analytical chemistry

Article

pubs.acs.org/ac

Trypsin-Catalyzed N-Terminal Labeling of Peptides with Stable Isotope-Coded Affinity Tags for Proteome Analysis

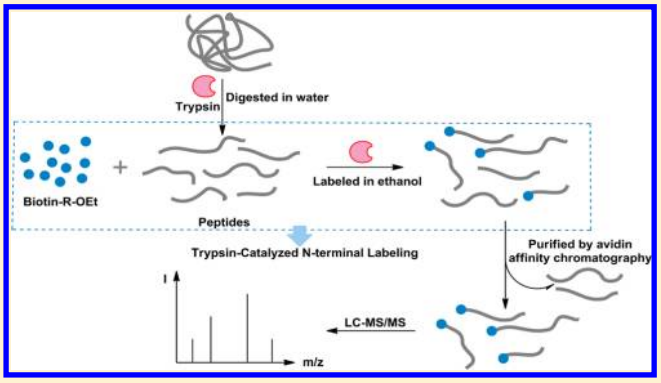
Yanbo Pan,[†,‡] Mingliang Ye,*[,†] Hao Zheng,[†,‡] Kai Cheng,[†,‡] Zhen Sun,[†,‡] Fangjie Liu,[†,‡] Jing Liu,[†,‡] Keyun Wang,[†,‡] Hongqiang Qin,[†] and Hanfa Zou*[,†]

[†]Key Laboratory of Separation Sciences for Analytical Chemistry, National Chromatographic Research and Analysis Center, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China

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

Supporting Information

ABSTRACT: An enzymatic approach to label peptide N-termini with isotope-coded affinity tags is presented. This method exploits the high activity of trypsin for peptide synthesis in organic solvents. A cosubstrate containing a stable isotope-coded Arg residue and a biotin tag was synthesized. When the cosubstrate was incubated with tryptic peptides and trypsin in ethanol solution, the stable isotope-coded affinity tag was specifically coupled onto the N-termini of peptides via the formation of new peptide bonds. The labeled peptides were specifically enriched by avidin affinity chromatography and then were submitted to liquid chromatography−tandem mass spectrometry (LC/MS/MS) for quantification. This enrichment step effectively reduced the interference by unlabeled peptides. The excellent performance of this approach was demonstrated by labeling standard peptides as well as a mouse liver digest. In addition to one amino acid residue, a few dipeptide tags were also introduced to the N-termini of peptides successfully by this enzymatic approach. It was found that the identifications for samples labeled with these tags were highly complementary. Coupling a short sequence tag onto peptides could be an effective approach to improve the coverage for proteome analysis.



Quantification of proteins by stable isotopic labeling of tryptic peptides has been widely used in proteomics. Chemical labeling and metabolic labeling are the most commonly used methods to introduce stable isotopes for quantitative proteomics.[1,2] Metabolic labeling strategies such as stable-isotope labeling by amino acids in cell culture (SILAC) are mainly applicable to living cells,[3] and costly heavy amino acids limit its wide application. Chemical labeling is much cheaper and is applicable to analyze almost any proteomics sample, including body fluids and tissues. The popular chemical labeling approaches, for example, dimethyl labeling, label the primary amines on peptides. Thus both the primary amines on the N-termini and the side chains of lysine residue are labeled, which results in the labeling of multiple tags per peptide. This is acceptable for quantification of proteins in most cases. However, it may be not applicable to quantify some types of post-translational modifications (PTMs) on lysine. The lysine side chain is a target of different PTMs, including methylation and acetylation.[4] The dimethyl labeling method cannot distinguish in vivo dimethylation of lysine from the modification resulting from chemical labeling. This will be not a problem if a terminal labeling approach that does not modify protein side chains is available. Furthermore, the labeling of side chains on peptides for those conventional approaches does not benefit spectral interpretation, while the terminal labeling approach has important applications in peptide de novo sequencing. Terminal labeling of peptides could either simplify the spectra or distinguish the fragmentation ion series that facilitate direct reading the sequence from the spectra.[5−7] Because different peptides have different fragmentation behaviors, coupling oligopeptide to tryptic peptides could be a promising labeling approach to improve proteomics coverage. This type of labeling must be terminal-directed. Otherwise peptides with branched chains are formed that cannot be easily identified. Therefore, it is of interest to develop new approaches to specifically label peptide termini.

Up to now, two chemical labeling approaches have been developed to specifically label peptide N-termini. The first one is achieved by blocking primary amine groups on lysine side chains via guanidination, followed by labeling N-terminal amine groups with amine-reactive isotope-coded reagents.[8] The guanidination reaction is performed at pH 11 and 65 °C. Such harsh reaction conditions may lead to peptide degradation, unexpected modification of side chains, or loss of some labile PTMs. The second method is much more straightforward. It makes use of the difference in pK_a values of these two types of amine groups.[9]

Received: September 28, 2013
Accepted: December 19, 2013
Published: December 19, 2013

   dx.doi.org/10.1021/ac403060d | *Anal. Chem.* 2014, 86, 1170−1177

Selective labeling of primary amine groups on peptide N-termini is achieved by carefully controlling the pH for reaction. However, such a method might lack enough specificity.

Trypsin is often used as the protease for protein digestion in proteomics analysis. However, it also has high activity for peptide synthesis in organic solvents.[10−12] As early as 1977, trypsin was used as a catalyst for peptide synthesis.[13] Recently, isotope-coded amino acids were covalently linked to tryptic peptides for quantitative proteomics by using trypsin as a ligase.[5] The reaction was demonstrated to be specifically N-terminal-directed and proceeded under mild conditions. However, due to the low labeling efficiency, many unlabeled peptides were present in the labeled sample, which seriously compromised the proteomics analysis performance. In the classic ICAT (isotope-coded affinity tags) labeling approach, only the labeled peptides were selected and enriched for protein analysis, which eliminated the interference of unlabeled peptides on analysis.[14] Inspired by this method, we developed an enzymatic approach to label peptide N-termini with isotope-coded affinity tags. After purification by affinity chromatography, only the isotope-labeled peptides were submitted to reversed-phase liquid chromatography−tandem mass spectrometry (RP LC/MS/MS) analysis. In addition to one amino acid residue, some dipeptides with affinity tags were also introduced to the N-termini of peptides successfully. It was found that different tags had different influences on fragmentation, and thus highly complementary identification results were obtained. To the best of our knowledge, this is the first time that short peptide sequences have been ligated to the N-termini of tryptic peptides for proteome analysis.

