Published in final edited form as:

Rapid Commun Mass Spectrom. 2014 March 30; 28(6): 635–644. doi:10.1002/rcm.6820.

Solid-phase N-terminal peptide enrichment study by optimizing trypsin proteolysis on homoarginine modified proteins by mass spectrometry

Saiful M. Chowdhury^{1,*}, Gerhard R. Munske², Jonathon Yang¹, Daria Zhukova¹, Hamilton Nguen¹, and James E. Bruce^{3,*}

¹Department of Chemistry and Biochemistry, University of Texas at Arlington, Arlington, TX

²School of Molecular Biosciences, Washington State University, Pullman, WA

³Department of Genome Sciences, University of Washington, Seattle, WA

Abstract

Rationale—Proteolytic cleavages generate active precursor proteins by creating new N-termini in the proteins. A number of strategies recently published regarding the enrichment of original or newly formed N-terminal peptides using guanidination of lysine residues and amine reactive reagents. For effective enrichment of N-terminal peptides, the efficiency of trypsin proteolysis on homoarginine (guanidinated) modified proteins must be understood and simple and versatile solid-phase N-terminal capture strategies should be developed.

Methods—We present here a mass spectrometry-based study to evaluate and optimize the trypsin proteolysis on a guanidinated modified protein. Trypsin proteolysis was studied using different amount of trypsin to modified protein ratios. To capture the original N-termini, after guanidination of proteins, original N-termini were acetylated and the proteins were digested with trypsin. The newly formed N-terminal tryptic peptides were captured with a new amine reactive acid-cleavable solid-phase reagent. The original N-terminal peptides were then collected from the supernatant of the solution.

Results—We demonstrated a detailed study of the efficiency of enzyme trypsin on homoarginine modified proteins. We observed that the rate of hydrolysis of homoarginine residues compared to their lysine/arginine counter parts were slower but generally cleaved after an overnight digestion period depending on the protein to protease concentration ratios. Selectivity of the solid-phase N-terminal reagent was studied by enrichment of original N-terminal peptides from two standard proteins, ubiquitin and RNaseS.

Conclusion—We found enzyme trypsin is active in guanidinated form of protein depending on enzyme to protein concentrations, time and the proximity of arginine residues in the sequence. The novel solid-phase capture reagent also successfully enriched N-terminal peptides from the standard protein mixtures. We believe this trypsin proteolysis study on homoarginine modified proteins and our simple and versatile solid-phase capture strategy could be very useful for enrichment and sequence determination of proteins N-termini by mass spectrometry.

Corresponding author: Saiful M. Chowdhury, PhD, Chemistry and Biochemistry, Box 19065, University of Texas at Arlington, 700 Planetarium Place, Rm 130, Arlington, TX 76019, schowd@uta.edu. James E. Bruce, Ph. D, Department of Genome Sciences, Box 358050, University of Washington, 815 Mercer Street, Seattle, WA 09109, jimbruce@u.washington.edu.

Introduction

Proteolytic cleavage is one of the most common types of posttranslational modifications. Some proteins generate their active precursor after proteolytic cleavage of certain amino acid residues. The most common proteolytic cleavage is the removal of the initial methionine residue from the N-terminus¹. Another example of proteolytic cleavage is the class of proteins synthesized by ribosome and associated with the membranes of the endoplasmic reticulum (ER). The N-termini of these classes of proteins contain sequences, which are commonly known as signal sequences or signal peptides². The signal peptides generally consist of 13–36 hydrophobic residues, which are typically removed following passage through ER. The removal of the signal peptide is catalyzed by an enzyme called signal peptidase, which generates new N-termini of the proteins³. Proteins are also sometimes endo-proteolytically cleaved, which results in the generation of biologically active proteins. Thus, N-terminal sequence of a protein is very important with respect to protein structure and function. Studying the sequence of N-termini of proteins can help us determine whether or not initiator methionine residues or signal peptides are clipped off during N-terminal processing of proteins. It can also assist us in clarifying the type and location of modifications present on the N-terminus. In addition, if proteins are endo-proteolytically cleaved, their new N-termini can reveal the protease activity sites on the proteins.

Several proteomics approaches recently published sort N-termini and identifications of protein through the use of mass spectrometry. For selective labeling of the N-termini of proteins, most methods block the lysine primary amine with acetylation reaction. After that, proteins are digested with trypsin and N-terminal peptides are enriched utilizing selective chemical reactive groups targeted for the original N-termini of proteins. Gevaert *et al.* developed a two-stage chromatographic method to reduce the sample complexity before sorting free N-terminal peptides⁴. The method results in the blocking of the free N-termini of proteins by an acetylation reaction. After tryptic digestion of proteins the primary RP-HPLC fractionation of peptide mixture was carried out. The peptide mixtures were then treated with 2, 4, 5-trinitrobenzenesulphonic acid (TNBS), which results in a strong hydrophobic shift for peptides in the second HPLC separation. McDonald *et al.* reported a method to enrich N-terminal peptides utilizing acetylation as a blocking reaction⁵. However, rather utilizing two-stage chromatographic method in their approach, the new N-termini generated after tryptic digestion were biotinylated and then removed from the solution by binding with immobilized avidin beads. The original N-termini (that were acetylated, but not biotinylated) remained in the solution, were collected from the flow-through of the avidin column and were analyzed by MALDI mass spectrometry.

