analytical chemistry

Article

pubs.acs.org/ac

Specific Enrichment of Peptides with N-Terminal Serine/Threonine by a Solid-Phase Capture-Release Approach for Efficient Proteomics Analysis

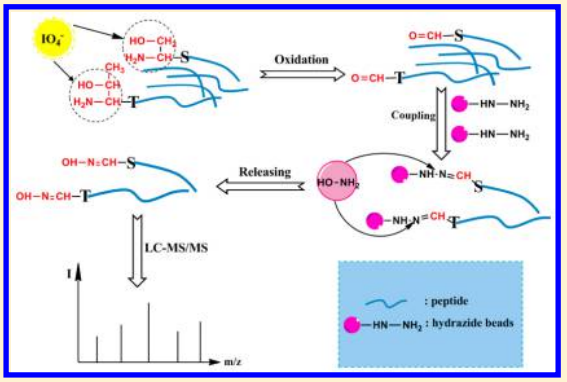
Yating Yao,[†,‡] Junfeng Huang,[†,‡] Kai Cheng,[†,‡] Yanbo Pan,[†,‡] Hongqiang Qin,[†] Mingliang Ye,[*,†] and Hanfa Zou[*,†]

[†]Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences (CAS), Dalian, 116023, China

[‡]University of Chinese Academy of Sciences, Beijing, 100049, China

Supporting Information

ABSTRACT: A problem for "shot-gun" proteomics is that the peptides generated in the proteolysis step overwhelm the analytical capacity of current LC−MS/MS systems. A straightforward approach to overcome this problem is to reduce the sample complexity by isolating the representative peptides of each protein. In this study, we presented a facile solid-phase capture-release approach to selectively enrich the peptides with N-terminal serine/threonine from protein digests. This method exploited the highly efficient reaction between an aldehyde group and a hydrazine group. The excellent performance of this approach was validated using synthetic peptides as well as complex protein digests. It was found that high enrichment specificity could be obtained and the identifications for complex samples with and without enrichment were highly complementary. Besides, the enrichment of peptides with serine/threonine adjacent to different protease cleavage sites demonstrated that our method was able to enrich peptides from protein digests in a sequence specific way. As a result, this new approach provides a simple way to reduce sample complexity and facilitates the identification of low-abundance proteins.



Characterization of a protein mixture by analysis of peptides released from the proteins through proteolysis, referred to as "shot-gun" proteomics, has become the mainstream technology in proteome research.[1,2] A problem for this technology is that the peptides generated in the proteolysis step overwhelm the analytical capacity of current liquid chromatography−tandem mass spectrometry (LC−MS/MS) systems, both in dynamic range and in number.[3,4] Take the mouse proteome as the example, in silico trypsinization of all proteins (51 221) in the proteome leaded to generation of as many as 3,107,991 peptides according to our computation. That is to say, each protein can produce 60 peptides on average. In addition, because of the incomplete proteolysis in a practical situation, more peptides will be generated for each protein. Therefore, the number of species, i.e., peptides, in the digest will increase at least 2 orders of magnitude compared to the original protein mixture.[5] As a consequence, the huge number of peptides derived from high-abundance proteins make the detection and identification of low-abundance proteins extremely difficult.

To address this issue, a straightforward approach is to reduce the sample complexity by isolating the representative peptides of each protein. In this context, peptides containing a specific type of amino acid residue could be selectively enriched from complex peptide mixture for proteome analysis. A number of amino acid residues, such as cysteine,[6−9] tryptophan,[10−12] methionine,[13,14] and histidine,[15,16] have been selectively targeted to reduce the complexity of the proteome sample. An advantage of targeting specific amino acid residues is that knowledge of their presence can improve the reliability of protein identification.[17]

N-terminal serine/threonine in peptide is an interesting target for chemical tagging because of its special structure.[18] The 1,2-amino alcohol structure at the N-terminal serine/threonine peptides can be oxidized by sodium periodate into aldehyde groups, which is similar to the oxidation of the cis-diol structure in glycopeptides.[19−21] Hence, the solid phase hydrazide chemistry method applying for the enrichment of glycopeptides may also be used for the enrichment of peptides with N-terminal serine/threonine in the protein digests.[22,23] In this study, we exploited the reaction of an aldehyde group with a hydrazine group for the development of a solid-phase capture-release method to isolate N-terminal serine/threonine peptides from the complex peptide mixture. After enriched by hydrazide

Received: July 19, 2015
Accepted: October 15, 2015
Published: October 15, 2015

DOI: 10.1021/acs.analchem.5b02711
Anal. Chem. 2015, 87, 11353−11360

beads, only the peptides with N-terminal serine/threonine were submitted to reverse-phase (RP) LC−MS/MS analysis. The method was first optimized with standard peptides containing N-terminal serine/threonine and standard peptide mixtures. This optimized enrichment method was then successfully applied to tryptic and Lys-C digests of a mouse liver protein and tryptic digest of human serum. It was found that this method exhibited an extremely high specificity for the peptides containing N-terminal serine/threonine and thus highly complementary identification results were obtained. Additionally, the enrichment of peptides with serine/threonine adjacent to different protease cleavage sites also demonstrated that our method was able to enrich peptides from protein digests in a sequence specific way. To the best of our knowledge, this is the first time that hydrazide beads were applied to the enrichment of peptides with N-terminal serine/threonine from complex protein digests for proteome analysis.