EXPERIMENTAL SECTION

Materials. L-Arginine, trifluoroacetic acid (TFA), 2,5-dihydroxybenzoic acid (2,5-DHB), dithiothreitol (DTT), iodoracetamide (IAA), biotin *N*-hydroxysuccinimide ester (biotin-NHS), and trypsin [bovine, treated with L-(tosylamido-2-phenyl)ethyl chloromethyl ketone (TPCK)] were obtained from Sigma−Aldrich; $^{13}C_6$-L-arginine and monomeric avidin agarose were purchased from Thermo Electron; peptides (ALKAADTIGYPVMIR, VGKANEELAGVVAEVQK, YVKAGPWTPEAAVEHPEAVR VIFIEHAKRKG, KGGAKRHRKVLRD, GR, HR, and PR) were synthesized by ChinaPeptides Co. Ltd. (Shanghai, China); formic acid (FA) was provided by Fluka (Buchs, Germany); acetonitrile (ACN, HPLC-grade) was purchased from Merck (Darmstadt, Germany); all other chemicals including thionyl chloride ($SOCl_2$), methanol, and ethanol were of chromatographic grade. All the water used in this experiment was prepared on a Milli-Q system (Millipore, Bedford, MA).

Preparation of Tryptic Digests. Adult female C57 mice were purchased from Dalian Medical University (Dalian, China). The procedures for preparation of mouse liver lysate and trypsin digestion were performed as previously reported.[15] Protein samples (1 mg) were dissolved in 1 mL of denaturing buffer (pH 8.0) containing 8 M urea and 50 mM Tris-HCl. To the obtained protein solution, 20 μL of 50 mM DTT was added. The disulfide bond of protein was opened by incubation for 1 h at 56 °C. Then 40 μL of 50 mM IAA was added and the obtained solution was incubated for an additional 40 min at room temperature in the dark. After that, the mixture was diluted 8-fold with 50 mM Tris-HCl (pH 8.0) and incubated for 16−20 h at 37 °C with trypsin at an enzyme/substrate ratio of 1:40 (w/w) to produce a proteolytic digest. Finally, the tryptic peptides were desalted with a homemade C18 solid-phase extraction (SPE) column. All of the resulting peptide samples were then stored in the freezer at −30 °C for further usage.

Synthesis of Stable Isotope-Coded Affinity Tags. The synthesis of biotin-R-OEt began with biotinylation of arginine (R).[16] Arginine (1 mg) was dissolved in 300 μL of 0.25 M triethyl ammonium bicarbonate (TEAB)/75% ethanol, and then biotin *N*-hydroxysuccinimide ester was added (2.4 mg) and allowed to react at room temperature for 30 min. Residual reagent was quenched by adding water and allowing excess reagent to completely hydrolyze over an additional 30 min. The carboxyl group on the biotinylated arginine (biotin-R-OH) was then esterified with ethanol. Anhydrous ethanol with thionyl chloride (1.5 equiv) was refluxed at −10 °C for 2 h, biotin-R-OH was added to the ethanol solution and incubated at −10 °C for 2 h, and then the temperature gradually rose to room temperature and the mixture was refluxed at 60 °C for 6−8 h until the disappearance of solid. The generated crude product biotin-R-OEt was lyophilized and recrystallized twice in anhydrous ethanol−diethyl ether, and the purity was confirmed by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS; Figure S1, Supporting Information). The light and heavy stable isotopic biotin-R-OEt were derived from $^{12}C_6$-L-arginine and $^{13}C_6$-L-arginine, respectively. Biotin-GR-OEt, biotin-HR-OEt, and biotin-PR-OEt were synthesized in the same way by using the synthesized dipeptides from ChinaPeptides.

Trypsin-Catalyzed Enzymatic Labeling and Affinity Purification of Labeled Peptides. For labeling of tryptic peptide derived from synthetic peptide ALKAADTIGYPVMIR, 10 μg of this peptide was first digested by IM-trypsin (0.01 mg of IM-trypsin containing 0.5 μg of trypsin, see Supporting Information for details) in 50 μL of an aqueous solution (50 mM Tris-HCl, pH 8.0) for 4 h. Then 10 μg of the cleaved peptide was dried down and labeled with light biotin-R-OEt (0.16 nmol) by use of IM-trypsin (0.1 mg containing 5 μg of trypsin) in 50 μL of ethanol solution containing 4% aqueous solution of 0.1 M Tris-HCl (pH 8.0) for 6 h. After removal of the IM-trypsin by magnetic force, the labeled peptides were lyophilized for isolation of biotin-containing peptides. The dipepides biotin-GR-OEt, biotin-HR-OEt, and biotin-PR-OEt were labeled with the synthetic peptide ALKAADTIGYPVMIR in the same way.

For the labeling of tryptic peptides from mouse liver proteins, the same amount of tryptic digests of proteins (200 μg each) were separately labeled with light (biotin-R-OEt, 32 mM) and heavy [biotin-($^{13}C_6$)R-OEt, 32 mM] Arg tags in 50 μL of ethanol containing 4% (v/v) aqueous solution (0.1 M Tris-HCl buffer, pH 8.0). The two reactions were initiated by addition of the same IM-trypsin (containing 50 μg of trypsin) and incubated at 30 °C for about 10 h, and after removal of the IM-trypsin by magnetic force, the stable isotope-labeled peptides were 1:1 mixed and lyophilized.

