Summer 8-6-2013

Approaches for Improved Positional Proteomics

Yanjie Jiang
University of New Orleans, yjiang1@uno.edu

07-11-2013

Approaches for Improved Positional Proteomics[1]

Yanjie Jiang

Department of Chemistry-UNO, yjiang1@uno.edu

Approaches for Improved Positional Proteomics

A Dissertation

Submitted to the Graduate Faculty of the
University of New Orleans
in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Chemistry

By

Yanjie Jiang

Bachelor of Science. Nankai University, P. R. China, 1996
Master of Science. Dalian Institute of Chemical Physics, CAS, P. R. China, 1999
Master of Science. University of New Orleans, LA, USA, 2004

August, 2013

Acknowledgement

I would like to express my deepest appreciation to all those who provided me the possibility to complete this thesis. I would like to thank my advisor Dr. Richard Cole for his guidance and encouragement during this work. I owe my biggest thank you to Dr. Jonathan Lansing for his constructive advices in the process of this study and his tremendous help in polishing the writing. I would also like to thank the committee member Dr. Mark Trudell and Dr. Yang Cai for taking time reviewing and evaluating this study.

I am grateful to Dr. James Madsen for and Dr.Victor Farutin for their collaboration. I also want to thank Dr. Rick Sachleben and Dr. John Schaeck for their valuable discussion on chemistry reaction. I would like to thank my colleagues who supported me one way or the other.

I thank the support I received from Momenta Pharmaceuticals while conducting this research. I thank the graduate committee at department of chemistry in University of New Orleans to accept me in the graduate program.

This achievement is not possible without the support of my family. This is tribute to my husband, my two daughters and my parents.

Table of Contents

Abstract XI

Chapter 1. Introduction 1

1.1 Introduction of Proteomics 1

1.2 Isobaric Labeling in Proteomics 2

1.3 Positional Proteomics 5

1.3.1 Positional Proteomics approach for N-terminal analysis 8

1.3.2 Positional Proteomics approach for C-terminal analysis 17

1.3.3 Positional Proteomics approach for both N and C-terminal analysis 21

1.3.4 Thesis overview 23

1.4 References 24

Chapter 2 Experiments toward a streamlined workflow for positional proteomics 33

2.1 Overview of method 33

2.2 Consideration for development 33

2.3 Optimization 34

2.3.1 Guanidination reaction 34

2.3.2 SPITC labeling 36

2.3.3 Isolation/Purification 37

2.3.4 Trypsin digestion 39

2.3.5 Enrichment 39

2.3.6 Mass spectrometer selection 42
2.3.7 Database searching 43
2.4 Summary 44
2.5 References 44
Chapter 3 High Fidelity Approach for Proteomic Scale Enrichment and Identification of N-termini 47
3.1 ABSTRACT 47
3.2 INTRODUCTION 48
3.3 MATERIALS AND METHODS 51
3.3.1 Materials 51
3.3.2 Protein denaturation, reduction and alkylation 51
3.3.3 Guanidination 51
3.3.4 SPITC modification of N-termini 52
3.3.5 Trypsin digestion 52
3.3.6 N-terminal enrichment by ERLIC 53
3.3.7 LC-MS/MS analysis 54
3.3.8 Peptide identification 54
3.4 RESULTS AND DISCUSSION 55
3.4.1 Overall workflow description 55
3.4.2 Chemical modification of free amine groups in proteins 59
3.4.3 Trypsin digestion specificity after guanidination 62

3.4.4 Effectiveness of N-terminal enrichment..... 62
3.4.5 Tandem mass spectra comparison between CID and HCD 65
3.4.6 N-termini analysis of E. coli cell lysate..... 67
3.4.7 Comparison of bound and flow-through fractions of E. coli cell lysate.. 75
3.5 CONCLUSIONS..... 81
3.6 REFERENCES..... 81
Chapter 4 iTRAQ Labeling of N-terminal Amines in Complex Samples and Its Application in Protease Substrate Degradomics 85
4.1 Abstract 85
4.2 Introduction..... 86
4.3 Experiment 89
4.3.1 Materials..... 89
4.3.2 Reduction and alkylation..... 89
4.3.3 Guanidination 89
4.3.4 AspN-digestion..... 90
4.3.5 iTRAQ labeling..... 90
4.3.6 Trypsin digestion..... 91
4.3.7 N-terminal enrichment by NHS-activated agarose resin 91
4.3.8 LC-MS/MS..... 92
4.4 Results and Discussion 94
4.4.1 Strategy for the identification of N-terminal proteolytic peptides 94

4.4.2 E. coli cell lysate results using the proposed workflow..... 96
4.5 Summary..... 99
4.6 References..... 100
Chapter 5 Summary..... 104
References..... 106
Appendix..... 108
VITA..... 163

List of Figures

Figure 1.1 Scheme for chemical labeling of ICAT and iTRAQ**Error! Bookmark not defined.**5

Figure 1.2 Illustration of MS and MS/MS of iTRAQ labeled peptide ..... **Error! Bookmark not defined.**5

Figure 1.3 Typical positional proteomics workflow **Error! Bookmark not defined.**7

Figure 1.4 Scheme for N-TAIL workflow ..... **Error! Bookmark not defined.**13

Figure 1.5 Scheme for COFRADIC for N-terminal peptides **Error! Bookmark not defined.**16

Figure 1.6 scheme of proteome wide C-termini analysis by Overall lab ..... **Error! Bookmark not defined.**20

Figure 2.1 Comparison between before (A) and after (B) enrichment by NHS activated agarose resin spin column for BSA sample. N-terminal peptides (SPITC-DTHK(guandinyl)SEIAHR m/z= 484.20^{3+}). ..... **Error! Bookmark not defined.**39

Figure 2.2 Workflow for database searching ..... **Error! Bookmark not defined.**43

Figure 3.1 Flow chart for N-terminal identification by SPITC modification and ERLIC enrichment. ..... **Error! Bookmark not defined.**54

Figure 3.2 Schematic illustration for the guanidination reaction**Error! Bookmark not defined.**58

Figure 3.3 Schematic illustration of SPITC modification**Error! Bookmark not defined.**58

Figure 3.4 Fraction collections for trypsin digested *E.coli* cell lysate using electrostatic repulsion hydrophilic interaction (ERLIC) chromatography; (A) UV detection at 215 nm and (B) UV detection at 280 nm.**Error! Bookmark not defined.**62

Figure 3.5 MS/MS spectra for BSA peptide SPITC-DTHK(guandinyl)SEIAHR m/z= 484.20^{3+} by (A) CID and (B) HCD. SPITC is attached to N-terminal ions labeled in red. ..... **Error! Bookmark not defined.**64

Figure 3.6 MS/MS spectra for N-terminal peptides without guanidinyl lysine; (A) sequence MNIIEANVATPDAR, M1-SPITC protein ID P61714 (6,7-dimethyl-8-ribityllumazine synthase OS=*Escherichia coli*) and (B) sequence AVTNVAELNALVER, A2-SPITC protein ID P0A9Q7 (Aldehyde-alcohol dehydrogenase OS=*Escherichia coli*). SPITC is attached to N-terminal ions labeled in red..... **Error! Bookmark not defined.**66

Figure 3.7 MS/MS spectra of N-terminal peptide containing guanidinyl lysine; (A) AGK(guanidinyl)AGVEVDDR protein ID P0A9P0 (dihydrolipoyl dehydrogenase OS=*Escherichia coli*) (B) VVINK(guanidinyl) DTTTIIDGVGEEAAIQGR protein ID P0A6F5 (60 kDa chaperonin OS=*Escherichia coli*) and (C) LDMLNEELSDK(guanidinyl)ER protein ID P63284(Chaperone protein ClpB OS=*Escherichia coli*). SPITC is attached to N-terminal ions labeled in red. ... **Error! Bookmark not defined.**67

Figure 3.8 MS/MS spectra for SPITC peptide ELVTAAKLGGGDPDANPR (protein ID P0A8A0; (A) containing unmodified lysine and (B) containing guanidinyl Lysine. ..... **Error! Bookmark not defined.**69

Figure 3.9 MS/MS spectra for N-terminal peptides of protein P0A850 (Trigger factor OS=Escherichia coli); (A) P0A850from1to13, (B) P0A850from244to255, and (C) P0A850from322to334. SPITC is attached to N-terminal ions labeled in red..... **Error! Bookmark not defined.**71

Figure 3.10 Overlaps between protein accessions identified in flow-through and bound fractions; (A) all proteins identified in each of the fractions and (B) all proteins identified in flow-through fractions versus those detected in bound

fraction as labeled with SPITC on N-terminii. Areas are approximately proportional to the counts shown in the figures..... **Error! Bookmark not defined.**74

Figure 3.11 Sums of PSMs for all proteins (A) in flow-through fraction and (B) those in bound fraction that were identified as N-terminal SPITC labeled. In each case, proteins were split into two groups those that were also present in the other set (N-terminal SPITC in bound fraction or all proteins in flow-through fraction respectively) or only present in the given set. Proteins present in both sets tend to have higher PSM sums, suggesting their higher abundance.**Error! Bookmark not defined.**75

Figure 3.12 fractions of PSMs assigned to peptides starting from different ranges of positions in original protein sequences. ..... **Error! Bookmark not defined.**77

Figure 4.1 Scheme for N-terminal Proteolytic Peptides by iTRAQ labeling.... **Error! Bookmark not defined.**92

Figure 4.2 iTRAQ labeled N-terminal peptides (a) iTRAQ-MFPEYR from protein P0ACW6 (Uncharacterized protein YdcH OS=*Escherichia coli*), position 1-6 (b) iTRAQ-DVQVFTR from protein P0A8V2(DNA-directed RNA polymerase subunit beta OS=*Escherichia coli*), position 930-936.**Error! Bookmark not defined.**95

List of Tables

Table 2.1 Guanidination optimization..... **Error! Bookmark not defined.**35

Table 2.2 Evaluation of SPITC reaction conditions **Error! Bookmark not defined.**36

Table 3.1 Counts and percentages of peptides and corresponding PSMs annotated as labeled with SPITC (on N-termini and lysines) and not in bound and flow-through fractions of *E. coli* cell lysate..... **Error! Bookmark not defined.**73

Abstract

Positional proteomics is emerging as an attractive technique to characterize protein termini, which play important biological roles in cells. Even with the advances in past decades, there still are areas for improvement. This thesis focuses on improving data quality and assignment confidence in positional proteomics.

A novel workflow was designed for the large-scale identification of protein N-terminal sequences. 4-sulfophenyl isothiocyanate (SPITC) is used for N-termini sulfonation; Upon higher energy collisional dissociation (HCD), SPITC peptides in electrospray ionization ESI generate predominately y-type ion series; such simplification of spectra enables the identification of N-termini with high fidelity. The presence of b_1 + SPITC product ions upon HCD furthers the confidence for N-terminal identifications. Secondly, sulfonated N-terminal peptides possess one negative charge site at low pH, which was exploited to enrich the SPITC modified N-terminal peptides by electrostatic repulsion hydrophilic interaction (ERLIC) chromatography. Such enrichment process allows both N-termini enriched and N-termini deficient fractions to be collected and analyzed by LC-MS/MS. This method was applied to an *E. coli* cell lysate, identifying approximately 350 N-terminal peptides (85% represented neo-N-termini from protein degradation and 15% from leading methionine excision). These N-terminal peptides represented

274 distinct E.coli proteins, 224 of which were also identified in the analysis of flow-through fractions from internal peptides.

Another approach we took to boost the identification confidence is by exploiting iTRAQ (isobaric tag for relative and absolute quantitation) in the positional proteomics workflow. This approach allows for multiplexed comparison between different samples, and thus is well-suited for degradadomics analyses where degraded samples are compared to control samples. Both control and protease treated sample are labeled by different tags which allows direct comparison of protein N-termini with neo-N-termini. In addition, samples are analyzed duplicate by labeling with two tags, aiming for quick validation of peptides by internal replicates. In this study, Asp-N digested *E.coli* cell lysate is taken as a model system. A total of 500 N-terminal peptides, corresponding to 370 proteins, were identified with high confidence in one experiment, with 87% of those proteolytic products matching the expected protease digestion specificity, validating the assignment accuracy of this approach.

Keywords: Positional proteomics; Sulfonation; SPITC; ITRAQ modification.

Chapter 1. Introduction

1.1 Introduction of Proteomics

Proteomics is the large-scale study of proteins particularly their structures and functions.[1,2] Proteomics is the next step in the study of biological systems after genomics and transcriptomics, since proteins are vital parts of living organisms as the main components of the physiological metabolic pathways of cells. Mass spectrometry (MS) has increasingly become a key technology for protein sequence analysis in the past decade, especially for complex mixtures, due to significant advances in instrumentation, sample preparation techniques and data interpretation algorithms.

Large scale, "bottom-up" (also called "shotgun") characterization of cellular proteomes is the most widely adapted method in proteomics study.[3,4,5,6,7,8,9,10] A typical workflow for high-throughput bottom-up characterization is composed of protein extraction from cells, enzymatic digestion, LC-MS/MS analysis and database searching which correlate MS/MS spectra with sequence and ultimately the parent proteins.

Proteome systems typically encompass thousands of individual components present in concentrations ranging over several orders of magnitude;[11] therefore, to reduce the complexity, multidimensional liquid chromatography (MDLC) is

generally used for fractionation prior to MS analysis.[12,13] The combination of online/offline strong cation exchange (SCX) with reversed-phase (RP) chromatography is the most widely used MDLC due to the good orthogonality of the methods.[14,15,16] Peptides are separated by charge in SCX and by hydrophobicity in RP. The use of high concentrations of salt in SCX may cause problems in downstream analysis; therefore, online desalting of the collected fractions from SCX with the trap column is the common practice,[17] even with the drawback of a high tendency for autosampler blockage and quick deterioration of the trap column. Extensive separation of components aids in the protein coverage; it is now possible to identify more than 10,000 proteins from human cells.[18,19]

1.2 Isobaric Labeling in Proteomics

A major aspect of proteome research is quantitative proteomics aiming at measuring relative changes in proteins expressed in cells or tissues of different states, e.g. healthy versus disease state.[20,21,22,23,24] Measurement of relative changes is simplified when two or more analytes can be labeled distinctly, combined, and analyzed as a mixture. Labeling schemes based on isotopic labeling, in which all labeled versions of a single peptide co-elute from LC, are especially powerful, as the relative abundance for this peptide corresponds directly to the relative signal intensity due to identical ionization efficiencies. Accordingly, a significant effort

has been made to develop stable isotopic labeling methods to facilitate downstream MS analysis for direct comparison.[25,26,27,28]

Isotopic labeling with stable isotopes is a well-known method for "tagging" specific proteins. Such metabolic labeling is applicable to cell culture, e.g., through growth in isotope-labeled media (e.g., ^{15}N media[29,30]). Another approach is stable isotope labeling by essential amino acids in cell culture (SILAC)[31,32] which relies upon addition of intact isotopically-labeled amino acids. The most commonly used stable isotope-encoded amino acids are $^{13}C_6$-lysine or $^{13}C_6$-arginine.[33,34,35,36] The biggest advantage of SILAC compared to other isotope labeling techniques is that the SILAC technique offers minimum technical variations in sample processing due to the fact that the isotope is introduced into the cell culture, the earliest possible sample mixing stage.

Most *in vitro* labeling techniques are based on the formation of a covalent bond between the labeling reagent and the specific functional groups in polypeptides. Isotope-coded affinity tag (ICAT) is among the first of such applications reported in 1999.[37] ICAT consists of three elements: an iodoacetamide group to modify cysteine residues, an isotopically coded linker, and a biotin tag for the affinity isolation of labeled proteins/peptides (refer to figure 1.1). The process starts with ICAT labeling, followed by trypsin digestion, then, the ICAT labeled peptides are enriched by affinity chromatography via a biotin tag, resulting in

reduction of sample complexity. The concept is innovative and widely accepted.[38] However, the method provides limited coverage over the proteome, as the cysteine content in proteins is fairly low giving poor coverage of the digested peptides, and many proteins have no cysteine.

Among all the isotopic labeling techniques, the iTRAQ (isobaric tags for relative and absolute quantification) method shows significant advantages.[39,40,41,42] The iTRAQ reagent reacts with proteolyzed peptides to form an NHS ester derivative with primary amino groups. Differentially labeled peptides appear as single peaks in MS scans due to the isobaric mass design of the iTRAQ reagent (refer to figure 1.1); such multiplication of peptide abundance results in improvement of sensitivity. When subjected to MS/MS, the isotope encoded reporter ions provide relative quantitative information on peptides and ultimately on proteins (refer to figure 1.2). In a complex mixture, iTRAQ samples subjected to independent data acquisition in LC-MS/MS have a tendency to allow identification of only high abundance proteins in a traditional proteomics workflow. Both 4-plex and 8-plex versions of iTRAQ are commercially available.

In this thesis, effort has been made to incorporate iTRAQ with positional proteomics to monitor protease substrates in *E. coli* cell lysates.

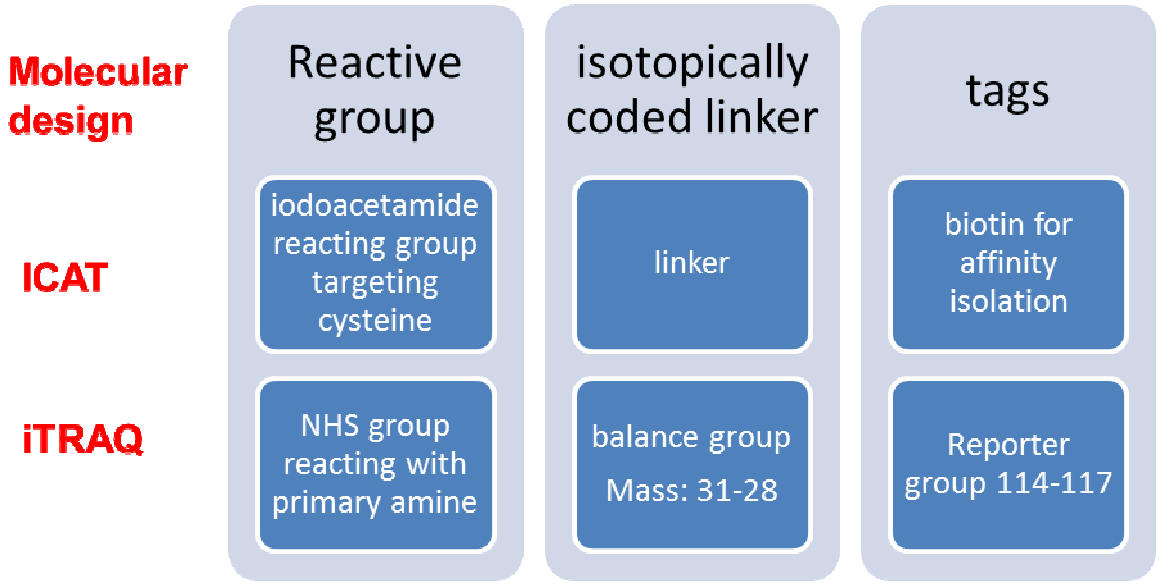


Figure 1.1 Scheme for chemical labeling of ICAT and iTRAQ

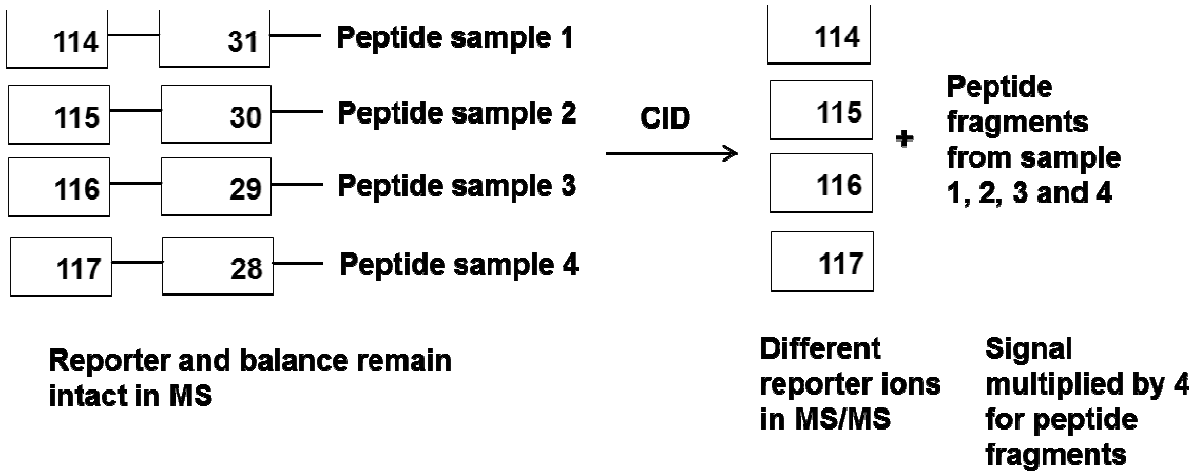


Figure 1.2 Illustration of MS and MS/MS of iTRAQ labeled peptide

1.3 Positional Proteomics

The concept of positional proteomics is that a protein can be identified by a single, position-defined peptide, with the two most obvious positional locations

within every protein being the N- and C-termini.[43] Sample complexity is dramatically reduced in N-terminal or C-terminal enriched samples of proteolytic digests. There are two driving forces in this field to motivate the advance of this technology. One is the proteome annotation,[44] where the termini play important roles in protein function, mutation and post translational modification.[45,46] The other is the development of degradomics,[47,48] which focuses on the elucidation of protease substrate and cleavage sites. The newly generated termini after protease treatment are called neo-N-termini or neo-C-termini, in order to differentiate them from mature (innate) protein termini.

To identify terminal peptides, the normal proteomics workflow is no longer suitable due to the fact that the terminal peptides are buried in the sea of internal tryptic peptides. The sample complexity makes it difficult to select and detect the terminal peptides during MS/MS acquisition. Therefore, terminal peptide enrichment prior to MS analysis is essential in such workflow development. Enrichment greatly simplifies the proteome by using single terminal peptides for protein identification, which increases dynamic range and proteome coverage for low abundant proteins.

It is essential to differentiate the termini peptides from the internal peptide in positional proteomics, since the same functional groups that define the protein termini in the sample, *i.e.* α-amine and carboxyl groups are presented by internal

proteolyzed peptides. In addition, primary amine and carboxylic groups are present in the side chains of lysine and acidic amino acid residues, respectively. The following paragraphs summarize strategies that have been applied in positional proteomics.

Positional proteomics workflow can be roughly divided into the following modules shown in figure 1.3: (1) labeling of termini, (2) proteolysis, (3) enrichment, and (4) LC-MS/MS analysis.

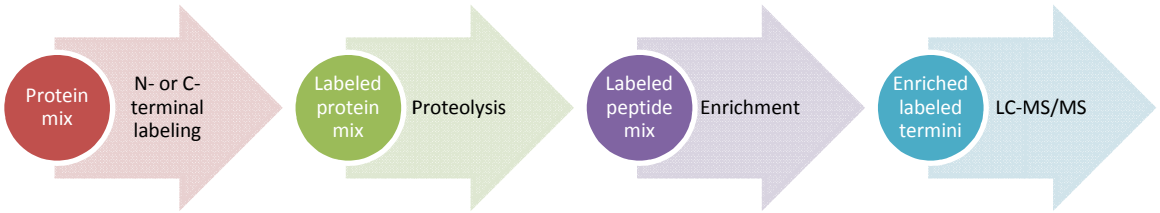


Figure 1.3 Typical positional proteomics workflow

The enrichment process can be categorized as either a positive selection approach or a negative selection approach. The former modifies the termini with a tag that enables targeted enrichment of terminal peptides from the digested mixture. The latter takes advantage of the newly generated functional groups (primary amine for N-termini and carboxylic acid for C-termini) after digestion, which can be conjugated to another matrix and depleted, resulting in the enrichment of the targeted termini.

1.3.1 Positional Proteomics approach for N-terminal analysis

1.3.1.1 Enrichment of protein N-termini by positive selection using biotinylation

A method using biotinylation to positive select N-terminal peptides was reported by Kuhn *et al* in 2003.[49] This protein tag technology was later on adapted and further developed by the Salvesen lab to identify protease substrates.[50] In this method, the side chains of lysine are first blocked by guanidination, and then free alpha-amines (protein N-termini) are reacted with a disulfide-linked biotin derivative (sulfosuccinimidyl 2-(biotinamido)-ethyl-1, 3-dithiopropionate (sulfo-NHS-SS-biotin)). After trypsin digestion, the peptide mixture is incubated with immobilized streptavidin beads, which retains the N-terminal peptide. The last step of sample preparation is to elute the N-terminal peptide by reducing the disulfide bond. This workflow was applied to study *in vivo* constitutive proteolysis in *E. coli*, yeast, mouse and human samples and determine the specificity of methionine aminopeptidases, signal peptidases and mitochondrial peptidases.

Another biotinylation protocol was reported by Wells *et al.* based on subtiligase in 2008[51]. An engineered variant of subtilisin called subtiligase was used, which shows absolute selectivity for ligation of alpha-amines, forming a biotinylated peptide ester containing a tobacco etch virus (TEV) protease cleavage site. Proteins

are then denatured, reduced and alkylated before trypsin digestion. Then, the biotinylated N-terminal peptides are captured by immobilized streptavidin, resulting in the separation from internal and C-terminal peptides. The N-terminal peptides are eventually released by TEV digestion, with tagging of Ser-Tyr-dipeptides at the N-termini as a signature for identification by LC-MS/MS. The subtiligase method was validated by analysis of the proteome of living and apoptotic Jurkat cells. It was reported that 333 unique cleavage sites in 282 proteins were identified after aspartic acid residues and were therefore linked to caspase activity in etoposide treated apoptotic Jurkat cells. A separate study using this technique combined with SILAC in cell culture to identify caspase-1 substrates *in vitro* and in cell-based inflammation models was also reported.[52]

A third biotinylation method incorporated Edman degradation in the workflow [53] . After protein is denatured, reduced and alkylated, phenyl isothiocyanate (PITC) is used to block all primary amines in the proteome. Then, trifluroacetic acid is used to break the peptide bond between the first and second amino acid of PITC modified proteins, while PITC modified ε-amines (*i.e.* lysine side chains) remain intact in this treatment. The next steps are similar to the aforementioned method. The newly generated free α-amine is biotinylated with sulfo-NHS-SS-biotin, followed by trypsin digestion. Streptavidin is used to trap biotinylated peptides, which is further released by disulfide bond reduction. The

enriched N-terminal peptides in this case are one amino acid shorter than the true N-termini. This approach was validated by a mixture of known proteins and applied to characterize the constitutive N-terminal processing events in Jurkat cells. Both known and new caspase substrates were identified in the cisplatin-induced apoptosis of Jurkat cells.

A fourth biotinylation-related method in degradomics is called PICS (proteomic identification of protease cleavage) developed by the Overall lab[54]. This approach differs from the other approaches in that it is peptide centric and not protein/substrate centric. First, proteins are digested by trypsin or endoproteinase Glu-C, and then all primary amines are blocked by methylation. The peptide mixture is treated as a library and incubated with a protease of interest. Neo-N-terminal peptides are biotinylated and affinity-selected, followed by LC-MS/MS analysis. This approach was applied to profile serine (thrombin, neutrophil elastase, cathepsin G), aspartic (HIV-1 protease) and cysteine proteases (cathepsin K, caspase-3, caspase-7).[55]

The advantage of positive selection using biotinylation is the high efficiency of enrichment due to the high specificity between biotin-streptavidin. A drawback of this type of technique may arise from amino acid bias during the biotinylation reaction. For example, the secondary α-amino group of proline is less reactive compared to the primary α-amino group of other amino acids.[56] In addition, in

positive selection, mature proteins with acetylation, dimethylation or cyclization at N-termini will not be detected.

1.3.1.2 Enrichment of protein N-termini by negative selection using biotinylation

McDonald *et al*[43] reported a protocol based on biotinylation for negative selection in 2005. Here, all the primary amines in a proteome sample are acetylated first, followed by trypsin digestion. The internal peptides and C-terminal peptides containing an α-amino group at a newly generated N-termini are biotinylated and removed by passing through immobilized streptavidin beads. The simplification effect was shown in chicken skeletal muscle and *E. coli* cell lysate.

1.3.1.3 Enrichment of protein N-termini by negative selection using amine reactive reagent

There are a few reports using amine reactive reagents to scavenge the newly generated internal peptides after digestion, thus enriching N-terminal peptides. The basic workflow is as follows: (1) The primary amines in proteins are acetylated or dimethylated, followed by trypsin digestion. (2) The internal peptides and C-terminal peptides are scavenged by amine reactive reagent and the flow through, which contains enriched N-terminal peptides, is collected.

McDonald *et al*[57] used NHS activated Sepharose beads to scavenge internal peptides in 2006 and reported the identification of about 300 *E. coli* protein N-termini.

Mikami and Takao used an isocyanate resin to capture internal peptides.[58] The performance of isocyanate resin was demonstrated by applying it to several peptide mixtures, including proteolytic digests.

Kleifeld *et al*[47] employed a self-synthesized dendritic polyglycerol aldehyde polymer to capture tryptic and C-terminal peptides, which can be conveniently removed by centrifugation. It is claimed that the binding of tryptic peptides of such polymer is up to 2.5 mg peptide/ mg polymer, a more than ten-fold improvement in capacity over amine reactive resins. This approach incorporated with isotopic labeling under various formats is a valuable venue for N-terminal positional proteomics research and it is further explained in the following section of N-TAILS.

1.3.1.4 Protein N-terminal identification by N-terminal amine based isotope labeling of substrates (N-TAILS)

A detailed and streamlined protocol of N-TAILS was described by the Overall lab in 2011,[59] based on previous development and applications by the same lab.[45,60,61] The highlight of TAILS is that it uses negative selection to enrich for all

N-terminal peptides and uses primary amine labeling-based quantification as the discriminating factor. Labeling is versatile; the authors elaborate dimethylation-TAILS, SILAC-TAILS and iTRAQ-TAILS in the current protocol. This method is suited to many applications, including biochemical and cell culture analysis *in vitro* as well as analysis of tissue samples from animal and human sources *in vivo*.

The TAILS workflow is composed of the following steps (refer to figure 1.4): protein collection and proteolysis by the test protease; isotopic labeling and primary amine blocking followed by tryptic digestion; negative selection by a high efficiency polymer (dendritic polyglycerol aldehyde polymers); identification of N-terminal peptides by LC-MS/MS; identification of protease substrates by the sequence of the cleavage sites, or loss of cleaved natural N-terminal peptides.

To improve coverage, it is recommended to employ two or more digesting proteases other than trypsin, e.g. Glu-C or chymotrypsin.

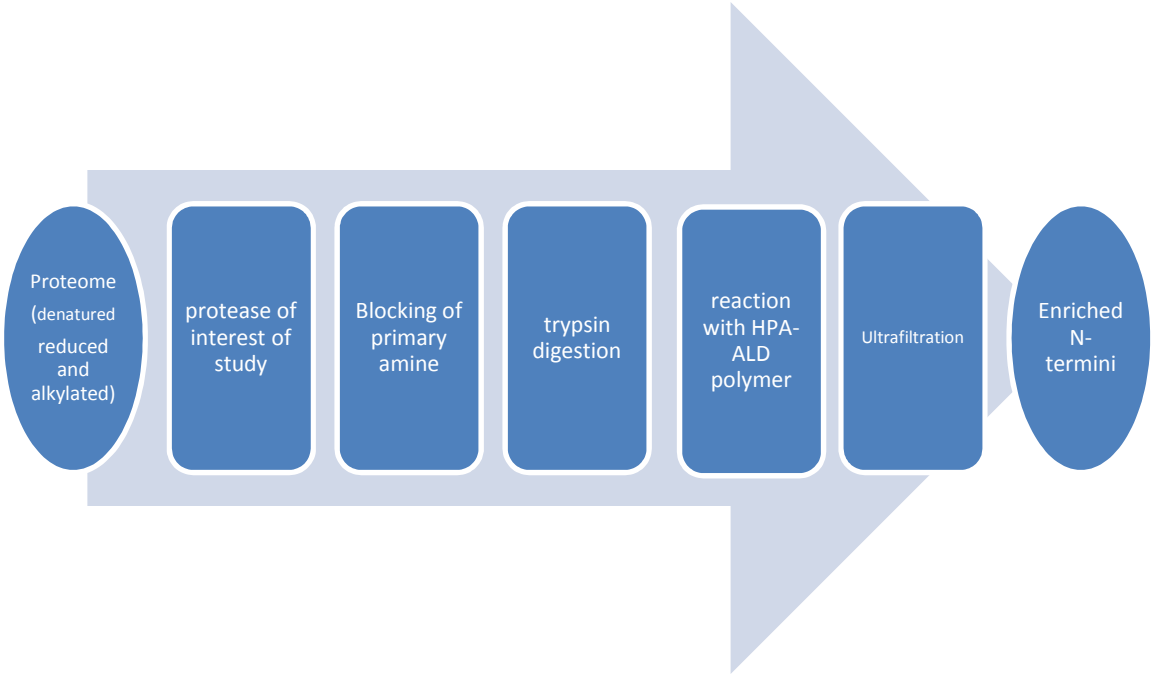


Figure 1.4 Scheme for N-TAIL workflow [59]

1.3.1.5 Enrichment of protein N-termini using phosphor tagging (PTAG) and TiO2-based depletion

Mommen *et al*[62] described a protocol to enrich protein N-terminal peptides using phospho tagging (PTAG) and titanium dioxide (TiO_2) affinity chromatography. Primary amino groups in proteins are initially dimethylated with formaldehyde, followed by digestion using trypsin, chymotrypsin and endoproteinase Glu-C. The newly formed internal peptides are modified with the PTAG reagent glyceraldhyde-3-phosphate in nearly perfect yields (>99%). The resulting phosphopeptides are removed by binding onto TiO_2 affinity column. This method allowed identification of 753 N-terminal peptides, corresponding to 428

proteins, in *N. meningitides* and 928 N-terminal peptides, corresponding to 572 proteins, in *S. cerevisiae*.