EXPERIMENTAL SECTION

Materials and Reagents. Formic acid (FA) was obtained from Fluka (Buches, Germany). Trifluoroacetic acid (TFA), dithiothreitol (DTT), iodoacetamide (IAA), 2,5-dihydroxyl benzoic acid (DHB), sodium periodate ($NaIO_4$), sodium chloride, hydroxylammonium chloride, and trypsin (bovine, TPCK-treated) were obtained from Sigma-Aldrich. Endoproteinase Lys-C was purchased from Wako Pure Chemical Industries, Ltd. Peptides (SIINFEKL, TRNYYVRAVL, GRRNSIGK, EESLEDSDVDADF) were synthesized by ChinaPeptides Co. Ltd. (Shanghai, China). Hydrazide sepharose resin was purchased from Bio-Rad (Hercules, CA). PNGase F was purchased from New England Biolabs (Ipswich, MA). Acetonitrile was chromatographic grade from Merck (Darmstadt, Germany). Ammonium bicarbonate and urea were from Bio Basic (Ontario, Canada). The centrifugal filter units (Amicon Ultra-0.5) were purchased from Millipore. Sep-Pak C18 cartridges were obtained from Waters (Milford, MA). Fused silica capillaries with 75 μm i.d. and 200 μm i.d. were provided by Polymicro Technologies (Phoenix, AZ). Other chemicals and reagents were either of analytical grade or of better grade. Pure water used in all of the experiments was purified with a Milli-Q system (Millipore, Bedford, MA).

Preparation of Protein Sample. Adult female C57 mice were purchased from Dalian Medical University (Dalian, China). The mouse liver proteins were prepared as previously described.[24] Briefly, the liver tissues were lysed in ice-cold homogenization buffer consisting of 8 M urea, 1% Triton X-100 (v/v), 65 mM DTT, 1 mM EDTA, 0.5 mM EGTA, 1% protease inhibitor cocktail, phosphatase inhibitors (1 mM NaF, 1 mM Na_3VO_4, 1 mM $C_3H_7Na_2O_6P$, 10 mM $Na_4O_7P_2$), and 40 mM Tris-HCl at pH 7.4. After homogenized by a Potter-Elvehjem homogenizer with a Teflon piston, sonicated for 100 W × 30 s, and centrifuged at 25 000 g for 1 h, the protein concentration was determined by Bradford assay. The extracted proteins were precipitated by chloroform/methanol precipitation. After washing with methanol, the pellets were resuspended in the denaturing buffer containing 8 M urea and 50 mM Tris-HCl (pH 8.0), the protein concentration was determined again by Bradford assay. Protein samples were stored at −80 °C until further usage.

Human serum sample was provided by Second Affiliated Hospital of Dalian Medical University (Dalian, China) according to the standard protocols and stored at −80 °C before analysis.

Protein Digestion. The procedures for protein digestion were performed as follows. The proteins were first reduced by 20 mM DTT at 60 °C for 1 h. Then 40 mM IAA was added and the obtained solutions were incubated for an additional 40 min in the dark at room temperature. After that, the mixtures were diluted 8-fold with 50 mM Tris-HCl (pH 8.0) and incubated at 37 °C overnight with trypsin or endoproteinase Lys-C at an enzyme to substrate ratio of 1/25 (w/w) or 1/50 (w/w), respectively. The resulting peptide solution was added to the filter units with a 10 kDa cutoff. The filter units were centrifuged at 14 000*g* for 15 min to remove the excess trypsin or endoproteinase Lys-C. Then PNGase F was add to the eluted solution at a protein to enzyme ratio of 1 mg: 500 U and incubated at 37 °C for 3 h for deglycosylation of the peptides. The resulting digests were desalted through SPE columns, dried down in a Speedvac Concentrator (Thermo-Fisher, San Jose, CA) and stored at −20 °C until usage.

Capture-and-Release Protocol. The enrichment of N-terminal serine/threonine peptides was similar to the hydrazide chemistry method to enrich glycopeptides.[22] Briefly, a total of 500 μg of the dried proteolyzed peptides were reconstituted in 400 μL of oxidation buffer (100 mM NaAc, 150 mM NaCl, pH 5.5). A volume of 8 μL of $NaIO_4$ (100 mM) was added to the solution to reach a final concentration of 2 mM. The reaction was kept in the dark for 30 min and quenched by the addition of 8 μL of $Na_2S_2O_3$ (200 mM). Then the oxidized peptides were added to 80 μL of Affi-Gel Hz hydrazide beads washed with oxidation buffer in advance and agitated at 25 °C overnight. The beads were washed extensively and sequentially with sodium chloride (1.5 M), ACN/H_2O (80/20, v/v), and sodium chloride (0.9%) for washing hydrophilic, hydrophobic, and other nonspecific binding peptides, respectively. After that, the hydrazide beads were incubated in 400 μL of hydroxylamine-HCl buffer (200 mM, pH 5.0) at 37 °C overnight.[25,26] The supernatant was collected, desalted through SPE column, and dried down in a Speedvac concentrator. As for the peptide standards, the capture-and-release protocol is the same as the protocol described above except skipping the deglycosylation step.

MALDI-TOF MS Analysis. All the MALDI-TOF MS analyses were performed on an AB Sciex 5800 MALDI-TOF/TOF mass spectrometer (AB Sciex, CA) equipped with a pulsed Nd/YAG laser at 355 nm. All mass spectra reported were obtained in the reflex positive ion mode with delayed ion extraction. DHB (2,5-dihydroxybenzoic acid, 25 mg/mL in ACN/H_2O/H_3PO_4 (50/49/1, v/v/v)) was used as the matrix for the analysis of peptides. Sample aliquots (0.5 μL) and DHB (0.5 μL) matrix solution were spotted on the MALDI plate sequentially and dried at room temperature prior to MALDI-TOF MS analysis.