The labeled peptides were reconstituted in 2× PBS (pH 7.2) and incubated with monomeric avidin agarose at room temperature for 30 min, and then the agarose was washed with washing buffer 1 (Milli-Q H_2O) and washing buffer 2 (50 mM ammonium bicarbonate, pH 8.3, and 20% methanol in H_2O), respectively. The bound biotinylated peptides were eluted with elution buffer (30% ACN/0.4% TFA in H_2O), and the eluate was immediately lyophilized and analyzed by LC/MS/MS.

dx.doi.org/10.1021/ac403060d | *Anal. Chem.* 2014, 86, 1170−1177

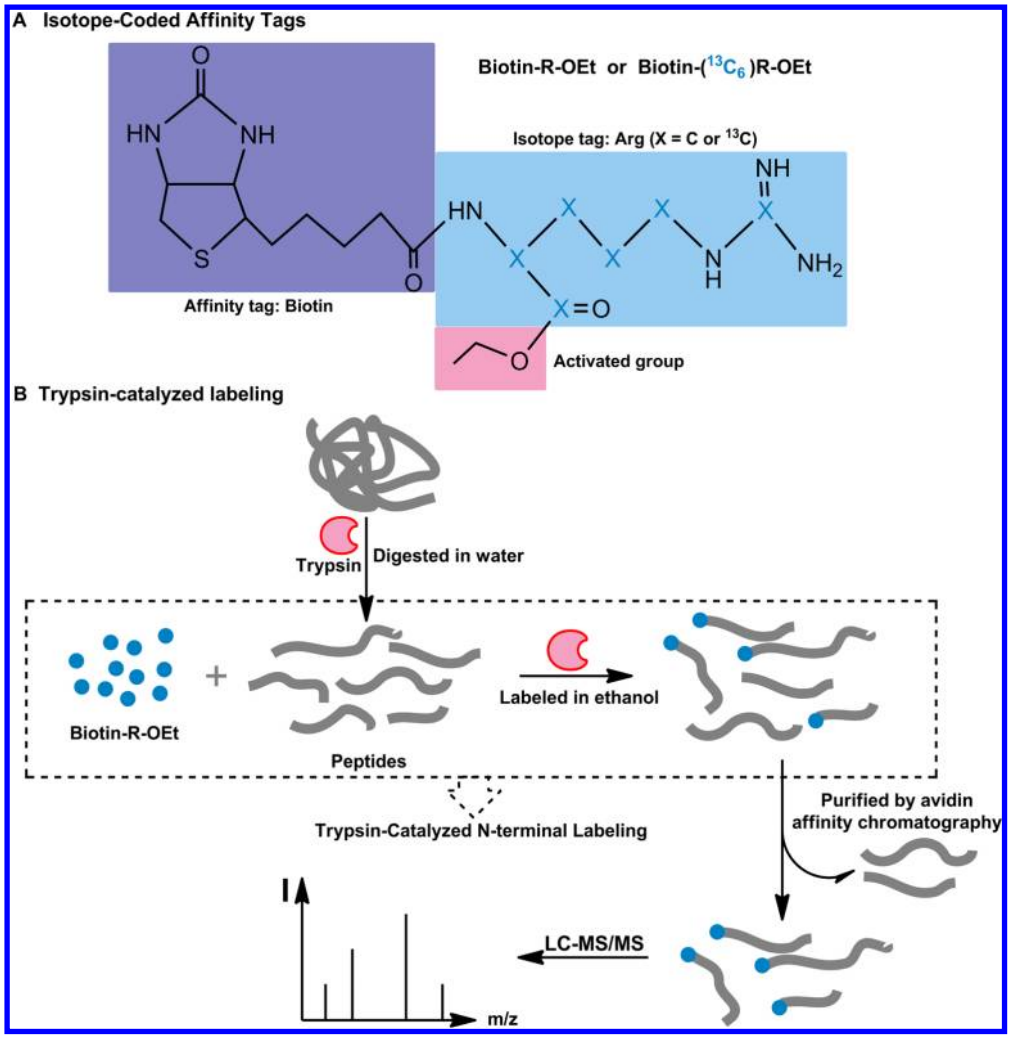


Figure 1. Schematic representation of trypsin-catalyzed N-terminal labeling. (A) Structure of the biotin-containing stable isotope tagging reagent, consisting of three elements: affinity tag (biotin), stable isotope tag (arginine molecule containing six $^{12}C_6$ or six $^{13}C_6$ indicated by X), and specific reactive group (ethyl ester). (B) Schematic for trypsin-catalyzed labeling approach.

MALDI-TOF MS Analysis. All MALDI-TOF MS analyses were performed on an AB Sciex TOF/TOF 5800 system mass spectrometer. All mass spectra reported were obtained in the reflex positive ion mode with delayed ion extraction. Sample aliquots (0.5 μL, 1 pmol) and 2,5-dihydroxybenzoic acid (DHB) matrix (0.5 μL) were spotted on the MALDI plate sequentially and dried at room temperature prior to MALDI-TOF MS analysis.

Nano LC/MS/MS Analysis. Reversed-phase (RP) LC/MS/MS system consisted of an LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific) with a nanospray source. A capillary column was first manually pulled to a fine point as spray tip and then packed with C18 AQ beads (3 μm, 120 Å, Michrom Bio Resources). Formic acid (0.1% v/v) in water and formic acid (0.1% v/v) in acetonitrile were applied as the mobile phase. 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 in one-dimensional (1D) RP LC/MS/MS. All MS and MS/MS spectra were acquired in data-dependent mode with the 10 most intense ions fragmented by collision-induced dissociation (CID) or electron-transfer dissociation (ETD).

Protein Identification and Quantification. Protein quantification was performed with MaxQuant (http://www.maxquant.org).[17] For stable isotope dimethyl labeling, the raw files were searched against a UNIPROT database of mouse (ftp.uniprot.org), carbamidomethylation on cysteine (+57.0215 Da) was set as a fixed modification, and oxidation of methionine (+15.9949 Da) was set as a variable modification. Peptide mass tolerance was set to 10 ppm and fragment mass tolerance was set to 0.5 Da. The peptide and protein false discovery rates (FDRs) were set to 0.01. Peptides were searched with fully tryptic cleavage constraints, and up to two missed cleavage sites were allowed. For quantification, lysine and peptide N-termini in light dimethylation (+28.03130 Da) and heavy dimethylation (+32.05641 Da) were set as light and heavy labels. The other settings were the same as the conventional search.