Guanidination of tryptic peptides with isomethyl urea is currently a routine procedure in proteomics research to increase the intensity of peptides in MALDI-TOF-MS^{6–9}. This modification is now widely being used in N-terminal peptide enrichment studies. The advantage of homoarginine modification of lysine is that it retains charge during mass spectrometric analysis whereas acetylation does not.

An N-terminal peptide enrichment method was reported for *de novo* sequencing of N-terminal fragment by PSD-MALDI-mass spectrometry that utilized guanidination reaction to modify lysine residues in the proteins to homoarginines¹⁰. The authors employed a biotinylated cysteic acid (BCA) to incorporate a sulfonic acid group into the N-termini of the proteins. The labeling procedure was performed after the proteins were blotted on a PVDF membrane which allowed efficient sample cleaning from the reagents. The original N-termini were then captured utilizing biotin-avidin affinity chromatography.

Recently “Overall group” reported an N-terminal peptide enrichment study using isotope labeled formaldehyde¹¹. Although formaldehyde is a good labeling reagent for primary amine but it is the most commonly used reagent for protein-protein and protein- DNA cross-linking¹². However, it is non-specific and reacts also with tyrosine and arginine residues to a certain extent. Moreover, it requires a NaCNBH₃ reduction step to add dimethyl in the lysine residues. Thus, blocking primary amine group with formaldehyde will be sample dependent and cross-linked products can also obstruct enrichment of N-termini. In their study, after labeling the primary amine, the digested peptides were captured using a HPG-ALD polymer leaving the N-terminal peptide in the supernatant. The HPG-ALD step also requires NaCNBH₃ reduction to form covalent linkage. Although, they demonstrated a successful enrichment study but the method requires several chemicals labeling steps and we believe will be sample dependents.

Zhang et al. reported a similar reverse purification of N-acetylated peptides using a commercial CNBr activated sepharose. This commercial resin has several drawbacks. A pH adjustment is necessary to avoid reactions of activated resin with α -NH₂ group of lysines. In addition, they also mentioned unexpected reaction of CNBr activated beads with histidine¹³. Another method recently reported by Kim et al. while our paper was in review¹⁴. The method utilized guanidination reaction of proteins followed by labeling of N-terminus by SATA (N-succinimidyl S-acylthioacetate) reagent. It also required a NH₂OH treatment to incorporate thiol group in the protein N-termini. The thiol containing N-termini was then captured with a thiol reactive DTT cleavable solid-phase reagent. There is strong concern about disulphide bond formation during trypsin digestion using this method and having DTT in the solution can also cleave disulfide bond in the resin, hence deactivating the resin for efficient capture. The paper also demonstrated the efficient conditions for guanidination of lysine in the proteins and mentioned about cleavage of homoarginyl residues by trypsin. No detailed study has been reported on the cleavage efficiency of homoarginine modified sites by trypsin. Although useful, all these methods utilized several additional separation and chemical labeling steps for enrichment of N-terminal peptides.

Enrichment with solid-phase reagent is emerging as an efficient strategy to reduce the sample complexity in mass spectrometry based proteomics experiment^{15–17}. The advantage of solid-phase reagent is that it helps to avoid several sample preparation steps. In addition, stringent washing conditions can be applied which can significantly reduce non-specific contaminants. At present, there are very few cleavable reagents are available for N-terminal peptide enrichment studies.

To develop an effective enrichment strategy of N-terminal peptides utilizing homoarginine modification, a detailed study must be needed for the trypsin cleavage efficiency of the homoarginine modified proteins. In addition a simple and versatile N-terminal solid-phase capture reagent will help to avoid several chemical labeling steps in currently existed methods. In this report, we demonstrated a novel acid-cleavable solid-phase N-terminal selective reagent and a simple strategy for selective enrichment of N-terminal peptides utilizing homoarginine modification procedure. A detailed study was also presented to evaluate the efficiency of trypsin proteolysis on homoarginine modified proteins. After optimization of trypsin cleavage on modified proteins, the newly formed N-termini were captured with an acid cleavable solid-phase reagent. The original N-termini, which were blocked by an acetylation reaction in a previous step, remained in the supernatant of the solution. The solid-phase capture reagent utilized was constructed with a primary amine reactive functional group and with an acid-cleavable group in the backbone. This allowed the secondary option of release of captured peptides from the beads. Initial validations of this solid-phase capture strategy were demonstrated by selective isolation of N-terminal peptides from a mixture of Ubiquitin and RNaseS proteins. We believe this study will

contribute in current proteomics research for understanding the trypsin proteolysis in homoarginine modified proteins as well as determination of the original N-terminal protein sequences and also new sequences generated by protease activities.