1.3.1.6 Enrichment of protein N-termini by combined fractional diagonal chromatography (COFRADIC)

In 2003, the N-terminal combined fractional diagonal chromatography (COFRADIC) technology introduced by the Gevaert lab[63] was the first positional proteomics technology by which N-terminal peptides were enriched by depleting other peptides (refer to figure 1.5). This method progressed in the following years[64,65,66,67,68]and the latest protocol was reported in 2011.[69] The protocol can be briefly described as the following: Before trypsin digestion, proteins undergo denaturation, reduction and alkylation. Then, all primary amines are blocked by trideutero-acetylation (to distinguish *in vitro* acetylation from *in vivo* acetylation) so that trypsin digestion will produce only Arg-ending peptides. The peptide mixture is incubated with glutamine cyclotransferases and pyroglutamyl amino peptidases, respectively, to convert N-terminal glutamines into pyroglutamates and remove these pyroglutamates from the peptide backbone. The above mixture is then loaded onto an SCX cartridge at low pH, where N-terminal blocked peptides are poorly retained and are collected in the flow through fraction since they carry one less positive charge compared to internal peptides at low pH. This peptide mixture is then separated in RP-HPLC and primary fractions are collected. Each of

the above collected fractions are treated with 2,4,6-trinitrobenzenesulfonic acid (TNBS), which reacts with the primary amine of any remaining internal peptides to introduce a very hydrophobic trinitrophenyl group with these peptides. Each treated fraction is put through the second RP-HPLC separation using identical chromatographic conditions as in the first separation, where the N-terminal peptides elute within the same time interval as during the primary run, however, the TNBS modified internal and C-terminal peptides carrying a large hydrophobic group elute much later in RP-HPLC, thus resulting in separation from the neo-N-terminal and protein N-terminal peptides.

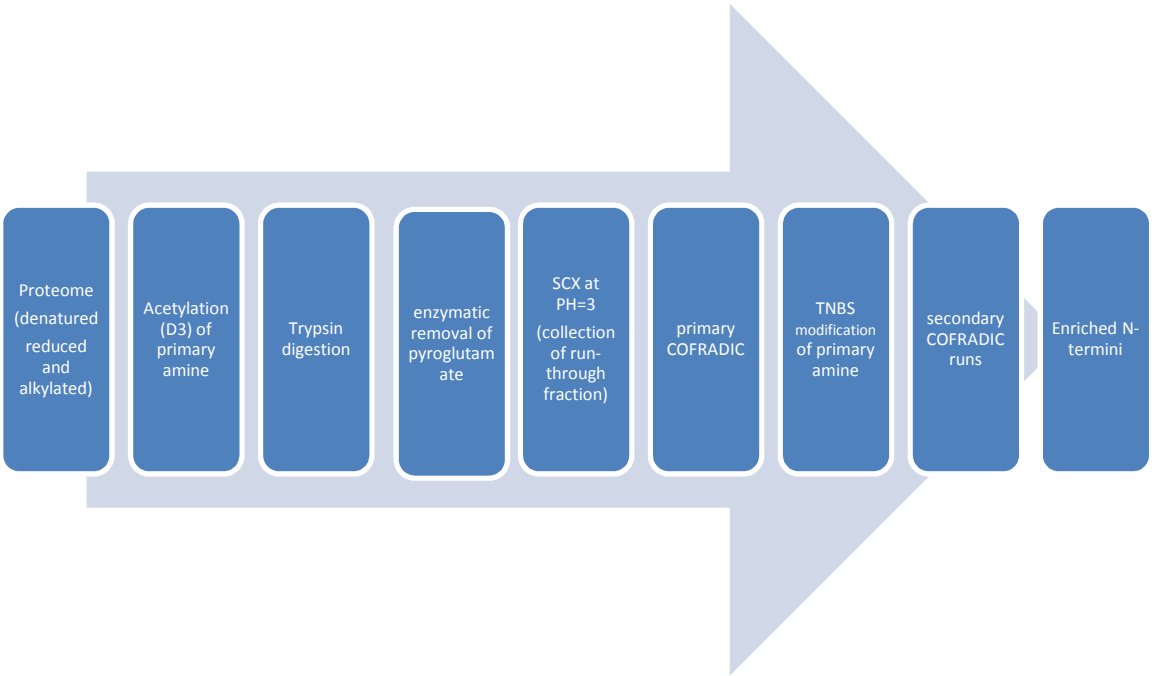


Figure 1.5 Scheme for COFRADIC for N-terminal peptides [63]

The COFRADIC technology has been applied in various biological systems and has been proved to be a reliable approach for protease-related studies. Stable isotopes for labeling purpose are introduced in the protocol in biological applications. Oxygen-18 was used to label peptides during trypsin digestion in a human Jurkat cell culture; a total of 93 *in vivo* protease-processed sites in 71 proteins associated with Fas-induced apoptosis were identified. Oxygen-18 labeling together with COFRADIC was also used to map the proteolytic process in anthracycline-induced acute myelogenous leukemia cell death.[66]

SILAC can also be incorporated into the COFRADIC workflow (refer to figure 1.5), Arginine is chosen for SILAC labeling, since tryptic digestion produces arginine-ending peptides. The control cells use regular arginine in cell culture while the sample cells use $^{13}C_6$ arginine in cell culture. Purified proteins in the heavy labeled cells are treated with protease, and then combined with the purified proteins in control cells. The following steps are the same as normal COFRADIC. More than 800 cleavage sites in 332 human and 282 mouse substrates for granzyme b were identified using SILAC combined with COFRADIC.[70]

1.3.2 Positional Proteomics approach for C-terminal analysis

Generally speaking, C-terminal analysis is not as widely available as the N-terminal analysis due to the formidable technical difficulties of selective activation of carboxylic acids. Current methods of isolating C-terminal peptides are predominantly affinity-based procedures. The application of mass spectrometry to C-terminal analysis in the literature related to our topics is as follows:

1.3.2.1 Use anhydrotrypsin-lysine affinity to isolate the C-terminal peptides

An elegant approach to isolate C-terminal peptides was reported in 2000.[71] Basically, a sample is digested with endoprotease Lys-C, and then anhydrotrypsin coupled onto agarose beads is applied to the digest sample. Anhydrotrypsin is a catalytically inert variant of trypsin capable of binding peptides with C-terminal lysine or arginine. Thus, the N-terminal and internal peptides are bound to anhydrotrypsin beads; after centrifuging to get rid of beads, the supernatant only contains the original C-terminal peptides for further analysis. However, this method is not suitable for proteins ending with lysine or arginine (~ 84% of proteins do not end with lysine or arginine). This approach lacks robustness; the amount of anhydrotrypsin beads was adjusted for each of the proteins investigated. The reason is that a small amount of anhydrotrysin beads is not sufficient to capture the internal peptides, while a large excess of beads will induce non-specific

binding of C-terminal peptides to the beads. For a complex system with a large dynamic range, the integrity of the results cannot be ensured.

1.3.2.2 Use of DITC resin to isolate C-terminal peptides

In an approach described by Kuyama *et al.*, a sample is initially digested with endoprotease Lys-C, then TMPP modification is selectively applied to cap the N-terminal amino group. The reagent for TMPP modification is succinimidyloxycarbonylmethyl tris (2,4,6-trimethoxyphenyl) phosphonium bromide (TMPP-Ac-Osu). p-Phenylenediisothiocyanate (DITC) resin is used to scavenge the lysine-containing peptides from a Lys-C digest. Isolated C-terminal peptides are then *de novo* sequenced using MALDI-MS/MS.[72] This method has been further optimized[73] to use diisothiocyanate coupled glass beads. This method is not suitable for proteins ending with lysine or arginine.

1.3.2.3 Use of polymer-based enrichment for C-terminal peptides

Overall *et al*[74] reported a polymer-based enrichment approach to profile protein C-terminal peptides. Briefly, the workflow is as follows (refer to figure 1.6): protein thiol groups are reduced and alkylated, carboxyl groups of C termini, aspartate and glutamate side chains are then protected by carbo-diimide-mediated (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride, EDC) and N-hydroxy succinimide-assisted condensation with 1 M excess of ethanolamine. The

derivatized proteins are then digested by trypsin. To prevent cross-reactivity, peptide concatamerization or cyclization, free amines are protected with methyl groups. The internal tryptic and N-terminal tryptic peptides are removed by coupling with polyallylamine (MW ~56000, sigma) mediated with EDC. Ultrafiltration is used to remove the polymer, retaining the enriched C-terminal peptides in solution. LC-MS/MS follows to analyze the sample.

This approach is the first that allows proteome-wide C-terminal analysis. In the same report, C-terminal amine-based isotope labeling of substrates (C-TAILS) was also described, in which heavy isotope formaldehyde is used during both reductive methylation steps. C-TAILS was tested using a Glu-C digested *E. coli* cell lysate model system. The result showed that more than 90% of the peptides identified had C-termini corresponding to the Glu-C cleavage site.

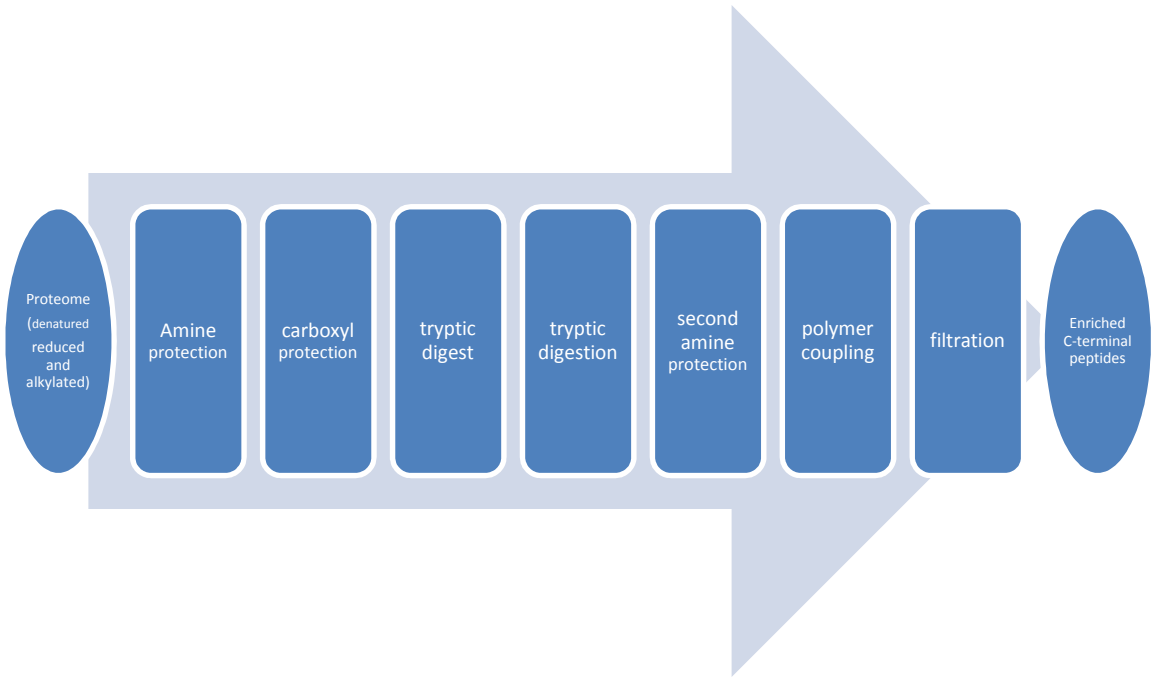


Figure 1.6 scheme of proteome wide C-termini analysis by Overall lab [74]

1.3.3 Positional Proteomics approach for both N and C-terminal analysis

1.3.3.1 SCX to selectively enrich in vivo acetylated protein N-terminal and C-terminal peptides

This method requires no chemical derivatization. The sample is trypsin digested and SCX is run at pH 3 to fractionate the sample.[44] The majority of C-terminal and acetylated N-terminal peptides elute very early in the chromatogram, while internal peptides elute much later. This separation is due to a lack of net positive charge in both C-terminal and acetylated N-terminal peptides at pH 3, while internal peptides have a net positive charge due to the unblocked amino terminus and the Lys or Arg side chain. However, this separation will not work in the following two circumstances:

- The C-terminal ends with lysine or arginine.
- Trypsin digestion has a missed-cleavage, resulting in peptides with multiple basic residues.

In such circumstances, the terminal peptides will possess at least two free amines, thus their charge properties change and they are no longer able to be separated from the internal peptides. The author also pointed out that the selective recovery of acetylated N-terminal and C-terminal peptides depends on the trypsin digestion integrity.

1.3.3.2 Combined fractional Diagonal Chromatography (COFRADIC) for simultaneous N and C terminal proteomics

In 2010, the Gevaert lab[48] published a COFRADIC workflow for both N and C- terminal peptide identification in a single sample preparation. This COFRADIC workflow follows the procedures described in N-terminal COFRADIC, except that after the primary fraction collection, peptides are reacted with an N-hydroxy-succinimide ester of butyric acid, butyrylating the C-terminal peptides. Such modification allows C-termini to elute 4-12 min after N-termini in secondary RP-HPLC. The pooled N-terminal and C-terminal peptides are ready for LC-MS/MS analysis. The authors used this COFRADIC based approach to study processing by the human endoprotease granzyme B in K-562 cell lysates. SILAC of $^{13}C_6$ arginine

cell lysate is used as substrate pool for granzyme B, with $^{12}C_6$ arginine cell lysate as control. Thus, upon trypsin digestion, neo-N-terminal peptides can be differentiated from the N-termini background of the cell lysate. C-termini are differentiated by using N-hydroxysuccinimide (NHS) esters of $^{12}C_4$ or $^{13}C_4$ butyric acid. In this study, a total of 1621 annotated N-termini and 760 annotated C-termini were identified, with 334 neo-C termini resulting from granzyme B processing and 16 neo-C termini resulting from carboxypeptidase A4 processing.

1.3.4 Thesis overview

With all the efforts and advances in positional proteomics, the inherent main drawback of such techniques still needs to addressed, which is how to validate the identified substrates due to the "one hit wonder" in such technique. The strength of positional proteomics arise from the dramatically reduced complexity of proteome samples, however, the strength turns into weakness if the question is asked how you can validate the results. In this thesis, we propose an approach which separates the enriched N-termini fraction for positional proteomics from the peptide mixtures, while preserving the rest of the peptide mixture. Therefore, MS analysis of the remaining peptide mixtures can serve as validation for the results of positional proteomics. Such direct experimental validation is meaningful and convincing considering both analyses use the same starting material. The above goal is achieved by utilizing 4-sulfophenyl isothiocyanate (SPITC) to modify the

N-termini and separate the SPITC modified peptides from others using electrostatic repulsion hydrophilic interaction chromatography (ERLIC). Moreover, the confidence of N-termini identification is further strengthened by exploiting the unique fragmentation behavior of SPITC peptides, thus we conclude that our approach offers high fidelity assignment of N-terminal peptides. This work is presented in chapter 3.

Another attempt to simultaneously validate the results of positional proteomics is by use of iTRAQ-4plex. In this study, both control and protease treated samples are labeled by different tags which allows direct comparison of protein N-termini with neo-N-termini. In addition, samples are analyzed in duplicate by labeling with two tags (i.e. tag 116 and tag 117), aiming for quick validation of peptides by internal replicates. A new workflow is designed which incorporates iTRAQ into positional proteomics to study the specificity of protease. This work is presented in chapter 4.

Experimental optimization is shown in chapter 2. A summary of study is presented in chapter 5.

1.4 References

[1] Anderson N.L.; Anderson N.G. Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* 1998, 19 (11): 1853–61.

[2] Blackstock W.P.; Weir M.P. Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* 1999, 17 (3), 121–7.

[3] de Godoy, L. M. F.; Olsen, J. V.; Cox, J.; Nielsen, M. L.; Hubner, N. C.; Froehlich, F.; Walther, T. C.; Mann, M. *Nature* 2008, *455*, 1251-1254.

[4] Cramer, Grant R.; Van Sluyter, Steve C.; Hopper, Daniel W.; Pascovici, Dana; Keighley, Tim; Haynes, Paul A.; Proteomic analysis indicates massive changes in metabolism prior to the inhibition of growth and photosynthesis of grapevine (Vitis vinifera L.) in response to water deficit. BMC Plant Biology (2013), 13, 49.

[5] Shimwell, N. J.; Bryan, R. T.; Wei, W.; James, N. D.; Cheng, K. K.; Zeegers, M. P.; Johnson, P. J.; Martin, A.; Ward, D. G.; Combined proteome and transcriptome analyses for the discovery of urinary biomarkers for urothelial carcinoma. British Journal of Cancer (2013), 108(9), 1854-1861.

[6] Alves, G.; Yu, Y.K.; Improving Peptide Identification Sensitivity in Shotgun Proteomics by Stratification of Search Space. Journal of Proteome Research (2013), 12(6), 2571-2581.

[7] Cantor, D.; Slapetova, I.; Kan, A.; McQuade, L. R.; Baker, M. S.; Overexpression of αvβ6 integrin alters the colorectal cancer cell proteome in favor of elevated proliferation and a switching in cellular adhesion That Increases Invasion.

[8] Yates, John R.; The Revolution and Evolution of Shotgun Proeomics for Large-Scale Proteome Analysis. Journal of the American Chemical Society (2013), 135(5), 1629-1640.

[9] Claassen, M.; Inference and validation of protein identifications. Molecular & Cellular Proteomics (2012), 11(11), 1097-1104.

[10] Vaudel, M.; Sickmann, A.; Martens, L.; Current methods for global proteome identification. Expert Review of Proteomics (2012), 9(5), 519-532.

[11] Teng, P.; Bateman, N.W; Hood, B.L. and Conrads T.P.; J. of proteome research 2010, 9(12), 6091-6100.

[12] Di Palma, S.; Hennrich, M. L.; Heck, A. J. R.; Mohammed, S.; Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. Journal of Proteomics (2012), 75(13), 3791-3813.

[13] Motoyama, A.; Yates, J. R.; Multidimensional LC Separations in Shotgun Proteomics. Analytical Chemistry (Washington, DC, United States) (2008), 80(19), 7187-7193.

[14] Fang, X.; Balgley B.; Wang W.; Park D.M.; Lee C.S.; comparison of multidimensional shotgun technologies targeting tissue proteomics. Electrophoresis 2009, 30 (23), 4063-70.

[15] Huang, E. L.; Orsat, V.; Shah, M. B.; Hettich, R. L.; VerBerkmoes, N. C.; Lefsrud, M. G.; The temporal analysis of yeast exponential phase using shotgun proteomics as a fermentation monitoring technique. Journal of Proteomics (2012), 75(17), 5206-5214.

[16] Kang, D.; Nam, H.; Kim, Y-S.; Moon, M. H.; Dual-purpose sample trap for on-line strong cation-exchange chromatography/reversed-phase liquid chromatography/tandem mass spectrometry for shotgun proteomics. Application to the human Jurkat T-cell proteome. Journal of Chromatography A (2005), 1070(1-2), 193-200.

[17] Unwin, R.D.; Griffiths J.R.; Whetton A.D.; Simultaneous analysis of relative protein expression levels across multiple samples using iTRAQ isobaric tags with 2D nano LC-MS/MS. Nat Protoc. 2010, 5(9), 1574-1582.

[18] Nagaraj N.; wisniewski J.R.; Geiger T.; Cox J.; Kircher M.; Kelso J.; Paabo S. and Mann M.; Deep proteome and transcriptome mapping of a human cancer cell line. Mol Syst Biol 2011, 7, 1-8.

[19] Beck M.; Schmidt a.; Malmstroem J.; Claassen M.; Or A.; Szymborska A.; Herzog F.; Rinner O.; Ellenberg J.; Aebersold R.; The quantitative proteome of a human cell line. Mol Syst Biol 2011, 7, 1-8.

[20] Dihazi, H.; Prognosis markers for metastatic renal cell carcinoma: quantitative proteomics approach. Expert Review of Proteomics (2013), 10(1), 21-24.

[21] Dagley, L. F.; Emili, A.; Purcell, A. W.; Application of quantitative proteomics technologies to the biomarker discovery pipeline for multiple sclerosis. Proteomics: Clinical Applications (2013), 7(1-2), 91-108.

[22] Marcilla, M.; Albar, J. P.; Quantitative proteomics: A strategic ally to map protein interaction networks. IUBMB Life (2013), 65(1), 9-16.

[23] Holman, S. W.; Sims, P. F. G.; Eyers, C. E.; The use of selected reaction monitoring in quantitative proteomics. Bioanalysis (2012), 4(14), 1763-1786.

[24] Shi, T.; Su, D.; Liu, T.; Tang, K.; Camp, D. G.; Qian, W-J.; Smith, R. D.; Advancing the sensitivity of selected reaction monitoring-based targeted quantitative proteomics. Proteomics (2012), 12(8), 1074-1092.

[25] Kovanich, D.; Cappadona, S.; Raijmakers, R.; Mohammed, S.; Scholten, A.; Heck, A.J. R. Applications of stable isotope dimethyl labeling in quantitative proteomics. Analytical and Bioanalytical Chemistry (2012), 404(4), 991-1009.

[26] Christoforou, A. L.; Lilley, K. S.; Isobaric tagging approaches in quantitative proteomics: the ups and downs. Analytical and Bioanalytical Chemistry (2012), 404(4), 1029-1037.

[27] Arsova, B.; Kierszniowska, S.; Schulze, W. X.; The use of heavy nitrogen in quantitative proteomics experiments in plants. Trends in Plant Science (2012), 17(2), 102-112.

[28] Christoforou, A.; Lilley, K. S.; Taming the isobaric tagging elephant in the room in quantitative proteomics. Nature Methods (2011), 8(11), 911-913.

[29] Joost, W. G.; Bastiaan, B. J. T.; Jeroen, K.; Metabolic labeling of model organisms using heavy nitrogen (15N). Methods in molecular biology (Clifton, N.J.) (2011), 753, 29-42.

[30] Rayavarapu, S.; Coley, W.; Cakir, E.; Jahnke, V.; Takeda, S.; Aoki, Y.; Grodish-Dressman, H.; Jaiswal, K. J.; Hoffman, P. E.; Brown, J. K.; et al. Identification of Disease Specific Pathways Using in Vivo SILAC Proteomics in Dystrophin Deficient mdx Mouse. Molecular & cellular proteomics: MCP (2013), 12(5), 1061-73.

[31] Munday, D. C.; Surtees, R.; Emmott, E.; Dove, B. K.; Digard, P.; Barr, J. N.; Whitehouse, A.; Matthews, D.; Hiscox, J. A.; Using SILAC and quantitative proteomics to investigate the interactions between viral and host proteomes. Proteomics (2012), 12(4-5), 666-672.

[32] Cox, J.; Matic, I.; Hilger, M.; Nagaraj, N.; Selbach, M.; Olsen, J. V.; Mann, M.; practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. Nature Protocols (2009), 4(5), 698-705.

[33] Zanivan, S.; Krueger, M.; Mann, M.; In vivo quantitative proteomics: the SILAC mouse. Methods in Molecular Biology (New York, NY, United States) (2011), 757(Integrin and Cell Adhesion Molecules), 435-450.

[34] Austin, J. R.; Kuestner, E. R.; Chang, K. D.; Madden, R. K.; Martin, B. D.; SILAC Compatible Strain of Pichia pastoris for Expression of Isotopically Labeled Protein Standards and Quantitative Proteomics. Journal of Proteome Research (2011), 10(11), 5251-5259.

[35] Nie, A-Y.; Zhang, L.; Yan, G-Q.; Yao, J.; Zhang, Y.; Lu, H-J.; Yang, P-Y.; He, F-C.; In Vivo Termini Amino Acid Labeling for Quantitative Proteomics. Analytical Chemistry (Washington, DC, United States) (2011), 83(15), 6026-6033.

[36] Geiger, T.; Wisniewski, R. J.; Cox, J.; Zanivan, S.; Kruger, M.; Ishihama, Y.; Mann, M.; Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. Nature Protocols (2011), 6(2), 147-157.

[37] Gygi S.P.; Rist B.; Gerber S.A.; Turecek F.; Gelb M.H.; Aebersold R.; Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. 17, 994-999.

[38] Aerbersold, R.; Gygi, P. S.; Griffin, J. T.; Han, K. M. D.; Yelle, J. M.; The isotope-coded affinity tag reagent method for quantitative proteomics. American Genomic/Proteomic Technology (2001), 1(1), 22, 24, 26-27.

[39] Wu, W. W., Wang, G., Baek, S. J. & Shen, R. F. Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. J Proteome Res 5, 651-8 (2006).

[40] Wiese S.; Reidegeld K.A.; Meyer H.E. and Warscheid B.; Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. Proteomics, 2007, 340-350.

[41] Noirel, J.; Evans, C.; Salim, M.; Mukherjee, J.; Ow, S. Y.; Pandhal, J.; Pham, T. K.; Biggs, A. C.; Wright, C. P.; Methods in quantitative proteomics: setting iTRAQ on the right track. Current Proteomics (2011), 8(1), 17-30.

[42] Wang, W-S.; Liu, X-H.; Liu, L-X.; Jin, D-Y.; Yang, P-Y.; Wang, X-L.; Identification of proteins implicated in the development of pancreatic cancer-associated diabetes mellitus by iTRAQ-based quantitative proteomics. Journal of Proteomics (2013), 84, 52-60.

[43] McDonald, L. and Beynon R.J.; Positional proteomics: selective recovery and analysis of N-terminal proteolytic peptides. Nature Methods, 2005 2(12), 955-957.

[44] Dormeyer W.; Mohanned S.; Breukelen B.; Krijgsveld J. and Heck A.J.R.; Targeted analysis of protein termini. Journal of proteome research 2007, 6, 4634-4645.

[45] Ajami K.; Pitman M.R.; Wilson C.H.; Park J.; Menz R.I.; Starr A.E.; Cox J.H.; Abbott C.A.; Overall C.M.; Gorrell M.D.; Stromal cell-derived factors 1alpha and 1beta, inflammatory protein-10 and interferon-inducible T cell chemo-attractant are novel substrates of dipeptidyl peptidase 8. FEBS lett 2008, 582, 819-825

[46] McQuibban G.A.; gong J.H.; Wong J.P.; Wallace J.L.; Clark-Lewis I.; Overall C.M.; Matrix metalloproteinase processing of monocyte chemoattractant proteins generates CC chemokine receptor antagonists with anti-inflammatory properties in vivo. Blood 2002, 100, 1160-1167

[47] Kleifeld O.; Doucet, A.; Keller, U.; Prudova, A.; Schilling, O.; Kainthan, R.; Starr, A.; Foster, L.J.; Kizhakkedathu, J.; Overall, C.M.; Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products 2010, Nature Biotech. 28(3), 281-288.

[48] Van Damme P.; Staes A.; Bronsoms S.; Helsens K.; Colaert N.; Timmerman E.; Aviles F.X.; Vandekerckhove J. and Gevaert K; Complementary positional proteomics for screening substrates of endo- and exoproteases; Nature Methods, 2010, 7(7) 512-515.

[49] Kuhn K.; Thompson A.; Prinz T.; Muller J. et al.; Isolation of N-terminal protein sequence tags fro cyanogen bromide cleaved proteins as a novel approach to investigate hydrophobic proteins. J. Proteome Res. 2003, 2, 598-609.

[50] Timmer J.C.; Enoksson M.; Wildfang E.; Zhu W.; Igarashi Y.; Denault J-B.; Ma Y.; Dummitt B.; Chang Y-H.; Mast A.E.; Eroshkin A.; Simth J.W.; Tao W.A. and Salvesen G.S.; Profiling constitutive proteolytic events in vivo.; Biochem. J., 2007, 407, 41-48.

[51] Mahrus S.; Trinidad J.C.; Barkan D.T.; Sali A.; Burlingame A.L. and Wells J.A.; Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. Cell, 2008, 134, 866-876.

[52] Agard N.J.; Maltby D.; Wells J.A.; Inflammatory stimuli regulate caspase substrate profiles. Mol. Cell Proteomics, 2010, 9, 880-893.

[53] Xu G.; Shin S.B.Y. and Jaffrey S.R.; Global profileing of protease cleavage sites by chemoselective labeling of protein N-termini. Proc Natl Acad Sci USA 2009, 106, 19310-19315.

[54] Schilling O.; Overall C.M.; Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. Nat Biotechnol 2008, 26, 685-694

[55] Schiling O.; Overall C.M.; Proteome-derived, database-searchable peptide libraries for identifying proteoase cleavage sites. Nat. biotechnol. 2008, 26, 685-694.

[56] Burger W.C.; Interference by carbonyl compounds in the trinitrobenzenesulfonic acid method for amino groups. Anal. Biochem. 1974, 57, 306-309.

[57] McDonald L.; Robertson D.H.; Hurst J.L.; Beynon R.J.; Positional proteomics: preparation of amino-terminal peptides as a strategy for proteome simplification and characterization. Nat. Protoc. 2006, 1, 1790-1798.

[58] Mikami T. and Takao T.; Selective isolation of N-blocked peptides by isocyanate-coupled resin. Anal. Chem. 2007, 79, 7910-7915.

[59] Kleifeld O.; Doucet A.; Prudova A.; auf dem Keller U.; Gioia M.; Kizhakkedathu J.N. and Overall C.M.; Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. Nature protocol 2011, 6, 10, 1578-1611.

[60]Prudova A.; auf dem Keller U.; Butler G.S. and Overall C.M.; Multiplex N terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics. Mol. Cell. Proteomics 2010, 9, 894-911.

[61] Auf dem Keller U.; Prudova A.; Gioia M.; Butler G.S. and Overall C.M.; A statistic-based platform for quantitative N terminome analysis and identification of protease cleavage products. Mol. Cell. Proteomics, 2010, 9, 912-927.

[62] Mommen, G.P.M; Waterbeemd, B.V.D.; Meiring H.D.; Kersten, G.; Heck, A.J.R and Jong, P.J.M; unbiased selective isolation of protein N-terminal peptides from complex proteome samples using phosphor tagging (PTAG) and TiO2-based depletion. MCP, published on June 22, 2012.

[63] Gevaert K.; Goethals M.; Martens L.; Van Damme J.; Staes A.; Thomas G.R.; Vandekerchhove J.; exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. Nat. Biotechnol. 2003, 21, 566-569.

[64] Van Damme, P.; Martens, L.; Van Damme, J.; Hugelier, K.; Caspase-specific and non-specific in vivo protein processing druing Fas-induced apoptosis. Nat. Methods 2005, 2, 771-777.

[65] Staes A.; Van Damme P.; Helsens K.; Demol H. et al Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC) Proteomics 2008, 8, 1362-1370.

[66]Gausdal, G.; Gjertsen, B.T.; McCormack, E.; Van Damme, P. et al.; Abolition of stress induced protein synthesis sensitizes leukemia cells to anthracycline –induced death. Blood 2008, 111, 2866-2877.

[67] Impens, F.; Van Damme, P.; Demol, H.; Van Damme, J.; et al.; Mechanistic insight into taxol-induced cell death. Oncogene 2008, 27, 4580-4591.

[68] Arnesen T.; Van Damme P.; Rolevoda B.; Helsens K. et al; Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. Proc. Natl. Acad. Sci. USA, 2009, 106, 8157-8162.

[69] Staes A.; Impens F.; Van Damme P.; Ruttens B.; Goethals M.; Demol H.; Timmerman E.; Vandederchhove J.; Gevaert K.; Selecting protein N-terminal peptides by combined fractional diagonal chromatography. Nat. Protoc. 2011, 6, 1130-1141.

[70] Van Damme, P.; Maurer-Stroh S.; Plasman K.; van Durme J. et al.; Analysis of protein processing by N-terminal proteomics reveals novel species-specific substrate determinants of granzyme B orthologs. Mol. Cell. Proteomics 2009, 8, 258-272.

[71] Sechi S. and. Chait B.T.; A method to define the carboxyl terminal of proteins. Analytical Chemistry 2000, 72, 3374-3378.

[72] Hiroki Kuyama et al A simple and highly successful C-terminal sequence analysis of proteins by mass spectrometry Proteomics 2008, 8, 1539-1550.

[73] Sonomura, K. et al the specific isolation of C-terminal peptides of proteins through a transamination reaction and its advantage for introducing functional groups into the peptide. Rapid Commun. Mass Spectrom. 2009, 23, 611-618.

[74] Schilling O.; Barre O. Huesgen P.F. and Overall C.M.; proteome-wide analysis of protein carboxy termini: C terminomics. Nature Methods, 2010, 7 (7), 508-511.

Chapter 2 Experiments toward a streamlined workflow for positional proteomics

2.1 Overview of method

The workflow starts by converting lysine to homoarginine by guanidination, followed by sulfonation of N-termini by 4-sulfophenyl isothiocyanate (SPITC). After trypsin digestion, electrostatic repulsion hydrophilic interaction (ERLIC) chromatography is used to enrich SPITC modified peptides. Both the flow-through fraction (containing internal and C-terminal peptides) and eluted fraction (containing SPTIC modified N-terminal peptides) are analyzed by LC-MS/MS. A specialized N-terminal database with sequentially trimmed N-termini is used to identify N-terminal peptides from degraded proteins.

2.2 Consideration for development

The two critical aspects for any positional proteomics workflow are terminal labeling and enrichment. To identify minor species in a complex mixture, some sort of enrichment for the minor species is required. There are various approaches and efforts described for this topic in the literature as summarized in chapter 1. N-terminal enrichment approaches can be categorized into three pathways, namely chemically or enzymatically positive selection of N-termini,[1,2] negative selection (scavenging the internal and C-terminal peptides)[3] and the application of ion

exchange chromatography[4] based on the charge differences among N-terminal , C-terminal and internal peptides. Both negative and positive selection approaches were tested in our study, the results of which are described later in this chapter.

The second critical aspect is the manner of labeling of the protein termini. Here we choose to use SPITC for labeling N-termini. SPITC derivatization is predominantly applied to proteolytic peptides in the literature,[5] so our development effort was focused on modifying the protocol to make it work on complex mixtures of intact proteins with high efficiency. Guanidination was performed prior to SPITC, which converted lysine to homoarginine by blocking the side chain of lysine. Therefore, SPITC derivatization only occurred at the N-terminal site and was assigned as such during data interpretation.

Besides the two aspects mentioned above, the overall workflow was streamlined, particularly with respect to protein/peptide purification steps.

2.3 Optimization

Experiments designed to optimize each step are presented in this section in the same order as the steps in the workflow. These are: guanidination reaction, SPITC reaction, protein purification, trypsin digestion, enrichment, LC-MS/MS and database searching.