For trypsin-catalyzed enzymatic labeling, the raw files were searched against the corresponding peptide-centric database (PC-DB, derived from a UNIPROT database of mouse) according to our previous report.[5] This is because the conventional search strategy using protein databases lacks sensitivity because the extra fragment ions derived from the Arg tag are not considered, while the peptide-centric databases (PC-DBs) generated from protein databases can circumvent this problem. In brief, the PC-DBs were constructed as follows: mouse protein sequences downloaded from the UNIPROT databases were first digested in silico with trypsin, and the desired peptide sequences were then generated by concatenat-

dx.doi.org/10.1021/ac403060d | Anal. Chem. 2014, 86, 1170−1177

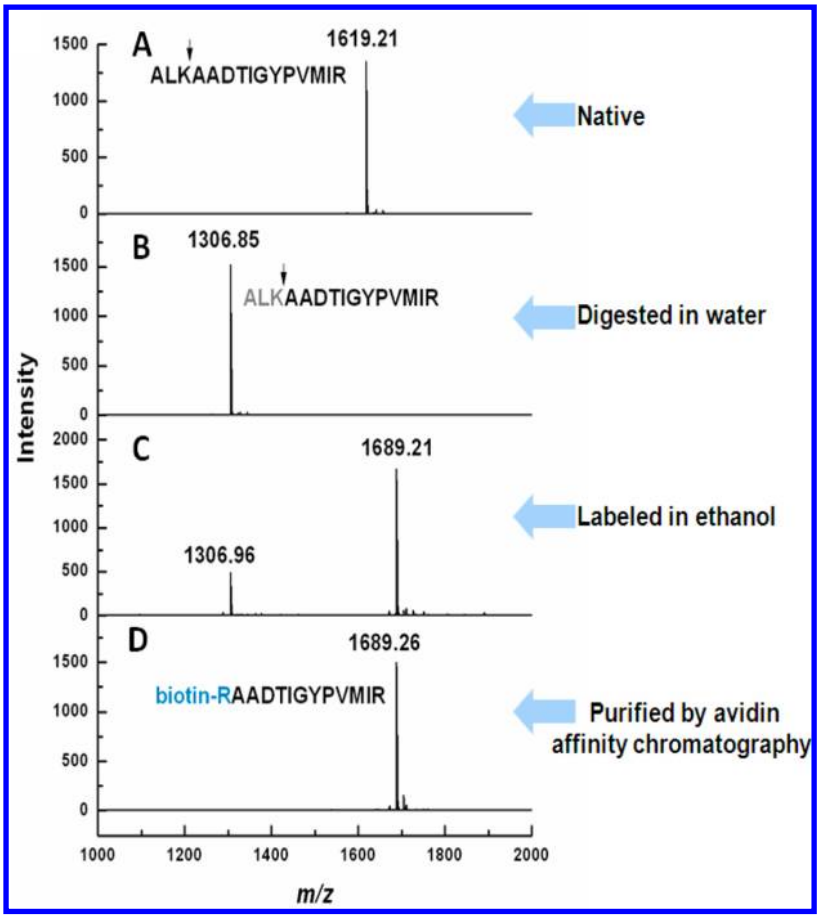


Figure 2. Example of trypsin-catalyzed N-terminal labeling with a synthetic peptide. Peptide ALKAADTIGYPVMIR (10 μg) was first digested by IM-trypsin (0.01 mg of IM-trypsin containing 0.5 μg of trypsin) in 50 μL of an aqueous solution (50 mM Tris-HCl, pH 8.0) for 4 h. Then 10 μg of the cleaved peptide was dried down and labeled with light biotin-R-OEt (0.16 nmol) by use of IM-trypsin (0.1 mg containing 5 μg of trypsin) in 50 μL of ethanol solution containing 4% aqueous solution of 0.1 M Tris-HCl (pH 8.0) for 6 h. MALDI-TOF MS analysis at each step in the process is provided.

ing an arginine at each N-terminus of tryptic peptide sequence, and each new peptide sequence was listed as a separate FASTA entry in the PC-DB. To identify the unlabeled peptides, each in silico digested tryptic peptide sequence was also listed as a separate entry in the PC-DB. Carbamidomethylation on cysteine (+57.0215 Da) was set as a fixed modification, oxidation of methionine (+15.9949 Da), and biotinylation (+226.0776 Da) of the N-terminus of the peptide were set as variable modifications. Initial peptide mass tolerance was set to 10 ppm and fragment mass tolerance was set to 0.5 Da. The peptide and protein false discovery rates (FDRs) were set to 0.01. An imaginary enzyme that would cleave C-terminal to an amino acid nonexistent in the database was used to avoid further digestion. For quantification, light arginine and heavy arginine ($^{13}C_6$-Arg, +6.0201 Da) were set as light and heavy labels, respectively. Protein ratios were calculated as the medium of all the peptide ratios. The other settings were the same as the conventional database search.

RESULTS AND DISCUSSION

This method exploits the ligase activity of trypsin. Trypsin was used as a catalyst to label tryptic peptides. An isotope-coded cosubstrate containing affinity tag, that is, isotope-coded affinity tag (ICAT), was synthesized for trypsin-catalyzed labeling (Figure 1A). Briefly, the primary amine group of arginine (Arg) containing either six ^{12}C ($^{12}C_6$) or six ^{13}C atoms ($^{13}C_6$) reacted with biotin-NHS to form the biotin affinity tag. The carboxyl group on Arg was activated to form ethyl ester. The modified Arg, biotin-R-OEt, was then used as the cosubstrate for trypsin-catalyzed labeling. The labeling strategy for quantitative protein analysis includes the following sequential steps (Figure 1B): (1) Proteins to be compared are first digested by free trypsin as a protease in an aqueous solution (1 M urea/50 mm Tris-HCl, pH 8.0) as in the conventional proteomics analysis. (2) The generated tryptic peptides are lyophilized and dissolved in an ethanol solution containing 4% of an aqueous buffer (0.1 M Tris-HCl, pH 8.0), and then to the above solution are added isotopically light (biotin-R-OEt) or heavy [biotin-($^{13}C_6$)R-OEt] forms of the cosubstrates and IM-trypsin for the trypsin-catalyzed ligation. (3) The tagged (biotin-containing) peptides are selectively isolated by avidin affinity chromatography. (4) Finally, the isolated peptides are analyzed by RP LC/MS/MS for identification and quantification. It should be mentioned that the enzyme trypsin was used twice here. First, trypsin acted as the protease for digestion as in conventional proteomics studies. Second, trypsin acted as the ligase to couple the tag to peptide N-termini. The shift between these two functions was achieved by using different reaction conditions and different substrates.