Materials and Methods

Ubiquitin (bovine erythrocytes), β -casein (bovine milk), RNase S proteins (bovine pancreas) were purchased from Sigma (St. Louis, MO, USA) and isomethylurea hemisulfate was obtained from Acros Organics (Morris Plains, NJ, USA). Sequence grade modified trypsin was used for proteolysis and purchased from Promega (Madison, WI). The samples were desalted by ZipTip (C4, C18) as well as SepPack (C18) desalting column (Water, Milford, MA, USA). Mass spectra were obtained with a Bruker Daltonics, Omnistar MALDI-TOF mass spectrometer in reflectron or linear mode, unless otherwise specified. All FTICR-MS spectra were obtained with Bruker Daltonics, APEX Q-FTICR mass spectrometer by direct infusion of samples in a nano ESI tip made with a fused capillary (360 μ m o.d and 20 μ m i.d). The capillary tip was etched with 49% HF. Shimadzu MALDI-IT/TOF was also utilized to reproduce some of the FTICR mass spectra obtained for studying the tryptic cleavage of homoarginine modified proteins. For MS/MS studies, precursor peptides were isolated ($\pm$ 1Da) in the quadrupole and fragmented in the hexapole and ions were transmitted to ICR cell. Hexapole collisional trap voltage was kept around 25–30 kV for efficient fragmentation. All the spectra were saved with 512k data points from 10 scans. The FT-ICR spectra were processed with the software, ICR-2LS, developed by Pacific Northwest National Laboratory (PNNL) and externally calibrated using known standard peptides and peptides from the tryptic digest of β -casein¹⁸. MALDI-TOF and MALDI-IT/TOF mass spectra were obtained with the matrix α -cyano-4-hydroxycinnamic acid or DHB with 50% laser intensity and ESI-FTICR-MS spectra were obtained with 49% MeOH/H₂O/2% acetic acid solution.

Guanidination reaction of proteins

Modification of lysine to homoarginine was carried out by commonly used procedure improved by Beardsley *et al*⁶. Ubiquitin and β -casein (1 μ L, 1 mM) were taken separately into a micro centrifuge tube. 30 μ L of 7N NH₄OH was added to each tube. The pH of the solutions was maintained above ten. A 5 μ L of fresh isomethylurea (50 mg dissolved in 50 μ L of H₂O) was then added to each tube. The reactions were conducted at 65°C for 30 min. A solution of 1% TFA was added to stop the reactions. The reaction solutions were evaporated to dryness under vacuum. The lysine modified proteins were reconstituted in 20 μ L of water/0.1% TFA. Excess isomethylurea was removed with SepPack (C18) column. The protein solutions were evaporated to dryness under vacuum and reconstituted in a 50 mM, 500 μ L NH₄HCO₃ solution. Trypsin solution was added by maintaining the following proteases to protein ratios: 1:250, 1:100, and 1:25. The digestions were performed over 12h periods. The NH₄HCO₃ salts were removed from the digestion solutions by evaporating the solutions under vacuum for overnight. Time course studies of trypsin proteolysis were done under same conditions described as above using 1:50 proteases to protein ratio.

Synthesis of the solid-phase reagent

Aminopropyl modified glass beads (Sigma, 0.162 μ mol/g) were reacted with Fmoc Rink-OH (NovaBiochem) in an Applied Biosystems 431 peptide synthesizer using standard peptide synthesis chemistry. Next, the Fmoc group was removed and the amine group was modified with succinyl anhydride to introduce a carboxyl group. The activated ester was formed using DCC (Dicyclohexylcarbodiimide, Applied Biosystems) and N-hydroxysuccinimide was coupled using common peptide synthesis protocol. A solid-phase

reagent of similar structure utilizing maleimide group was reported previously, using PEG as solid support¹⁹.

Selectivity of the solid-phase reagent

β -casein tryptic digest (1 μ L, 100 μ M) was placed in a microcentrifuge tube. The conversion of lysine residues to homoarginines was achieved by the protocol described in the experimental section. After conversion, the solution was dried in a vacuum to remove NH_4OH from the solution. The solution was reconstituted in 200 μ L of water, and isomethylurea salt was removed with a SepPack (C18) desalting column. After complete removal of salt, the solution was dried in a vacuum and then reconstituted in 200 μ L of $\text{MeOH}:\text{H}_2\text{O}$ (1:1) water solution. Acetic anhydride, 10 μ L was added in the solution. The reaction was carried out for 10 min. at room temperature. The N-term acetylated peptides were then dried in vacuum and 1 μ L Bradykinin (100 μ M) was spiked into the solution (1:1 ratio). The peptides were redissolved in 200 μ L of PBS (pH 7.2). In this solution, a 10 mg of solid-support reagent was added and the reaction solution was stirred for 1h. The beads were then washed several times with water, isopropanol and acetonitrile. The peptides captured on the beads were released by treating the beads with a solution of TFA: CH_2Cl_2 (80:20). The supernatant solution was collected and evaporated to dryness. The sample was reconstituted in a methanol:acetic solution to obtain ESI-FTICR-MS spectra.

Selective enrichment of blocked peptides

Bovine ubiquitin (85 μ g) was modified with isomethylurea as described above. After desalting, the protein was reconstituted in 200 μ L water. The solution was then fractionated into two equal volumes. In one fraction all N-terminal peptides were acetylated as described above. Both fractions are combined. The capture of free N-terminal peptides was achieved with the solid-phase reagent as described above. The supernatant was then collected and MALDI-TOF mass spectra were obtained.