2.3.1 Guanidination reaction

The purpose of guanidination is to selectively cap the side chain of lysine, thus leaving only the N-terminal amine for the next step in the SPITC reaction. The reaction is selective, except when glycine is the N-terminal amino acid; in such cases the primary amine group of glycine can also be modified.[6] The guanidination reaction is performed as reported by Reilly et al.[7]

A standard peptide with the sequence TNEIVEEQYTPQSLATLESVFQELGK (m/z 2952.4; m/z for guanidination product is 2994.4) was used to monitor this reaction. The conversion rate was computed as the area of m/z 2994.4 divided by the sum of the areas of m/z 2952.4 and m/z 2994.4 obtained from LC-MS. The concentration for isourea was 300 mM, the starting material of peptide was 0.12 mM and only 1 μL was used for each reaction. 10 μl NH_4OH with 15 μL isourea is the recommended condition by the kit vendor (Sigma Aldrich). The conversion was around 90%. Doubling the use of isourea or hydroxide did not increase the conversion rate. Due to the downstream SPITC derivatization of primary amines, the introduction of a large quantity of ammonium ions is undesirable, so NH_4OH was replaced with 10 μL of 0.1 M NaOH. While this change in the base caused the conversion rate to drop significantly to around 25%, further study determined that titrating the reaction with NaOH to pH 10 produced a comparable conversion rate to that for the reaction using NH_4OH (Table 2.1).

Table 2.1 Guanidination optimization

base	isourea (μl)	conversion rate
NH_4OH-10 μl	15	85.7%
NH_4OH-10 μl	15	92.0%
NH_4OH-10 μl	30	85.0%
NH_4OH-20 μl	30	89.6%
NaOH (0.1 M)-10 μl	30	24.6%
NaOH (0.1M)-10 μl	30	25.2%
NaOH pH 10	15	89.7%
NaOH pH 10	15	87.4%

Notably, the reaction completeness is 90% in our measurement using LC-MS/MS, which is lower than in a reported study by MALDI-MS.[7] Such discrepancy may arise from the mass spectrometer employed. In MALDI-MS, the high gas-phase basicity of guanidinated lysine facilitates proton capture and charging in the positive ion mode. Thus, ionization of guanidinated lysine peptides occurs more efficiently than for the non-guanidinated lysine-containing peptides that are characterized by a lower gas-phase basicity. When the two are present as a mixture, the former can capture protons more efficiently, and thereby, the signal of the latter may be suppressed. In contrast, in LC-ESI-MS, the non-guanidinated standard peptide and the guanidinated reaction product were separated by LC before MS detection, facilitating detection of the minor species.

2.3.2 SPITC labeling

The SPITC reaction protocol followed the description by the Cotter lab.[8] Specifically, a solution of SPITC was freshly prepared at 50 mg/mL in 20 mM sodium bicarbonate buffer (pH 9). The guanidinated protein mixture was adjusted to pH 9 by addition of NaOH (1 N). A 100-fold molar excess of SPITC was added to the mixture. The mixture was then incubated at 65 °C for 1 hour.

Using the peptide AAAAK (1 mg/mL in water) as a model compound, the SPITC modification (50 mg/mL in 20 mM sodium bicarbonate buffer (pH 9)) was 100% complete as determined by LC-MS after incubating at 65 °C for one hour with all tested conditions listed in Table 2.2. This reaction is robust as long as ammonium-based buffer is avoided.

Table 2.2 Evaluation of SPITC reaction conditions

AAAAK (µl)	SPITC (µl)	conversion
10	20	100%
10	40	100%
10	60	100%
10	80	100%

2.3.3 Isolation/Purification

It is easy to test reaction conditions with peptides, where the contaminant reagents can be diverted to waste by an on-line valve switch in HPLC or be

removed by off-line C18 cartridge clean-up. When dealing with protein purification, it is common to use trichloro-acetic acid (TCA) or ice cold acetone to precipitate protein into pellet and to follow with a wash. However, once the proteins aggregate into pellets, it is extremely difficult to bring them back into solution. While solubilization can be promoted using urea or guanidine hydrochloride, these denaturants are problematic for the downstream trypsin digestion and require additional cleanup efforts. Therefore, we endeavored to find an approach that avoided protein precipitation entirely. Alternatively, the excess reagents can be removed by dialysis. However, this method usually takes days to complete and is therefore undesirable.

A desalting Zeba column (7k MWCO) from Thermo Scientific was used to isolate/purify proteins in our workflow, allowing the proteins to remain in solution all along the process. The success of this approach may arise from the increase of the basicity of lysine due to the guanidination reaction; this increased basicity promotes the solubility of denatured proteins in mild buffer. Following the protocol depicted in the product booklet, Zeba spin desalting columns are buffer exchanged 4 times with 300 μL of 50 mM ammonium acetate (or other buffer suitable for the following enzymatically digestion) by centrifuging for 1 min at 1500 g, then no more than 130 μL of protein is loaded on top of the resin bed, and

samples are collected after centrifuging for 2 min at 1500 g. The whole process takes less than 10 min. We adapted this process in the final workflow.

2.3.4 Trypsin digestion

Trypsin (Promega, sequence grade) is added at an enzyme/substrate ratio of 1:50 to the buffer exchanged sample. The sample is then incubated at 37 degrees overnight. No further modification was made to the digestion condition.

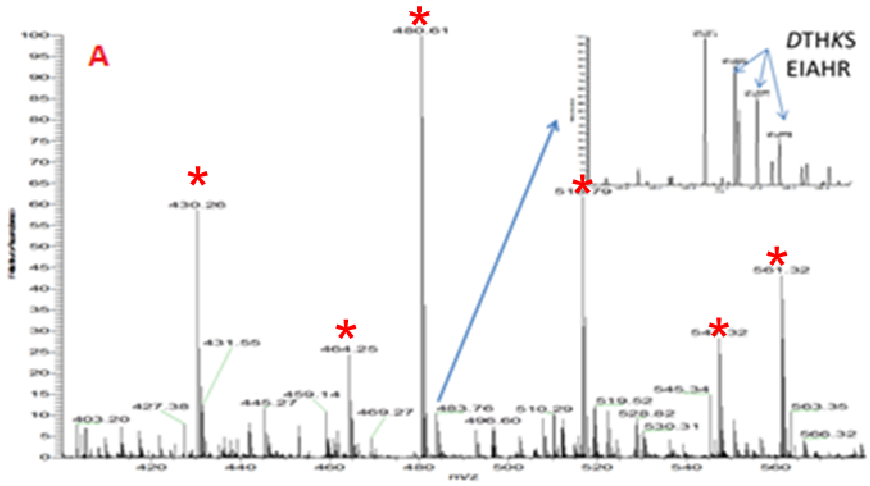
2.3.5 Enrichment

The enrichment step occurs after trypsin digestion. Two enrichment methods were tested. In the end, the iTRAQ application (Chapter 4) used NHS activated agarose beads for enrichment, while the SPITC approach (Chapter 3) used the ERLIC method.

2.3.5.1 N-termini enrichment by NHS activated agarose spin column

We first tested negative selection by scavenging the internal and C-terminal peptides through commercially available NHS activated agarose spin columns (Thermo Scientific), due to the simplicity of this approach. Following the protocol in the brochure, the whole process took one hour. After desalting using C18 cartridge, the sample was analyzed using LC-MS/MS. The effectiveness was demonstrated in Figure 2.1 using BSA (50 μg starting material) as a testing protein. Before enrichment, the modified N-termini were at a low abundance level

(insert of the top figure in figure 2.1) compared to other peptides when averaging the total ion chromatogram (TIC) over the range of HPLC elution. After enrichment, the abundance of N-terminal peptides was raised and became present as one of the major components in the TIC over the full HPLC elution range. The six most abundant m/z values before enrichment including m/z at 480.61, 516.79, 430.26, 561.32, 547.32 and 464.25 are marked in Figure 2.1A, the counterparts of these six peptides are also marked in Figure 2.1B. Significant intensity drops are observed for all six peptides.



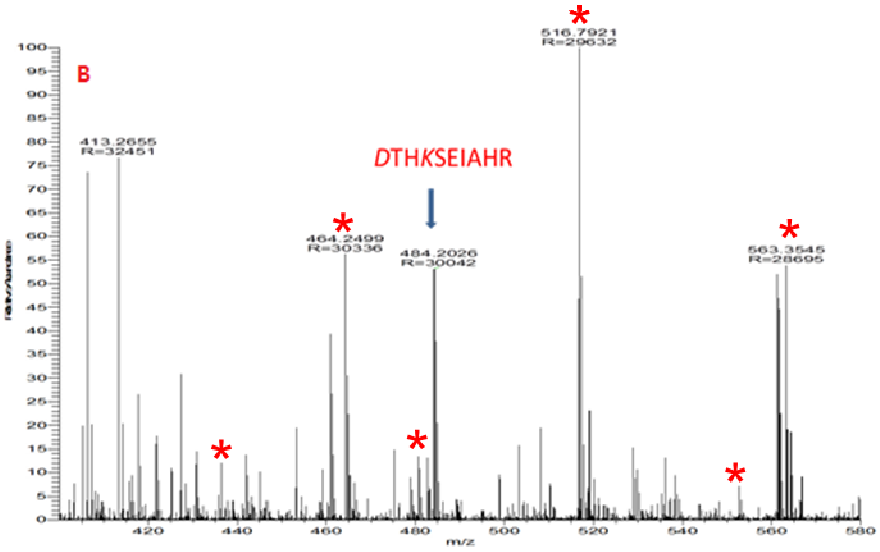


Figure 2.1 Comparison between before (A) and after (B) enrichment by NHS activated agarose resin spin column for BSA sample. N-terminal peptides (SPITC-DTHK(guandinyl)SEIAHR m/z= 484.20^{3+}).

This approach also worked well for the iTRAQ-based application (Chapter 4). However, unsatisfactory results were obtained using the NHS-activated agarose resin spin column to process an *E. coli* cell lysate sample with SPITC modified peptides. For 100 µg of *E. coli* starting material, we were able to obtain 62 annotated N-terminal peptides. A second round of enrichment generated a total of 13 annotated N-terminal peptides, while a third round of enrichment produced only 9 annotated N-terminal peptides. Non-specific interactions of peptides with the NHS-activated resin may be the reason for the loss of N-terminal peptides during the resin enrichment. Due to the inefficiency of this approach, we switched to charge-based enrichment for SPITC modification.

2.3.5.2 SPITC peptide enrichment by ERLIC

We refer to the protocol by Alpert[9] for ERLIC-based enrichment. A PolyWAX LP, 5-μm, 300 Å, 2.1 mm x 50 cm (Poly LC Inc., Columbia, MD) column was used, with mobile phase A as 20 mM ammonium formate, pH 2.2 with 70% acetonitrile; and mobile phase B as 900 mM ammonium formate, pH 2.2 with 10% acetonitrile. To facilitate downstream MS analysis, the column was used in a solid phase extraction fashion and only two fractions were collected. Therefore, 100% A was applied at the beginning for 3 minutes for the first fraction. Then, the mobile phase was switched to 100% mobile phase B with a one minute gradient and held for 10 minutes for the second fraction. The salt is volatile and is removed by vacuum centrifuge.

2.3.6 Mass spectrometer selection

High mass accuracy of parent ions and fragment ions is extremely important for accurate peptide sequence identification. Both Orbitrap XL (Thermo) and QExactive (Thermo) were used in this study. With the Orbitrap XL instrument, a typical workflow involves a survey scan at high resolution mode (R=30,000, FT mode), followed by 10 MS/MS scans at low resolution mode (trap mode). In the follow-up database search, mass accuracy for the parent ions was set to be 10 ppm and for the fragment ions at 0.8 Da. Such low mass accuracy of fragment ions may very well contribute to the false identification of peptide and proteins in our opinion. It is possible to acquire data in high resolution mode for both survey scan

and MS/MS scan using Orbitrap XL, however, the cycle time is accordingly prolonged and the duty cycle is reduced drastically, which means much less input into the database, leading to less peptide and protein identification. This feature of the Orbitrap XL instrument impairs its application in complex system (e.g. proteomics) in our proposed workflow.

Instead, QExactive is much better fit to the proposed workflow. With its high speed (5X faster than Orbitrap XL), both survey scans and MS/MS scans can be performed in high resolution mode. In the final set-up, data is acquired using QExactive, with one survey scan followed by 12 MS/MS scans at FT mode for both types of experiments. HCD mode of activation is employed for activation of ions. An example is shown in figure 3.5 (Chapter 3) that demonstrates that HCD offers much more signature ions in MS/MS spectra for peptide sequencing.

2.3.7 Database searching

The analysis of mass spectrometric raw data was carried out using Proteome Discovery 1.3 software (Thermo Fisher Scientific, Bremen, Germany) applying standard settings unless otherwise noted. The *E. coli* K12 strain database (Uniprot) prepared in the fashion similar to that described by Dormeyer *et al*[10] was used for searching. Specifically, all protein sequences in the database were cut at arginine, and for each resulting peptide, multiple sequence database entries were generated

that represented sequential truncations of N-terminal amino acids. Only sequences with 6 amino acids or longer were retained in the database. The workflow utilizing both SEQUEST and Mascot search was used as shown in Figure 2.2, with the convergence handled by Percolator. Decoy database searches were performed with False Discovery Rate (FDR) tolerances set to 5% and 1% for modest and high confidence filtering settings, respectively. Other settings are depicted in Chapter 3.

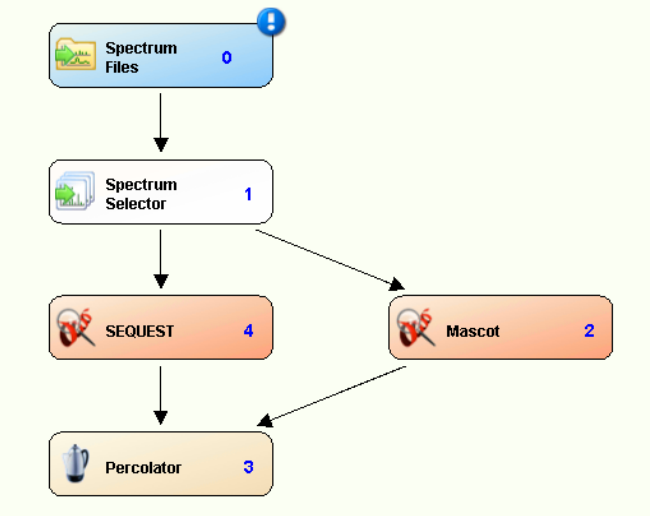


Figure 2.2 Workflow for database searching

2.4 Summary

This chapter described method development for several key steps in the workflow. The application of these methods is shown in chapter 3 and chapter 4.

2.5 References

[1] Timmer J.C.; Enoksson M.; Wildfang E.; Zhu W.; Igarashi Y.; Denault J-B.; Ma Y.; Dummitt B.; Chang Y-H.; Mast A.E.; Eroshkin A.; Smith J.W.; Tao W.A. and Salvesen G.S.; Profiling constitutive proteolytic events in vivo.; Biochem. J., 2007, 407, 41-48.

[2] Xu G.; Shin S.B.Y. and Jaffrey S.R.; Global profiling of protease cleavage sites by chemoselective labeling of protein N-termini. Proc Natl Acad Sci USA 2009, 106, 19310-19315.

[3] Kleifeld O.; Doucet A.; Prudova A.; auf dem Keller U.; Gioia M.; Kizhakkedathu J.N. and Overall C.M.; Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. Nature protocol 2011, 6, 10, 1578-1611.

[4] Staes A.; Impens F.; Van Damme P.; Ruttens B.; Goethals M.; Demol H.; Timmerman E.; Vandederchhove J.; Gevaert K.; Selecting protein N-terminal peptides by combined fractional diagonal chromatography. Nat. Protoc. 2011, 6, 1130-1141.

[5] Wang, D; Kalb S. R. and Cotter, R.J.; Improved procedure for N-terminal sulfonation of peptides for matrix-assisted laser desorption/ionization post-source decay peptide sequencing, Rapid Commun. Mass Spectrom. 2004, 18, 96-102

[6] Hale, J. E. et al., increased sensitivity of tryptic peptide detection by MALDI-TOF mass spectrometry is achieved by conversion of lysine to homoarginine. Anal. Biochem., 2000, 287, 110-117

[7] Beardsley, R and Reilly, J. P; Optimization of guanidination procedures for MALDI mass mapping. Anal.Chem 2002, 74, 1884-1890

[8] http://www.hopkinsmedicine.org/institute_basic_biomedical_sciences/research_centers/high_throughput_biology_hit/technology_center_networks_pathways/pdfs/protocols/COTTER_N_term_derivatiz.pdf

[9] Alpert, A.; Mitulovic G. and Mechtler, K.; Isolation of tryptic phosphopeptides by ERLIC (electrostatic repulsion-hydrophilic Interaction chromatography). HPLC 2008 poster P-2412-W

[10] Dormeyer, W.; Mohammed, S.; Breukelen, B.; Krijgsveld, J.; Heck, A. J.R. Targeted Analysis of protein termini 2007, 6, 4634-4645

Chapter 3 High Fidelity Approach for Proteomic Scale Enrichment and Identification of N-termini

3.1 ABSTRACT

A novel workflow was designed for the large-scale identification of protein N-terminal sequences. The workflow started with converting lysine to homoarginine by guanidination, followed by reaction with sulfonation of N-termini by 4-sulfophenyl isothiocyanate (SPITC). Upon trypsin digestion, the N-terminal peptides possessed one negatively charged sulfonate site at low pH, while all amino acids of internal and C-terminal peptides were neutral or positively charged. This difference was exploited to enrich the SPITC modified N-terminal peptides by electrostatic repulsion hydrophilic interaction (ERLIC) chromatography in which the internal and C-terminal peptides eluted at the void volume, and the SPITC peptides were retained in the column due to the hydrophilicity and electrostatic attraction of the sulfonyl group to the stationary phase. Both the flow-through fraction (containing internal and C-terminal peptides) and eluted fraction (containing SPTIC modified N-terminal peptides) were analyzed by LC-MS/MS. A specialized N-terminal database with sequentially trimmed N-termini was used to identify N-terminal peptides from degraded proteins. Upon HCD, SPITC peptides in ESI generated predominately y-type ion series, similar to what has been

previously noted for MALDI MS/MS spectra of N-sulfonated species; such simplification of spectra enables the identification of N-termini with high confidence. The appearance of b_1 + SPITC product ions upon HCD further boosts the confidence for N-terminal identifications. This method was applied to an *E. coli* cell lysate, thus allowing the identification of approximately 350 high confidence N-terminal peptides (85% represented neo-N-termini from protein degradation and 15% from leading methionine excision). These N-terminal peptides represented 274 distinct *E. coli* proteins, 224 of which were also identified in the analysis of flow-through fractions from internal peptides. The approach presented here resulted in an approximately 10-fold enrichment of N-terminal peptides, and greatly improved confidence for each MS/MS assignment over other positional proteomics approaches.

3.2 INTRODUCTION

Protein isoforms at N-termini and C-termini are diverse for several reasons. They may be generated during the course of protein biosynthesis due to alternative initiation of transcription within genes, by transcription processes that are independent of annotated gene boundaries,[1-2] from post translational modification by mRNA splicing[3] or by enzymatic protein processing.[4-5] The structural changes occurring at protein termini often correlates with function alteration. Many human diseases are associated with proteolytic truncation or terminal modification;[4-7]

therefore, it is essential to characterize protein N-termini and C-termini for functional annotation of proteomes of interest.

Mass spectrometry (MS) driven techniques show advantages in such applications due to their sensitivity, speed and selectivity in identification of proteins (even in complex systems such as a whole cell lysate or serum samples). Several strategies have been reported to selectively isolate protein N-terminal peptides. The technique called combined fractional diagonal chromatography (COFRADIC), developed by the Gevaert group, has been applied to map protease processed sites in human Jurkat cell cultures[8], anthracycline-induced acute myelogenous leukemia cells[9], and human A549 nonsmall-cell lung carcinoma cells[10]. In this method, the free amine of N-termini and lysine side-chains were tagged by acetylation with either a natural or isotopically labeled reagent, followed by trypsin digestion. N-termini (including the original N-termini of proteins and neo-N-termini generated upon proteolytic cleavage) and C-termini were enriched by strong cation exchange chromatography at low pH. Secondary enrichment was also performed using TNBS derivatization. The hydrophobicity of TNBS causes C-termini peptides to elute at later retention times on reversed-phase columns, whereas N-terminal peptides elute earlier.

In an alternative method, McDonald *et al.* reported a protocol for positional proteomics where free amine groups in proteins were first blocked by acetylation,

followed by N-termini enrichment by NHS-activated sepharose resin and MS analysis.[11] Recently, the terminal amine isotopic labeling of substrates (TAILS) method was reported by Kleifeld *et al.* for identifying acetylated, cyclized N-termini and matrix metalloproteinase (MMP)-2 cleavage sites in mouse fibroblast secretomes[12]. Here, reductive dimethylation reactions were performed on N-termini and lysine side-chains; internal and C-terminal peptides were covalently bonded with dendritic polyglycerol aldehyde polymers, which were later removed by centrifugation. Another approach utilizing unbiased selective isolation of N-terminal peptides using phospho-tagging (PTAG) and TiO_2-based depletion was shown to be effective for N-terminal identification. In this approach, dimethylation was first performed to label N-terminal and lysine side chain amino groups at the protein level. Proteins were then enzymatically digested, and the newly formed internal peptides were modified with PTAG reagent and further depleted from the solution by binding onto TiO_2. The high conversion rate of PTAG and high selectivity of TiO_2 toward phospho-peptides makes this approach very attractive.

Frequently, the value of positional proteomics is questioned due to the reliance on single peptides, or so-called 'one hit wonders'[13] to identify whole proteins, as this approach is expected to be susceptible to many false positives. Previous efforts to reject the false identifications involved tailoring the search space in the database[14-15] or building post-processing tools[16-17]. The work herein

aims to reduce the number of false identifications by improving MS/MS spectral quality. We utilize 4-sulfophenyl isothiocyanate (SPITC), which serves to simultaneously aid the generation of more diagnostic MS/MS spectra and improve N-terminal enrichment.

3.3 MATERIALS AND METHODS

3.3.1 Materials

E. coli cell lysate was obtained from McLab (San Francisco, CA in USA). Bovine serum albumin (BSA), 4-sulfophenyl isothiocyanate, dithiothreitol and iodoacetamide were purchased from Sigma Aldrich (St. Louis, MO in USA). Trypsin Gold (mass spectrometry grade) was obtained from Promega (Madison, WI in USA), whereas the Zeba spin desalting column (7K MWCO) was from Pierce Biotechnology(Rockford, IL in USA)

3.3.2 Protein denaturation, reduction and alkylation

Sample, consisting of either 100 μg of BSA or 500 μg of *E. coli* cell lysate was dissolved to 5 mg/mL in 6 M urea and 20 mM sodium phosphate buffer (pH 8). 50 mM dithiothreitol (DTT) was used to reduce disulfide bonds by incubating for 1 hr at 37 °C. Free sulfhydryls were alkylated with 50 mM iodoacetamide (IAM) at 25°C for 30 min. The unreacted IAM was quenched with DTT.

3.3.3 Guanidination

The guanidination kit was obtained from Sigma-Aldrich. An *o*-methylisourea hemisulfate solution was prepared at 6 mg/mL in water. The protein solution was adjusted to pH 10 by addition of NaOH (1 N) and a 100-fold molar excess of *o*-methylisourea hemisulfate was added. The mixture was then incubated at 65 °C for 30 min. The reaction was quenched by formic acid addition to pH 3.

3.3.4 SPITC modification of N-termini

A solution of SPITC was freshly prepared at 50 mg/mL in 20 mM sodium bicarbonate buffer (pH 9). The guanidinated protein mixture was adjusted to pH 9 by addition of NaOH (1N). A 100-fold molar excess of SPITC was added to the mixture. The mixture was then incubated at 65 °C for 1 hr.

A Zeba spin desalting column (7K MWCO) was used to clean the sample. Briefly, the spin column was buffer exchanged with 50 mM ammonium acetate by centrifuging four times at 1500 *g* for one min. A 130 µL aliquot of the sample was applied on top of the compact resin bed and centrifuged at 1500 *g* for 2 min to collect the sample.

3.3.5 Trypsin digestion

Trypsin was added to the buffer exchanged sample at an enzyme : substrate ratio of 1:50, and the sample was then incubated at 37 °C overnight. The reaction was quenched with addition of formic acid to a pH of 2 to 3. To prepare for ERLIC

injection, acetonitrile was added to the sample to reach 70% acetonitrile in the final solution.

3.3.6 N-terminal enrichment by ERLIC

An Agilent 1200 HPLC system, and a PolyWAX LP, 5-μm, 300 A, 2.1mm x 50cm (Poly LC Inc., Columbia, MD) column were used for ERLIC. The stationary phase consists of a silica-based material with an adsorbed, cross-linked coating of linear polyethyleneimine. The flow rate was 300 μL/min and the column temperature was 25 °C. The UV detection wavelengths were set to 215 nm and 280 nm. Mobile phase A was 20 mM ammonium formate, pH 2.2, with 70% acetonitrile. Mobile phase B was 900 mM ammonium formate, pH 2.2, with 10% acetonitrile. The column was used in a solid phase extraction fashion and only two fractions were collected. Therefore, 100% A was applied at the beginning for 3 minutes to obtain the first fraction. Then, the mobile phase was switched to 100% mobile phase B with a one minute gradient and held for 10 minutes to obtain the second fraction. The first fraction (the flow through) predominately contained neutral and positively charged peptides; the second fraction is enriched in SPITC labeled peptides and other negatively charged peptides. Both fractions were dried using a SpeedVac evaporator (Thermo Electron) and then re-dissolved with 30 μL of water and subjected to LC-MS/MS analysis or stored at -20 °C for further analysis.

3.3.7 LC-MS/MS analysis

LC-MS/MS was performed on a Dionex RSLC nano system (Thermo Scientific, San Jose, CA) coupled to a Q Exactive orbitrap mass spectrometer (Thermo Scientific, San Jose, CA). An Easy Spray column from Thermo Scientific (50 cm* 75 μm, pepmap C18, 2 μm particle) was used with mobile phase A (0.1% formic acid in water) and mobile phase B (0.1% formic acid in acetonitrile). Peptides were eluted over 180 min by linearly increasing mobile phase B from 0 to 30%. Mass spectrometry data were acquired automatically using Xcalibur software (Thermo Scientific) in data dependent acquisitions, with both survey scans and MS2 scans in the FT mode (resolution set at 35000 for parent ions and 17500 for fragment ions). Each survey scan was followed by 12 MS/MS scans. MS/MS spectra were acquired in the HCD mode using a normalized collision energy (NCE) of 25.

3.3.8 Peptide identification

Peptides were identified using the Thermo Proteome Discovery 1.3 software package (Thermo Scientific). Peptide and protein identifications were obtained using Sequest and Mascot algorithms. The *E. coli* K12 strain database prepared in a fashion similar to that described by Dormeyer *et al*[14] was used for searching. Specifically, all protein sequences in the database were cut at arginines, and for each resulting peptide, multiple sequence database entries were generated that

represented sequential truncations of N-terminal amino acids. Only sequences 6 amino acids or longer were retained in the database. Mass tolerance for parent ions was set to 10 ppm and 0.1 Dalton for fragment ions. Trypsin cleavage only at arginine residues was applied with up to 1 missed cleavage in the search. Carboamidomethylation of cysteine (+57.021Da) was set as a static modification. Dynamic modification included guanidination for lysine (+42.022 Da) and N-terminal or (to account for incomplete guanidination) lysine modification with SPITC (+215.971Da). Percolator®, embedded in the proteome discovery software package, was used for peptide filtering, with a peptide false discovery rate (FDR) threshold set to 0.01.

3.4 RESULTS AND DISCUSSION

The strategy of the developed method was to first convert lysine to homoarginine by guanidination. In a second step, N-termini were sulfonated with 4-sulfophenyl isothiocyanate (SPITC), thus permitting enrichment based on the newly acquired N-terminal negative charge. The downstream analysis strategy was to employ HCD activation for ion dissociation, coupled with high resolution MS/MS acquisitions.

3.4.1 Overall workflow description

As depicted in the flow diagram in Figure 1, proteins were first denatured, reduced and alkylated. Prepared proteins were then subjected to guanidination resulting in conversion of lysine residues to homoarginine. Afterwards N-termini were sulfonated by reaction with 4-sulfophenyl isothiocyanate (SPITC), which has been shown to be useful for a variety of proteome applications. [22-24] Following trypsin digestion, SPITC modified N-terminal peptides were enriched by electrostatic repulsion hydrophilic interaction (ERLIC) chromatography. Both the flow-through fraction (containing predominantly internal and C-terminal peptides) and eluted fraction (enriched for SPTIC modified N-terminal peptides) were analyzed by LC-MS/MS.

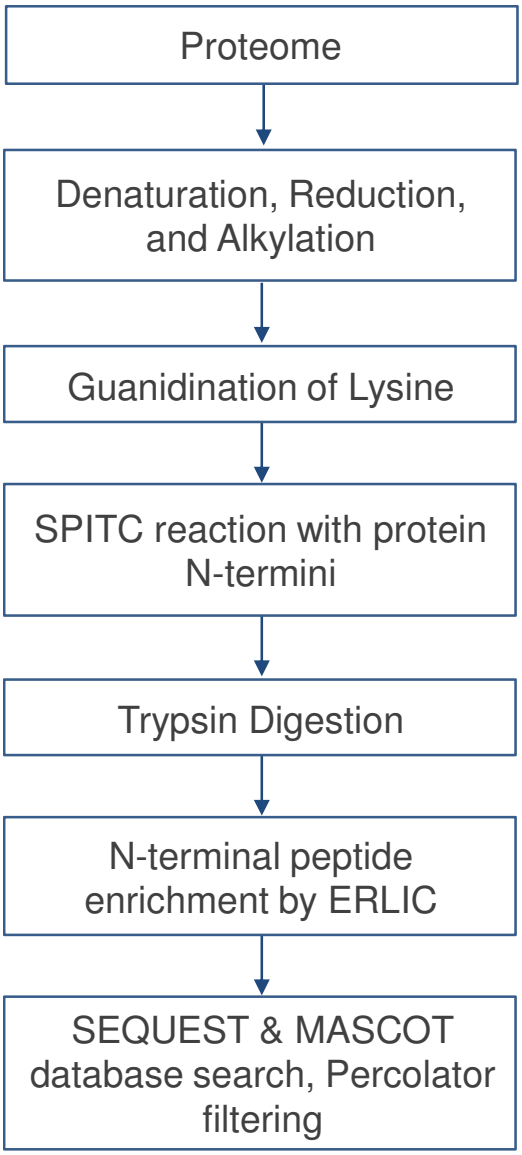


Figure 3.1 Flow chart for N-terminal identification by SPITC modification and ERLIC enrichment.

The above described derivatization procedure offers several advantages. First, 4-sulfophenyl isothiocyanate (SPITC) introduces a negatively charged sulfonic acid group to the N-teruminus. Upon dissociation, the negative charge neutralizes N-terminal ions (a and b ions), thereby promoting the generation and survival of C-terminal ions (y ions). Therefore, the overall effect of SPITC

derivatization is to improve tandem mass spectra by generating predominantly y-type ions. An additional benefit is that often the [b_1 + SPITC] product ion is present, which confirms that modification of the N-terminus had indeed occurred. These spectral features/characteristics increase confidence in identifying N-terminal peptides and allow discrimination between SPITC derivatization occurring at the desired N-terminus position from those potentially occurring at lysine side-chains.

A second advantage of SPITC derivatization is that the negative charge obtained upon sulfonylation allows differentiation of N-terminal peptides from other peptides in the mixture, and therefore can be used to selectively enrich N-terminal peptides. ERLIC enrichment entails N-terminal peptides retention on the column due to hydrophilic interaction and electrostatic attraction, while all the other peptides elute at close to the void volume due to electrostatic repulsion.

A third advantages related to the guanidination reaction that the positive charge from lysine is retained, thus preserving solubility and ionization efficiency. The guanidination reaction converts lysine to homoarginine, conserving the positive charge, as opposed to other chemical modifications such as dimethylation or acetylation where the side chain of lysine is converted to an uncharged group. The guanidination reaction increases the basicity of lysine thus promoting the solubility of denatured proteins, and increasing the charged character of the

peptides which leads to improved ionization efficiency. The relatively high solubility of these modified proteins in mild buffer solutions facilitates clean-up prior to trypsin digestion. In the current protocol, excess reagent was removed and buffer was exchanged directly to the trypsin digestion buffer (50 mM ammonium acetate) by Zeba Spin desalting column (7K MWCO) in 10 min.

A fourth advantage of the combination of guanidination and SPITC derivatization in terms of streamlining the workflow is that the modification reactions are relatively rapid, as it takes 30 min to complete the guanidination reaction, and one hour to finish the SPITC modification. This time scale is comparable to that of the acetylation reaction (2h), but much shorter than the dimethylation reaction (overnight). In addition, there is no need for clean-up until all the reactions are completed, because any excess reagent from one step does not interfere with the next step.

Lastly, the current workflow obviates more laborious procedures such as protein precipitation and peptide desalting, making it more convenient for higher sample throughput. Instead of precipitating proteins, Zeba Spin desalting columns (7K MWCO) were used to clean up samples, enabling numerous samples to be processed in 10 min using a centrifuge.

3.4.2 Chemical modification of free amine groups in proteins

After denaturing, reduction and alkylation of sulfuryl groups in proteins, the free amine groups were modified by a stepwise procedure. First, selective modification was performed on the ε-amine of the side chain of lysine to convert lysine to homoarginine (the reaction is shown in Figure 3.2). It has been reported that the reaction can be completed in approximately 5 min, because after this time interval, no signals from unguanidinated-lysine containing peptides were observed by MALDI-MS.[19] The reaction efficiency of this reaction was tested using the model peptide TNEIVEEQYTPQSLATLESVFQELGK and it was found by LC-ESI-MS that the reaction proceeded to approximately 90% completion under optimum conditions. After guanidination, SPITC modification (Figure 3.3) was performed. Using the peptide AAAAK as a model compound, SPITC modification proceeded to 100% completion upon incubating at 65 °C for one hour with 20 mM sodium bicarbonate as buffer.