A standard peptide ALKAADTIGYPVMIR, which contains one trypsin cleavage site, was used to demonstrate this workflow (Figure 2). The peptide (10 μg) was digested by trypsin in aqueous buffer, which generated a new tryptic peptide (AADTIGYPVMIR) (compare Figure 2 panels A and B). To couple the affinity tag on the tryptic peptide, the digest was

dx.doi.org/10.1021/ac403060d | Anal. Chem. 2014, 86, 1170−1177

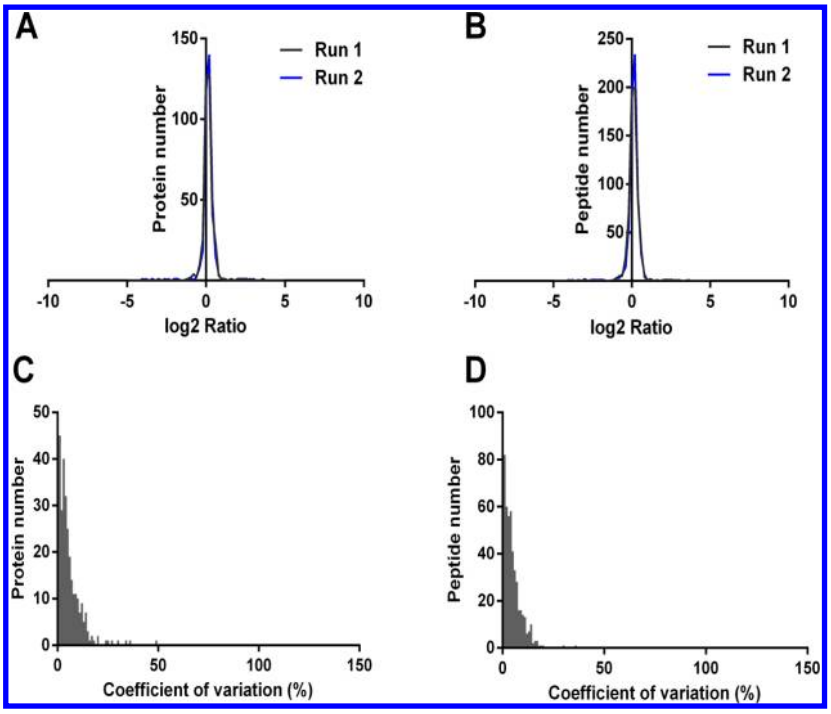


Figure 3. Distributions of quantified (A) protein and (B) peptide log2 ratios and histograms of the coefficient of variation of two technical replicate analysis of (C) protein and (D) peptide ratios between the two runs.

incubated with biotin-R-OEt (0.16 nmol) and immobilized trypsin (IM-trypsin) in ethanol solution containing 4% of an aqueous buffer (0.1 M Tris-HCl, pH 8.0) for 6 h. A mass shift of 382 Da was observed, indicating the addition of a biotin-R tag to the peptide. However, the unreacted peptide was also detected (compare Figure 2 panels B and C), indicating this trypsin-catalyzed reaction did not go to completion. To eliminate the possible interference of unreacted peptides on detection of labeled peptides, avidin affinity chromatography was applied to enrich the biotin-R-labeled peptides. In this study, the labeled peptides were captured by avidin beads and released with 30% ACN/0.4% TFA. Finally, only the labeled peptide containing a biotin-R tag was observed (Figure 2D). We also confirm the feasibility of the labeling approach with other two standard peptides, VGKANEELAGVVAEVQK and YVKAGPWTPEAA-VEHPEAVR. It was found that the tryptic peptides derived from these two peptides were successfully labeled with the affinity tag in ethanol solution as above (Figure S2, Supporting Information). The tryptic peptide (ANEELAGVVAEVQK, m/z = 1456.97) has one N-terminal primary amine and one side-chain primary amine on the lysine residue (Figure S2B, Supporting Information). The m/z values of the peptide labeled with one and two tags are 1839.15 and 2221.33, respectively. Only the peak corresponding to the peptide with one tag was observed in the MALDI-TOF MS spectrum, indicating this enzymatic approach labels only one tag per peptide. The MS/MS spectrum of this labeled peptide (Figure S2D, Supporting Information) indicated that the tag is coupled to the peptide N-terminus. Modification of lysine side chain is not observed in the above example. These results are consistent with the report that the trypsin-catalyzed labeling is specific for the primary amine on peptide N-terminus; other primary amine groups on the peptides except the N-terminal one cannot be labeled.[5] Tris-HCl buffer cannot be used for conventional amine-directed labeling because its primary amine also reacts with the labeling agents. However, Tris-HCl buffer was used here and no reaction was observed. This indicated that the enzymatic labeling is highly stereospecific. The sensitivity of the method was investigated with these three standard peptides (Figure S3, Supporting Information), and it was found that as little as 300 fmol of synthetic peptides could be labeled and detected.

The labeling strategy was further validated by labeling tryptic peptides from mouse liver proteins. Briefly, tryptic peptides (100 μg) from mouse liver tissues were labeled with biotin-R tag by incubation with biotin-R-OEt (1.6 nmol) and IM-trypsin (1 mg containing 50 μg of trypsin) in 50 μL of ethanol solution for 12 h. Then the biotin-R-labeled peptides were purified by avidin affinity chromatography for RP LC/MS/MS analysis. The acquired data were searched with MaxQuant by use of the peptide-centric database (PC-DB). This search resulted in identification of 788 unique peptides with a false positive rate less than 1%. Among these peptides, >97% were found to be labeled with the affinity tags, indicating the effectiveness of this labeling strategy. Analysis of unlabeled sample with the same LC/MS system led to the identification of 817 unique peptides, indicating that labeling did not compromise the performance for peptide identification. For comparison, the labeled sample without affinity purification was also analyzed in the same way, which resulted in the identification of 901 unique peptides. Among them, only 166 (18.4%) were labeled. After purification, the number of labeled peptides increased by 4.7-fold, indicating that interference by unlabeled peptides was almost eliminated. Though other blocking groups on the synthesized cosubstrate (the group coupled on the Arg residue via the reaction with the primary amine) have higher labeling efficiency, there are still about 10% of peptides unlabeled.[5] The enzymatic reaction cannot go to completion and the unlabeled peptides have serious interference with detection of labeled peptides. The affinity tag approach presented in this study effectively overcomes this problem.

