a tenfold dilution with 50 mM ammonium bicarbonate was directly performed to reduce the concentration of urea to below 1 M after carbamidomethylation. Next, $CaCl_2$ was added into the diluted samples to a final concentration of 1 mM. Then, the samples were digested using trypsin at a trypsin-to-protein ratio of 1:50 (w/w) for 12 h at 37 ℃. The digested products were acidified to below pH 3.0 with trifluoroacetic acid (TFA) followed by centrifugation at 14,000 g for 20 min before desalting the supernatant using C18 column (Sep-Pak Vac C18, Waters Corporation) according to manufacturer's instructions. Finally, the purified peptides were dried by vacuum centrifugation and stored at −80 ℃ for further use.

Enrichment of N-glycopeptides

The digested peptides were enriched with ZIC-HILIC material using the procedure reported by Ma with brief modification. Approximately 100 μg peptides were dissolved in 20 μL loading buffer of 80% (v/v) acetonitrile and 0.5% (v/v) formic acid and loaded onto an in-house zwitterionic Zic-HILIC tip containing 2 mg of Zic-HILIC particles (10 μm, 200 Å; Sequant/Merck) packed onto a C8 disk (Empore). The flow-through was collected and passed back through the tip for additional four times. The ZIC-HILIC tip was washed with 100 μL wash buffer of 80% (v/v) acetonitrile and 1% (v/v) formic acid for five times. The bound peptides were eluted with 1% (v/v) formic acid for three times and dried by vacuum centrifugation. The elute was further deglycosylated with PNGase F in 50 mM ammonium bicarbonate overnight at 37 ℃ and dried again.

LC-MS/MS analysis

LC-MS/MS experiments were performed on an HPLC system composed of two LC-20AD nano-flow LC pumps, an SIL-20AC auto-sampler and an LC-20AB micro-flow LC pump (all Shimadzu, Tokyo, Japan) connected to an LTQ-Orbitrap mass spectrometer (ThermoFisher, San Jose, CA). Sample was loaded onto a CAPTRAP column (0.5 x 2 mm, MICHROM Bioresources, Auburn, CA) in 5 min at a flow rate of 10 μL/min. The sample was subsequently separated by a C18 reverse-phase column (0.075 x 150 mm, packed with 3 μm Aeris C18 particles, Phenomenex) at a flow rate of 300 nL/min. The mobile phases were 0.1% formic acid as the loading phase and 4% acetonitrile in 0.1% formic acid (phase A) and 96% acetonitrile with 0.1% formic acid (phase B). To achieve proper separation, a 60 min linear gradient from 5% to 40% phase B was employed. The separated sample was introduced into the mass spectrometer via nano electrospray. The spray voltage was set at 2 kV and the heated capillary at 180 ℃. The mass spectrometer was operated in data-dependent mode and each cycle of duty consisted one full MS survey scan at the mass range 300~1600 Da with resolution power of 100,000 using the Or

bitrap section, followed by MS/MS experiments for 10 strongest peaks using the LTQ section. The AGC expectation during full-MS and MS/MS were 1000000 and 10000, respectively. Peptides were fragmented in the LTQ section using collision-induced dissociation with helium and the normalized collision energy value set at 35%. Only 2+ and 3+ peaks were selected for MS/MS run and previously fragmented peptides were excluded for 60 s.

Data analysis

Raw data files were searched using pFind studio 2.8[23] against SwissProt_2015_03 database (Homo sapiens species or bovine or Saccharomyces cerevisiae or mouse) with static modification of Carbamidomethyl (Cys), dynamic modification of Oxidation (Met), and Acetylation (N-Terminal). Semi-trypsin or full trypsin was selected as the enzyme and two missed cleavages were allowed. The mass tolerance was set to 20 ppm for the precursor ions and 0.5 Da for the fragment ions. A false discovery rate (FDR) of 1% was estimated using concatenated forward-reverse database search at the peptide-spectrum match (PSM) level.

Results and discussion

Overview of experimental design

Figure 1 exhibits the schematic diagram of experimental design in this study. Nonspecific trypsin cleavages are often introduced in the course of sample preparation in shotgun proteomics. To explore the extent of nonspecific trypsin cleavages within different biological systems, we selected five representative specimens including standard protein mixture, yeast whole-cell lysate, human serum, human cancer cell extracts and mouse brain homogenates. Afterwards, to control the nonspecific cleavage events, four parameters were investigated respectively: denaturing agents (SDS or urea) and storage time (0 month or 7 months), trypsin types (Sequencing Grade Modified Trypsin, Mass spectrometry Grade Modified Trypsin or unmodified TPCK treated trypsin), enzyme-to-substrate ratio (1:5, 1:20, 1:50 or 1:100), as well as protein concentration during digestion (0.5 μg/μL, 0.25 μg/μL, 0.1 μg/μL or 0.05 μg/μL). The digestion conditions and corresponding nonspecific tryptic peptides percentages are summarized in Table 1 and the detailed information for each digestion are listed in supplementary table. Finally, the optimized digestion protocol was further applied to control nonspecific trypsin cleavages in N-glycoproteomic study.