LC−MS/MS Analysis. The lyophilized peptides were resuspended in 0.1% FA. For the untreated samples, an equivalent of approximately 2 μg of total protein was submitted to analysis. For samples obtained from the enrichment procedure, about 100 μg of total protein as starting material was used. RPLC−MS/MS analysis was performed using a quaternary surveyor MS pump (Thermo, San Jose, CA) and LTQ-Orbitrap Velos (Thermo, San Jose, CA). For the capillary separation column, one end of the fused-silica capillary was first manually pulled to a fine point with a spray tip and then packed in-house with C18 AQ particles (3 μm, 120 Å). For the RPLC separation, formic acid (0.1%, v/v) in water and formic acid (0.1%, v/v) in acetonitrile were used as the mobile phase A and

DOI: 10.1021/acs.analchem.5b02711
Anal. Chem. 2015, 87, 11353−11360

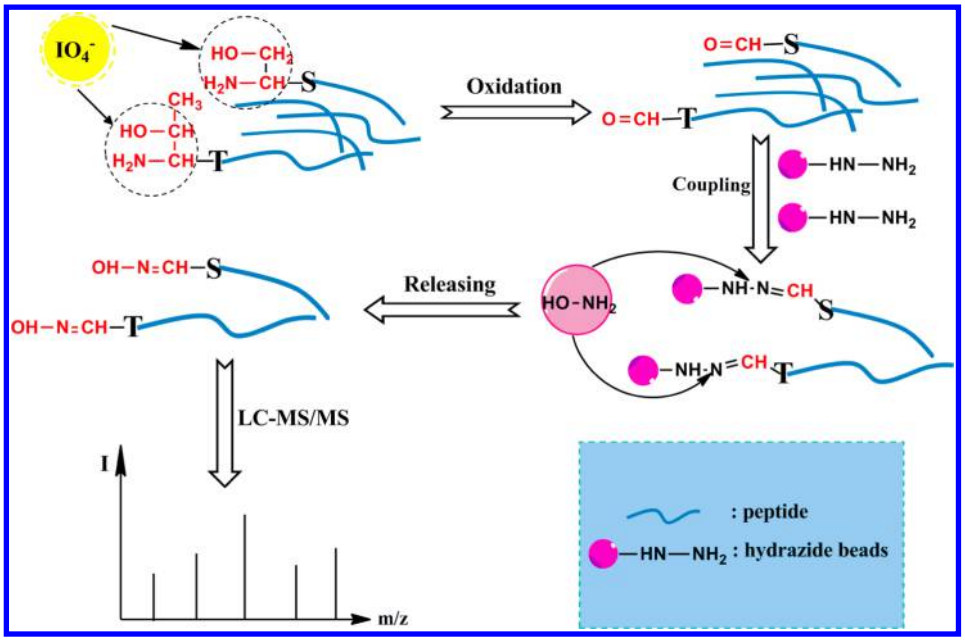


Figure 1. Schematic for the enrichment of N-terminal serine/threonine peptides by the solid-phase capture-release approach.

B, respectively, and the flow rate was adjusted to ~300 nL/min after splitting. Gradient elution from 5% to 35% (v/v) of the 0.1% (v/v) formic acid in acetonitrile in 120 min was performed to elute each sample.

The LTQ-Orbitrap Velos mass spectrometer (Thermo, San Jose, CA) was operated in data-dependent MS/MS acquisition mode. A spray voltage of 2.2 kV was applied between the spray tip and MS interface. The temperature of the ion transfer capillary was set as 250 °C. The full mass scan acquired in the Orbitrap mass analyzer was from m/z 400 to 2000 with a resolution of 60 000. The 20 most intense ions from the full scan were fragmented by collision-induced dissociation (CID) in the LTQ. The dynamic exclusion function was set as follows: repeat, 2; duration, 30 s; exclusion duration, 60 s. System control and data collection were carried out by Xcalibur software.

Database Searching and Data Analysis. Protein identification was performed with MaxQuant (http://www.maxquant.org).[27] Raw data files were searched against non-redundant mouse protein database or human protein database downloaded from UniProt (ftp.uniprot.org). The parameters were set as follows: precursor-ion mass tolerance, 20 ppm; fragment-ion mass tolerance, 6 ppm; protease, trypsin, or Lys-C, two missed cleavages. The peptide and protein false discovery rates (FDRs) were set to 0.01. Carbamidomethylation on cysteine (C, + 57.0215 Da) was set as a fixed modification for all the searches. One or more following variable modifications were set: oxidation on methionine (M, + 15.9949 Da), deamidation on asparagine (N, + 0.9858 Da), oximation on N-terminal serine (S, −16.0313 Da), and threonine (T, − 30.0470 Da). The other settings were the same as the conventional search. Venn diagrams were all generated automatically by the Venn Diagram Plotter program (http://omics.pnl.gov/software/venn-diagram-plotter). Sequence logos were automatically generated by the WebLogo (http://weblogo.berkeley.edu/logo.cgi). The raw sequences for WebLogo analysis were centered at the cleavage site. The N- or C-terminal sequences that could not be extended were excluded. The protein details obtained were analyzed for the subcellular localization, molecular function, or biological process by PANTHER (Protein ANalysis THrough Evolutionary Relationships) classification system (http://www.pantherdb.org/).[28]

RESULTS AND DISCUSSION

Enrichment of N-Terminal Serine/Threonine Peptides by a Solid-Phase Capture-Release Approach. The approach presented here to enrich N-terminal serine/threonine peptides originated from the hydrazide chemistry method widely used in the enrichment of glycopeptides. As shown in Figure 1, the peptides are first incubated with $NaIO_4$ which leads to the formation of aldehyde groups on N-terminal serine/threonine. Then the peptides are incubated with hydrazide beads which resulted in the specific capture of N-terminal serine/threonine peptides onto the beads due to the generation of hydrazine-peptide conjugates. The bonds between the beads and the captured peptides are sensitive to hydroxylamine. Therefore, the captured peptides can be selectively released by hydroxylamine chloride.[25,26] The released peptides are submitted to 1D RPLC−MS/MS for identification. The mass difference between the original peptides and the finally enriched peptides provided the evidence to identify the peptides with N-terminal serine/threonine.