Enrichment of N-terminal peptides

RNase S (5 μ L, 1mM), ubiquitin (10 μ L, 1 mM), were placed in a microcentrifuge tube. Performic acid was prepared by mixing 100 μ L of 30% (v/v) H_2O_2 solution with 400 μ L of 88% formic acid. After that solution was kept at room temp. for 30 min and cooled to 0°C. A 200 μ L solution of performic acid was added and the reaction was kept at 4 °C for 4h. The reaction solution was evaporated to dryness. The proteins were guanidinated and acetylated as described before. After acetylation, proteins were digested with trypsin overnight using 1:100 protease to protein ratio. The digested solution was then captured with the solid phase beads for 2h. Every 30 min, fresh 5 mg of new beads were added for complete capture of peptides with unblocked N-termini. The supernatant was collected and desalted with a SepPack desalting column. MALDI-TOF and ESI-FTICR-MS spectra were obtained.

Results and discussion

To develop a simple procedure for selective enrichment of N-terminal peptides, we first utilized performic acid oxidation to cleave and oxidize disulphide bonds in the proteins. After that all the lysine residues in the proteins were modified to homoarginines by guanidination reaction. In the next step, original N-termini of proteins were blocked by an acetylation reaction. The homoarginine modified proteins were then digested with trypsin. The resulting tryptic digest should contain original N-terminal peptides, which will be in the solution as N-terminal acetylated peptides. The newly formed N-termini due to tryptic cleavage are then captured with a primary amine reactive solid-phase acid-cleavable reagent (Figure 1). The solid-phase reagent contains an acid-cleavable bond, which also gave the option of releasing the captured tryptic peptides from the bead. The original N-terminal

peptides were collected from the supernatant of the solution and identified by mass spectrometry (Scheme 1).

Trypsin cleavage of homoarginine modified proteins has been reported in early literature^{20,21}. At this time, no proteomics study has taken this point in consideration when working on this modification. Most of the studies utilized this modification on peptide levels^{22,23}. To consider trypsin-cleaved homoarginine terminated peptides for our enriched N-terminal peptides search, we performed trypsin digestion study of two different homoarginine modified proteins, ubiquitin and β -casein. At first, a small protein, bovine ubiquitin (~8.6 kDa) was used for this study. After modification of lysines to homoarginines, the modification reagent isomethylurea was removed using ZipTip purification (C18, C4 reverse phase) (Millipore, Billerica, MA). After loading the sample on the ZipTip, several washing steps were done to remove isomethylurea hemisulfate salts completely from the solution. The salt removal was also achieved through the use of a SepPack (C18) desalting column. Tryptic digestion of homoarginine modified proteins was performed under various protease to protein ratios (1:250, 1: 100, and 1:25). In the ESI-FTICR-MS spectrum from low protease-to-protein ratio experiments (1:250), all the peptides we observed in the mass spectrum were generated from the C-terminal tryptic cleavage at arginine residues (Figure 2, A). It was obvious, trypsin preferred cleavage at arginine residues with one or two missed cleavages at the homoarginine sites. The list of peptides generated is shown in supplementary table S1. It is clear from the list of peptides in table S1 that in a 12h digestion period trypsin preferred cleavage on the arginine residues in the modified proteins (Figure 2, A).

Increasing protease to protein ratio (1:100) for digestion (12h) generated several homoarginine terminated peptides (Figure 2, B). Several peptides were also generated with C-terminal arginine residues with one or two missed cleavages at the homoarginine sites. The peptides generated from this protease to protein ratio are listed in table S2. The tryptic peptide residue 1–6 (MQIFVK*), at m/z 765.43, was modified to homoarginine (m/z 807.45) and cleaved by trypsin. MS/MS analysis of the m/z 807.45 confirmed the cleavage at the homoarginine sites (Figure 2, A, also supplementary figure S1, A). The most intense peak in the spectrum shown in Figure 2, C, was isolated (m/z 694.88, +2), and fragmented in the collisional cell (Figure S1, B). The sequence was confirmed for the peptide LIFAGK*QLEDGR, which was produced by trypsin cleavage at the arginine residue with one missed cleavage at the homoarginine modified lysine. The homoarginine terminated peptide was also observed for this peptide in the ESI-FTICR-MS spectrum of protein digest (accurate m/z 690.43, +1, LIFAGK*). This peptide (m/z 690.43) was not observed in low protease to protein ratio (1:250) but the peptide LIFAGK*QLEDGR (m/z 698.88) was observed (Figure 2, A).

We performed another tryptic digestion of modified protein with higher protease to protein ratio (1:25). The complete list of peptides generated from the same digestion time (12h) showed more homoarginine terminated peptides with no missed cleavages (Figure 2, C). The generation of more homoarginine terminated peptides through an increase in the protease concentration is shown in table S3. Complete sequence coverage for modified protein ubiquitin was also observed with no missed cleavage in both homoarginine and arginine residues when using a ratio of 1:25. The peptides observed with one or two missed cleavages at 1:100 protease to protein ratio, also showed cleavages at the homoarginine residues. One of these peptides is shown in Figure 2, B, C inset. With a 1:100 ratio, the peptide TLTGK*, m/z 561.34 (+1) (Figure 2, B **inset**) was observed. Following an increase in the concentration (1:25), the homoarginine terminated peptide, TLSYNIQK*, m/z 562.29 (+2) appeared in that range (Figure 2, C **inset**).