The extent of nonspecific trypsin cleavages within different biological systems

To determine the extent of nonspecific trypsin cleavages in representative specimens, we performed separate trypsin digestions on the five samples under the same digestion conditions. Digestion products were analyzed by LC-MS/MS in triplicate and the resulting MS/MS spectra were searched with a partially tryptic setting to pick up fully tryptic peptides and partially tryptic ones. Peptides without tryptic termini were not considered in this paper because of the low identification rates in database search without enzyme constraint (data not shown). As shown in Figure 2A, a substantial number of partially tryptic peptides were identified in all sample types. In both unique peptide level and spectra level, the percentage of observed partially tryptic peptides among different biological sample systems differed greatly. For standard proteins, partially tryptic peptides represented a large portion (~50%) of the total peptide identifications (Figure 2B). Nevertheless, for each standard protein, the proportions of partially tryptic peptides varied from 40% to 86% (Supplementary Figure 1). The great differences between standard proteins were likely due to the characteristics of proteins used.

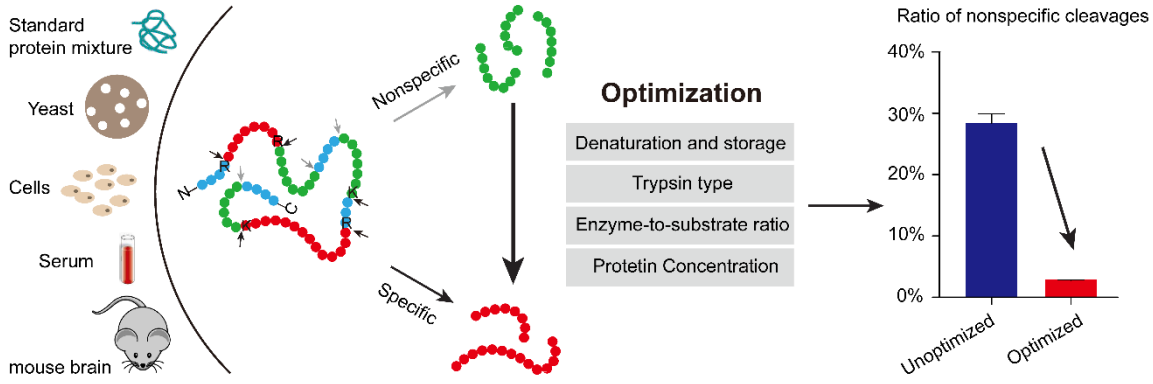

Figure 1. Schematic diagram of experimental design. Digesting proteins from various biological samples is usually performed using trypsin due to its high specific cleavage activity at C-terminal to lysine or arginine. But in actual situation, nonspecific trypsin cleavages are concurrently generated with specific species due to multiple potential factors. To control these nonspecific trypsin cleavages, four digest conditions were selected and optimized to reduce the ratio of nonspecific tryptic peptides in digest products.

Table 1. Summary of the Digestion Conditions and Corresponding Nonspecific Tryptic Peptides Percentage

No.	Samples	Trypsin digest conditions					Peptides/%[d]	Spectra/%[e]
		Denaturing agents	Storage time/month	Trypsin type	Enzyme-to-substrate ratio	Concentration/μg/μL		
1	Standard prot	SDS	0	1	1:50	0.5	47.3	41.1
2	Yeast	SDS	0	1	1:50	0.5	18.7	16.9
3	Serum	SDS	0	1	1:50	0.5	19.9	13.9
4	Cell lines	SDS	0	1	1:50	0.5	6.7	6.0
5	Mouse brain	SDS	0	1	1:50	0.5	5.5	5.0
6		SDS	0	1[a]	1:50	0.25	3.5	2.9
7		SDS	7	1	1:50	0.25	3.6	2.9
8		Urea	0	1	1:50	0.25	5.1	4.4
9		Urea	7	1	1:50	0.25	7.8	6.5
10		SDS	8	1	1:50	0.25	4.6	3.6
11		SDS	8	2[b]	1:50	0.25	3.3	2.4
12		SDS	8	3[c]	1:50	0.25	5.6	4.5
13	Mouse brain	SDS	8	1	1:5	0.5	10.9	9.5
14		SDS	8	1	1:20	0.5	5.4	4.3
15		SDS	8	1	1:50	0.5	4.6	3.6
16		SDS	8	1	1:100	0.5	3.9	3.1
17		SDS	8	1	1:50	0.5	4.2	3.1
18		SDS	8	1	1:50	0.25	4.2	2.9
19		SDS	8	1	1:50	0.1	3.8	2.8
20		SDS	8	1	1:50	0.05	3.0	2.5
21	Optimized	SDS	0	2	1:50	0.05	2.8	2.4
22	Unoptimized	Urea	7	3	1:5	0.5	28.4	27.1