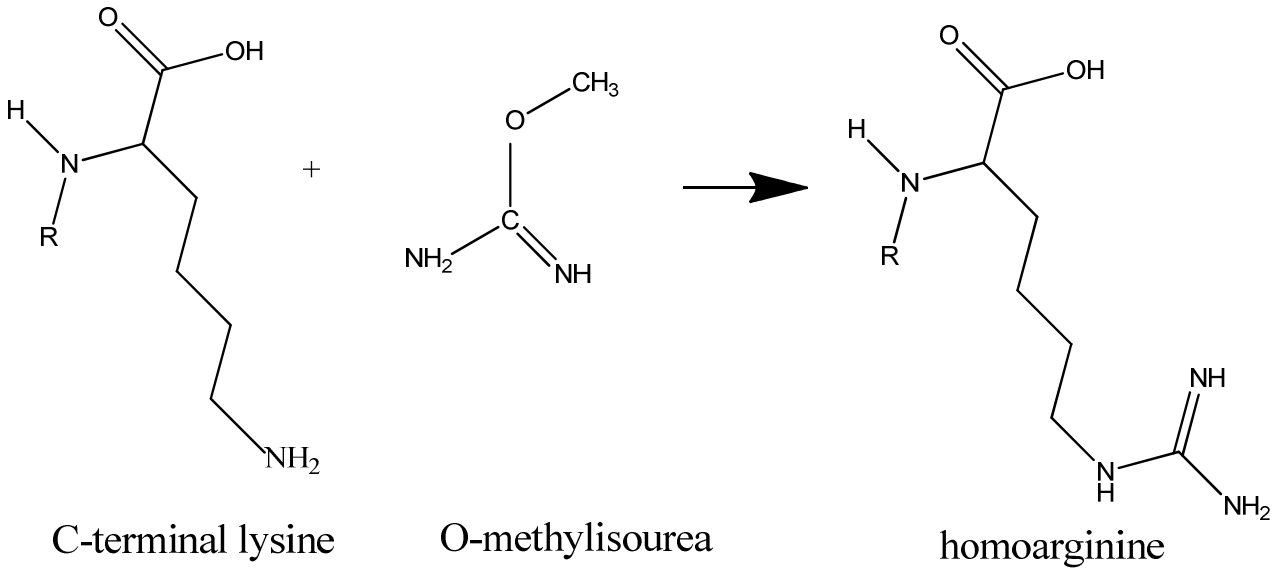


Figure 3.2 Schematic illustration for the guanidination reaction

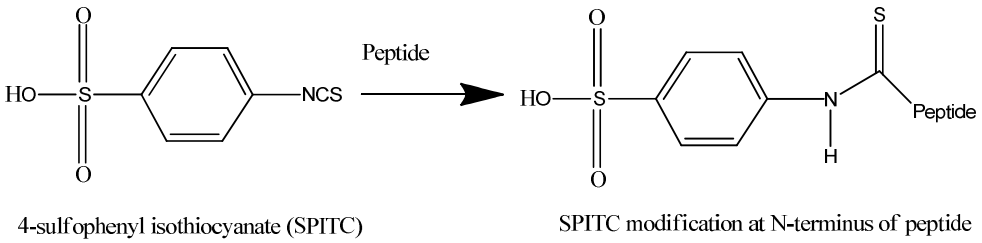


Figure 3.3 Schematic illustration of SPITC modification

3.4.3 Trypsin digestion specificity after guanidination

Upon guanidination, lysine was converted to homoarginine. The selectivity of trypsin digestion at homoarginine was evaluated using BSA. As expected, the cleavage capacity at the now modified lysine was dramatically reduced. Taking the N-terminal peptide of BSA as an example, the sequence for zero missed cleavages is DTH*K* (m/z 757.2 $^{+}$, lysine converted to homoarginine, N-terminus modified by SPITC), the sequence for one missed cleavage is DTH*K*SEIAHR (m/z 725.8 $^{2+}$ and 484.2 $^{3+}$; lysine converted to homoarginine; N-terminus modified by SPITC). The ratio of the peak area (the combination of doubly and triple charged peaks) in the extracted ion chromatogram for the one missed cleavage product versus the area for the zero missed cleavage product was 116 : 1. In a control sample, trypsin digestion for BSA without guanidination indicated that the peak area ratio for the same pair was 0.45 : 1. The conversion of lysine to homoarginine therefore hindered the trypsin activity on this site, and the majority of trypsin digested product of guanidinated protein resulted in peptides ending with arginine at the C-termini.

3.4.4 Effectiveness of N-terminal enrichment

The SPITC modification introduces a sulfonic acid group to N-terminal peptides which remains negatively charged even at low pH. We viewed enrichment/purification of these sulfonated N-terminal peptides by ERLIC as an

attractive strategy based on reports that tryptic peptides containing only one phosphate can be separated from peptides with no phosphates[20]. In our approach, an anion exchange column is essentially used in hydrophilic mode. At pH 2.2, basic amino acids and the primary amines of N-termini possess positive charges, the carboxylic acid sites at C-termini and on glutamic acid and aspartic acid side chains are neutral. Phosphate groups, however, are negatively charged. Thus, by operating in the hydrophilic mode, the hydrophilicity of phosphate groups together with their electrostatic attraction to positively charged sites on the stationary phase overcomes the electrostatic repulsion experienced by protonated amino acid sites on the peptides, and allows separation of phosphopeptides from peptides without such anionic groups. We reasoned that SPITC peptides (bearing sulfonate groups) will interact similarly with the stationary phase as peptides with single phosphate groups. However, since phosphorylated peptides are generally present in much lower abundances than N-terminal peptides, their interference should be minimal when attempting to identify N-terminal peptides. To our knowledge, this is the first application of the ERLIC technique for sulfated peptide enrichment from non-sulfated peptides, and the approach could prove to be useful for analyzing important post-translational modifications such as tyrosine sulfation.

Another aspect worth mentioning concerning ERLIC is that an ammonium formate based mobile phase system was used, which simplifies sample clean-up.

For each ERLIC analysis, two fractions were collected. The first fraction contained the internal and C-terminal peptides, which were positively charged at pH 2.2 and were repelled from the column, and thus eluted with the void volume. It is reasonable to expect much higher absorbance from the first fraction than from the second fraction, since the N-termini only constitute a small portion of the peptide pool. We did observe a large absorbance at 215 nm for the first fraction; however, the high concentration of ammonium formate (900 mM) skewed the UV absorbance reading at 215 nm and no quantitative data was obtainable in terms of the ratio of these fractions. Due to the presence of the benzyl group introduced by SPITC derivatization, a high UV absorbance at 280 nm was observed for the second fraction, demonstrating that these peptides are enriched with SPITC peptides as shown in Figure 3.4. Additional comparisons can be found in the section describing peptide identification results below.

Another benefit of ERLIC enrichment was that the internal peptides and C-terminal peptides were preserved, thus allowing their analysis by LC-MS/MS. The obtained information can serve as cross-validation of the N-termini detected from the SPITC fraction. The correlation between fraction 1 and fraction 2 helps address the “one-hit wonder” problem in positional proteomics. Examples are provided in subsequent sections below.

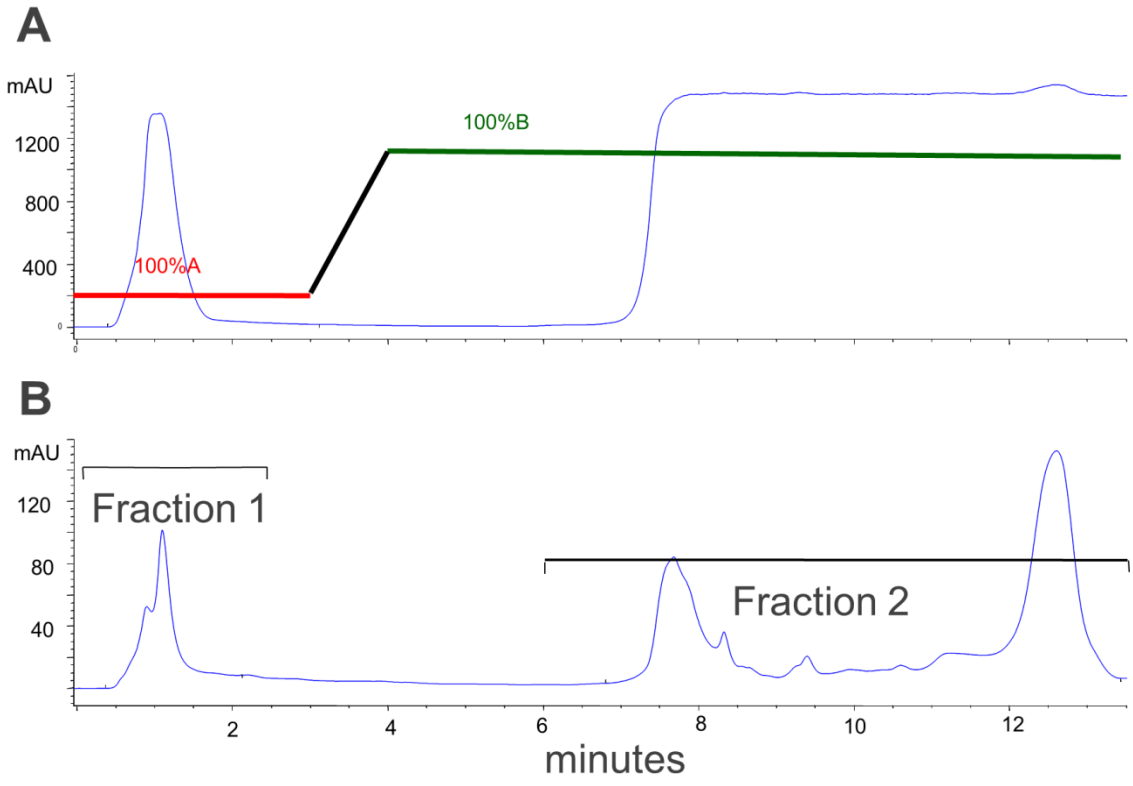


Figure 3.4 Fraction collections for trypsin digested *E.coli* cell lysate using electrostatic repulsion hydrophilic interaction (ERLIC) chromatography; (A) UV detection at 215 nm and (B) UV detection at 280 nm.

3.4.5 Tandem mass spectra comparison between CID and HCD

To assess the differences in fragmentation patterns between CID and HCD, spectra from multiply charged ions of SPITC modified peptides were subjected to CID using an LTQ OrbiTrap XL and to HCD using a Q Exactive. Dissociation spectra of the BSA N-terminal peptide acquired by CID and HCD are shown in Figure 3. Upon CID dissociation, several y ions were generated as marked in Figure 3.5a. In comparison, HCD dissociation yielded a much higher quality

tandem mass spectrum (Figure 3.5b), composed of a complete set of y fragment ions. The only significant N-terminal fragment ion observed was b_1 + SPTIC, which pinpoints the SPITC modification to the N-terminus. This process is proposed to be promoted by the nucleophilic sulfur atom of the SPITC moiety attacking the carbonyl oxygen atom of the adjacent amino acid, resulting in an Edman-type degradation.[25,26] The lack of other N-terminal ions illustrate that SPTIC derivatization at the peptide N-termini can effectively neutralize the positive charge of the N-terminal ions (a and b ions) upon activation, thus yielding spectra rich in y ions for peptide sequencing. Besides improving spectral quality, the generation of high mass accuracy fragment ions is another positive attribute of HCD on the Q Exactive Orbitrap. This improved mass accuracy of fragment ions improves confidence in peptide filtering via the semi-machine learning Percolator algorithm.

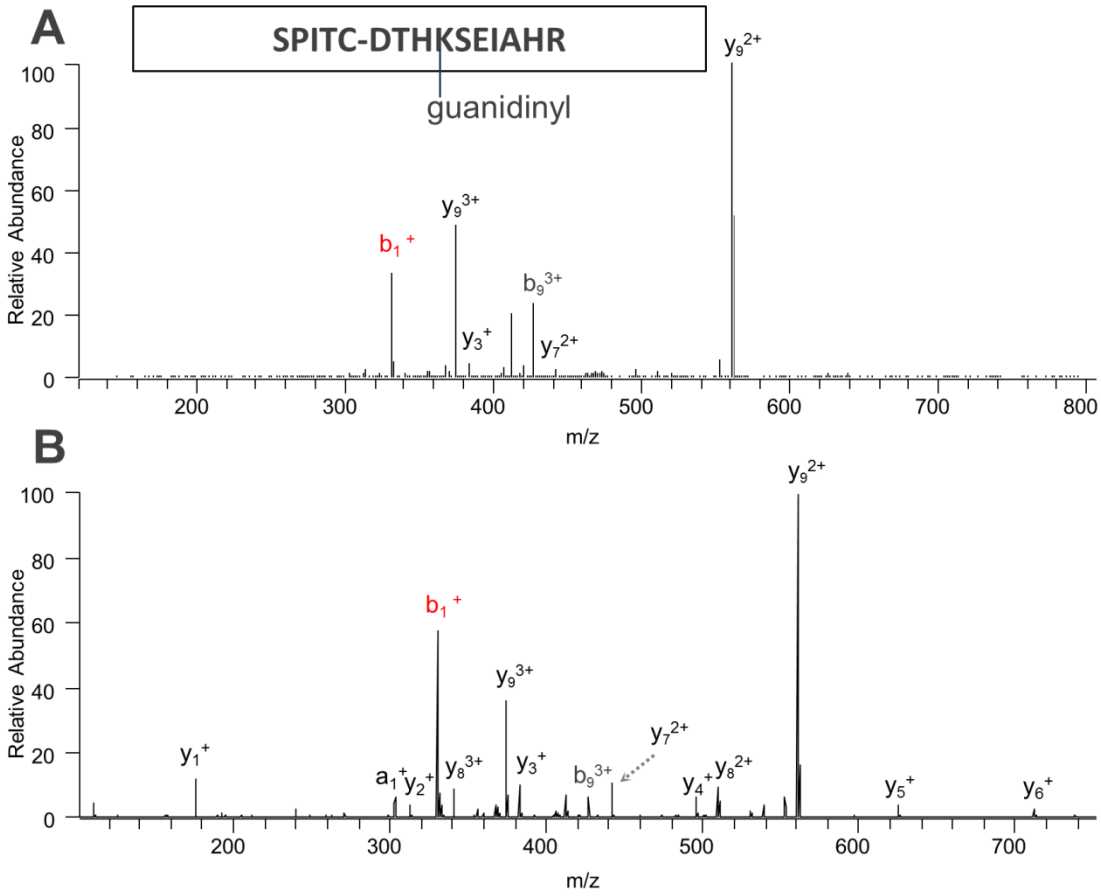


Figure 3.5 MS/MS spectra for BSA peptide SPITC-DTHK(guandinyl)SEIAHR m/z= 484.20^{3+} by (A) CID and (B) HCD. SPITC is attached to N-terminal ions labeled in red.

3.4.6 N-termini analysis of E. coli cell lysate

The workflow as described above was applied to an *E. coli* cell lysate. Sample preparation followed by LC-MS/MS analysis was performed in triplicate and the results reported below represent pooled data acquired from the three runs.

A total of 358 N-terminal SPITC-labeled high confidence peptides were identified in the bound fraction with a false discovery rate (FDR) less than 0.01 (assessed by Percolator). These peptides represented 274 proteins annotated in the *E. coli* database. We found that the presence and location of guanidinyl lysine (homoarginine) played an important role in terms of the appearance of MS/MS spectra, with higher quality MS/MS spectra obtained for peptides without guanidinyl lysine. For those N-terminal peptides without guanidinyl lysine, the MS/MS spectra were dramatically simplified and were composed of predominately y-type ion series, as shown in Figure 3.6. There were 315 cases belonging to the category of those 358 high confidence N-terminal SPITC labeled peptides. The remaining 43 peptides contained guanidinyl lysine, and the peptide fragmentation profile changed case-by-case as shown in Figure 3.7. The spectra contained both y and b ions for sequence VVINK(guanidinyl) and DTTTIIDGVGEEAAIQGR as labeled in Figure 3.7a, the possible explanation might be that homoarginine located towards the N-terminal side of peptide may aid in the survival of b ions. Whereas the spectra contained predominately the y-ion series for peptide sequence LDMLNEELSDK (guanidinyl)ER .

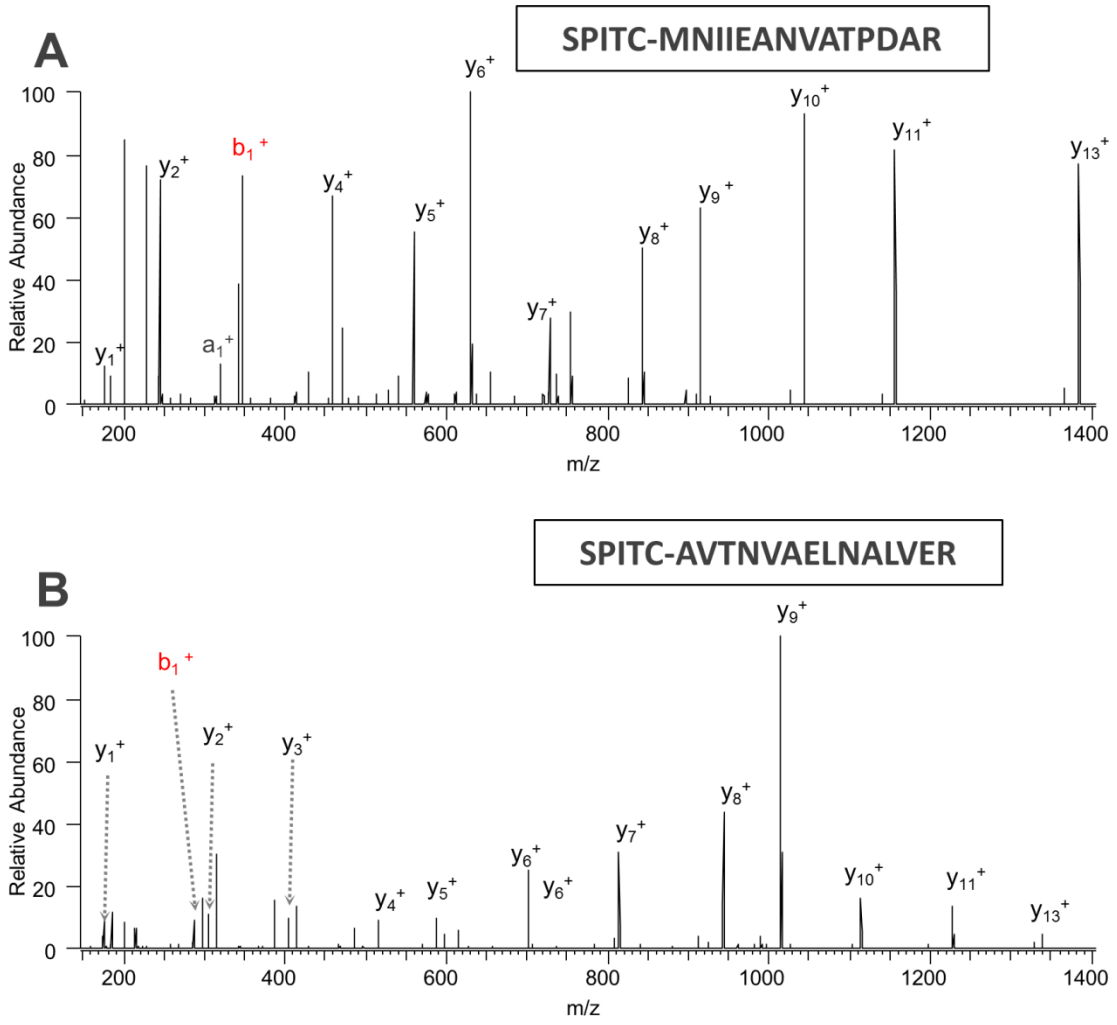


Figure 3.6 MS/MS spectra for N-terminal peptides without guanidinyl lysine; (A) sequence MNIIEANVATPDAR, M1-SPITC protein ID P61714 (6,7-dimethyl-8-ribityllumazine synthase OS=*Escherichia coli*) and (B) sequence AVTNVAELNALVER, A2-SPITC protein ID P0A9Q7 (Aldehyde-alcohol dehydrogenase OS=*Escherichia coli*). SPITC is attached to N-terminal ions labeled in red.

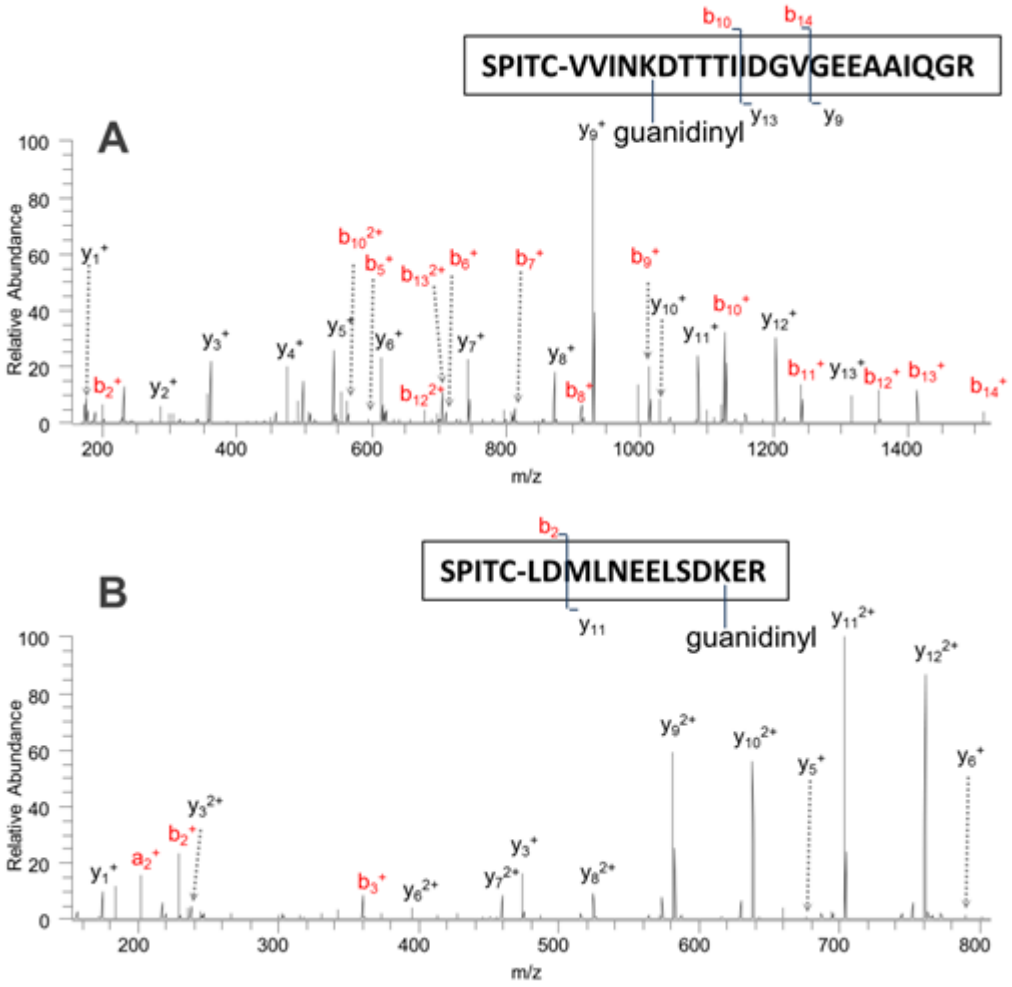


Figure 3.7 MS/MS spectra of N-terminal peptide containing guanidinyl lysine; (A) VVINK(guanidinyl) DTTTIIDGVGEEAAIQGR protein ID P0A6F5 (60 kDa chaperonin OS=*Escherichia coli*) and (B) LDMLNEELSDK (guanidinyl)ER protein ID P63284(Chaperone protein ClpB OS=*Escherichia coli*). SPITC is attached to N-terminal ions labeled in red.

There were occasions, however, when two sequences were identified for the same peptide species, the difference being that one had a guanidinyl lysine and the other had an unmodified lysine (see Figure 3.8). Both peptide species were

modified at the N-termini by SPITC. The y-ion series in both spectra showed similar patterns and intensities up to the y_{11} ion (just after lysine). For the guanidinyl modified peptide, no further y ions were detected, while for the peptide with no lysine modification, y_{12}, y_{13}, y_{14} and y_{15} were also present in the MS/MS spectra. The assignments of y_{12} to y_{15} were valuable since this clarified that the SPITC modification occurred at the N-termini, not at the lysine, whose side-chain failed to be modified by the guanidination reaction.

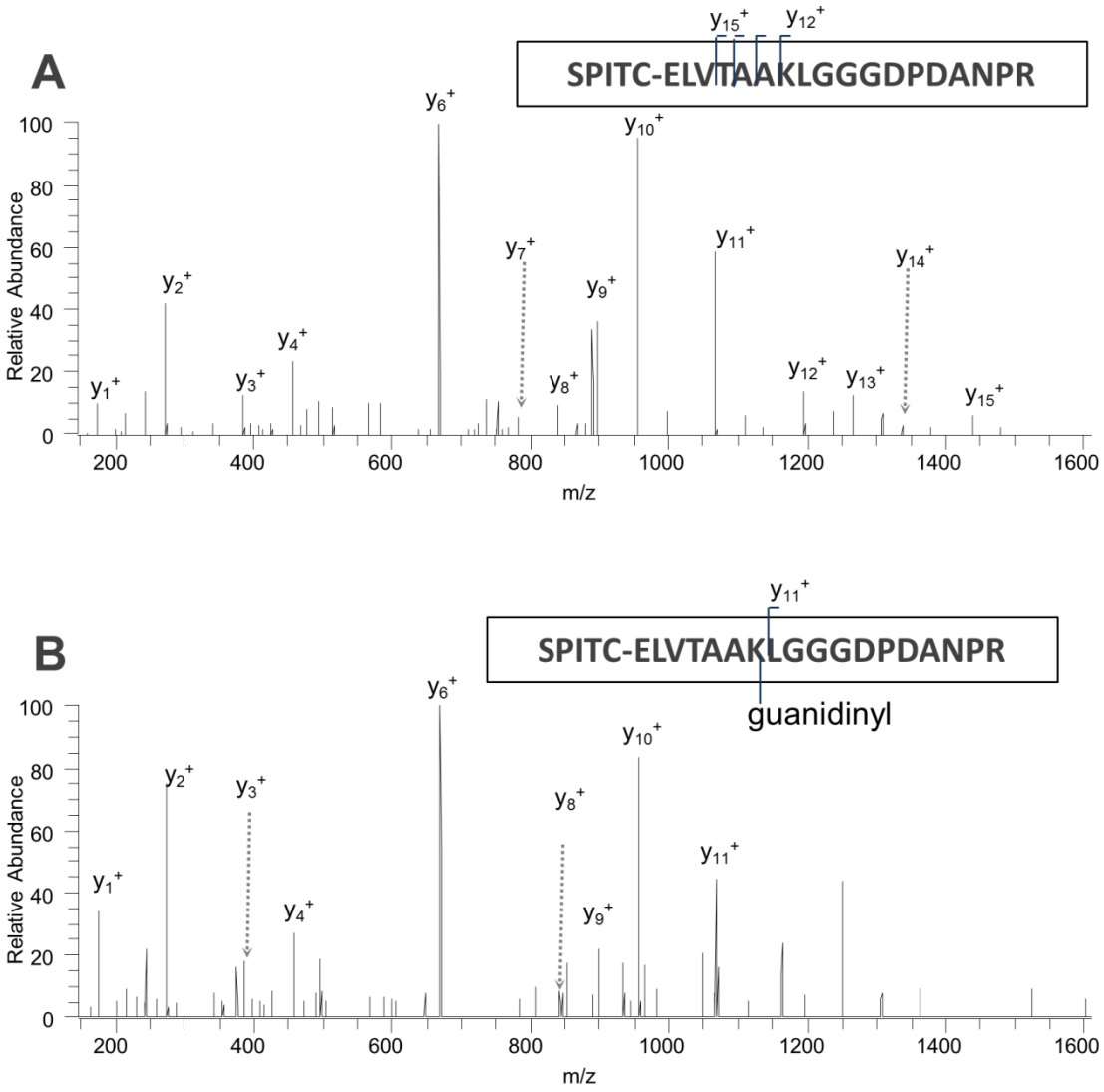


Figure 3.8 MS/MS spectra for SPITC peptide ELVTAAKLGGGDPDANPR (protein ID P0A8A0; (A) containing unmodified lysine and (B) containing guanidinyl Lysine.

A detailed analysis of the data showed that a single protein may present multiple N-termini. This may sound counter-intuitive, but cellular systems are complex with the coexistence of various enzymes, which can decompose proteins. One example is given for protein P0A850 (Trigger factor OS=Escherichia coli), in which three N-termini were observed with high confidence by our approach. The three peptides were P0A850from1to13, P0A850from244to255 and P0A850from322to334. The annotated peptide (1-13) was in low abundance compared to the other two peptides, judging by PSMs (a higher number correlates with higher abundance; PSMs for AA1-13 is 1, while PSMs for the other peptides are 5). In addition, the assignment of N-termini can be assured by the quality of each spectrum. The spectral simplification effect of the SPITC modification was present in all three N-termini of P0A850 as depicted in Figure 3.9. Therefore, we conclude that all these three peptides are N-terminal peptides derived from protein P0A850.

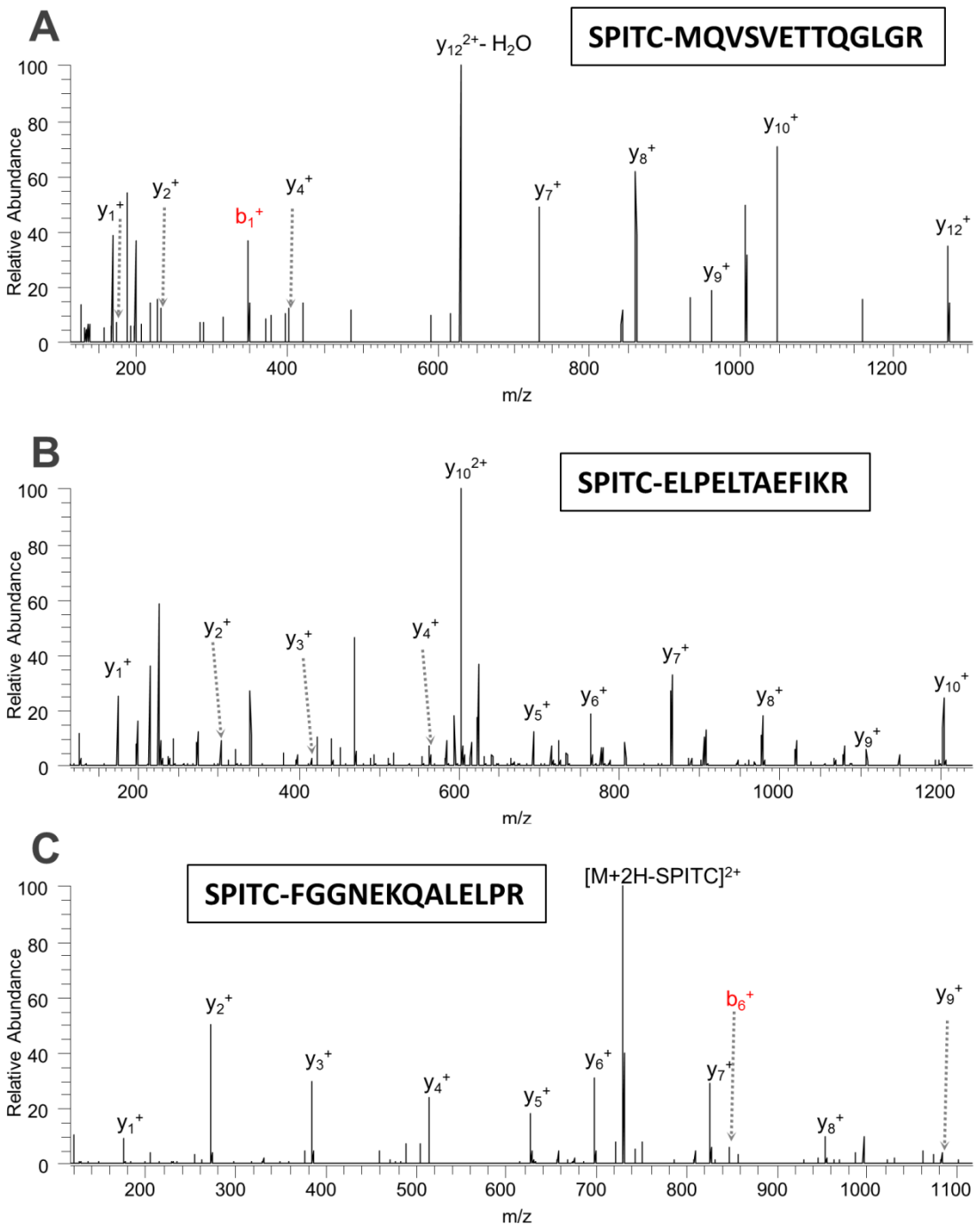


Figure 3.9 MS/MS spectra for N-terminal peptides of protein P0A850 (Trigger factor OS=Escherichia coli); (A) P0A850from1to13, (B) P0A850from244to255, and (C) P0A850from322to334. SPITC is attached to N-terminal ions labeled in red.

3.4.7 Comparison of bound and flow-through fractions of E. coli cell lysate

As we described previously, the chemically-modified protein digest was fractionated into two parts. One fraction was unbound with ERLIC and the other was bound with ERLIC (N-terminal enriched). The unbound fraction was also collected and the identified proteins were used to examine the crossover with the identified N-terminal peptides. Three sample preparations followed by LC-MS/MS analyses were performed for the *E. coli* cell lysate using the described workflow. The findings reported below represent pooled results of peptide identification for these data.

Table 3.1 presents counts and percentages of peptides (and corresponding PSMs) determined to be labeled with SPITC (on N-termini and lysines) found in both the bound and flow-through fractions. Consistent with the observations reported above on the basis of UV absorbance, the overall intensity of the signal is ~3-4x higher in the flow-through fraction as compared to the one bound to ERLIC. At the same time relative enrichment for N-terminal SPITC tagged peptides is about 10x higher in the ERLIC bound fraction as compared to that in the flow-through (e.g. 14% of peptides in bound fraction were identified as SPITC labeled at N-termini as compared to ~1% in the flow-through fraction). Similar

observations can be made on the basis of PSM counts, which are approximate indicators of peptide abundance in the mixture – those determined to be labeled with SPITC at the N-terminal account for about 10% of the total PSM count in the bound fraction versus ~1% in the flow-through fraction.