Modification with isomethylurea for larger peptides is slower as reported by Beardsley *et al*⁶⁶. Complete modification of ubiquitin was attempted by increasing the reaction time to 30 min. We found most of the lysine-containing peptides were modified by this guanidination procedure. A few lysine residues were found partially modified. The ratio of some of the modified and unmodified peak varies from 1:7 to 1:9 ratios. We did not try to increase reaction time more than 30 min due to the fact that in longer reaction time some of the amino acid residues besides lysine can also be guanidinated by isomethylurea.

The data from three different tryptic digestions of modified ubiquitin with various proteases to protein ratios suggested that increasing the protease concentration generated more peptides involving the subsequent hydrolysis of homoarginine sites. The number of C-terminal homoarginine peptides observed with low protease to protein ratio was zero whereas increasing the concentration of protease generated almost three times the number of homoarginine terminated peptides. With higher concentration of protease, the number of observed arginine and C-terminal homoarginine peptides both were increased. Our study suggests that under high concentration of protease to protein ratio within a 12h time period, enzyme trypsin is active on guanidinated form of protein.

In order to better understand the rate of trypsin-cleavage at homoarginine and arginine residues in modified ubiquitin, we performed a time-course study on homoarginine modified ubiquitin. Homoarginine modified protein ubiquitin was treated with trypsin at 1:50 protease to protein ratio. Aliquots were taken from the solution at 0.5 h, 1.5 h, 2.5 h and 3.5 h. Trypsin activity was stopped by adding 1 μ L 1% TFA. At 30 min, the MALDI-TOF-MS spectra showed that most of the peptides generated were cleaved at arginine residues. A homoarginine terminated peptide (m/z 807.45, +1) of very low intensity was also observed (Figure S2). We have observed that intensity of this peptide was increased with reaction time. Although the sample concentration varies from point to point in the MALDI plate, several laser shots (50 shots) were utilized at several positions to maximize the intensity for every time point. The time-course proteolysis study with MALDI-TOF mass spectrometer suggested that trypsin activity is slower in homoarginine residues than in arginine residues within the protein. To confirm this observation with ESI-FTICR-MS, we acquired mass spectra of the tryptic digest of modified protein after 30 min at 1:50 protease to protein ratio. We also collected the mass spectra of unmodified tryptic digest of ubiquitin at the same ratios and the same time. The spectra revealed that in 30 min, the unmodified ubiquitin showed several cleavages at lysine and arginine residues whereas the modified protein showed cleavages mostly at arginine residues (Figure S3). The MALDI and ESI time-course proteolysis studies both suggested that trypsin hydrolysis at homoarginine residues is slower than that at lysine and arginine residues. With a 1:25 protease to protein ratio, we observed a very low intensity of the peptide residues, 12–27, TITLEVEPSDTIENVK*, but observed a high intensity signal in MS spectrum for peptide TITLEVEPSDTIENVK*AK*. This observation suggests that trypsin activity in two nearby homoarginine sites is also slower.

To further confirm the trypsin cleavage of modified homoarginine sites by trypsin, we performed the same study on a different protein β -casein (~24 kDa) using 1:100 protease to protein ratio. Both ESI-FTICR-MS and MALDI-TOF-MS spectra were obtained. The cleavage at homoarginine residues was evident from the MALDI-TOF-MS spectrum, at m/z 822.53 (residues 185–191, VLPVPQK*), 957.48 (residues 113–120, VK*EAMAPK*), 998.52 (residues 41–47, INK*K*IEK*) and 1097.53 (residues 121–128, HK*EMPFPK*) (Supplementary Figure S4).

These results are in agreement with study by Chauvet *et al.* on oxidized and modified bovine trypsin inhibitor as presented in early literature. We suggest, as also suggested by Chauvet *et al.* in early literature that the tryptic cleavage of the homoarginyl bonds depends on

enzyme concentration, time, proximity of arginine residues, and the position of two nearby homoarginine sites. The selectivity of trypsin for particular homoarginine residues could be attributed to neighboring amino acids, which was not studied in this research.

The goal of this study to include the trypsin-cleaved homoarginine terminated peptides in our N-terminal peptide search; a more detailed comprehensive study is necessary to understand the influence of nearby arginine and other amino acid residues. We observed very good protein coverage of homoarginine modified proteins with trypsin in 1:100 protease/protein ratios (Figure 2, B). For that reason, we did not want to use the highest concentration of protease (Figure 2, C), because we thought it would generate more new tryptic peptides to be captured by the beads for N-terminal peptide purification study. To make less complicated tryptic peptide mixtures, we utilized optimum cleavage condition 1:100 protease to protein ratio for subsequent N-terminal peptides purification study but considered the tryptic cleavage of homoarginine modified peptides in our search. In addition, this helped to generate more charged peptides due to the missed cleavages in the homoarginine sites, where as other N-terminal selective methods neutralized the charges on all lysine residues due to acetylation.