First, we used two synthetic peptides (SIINFEKL, TRNYYVRAL) containing N-terminal serine and N-terminal threonine, respectively, to optimize the oxidation reaction conditions. It is well-known that the concentration of sodium periodate plays a key role in the oxidation reaction.[29,30] Specifically, a low concentration may not sufficient to completely convert the 1,2-amino alcohols at N-termini to aldehydes, while a high concentration may result in significant side reactions.[31−33] On the basis of our previous study,[25] we further investigated the effect of four $NaIO_4$ concentrations of 0.2, 0.5, 1, and 2 mM on the oxidiation of the synthetic peptides containing N-terminal serine/threonine. As a control, the N-terminal serine/threonine peptides without $NaIO_4$ treatment were analyzed in the same way. The resulting

DOI: 10.1021/acs.analchem.5b02711
Anal. Chem. 2015, 87, 11353−11360

MALDI spectra were shown in Figures S1 and S2. It was observed that the original peaks of the peptides disappeared gradually with the increasing of the concentration of $NaIO_4$. When the concentration increased to 1 mM, the synthetic N-terminal threonine peptide TRNYYVRAL was fully converted to its oxidized derivate (− 45 Da). However, 1 mM $NaIO_4$ was not sufficient for the complete oxidation of synthetic peptide containing N-terminal serine (SIINFEKL). When the peptide was incubated with 2 mM $NaIO_4$, the N-terminal oxidized derivate (− 31 Da) and its hydrate (− 13 Da) were dominated in the MALDI mass spectrum. Therefore, we adopted 2 mM as the optimized concentration of $NaIO_4$ to perform the following experiments. Other reaction conditions including the amount of starting material and hydrazide beads, oxidation time, and temperature, etc. remained the same with previous study and were introduced in detail in the Experimental Section.

Second, we tested the feasibility of the method for the enrichment of N-terminal serine/threonine peptides. The synthetic peptides containing N-terminal serine/threonine were oxidized with 2 mM sodium periodate and were then captured onto the hydrazide beads. After hydroxylamine chloride treatment, the peptides were collected and submitted to MALDI analysis. As illustrated in Figure 2 and Figure S3, the

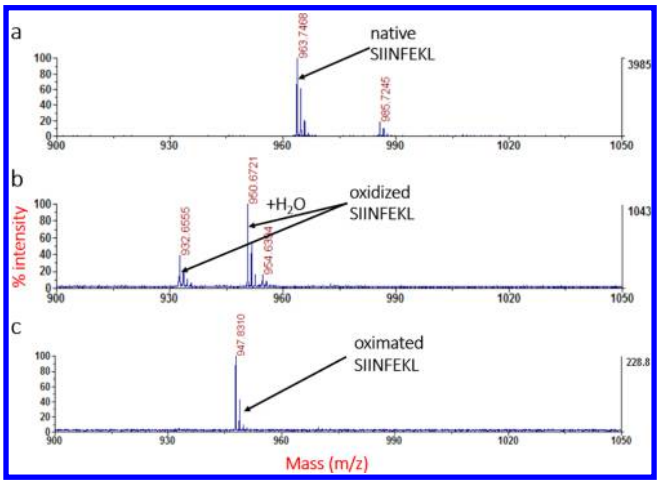


Figure 2. Enrichment of the standard peptide SIINFEKL with N-terminal serine: (a) original peptide, (b) peptide oxidized at 2 mM $NaIO_4$, and (c) peptide released from hydrazide beads with hydroxylamine chloride.

synthetic peptides were converted to their oxime forms with the m/z values of 947 (− 16 Da) and 1224 (− 30 Da), respectively. Although the oxidation of standard peptide with N-terminal Ser resulted in the formation of the aldehyde form and the hydrate form, the release of this peptide showed only one major peak, which indicated no other side reaction was generated. As a whole, all the results discussed above confirmed that the peptides were modified as expected. Therefore, applying this method to enrich N-terminal serine/threonine peptides from complex samples is theoretically feasible.

The efficiency of this method was then investigated. Given the N-terminal serine/threonine peptides were not recovered in their native forms, we investigated the loss of the targeted peptides to indirectly measure the efficiency of this strategy. Three key steps, namely, oxidation, coupling, and release, finally determined the efficiency of this method. The oxidation step, leading to the formation of aldehyde groups on N-terminal serine/threonine, could reach complete as long as the concentration of sodium periodate is sufficient (Figures S1 and S2). The coupling step is related to the reaction of an aldehyde with a hydrazide, which was demonstrated to be very effective and could be achieved quantitatively.[18] As to the release step, if insufficient eluent was used, then part of the target peptides would remain on hydrazide beads, resulting in the loss of N-terminal serine/threonine peptides. Herein, we used a standard peptide mixture containing the two N-terminal serine/threonine peptides (SIINFEKL, TRNYYVRAL) to investigate if the captured peptides could be released completely under the optimized conditions. All the procedures remained the same as before. After collecting the eluent, the remaining hydrazide beads were again incubated with hydroxylamine chloride solution at the same conditions. The two supernates successively collected were submitted to MALDI analysis, respectively. As shown in Figure S4, no peaks corresponding to any peptides were found in the spectrum for the second eluent, while the peaks corresponding to the oximated N-terminal serine/threonine peptides were dominated in the spectrum for the first eluent, which clearly demonstrated the release step by the first elution was complete. All of these above evidence indicated that the high efficiency of this enrichment method could be obtained. However, it should be mentioned that the one centrifugation step and two desalting steps for the processing of complex biological samples may cause severe loss of samples, which may lead to a low efficiency for the sample preparation.

We also evaluated the specificity of this enrichment strategy. A mixture of four standard peptides, one containing N-terminal serine (SIINFEKL), one containing N-terminal threonine (TRNYYVRAL) and two containing neither N-terminal serine nor threonine (GRRNSIGK, EESLEDSDVDADF) as negative controls, was used. The mixture was subjected to the same oxidization, coupling, and release procedures as described above. The enriched peptides were submitted to MALDI analysis. The MALDI mass spectra of the mixture before and after enrichment were shown in Figure S5. The untreated peptide mixture showed four peaks, corresponding to the four standard peptides. After our enrichment process, only two peaks representing the oximated N-terminal serine/threonine peptides could be identified. It should be noted that no undesired side reaction product peaks were observed in the spectra. So above data preliminary demonstrated that our enrichment strategy could selectively enrich peptides with N-terminal serine/threonine.