Table 3.1 Counts and percentages of peptides and corresponding PSMs annotated as labeled with SPITC (on N-termini and lysines) and not in bound and flow-through fractions of *E. coli* cell lysate.

	Peptides:			PSMs:		
	Lysines	N-Term	Other	Lysines	N-Term	Other
Bound	301 (11%)	358 (14%)	1965 (75%)	949 (12%)	918 (11%)	6358 (77%)
Flowthrough	124 (2%)	108 (1%)	7270 (97%)	216 (1%)	164 (1%)	23462 (98%)

The pie chart for calculated pI of identified peptides for bound and unbound fractions are shown in Figure 3.10. The bound fraction (A) contains mostly acidic peptides; since 71% of the peptide population is composed of peptides with pI lower than 4.99, with 10% contributed by pI lower than 3.99 and 61% by pI between 4-4.99. The unbound fraction (B) shows 38% of peptides with pI lower than 4.99, with only 1% from peptides with pI lower than 3.99. This agrees well with the ERLIC theory that acidic peptides have weaker electrostatic repulsion relative to the basic peptides, and they therefore exhibit higher retention on the column.

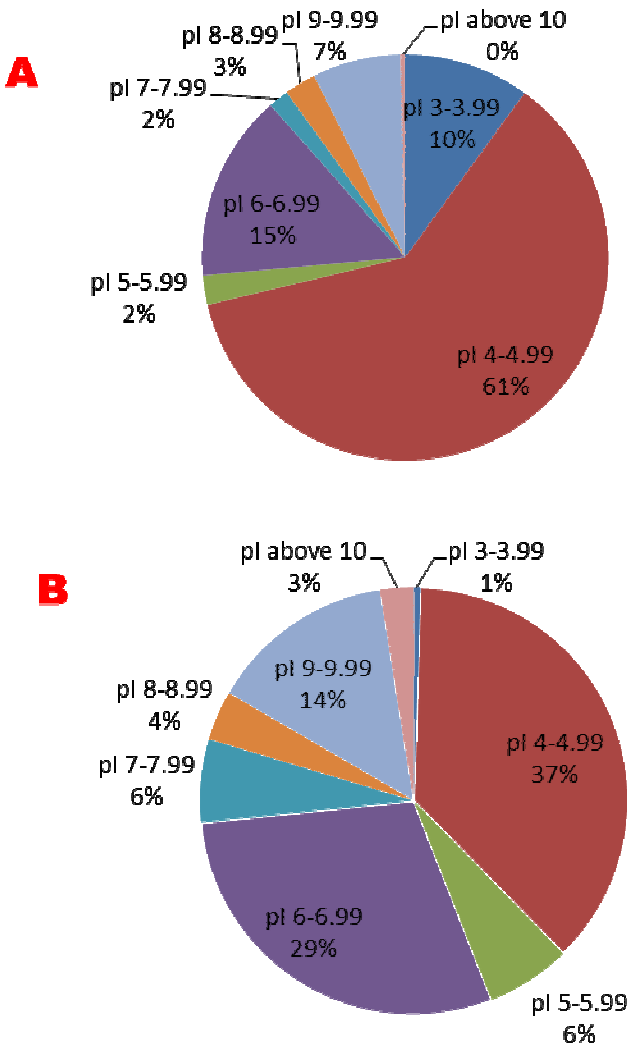


Figure 3.10 pie chart for calculated pI of identified peptides for (A) bound fraction and (B) unbound fraction

Figure 3.11 shows the overlaps between identified proteins detected in the flow-through and bound fractions. The majority of the proteins detected in the bound fraction irrespective of their SPITC labeling status were also detected in the flow-through fraction. Similarly, the majority (~80%) of the proteins identified as labeled with SPITC on the N-termini were also detected in the flow-through

fractions with peptides lacking the SPITC label. Proteins identified both in flow-through fractions and by N-terminal SPITC labeled peptides in the bound fraction tend to have higher PSM counts indicative of their higher abundances as compared to those that were present only in one of these two sets (Figure 3.12). Comparison of the peptides identified in the flow-through and bound fractions provides an additional level of support for the reliability of detection of the enriched N-terminal SPITC labeled peptides in the bound fraction.

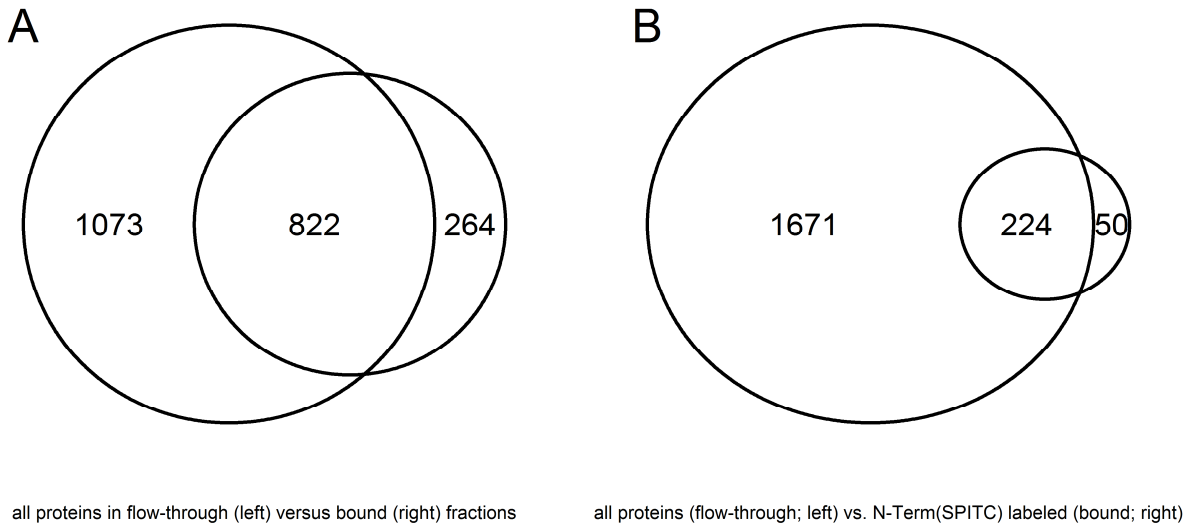


Figure 3.11 Overlaps between protein accessions identified in flow-through and bound fractions; (A) all proteins identified in each of the fractions and (B) all proteins identified in flow-through fractions versus those detected in bound

fraction as labeled with SPITC on N-termini. Areas are approximately proportional to the counts shown in the figures.

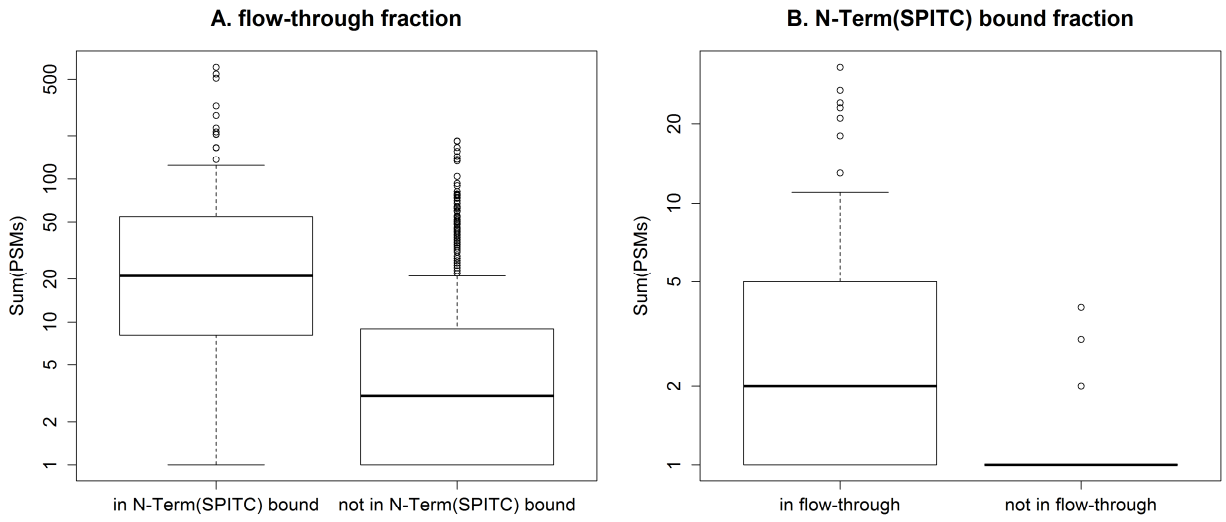


Figure 3.7 Sums of PSMs for all proteins (A) in flow-through fraction and (B) those in bound fraction that were identified as N-terminal SPITC labeled. In each case, proteins were split into two groups those that were also present in the other set (N-terminal SPITC in bound fraction or all proteins in flow-through fraction respectively) or only present in the given set. Proteins present in both sets tend to have higher PSM sums, suggesting their higher abundance.

3.5 CONCLUSIONS

The method presented herein shows many improvements over existing positional proteomics approaches. The use of SPITC modification at the N-termini enables the generation of high quality MS/MS spectra for high fidelity assignments, and enrichment due to the addition of the negatively-charged sulfate group. The ERLIC separation was advantageous compared to free amine scavenging approaches, due to the lower losses and the isolation of internal and C-terminal peptides for cross-validation of N-terminal assignments.

3.6 REFERENCES

1. Bertone, P.; Stolc, V.; Royce, T. E.; Rozowsky, J.S.; Urban, A.E.; Zhu, X.; Rinn, J. L.; Tongprasit, W.; Samanta, M.; Weissman, S.; Gerstein, M.; Snyder, M. Global identification of human transcribed sequences with genome tiling arrays. Science 2004, 306 (5705), 2242-2246
2. Emanuelsson, O.; Nagalakshmi, U.; Zheng, D.; Rozowsky, J.S.; Urban, A.E.; Du, J.; Lian, Z. ; Stolc, V.; Wissman, S.; Synder, M.; Gerstein, M.; Assessing the performance of different high density tiling microarray strategies for mapping transcribed regions of the human genome. Genome Res. 2006
3. Kuroyanagi, H.; Kobayashi, T.; Mitani, S.; Hagiwara, M. Transgenic alternative-splicing reporters revel tissue-specific expression profiles and regulation mechanisms in vivo. Nat. Methods 2006, 3(11), 909-915
4. Ducet, A.; Butler, G.S.; Rodriguez, D.; Rudova, A.; Overall, C.M.; Metadegradomics: tovard in vivo quantitative degradomics of proteolytic post-translational modifications of the cancer proteome. Mol. Cell. Proteomics 2008, 7, 1925-1951

5. Quesada, V.; Ordonez, G.R.; Sanchez, L.M.; Puente, X.S.; Lopez-Otin, C.; The Degradome database: mammalian proteases and diseases of proteolysis. Nucleic Acids Res. 2009, 37, D239-D243
6. Opferman, J.T.; Korsmeyer, S.J. Apoptosis in the development and maintenance of the immune system. Nat. Immunol. 2003, 4(5), 410-415
7. Rao, J. S. Molecular mechanisms of glioma invasiveness: the role of proteases. Nat. Rev. Cancer 2003, 3(7), 489-501
8. Van Damme, P.; Martens, L.; Van Damme, J.; Hugelier, K.; Caspase-specific and non-specific in vivo protein processing during Fas-induced apoptosis. Nat. Methods 2005, 2, 771-777
9. Gausdal, G.; Gjertsen, B.T.; McCormack, E.; Van Damme, P. et al.; Abolition of stress induced protein synthesis sensitizes leukemia cells to anthracycline –induced death. Blood 2008, 111, 2866-2877
10. Impens, F.; Van Damme, P.; Demol, H.; Van Damme, J.; et al.; Mechanistic insight into taxol-induced cell death. Oncogene 2008, 27, 4580-4591
11. McDonald L.; Beynon R.J.; positional proteomics: preparation of amino-terminal peptides as a strategy for proteome simplification and characterization 2006, 4, 1790-1798
12. Kleifeld O.; Doucet, A.; Keller, U.; Prudova, A.; Schilling, O.; Kainthan, R.; Starr, A.; Foster, L.J.; Kizhakkedathu, J.; Overall, C.M.; Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products 2010, Nature Biotech. 28(3), 281-288
13. Veenstra, T.D.; Conrads, T.P.; Issaq, H.J.; what to do with one hit wonders? Electrophoresis 2004, 25, 1278-1279
14. Dormeyer, W.; Mohammed, S.; Breukelen, B.; Krijgsveld, J.; Heck, A. J.R. Targeted Analysis of protein termini 2007 6, 4634-4645
15. Martens, L.; Vandekerchkhove, J.; Gevaert, K.; DBToolkit: processing protein databases for peptide-centric proteomics. Bioinformatics 2005, 21, 3584-3585
16. Keller, A.; Nesvizhskii, A.I.; Kolker, El; Aebersold , R.; Empirical statistical model to estimate the accuracy of peptide idnetifications made by MS/MS and database search. Anal. Chem. 2002, 74, 5383-5392

17. Helsens, K.; timmerman, E.; Vandekerchhove, J.; Gevaert, K.; Martens, L.; peptizer, a tool for assessing false positive peptide identifications and manually validating selected results. Mol. Cell. Proteomics 2008, 7, 2364-2372
18. Beardsley, R and Reilly, J. P; Optimization of guanidination procedures for MALDI mass mapping. Anal.Chem 2002, 74, 1884-1890
19. Wang, D; Kalb S. R. and Cotter, R.J.; Improved procedure for N-terminal sulfonatiion of peptides for matrix-assisted laser desorption/ionization post-source decay peptide sequencing, Rapid commun. Mass Spectrom. 2004, 18, 96-102
20. Alpert, A.; Mitulovic G. and Mechtler, K.; Isolation of tryptic phosphopeptides by ERLIC (electrostatic repulsion-hydrophilic Interaction chromatography). HPLC 2008 poster P-2412-W
21. Mommen, G.P.M; Waterbeemd, B.V.D.; Meiring H.D.; Kersten, G.; Heck, A.J.R and Jong, P.J.M; unbiased selective isolation of protein N-terminal peptides from complex proteome samples using phosphor tagging (PTAG) and TiO2-based depletion. MCP, published on June 22, 2012
22. Horvatic, A; Dodig, I; Vuletic, T; Pavokovic, D; Hamersak, Z; Butorac, A and Cindric, M.; Comparison between enhanced MALDI in-source decay by ammonium persulfate and N- or C-terminal derivatization methods for detailed peptide structure determination. Anal. Chem, 2013, 85(8), 3940-3947
23. Bartel, S; Doellinger, J; Darsow, K; Bourquain, D; Buchholz, R; Nitsche, A and Lange, H.A.; proteome analysis of Vaccinia Virus IHD-W-infected HEK 293 cells with 2-*dimensional gel electrophoresis and MALDI-PSD-TOF MS of on solid phase support N-terminally sulfonated peptides. Virology Journal 2011, 8, 380
24. Madsen, J. and Brodbelt, J.; Simplifying fragmentation patterns of multiple charged peptides by N-terminal derivatization and electron transfer collision activated dissociation. Anal. Chem. 2009, 81(9), 3645-3653
25. Lee, Y.H.; Kim, M.S.; Min, H.K.; Lee, S.W.; Highly informative proteome analysis by combining improved N-terminal sulfonation for de novo peptide sequencing and online capillary reverse-phase liquid chromatography/tandem mass spectrometry. Proteomics 2004, 4, 1684-1694

26. Summerfield, S.G.; Bolgar, M.S. and Gaskell S.J.; Promotion and stabilization of b1 ions in peptide phenylthiocarbamoyl derivatives: Analogies with condensed-phase chemistry. Journal of mass spectrometry 1997, 32, 225-231

Chapter 4 iTRAQ Labeling of N-terminal Amines in Complex Samples and Its Application in Protease Substrate Degradomics

4.1 Abstract

A positional proteomics strategy for proteolytic cleavage sites is presented based on iTRAQ labeling at the N-termini of peptides followed by enrichment on a NHS-activated agarose spin column. After substrates are treated with protease, a guanidination reaction is used to block the primary amine of the lysine side chain; subsequent reaction with iTRAQ reagents labels only the primary amine of the N-termini. A control sample is prepared following the same steps, except that there is no protease treatment. All iTRAQ labeled samples are then pooled and treated with trypsin. The newly trypsin-digested sample is a mixture of peptides of various characteristics: N-terminal peptides (N-termini capped with iTRAQ reagent), internal peptides (free N-termini and Arg as C-termini) and C-terminal peptides (free N-termini). The latter two portions, which contain primary amines at the N-termini, can be scavenged by NHS-activated agarose beads, resulting in enrichment of iTRAQ labeled peptides. Samples prepared in this manner are subjected to LC-MS/MS analysis and database searching, thus, the substrate of the protease can be revealed. Since the N-termini of protease cleavage are preserved by iTRAQ labeling, the proteomic scale analysis of N-termini will disclose the

cleavage sites of protease. In this report, Asp-N digested *E.coli* cell lysate is taken as a model system. A total of 764 N-terminal peptides, corresponding to 377 proteins, are identified in one experiment, with 91% of those proteolytic products matching the expected protease digestion specificity. Our results suggest that iTRAQ in combination with N-terminal proteomics is useful for the identification of the proteolytic cleavage sites in complex systems as well as for the establishment of an *in vivo* proteolytic signature profile.

4.2 Introduction

The majority of proteomics studies appearing in the recent literature employ a "shotgun" approach,[1-3] where the proteins are first digested into peptides, typically using trypsin as protease, that are subsequently analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Proteome systems typically encompass thousands of individual components present in different concentrations varying over several orders of magnitude.[4] This poses a major challenge to achieving high proteome coverage even when employing highly advanced instrumentation and workflows. Positional proteomics[5-7] has drawn attention to the concept that a single peptide with a well-defined position (typically N-terminal or C-terminal) can be used to characterize the entire protein. Using this

principle, the sample complexity can be drastically decreased and the recovery, especially for low abundant proteins, is expected to increase. N-terminal positional proteomics has found value in another subfield of proteomics called degradomics[8-9] that focuses on identifying proteases and their corresponding substrates. In the human genome, 570 genes are predicted to encode for proteases. The substrates for half of those proteases are unknown, while processing of known targets of the other half is not well characterized.[10] To fully understand the physiological and pathological implications of proteases, identifying their substrates and cleavage sites is a first step forward in this unknown territory. N-terminal positional proteomics has been shown to be a valuable tool in this field.[11-13]

The essence of positional proteomics lies in the ability to differentiate selected positional peptides from other redundant peptides via the use of chemical or enzymatic modification. The relatively high reactivity of the N-terminal primary amine compared to the less reactive C-terminal carboxylic acid is a major reason that the numbers of N-terminal proteomics reports far exceed those of the latter in this subfield of proteomics. Due to the robustness and simplicity of reaction, as well as the availability of isotopic forms of reagents, the most widely used chemical modifications in N-terminal proteomics are dimethylation[14-15] and acetylation[16-17]. In dimethylation, catalyzed by sodium cyanoborohydride, formaldehyde is used to simultaneously label all free amino groups, i.e., the

primary amine of lysine, as well as those of each N-terminus. In-solution, online and on-column protocols for stable isotope dimethyl labeling were described in detail by Boersema *et al*[15] for sample amounts ranging from sub-micrograms up to milligrams. An example of acetylation was described by McDonald *et al*[16] employing sulfo-NHS acetate. Similar to the dimethylation reaction, acetylation blocks all primary amines in peptides and proteins.

Isobaric tagging for relative and absolute quantitation (iTRAQ) labeling is a useful tool in protein biomarker expression analysis.[18-20] The reagent consists of a reporter group, a balance group and an amino reactive group, enabling isotopic arrangement that permits 4-plex or 8-plex reagents. The unique feature of the iTRAQ reagent is that peptides labeled with different reagents have the same nominal mass, while upon MS/MS, each individual label produces an exclusive reporter ion, which permits relative quantitation of the peptide(s) in each sample. Changes in biomarker levels are observable in the peptide reporter ions generated by MS/MS, thus allowing one to pinpoint the relevant proteins.

A strategy using iTRAQ as the labeling reagent in N-terminal proteomics is explored in this study. iTRAQ is used for two purposes; firstly, to serve as tagging reagent to differentiate N-terminal peptides from internal peptides, and secondly, to provide direct verification for N-terminal peptides when duplicate samples are tagged with different iTRAQ labels.

4.3 Experiment

4.3.1 Materials

E. coli cell lysate was obtained from McLab (San Francisco, CA in USA). Guanidination kit, dithiothreitol (DTT) and iodoacetamide (IAM) were purchased from Sigma Aldrich (St. Louis, MO in USA). iTRAQ 4-plex was obtained from Applied Biosystems (Framingham, MA in USA). Trypsin Gold (mass spectrometry grade) was obtained from Promega (Madison, WI in USA). Endoproteinase Asp-N (sequencing grade) was purchased from Roche (Indianapolis, IL in USA). Zeba spin desalting column (7K MWCO) and the NHS-activated agarose spin column were obtained from Pierce Biotechnology (Rockford, IL in USA).

4.3.2 Reduction and alkylation

2 mg of *E. coli* cell lysate was dissolved with 200 μL of Urea (6 M, with 0.1 M phosphate buffer, pH 7.5). The solution was incubated at 37 °C for 30 min. A 40 μL aliquot of DTT (0.5M) was added and the mixture was incubated at 37 °C for 30 min. A 100 μL aliquot of IAM (0.5 M) was added and the sample was incubated at room temperature in the dark for 30 min. Another 10 μL aliquot of DTT was added to quench the alkylation reaction.

4.3.3 Guanidination

The guanidination kit was obtained from Sigma-Aldrich. The concentration of *o*-methylisourea hemisulfate was prepared at 6 mg/mL in water. The protein sample was adjusted to pH 10 by NaOH (1N) after reduction and alkylation and 100-fold molar excess of *o*-methylisourea hemisulfate was added. The mixture was then incubated at 65 °C for 30 min.

A 50 μL aliquot of urea (6M at pH 7.5) was added to make a 5 mg/mL protein solution. A Zeba spin desalting column was first buffer exchanged with 100 mM sodium phosphate buffer (pH 7.5) (300 μL of buffer each time, spin down for 1 min at 1000 g, 4 times in total for each spin column). Two separate Zeba spin columns were used, with each loaded with 120 μL of protein sample. The flow-through of one sample was used for Asp-N digestion, while the flow-through of the other was labeled with iTRAQ directly as a control.

4.3.4 AspN-digestion

6 μg of endoproteinase Asp-N was dissolved in 120 μL of water. The enzyme was mixed with 60 μg of cell lysate (after guanidination and Zeba spin desalting clean-up), and the mixture was incubated at 37 °C for 18 h. The sample was freeze-dried prior to iTRAQ labeling.

4.3.5 iTRAQ labeling

Two 25 μg control portions of cell lysate (after guanidination and Zeba spin desalting clean-up) were each diluted with 20 μL of urea (6M, with 0.1 M phosphate buffer, pH 8.5) in separate vials. Each cell lysate vial was mixed with the iTRAQ reagent dissolved in 70 μL ethanol; one vial was reacted with iTRAQ tag 114 and the other with iTRAQ tag 115. The mixture was incubated for 2 hr at room temperature.

Two replicates of 25 μg of AspN digested cell lysate were each diluted with 20 μL of phosphate buffer (pH 8.5) in separate vials. Each digested sample vial was mixed with the iTRAQ reagent dissolved in 70 μL ethanol; one vial was reacted with iTRAQ tag 116 and the other with iTRAQ tag 117. The mixture was incubated for 2 hr at room temperature.

All the iTRAQ labeled samples were pooled into a single vial and dried by centrifuging. The sample was re-suspended in 196 μL of sodium bicarbonate (0.1 M).

4.3.6 Trypsin digestion

4 μg of trypsin (1 μg/μL in water) was added to the pooled iTRAQ solution. The solution was incubated at 37 °C overnight.

4.3.7 N-terminal enrichment by NHS-activated agarose resin

PBS buffer (200 μL, 0.1 M sodium phosphate and 0.3 M sodium chloride, pH 7.2) was added to the trypsin-digested solution, which was directly added to NHS-activated agarose dry resin. After mixing the sample end over end for 1 hr, the spin column was centrifuged at 1000 g for 1 min and the flow through was collected; this flow-through contained the N-termini enriched portion. The resin was further washed twice with 400 μL of PBS buffer, with collection of the flow through. These three flow-through portions were combined and desalted by C18 cartridge. The salt was washed out with 0.1% formic acid in water and the N-terminal peptides were eluted with 0.1% formic acid in 80% acetonitrile in water. After drying the solvent by centrifuge, the N-terminal peptides were re-dissolved with 40 μL of water and subjected to immediate LC-MS/MS analysis or stored at −20 °C for further analysis.

4.3.8 LC-MS/MS

LC-MS/MS was performed on a capillary Dionex U3000 HPLC system (Thermo Scientific, San Jose, CA) coupled to a QExactive Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA). A Targa C18 column from Higgins Analytics (150 cm* 300 μm, 3 μm particle) was used with mobile phase A (0.1% formic acid in water) and mobile phase B (0.1% formic acid in acetonitrile). Peptides were eluted over 90 min by increasing mobile phase B from 0 to 30% linearly. Mass spectrometry data were acquired automatically using Xcalibur

software (Thermo Scientific) in data dependent acquisitions, with both survey scans and MS2 scans in the FT mode (resolution set at 35000 for parent ions and 17500 for fragment ions). Each survey scan was followed by 12 MS/MS scans. MS/MS spectra were acquired in HCD mode with normalized collision energy (NCE) of 25.

Peptides were identified by the Thermo Proteome Discovery 1.3 software package (Thermo Scientific). Peptide and protein identifications were obtained using Sequest and Mascot algorithms. The *E. coli* K12 strain database prepared in a fashion similar to that described by Dormeyer *et al*[23] was used for searching. Specifically, all protein sequences in the database were cut at arginine (C-terminal); for each resulting peptide (i.e., the series of neo-N-terminal peptides created as a result of these cleavages), multiple sequence database entries were generated that represented sequential truncations of each N-terminal amino acids. Only sequences 6 amino acids or longer were retained in the database. The mass tolerance for parent ions was set to be 10 ppm and 0.1 Dalton for fragment ions. Trypsin cleavage was applied with up to 1 missed cleavage in the search. Carboamidomethyl of cysteine (+57.021Da) was set as a static modification. Dynamic modification included guanidination for lysine (+42.022 Da) and N-terminal modification with iTRAQ (+144.102 Da). Percolator (Thermo Scientific) was used for peptide filtering, with a peptide FDR threshold set to 0.01.

4.4 Results and Discussion

Enoksson *et al*[21] reported the identification of ten caspase-3 cleavage sites in a mixture of 7 purified recombinant *E. coli* proteins using iTRAQ, with all sites corresponding to caspase-3 previously reported cleavage sites. Prudova *et al.*[22] incorporated iTRAQ whole protein labeling with terminal amine isotopic labeling (iTRAQ-TAILS) to enrich the N-terminome by negative selection. By this approach, substrate degradomes of two closely related matrix metalloproteinase MM2 and MM9 were found to be significantly different. We have employed Asp-N digested *E. coli* cell lysate as a model system, where a protein database already exists and enzyme specificity is mostly known. Our results demonstrate that this combination of iTRAQ with N-Terminal proteomics can be used for the identification of the proteolytic cleavage sites in complex system and for the establishment of proteolytic signature profile *in vivo*.

4.4.1 Strategy for the identification of N-terminal proteolytic peptides

The strategy to identify N-terminal proteolytic peptides is illustrated in figure 4.1 using Asp-N as an example. The substrate was *E. coli* cell lysate. iTRAQ tags 114 and 115 were used to label the cell lysate sample, while iTRAQ tags 116 and 117 were used to label the Asp-N digested cell lysate sample. Prior to iTRAQ labeling, the side chain of lysine in both samples is blocked by the guanidination reaction. After pooling iTRAQ labeled samples, trypsin digestion

was performed, followed by N-terminal enrichment employing a NHS agarose spin column. LC-MS/MS data was then acquired. In the current workflow, peptides can be identified as belonging to the protein N-termini if all four tags are visible in the mass spectrum, whereas the peptides originating from Asp-N cleavage will only appear with two tags (116 and 117). Thus, peptides without the 116 and 117 labels most likely originate from internal peptides which survive the NHS-activated agarose enrichment process. Replicate samples with different tags allow simultaneous validation of identified peptides. Therefore, peptides with only one tag are ignored in the final peptide counting.

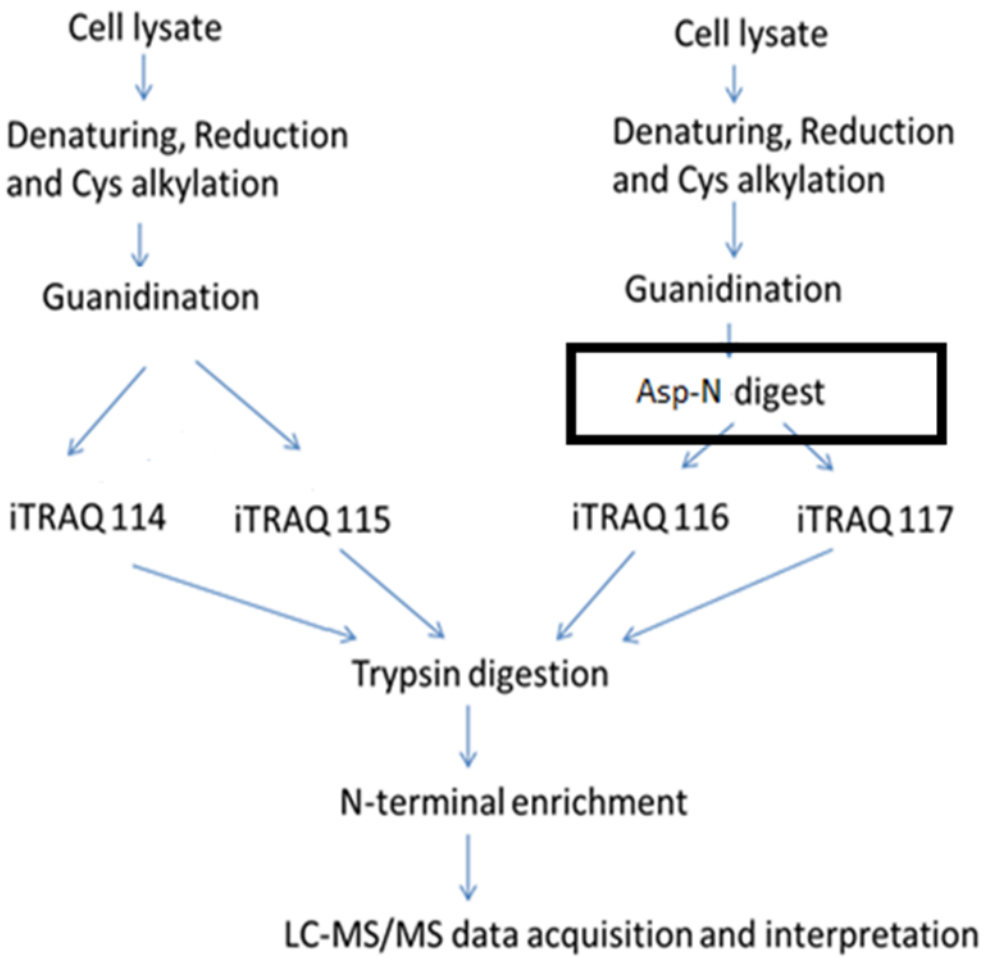


Figure 4.1 Scheme for N-terminal Proteolytic Peptides by iTRAQ labeling

4.4.2 E. coli cell lysate results using the proposed workflow

A total of 100 µg of *E. coli* cell lysate was used in the current experiment with one quarter reacted with each iTRAQ tag. The samples were prepared according to the workflow as in figure 4.1. The result table generated by Proteome Discovery 1.3 software (Thermo Scientific) was first filtered by protein accessions, then by modifications (containing N-terminal modified with iTRAQ) and lastly by

the confidence level of identification. A total of 500 N-terminal iTRAQ labeled peptides with *E. coli* protein accessions were identified with medium to high confidence. These peptides represented 370 proteins found in the *E. coli* database.

Endoproteinase Asp-N is known to cleave peptide bonds N-terminally at aspartic and cysteic acid.[24-26] Ingrosso *et al.* reported that Asp-N can also cleave at glutamyl residues.[27] Thus, the neo-N-terminal peptides produced by endoprotease digestion are expected to start predominately with Asp (D) and a lesser amount of Glu (E). Of those 764 N-termini iTRAQ peptides identified, 481 peptides had N-terminal Asp. We also found 215 peptides with N-terminal Glu. Altogether, peptides starting with Asp and Glu account for 91% of N-terminal peptides identified. There are 6 peptides that start with methionine, and 17 peptides are identified as annotated peptide sequences obtained by stripping off the first methionine, thus, annotated protein N-termini account for 3% of peptides identified. The origin of the other 6% is unclear.

An example of the annotated N-terminal peptide is presented in Figure 4.2a. This peptide sequence is identified with high confidence to be iTRAQ-MFPEYR by Proteome Discovery. With MFPEYR as the N-terminal peptide, the corresponding protein is determined to be P0ACW6 (Uncharacterized protein YdcH OS=*Escherichia coli*). The inset of Figure 2a shows the four reporter ions, further validating that this peptide is not from Asp-N digestion.

An example of an Asp-N digested neo-N-terminal peptide is shown in Figure 4.2b. This peptide sequence is identified to be iTRAQ-DVQVFTR through software assignment of both b and y series ions. This peptide is determined to be the 930 to 936 heptamer in protein P0A8V2 (DNA-directed RNA polymerase subunit beta OS=*Escherichia coli*). As shown in the inset in Figure 2b, only two reporter ions (m/z 116, 117) are present for this peptide, validating that this peptide is the product of Asp-N digestion.