The selectivity of the solid phase reagent was first studied. β -casein was digested with trypsin and the C-terminal lysines of the resultant peptides were then converted to homoarginines utilizing commonly used reaction protocol. After that, all the new termini in the peptides were blocked by an acetylation reaction. A standard peptide, Bradykinin was spiked in the solution keeping the 1:1 ratio of peptide mixture. The ESI-FTICR-MS spectrum of the solution was obtained (Figure 3, A). In the whole mixture, the only primary amine available was that of the N-terminal of spiked Bradykinin peptide; N-termini of all the tryptic peptides from the β -casein were blocked by acetylation. The reaction mixture was then treated with 10 mg of solid-phase reagent for 1h. The beads were then washed several times with acetonitrile, isopropanol and water. The washed beads were then treated with TFA:CH₂Cl₂ solution (80:20) to release the peptide from the beads. MALDI-TOF and ESI-FTICR-mass spectra were obtained. ESI-FTICR mass spectra showed that Bradykinin was selectively isolated from this complex mixture with added linker mass from the reagent (Figure 3, B). MS/MS of this peptide confirmed the addition of a linker mass in the N-terminus of the standard peptide Bradykinin (Figure 3, C).

The reagent was then further tested with a 1:1 mixture of tryptic peptides of acetylated and un-acetylated ubiquitin. After performic acid oxidation, ubiquitin was digested with trypsin. The resultant peptides were separated into two equal volumes. Both fractions were treated with isomethylurea to convert C-terminal lysines to homoarginines. Following this conversion, all the N-termini of one fraction were blocked with acetylation reaction. Both fractions were combined and treated with the solid-phase reagent. A MALDI-TOF mass spectrum was obtained before solid-phase enrichment. After solid phase capture, supernatant was collected. The MALDI-TOF mass spectrum showed that most of the unacetylated peptides were captured by the solid phase reagent, leaving the blocked N-termini in the supernatant (Supplementary Figure S5).

The selectivity studies performed showed the effectiveness of this solid phase reagent to enrich N-terminal peptides. To demonstrate the utility of our reverse solid-phase purification method to selectively enrich N-terminal peptides of the proteins (Scheme 1), we processed two proteins Ubiquitin and RNase S. The protein mixtures were treated with performic acid to convert sulfhydryl to cysteic acid. Lysines in the proteins were then converted to homoarginines. The extent of guanidination reaction to N - terminal primary amines has been reported by Beardsley *et al*⁶. N - terminal reactivity has been reported if a glycine is situated at the N - termini of proteins. However, even the protein N-terminal amine does

react by guanidination, method will still result in enrichment of the N-terminal tryptic peptides. Guanidination, like acetylation, effectively blocks the protein N-terminus from subsequent solid-phase capture. The original N-termini were then blocked with an acetylation reaction. The N-termini acetylated and homoarginine modified proteins were digested with trypsin with 1:100 protease to protein ratio overnight. An N-terminus acetylated Bradykinin was spiked in the digest mixture. The resultant peptide mixture was then treated with the solid-phase reagent beads for 2h. Every 30 min fresh beads (5 mg) were added to the reaction solution for complete capture of the N-terminal un-blocked tryptic peptides. After 2h, supernatant was collected and a MALDI-TOF mass spectrum was obtained. The MALDI-TOF mass spectrum showed the enrichment of N-terminal peptides of ubiquitin and RNaseS protein and also spiked N-terminal blocked Bradykinin from the digest mixture (Figure 4). The N-terminus of ubiquitin was enriched by trypsin cleavage at the homoarginine modified lysine. The methionine residue was converted to sulfone due to performic acid treatment and N-terminal primary amine was found acetylated [Ac-M*(O2)QIFVK*]. Spiked, N-term acetylated Bradykinin as well as an acetylated degradation product of Bradykinin was observed in the spectrum (Figure 4). A MALDI-TOF mass spectrum of our control sample of Ac-bradykinin showed the presence of both species (Supplementary Figure S6) and ESI-FTICR-MS analysis also confirmed their presence (Supplementary Figure S7). The identity of the ions at m/z 1276.65 shown in Figure 4 was determined by ESI-FTICR-MS/MS analysis. This peptide was confirmed to be the N-terminal tryptic peptide of RNaseS protein and also called S- peptide (Figure 5). The sequence of S-peptide in the RNaseS protein is KETAAAKFERQHMSSTSA. This peptide was cleaved by trypsin at the R10 residue and both lysine residues were converted to homoarginines (Ac-K*ETAAAK*FER), with the N-terminus blocked by an acetyl group as expected. This result illustrates that this reverse solid-phase purification strategy effectively isolated N-terminal peptides from a moderately complex protein digest mixture.

Conclusions

In this study we show clear indication of cleavage of homoarginine modified sites by trypsin and we also demonstrated that cleavage of modified sites depends on enzyme concentration, time, proximity of arginine residues, and the position of two nearby homoarginine sites. We also demonstrated a simple solid-phase reverse purification strategy utilizing a novel N-terminal selective reagent. The use of solid-phase reagent facilitated the avoidance of additional N-biotinylation and avidin capture steps of new N-termini in the sample mixture. In addition, with this solid-phase purification strategy, profiling of post translationally modified (such as acetylated, methylated, pyroglutamic acid, etc.) N-termini are also feasible as shown in figure S5. At present there is a shortage of efficient solid-phase cleavable capture reagents. Most of the solid-phase reagents currently available are DTT cleavable and have significant disadvantages in proteomics applications. The option of acid cleavage has tremendous advantages for studying the signal sequences or protease cleavage sites in gel-based study. It is extremely difficult to identify the N-terminal sites with current strategies as N-terminal peptide identification solely depends on MS/MS of one peptide. The option of acid cleavage will reconfirm or indicates the cleavage sites by generating peptides after that region. In addition, we showed enrichment by reverse solid-phase purification in peptide level. This reagent can also be utilized to enrich N-terminal peptides directly (without acetylation) from the protein level using the acid-cleavage option.