Enrichment of N-Terminal Serine/Threonine Peptides from Tryptic Digest of Mouse Liver Extract. After demonstrating the feasibility of this method, we then applied the solid-phase capture-release approach to selectively enrich N-terminal serine/threonine peptides from complex peptide mixture. Different from standard peptide mixture, a complex digest derived from total cell lysate may contain many glycopeptides. The existence of glycopeptides will hamper the identification of N-terminal serine/threonine peptides as the diol groups of the sugar moieties in glycopeptides can also be oxidized by sodium periodate into aldehyde groups, thus the aldehyde groups from glycopeptides would compete with aldehyde groups from N-terminal serine/threonine in coupling with hydrazide. With the fact that about a thousand glycoproteins were identified from complex proteome sample,[25] obviously the influence of glycopeptides cannot be negligible in the enrichment and identification of N-terminal serine/threonine for such samples. In order to avoid the interference

DOI: 10.1021/acs.analchem.5b02711
Anal. Chem. 2015, 87, 11353−11360

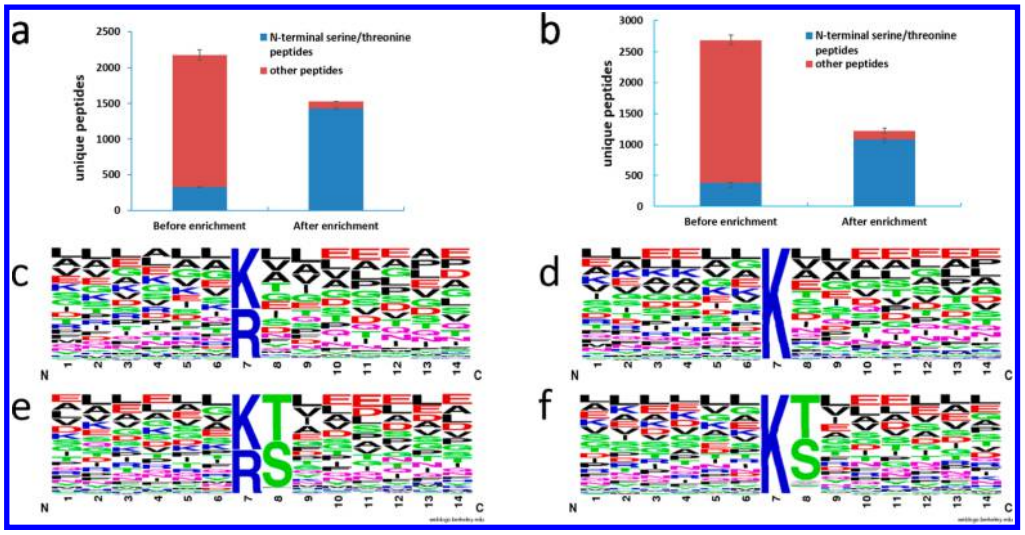


Figure 3. Specificity of peptides enriched from mouse liver digests: (a,b) number of peptides identified before and after enrichment, (c,d) sequence logo of the identified peptides before enrichment, and (e,f) sequence logo of the identified peptides after enrichment; (a,c,e) trypsin digestion and (b,d,f) Lys-C digestion.

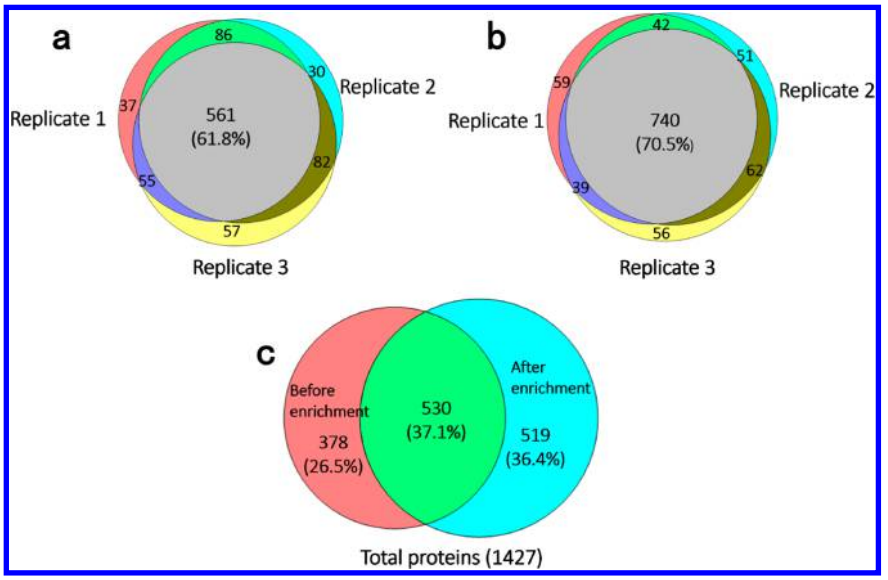


Figure 4. Complementary protein identifications achieved by selective enrichment of N-terminal serine/threonine peptides from mouse liver tryptic digests: (a) overlap of protein identifications among three replicated experiments before enrichment, (b) overlap of protein identifications among three replicated experiments after enrichment, and (c) overlap of total protein identifications before and after enrichment.

of glycopeptides, the proteins extracted from mouse liver tissues were deglycosylated by PNGase F following the trypsin digestion and subsequently subjected to this new peptide enrichment method. The resulting peptides were analyzed by 1 D RPLC−MS/MS. For comparison, the deglycosylated protein digest without enrichment was also analyzed in the same way. As shown in Figure 3a, the ratio of the identified peptides with N-terminal serine/threonine reached up to 93.5% (1429/1528), while this ratio was only 15.0% (326/2177) before the enrichment strategy was applied. Clearly these data demonstrated the high specificity of the method for the enrichment of N-terminal serine/threonine peptides.