[a] sequencing grade modified trypsin. [b] mass spectrometry grade modified trypsin. [c] unmodified trypsin. [d,e] average percentage of partially tryptic peptides

The proportions of nonspecific cleavage products were smaller in complex samples. However, partially tryptic peptides still accounted for 18.7% and 19.9% of all identified sequences in yeast and serum, respectively. The high level of nonspecific trypsin cleavage in serum can be attributed to the nature of serum sample which contains numerous secreted protease. Meanwhile, serum has much greater dynamic range of protein concentrations compared with cell and tissue samples, where 6.7% and 5.5% of partially tryptic peptides were observed from cancer cell lines and mouse brain, respectively.

Our data showed that with the increase of sample complexity, the level of nonspecific trypsin cleavage decreased. The detectability of partially tryptic peptides was largely dependent on the overall sample complexity due to the much lower abundance of partially tryptic sequences compared to fully tryptic peptides. In standard proteins, peptide truncation might be caused by various chemical mechanisms during sample preparation, storage or mass spectrometer detecting, and are not the result of proteases in the sample or other biological processes. In-source fragmentation has been reported to be one main reason for the large proportion of partially tryptic peptides in standard proteins because of the increased probability of low-abundance ions to be selected for MS/MS fragmentation. For yeast proteome, sequencing of low abundant partially tryptic peptides was easier compared to other complex samples due to its lower complexity. In human serum digests, the identified partially tryptic peptides were mostly from highly abundant proteins, for example, Serum albumin, Lactotransferrin, Apolipoprotein A-II (Supplementary Table 2). With the plasma protein concentrations in the HPP database as a reference,[24] we checked the absolute concentrations of proteins occurring nonspecific trypsin cleavage, seventy proteins were found in the HPP database. We listed the concentrations of these proteins in supplementary table 2_Human serum and further drew a chart of protein concentration distribution (supplementary Fig. 2). The chart shows the concentrations of proteins found in the HPP database spanned six orders of magnitude, ranging from 1.6×10^6 ng/ml for serum albumin (P02768) down to 1.4 ng/ml for Apolipoprotein (a) (P08519). The concentrations of seventy percent of these proteins were above 10^4 ng/mL, which indicated that the partially tryptic peptides in human serum were mainly from relatively high abundant proteins. These partially tryptic peptides of highly abundant proteins were competitive in abundance to those tryptic peptides from mid-or low-abundant proteins, thus making these partially tryptic peptides more detectable. In the case of human cancer cell extracts and mouse brain lysates, only approximate 5% of nonspecific cleavage products were identified. Due to the limit of dynamic range of mass spectrometer, in these highly complex samples, only the highest intensity ions are fragmented, whereas ions of low intensity were unselected even though their signal suited with the nominal dynamic range of the used instrument. Overall, the

extent of detectable partially tryptic peptides were dependent on the characteristics of the sample.

The cleavage patterns of nonspecific tryptic peptides from yeast, serum, cell lines and mouse brain were investigated separately (the detailed information of nonspecific tryptic peptides are listed in Supplementary Table 2). Using Icelogo,[25] we analyzed a ±5 residue region around the nonspecific cleavage sites (Position 5). The analysis revealed high frequency of phenylalanine (F) and tyrosine (Y) at left side of the cleaved peptide bond in all samples, suggesting chymotryptic activity (Figure 3). An apparent enrichment of aspartic (D) and asparagine (N) at left side of the cleaved peptide bond in serum, cell lines and mouse brain was also observed. Trypsin has previously been reported to cleave polypeptides at the C-terminus of asparagine.[26] However, in the case of yeast, the cleavage pattern was different from the other three samples with no significant enrichment of D or N. In addition, the frequency of a specific amino acid at the cleavage sites differed greatly among the samples, indicating unique nonspecific cleavage patterns in different biological sample systems.

In summary, nonspecific cleavage events commonly occurred in diverse biological samples with different extents. So, it is necessary to improve cleavage specificity of trypsin through optimizing conditions during sample preparation.