The limitation of solid-phase capture steps is that they require more samples but the advantage is less nonspecific contaminants in the samples. Another limitation of solid-phase capture reagents is polymer contamination from the solid-phase supports, which are more prevalent in polystyrene or peg-based solid supports. We utilized primary amine functionalized control pore glass (CPG) for synthesis of the reagent for less polymer

contamination but for more efficient sample cleaning, magnetic bead can also be used. Solid-phase approaches are extremely useful to be adopted for high-throughput applications but there are some potential limitations in sample requirements side. We showed here applicability of a solid-phase dual functional reagent (primary amine reactive and acid-cleavable) which we think could potentially be adopted in high-throughput applications. At this moment we can analyze sample of concentration in the range of 50 to 100 pmol by spiking peptide in a protein mixtures. Choice of solid-phase and also their pore sizes are very important parameter for enrichment efficiencies of peptides. A future study with the emphasis on high-throughput application with more complex samples are in progress. We believe these simple solid-phase capture strategy and trypsin proteolysis study on homoarginine modified proteins could be very useful for enrichment and sequence determination of N-termini of proteins by mass spectrometry.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the Office of Science (BER), U.S. Department of Energy, and Grant No. DE-FG02-04ER63924; NIH-NCRR, Grants No. 1 S10 RR017805-01, 1R01RR023334; and The Murdock Charitable trust. Author also acknowledges the start-up fund from the department of chemistry and biochemistry and mass spectrometry research core "Shimadzu Center for Advance Analytical Chemistry (SCAAC), which is operated by Shimadzu Institute for Research Technologies at University of Texas at Arlington.

References

1. Bradshaw RA, Brickey WW, Walker KW. N-terminal processing: the methionine aminopeptidase and N alpha-acetyl transferase families. Trends in biochemical sciences. 1998; 23:263–7. [PubMed: 9697417]
2. Paetzel M, Karla A, Strynadka NC, Dalbey RE. Signal peptidases. Chemical reviews. 2002; 102:4549–80. [PubMed: 12475201]
3. Matsumi R, Atomi H, Imanaka T. Identification of the amino acid residues essential for proteolytic activity in an archaeal signal peptide peptidase. The Journal of biological chemistry. 2006; 281:10533–9. [PubMed: 16484219]
4. Gevaert K, Goethals M, Martens L, Van Damme J, Staes A, Thomas GR, Vandekerckhove J. Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. Nature biotechnology. 2003; 21:566–9.
5. McDonald L, Robertson DH, Hurst JL, Beynon RJ. Positional proteomics: selective recovery and analysis of N-terminal proteolytic peptides. Nature methods. 2005; 2:955–7. [PubMed: 16299481]
6. Beardsley RL, Reilly JP. Optimization of guanidination procedures for MALDI mass mapping. Analytical chemistry. 2002; 74:1884–90. [PubMed: 11985322]
7. Brancia FL, Oliver SG, Gaskell SJ. Improved matrix-assisted laser desorption/ionization mass spectrometric analysis of tryptic hydrolysates of proteins following guanidination of lysine-containing peptides. Rapid Commun Mass Spectrom. 2000; 14:2070–3. [PubMed: 11085420]
8. Pitteri SJ, Reid GE, McLuckey SA. Affecting proton mobility in activated peptide and whole protein ions via lysine guanidination. Journal of proteome research. 2004; 3:46–54. [PubMed: 14998162]
9. Warwood S, Mohammed S, Cristea IM, Evans C, Whetton AD, Gaskell SJ. Guanidination chemistry for qualitative and quantitative proteomics. Rapid Commun Mass Spectrom. 2006; 20:3245–56. [PubMed: 17019669]
10. Yamaguchi M, Nakazawa T, Kuyama H, Obama T, Ando E, Okamura TA, Ueyama N, Norioka S. High-throughput method for N-terminal sequencing of proteins by MALDI mass spectrometry. Analytical chemistry. 2005; 77:645–51. [PubMed: 15649066]