The feasibility of this method for quantitative proteomics was investigated with complex proteomics samples. Proteins (400

dx.doi.org/10.1021/ac403060d | *Anal. Chem.* 2014, 86, 1170−1177

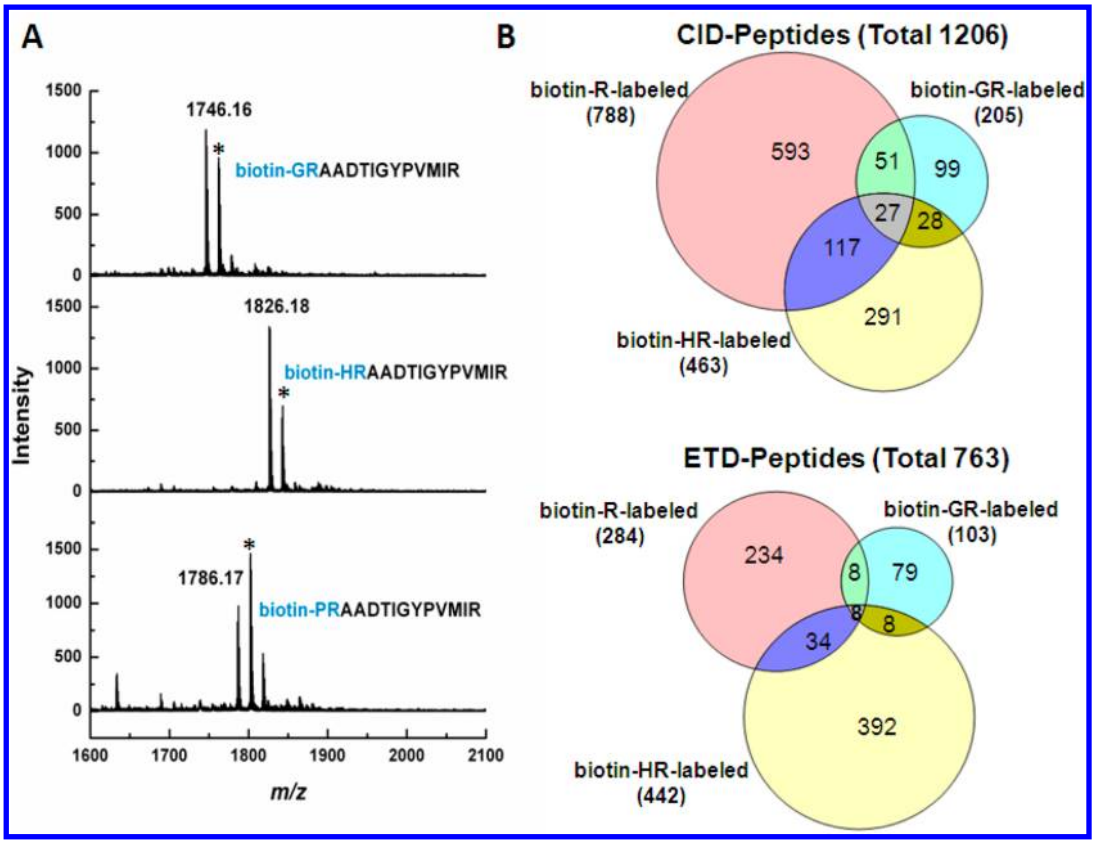


Figure 4. (A) Three dipeptides, GR, HR, and PR, were attached to N-termini of tryptic peptides through trypsin-catalyzed N-terminal labeling (peaks with oxidized methionine are marked with asterisks). (B) Overlaps of peptide and protein identifications among three different labels in CID and ETD modes.

μg) from mouse liver tissues were digested with trypsin, and then the same amount of tryptic peptides (200 μg each) was separately labeled with biotin-R-OEt and biotin-($^{13}C_6$)R-OEt, respectively. They were combined in a ratio of 1:1, and then the stable isotope biotin-R-labeled peptides were enriched by avidin affinity chromatography followed by 1D RP LC/MS/MS analysis. The acquired MS data were searched with MaxQuant in a mouse PC-DB for identification and quantification. The two replicate runs led to quantification of 377 and 364 proteins, respectively. The protein and peptide ratio distributions from the two runs were narrow (Figure 3A,B), and the average percentages of proteins and peptides within 2-fold abundance change, that is, the percentages of proteins and peptides with log2 ratio within [−1, 1], were 97.0% and 97.6%, respectively (Table S1, Supporting Information). The percentages of proteins and peptides within 2-fold abundance change were reported to be 85.5% and N/A for quantification by label-free methods, 93.5% and 86% for metabolic labeling, 98.5% and 98% for isobaric tags for relative and absolute quantitation (iTRAQ), and 99.5% and 99% for tandem mass tags (TMT).[18] These data indicated that the accuracy of this method is similar to other chemical labeling methods but better than label-free metabolic labeling methods. The percentage of protein and peptide ratio difference for a protein and peptide between the two runs in this study was determined (Figure 3C,D). It was found that the coverage for 50%, 30% and 20% variation was 99.6%, 98.6% and 97.2% for protein ratio, respectively. For peptide ratio, these percentages were 100%, 99.8%, and 99.6%, respectively. The reliability of iTRAQ based on technical, experimental, and biological replicate analysis was systematically assessed by Wright and co-workers,[19] and it was found that the technical replicate analysis produces only a coverage level of 95% at a cutoff point of 30% variation. In this study, the coverage for protein and peptide were 97.2% and 99.6% even at a cutoff point of 20% variation. The above data indicate that this method has excellent quantification performance at both the protein and peptide level.

Up to now, a variety of functional groups have been coupled onto tryptic peptides for proteomics analysis.[20−23] However, short peptides were never reported to label tryptic peptides. Coupling an oligopeptide to the N-termini of tryptic peptides via formation of peptide bonds by chemical approach is extremely challenging. Protection and deprotection of the side chains are required and some harsh reaction conditions are inevitably used, while the trypsin-catalyzed approach presented in this study is able to couple oligopeptides to the N-termini of tryptic peptides under mild conditions without any side-chain protection. To demonstrate the feasibility by labeling peptides with oligopeptides with this enzymatic approach, three cosubstrates derived from three dipeptides, biotin-GR-OEt, biotin-HR-OEt, and biotin-PR-OEt, were synthesized. These oligopeptide cosubstrates were separately incubated with a tryptic peptide (AADTIGYPVMIR) in the presence of IM-trypsin for enzymatic labeling. As shown in Figure 4A, three dipeptides with biotin affinity tags were ligated to the N-termini of the peptide successfully. Next we labeled tryptic peptides from mouse liver proteins with the dipeptide cosubstrates biotin-GR-OEt and biotin-HR-OEt, respectively. For comparison, the same sample was also labeled with biotin-R-OEt. After enrichment by avidin column, the labeled peptides were