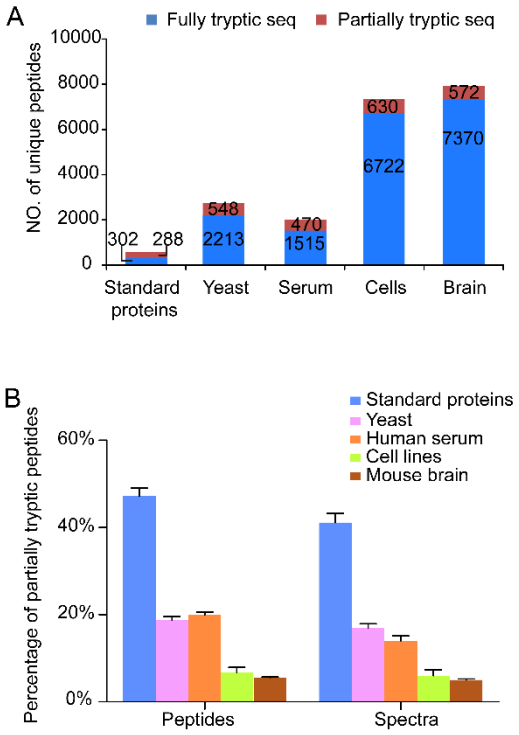

Figure 2. Partially tryptic peptides are observed at substantial levels for different biological sample systems. (A) Fully and partially tryptic sequences identified in the five types of samples at the unique peptide level. (B) Percentage of partially tryptic peptides at the unique peptide and spectra level for the five types of samples. Error bar represents mean ± standard deviation across three replicates.

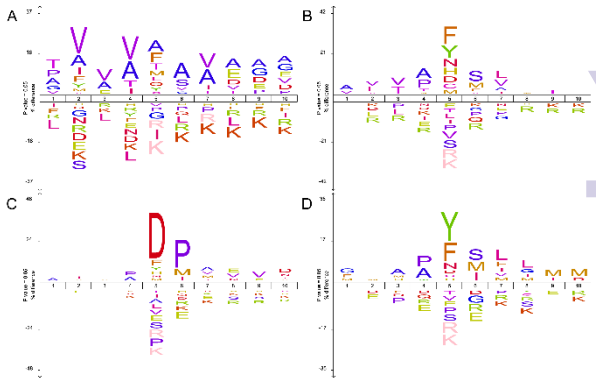

Figure 3. Icelogo analysis of nonspecific cleavage events within yeast (A), serum (B), cell lines (C) and mouse brain (D). Position 5 is at left side of the nonspecific cleavage sites.

Optimization of digestion conditions for controlling nonspecific trypsin cleavages

To explore key factors that affect cleavage specificity of try[illegible], we evaluated several major parameters including denaturing agents and storage time, trypsin type, enzyme-to-substrate r[illegible] and protein concentration during digesti

Denaturing agents and protein storage time

To cleave the embedded regions of native proteins effectively, it is essential to unfold secondary and tertiary structures to e[illegible]pose embedded cleavage sites. Denaturing agents are usual[illegible] used to facilitate the proteolytic digestion. Initially, we eva[illegible]ed two kinds of widely used denaturing agents: SDS and ure[illegible] Meanwhile, storage time of proteins after extraction from fre[illegible]h tissues was also investigated. As shown in Figure 4A, the proportions of partially tryptic peptides from proteins denatured [illegible]y SDS is lower than that by urea, no matter the proteins were stored for 0 month or 7 month. This is probably because th[illegible]t SDS is more efficient in disrupting the advanced structure o[illegible] proteins such as various endogenous proteases present in co[illegible]plex biological samples. However, it is noted that in-solution depletion of SDS is difficult and the organic precipitation method could induce inevitable sample loss. In addition, w[illegible] storage time was prolonged, the proportions of nonspecific tryptic peptides increased both in the case of SDS and urea but the increment for SDS digests was much lower than that [illegible] urea. The results suggested that SDS is a better option for co[illegible]trolling nonspecific trypsin cleavages and the storage ti[illegible] should not be too long.

Trypsin type

Trypsin is available from many sources in different qualiti[illegible] and costs. The detailed information of three types of comme[illegible]cial trypsin often used in our laboratory are listed in supplemen[illegible]tary table 3, with priority of trypsin 1 (sequencing grade mod[illegible]fied trypsin) and trypsin 2 (mass spectrometry grade modified trypsin) for in-gel digestion. Trypsin 3 (unmodified tryp[illegible]) was usually used for preliminary tests or digestion of protei[illegible]

in large amounts because of its low price. In comparison of specificity for the three trypsins, trypsin 2 showed the best results with a percentage of 96.7% for fully tryptic peptides (Figure 4B). The cleavage specificity of trypsin 1 was lower than trypsin 2. The partially tryptic peptides accounted 5.6% in trypsin 3. The particularly high nonspecific cleavage activity of trypsin 3 was probably due to the contamination of chymotrypsin present in trypsin preparations or the pseudotrypsin activity from auto-proteolytic trypsin in its unmodified state. It has been demonstrated that unmodified trypsin is subject to auto-proteolysis, generating fragments that interfere the protein sequencing. In addition, auto-proteolysis can result in the generation of pseudotrypsin, which has been shown to exhibit chymotrypsin-like specificity.[19] To reduce side reactions, trypsin is usually modified by reductive methylation to retard autolytic digestion, but incompletely. Overall, modified trypsins provide greater control over nonspecific trypsin cleavage than unmodified ones. So, the quality of trypsin is a valuable factor in controlling nonspecific trypsin cleavages in shotgun proteomics.