11. Kleifeld O, Doucet A, auf dem Keller U, Prudova A, Schilling O, Kainthan RK, Starr AE, Foster LJ, Kizhakkedathu JN, Overall CM. Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nature biotechnology*. 2010; 28:281–8.
12. Chowdhury SM, Shi L, Yoon H, Ansong C, Rommereim LM, Norbeck AD, Auberry KJ, Moore RJ, Adkins JN, Heffron F, Smith RD. A method for investigating protein-protein interactions related to salmonella typhimurium pathogenesis. *Journal of proteome research*. 2009; 8:1504–14. [PubMed: 19206470]
13. Zhang X, Ye J, Engholm-Keller K, Hojrup P. A proteome-scale study on in vivo protein Nalpha-acetylation using an optimized method. *Proteomics*. 2011; 11:81–93. [PubMed: 21182196]
14. Kim JS, Dai Z, Aryal UK, Moore RJ, Camp DG 2nd, Baker SE, Smith RD, Qian WJ. Resin-assisted enrichment of N-terminal peptides for characterizing proteolytic processing. *Analytical chemistry*. 2013; 85:6826–32. [PubMed: 23772796]
15. Galan JA, Guo M, Sanchez EE, Cantu E, Rodriguez-Acosta A, Perez JC, Tao WA. Quantitative analysis of snake venoms using soluble polymer-based isotope labeling. *Molecular & cellular proteomics : MCP*. 2008; 7:785–99. [PubMed: 18089550]
16. Qian WJ, Goshe MB, Camp DG 2nd, Yu LR, Tang K, Smith RD. Phosphoprotein isotope-coded solid-phase tag approach for enrichment and quantitative analysis of phosphopeptides from complex mixtures. *Analytical chemistry*. 2003; 75:5441–50. [PubMed: 14714534]
17. Zhou H, Ranish JA, Watts JD, Aebersold R. Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. *Nature biotechnology*. 2002; 20:512–5.
18. Anderson, GABJE., editor. ICR-2LS. Pacific North West National Laboratory; 1995.
19. Chowdhury SM, Munske GR, Siems WF, Bruce JE. A new maleimide-bound acid-cleavable solid-support reagent for profiling phosphorylation. *Rapid Commun Mass Spectrom*. 2005; 19:899–909. [PubMed: 15739244]
20. Chauvet J, Acher R. Tryptic cleavage of homoarginyl bonds in the oxidized guanidinated derivative of a bovine trypsin inhibitor (Kunitz inhibitor). *FEBS letters*. 1971; 18:265–267. [PubMed: 11946136]
21. Seidl DS, Liener IE. Guanidination of the Bowman-Birk soybean inhibitor: evidence for the tryptic hydrolysis of peptide bonds involving homoarginine. *Biochemical and biophysical research communications*. 1971; 42:1101–7. [PubMed: 5550799]
22. Enoksson M, Li J, Ivancic MM, Timmer JC, Wildfang E, Eroshkin A, Salvesen GS, Tao WA. Identification of proteolytic cleavage sites by quantitative proteomics. *Journal of proteome research*. 2007; 6:2850–8. [PubMed: 17547438]
23. Lemmel C, Weik S, Eberle U, Dengjel J, Kratt T, Becker HD, Rammensee HG, Stevanovic S. Differential quantitative analysis of MHC ligands by mass spectrometry using stable isotope labeling. *Nature biotechnology*. 2004; 22:450–4.

Figure 1.
Primary amine reactive solid-phase reagent. CPG – control pore glass bead

Figure 2.

ESI-FTICR-MS spectra of modified ubiquitin after tryptic digestion with different protease to protein ratios (1:250, 1:100, 1:25). A) ESI-FTICR-MS spectrum of modified ubiquitin under 1:250 protease/protein ratio digestion. B) ESI-FTICR-MS spectrum of modified ubiquitin under 1:100 protease/protein ratio digestion. In the inset, one homoarginine peptide is shown (m/z 561.34, TLTGK*). C) ESI-FTICR-MS spectrum of modified ubiquitin from 1:25 protease/protein ratio digestion. In the inset the m/z range 561–564 is shown. The additional homoarginine terminated peptide (m/z 562.29, +2, TLSDYNIQK*) also appeared in that range. This peptide was not observed at 1:100 protease/protein ratios (see Figure 1, B inset). HR- homoarginine terminated peptide and in the inset K* - homoarginine modified lysine. See figure S8 for MALDI-IT-TOF MS spectrum.

Figure 3.

Selectivity of the solid-phase reagent. A) ESI-FTICR-MS spectrum of the β -casein tryptic peptides, guanidinated and N-termini were acetylated. Bradykinin, a standard peptide, was spiked in the digest mixture. B) ESI-FTICR- mass spectrum of the peptide, enriched from the digest mixture after solid-phase capture and cleavage (1056.5614 (bradykinin mw) – H (from N-terminal) + 100.0399 (linker mass after acid cleavage) + 2H⁺)/2 = 580.3041) C) ESI-FTICR-MS/MS of the peptide confirmed the labeling on the N-terminus. * denotes noise peaks.

Figure 4. MALDI-TOF mass spectrum of the N-terminal peptides enriched from RNase S and ubiquitin tryptic digest by this solid-phase capture strategy. Spiked N-terminal acetylated Bradykinin also showed a fragment from the loss of C-terminal arginine residue. K*-homoarginine (HR) modified lysine. See figure S7 for ESI-FTICR-MS spectrum.

Figure 5. ESI-FTICR-MS/MS of the N-terminal peptide of m/z 638.86. The assigned fragments confirm the peptide sequence to be that of the N-terminal tryptic peptide from the S peptide of the RNaseS protein. K*-homoarginine (HR) modified lysine.

Scheme 1.

A solid-phase reverse purification strategy used to selectively isolate N-terminal peptides.
 HR - homoarginine, R - arginine, Ac - acetyl group.