dx.doi.org/10.1021/ac403060d | Anal. Chem. 2014, 86, 1170−1177

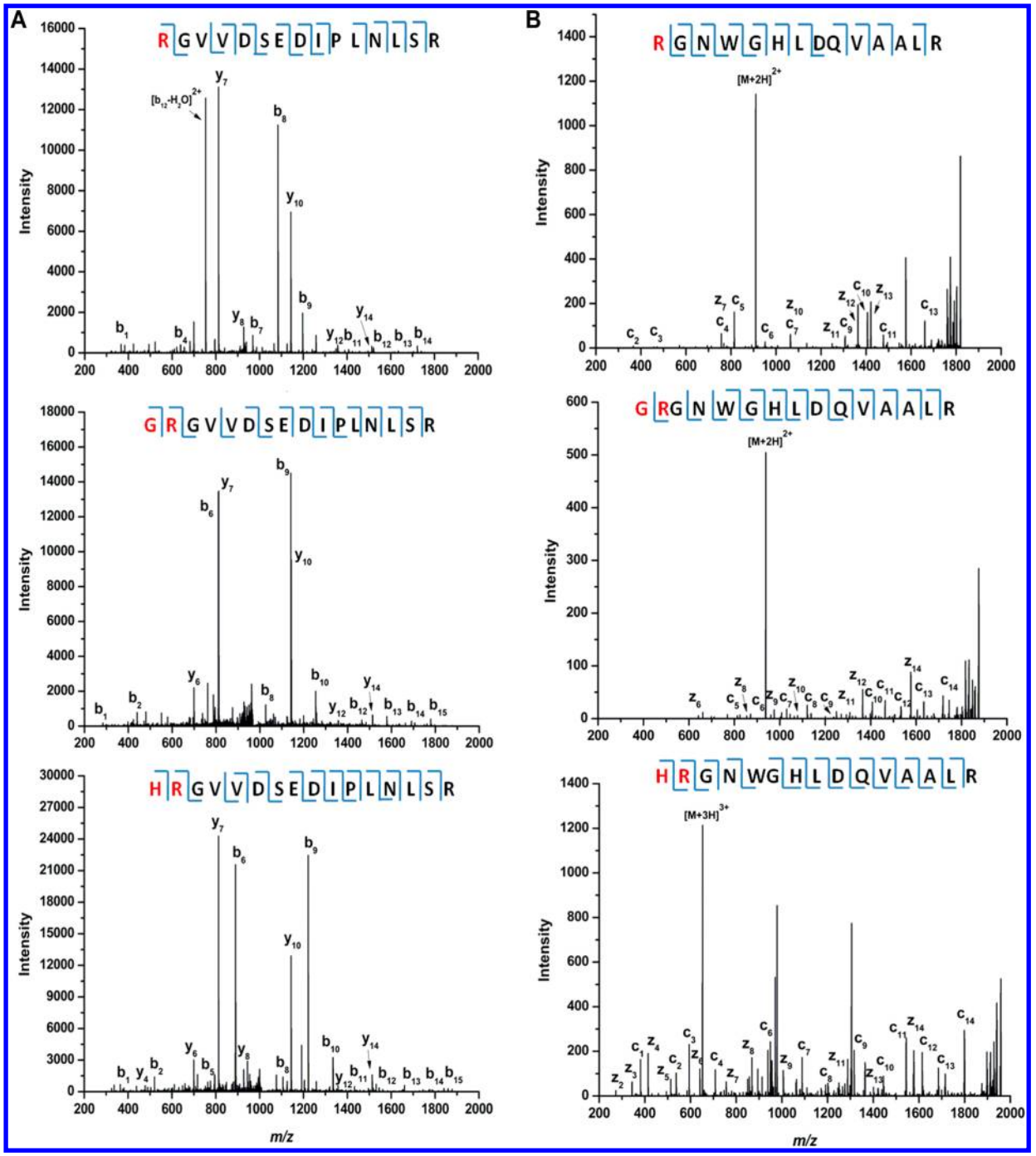


Figure 5. MS/MS spectra of the same peptide labeled with biotin-R, biotin-GR, and biotin-HR. (A) CID mode; (B) ETD mode.

analyzed on an LTQ-Orbitrap mass spectrometer in CID and ETD mode, respectively. Figure 4B gives the comparison of identification results for these three labeled samples. In the CID data set, biotin-R-OEt yielded the highest number of peptide identifications (788), biotin-HR-OEt yielded the second (463), and biotin-GR-OEt yielded last (only 205). In the ETD data set, biotin-HR-OEt yielded the highest number of peptide identifications (442), followed by biotin-R-OEt (284) and biotin-GR-OEt (103). The overlaps of peptide identifications between every two labeled samples were all <20%. However, the overlaps of two experiments for labeling with the same tag were as high as >51% (Figure S4, Supporting Information). The significant low overlap indicated that the identifications achieved by labeling with different tags are highly complementary.

Due to significant differences in physicochemical properties for these tags, the coupling of these tags onto peptides may significantly change their fragmentation behavior. Spectra for the same peptide identified with different tags for either ETD or CID are given in Figure 5. The patterns for theses spectra were significantly different (see Figure S5 in Supporting Information for matched ions and searching scores). In addition to the difference in peak intensity, the fragment ions were also different. For ETD spectra, the biotin-HR-labeled peptide yields many more fragment ions, and so the highest searching score of 78 was obtained. This is consistent with the fact that the highest

dx.doi.org/10.1021/ac403060d | Anal. Chem. 2014, 86, 1170−1177

number of peptide identifications was achieved for biotin-HR-labeled sample detected in ETD mode. ETD prefers the fragmentation of peptides with high charge states.[24] The biotin-HR tag is able to increase the charges of peptide ions and thereby improves the fragmentation of labeled peptides in ETD. To compare the difference in fragmentation behavior on a large scale, the fragment ion percentage distribution for all identified peptides is given in Figure S6 and Table S2 (Supporting Information). For identification with CID spectra, 44.5%, 56.3%, and 65.0% of peptides identified from biotin-R-, biotin-GR-, and biotin-HR-labeled samples, respectively, were found to have a percentage of b-ions (number of matched b-ions divided by the total number of b- and y-ions) over 50%. For identification with ETD spectra, 55.2%, 36.9%, and 61.9% of peptides identified from biotin-R-, biotin-GR-, and biotin-HR-labeled samples were found to have a percentage of c-ions over 50%. In general, biotin-HR-labeled peptides tend to generate more N-terminal fragment ions (b- or c-ions) for both dissociation modes. When ion distributions for the biotin-R- and biotin-GR-labeled peptides are compared, biotin-R-labeled peptides generate more N-terminal fragment ions in ETD, while biotin-GR-labeled peptides generate more N-terminal fragment ions in CID. The above data clearly indicate that coupling of these tags onto peptides significantly changes their fragmentation behavior. The high complementary identification results also confirm the allegation above. Different peptides have different fragmentation behaviors. Ligation of different oligopeptides to tryptic peptides is presented as a promising method to tailor the fragmentation of tryptic peptides. The enzymatic approach developed in this study enables the specific ligation of oligopeptides to tryptic peptides.