Enzyme-to-substrate ratio

To assess the influence of enzyme-to-substrate ratio on nonspecific trypsin cleavage, equal amounts of proteins were digested with four trypsin-to-protein ratios of 1:5, 1:20, 1:50 and 1:100. As shown in Figure 4 C, The percentages of partially tryptic peptides were in the order of 1:5 > 1:20 > 1:50 > 1:100. The results indicated that nonspecific trypsin cleavages increased with increasing amounts of enzyme. Although the ratio of partially tryptic peptides was the lowest in 1:100, the missed cleavages in it was the highest (Supplementary figure 2), so we still chose 1:50 as the optimized condition. Excess enzyme substantially decreased the cleavage specificity of trypsin (only ~90% in the 1:5 digests), suggesting that enzyme-to-substrate ratio is also a significant factor that influences how specific a trypsin is. As the autolytic rate of most exocrine proteases, such as trypsin, is concentration dependent, high enzyme concentration may result in enzyme being more likely to encounter enzyme instead of substrates and thereby produce pesudotrypsin activity.[27]

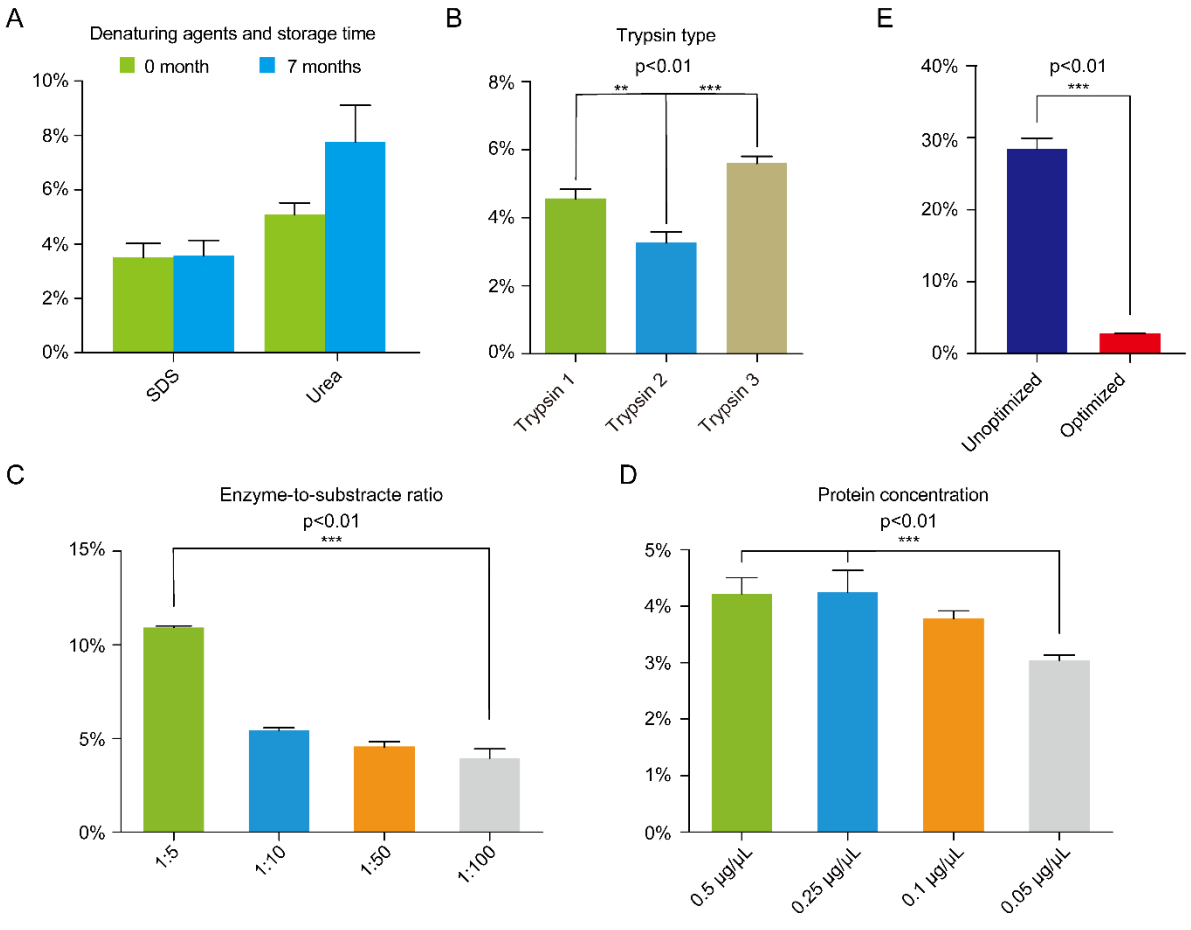

Figure 4. Optimization of digestion conditions to reduce the percentage of partially tryptic peptides. (A) Denaturing agents and storage time, (B) Trypsin type, (C) Enzyme-to-substrate ratio, and (D) Protein concentration during digestion. (E) Comparison between optimized and unoptimized digestion conditions. Error bar represents mean ± standard deviation across three replicates.