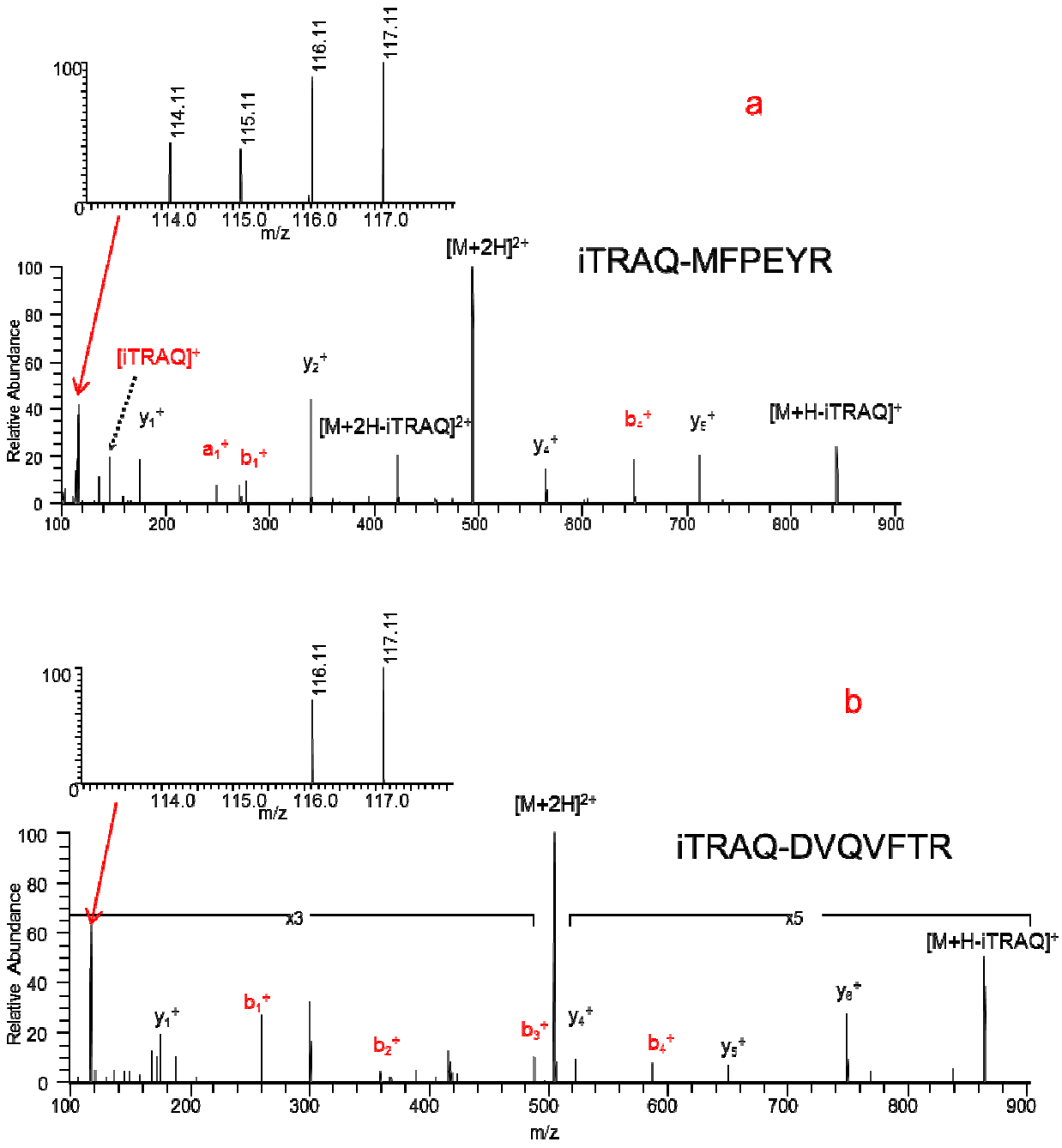


Figure 4.2 iTRAQ labeled N-terminal peptides (a) iTRAQ-MFPEYR from protein P0ACW6 (Uncharacterized protein YdcH OS=*Escherichia coli*), position 1-6 (b) iTRAQ-DVQVFTR from protein P0A8V2(DNA-directed RNA polymerase subunit beta OS=*Escherichia coli*), position 930-936.

4.5 Summary

An Asp-N digested *E. coli* cell lysate was employed as a model system to demonstrate the application of a novel workflow to study proteolytic cleavage sites and substrates. This workflow differentiates N-termini and neo-N-termini by iTRAQ, and enriches these species for improved sensitivity by negative selection using NHS agarose beads. A total of 764 N-terminal peptides were identified in one experiment, with 89% of those matching the protease digestion specificity. Those 764 peptides correspond to 377 proteins. Compared with dimethylation or acetylation modification, the obvious advantage of using iTRAQ as the labeling reagent is that 4 or 8 samples can be compared simultaneously with sensitivity improvement, which makes it an attractive approach for the establishment of *in vivo* proteolytic signatures.

4.6 References

1. Tracz D.M.; McCorrister S.J.; Chong P.M; Lee D.M.; Corbett C.R and Westmacott G.R.; A simple shotgun proteomics method for rapid bacterial identification. J. of microbiological methods 2013, 94(1), 54-7
2. Hammer E.; Darm K. and Volker U.; Characterization of the human myocardial proteome in dilated cardiomyopathy by label free quantitative shotgun proteomics of heart biopsies. Methods in molecular biology 2013, 1005, 67-76
3. Fanayan S.; Smith J.T.; Sethi M.L; Cantor D.; Goode R.; Simpson R.J; Bake M.S; Hancock W.S and Nice E.; J. of proteome research 2013, 12(1), 89-96
4. Teng P.; Bateman, N.W; Hood, B.L. and Conrads T.P.; J. of proteome research 2010, 9(12), 6091-6100

5. Mommen G.P.M; van de Waterbeemd B.; Meiring H.D.; Kersten G.; Heck. A.J.R. and de Jong A.P.J.M.; Unbiased selective isolation of protein N-terminal peptides from complex proteome samples using phosphor tagging (PTAG) and TiO2-based depletion. Molecular and cellular proteomics 2012, 11(9), 832-842
6. Van D.; Petra S.A.; Bronsoms S.; Helsens K; Colaert N.; Timmerman E.; Aviles F.X.; Vandekerchhove J. and Gevaert K.; Nature methods 2010, 7(7), 512-515
7. Schilling O.; Barre O.; Huesgen P.F. and Overall C.M.; Nature methods 2010, 7(7), 508-511
8. Doucet A.; Butler G.S.; Rodriguez D.; Prudova A.; Overall C.M.; Metadegradomics: towards in vivo quantitative degradomics of proteolytic post-translational modifications of the cancer proteome. Molecular and Cellular Proteomics 2008, 7(10) 1925-1951
9. Wejda M; Impens F; Takahashi N. Van Damme P.; Gevaert K. and Vandenabeele P.; J. of Biological Chemistry 2012, 287(41), 33983-33995
10. Lopez-otin, C. and Overall. C.M. Protease degradomics: a new challenge for proteomics. Nat. Rev. Mol. Cell Biol. 2002, 3, 509-519
11. Impens F.; Colaert N.; Helsens K.; Plasman K.; van Damme P.; Vandekerckhove J. and Gevaert K.; MS-driven protease substrate degradomics, Proteomics 2010, 10, 1284-1296
12. Huesgen P.F. and Overall C.M. N- and C-terminal degradomics: new approaches to reveal biological roles for plant proteases from substrate identification, physiologia plantarum, 2012, 145, 5-7
13. Impens F.; Vandekerckhove J. and Gevaert K.; who gets cut during cell death? Current opinion in cell biology. 2010, 22, 859-864
14. Kleifeld O.; Doucet A.; Keller U.; Prudova A.; Schilling O.; Kainthan R.K.; Starr A.E.; Foster L.J.; Kizhakkedathu J.N. and Overall C.M.; nature biotechnology 2010, 28(3), 281-288
15. Boersema P.J.; Raijmakers R.; Lemeer S.; Mohanned S. and Heck A.J.R.; Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nature protocols 2009, 4(4), 484-494

16. McDonald L. and Beynon R.J.; Positional proteomics: preparation of amino-terminal peptides as a strategy for proteome simplification and characterization. Nature protocols 2006, 1(4), 1790-1798
17. Staes A.; Van Damme P.; Helsens K.; Demol H.; Vandekerchhove J. and Gevaert K.; Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). Proteomics 2008, 8, 1362-1370
18. Martyniuk C.J.; Alvarez S. and Denslow N.D.; DIGE and iTRAQ as biomarker discovery tools in aquatic toxicology. Ecotoxicology and environmental safety 2012, 76, 3-10
19. Jing L.; Parker C.E.; Seo D.; Hines M.W.; Dicheva N.; Yu Y. Schwinn D.; Ginsburg G.S. and Chen X.; Discovery of biomarker candidates for coronary artery disease from an APOE-knock out mouse model using iTRAQ-based multiplex quantitative proteomics. Proteomics 2011, 11(14), 2763-2776
20. Tonack S.; Jenkinson C.; Cox T.; Elliott V.; Jenkins R.E.; Kitteringham N.R.; Greenhalf W.; Shaw V.; Michalski C.W. and Friess H.; iTRAQ reveals candidate pancreatic cancer serum biomarkers: influence of obstructive jaundice on their performance. British Journal of Cancer 2013, 108(9), 1846-1853
21. Enoksson M.; Li J.; Ivancic M.M.; Timmer J.C.; Wildfang E.; Eroshkin A.; Salvesen G.S. and Tao W.A.; Identification of proteolytic cleavage sites by quantitative proteomics. J. of proteome research 2007, 6, 2850-2858
22. Prudova A.; auf dem Kekker U.; Butler G.S. and Overall C.M.; Multiplex N-terminome analysis of MMP2 and MMP9 substrate degradomes by iTRAQ-TAILS quantitative proteomics. Molecular & Cellular Proteomics 2010, 9(5), 894-911
23. Dormeyer, W.; Mohammed, S.; Breukelen, B.; Krijgsveld, J.; Heck, A. J.R. Targeted Analysis of protein termini 2007 6, 4634-4645
24. Noreau J. and Drapeau G.R.; J. Bacteriol. 1979, 140, 911-916
25. Drapeu G.R. J. Biol. Chem. 1980, 255, 839-840
26. Fischer S.; Geub U.; Schaffer M.; Krebe G.B. and Drapeau G.R. J. Protein Chem 1988, 7, 225-225

27. Ingrosso D.; Fowler A.V.; Bleibaum J. and Clarke S.; Specificity of endoproteinase Asp-N (Pseudomonas fragi): Cleavage at glutamyl residues in two proteins. Biochemical and biophysical research communications 1989 , 162(3), 1528-1534

Chapter 5 Summary

The concept of positional proteomics is that a protein can be identified by a single, position-defined peptide, with the two most obvious positional locations within every protein being the N- and C-termini. The strength of positional proteomics arises from the dramatically reduced complexity of proteome samples, however, the strength turns into weakness in protein identification due to risks inherent in identifying a protein on the assignment of a single peptide's fragmentation pattern, the so called "one hit wonder" problem. In this thesis, we focused on developing methods to identify peptides and proteins with higher confidence, by ensuring high quality spectra and/or utilizing internal validation.

An approach was proposed where the N-termini fraction for positional proteomics from the peptide mixtures is enriched, while the rest of the peptide mixture is preserved. Therefore, MS analysis of the remaining peptide mixtures can serve as validation for the results of positional proteomics. In the process, 4-sulfophenyl isothiocyanate (SPITC) is used to modify the N-termini and the SPITC modified peptides are separated from others using electrostatic repulsion hydrophilic interaction chromatography (ERLIC). Moreover, the confidence of N-termini identification is further strengthened by exploiting the unique fragmentation behavior of SPITC peptides, thus we conclude that our approach

offers high fidelity assignment of N-terminal peptides. This work was presented in chapter 3.

Another attempt to simultaneously validating the results of positional proteomics is to use iTRAQ to modify the peptide N-termini and this workflow was tested for its application to identify protease substrates. In this study, both control and protease treated sample were labeled by different tags, allowing direct comparison of protein N-termini with neo-N-termini. In addition, samples were analyzed in duplicate by labeling with two tags (*e.g.* tag 116 and tag 117), aiming for quick validation of peptides by internal replicates. Compared with dimethylation or acetylation modification, the dramatic advantage of using iTRAQ as labeling reagent is that up to 4 or 8 samples can be compared simultaneously with sensitivity improvement, which makes it an attractive approach for the establishment of proteolytic signatures *in vivo*. This work was presented in chapter 4.

The proof of concept of these approaches was demonstrated by application to *E. coli* cell lysates. Further modification or workflow improvement may be needed for biological application. One limitation of the current workflow is that it will not detect proteins with N-terminal modifications such as acetylation or dimethylation, since SPITC only reacts with N-terminal amines. It is feasible to use an appropriate enzyme to treat the protein mixture prior to SPITC modification

in order to catch these protein N-termini.[1] In addition, isotope labeled SPITC can be used for direct sample comparison.[2]

Furthermore, it is common in proteomics applications to use of multiple enzymes for cleavage in order to improve protein coverage. [3] This approach is applicable here: instead of using only trypsin for digestion, the same sample can be treated with a combination of chymotrypsin, trypsin and/or Glu-C, thus improving the chances to produce peptides with the right sizes for high-fidelity MS assignment.

Another application of the current concept is in the field of C-terminal proteomics. The negative charge introduced by SPITC at peptide N-termini, causes the neutralization of N-terminal ions (a and b ions) upon HCD, thereby promoting the generation and survival of C-terminal ions (y ions). Therefore, the overall effect of SPITC derivatization is to improve tandem mass spectra by generating predominantly y-type ions. Similarly, study has shown that Lys-N digested peptides generate predominantly b ions in tandem mass spectra[4]. Thus, high fidelity for C-termini in proteomics scale should be achievable by incorporating Lys-N digestion. Such approach is going to be a valuable alternative for the currently under-utilized C-terminal proteomics.

References

[1]Staes A.; Impens F.; Van Damme P.; Ruttens B.; Goethals M.; Demol H.; Timmerman E.; Vandederchhove J.; Gevaert K.; Selecting protein N-terminal peptides by combined fractional diagonal chromatography. Nat. Protoc. 2011, 6, 1130-1141.

[2] Guillaume E.; Panchaud A.; Affolter M.; Desvergnes V. and Kussmann M.; Differentially isotope-coded N-terminal protein sulphonation: combining protein identification and quantification; Proteomics, 2006, 6, 2338-2349

[3] Mommen, G.P.M; Waterbeemd, B.V.D.; Meiring H.D.; Kersten, G.; Heck, A.J.R and Jong, P.J.M; unbiased selective isolation of protein N-terminal peptides from complex proteome samples using phosphor tagging (PTAG) and TiO2-based depletion. MCP, published on June 22, 2012.

[4] Taouatas N.; Drugan M.M.; Heck A.JR. and Mohammed S.; Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase; *Nature Methods*, 2008, **5**, 405 - 407 (2008)

Appendix

Appendix table 1 N-terminal peptides by SPITC modification

sequence	PSMs	protein accessions	Modification	MH+(Da)
ADAQK	3	P0AES9from22to110	N-Term(SPITC)	747.2444
ISQEVR	2	P30136from593to598	N-Term(SPITC)	946.3796
GQINIPEGR	1	P39321from1066to1074	N-Term(SPITC)	1198.491
KSDQNVR	8	P0A7L0from54to60	N-Term(SPITC)	1061.414
KTGNTPDGR	8	P09373from616to624	N-Term(SPITC)	1160.45
QATKDAGR	7	P0A6Y8from152to159	N-Term(SPITC)	1061.416
KPVNDLSR	12	P0A9M8from685to692	N-Term(SPITC)	1143.498
GLSAKSFDGR	18	P62399from116to125	N-Term(SPITC)	1252.506
PWFIKALR	1	P33341from750to757	N-Term(SPITC)	1245.582
SPDPFER	1	P37653from394to400	N-Term(SPITC)	1062.368
EAAEQAKR	2	P0A705from135to142	N-Term(SPITC)	1117.446
TFYAQLR	1	P75860from143to149	N-Term(SPITC)	1113.453
MATKDGR	2	P65294from26to32	N-Term(SPITC)	993.363
VKLDEAGVR	4	P0AGE9from84to118	N-Term(SPITC)	1201.53
KAPAEPQR	2	P0A9P0from256to263	N-Term(SPITC)	1111.465

sequence	PSMs	protein accessions	Modification	MH+(Da)
FTESEVISR	1	P75990from276to284	N-Term(SPITC)	1282.521
LKNENPR	4	P0ACW6from12to18	N-Term(SPITC)	1085.456
VMVLDIDEER	2	P0AG67from330to339	N-Term(SPITC)	1433.567
ENNAQTTNESAGQK	1	P0AFH8from29to122	N-Term(SPITC)	1706.644
KAIGEAK	6	P0ABT2from134to153	N-Term(SPITC)	931.4012
VLIHPASLPEER	1	P28722from128to139	N-Term(SPITC)	1575.724
LFLGVDAIDLER	2	P0ACK2from183to194	N-Term(SPITC)	1575.716
VAEFFGKEPR	2	P0A6Y8from353to362	N-Term(SPITC)	1394.595
VEQKVAR	5	P25553from337to343	N-Term(SPITC)	1044.464
GIQPEEVER	1	P37353from361to369	N-Term(SPITC)	1271.504
KLTETDQR	2	P0A7F3from34to41	N-Term(SPITC)	1205.493
AIQSEKAR	5	P0A7U7from11to18	N-Term(SPITC)	1117.479
KLEAGDEVR	6	P0AA39from59to67	N-Term(SPITC)	1231.512
VDALTNEVR	1	P30855from260to268	N-Term(SPITC)	1231.507
PADAIDR	1	P76473from532to538	N-Term(SPITC)	972.357
AVKAPGFGDR	1	P0A6F5from275to284	N-Term(SPITC)	1232.52
TQQESLQKALR	1	P00968from390to400	N-Term(SPITC)	1516.695

sequence	PSMs	protein accessions	Modification	MH+(Da)
KSDNPSITR	8	P0A6H1from192to200	N-Term(SPITC)	1232.508
LLMEDLNDGLR	1	P09053from18to28	N-Term(SPITC)	1503.619
AIVKADNR	2	P0A6V8from201to208	N-Term(SPITC)	1101.479
TTQTVSGR	2	P0AGI1from2to9	N-Term(SPITC)	1064.415
KPATAGMENR	4	P15288from59to68	N-Term(SPITC)	1289.504
AQNTAAYIGNGDR	3	P06996from226to238	N-Term(SPITC)	1565.611
STIEER	7	P0A6A8from2to7	N-Term(SPITC)	949.3413
ELKDAGADR	2	P00968from427to435	N-Term(SPITC)	1189.463
ILQNEDKR	1	P13035from247to254	N-Term(SPITC)	1230.523
VDKDAVSR	7	P0A8M0from292to299	N-Term(SPITC)	1104.448
TSHVEYDTPTR	5	P0CE47from65to75	N-Term(SPITC)	1520.566
IDAGKTTTTER	7	P0A6M8from19to29	N-Term(SPITC)	1407.585
VQAIEKNR	3	P0A6H5from113to120	N-Term(SPITC)	1172.522
KWTDQSGQDR	2	P0AGE0from88to97	N-Term(SPITC)	1435.542
VTVKTDGGPR	3	P0AEG8from7to16	N-Term(SPITC)	1244.548
EAAATAGEKEDAPR	2	P45577from123to136	N-Term(SPITC)	1630.649
AKASQLDEAR	2	P62620from354to363	N-Term(SPITC)	1303.544

sequence	PSMs	protein accessions	Modification	MH+(Da)
SPEQQKVVDR	4	P0ADI4from102to111	N-Term(SPITC)	1400.588
LPDINNEASLR	2	P75961from64to74	N-Term(SPITC)	1456.628
ILSKDEGGR	7	P0CE47from311to319	N-Term(SPITC)	1189.502
KDLTAADGQTR	4	P77395from245to255	N-Term(SPITC)	1390.576
KASVEIDR	2	P0A7L3from85to92	N-Term(SPITC)	1132.479
STGTSGAGSDVEKVR	6	P0A9M8from592to606	N-Term(SPITC)	1665.686
AEQAALQADKR	1	P0AFK0from144to154	N-Term(SPITC)	1415.618
KSPFDSGGR	2	P07012from204to212	N-Term(SPITC)	1165.436
EAAGSALKGDR	2	P0ACB2from206to216	N-Term(SPITC)	1289.525
LQQVGDKPR	4	P0AFG0from115to123	N-Term(SPITC)	1255.56
SQGQEEAEKLR	1	P0ABC3from240to250	N-Term(SPITC)	1489.607
EQEAAELKR	2	P0A705from177to185	N-Term(SPITC)	1288.533
ALGKDDEVR	10	P09373from674to682	N-Term(SPITC)	1217.492
QLSNDVNAMR	1	P69776from43to52	N-Term(SPITC)	1362.511
IPEKSVSQSDR	1	P0ABH9from436to446	N-Term(SPITC)	1460.618
SNQFGDTR	6	P0AEH5from2to9	N-Term(SPITC)	1139.385
KLLDEGR	14	P0CE47from264to270	N-Term(SPITC)	1045.446

sequence	PSMs	protein accessions	Modification	MH+(Da)
EVKELTER	5	P0ADG7from108to115	N-Term(SPITC)	1218.514
VAKIYGDR	3	P0AFG3from574to581	N-Term(SPITC)	1136.497
VYKNYDPR	3	P0ABH7from308to315	N-Term(SPITC)	1269.502
DAFTLVLAR	3	P76254from233to241	N-Term(SPITC)	1220.539
GDKSMALR	3	P02359from112to119	N-Term(SPITC)	1092.429
DLLKEQNNR	2	P0A742from127to135	N-Term(SPITC)	1344.571
ALQSSINEDKAH	2	P0AD24from64to75	N-Term(SPITC)	1527.62
ISIKDTR	4	P22259from450to456	N-Term(SPITC)	1047.466
AENQYYGTGR	6	P0A7X3from2to11	N-Term(SPITC)	1373.492
NIDADKVNPR	2	P0AFG8from876to885	N-Term(SPITC)	1356.567
IQAEKSQQSSY	6	P69913from51to61	N-Term(SPITC)	1483.582
VKAPGFGDR	1	P0A6F5from276to284	N-Term(SPITC)	1161.484
HAQAEGIR	2	P0AFI0from40to47	N-Term(SPITC)	1096.438
QLGDKPADVR	8	P15254from849to858	N-Term(SPITC)	1313.563
MELSSLTAVSPVDGR	3	P0AB89from1to15	N-Term(SPITC)	1776.762
VDGTKPVAEVR	3	P69441from196to206	N-Term(SPITC)	1385.627
TLGADALEPKR	1	P0A9W3from537to547	N-Term(SPITC)	1385.624

sequence	PSMs	protein accessions	Modification	MH+(Da)
TIVR	1	P32717from253to270	N-Term(SPITC)	703.2896
ISNVELSKR	2	P0ACJ0from29to37	N-Term(SPITC)	1260.577
IAPLDADR	2	P16685from116to123	N-Term(SPITC)	1085.442
AISTIAESKR	3	P69908from22to31	N-Term(SPITC)	1290.58
KGSENYALTTNQGVR	6	P21179from73to87	N-Term(SPITC)	1852.791
ITKEDIER	3	P0ABK2from366to373	N-Term(SPITC)	1218.516
VKAEMENLR	1	P09372from65to73	N-Term(SPITC)	1304.554
AKDGLALSSR	4	P31663from180to189	N-Term(SPITC)	1232.543
TPDQVKEIAR	5	P33570from528to537	N-Term(SPITC)	1371.609
YAPNAKDLAGR	4	P0AC41from276to286	N-Term(SPITC)	1390.59
LAKEDPSFR	15	P0A6M8from438to446	N-Term(SPITC)	1277.533
IAAGADISKAAAGR	2	P0ACF4from10to23	N-Term(SPITC)	1486.683
LLKAANTGPHAAR	1	P0ADH5from21to33	N-Term(SPITC)	1534.728
GIPADKISAR	3	P0A910from289to298	N-Term(SPITC)	1242.56
KAELESAALNAR	8	P08312from75to86	N-Term(SPITC)	1487.656
QTVDEALKDAQTR	6	P0AEX9from381to393	N-Term(SPITC)	1689.727
IDDGELHGESPGDR	1	P64503from12to25	N-Term(SPITC)	1711.632

sequence	PSMs	protein accessions	Modification	MH+(Da)
PAITEDEIR	1	P39283from8to16	N-Term(SPITC)	1258.506
AESFTTTNR	2	P0ADN2from2to10	N-Term(SPITC)	1241.454
SHLDEVIAR	3	P64455from2to10	N-Term(SPITC)	1254.532
IKEAFDTGVR	3	P0A853from155to164	N-Term(SPITC)	1350.577
AFDQIDNAPEEKAR	17	P0CE47from46to59	N-Term(SPITC)	1818.737
VAQVDKTAVDTYR	2	P36683from830to842	N-Term(SPITC)	1680.735
NLTGKEADAALGR	3	P0A825from327to339	N-Term(SPITC)	1530.672
LSPKESEVLR	5	P69407from151to160	N-Term(SPITC)	1372.628
NKDGIPAVVER	22	P60422from70to80	N-Term(SPITC)	1412.618
GIGPAYEDKVAR	9	P0A7D4from133to144	N-Term(SPITC)	1490.635
ILDPTKVTR	3	P0A6F5from493to501	N-Term(SPITC)	1257.597
AEQTVEAPSVDAR	1	P08506from28to40	N-Term(SPITC)	1587.639
NEPDAVAEKLAR	3	P0A8L1from9to20	N-Term(SPITC)	1527.649
AKIVDEIGLPR	2	P25553from189to199	N-Term(SPITC)	1425.69
QYIEKDAALER	1	P63284from321to331	N-Term(SPITC)	1550.666
FNSLTKEQQQDVITR	5	P09373from740to754	N-Term(SPITC)	2021.906

sequence	PSMs	protein accessions	Modification	MH+(Da)
KFEELVQTR	12	P0A6Y8from528to536	N-Term(SPITC)	1364.587
MSTIEER	4	P0A6A8from1to7	N-Term(SPITC)	1080.382
FNMHLWLSPEIAR	1	P39172from140to152	N-Term(SPITC)	1828.789
VGFFNPIASEKEEGTR	2	P0A7T3from36to51	N-Term(SPITC)	1995.85
KENNLGADVVLR	6	P07118from456to467	N-Term(SPITC)	1542.698
LGISTLDDVLTDIR	1	P77541from59to72	N-Term(SPITC)	1745.803
MQIDSKPEELDR	3	P63284from403to414	N-Term(SPITC)	1675.689
VAIFNAATGKADR	4	P60624from70to82	N-Term(SPITC)	1548.701
VKAALELAEQR	11	P0A7V8from155to165	N-Term(SPITC)	1442.674
TLTQEDVEALEKR	3	P0A8M3from111to123	N-Term(SPITC)	1746.758
FLSALAGENDPEAKR	3	P04079from294to308	N-Term(SPITC)	1832.789
ELVTAAKLGGGDPDANPR	2	P0A8A0from31to48	N-Term(SPITC)	1995.896
VLQEAADKSNPLIER	1	P0A7B1from21to35	N-Term(SPITC)	1897.888
GVTVDKMTELR	5	P0A7J3from32to42	N-Term(SPITC)	1463.632
ILENGEVKPLDVK	1	P0A6F9from48to97	N-Term(SPITC)	1668.804
KGFIDVEQVR	4	P37744from261to270	N-Term(SPITC)	1405.62

sequence	PSMs	protein accessions	Modification	MH+(Da)
VKLDTTGLIDR	2	P0A953from52to62	N-Term(SPITC)	1445.673
IETLNFMEPQSAADLIR	1	P0ABZ1from107to123	N-Term(SPITC)	2162.957
LFGKPEIDGSR	4	P33221from352to362	N-Term(SPITC)	1433.615
GITINTSHVEYDTPTR	2	P0CE47from60to75	N-Term(SPITC)	2018.863
AYADDKAIVGGIAR	6	P0ABD5from94to107	N-Term(SPITC)	1634.737
ESYTKEDLLASGR	5	P0A6Q3from6to18	N-Term(SPITC)	1683.689
ILLDTKGPEIR	6	P0AD61from63to73	N-Term(SPITC)	1469.715
IFDLGNVIVDIDFNR	3	P0A8Y3from4to18	N-Term(SPITC)	1964.899
AKVDQLSNDVNAMR	8	P69776from39to52	N-Term(SPITC)	1775.761
NILNELQKDGR	2	P0ACJ0from18to28	N-Term(SPITC)	1514.672
ATYYSNDFR	3	P0A6N4from2to10	N-Term(SPITC)	1351.468
DKLEPYFTEGR	2	P00961from623to633	N-Term(SPITC)	1569.64
LGEYLKPLAER	2	P0A9S5from21to31	N-Term(SPITC)	1503.696
KDDTIPAIISHDE	2	P0AE18from252to264	N-Term(SPITC)	1668.685
VVSMPSTDAFDKQDAAYR	2	P27302from580to597	N-Term(SPITC)	2215.917

sequence	PSMs	protein accessions	Modification	MH+(Da)
TLNDAVEVKHADNTLTFGPR	8	P0AG55from36to55	N-Term(SPITC)	2413.084
LTTDKGEWLLYR	2	P00961from108to119	N-Term(SPITC)	1709.752
GFAVTPPELTKDDER	3	P62707from118to132	N-Term(SPITC)	1889.79
AEHWIDVR	5	P23857from20to27	N-Term(SPITC)	1240.49
SEFITVAR	6	P0ABA4from2to9	N-Term(SPITC)	1137.468
QAGELQEKLIAVNR	4	P0A7W1from5to20	N-Term(SPITC)	1783.841
SQVSTEFIPTR	4	P0AEZ9from2to12	N-Term(SPITC)	1479.617
FGVPLVR	1	P00864from386to392	N-Term(SPITC)	1002.461
NKVTDAEIAEVLAR	6	P63284from529to542	N-Term(SPITC)	1743.81
SGITFSQELKDSGMR	2	P0A953from31to45	N-Term(SPITC)	1870.773
VGGSTYQVPVEVR	2	P02359from80to92	N-Term(SPITC)	1605.692
ALYDSEKDAYLIGR	2	P22106from125to138	N-Term(SPITC)	1828.779
LASSPSPLNPGTNVAR	2	P0AE22from24to39	N-Term(SPITC)	1795.801
WQTLSAKSFPLPR	1	P45423from129to141	N-Term(SPITC)	1745.826
TSGGELDKLLAAGR	1	P16700from324to337	N-Term(SPITC)	1602.735

sequence	PSMs	protein accessions	Modification	MH+(Da)
KIDGIPALLDR	16	P37769from68to78	N-Term(SPITC)	1425.683
PVITLPDGSQR	2	P0A8M3from2to12	N-Term(SPITC)	1397.628
QTFAQKAQAFEQDR	3	P0AEU7from92to105	N-Term(SPITC)	1882.78
INKALDFIAER	3	P04805from457to467	N-Term(SPITC)	1504.69
NIAEAASGIDKLVSR	3	P0ABU5from197to211	N-Term(SPITC)	1758.815
TELLNSSYDVSR	5	P16700from26to37	N-Term(SPITC)	1598.642
NVYIKEAFDTGVR	21	P0A853from152to164	N-Term(SPITC)	1726.758
MNIIEANVATPDAR	4	P61714from1to14	N-Term(SPITC)	1729.742
ALLPLVEEKADR	5	P07003from318to329	N-Term(SPITC)	1568.743
MSHLDEVIAR	2	P64455from1to10	N-Term(SPITC)	1385.565
SANIALVLYKDGER	3	P60422from88to101	N-Term(SPITC)	1763.814
SLVWDEAQKLTGR	3	P21179from301to313	N-Term(SPITC)	1717.78
AAILSQSQLTALFGQYR	1	P37177from151to167	N-Term(SPITC)	2081.971
IWDSTDALELKEVPER	3	P0A9P0from162to177	N-Term(SPITC)	2115.927
IIDQEVKALIER	3	P0AAI3from544to555	N-Term(SPITC)	1641.791
MYQDLIR	1	P63224from1to7	N-Term(SPITC)	1153.45

sequence	PSMs	protein accessions	Modification	MH+(Da)
VVAILLNDEVR	1	P21179from602to612	N-Term(SPITC)	1455.691
LGEGDKVVSLIVPR	2	P0AES4from726to739	N-Term(SPITC)	1696.836
EQIIFPEIDYDKVDR	3	P62399from134to148	N-Term(SPITC)	2094.907
TSENPLLALR	4	P0A9J8from2to11	N-Term(SPITC)	1328.594
MESLTLQPIAR	2	P0A6D3from1to11	N-Term(SPITC)	1473.655
VQELAEKLYSELR	2	P16659from489to501	N-Term(SPITC)	1792.823
YLDVSTLKELAR	1	P0A784from131to142	N-Term(SPITC)	1622.76
AVTNVAELNALVER	4	P0A9Q7from2to15	N-Term(SPITC)	1713.78
LDELNNVDDFR	1	P32171from441to451	N-Term(SPITC)	1564.619
MENYLIDNLDR	2	P0ACI6from1to11	N-Term(SPITC)	1610.635
MQDPIADMLTR	1	P0A7W7from3to13	N-Term(SPITC)	1505.591
MSVVPVADVLQGR	2	P0A8M0from1to13	N-Term(SPITC)	1585.719
QLPDKAIDLIDEAASSIR	2	P63284from385to402	N-Term(SPITC)	2170.018
TMYATLEEAIDAAR	1	P25738from2to15	N-Term(SPITC)	1769.728
MDALELLINR	1	P0ACY1from1to10	N-Term(SPITC)	1402.618
STQLDPTQLAIEFLR	1	P64476from2to16	N-Term(SPITC)	1946.898