CONCLUSIONS

In this study, we present an enzymatic approach to label peptides with stable isotope-coded affinity tags for proteome analysis. Using a sulfhydryl-reactive group, the conventional stable isotope-coded affinity tag approach (ICAT approach) targets the side chains of cysteine residues on peptides/proteins; while in this trypsin-catalyzed approach, the tags were specifically linked to the N-termini of tryptic peptides. This method effectively reduced the interference for detection of labeled peptides by unlabeled peptides. In addition to Arg residue, oligopeptides with C-terminal Arg residue were introduced to the N-termini of tryptic peptides for the first time in proteomics analysis. It was found that these tags significantly changed the fragmentation behavior of peptides in MS, especially in the ETD mode. Therefore, coupling a short sequence tag onto peptides could be an effective approach to improve the coverage for proteome analysis.

ASSOCIATED CONTENT

Supporting Information

Additional text describing preparation of immobilized trypsin; six figures showing synthesis of isotope-coded affinity tags, examples of trypsin-catalyzed N-terminal labeling, sensitivity of the method, overlaps of identification, MS/MS spectra of labeled peptides, and b- and c-ion percentage distributions; and two tables listing protein quantification results and percentages of identified peptides. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*Phone +86-411-84379620; fax +86-411-84379620; e-mail mingliang@dicp.ac.cn.

*Phone +86-411-84379610; fax +86-411-84379620; e-mail hanfazou@dicp.ac.cn.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank PNNL and the omics.pnl.gov Web site for Venn diagrams. This work was supported by the China State Key Basic Research Program Grant (2013CB911202, 2012CB910101, and 2012CB910604), the Creative Research Group Project of NSFC (21321064), the National Natural Science Foundation of China (21235006, 21275142, and 81161120540), and National Key Special Program on Infection diseases (2012ZX10002009-011) and Analytical Method Innovation Program of MOST (2012IM030900).

REFERENCES

(1) Ong, S. E.; Mann, M. *Nat. Chem. Biol.* **2005**, *1*, 252−262.
(2) Yates, J. R., III. *J. Am. Chem. Soc.* **2013**, *135*, 1629−1640.
(3) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Mol. Cell. Proteomics* **2002**, *1*, 376−386.
(4) Jensen, O. N. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 391−403.
(5) 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.
(6) Sergeant, K.; Samyn, B.; Debyser, G.; Van Beeumen. *J. Proteomics* **2005**, *5*, 2369−2380.
(7) Wilson, J. J.; Brodbelt, J. S. *Anal. Chem.* **2006**, *78*, 6855−6862.
(8) Ji, C.; Guo, N.; Li, L. *J. Proteome Res.* **2005**, *4*, 2099−2108.
(9) Qin, H.; Wang, F.; Zhang, Y.; Hu, Z.; Song, C.; Wu, R. a.; Ye, M.; Zou, H. *Chem. Commun.* **2012**, *48*, 6265−6267.
(10) Wehofsky, N.; Haensler, M.; Kirbach, S. W.; Wissmann, J. D.; Bordusa, F. *Tetrahedron: Asymmetry* **2000**, *11*, 2421−2428.
(11) Bordusa, F. *Chem. Rev.* **2002**, *102*, 4817−4868.
(12) Wehofsky, N.; Koglin, N.; Thust, S.; Bordusa, F. *J. Am. Chem. Soc.* **2003**, *125*, 6126−6133.
(13) Oka, T.; Morihara, K. *J. Biochem.* **1977**, *82*, 1055−1062.
(14) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994−999.
(15) Song, C.; Wang, F.; Ye, M.; Cheng, K.; Chen, R.; Zhu, J.; Tan, Y.; Wang, H.; Figeys, D.; Zou, H. *Anal. Chem.* **2011**, *83*, 7755−7762.
(16) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S. *Mol. Cell. Proteomics* **2004**, *3*, 1154−1169.
(17) Cox, J.; Mann, M. *Nat. Biotechnol.* **2008**, *26*, 1367−1372.
(18) Li, Z.; Adams, R. M.; Chourey, K.; Hurst, G. B.; Hettich, R. L.; Pan, C. *J. Proteome Res.* **2012**, *11*, 1582−1590.
(19) Gan, C. S.; Chong, P. K.; Pham, T. K.; Wright, P. C. *J. Proteome Res.* **2007**, *6*, 821−827.
(20) Brittain, S. M.; Ficarro, S. B.; Brock, A.; Peters, E. C. *Nat. Biotechnol.* **2005**, *23*, 463−468.
(21) Tian, R.; Li, L.; Tang, W.; Liu, H.; Ye, M.; Zhao, Z. K.; Zou, H. *Proteomics* **2008**, *8*, 3094−3104.
(22) Sohn, C. H.; Lee, J. E.; Sweredoski, M. J.; Graham, R. L. J.; Smith, G. T.; Hess, S.; Czerwieniec, G.; Loo, J. A.; Deshaies, R. J.; Beauchamp, J. *J. Am. Chem. Soc.* **2012**, *134*, 2672−2680.
(23) Gu, C.; Shannon, D. A.; Colby, T.; Wang, Z.; Shabab, M.; Kumari, S.; Villamor, J. G.; McLaughlin, C. J.; Weerapana, E.; Kaiser, M. *Chem. Biol.* **2013**, *20*, 541−548.
(24) Molina, H.; Horn, D. M.; Tang, N.; Mathivanan, S.; Pandey, A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2199−2204.