We further compared the proteins identified with and without enrichment (Figure 4). Compared with the untreated sample, the sample subjected to enrichment strategy was identified about 15% more protein even though only about 60% peptides were identified. The overlaps of the three replicate experiments in untreated and enriched samples were about 60% to ~70%, while the overlap between the two methods was only about 37%. This low overlap indicated that the identifications achieved by the enrichment strategy were highly complementary to the identifications from the untreated samples. What needed to be emphasized here is that about 500 proteins were newly identified after the enrichment. Because of their low abundance in the sample, it is difficult to identify these proteins by direct analysis of the sample. However, our method is able to remove many redundant peptides derived from high-abundance proteins. Clearly the selective enrichment of peptides with N-terminal serine/threonine from digests is able to reduce the complexity of protein samples and thus allowed the proteome analysis to identify more low-abundance proteins.

Analysis of Mouse Liver Proteome by Using Different Proteases. In proteome samples, the N-terminal serine/threonine peptides are generated by digestion with a protease.

DOI: 10.1021/acs.analchem.5b02711
Anal. Chem. 2015, 87, 11353−11360

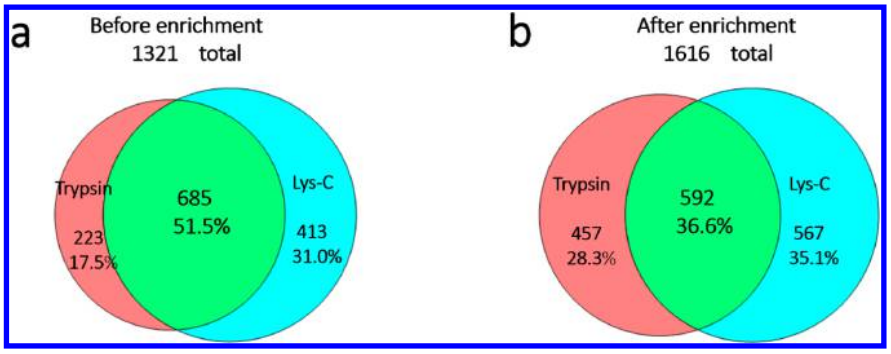


Figure 5. Complementary protein identifications for mouse liver samples digested by trypsin and endoproteinase Lys-C: (a) overlap of protein identifications before enrichment and (b) overlap of protein identifications after enrichment.

Thus, this method enables the enrichment of proteolyzed peptides that have either serine or threonine adjacent to a protease cleavage site. If the sample is digested by trypsin, then this method should be able to enrich K/R.S/T containing peptides. The sequence logos centered with N-terminal sites for the peptides identified from mouse liver deglycosylated tryptic digest before and after enrichment were generated by the WebLogo (Figure 3c,e). It is obvious that the obtained peptides after enrichment were dominated by K/R.S/T containing peptides. Clearly our method is able to enrich peptides from protein digest in a sequence specific way.

The reduction of the sample complexity through our solid-phase capture-release approach was investigated by in silico digestion of mouse proteome with different proteases (Table S1). The data indicated that the K/R.S/T containing peptides derived from trypsin digestion still covered 88% of the mouse proteome while the number of peptides was reduced by 7-fold. Furthermore, an average of 8 peptides for each protein ensures confident analysis of mouse proteome. According to Table S1, this enrichment method combined with other proteases can also lead to significant reduction in sample complexity while keeping the proteome analysis coverage high.

On the basis of the excellent performance of our solid-phase capture-release approach targeting K/R.S/T containing peptides, we believe that this method is also qualified for the enrichment of proteolyzed peptides that have either serine or threonine adjacent to other protease cleavage sites. The protein samples from mouse liver digested by endoproteinase Lys-C were treated with our enrichment process. Afterward, the obtained peptides were submitted to 1 D RPLC−MS/MS analysis. As expected, the results were somewhat alike to that of tryptic peptides. The first and foremost, selection of K.S/T containing peptides from a digest could lead to a reduction of sample complexity. The ratios of identified peptides to proteins were determined. This ratio was from 3.51 (3853/1098) before enrichment to 1.70 (1966/1159) after enrichment, which demonstrated the enrichment strategy could significantly reduce the complexity of the sample. The specificity of the approach was also investigated. As shown in Figure 3b, the ratio of the identified K.S/T containing peptides reached up to 88.6% (1083/1223), while this ratio was only 14.1% (380/2690) for the untreated sample, which is close to its proteome level before the enrichment strategy was applied (Table S1). The sequence logos (Figure 3d,f) further demonstrated the high specificity as well as the sequence specific enrichment feature of our method. The last, but not the least, the enrichment of K.S/T containing peptides facilitated the identification of some low-abundance proteins. The results of identified proteins were summarized in Figure S6. After enrichment, 582 proteins were newly identified. The overlap between the samples before and after enrichment was only about 35%, which is close to the results obtained from the enrichment of K/R.S/T containing peptides. This low overlap once again indicated that the identifications achieved by the enrichment strategy are highly complementary to the identifications from the samples without any treatment.

As the identification results achieved by the enrichment approach are highly complementary to the results obtained from the samples before enrichment, it is of interest to compare the identifications obtained from the two experiments with enrichment, in which the samples were subjected to digestion by trypsin and endoproteinase Lys-C, respectively. Figure 5 showed that before enrichment, the overlap of identified proteins between the two samples was 51.5%; however, the overlap was only 36.6% after enrichment. This low overlap manifested that without enrichment, the complementarity of the identification results was not so outstanding even when different proteases were applied to the digestion process due to the presence of a huge amount of redundant peptides. Nevertheless, enrichment of peptides containing specific sequences could achieve a highly complementary identifications. If other proteases were selected to accomplish the protein digestion, the identification results would be more complementary and more low-abundance proteins could be identified.