In addition, the effects of contaminating proteases such as chymotrypsin, can become more prominent as the amounts of trypsin rise. Our results demonstrate that appropriate ratio of enzyme-to-substrate can effectively minimize nonspecific trypsin cleavages.

Protein concentration during digestion

As the concentration of protein during digestion has an impact on the interaction of enzyme and substrate, we speculate that nonspecific trypsin cleavage may be also related to protein concentration. To investigate the speculation, we digested the same amounts of proteins from mouse brain in four concentrations (0.5 μg/μL, 0.25 μg/μL, 0.1 μg/μL, 0.05 μg/μL) with an enzyme-to-substrate ratio of 1:50. As shown in Figure 4D, the percentages of identified partially tryptic peptides were higher when protein concentration was 0.5 μg/μL or 0.25 μg/μL, indicating protein concentration influenced digest specificity. When protein concentration was diluted, the proportion of partially tryptic peptides slightly declined from 4.2% to 3.0%. The best condition to reduce nonspecific trypsin cleavage was 0.05 μg/μL. As the amounts of enzyme and substrate unchanged, lower concentration provided enough space for random collision for them. So protein concentration in digestion is another important factor that influences nonspecific trypsin cleavage.

Comparison of optimized and unoptimized digestion conditions

Besides the digestion conditions discussed above, we have also tested other parameters in the preliminary experiments, including $CaCl_2$ (add or not), pH (7, 8 or 9), digestion buffer (50 mM ammonium bicarbonate or 0.1 M Tris buffer) as well as digestion time (4 h or 16 h). However, these digestion conditions were found to have no significant influence on cleavage specificity (data not shown). Through condition screening, the optimal digestion conditions were developed as follows: protein storage time less than 1 month, SDS as the denaturing agent, mass spectrometry grade modified trypsin, enzyme-to-substrate ratio of 1:50 and protein concentration of 0.05 μg/μL. The unoptimized conditions were harsher than the routine digestion procedure. We meant to illustrate the bad consequences made by careless sample preparation with this unoptimized condition. In fact, with the routine procedure, the ratios of the non-specific tryptic digestion in different samples varied from 8% in mouse brain to 20% in serum. Under optimized experimental conditions, the ratio of partially tryptic peptides in the total identifications significantly decreased to 2.8% (Figure 4E), contrasting sharply with that (28.4%) under unoptimized conditions (7 months, urea, trypsin 3, 1:5 and 0.5 μg/μL), indicating that nonspecific trypsin cleavage was closely related to digestion conditions.

Specificity and efficiency of the trypsin cleavage

As another important factor for protein digestion, the missed cleavages, which represent the efficiency of trypsin digestion, were also examined under different digestion conditions. The partially tryptic cleavage rate, missed cleavages rate and spectrum identification rate were plotted in supplementary figure 2. Negative correlation was noticed between the partially tryptic cleavage rate and the missed cleavages rate, indicating that balancing the specificity and efficiency of the trypsin cleavage is a big challenge. Since our goal is to control the nonspecific trypsin cleavages, thereby reducing unfavorable influence on qualitative and quantitative proteomics, so the efficiency of the tryptic digestion was sacrificed to some extent. Fortunately, when we examined the spectrum identification rate and peptide identification number, we found that the optimized conditions to reduce the non-specific trypsin cleavage did not made against to the protein identification.

Impact of nonspecific trypsin cleavages on N-glycoproteomic Study

Post-translational proteins, such as glycoproteins, play an important role in many cellular process but are present in relatively low abundance. Thus, identification of glycoprotein usually needs enrichment of glycoproteins or glycopeptides. However, glycopeptide enrichment step may introduce additional nonspecific trypsin cleavages. On the other hand, after enrichment, the resulting peptide populations become much simpler than original products of digestion and are more likely to be influenced by nonspecific trypsin cleavages. In this case, controlling nonspecific trypsin cleavages is of greater importance to glycoproteomic study. In this study, we used a popular and efficient method (ZIC-HILIC) to enrich N-glycopeptides and further detached the N-glycans to reduce ion suppression from glycans. Enrichment of N-glycopeptides using trypsin digests under optimized and unoptimized digestion conditions were conducted in triplicate. Resulting MS/MS spectra were analyzed with two search modes: partially tryptic search (Figure 5A) and fully tryptic search (Figure 5B). As shown in Figure 5A, the percentage of partially tryptic peptides in enriched mixtures of peptides and glycopeptides had sharply declined under optimized experimental conditions. In addition, the average percentage of partially tryptic peptides in the enriched peptide populations under optimized and unoptimized digestion conditions were 12.7% and 46.7% respectively, which increased by 10% and 18.3% relative to corresponding unenriched digests shown in Figure 4E. These results showed an increase in the ratio of partially tryptic peptides after enrichment, which was likely due to glycopeptide enrichment or simpler sample system. In partially tryptic search, the optimized digestion conditions have better results with more fully tryptic N-glycopeptides and less partially tryptic ones than that in unoptimized conditions. Also, in fully tryptic search (Figure 5B), more N-glycopeptides in optimized conditions were identified. Moreover, under fully tryptic search mode, the number of fully tryptic N-glycopeptides in both unoptimized and optimized conditions was larger than that under partially tryptic search mode. Partially tryptic N-glycopeptides identified in database searches are considered to be less confident than fully tryptic N-glycopeptides. Because in nonspecific database searches, partially tryptic sequences greatly outnumber tryptic peptides considered by search algorithm, it is far more likely that partially tryptic sequences will be in-