sequence	PSMs	protein accessions	Modification	MH+(Da)
ILLNDEVR	1	P21179from605to612	N-Term(SPITC)	1186.526
ELLMALR	1	P0AF10from111to117	N-Term(SPITC)	1060.454
VKAQDVQR	1	P68919from72to79	N-Term(SPITC)	1158.506
GTVVTGR	1	P0CE47from225to231	N-Term(SPITC)	904.3669
NDDAKAVQR	2	P0AG16from484to492	N-Term(SPITC)	1231.486
KTLAEGQR	2	P0A972from42to49	N-Term(SPITC)	1117.48
KAIATPDR	2	P0ABJ1from221to228	N-Term(SPITC)	1086.475
TTLSTDPKR	1	P22259from230to263	N-Term(SPITC)	1233.531
EEIKEVAPHR	2	P0A6H1from50to59	N-Term(SPITC)	1422.611
ATVSMR	2	P0A7V0from2to7	N-Term(SPITC)	879.3162
IISAKDHGDSFR	1	P37902from181to192	N-Term(SPITC)	1560.66
NAVVTAADKASANR	1	P00805from153to166	N-Term(SPITC)	1602.713
LTGKEADAALGR	2	P0A825from328to339	N-Term(SPITC)	1416.623
KDLANAIR	2	P33570from4to11	N-Term(SPITC)	1115.505
GAPTITKDGVSVAR	1	P0A6F5from45to58	N-Term(SPITC)	1586.735
MNEQYSALR	1	P00864from1to9	N-Term(SPITC)	1326.493
TTIVSVR	2	P0A7B8from2to8	N-Term(SPITC)	990.436

sequence	PSMs	protein accessions	Modification	MH+(Da)
TDLTAQEPAWQTR	1	P33599from2to14	N-Term(SPITC)	1731.708
VDYPLPPTGSR	1	P76193from24to34	N-Term(SPITC)	1416.593
VTYPLPTDGSR	1	P0AAX8from25to35	N-Term(SPITC)	1420.592
MQVSVETTQGLGR	1	P0A850from1to13	N-Term(SPITC)	1620.684
DLQSIADYPVKVR	1	P23830from420to432	N-Term(SPITC)	1718.789
LIVIDFIDMTPVR	5	P21513from342to354	N-Term(SPITC)	1746.811
TTIVDSNLPVAR	1	P39160from2to13	N-Term(SPITC)	1500.674
DWYVVDATGKTLGR	3	P0AA10from14to27	N-Term(SPITC)	1795.766
MFEINPVNNR	2	P07012from1to10	N-Term(SPITC)	1448.569
SAQPVDIQIFGR	1	P0ADS2from2to13	N-Term(SPITC)	1545.698
MIDTTLPLTDIHR	1	P22333from1to13	N-Term(SPITC)	1740.777
TQTLSQLENSGAFIER	1	P33195from2to17	N-Term(SPITC)	2008.876
SVVPVADVLQGR	4	P0A8M0from2to13	N-Term(SPITC)	1454.668
SQNVYQFIDLQR	5	P09832from2to13	N-Term(SPITC)	1725.732
TSLVVPGLDTLR	1	P0AAY6from2to13	N-Term(SPITC)	1485.712
SNDVIQDDVFR	1	P52643from265to282	N-Term(SPITC)	1522.609
GVEVIAVDR	1	P33221from35to43	N-Term(SPITC)	1172.509

sequence	PSMs	protein accessions	Modification	MH+(Da)
VDAYDR	1	P77690from213to218	N-Term(SPITC)	953.3083
ESLLLPSPFDYSR	1	P76257from418to430	N-Term(SPITC)	1738.743
IDPR	1	P08244from27to39	N-Term(SPITC)	715.2537
EQWDAQR	1	P37672from77to83	N-Term(SPITC)	1147.395
ISGADLTR	1	P37610from18to25	N-Term(SPITC)	1047.422
GDEAYSGSR	1	P0A853from70to78	N-Term(SPITC)	1156.363
HGESQWNKENR	1	P62707from11to21	N-Term(SPITC)	1599.596
QAKDVAESDR	2	P36683from371to380	N-Term(SPITC)	1333.507
EQEKGLDR	2	P0A853from118to125	N-Term(SPITC)	1189.459
PEAIPELLER	1	P0AFD1from155to164	N-Term(SPITC)	1381.617
VIYVAADR	1	P76157from14to35	N-Term(SPITC)	1121.482
EDEGLADR	1	P0AE08from113to120	N-Term(SPITC)	1119.38
PDNAGILR	1	P46879from58to65	N-Term(SPITC)	1070.442
PQGQLQDIER	1	P15043from356to365	N-Term(SPITC)	1398.573
YEGIDER	1	P0A873from115to121	N-Term(SPITC)	1096.37
EQQVALR	1	P71229from139to145	N-Term(SPITC)	1058.442
EHLSQEVLGKR	1	P0AE88from179to189	N-Term(SPITC)	1510.672

sequence	PSMs	protein accessions	Modification	MH+(Da)
YPEGTKLTGR	2	P0AG67from274to283	N-Term(SPITC)	1336.564
VGVETFKAEVER	2	P17846from314to325	N-Term(SPITC)	1578.679
EFYEKPTTER	1	P68679from36to45	N-Term(SPITC)	1514.59
VLEEKGFR	1	P0A9L8from242to249	N-Term(SPITC)	1192.509
SQLNYSEENLKQAR	1	P21888from300to313	N-Term(SPITC)	1894.808
AEGLFKEER	1	P0AB77from18to26	N-Term(SPITC)	1293.522
EDGTIDFDDGSKTENTR	1	P22259from317to333	N-Term(SPITC)	2114.791
TYDDDPTKYQDLR	2	P0AEM9from175to187	N-Term(SPITC)	1844.704
LDYSKGLPER	1	P31677from264to273	N-Term(SPITC)	1392.587
IKEEDFIDR	1	P0AES4from572to580	N-Term(SPITC)	1379.553
YAIDAEKIGR	1	P37759from304to313	N-Term(SPITC)	1350.578
IINGEVPEGLKGR	1	P63284from223to235	N-Term(SPITC)	1596.74
AKITVPVDATEEQVR	2	P07813from812to826	N-Term(SPITC)	1870.878
HGYAFNELDLGKR	1	P0ADN2from38to50	N-Term(SPITC)	1734.735
AGEGAKVIELQGIAGTSAAR	2	P02925from126to164	N-Term(SPITC)	2114.006
LTLDLGGEKR	2	P00959from601to	N-Term(SPITC)	1316.59

sequence	PSMs	protein accessions	Modification	MH+(Da)
		610		5
EHWIDVR	1	P23857from21to27	N-Term(SPITC)	1169.444
VPGLDFKR	1	P15639from374to381	N-Term(SPITC)	1146.506
SIVPNALGKDDEVR	1	P09373from669to682	N-Term(SPITC)	1727.764
IAFVNKMDR	1	P0A6M8from138to146	N-Term(SPITC)	1308.554
FDPEFEKISR	1	P13029from382to391	N-Term(SPITC)	1482.6
LFADEPTGNLDR	1	P0A9T8from168to179	N-Term(SPITC)	1562.616
ATVNQLVR	1	P0A7S3from2to9	N-Term(SPITC)	1115.497
SLIGPDGEQYKLPR	1	P0A9Q1from150to163	N-Term(SPITC)	1787.8
AIGEAKDDDTADILTAASR	2	P0ABT2from134to153	N-Term(SPITC)	2147.908
EAEGQDFQLYPGELGKR	1	P0A8P3from13to29	N-Term(SPITC)	2151.915
EITASLVKELR	1	P0A6P1from3to13	N-Term(SPITC)	1473.705
GSLPIALDEVITDGHKR	2	P0A9Q7from465to481	N-Term(SPITC)	2035.951
QLNITGNNVEITEALR	1	P0AFX0from2to17	N-Term(SPITC)	1999.923
VYSKLTSENPIDLVR	2	P0A991from259to273	N-Term(SPITC)	1948.91

sequence	PSMs	protein accessions	Modification	MH+(Da)
ELLSQYDFPGDDTPIVR	1	P0CE47from156to172	N-Term(SPITC)	2179.939
EVTYGDVTLDFGKPFEK	2	P0A8N3from310to330	N-Term(SPITC)	2159.935
LYTTNADGELITIDTADNKILSR	2	P76116from199to221	N-Term(SPITC)	2752.272
ETEGQALKALIEQR	1	P23839from146to159	N-Term(SPITC)	1800.825
ANQANIPVITLDR	1	P02925from103to115	N-Term(SPITC)	1639.75
SGIIELPADAPIGTDIR	1	P07395from132to148	N-Term(SPITC)	1952.915
SNQEPATILLIDDHPMLR	1	P0AF28from2to19	N-Term(SPITC)	2278.026
QLIDGIKDLAVQYR	2	P0AB89from149to162	N-Term(SPITC)	1846.874
LPDWDAKIANLSGGER	2	P0A9W3from153to168	N-Term(SPITC)	1956.843
IVGLEIGADDYIPKPFNPR	1	P0AA16from92to110	N-Term(SPITC)	2329.095
VTPTVLQKGR	1	P34749from306to315	N-Term(SPITC)	1313.636
SVLQVLHIPDER	1	P0A6K3from2to13	N-Term(SPITC)	1620.748
EFEKAAASR	1	P00448from116to124	N-Term(SPITC)	1223.477

sequence	PSMs	protein accessions	Modification	MH+(Da)
CGIVGAIAQR	3	P17169from2to11	N-Term(SPITC); C1(Carbamidomethyl)	1259.528
CAQVEALEIIVTAMLR	1	P0AAN9from23to38	N-Term(SPITC); C1(Carbamidomethyl)	2031.946
SQVQSGILPEHCR	3	P76536from2to14	N-Term(SPITC); C12(Carbamidomethyl)	1725.713
AASEAVKDAALSCDQFFVNHR	3	P0A7V0from75to95	N-Term(SPITC); C13(Carbamidomethyl)	2551.065
LCNQKGVER	1	P0AAJ8from176to184	N-Term(SPITC); C2(Carbamidomethyl)	1318.544
ECISENQILKR	2	P0A9C0from70to80	N-Term(SPITC); C2(Carbamidomethyl)	1604.679
TICPDQKGLIAR	1	P37051from11to22	N-Term(SPITC); C3(Carbamidomethyl)	1586.704

sequence	PSMs	protein accessions	Modification	MH+(Da)
QVCAKLSASASDYLR	1	P77439from218to232	N-Term(SPITC); C3(Carbamidomethyl); K5(Guanidinyl)	1925.836
ALPCPVR	1	P05827from172to178	N-Term(SPITC); C4(Carbamidomethyl)	1027.425
FLTCGSVDDGKSTLIGR	1	P23845from30to46	N-Term(SPITC); C4(Carbamidomethyl)	2040.878
GALDCSGVKDR	7	P0A7S3from100to110	N-Term(SPITC); C5(Carbamidomethyl)	1392.53
IVQACTQDKQANFK	2	P0AES9from83to110	N-Term(SPITC); C5(Carbamidomethyl)	1865.799
LLSLCGPFDDNIKQLER	1	P0A9K3from18to34	N-Term(SPITC); C5(Carbamidomethyl)	2233.006
KVEADCR	8	P60422from183to189	N-Term(SPITC); C6(Carbamidomethyl)	1092.396

sequence	PSMs	protein accessions	Modification	MH+(Da)
TGFAECAFEDAAR	1	P0A9U8from250to262	N-Term(SPITC); C6(Carbamidomethyl)	1659.58
LIIFLICLLVFWLWLHK	2	P0AFS1from167to188	N-Term(SPITC); C7(Carbamidomethyl)	2442.295
TFFAEFASCLTELQTR	1	P16694from675to690	N-Term(SPITC); C9(Carbamidomethyl)	2135.874
DHTASLGACEIPR	1	P0A8P1from192to204	N-Term(SPITC); C9(Carbamidomethyl)	1641.647
GKSVEEILGK	5	P0A7W1from158to167	N-Term(SPITC); K10(Guanidinyl)	1316.594
SATPEQELGKLPLGSR	1	P00864from684to699	N-Term(SPITC); K10(Guanidinyl)	1939.899
PLSLTLIPESKPGNGK	6	P0AEH1from282to319	N-Term(SPITC); K11(Guanidinyl)	1907.947
IWDSTDALELKEVPER	2	P0A9P0from162to177	N-Term(SPITC); K11(Guanidinyl)	2157.953

sequence	PSMs	protein accessions	Modification	MH+(Da)
ILENGEVKPLDVK	1	P0A6F9from48to97	N-Term(SPITC); K13(Guanidinyl)	1710.824
ALLYEETAESVEKR	1	P0ACG8from82to95	N-Term(SPITC); K13(Guanidinyl)	1894.817
SSGDPADQKYVELK	1	P0AEH5from27to42	N-Term(SPITC); K14(Guanidinyl)	1793.75
TPFAPIVNTATSLKPVR	1	P0A996from106to122	N-Term(SPITC); K14(SPITC)	2241.999
LANELSDAAENKGTAVK	4	P02359from120to138	N-Term(SPITC); K17(Guanidinyl)	1987.887
IASDNVLGGKIAGDYIAK	1	P02925from126to164	N-Term(SPITC); K18(Guanidinyl)	2061.991
TKSELIER	1	P0A6Y1from2to9	N-Term(SPITC); K2(Guanidinyl)	1232.537
AKSFIVR	1	P46850from318to324	N-Term(SPITC); K2(Guanidinyl)	1077.505

sequence	PSMs	protein accessions	Modification	MH+(Da)
GKTVINFDNAIIAAGSR	1	P0A9P0from132to148	N-Term(SPITC); K2(Guanidinyl)	2003.952
NKVTDAEIAEVLAR	1	P63284from529to542	N-Term(SPITC); K2(Guanidinyl)	1785.815
TKTGELSIHCTELR	1	P0A8N3from102to144	N-Term(SPITC); K2(Guanidinyl); C10(Carbamidomethyl)	1901.818
IHKNMSIER	1	P0ADG7from70to78	N-Term(SPITC); K3(Guanidinyl)	1384.591
AGKAGVEVDDR	1	P0A9P0from282to292	N-Term(SPITC); K3(Guanidinyl)	1373.563
AAKSDNGASNLLR	1	P0ABC7from366to378	N-Term(SPITC); K3(Guanidinyl)	1573.68
TDKLTSLR	1	P0A870from2to9	N-Term(SPITC); K3(Guanidinyl)	1190.524

sequence	PSMs	protein accessions	Modification	MH+(Da)
TTKLHVHDENNECGIGDVVEIR	1	P0AG63from41to62	N-Term(SPITC); K3(Guanidinyl); C13(Carbamidomethyl)	2792.2
SAKTENELEEIK	1	P30958from1032to1049	N-Term(SPITC); K3(Guanidinyl); K12(Guanidinyl)	1689.728
LQKIFPIR	1	P0A8G0from145to152	N-Term(SPITC); K3(SPITC)	1444.574
VIKHPHAVLLLDEIEK	2	P0ABH9from553to592	N-Term(SPITC); K3(SPITC)	2284.043
LSKVLPNPDNVELIR	1	P77397from257to271	N-Term(SPITC); K3(SPITC)	2136.939
DEKTSEWDER	1	P15640from320to329	N-Term(SPITC); K3(SPITC)	1724.502
ANIK	1	P52143from596to730	N-Term(SPITC); K4(Guanidinyl)	702.2695
SEFKNPER	1	P39277from208to215	N-Term(SPITC); K4(Guanidinyl)	1263.485

sequence	PSMs	protein accessions	Modification	MH+(Da)
PSEK	1	P21507from386to402	N-Term(SPITC); K4(Guanidinyl)	717.2314
APTKSLQQAAR	4	P0A6X1from391to401	N-Term(SPITC); K4(SPITC)	1600.605
ADAQK	1	P0AES9from22to110	N-Term(SPITC); K5(Guanidinyl)	789.2638
ADAQKAADNK	1	P0AES9from22to110	N-Term(SPITC); K5(Guanidinyl)	1288.511
VDGTKPVAEVR	1	P69441from196to206	N-Term(SPITC); K5(Guanidinyl)	1427.651
LMTDKYEIDAR	1	P09836from13to23	N-Term(SPITC); K5(Guanidinyl)	1611.652
GEVNKVIEQAR	1	P00956from817to827	N-Term(SPITC); K5(Guanidinyl)	1499.68
FNQQKNTLLVLSHESR	1	P39382from211to226	N-Term(SPITC); K5(Guanidinyl)	2171.009

sequence	PSMs	protein accessions	Modification	MH+(Da)
EALEKAEEAGVDLVEISPNAEPPVCR	1	P0A707from41to66	N-Term(SPITC); K5(Guanidinyl); C25(Carbamidomethyl)	3079.365
IIDFKTALTASGR	1	P27254from253to265	N-Term(SPITC); K5(SPITC)	1822.732
SLINTK	2	P0AE08from2to32	N-Term(SPITC); K6(Guanidinyl)	932.3962
DPTLAK	1	P52101from80to91	N-Term(SPITC); K6(Guanidinyl)	901.3502
WADASKADR	1	P09147from300to308	N-Term(SPITC); K6(Guanidinyl)	1276.471
NAVDAAKALGIDAR	1	P0ABP8from137to150	N-Term(SPITC); K7(Guanidinyl)	1641.731
APLMTDKYEIDAR	1	P09836from11to23	N-Term(SPITC); K7(Guanidinyl)	1779.736
ELVTAAKLGGGDPDANPR	1	P0A8A0from31to48	N-Term(SPITC); K7(Guanidinyl)	2037.898

sequence	PSMs	protein accessions	Modification	MH+(Da)
SNVPAELK	1	P0A6T9from2to18	N-Term(SPITC); K8(Guanidinyl)	1114.467
VTTEVTVKLLPKPPVAR	1	P0AEP9from219to235	N-Term(SPITC); K8(SPITC); K12(Guanidinyl)	2320.092
SLSTEATAK	2	P0ADZ4from2to17	N-Term(SPITC); K9(Guanidinyl)	1164.47
TIMSPWAAKR	1	P0ACY3from174to183	N-Term(SPITC); K9(Guanidinyl)	1417.627
SLLNVPAGK	1	P0A7A9from2to44	N-Term(SPITC); K9(Guanidinyl)	1155.526
AEITASLVK	1	P0A6P1from2to13	N-Term(SPITC); K9(Guanidinyl)	1188.537

Appendix table 2 N-terminal peptides of AspN digestion by iTRAQ labeling

Sequence	Protein Group Accessions	q-Value	XCorr	SpScore	MH+ [Da]	Delta Mass [PPM]
dLGEPLSLITESVFAR	P00350from272to287	0	5.24	798.95	1891.02463	-4.56
ePLSLITESVFAR	P00350from275to287	0.00372	1.73	115.06	1605.90691	3.83
dFAYQGFAR	P00509from211to219	0	2.19	310.47	1218.60478	-1.23
dDVIGTLAR	P00582from115to123	0.00284	2.12	218.27	1103.62175	0.13
dFVPYFR	P00864from674to683	0.00593	1.73	226.72	1087.57451	1.22
dDLYGIIR	P00956from35to42	0.00154	2.41	209.04	1108.6165	0.65
dQYPEILR	P00961from203to210	0.00154	2.25	220.89	1177.63628	-0.82
dLQTLTEEAVR	P00961from499to509	0	2.14	225.1	1418.76286	-1.26
dTIQAVLAR	P00961from544to552	0	3.02	302.3	1130.66985	0.85
dQSEQVPGMIER	P02358from13to24	0	3.72	615.72	1532.75383	0.25
eQVPGMIER	P02358from16to24	0.00372	1.99	164.02	1202.64055	3.88
dELETTFR	P02358from72to79	0	2.76	277.55	1154.58318	-1.47
aGEGAKVIELQGIAGTSAAR	P02925from130to164	0	3.77	430.35	2043.14243	3.42
gEGAKVIELQGIAGTSAAR	P02925from146to164	0.00154	3.92	306.13	1972.10703	4.41
dEMALGALR	P02925from216to224	0	3.14	517.59	1119.59758	-1.07
dAIFIEELR	P04079from428to436	0	3.03	534.27	1249.69658	1.44

dFNPSGIILSGGPESTTEENSPR	P04079from48to70	0	4.9	459.84	2548.20792	-9.58
dDGYLPEALLNYLVR	P04805from252to266	0	3.44	227.52	1895.01615	4.82
dTFIELVR	P05020from310to317	0	2.57	226.84	1136.64763	0.48
dDTLVPFLAGETVR	P05020from330to343	0	2.36	321.43	1676.89934	-1.29
dISEFAPR	P05055from547to554	0.00372	1.98	196.88	1078.56841	-0.4
dITELEAFR	P06959from433to441	0	2.99	334.34	1237.65849	0.09
dSFTIQPGER	P06968from106to115	0.00372	1.68	95.68	1293.6552	-3.29
dLTTVIENLR	P06989from190to199	0	2.74	228.58	1317.76201	6.57
dTLVLLGTQFPYR	P07003from267to279	0.00154	2.6	430.26	1666.92888	-2.1
dIGWGSQIR	P07012from316to324	0.00154	2.51	223.61	1175.63164	-1.02
dAFSVFR	P07014from199to205	0.00154	1.58	142.69	985.52544	-0.81
dEWQAVAPSWR	P07395from442to452	0	3.46	737.03	1488.73308	-4.05
dMVIFR	P08200from148to153	0.00372	1.99	232.75	924.51091	-2.52
dWGYQLAR	P08200from226to250	0	2.69	341.04	1152.5949	-0.71
dAFLQQILLR	P08200from283to292	0	2.47	355.24	1360.80742	-2.49
vAIKGPLTTPVGGGIR	P08200from97to112	0.00154	4.36	202.31	1680.03257	-0.14
dEFSMSAISIPR	P08839from521to532	0	3.3	501.78	1496.7603	1.89
dFLTLPGYR	P08997from522to530	0.00284	2.35	267.58	1225.67485	1

dWQNEVNVR	P09373from19to27	0	2.26	279.53	1303.65337	-1.27
dLENGVNLEQTIR	P09373from207to219	0	3.21	541.54	1644.8781	4.16
dDLAVDLVER	P09373from575to584	0	2.91	300.09	1288.68828	-1.67
dTSMGLTPLEGLVMGTR	P0A6A3from227to243	0	4.58	401.02	1921.97934	-4.67
dYTIELVR	P0A6B7from367to374	0.00154	1.98	123.77	1152.64238	0.34
dLPLIASNFR	P0A6F1from84to93	0.00702	1.98	217.55	1289.74407	5.24
dANIISVSQR	P0A6F3from25to34	0	3.04	481.78	1246.69084	-0.19
dILGTR	P0A6F3from425to430	0	1.75	99.17	818.4857	-4.17
dVEGEALATLVVNTMR	P0A6F5from253to268	0	4.87	648.32	1861.99161	3.59
dGVGEEAAIQGR	P0A6F5from334to345	0.00372	1.83	71.96	1345.69695	7.58
eGVVAGGGVALIR	P0A6F5from409to421	0.00803	2.2	189.62	1341.8073	4.69
dLGAAGGMGGMGGMGGMM	P0A6F5from531to548	0.00154	3.97	1048.54	1701.705	0.15
dGTTTATVLAQAIITEGLK	P0A6F5from87to118	0	4.46	433.01	2047.14841	2.03
dLIPELQGR	P0A6H5from317to325	0.00154	1.79	117.69	1184.6762	-2.76
dAGLNIAPFITLTR	P0A6J8from150to163	0	3.57	476.89	1645.95195	5.26
dPFVGNLTFFR	P0A6M8from327to337	0	3.21	452.9	1456.76238	-8.27
dAPIILER	P0A6M8from401to408	0	2.26	401.99	1070.63603	-0.46
dWMEQEQER	P0A6M8from51to59	0	2.16	161.05	1394.61455	-1.48

eVPLSEMFGYATQLR	P0A6M8from657to671	0	5.39	678.66	1884.97098	1.3
dEAPSNVAQAVIEAR	P0A6M8from688to702	0	4.8	599.86	1713.89371	0.57
dFTIEVER	P0A6M8from94to101	0	2.1	283.61	1152.60393	-1.47
eVTGFIR	P0A6P1from252to258	0.00154	2.5	257.16	965.55705	-0.5
dIELAIENMR	P0A6P1from33to42	0	3.66	706.5	1347.70903	-0.56
dAGYTAVISHR	P0A6P9from192to371	0	4.01	457.98	1333.69223	-7.3
eVHLEGGFVGMAAAPSGASTGSR	P0A6P9from24to46	0	4.05	501.8	2332.15097	-0.09
dLLASGR	P0A6Q3from12to18	0.00154	1.73	152.99	875.50902	-1.79
dEWIVTR	P0A6R0from30to36	0	2.08	218.53	1062.56755	-5.99
eLFFEEIR	P0A6X7from28to35	0	2.78	352.14	1226.6563	-1.11
dIPITAR	P0A6X7from70to76	0.00803	1.79	178.51	929.55614	-1.51
dPLAMQR	P0A6Y8from255to261	0.00372	1.92	252.19	974.52233	-2.61
aKLESLVEDLVNR	P0A6Y8from303to315	0	4.47	501.81	1629.93912	3.68
dVILVGGQTR	P0A6Y8from336to345	0	2.96	464.28	1201.7072	0.99
dNQSAVTIHVLQGER	P0A6Y8from431to445	0	4.93	593.09	1810.95536	-0.75
aADNKSLGQFNLDGINPAPR	P0A6Y8from448to467	0	4.69	549.42	2242.17899	2.39
eIFLR	P0A6Z3from25to33	0.00593	1.7	219	821.50188	-2.63
eQFMPNYLR	P0A6Z3from311to3	0.001	2.31	228.9	1341.679	1.24

	19	54		6	74	
dVIQLPAFLAR	P0A715from110to120	0.00372	2.64	578.64	1386.82781	0.96
dGLVEFMTSGPIVVSVLEGENAVQR	P0A763from61to85	0	4.61	360.27	2790.45646	2.17
dLLGATNPANALAGTLR	P0A763from88to104	0	5.27	588.36	1812.01627	1.52
dLLLAGYGGR	P0A796from274to283	0	2.48	313.81	1178.66838	-0.43
dVAEGELVVLENVR	P0A799from100to113	0	3.96	645.87	1685.92436	0.83
eGELVVLENVR	P0A799from103to113	0	2.77	398.29	1400.7935	2.16
dVLVPTPYPLQPGSVIR	P0A7A9from71to87	0	4.33	412.54	1995.1428	-0.33
dLIAIGSGGPYAQAAAR	P0A7B8from119to135	0.00284	2.58	296.01	1774.96086	0.06
eTMILR	P0A7D4from422to427	0.00154	2.32	181.29	906.5208	-3.31
eAGIPTQMER	P0A7D7from63to72	0.00702	1.94	122.31	1275.64739	-3.83
dAILVPGGFGYR	P0A7E5from345to356	0.00284	2.35	226.92	1408.773	-1.02
lITEWR	P0A7E5from415to420	0.00154	1.88	206.71	961.56279	0.18
dIALGAGGLPMGR	P0A7G6from49to61	0	2.91	546.29	1371.75786	0.31
mNQTLLSSFGTPFER	P0A7J0from1to15	0	2.03	91.75	1871.94951	0.74
dAFVGPTLIAYSMEHPGAAAR	P0A7J3from64to94	0	5.11	545.27	2318.15586	-8.66
dIEAMTR	P0A7J7from121to127	0.00154	1.54	74.12	979.50084	-3.02
dLVESAPAALK	P0A7K2from76to121	0	3.71	673.26	1257.72112	0.09

eLNTELLNLLR	P0A7M6from13to23	0	2.6	152.12	1471.86467	0.48
eQFNLR	P0A7M6from24to29	0.00154	2.05	256.71	950.52013	-1.42
eKSVEELNTELLNLLR	P0A7M6from8to23	0	4.25	466.53	2044.15464	4.91
aTVNQLVR	P0A7S3from2to9	0	2.32	340.54	1044.63042	-1.61
dSITSQLER	P0A7V3from118to126	0.00803	2.41	150.46	1192.63103	-1.56
dVVSIR	P0A7V8from141to146	0.00154	2.09	151.34	832.50206	-3.25
dYGVQLR	P0A7V8from50to56	0	2.48	291.96	994.54637	-1.32
aYGSTNPINVVR	P0A7W1from116to138	0.00284	1.82	137.51	1434.7874	0.95
sMQDPIADMLTR	P0A7W7from2to13	0	2.92	330.52	1521.75347	-1.73
dPIAMENAINAIPGVVTVGLFANR	P0A7Z0from178to201	0	3.35	187.96	2626.41025	-3.07
eNAINAIPGVVTVGLFANR	P0A7Z0from183to201	0	3.87	506.48	2099.17746	0.26
dLELTVR	P0A7Z4from259to265	0	1.99	125.06	989.57738	-1.3
dGIVVEYYGTPTPLR	P0A805from38to52	0	4.52	406.38	1823.97544	3.04
eEHIELIASENYTSPR	P0A825from27to42	0	3.71	596.41	2032.00522	-4.46
eVFLER	P0A825from300to305	0.00593	2.19	162.79	936.52965	-1.43
dPLTGPMPYQGR	P0A836from150to161	0.00803	2.12	140.83	1475.7487	1
eVGVNVPVVVR	P0A836from338to348	0	2.83	665.18	1310.7957	0.41
dFVLAMGQGR	P0A850from184to193	0	2.97	744.23	1237.65349	1.3

mQVSVETTQGLGR	P0A850from1to13	0	2.89	342.57	1549.81316	-2.08
dLEGLER	P0A853from172to178	0.00154	2.73	264.22	975.52489	-1.79
dWTIEQITR	P0A853from242to255	0	3.45	384.49	1305.6939	-1.48
eAEYKDWTIEQITR	P0A853from242to255	0	3.18	240.8	1925.97763	0.62
eGFPTYGGLEGGAMER	P0A853from302to317	0	5.07	585.78	1814.85477	0.47
dGMNLDWLAYR	P0A853from324to334	0	3.09	156.88	1497.72881	-1.85
dSGTGAVTQSMQAAMMR	P0A853from53to69	0	5.73	648.44	1885.86509	-4
dIGAQYIIIGHSER	P0A858from71to98	0	5.48	770.01	1715.9238	0.11
dLPFAQSR	P0A862from86to93	0.00154	2.36	275.5	1077.58293	-1.75
eAGVFLISPFVGR	P0A867from168to180	0	2.95	349.59	1535.88078	4.31
dATTNPSLILNAAQIPEYR	P0A870from31to49	0	5.39	725.8	2231.19358	4.85
dDLQAVMAMVR	P0A8E7from138to148	0	2.61	288.2	1392.70928	-3.04
dGVLENVPSAR	P0A8F0from85to95	0	2.68	336.99	1300.70122	-0.34
dAIIFGTPTR	P0A8G6from70to79	0	2.6	373.04	1234.6967	1.29
dLEVWIPAQNTYR	P0A8L1from342to354	0	3.32	422.67	1748.91606	1.92
dLENLPR	P0A8M0from183to189	0.00372	1.52	217.47	1000.55535	-2.92
dVLAPGIGEIIGGSQR	P0A8M0from381to396	0	4.01	293.4	1725.97062	2.96
dVLEFVR	P0A8M6from10to16	0.00154	2.53	367.42	1021.58208	-1.64

dLIELTESLFR	P0A8N3from289to299	0	2.09	198.77	1479.81828	-2.13
eAHLIQPTFITEYPAEVSPLAR	P0A8N3from381to402	0	4.82	470.63	2626.41721	5.16
eFFIGGR	P0A8N3from415to421	0.00284	2.17	192.94	969.52733	-4.12
dITGGLPR	P0A8T7from1129to1140	0.00154	1.82	143.47	972.56133	-2.09
dFLEGEQVEYSR	P0A8T7from1273to1284	0	2.55	203.73	1615.76409	-7.34
dNLQTETVINR	P0A8T7from699to709	0.00154	1.99	61.36	1446.76933	-1.01
dGSIIETPITANFR	P0A8T7from751to764	0.00154	2.68	185.63	1677.90422	4.46
dIVLNPLGVPSR	P0A8V2from1095to1106	0	3.39	506.87	1423.84697	2.9
dNLFVR	P0A8V2from181to197	0.00154	2.09	212.32	907.5128	-3.16
eYGFLETPYR	P0A8V2from583to592	0.00284	1.57	73.39	1418.70903	-1.48
dSILVSER	P0A8V2from814to821	0	2.33	180.48	1062.58928	-5.43
dVQVFTR	P0A8V2from930to936	0.00284	2.51	128.55	1008.56151	-1.82
dTGVSPVFAGGVEYAITPEIATR	P0A910from137to159	0	4.85	513.83	2494.30332	1.91
dGFVIAGGGGMVVVEELEHALAR	P0A953from227to249	0	4.5	390.65	2470.29197	-0.04
eIAAIR	P0A953from308to314	0	2.02	374.56	816.50719	-3.27
dIIPETLHQR	P0A993from299to308	0	2.48	188.15	1365.76398	-0.44
dEFETVGNTIR	P0A9A6from273to283	0	2.47	297.3	1424.71623	-1.03

dMGEEIGLATVYR	P0A9A9from45to57	0	2.47	172.98	1597.80583	0.43
dQAAFLGASIGR	P0A9C3from64to75	0	2.63	347.69	1349.73381	0.38
dAYIALR	P0A9C5from441to447	0	2.48	210.55	965.5584	0.9
dQSLYPANSVPAVVER	P0A9G6from108to123	0	2.83	292.9	1888.99407	0.87
dVTGVPTLLVAR	P0A9G6from221to232	0.00372	3.02	236.51	1384.8405	6.18
dELQVGMR	P0A9K9from88to95	0	2.56	202.32	1091.56462	-2.61
dVLNIFR	P0A9L8from146to152	0.00593	2.11	114.89	1020.59966	-0.08
dSAAAFGIEPR	P0A9M8from575to585	0	3.85	496.69	1277.6646	0.04
dAEVVLVEGLVPTR	P0A9M8from65to114	0	2.82	870.47	1640.94036	1.5
dLISIGPMLQGMR	P0A9M8from672to684	0	2.66	216.09	1574.85381	-1.03
dDIVYTIALTAIQSAQQQ	P0A9M8from697to714	0	3.33	269.71	2122.12369	2.33
dAVLVAIGR	P0A9P0from265to273	0	3.08	569.02	1057.6519	-0.58
eTATFPWAASGR	P0A9P0from375to386	0	2.98	364.62	1437.72685	-0.94
eQANVALMFLTGR	P0A9Q1from71to83	0	4.16	697.47	1593.85343	-2.78
eLTPAAVTGTLTTPVGR	P0A9Q9from315to331	0.00154	2.55	180.57	1828.03935	3.16
dLVVEIPR	P0A9R4from95to102	0	2.56	418.47	1084.65141	-0.7
dEAIINYVR	P0A9X4from195to203	0	2.7	253.87	1236.67082	-2.86
mFQQEVTITAPNGLHTR	P0AA04from1to17	0	4.82	396.12	2087.09031	1.89
dNLAFPLR	P0AAZ7from32to39	0.001	2.48	235.6	1089.622	1.62