Compared with the methods targeting one specific type of amino acid residue,[9,11,12,15] a significant advantage of our approach is sequence specific. By combining the specificities of N-terminal selective chemical reaction and protease catalyzed cleavage, our approach is able to target analysis of proteins containing two specific residues, e.g., K/R.S/T for trypsin digested sample. It is well-known that multiple residues with consensus sequence on proteins are closely related to protein functions.[34,35] We also performed gene ontology classification of proteins uniquely identified from the trypsin digested mouse liver sample with and without enrichment according to their biological processes, molecular functions, and cellular components (Figures S7−S9). Subtle differences were observed for these two sets of proteins. For example, the ratio of proteins in biological regulation process increased from 4.96% in samples before enrichment to 8.02% in samples after enrichment, the ratio of proteins identified with enzyme regulator activity increased about 4% after enrichment. Above results imply that the sequence specific enrichment approach has the potential to enrich a subset proteins with specific functions. To make this goal as a reality, more sequence specific enrichment approaches targeting more consensus sequences should be developed.

DOI: 10.1021/acs.analchem.5b02711
Anal. Chem. 2015, 87, 11353−11360

Application of the Solid-Phase Capture-Release Approach to Analyze Human Serum Proteome. To further investigate its performance, this approach was applied to analyze a more challenging sample, human serum. The presence of high-abundance proteins, such as IgG and albumin, makes it difficult to detect biomarkers which are usually low abundance in serum.[36] On the basis of all these results we have obtained in the previous section, we considered this method to be a valuable approach to the simplification of serum sample and the identification of low-abundance proteins. We have performed a series of experiments, in which the starting materials were used in a wide range, from 20 μg to 1 mg to selectively enrich N-terminal serine/threonine peptides generated by trypsin digestion of human serum sample. As shown in Figure 6, the number of identified proteins increased with

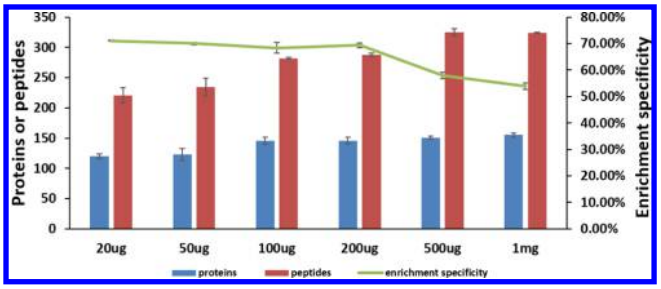


Figure 6. Identification results of human serum samples digested by trypsin after enrichment with different amounts of starting material. The above data were averaged from 3 replicates (error bars represent the standard deviation).

the amount of starting materials increasing. However, the enrichment specificity reduced dramatically when the amount was over 200 μg. We considered the reduction of enrichment specificity was due to the severe interferences of nonspecific binding peptides from high-abundance proteins when the starting protein amount was too high. At the same time, if the staring amount was too small, the peptides generated from low-abundance proteins might be lost during the sample preparation process, which resulted in a reduction of protein or peptide identifications. Accordingly, to obtain a trade-off between the number of proteins or peptides and enrichment specificity, we chose 200 μg as the starting material. For serum sample, this amount of sample is easy to be obtained as it corresponds to only about 3 μL of serum. However, for the analysis of rare sample such as tissues obtained by laser-capture microdissection, further optimization of the enrichment procedures to analyze minute amount of sample is required. For this challenging sample, the enrichment specificity of about 70% was obtained, which further indicated the excellence of this enrichment approach.

Generally speaking, the protein abundance could be reflected by the spectra counts in an experiment.[37−39] On this basis, we classified the identified proteins into 19 bins according to their spectra counts. The distributions of these proteins from untreated and enriched samples across the spectra counts were summarized in Figure S10. The overall distributions between the two samples were quite different, showing that the spectra counts changed a lot after enrichment. Compared with the results before enrichment, it can be noticed that the proteins identified after enrichment were mainly distributed in the low spectra count (<8) range, and the number of proteins with high spectra count (>32) decreased a lot. By combining with the protein identifications, we found that some high abundant proteins were also enriched and located in the low spectra count range. This is because of the existence of the K/R.S/T sequence in their sequences. Such specific sequences could be digested by trypsin to generate N-terminal serine/threonine peptides to be captured by our enrichment method. Although these proteins showed high abundance in the untreated sample, the spectra counts corresponding to them reduced significantly after enrichment. Taking the most abundant protein serum albumin as an example, the spectra counts decreased from 1583 before enrichment to 244 after enrichment, which decreased more than five times. This clearly illustrated that the proteins with less abundance would not be suppressed by the high-abundance proteins and further demonstrated that our enrichment approach have the potential to identify more low-abundance proteins in the serum sample.

CONCLUSION

In this study, we present a chemoselective method to enrich peptides containing N-terminal serine/threonine to simplify digests of complex protein samples. It is modified from the solid phase hydrazide chemistry method applying for the enrichment of glycopeptides. The excellent performance of this approach was validated using synthetic peptides as well as complex protein digests. The solid-phase capture-release format made this approach highly efficient and easy to manipulate. This new approach provides a simple way to reduce sample complexity and facilitates the identification of low-abundance proteins.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.5b02711.

Additional results and information (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: mingliang@dicp.ac.cn.

*E-mail: hanfazou@dicp.ac.cn. Phone: +86-411-84379610. Fax: +86-411-84379620.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work is funded in part by the China State Key Basic Research Program Grant (Grants 2013CB911202, 2012CB910101, and 2012CB910604) and the National Natural Science Foundation of China (Grants 21321064, 21275142, 21235006, 21525524, 21535008 and 81361128015).

REFERENCES

(1) Yates, J. R. *J. Am. Chem. Soc.* **2013**, *135*, 1629−1640.

(2) Yates, J. R.; Ruse, C. I.; Nakorchevsky, A. *Annu. Rev. Biomed. Eng.* **2009**, *11*, 49−79.

(3) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198−207.

(4) Nilsson, C. L.; Davidsson, P. *Mass Spectrom. Rev.* **2000**, *19*, 390−397.

(5) Zhang, H.; Yan, W.; Aebersold, R. *Curr. Opin. Chem. Biol.* **2004**, *8*, 66−75.