correctly assigned to MS/MS spectra when correct sequences are absent or when spectrum quality is insufficient to generate confident matches. Consequently, the identified partially tryptic peptides may be incorrect. Generally, the benefit of nonspecific searching is less pronounced for confident identification in proteomics. Thus in fully tryptic searches, it requires that the overwhelming majority of peptides present in a trypsin digest have tryptic ends. So, controlling nonspecific trypsin cleavages in enriched peptide populations is of great significance to the identification of N-glycopeptides. Nevertheless, in routine proteomic or glycoproteomic experiments, nonspecific trypsin cleavages are usually underappreciated by many researchers.

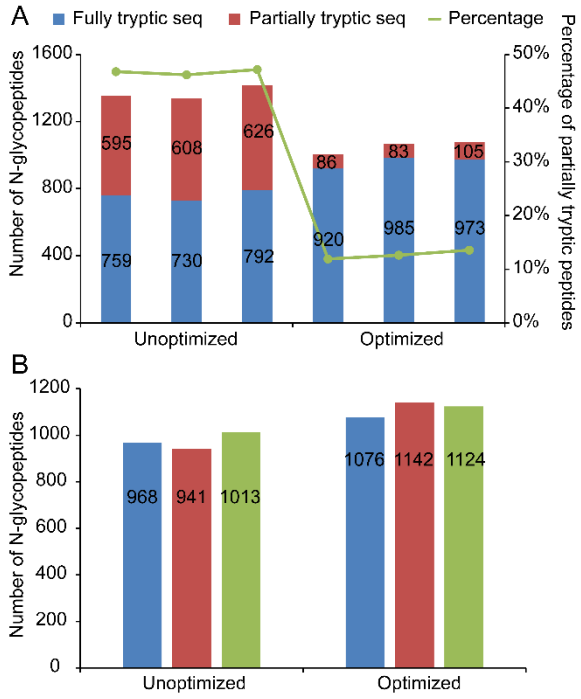

Figure 5. The analysis of N-glycopeptides using optimized and unoptimized experimental conditions in triplicate. (A) Number of fully and partially tryptic N-glycopeptides identified with partially tryptic search mode and correspongding percentage of partially trytic peptides. (B) Number of N-glycopeptdies identified with fully tryptic search mode.

Conclusions

In summary, we found nonspecific trypsin cleavages commonly existed in different sample systems with discrepant extent. Identifiable partially tryptic peptides in mass spectrometer were largely dependent on the complexity of samples. To control the nonspecific cleavages, several major digestion parameters were evaluated, and the optimized digestion conditions were as follows: protein storage time less than 1 month, SDS as denaturing agent, mass spectrometry grade modified trypsin, enzyme-to-substrate ratio of 1:50 and protein concentration of 0.05 μg/μL. The optimized experimental conditions significantly decreased the ratio of partially tryptic peptides in the total identifications from 28.4% to 2.8%. Furthermore, the optimized digestion protocol was applied to the study of N-glycoproteomics and the nonspecific cleavages in enriched mixtures were also sharply decreased. Our work demonstrates the importance of controlling nonspecific trypsin cleavages in both shotgun proteomics and glycoproteomics and provides a better understanding and standardization for routine proteomics sample treatment.

Acknowledgment

The work was supported by S973 Program (973 Program) (2013CB910503 2013CB910802) and MOST 863 Program (2014AA010902).

Authors Notes

*Corresponding Author
Phone: (+) 86-21-54237664. Fax: (+) 86-21-54237961.
Email: shenhuali@fudan.edu.cn and pyyang@fudan.edu.cn
[a.] Minhang Hospital and Institutes of Biomedical Sciences, Fudan University, Shanghai, 201199, P. R. China
[b.] Department of Chemistry, Fudan University, Shanghai, 200433, P. R. China
[c.] Department of Systems Biology for Medicine and School of Basic Medical Sciences, Fudan University, Shanghai, 200032, P. R. China
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