DOI: 10.1021/acs.analchem.5b02711
Anal. Chem. 2015, 87, 11353−11360

(6) Dai, J.; Wang, J.; Zhang, Y.; Lu, Z.; Yang, B.; Li, X.; Cai, Y.; Qian, X. *Anal. Chem.* **2005**, *77*, 7594−7604.
(7) Ren, D.; Julka, S.; Inerowicz, H. D.; Regnier, F. E. *Anal. Chem.* **2004**, *76*, 4522−4530.
(8) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994−999.
(9) Giron, P.; Dayon, L.; David, F.; Sanchez, J. C.; Rose, K. *J. Proteomics* **2009**, *71*, 647−661.
(10) Kuyama, H.; Watanabe, M.; Toda, C.; Ando, E.; Tanaka, K.; Nishimura, O. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1642−1650.
(11) Yu, Y.; Liu, M.; Yan, G.; He, Y.; Xu, C.; Shen, H.; Yang, P. *Talanta* **2011**, *85*, 1001−1006.
(12) Foettinger, A.; Leitner, A.; Lindner, W. *J. Proteome Res.* **2007**, *6*, 3827−3834.
(13) Shen, M.; Guo, L.; Wallace, A.; Fitzner, J.; Eisenman, J.; Jacobson, E.; Johnson, R. S. *Mol. Cell. Proteomics* **2003**, *2*, 315−324.
(14) Gevaert, K.; Van Damme, J.; Goethals, M.; Thomas, G. R.; Hoorelbeke, B.; Demol, H.; Martens, L.; Puype, M.; Staes, A.; Vandekerckhove, J. *Mol. Cell. Proteomics* **2002**, *1*, 896−903.
(15) Mesmin, C.; Domon, B. *J. Proteome Res.* **2014**, *13*, 6160−6168.
(16) Wang, S. H.; Zhang, X.; Regnier, F. E. *J. Chromatogr A* **2002**, *949*, 153−162.
(17) Sechi, S.; Chait, B. T. *Anal. Chem.* **1998**, *70*, 5150−5158.
(18) Geoghegan, K. F.; Stroh, J. G. *Bioconjugate Chem.* **1992**, *3*, 138−146.
(19) Bobbitt, J. M. *Adv. Carbohyd Chem.* **1956**, *11*, 1−41.
(20) Kung, K. K.; Wong, K. F.; Leung, K. C.; Wong, M. K. *Chem. Commun. (Cambridge, U. K.)* **2013**, *49*, 6888−6890.
(21) Chelius, D.; Shaler, T. A. *Bioconjugate Chem.* **2003**, *14*, 205−211.
(22) Nilsson, J.; Ruetschi, U.; Halim, A.; Hesse, C.; Carlsohn, E.; Brinkmalm, G.; Larson, G. *Nat. Methods* **2009**, *6*, 809−U826.
(23) Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R. *Nat. Biotechnol.* **2003**, *21*, 660−666.
(24) Pan, Y.; Ye, M.; Zhao, L.; Cheng, K.; Dong, M.; Song, C.; Qin, H.; Wang, F.; Zou, H. *Angew. Chem., Int. Ed.* **2013**, *52*, 9205−9209.
(25) Huang, J.; Qin, H.; Sun, Z.; Huang, G.; Mao, J.; Cheng, K.; Zhang, Z.; Wan, H.; Yao, Y.; Dong, J.; Zhu, J.; Wang, F.; Ye, M.; Zou, H. *Sci. Rep.* **2015**, *5*, 10164.
(26) Dirksen, A.; Yegneswaran, S.; Dawson, P. E. *Angew. Chem., Int. Ed.* **2010**, *49*, 2023−2027.
(27) Cox, J.; Mann, M. *Nat. Biotechnol.* **2008**, *26*, 1367−1372.
(28) Mi, H. Y.; Muruganujan, A.; Casagrande, J. T.; Thomas, P. D. *Nat. Protoc.* **2013**, *8*, 1551−1566.
(29) Nicolet, B. H.; Shinn, L. A. *J. Am. Chem. Soc.* **1939**, *61*, 1615−1615.
(30) Sklarz, B. *Q. Rev., Chem. Soc.* **1967**, *21*, 3.
(31) Previero, A.; Signor, A.; Bezzi, S. *Nature* **1964**, *204*, 687.
(32) Marondedze, C.; Turek, I.; Parrott, B.; Thomas, L.; Jankovic, B.; Lilley, K. S.; Gehring, C. *Cell Commun. Signaling* **2013**, *11*, 1.
(33) Stadtman, E. R.; Moskovitz, J.; Levine, R. L. *Antioxid. Redox Signaling* **2003**, *5*, 577−582.
(34) Puntervoll, P.; Linding, R.; Gemund, C.; Chabanis-Davidson, S.; Mattingsdal, M.; Cameron, S.; Martin, D. M. A.; Ausiello, G.; Brannetti, B.; Costantini, A.; Ferre, F.; Maselli, V.; Via, A.; Cesareni, G.; Diella, F.; Superti-Furga, G.; Wyrwicz, L.; Ramu, C.; McGuigan, C.; Gudavalli, R.; Letunic, I.; Bork, P.; Rychlewski, L.; Kuster, B.; Helmer-Citterich, M.; Hunter, W. N.; Aasland, R.; Gibson, T. J. *Nucleic Acids Res.* **2003**, *31*, 3625−3630.
(35) Amanchy, R.; Periaswamy, B.; Mathivanan, S.; Reddy, R.; Tattikota, S. G.; Pandey, A. *Nat. Biotechnol.* **2007**, *25*, 285−286.
(36) Anderson, N. L.; Anderson, N. G. *Mol. Cell. Proteomics* **2002**, *1*, 845−867.
(37) Ning, K.; Fermin, D.; Nesvizhskii, A. I. *J. Proteome Res.* **2012**, *11*, 2261−2271.
(38) Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G. *Mol. Cell. Proteomics* **2005**, *4*, 1487−1502.
(39) Liu, H. B.; Sadygov, R. G.; Yates, J. R. *Anal. Chem.* **2004**, *76*, 4193−4201.

DOI: 10.1021/acs.analchem.5b02711
Anal. Chem. 2015, 87, 11353−11360