References

1. J. V. Olsen, S. E. Ong and M. Mann, *Mol. Cell. Proteomics*, 2004, **3**, 608-614.
2. P. Picotti, R. Aebersold and B. Domon, *Mol. Cell. Proteomics*, 2007, **6**, 1589-1598.
3. D. Tsur, S. Tanner, E. Zandi, V. Bafna and P. A. Pevzner, *Proc. IEEE Comput. Syst. Bioinform. Conf.*, 2005, 157-166.
4. A. Guldbrandsen, H. Vethe, Y. Farag, E. Oveland, H. Garberg, M. Berle, K.-M. Myhr, J. A. Opsah, H. Barsnes and F. S. Berven, *Mol. Cell. Proteomics*, 2014, **13**, 3152-3163.
5. M. B. Strader, D. L. Tabb, W. J. Hervey, C. Pan and G. B. Hurst, *Anal. Chem.*, 2006, **78**, 125-134.
6. P. Alves, R. J. Arnold, D. E. Clemmer, Y. Li, J. P. Reilly, Q. Sheng, H. Tang, Z. Xun, R. Zeng and P. Radivoja, *Bioinformatics*, 2008, **24**, 102-109.
7. G. Alves and Y. K. Yu, *J. Proteome Res.*, 2013, **12**, 2571-2581.
8. J. E. Elias, W. Haas, B. K. Faherty and S. P. Gygi, *Nat. Methods*, 2005, **2**, 667-675.
9. H. Wang, H. Y. Tang, G. C. Tan and D. W. Speicher, *J. Proteome Res.*, 2011, **10**, 4993-5005.
10. M. S. Lowenthal, Y. Liang, K. W. Phinney and S. E. Stein, *Anal. Chem.*, 2014, **86**, 551-558.
11. S. Hildonen, T. G. Halvorsen and L. Reubsaet, *Proteomics*, 2014, **14**, 2031-2041.

12. J. S. Kim, M. E. Monroe, D. G. Camp, 2nd, R. D. Smith and W. J. Qian, *J. Proteome Res.*, 2013, **12**, 910-916.
13. K. Sergeant, C. Pinheiro, J. F. Hausman, C. P. Ricardo and J. Renaut, *J. Proteome Res.*, 2009, **8**, 3182-3190.
14. N. Gupta, S. Tanner, N. Jaitly, J. N. Adkins, M. Lipton, R. Edwards, M. Romine, A. Osterman, V. Bafna, R. D. Smith and P. A. Pevzner, *Genome Res.*, 2007, **17**, 1362-1377.
15. J. M. Burkhart, C. Schumbrutzki, S. Wortelkamp, A. Sickmann and R. P. Zahedi, *J. Proteomics.*, 2012, **75**, 1454-1462.
16. J. Bunkenborg, G. Espadas and H. Molina, *J. Proteome Res.*, 2013, **12**, 3631-3641.
17. P. L. Loziuk, J. Wang, Q. Li, R. R. Sederoff, V. L. Chiang and D. C. Muddiman, *J. Proteome Res.*, 2013, **12**, 5820-5829.
18. I. Chamrad, O. Strouhal, P. Rehulka, R. Lenobel and M. Sebela, *J. Proteomics*, 2011, **74**, 948-957.
19. V. Keil-Dlouhá, N. Zylber, J.-M. Imhoff, N.-T. Tong and B. Keil, *FEBS Lett.*, 1971, **16**, 291-295.
20. M. Kunitz and J. H. Northrop, *J. Gen. Physiol.*, 1934, **17**, 591-615.
21. S. J. Walmsley, P. A. Rudnick, Y. Liang, Q. Dong, S. E. Stein and A. I. Nesvizhskii, *J. Proteome Res.*, 2013, **12**, 5666-5680.
22. C. Ma, X. Zhao, H. Han, W. Tong, Q. Zhang, P. Qin, C. Chang, B. Peng, W. Ying and X. Qian, *Electrophoresis*, 2013, **34**, 2440-2450.
23. D. Li, Y. Fu, R. Sun, C. X. Ling, Y. Wei, H. Zhou, R. Zeng, Q. Yang, S. He and W. Gao, *Bioinformatics*, 2005, **21**, 3049-3050.
24. T. Farrah, E. W. Deutsch, G. S. Omenn, D. S. Campbell, Z. Sun, J. A. Bletz, P. Mallick, J. E. Katz, J. Malmström, R. Ossola, J. D. Watts, B. Lin, H. Zhang, R. L. Moritz and R. Aebersold, *Mol. Cell. Proteomics*, 2011, **10**, M110.006353. doi: 006310.001074/mcp.M006110.006353.
25. N. Colaert, K. Helsens, L. Martens, J. Vandekerckhove and K. Gevaert, *Nat. Methods*, 2009, **6**, 786-787.
26. L. Switzar, M. Giera and W. M. Niessen, *J. Proteome Res.*, 2013, **12**, 1067-1077.
27. Y. Lv, J. Zhang, H. Wu, S. Zhao, Y. Song, S. Wang, B. Wang, G. Lv and X. Ma, *Chem. Commun.*, 2015, **51**, 5959-5962.