		54		2	97	
dGIPVLLETEAR	P0AAZ7from40to51	0.00154	2.16	211.66	1456.81426	-1.68
dSLFWGEQTIER	P0AB14from35to72	0	3.03	360	1624.80889	-2.33
dTLAGIEATGVTQR	P0AB61from26to39	0	3.6	593.13	1575.85051	0.46
dGVLFTPPFTSSALPGITR	P0AB80from187to223	0	2.91	166.12	2120.14212	-5.96
dEVILPYWR	P0AB89from140to148	0	2.1	549.33	1334.72502	-1.03
dGFSAVEAIAPGVIER	P0ABB0from124to139	0	4.7	615.21	1774.95317	2.05
dSMIPIGR	P0ABB0from154to161	0.00372	1.68	80.55	1032.56694	0.2
eVQQQLGGGIVR	P0ABB4from42to53	0.00284	3.23	621.49	1427.81357	0.67
dMMIEILR	P0ABD3from118to125	0	2.09	232.03	1164.62627	-1.18
eAMGIIAPR	P0ABD5from249to257	0	2.62	300.86	1101.62346	-1.03
dVFIAGVGTGGTLTGVSR	P0ABK5from170to187	0	5.13	773.93	1851.02231	4.93
dNSLTIGHTPLVR	P0ABK5from7to19	0	4.17	301.24	1566.87668	0.48
dLVYQAIR	P0ABQ2from199to206	0.00372	1.84	155.55	1121.65056	2.81
dILTAASR	P0ABT2from146to153	0	2.61	423.61	990.5714	-2.55
eMLQNSPMALR	P0ABU0from220to230	0	2.17	223.38	1433.73699	-2.14
dVMNFLFNV	P0ABU2from355to363	0.00372	2.42	603.21	1242.63457	-0.2
eAVLTLLAISR	P0ABU5from21to31	0	2.97	322.95	1329.82414	-1.49
eILAQLSQYPVSTR	P0AC33from369to382	0	3.27	586.52	1748.97869	4.84

dAVPMTLGQEFR	P0AC38from194to205	0.00284	2.03	194.36	1507.77324	-0.14
dIPEFVR	P0AC38from43to49	0.00372	2.17	224.04	1019.56841	0.29
dLVVFGR	P0AC41from407to413	0.00372	2.01	207.16	949.56438	1.84
dDIQLFPLLR	P0AC59from171to180	0	2.78	344.83	1373.79729	1.79
dIVIEMYQR	P0AC69from85to93	0	2.34	257.96	1310.69328	-0.08
dVPLISVETLQR	P0ACC7from105to116	0	2.86	273.67	1513.88017	3.72
dDLVTAFR	P0ACE7from23to30	0.00154	1.78	208.96	1080.58574	1.16
dAVQLVGFGTFK	P0ACF0from40to55	0	2.61	305.76	1425.79973	7
dEIQSTETLIVLQNPIMR	P0ACI6from131to148	0	3.75	406.39	2244.20743	0.39
dVIAPVEVR	P0AD10from30to38	0.00154	2.6	248.13	1141.67302	-0.55
eMLIFQLGLR	P0AD33from85to94	0	2.73	284.59	1363.79473	1.46
eITPAIR	P0AD49from11to17	0.00154	2.3	250.82	943.56999	-3.39
dANFVEEVEEE	P0AD49from98to113	0.00154	3.1	949.96	1453.6458	-2.21
dFVAASFIR	P0AD61from187to195	0.00154	2.42	302.33	1169.6491	1.42
dVIEIR	P0AD61from199to204	0.00593	2.17	447.1	888.52855	-2.74
dGFAWIER	P0ADE8from80to87	0	1.98	160.77	1137.58452	-0.27
eENELVGIITGR	P0ADG7from126to137	0	3.8	409.6	1473.81035	2.37
dLQIIGGNVATAAGAR	P0ADG7from270to285	0	5.31	877.21	1670.93442	-0.07
eIIHQQMGG	P0ADG7from429to4	0.001	2.92	236.1	1425.777	-0.85

LR	39	54		7	99	
dPQTVLPTTTLR	P0ADG7from96to107	0.00154	2.23	249.36	1485.84502	1.2
dMWGPGLTR	P0ADI4from93to101	0.00372	1.68	144.54	1176.59734	-1.51
dIQIFGR	P0ADS2from7to13	0	1.87	230.73	992.567	-1.45
dDTWVTLR	P0ADU5from57to64	0	3.13	251	1149.60527	-0.59
mIQEQTMLNVADNSGAR	P0ADY3from1to17	0	4.59	390.14	2021.98349	-3.43
dGVPEELAR	P0ADY7from106to114	0	2.66	439.77	1129.59929	-1.41
dPQGIIQAIEVTAEGIGR	P0AE08from126to143	0	4.21	594.49	2011.10556	3.77
dAIFFVTGR	P0AE22from131to139	0	2.19	302.83	1169.6491	1.42
dTVLFSSPGFWR	P0AE22from72to83	0	2.91	777.62	1555.80889	1.56
dELLALIR	P0AE78from35to42	0	2.31	194.51	1086.66875	0.85
dMLIVPLR	P0AED0from133to140	0	3.12	367.95	1100.66448	-1.14
dDNFMSVVR	P0AEE5from36to44	0	1.99	201.81	1226.59941	-0.07
dIIETNLSSVFR	P0AEK2from105to116	0	2.98	520.61	1537.83623	-1.25
dILVNNAGITR	P0AEK2from81to91	0	2.65	473.77	1329.76555	0.72
dVAPSNLAIVGR	P0AEP3from206to217	0	2.76	766.86	1355.7824	1.58
dESFQPTAVGFAEALNNK	P0AES9from46to110	0	5.59	598.34	2082.02982	-0.07
dLTQIPK	P0AET2from76to108	0.00907	1.63	395.18	958.57231	-0.56
dEAIITTNPEVSSVR	P0AEZ3from137to151	0	3.67	806.46	1774.93608	1.03

sQVSTEFIPTR	P0AEZ9from2to12	0	3.35	342.22	1408.76079	1.16
eALTPVLAASLR	P0AF28from132to143	0.00702	2.22	152.46	1384.82744	-3.26
eAFFTEHNATFPAR	P0AF93from92to105	0	5.41	547.36	1781.87693	0.14
mIIGNIHNLQPWLPQELR	P0AF96from1to18	0	5.22	618.44	2316.27518	-2.37
eIVSPVVIGSVR	P0AFG3from324to335	0	2.83	338.98	1398.85088	2.34
dALSPAIR	P0AFG6from112to119	0.00372	2.49	316.72	986.57707	-1.96
dVSMAVSTPR	P0AFG6from268to277	0	2.78	335.33	1206.62871	-1.73
dWLQAIESVIR	P0AFG8from16to26	0	3.89	720.66	1473.82195	-0.11
dEQIWALNR	P0AFG8from354to362	0	2.67	296.33	1288.6729	-5.91
eISTTIAFVR	P0AFG8from483to501	0.00154	1.84	342.66	1280.73686	-0.08
dVYSVTSFTELAR	P0AFG8from755to767	0	2.53	382.61	1631.83904	-2.81
eETFQTLLVSR	P0AFM2from40to50	0	3.67	403.56	1466.8029	1.28
dTIAILR	P0AFQ7from137to143	0.00154	2.29	391.04	945.58769	-1.21
mQLNITGNNVEITEALR	P0AFX0from1to17	0	4.16	394.74	2060.11191	7.43
eQIIEMAR	P0AG16from377to384	0	2.24	330.2	1133.61211	-2.04
dSLYSIVQMPR	P0AG18from102to112	0	2.74	261.14	1452.77287	3.62
eNMGLENLAR	P0AG30from19to28	0	2.16	149.98	1290.66326	0.07
nMAGSLVR	P0AG44from23to30	0.00593	2.45	160.6	991.54943	-2.01
aTLLGLGLR	P0AG51from20to30	0.001	2.26	197.1	1057.686	-2.32

dNTLTFGPR	P0AG55from47to55	54 0	2.66	4 284.26	45 1164.61589	-0.82
dVVEIR	P0AG63from57to62	0.00154	2.49	375.73	874.51286	-2.83
dGVEGYLR	P0AG67from472to479	0	2.84	376.73	1052.55229	-0.84
dGFGETLLSR	P0AG67from77to86	0	2.95	391.07	1238.65398	0.27
dDVYEVVLR	P0AG86from53to61	0	2.27	242.35	1251.67497	0.76
dYTVAIGNPFGLGETVTSGIVSALGR	P0C0V0from188to213	0	3.36	169.26	2738.45084	-0.46
dASHLPFAQNISR	P0C8J6from104to116	0	4.1	998.95	1599.84079	0.57
dALYTNPAQAR	P0C8J6from151to161	0	3.28	535.11	1363.70964	-2.12
eAIFALAQIEQELIAPENR	P0C8J8from290to308	0	3.28	238.42	2299.24015	-2.28
eHILLGR	P0CE47from118to124	0.00154	2.55	207.72	981.5977	-2.41
dMVDDEELLELVEMEVR	P0CE47from139to155	0	4.32	459.31	2208.06003	1.28
eLVEMEVR	P0CE47from148to155	0.00154	2.03	218.21	1148.60881	-4.59
dSYIPEPER	P0CE47from197to205	0.00593	3.09	681.68	1249.6198	-1.75
aIDKPFLLPIEDVFSISGR	P0CE47from206to224	0	2.57	169.39	2261.27927	4.03
dVFSISGR	P0CE47from217to224	0	2.61	469.85	1024.55681	-1.42
eVYILSK	P0CE47from308to319	0.00372	2.32	257.37	995.59484	1.6
dDFIEALFAR	P10121from486to495	0	2.18	243.36	1340.69524	-4
dVLVTTASVR	P12758from106to1	0.003	2.46	392.7	1204.705	0.26

	15	72		6	98	
dTFYPGQER	P12758from160to168	0	2.02	184.27	1256.60454	-1.69
eIAQLGLR	P12758from80to87	0.00154	1.71	79.29	1043.6397	2.72
dLVFGSNSVLR	P13029from681to691	0.00284	2.63	300.03	1350.75871	3.72
dVAAPYIATGAR	P15254from1028to1039	0.00284	2.56	272.02	1348.73796	-0.06
dANSPVMHLVR	P15288from402to412	0.00154	2.46	254.86	1382.73236	-3.35
dQVGNILIR	P15288from50to58	0.00154	2.81	148.1	1171.69304	-2.05
dIGGPTMVR	P15639from127to135	0.00372	2.26	189.57	1089.58708	-1.03
dDNEGSLTLATR	P15639from161to172	0.00372	1.66	52.17	1435.71929	0.6
dPTSAFGGIIAFNR	P15639from311to324	0	4.93	794.51	1609.84751	-1.18
dAETAQAIISR	P15639from327to337	0	2.38	90.04	1318.70769	-3.44
dLGMVGAEELR	P15639from392to402	0	3.27	473.14	1333.68596	-6.14
dAFFPFR	P15639from476to482	0.00154	2.28	263.05	1043.5468	-0.17
dMVVVNLYPFAQTVAR	P15639from99to114	0	3.36	387	1967.05716	-0.44
eYQNFTEVEPALAYLR	P15640from121to136	0	4.84	444.76	2087.05552	-2.39
dAPEPIFGLAVTGIVPTER	P16456from143to161	0	4.65	914.42	2126.16922	1.82
dPNLLVGNETR	P16456from39to49	0.00154	2.04	185.21	1371.7365	-1.66
eQYGPELLR	P16659from87to95	0.00372	2.2	284.76	1248.66484	-7.62
eILLEGLR	P17169from15to22	0	1.59	147.18	1086.66496	-2.64

dQLALLPVTR	P17169from193to202	0	3.09	373.07	1269.7741	4.32
dSLFTSPEVAR	P19926from271to281	0	3.43	404.13	1365.7177	0.53
dALTLQAPAQR	P19926from373to383	0	3.26	463.18	1327.74883	-0.08
dGIGSLLPAAR	P21165from113to123	0	2.84	314.09	1213.70476	-1.04
dYLAVSFPR	P21599from191to199	0	2.85	630.03	1211.65593	-1.69
dVVMVAR	P21599from244to250	0.00372	1.85	126.32	933.53325	-1.56
dGVAAASEAVNLLR	P21599from433to446	0	4.24	305.93	1529.84314	-0.77
dYLVPSR	P21889from175to181	0.00593	1.59	68.25	993.5576	5.2
dVETSFMTAPQVR	P21889from233to245	0	3.14	308.4	1624.81865	1.6
dAYLIGR	P22106from128to138	0.00154	1.9	156.1	951.541	-0.92
eAYLPASVAWR	P22106from438to448	0.00154	2.75	313.02	1406.75749	-0.91
dALLENVTVR	P22259from307to316	0.00284	2.22	195.63	1273.73638	7.25
dAFGVLPPVSR	P22259from363to373	0.00154	2.88	390.39	1301.74419	5.28
dTGIFTGR	P22259from58to65	0	2.34	357.1	1010.5421	-0.51
eYTVAAPAAGLSR	P22333from290to302	0	2.7	227.86	1449.7874	1.16
dVSLPILVLTAR	P23836from70to81	0	2.54	340.63	1440.89995	3.75
eAPAPGITPELR	P24182from242to253	0.00154	2.41	127.91	1394.77434	-4
dEIGLPR	P25553from193to199	0.00372	1.92	147.86	943.53087	-6.29
dVVNPATEA	P25553from25to37	0	3.07	525.5	1514.834	0.49

VISR					15	
eYLQTQVVYLQS	P25553from468to479	0.00154	2.75	525.18	1614.85161	-1.13
tMYATLEEAIDAAR	P25738from2to15	0	4.63	749.77	1698.85527	1.44
eGIITSGTFSPTLGYSIALAR	P27248from309to329	0	5.01	540.78	2298.25522	2.22
dAEIALTR	P27302from267to274	0.00702	1.98	176.68	1032.58245	-1.98
dLTWEAFR	P30136from309to316	0.00154	1.52	199.67	1181.60954	-1.26
dYPTVAIGGISLAR	P30137from158to171	0.00907	2.03	93.15	1576.89287	4.71
dSVQWIER	P30137from20to27	0.00154	2.08	220.72	1176.61601	-0.71
tQLNPDIQLTALQQR	P30138from92to106	0.00372	2.57	84.76	1883.05057	-0.02
eQVLAAISLVR	P30871from249to262	0.00154	2.18	200.38	1342.80815	-9.85
dFQQLALIR	P31663from149to160	0	2.17	220.96	1247.73198	4.2
mLIIETLPLLR	P31663from1to11	0	2.65	263.99	1455.9168	2.71
dALAKPYR	P31678from68to75	0.00593	1.61	111.88	1077.62029	-0.85
dYNSYVGVIGIGR	P32132from213to225	0	2.69	281.94	1556.82121	-1.03
dEAVVLVPPIR	P32132from552to562	0	3.64	636.58	1351.81548	3.69
dELVEVTPTSIR	P32132from575to586	0	2.93	472.13	1502.81658	-3.72
tQTLSQLENSGAFIER	P33195from2to17	0	5.01	747.84	1938.0029	-3.05
dAFPVLYTGR	P33195from798to807	0.00284	2.22	310.71	1282.69707	1.54
eELQLPTSTYR	P33221from123to133	0	2.23	167.9	1480.78728	4.73

eSVLPSNVAAR	P33570from597to607	0.00284	2.1	146.84	1286.71404	-6.49
dVVQALNLMSVLNPR	P35340from134to148	0	4.25	695.71	1813.0206	2.45
dVLIVGSGPAGAAAAIYSAR	P35340from214to233	0	3.77	378.99	2003.11374	2.79
mLEEYR	P36683from1to6	0.00372	1.51	111.5	984.49822	0.25
eQVIAWVR	P37330from164to171	0	1.97	225.59	1144.66631	2.53
eLGYLVPQPER	P37330from87to97	0	3.11	536.9	1444.79778	1.54
dYLYLSGISLAILSPTSR	P37647from129to146	0	3.79	499.76	2113.16837	-0.8
dSFSAGYLAVR	P37647from264to274	0	3.47	550.33	1329.69451	-1
dTLNTSVYIAR	P37647from29to39	0	3.31	328.52	1396.75896	-0.14
dGYYLEPTILFGQNNMR	P37685from387to403	0	3.11	134.77	2175.06351	-3
dALIFMNFR	P37689from255to263	0	2.21	260.51	1270.67583	-1.2
dGGWLAR	P37769from247to253	0	2.12	137.96	918.49278	-2.72
dIVGINIVEPTETIEQVTALGR	P37769from36to57	0	5.5	885.69	2511.38633	1.48
dILVNNAGLIR	P37769from87to97	0	3.4	770.92	1341.8051	3.07
dITTYLLGSNAAAVVR	P37903from119to134	0	2.69	446.86	1808.01055	1.77
dFAVAVGNPFGLGQTATSGIVSALGR	P39099from166to191	0	2.84	180.03	2649.43283	6.48
dVVVIGSR	P39177from107to114	0.00154	2.01	149.63	988.59233	-2.35
dIEPGLVGGT	P39831from174to1	0	3.55	358.7	1833.952	1.22

EFSNVR	89			1	49	
dVLGGAPGIYSAR	P52061from75to87	0.00154	2.33	244.78	1419.77092	-2.98
sAGTYVQIVAR	P60422from157to167	0	2.65	267.74	1308.74602	2.22
dGAYVTLR	P60422from168to175	0	2.24	179.27	1038.57158	-2.24
dGIPAVVER	P60422from72to80	0.00154	2.67	553.06	1099.62517	-1.39
dGVSIPVTVIEVEANR	P60438from18to33	0	5.81	383.35	1842.01554	1.46
dENLFLAAR	P60723from154to162	0	3.5	451.45	1192.64714	-0.85
dFNEALVHQVVVAYAAGAR	P60723from22to40	0	4.88	557.58	2174.14396	-3.43
dAQSALTVSETTFGR	P60723from7to21	0	3.75	376.27	1726.87131	-3.14
dNYLGVVSLIR	P60785from201to211	0.00154	2.28	177.34	1392.80303	1.72
dVVLISAGVAR	P61889from71to81	0	3.18	543.2	1243.75798	4.03
eTILPR	P62707from167to172	0.00154	1.57	195.68	872.53575	-0.36
dSLGFQPNLR	P65367from63to86	0	2.37	272.66	1290.69731	0.89
mITGIQITK	P68066from1to30	0	2.67	609.5	1148.6873	0.4
eVPVEVKPEVR	P68066from56to66	0	4.09	434.03	1424.8306	2.62
vKVGDTVIEFDLPLLEEK	P69783from114to166	0	2.58	188.95	2188.23246	2.37
dELVALAVETGR	P69797from112to123	0	3	444.1	1416.7863	0.64
dTWGGSPFNAASR	P69797from67to79	0	2.41	203.59	1509.71636	-5.19
mTNLFVR	P69828from1to7	0	2.32	296.3	1024.575	-1.62

dTWPQALIAR	P69828from34to43	0	2.8	5 274.58	24 1314.73357	0.77
dYLGGQIGTFAINFSR	P69908from304to319	0.00154	2.19	124.31	1902.99504	4.25
dIVVMR	P69908from417to422	0.00154	1.84	192.24	876.51073	-2.86
ePMFFGQPVNVAR	P69924from16to28	0.00154	1.97	151.37	1635.85503	4.72
qFNVNVLTGSDGVLR	P75726from377to391	0.00907	1.89	67.6	1762.94382	-9.6
dVWVPVVESLR	P76014from132to142	0.00154	2.32	236.45	1442.81316	-2.18
sQQENALWLATSQSR	P76116from49to63	0	2.75	127.81	1862.95085	-0.41
eNALWLATSQSR	P76116from52to63	0	3.68	245.92	1519.80384	0.93
dFVYSWQR	P77348from114to121	0.00284	1.99	113.2	1244.61797	-3.18
dGFQGPQPEEAIR	P77395from91to103	0.00154	2.1	118.36	1587.78728	-3.15
eMMMEGLYGR	P77454from237to246	0	1.95	109.17	1360.6165	-3.97
vWLANPER	Q46851from2to9	0	2.09	308.97	1128.63127	-0.74
dVQTQVLLR	Q46868from53to61	0	2.59	198.74	1215.72209	0.37
dALQLAGR	Q57261from223to230	0.00284	2.05	255.67	987.58129	7.11
dQFPAQALAcELYK	P0A853from374to392	0	2.74	295.94	1797.90178	0.93
dAAAAAGVTcVIQPGGSIR	P15639from486to504	0	5.31	580.22	1958.03373	2.65
dFFQQIQLTQcTR	P0AC33from533to545	0	3.58	633.13	1828.9135	-1.99
dMLIQLEEEGLNVVPcAR	P33221from89to106	0	5.31	849.53	2230.14219	2.43

dcLIWAIGR	P06715from255to263	0	2.34	192.96	1247.66948	-2.5
dcLGIIGVAR	P07395from168to177	0	2.95	354.98	1217.68364	0.38
ecAFFSPLSITGR	P07395from334to346	0	3.12	432.13	1628.82829	1.28
dcQLFTQNLAR	P0A6J5from199to209	0	2.19	158.54	1509.7636	-0.22
dcYLAPLLWR	P0ACA3from158to167	0	2.34	200.69	1450.76506	-1.49
eScGVEIGIR	P0ABS1from115to124	0.00154	2.05	113.31	1263.64934	-2.31
iVcLIAAQIIAR	P39276from241to252	0.00372	1.83	64.92	1484.90732	-4.66
eVcWLTPEEITAR	P65556from137to149	0	2.66	304.65	1747.88701	1.46
dIVLcGFEYGR	P0A705from595to605	0	2.7	536.01	1472.7376	0.86
dGNAcVLLNNNSEQPIGTR	P0ADY3from80to98	0	5.18	849.09	2216.08738	-0.53
eEVGcAAVMPLGAPIGSNQGLETR	P30139from141to164	0	4.55	413.47	2600.30362	2.59
dLIMGcER	P0AEZ3from47to54	0.00372	1.67	179.6	1137.55217	-2.66
dMVPVEcVVR	P0A7D7from85to94	0.00593	2.27	262.89	1347.68596	-4.51
dEGIPAVcFK	P69908from372to398	0.00154	2.52	300.34	1279.64885	-1.84
dIVSNAScTTNcLAPLAk	P0A9B2from143to195	0	3.99	551.31	2121.05765	-0.55
dNLVFVINcNLQR	P0AFG8from252to264	0	4.6	1062.51	1748.92746	0.08
eEGLNVVPcAR	P33221from96to106	0	3.02	378.64	1387.7072	-6.29
dINEAIALLk	P0A7L0from22to53	0	2.93	324.58	1285.76702	2.71

eQAVEAVLAk	P0A850from405to432	0	2.3	210.65	1243.72161	4.04
dVSFGSFGLk	P0ADY7from25to38	0	2.29	377.99	1242.66008	-3
dDEVIVLTGk	P60624from8to22	0	3.14	447.41	1274.71184	0.54
dLVESAPAALk	P0A7K2from76to121	0	3.18	796.53	1299.74211	-0.53
eEQAVEAVLAk	P0A850from404to432	0	2.99	264.51	1372.76604	5
eLPLAAGALAk	P61714from59to84	0.00154	2.39	248.29	1239.76811	8.11
dLAVGTAAGQIkTGSMSR	P0A6P9from382to399	0.00702	2.5	68.74	1949.04819	4.5
eEFSLLPVVNYLk	P0A799from70to90	0	3.62	438.34	1736.97617	1.11
eFPEPVISIAVEPk	P0A6M8from410to437	0.00154	2.49	152.5	1740.97331	2.37
eTMPGWSESTFGVk	P0A7D4from374to389	0.00593	2.03	114.19	1741.8261	-6.56
dTITLLQMEIEELk	P0AF36from20to50	0	2.56	154.33	1862.00846	-0.86
dLGVEIPVEEVIFAQk	P0AD61from246to269	0	3.24	162.9	1972.08432	-3.44
dESFQPTAVGFAEALNNk	P0AES9from46to110	0	7.53	959.51	2124.05326	0.71
dkAGIVEFAQALSAR	P15639from16to30	0	3.81	410.24	1761.97574	-0.56
eTQkSTcTGVEMFR	P0CE47from241to263	0.00372	2.38	122.53	1859.89048	0.28
dILGkFPLPVEVIPMAR	P0A7Z0from124to140	0.00154	2.36	78.54	2081.22342	6.38
dDFLIkQ	P0ACF8from131to137	0.00154	2.33	266.04	1064.58867	-0.84
dIVLPkPR	P75897from128to135	0.00372	1.9	153.27	1123.7072	-3.11
iLFVGTk	P0A7V0from67to74	0.003	1.82	165.0	963.6122	-2.52

		72		4	3	
dLSLITk	P0ABT2from42to55	0.00154	1.82	84.24	975.60301	3.71
dDLYVFk	P0ADU5from73to93	0	2.53	198.69	1085.57915	0.45
dLTQIPk	P0AET2from76to108	0.00154	2.02	94.41	1000.59343	-1.22
eVYILSk	P0CE47from308to319	0.00154	1.44	127.29	1037.61467	-0.35
mIGLVGk	P60438from1to13	0	2.38	260.03	903.56279	2.51
eALLVALk	P0A7L0from190to234	0	1.78	343.33	1042.67754	-0.44
nFLVPQGk	P0A7R1from28to50	0.00702	1.72	129.26	1088.6353	-1.73
dPWVAIAkR	P0AG67from265to273	0	3.78	453.05	1241.72669	-0.59
dILTYTNk	P61175from34to84	0.00154	1.55	190.68	1153.63591	-1.14
sLLNVPAGk	P0A7A9from2to44	0	2.46	195.4	1084.66704	3.35
dEVIVLTGk	P60624from9to22	0.00593	1.62	158.28	1159.68706	2.46
mITGIQITk	P68066from1to30	0	2.35	168.52	1190.70867	0.02
dQILLFQVIAGG	O32583from55to66	0.0357	1.87	245.55	1417.82451	2.44
dPQQIQQIFNQY	P00805from337to348	0.0472	2.72	527.16	1665.83904	-0.09
dELLPVAR	P00864from66to73	0.021	1.87	193.2	1056.62322	2.21
dIVSLR	P00946from54to59	0.017	1.53	113.81	846.51708	-3.95
dFMLGR		0.0193	1.53	112.38	882.46672	0.49
dISSVISR	P04982from10to17	0.0121	1.94	97.09	1020.58422	-0.26

mLNPIVR	P05055from1to7	0.021	1.76	104.65	986.59612	-1.54
dVGTPTVWAAR	P07003from379to389	0.0283	1.6	56.81	1316.7116	-0.18
eMLPEVR	P0A6F3from213to219	0.0487	1.54	280.94	1017.55583	-0.01
dILAIVEA	P0A6F9from90to97	0.021	2.39	333.3	987.58702	-1.17
dGSLQGMLR	P0A6J8from113to121	0.0141	2.03	73.41	1120.58953	-4.01
mQAIPMTLR	P0A6W5from1to9	0.0131	1.68	243.82	1204.67046	0.24
dIESLPFLEAIR	P0A7E5from147to158	0.0381	1.72	52.11	1546.87078	4.61
dQAGLSASVN	P0A7L0from225to234	0.0381	1.6	206.84	1105.55925	-4.73
qLWIAR	P0A7L3from59to64	0.0358	1.72	133.24	930.56584	-2.36
dWLAYR	P0A853from329to334	0.0179	2.11	258.9	967.51036	-5.47
eLVSNLR	P0A858from21to27	0.0187	2.16	404.67	974.57646	-2.6
fGHELAQALR	P0A8C4from60to69	0.0488	2.09	161.06	1285.71611	-0.88
dVIPFPR	P0A8M0from453to459	0.0446	1.83	197.4	987.57244	-5.92
ePLFSNFGGR	P0A8R0from20to29	0.0357	1.56	101.69	1267.65508	-3.14
eLLNAGLGGSDNE	P0A8T7from1395to1407	0.0472	2.72	490.88	1432.70647	-0.75
eVLASLR	P0A953from21to27	0.0121	1.91	174.23	931.57512	2.07
eLEIVWNNIK	P0A9D4from5to18	0.0187	1.52	116.49	1401.80262	9.19
dPTLSIVR	P0A9Q7from519to526	0.0357	1.51	125.75	1044.62822	7.04
dIMAVGR	P0AAB6from208to2	0.021	1.59	115.5	905.4980	-5.87

	14			6	9	
dIPLLSR	P0AB89from163to169	0.0131	1.67	165.2	957.58855	-0.3
eMISLPGNR	P0ABB0from51to59	0.017	1.69	76.7	1160.62663	1.13
dATVVLSR	P0ABB4from317to324	0.0193	1.55	124.07	1004.58898	-0.59
dEPLVVIE	P0ABD8from149to156	0.017	2.2	617.74	1057.5949	1.17
mFPEYR	P0ACW6from1to6	0.021	1.86	143.39	986.49291	0.42
eQLQNILR	P0ACY1from24to31	0.0488	1.66	117.63	1157.68108	1.12
dSLSTPIGK	P0AD99from71to101	0.0431	1.56	19.35	1061.59441	-5.08
eLVGIITGR	P0ADG7from129to137	0.0209	1.54	197.01	1101.67888	0.14
dQIAELLR	P0AES4from483to490	0.0472	1.53	82.75	1101.64141	-0.85
eLNLNAVR	P0AF12from135to142	0.021	1.5	131.32	1072.62786	0.78
dDPEAVAFVTR	P0AFG3from425to435	0.0431	1.91	196.33	1363.70342	1.53
eVIPFGASLR	P0AFG3from891to900	0.0193	2.41	138.98	1232.71526	-0.48
mDLSQLTPR	P0AFZ3from1to9	0.0488	1.62	95.93	1204.65032	-1.01
dGFGFLR	P0AG30from60to66	0.0187	1.99	165.12	955.51378	-1.99
aISLSVR	P0AG67from514to520	0.0372	1.76	208.85	889.56188	-0.84
eNVGVLLR	P0CE47from273to280	0.0149	1.77	120.23	1043.63603	-0.79
gSPAMNLLTGR	P10907from236to246	0.0385	1.79	152.16	1260.68657	-1.92
dALTFELPEIIDLR	P11454from75to88	0.0357	1.75	32.24	1788.98252	-4.36

eLIPGSELWPR	P15254from1146to1156	0.0274	1.53	212.4	1440.79802	-1.82
dINPLYR	P18843from180to186	0.0458	1.51	78.26	1034.58049	1.45
tLGLPFIR	P19624from285to292	0.0488	1.67	99.95	1060.66326	-3.93
dVIIESVTVSE	P23869from154to164	0.0472	1.97	219.59	1334.72197	0.69
eLIPGGVNSPVR	P23893from13to24	0.0358	2.11	270.25	1381.79192	-2.88
dEIEIR	P25738from81to86	0.0149	1.64	134.01	918.50328	-2.05
dVDLLILSR	P27249from105to113	0.0274	1.56	14.05	1187.70622	-7.82
eLANAIR	P27302from6to12	0.0179	1.71	85.03	930.55181	-1.06
eFIAIR	P30136from163to168	0.021	1.87	191.37	892.53972	-1.61
dAIFFGR	P31142from220to226	0.0193	1.71	176.1	969.53227	0.97
iSASSLEGISYALDSTDILR	P31456from278to297	0.0357	1.8	38.71	2255.1771	-6.9
eIHLPVELR	P31665from28to36	0.0141	1.62	119.98	1249.74103	-1.09
dIVFGSAQR	P33195from264to272	0.0113	2.13	191.28	1136.61809	-3.39
dMPESVLVR	P36683from527to535	0.0426	1.7	76.74	1189.62944	-9.41
dAAAFWR	P37330from31to37	0.0488	1.61	67.44	980.50981	-1.14
hPEIPSLSLYSADIAFQR	P45763from205to222	0.021	1.76	51.19	2188.17258	7.66
dSEIPVAFGVLTTESIEQAIER	P61714from107to128	0.0357	1.92	118.11	2548.33891	3.4
eEGGYLTEAVR	P63284from657to667	0.0121	1.55	76.58	1367.69914	2.13

dPTQLAIEFLR	P64476from6to16	0.0193	1.5	121.01	1446.80962	-1.1
eTLPNTMFR	P69222from15to23	0.021	2.27	149.63	1252.65178	0.2
dIVTNEQIQK	P69924from140to150	0.0238	1.97	108.32	1331.72038	-9.18
vGIGGLEGR	P75745from130to142	0.0357	1.29	88.63	1001.58171	-8.19
dALWQLTQEVQA	P76423from251to262	0.0472	2.72	388.98	1545.80998	2.03
lNATLWWLNWFDGR	P76585from635to648	0.0464	2.22	155.24	1936.01474	6.3
eEILNAVSLAWVDGQSLR	P77243from192to209	0.0385	1.57	89.91	2144.15879	3.75
gEFLcKQGNEFGATTGR	P0A7D4from288to304	0.0121	1.65	26.8	2015.97672	0.12
dILIcR	P0A7E5from206to211	0.0113	1.86	134.28	933.53404	-0.7
dLGLVIAcLPYA	P0A7Q1from54to65	0.0357	2.23	389.54	1448.79936	1.04
eLAVVESFPTkIEGR	P0A707from154to168	0.0187	2.63	77.07	1861.03855	2.49
sFELPALPYAk	P0AGD3from2to58	0.021	1.62	101.83	1421.8001	3.71
dAVTAAGVEVAkSEVR	P0A7R1from101to116	0.0464	2.38	68.28	1787.97873	0.89
dASYQQAVNLLPEEkR	P0AF50from97to112	0.017	2.27	48.24	2047.0781	2.58
dkLVPEGIEGR	P0ADG7from410to420	0.0193	1.89	57.24	1398.78431	-1.26
dkSFGAPTITk	P0A6F5from41to58	0.0357	1.82	53.74	1392.77579	0.22
dPkSPFVTSGIR	P0A825from352to363	0.0141	2.19	128.13	1489.82773	-0.35
dEGkLFGSIGTR	P0A7R1from86to97	0.0446	2.45	77.07	1465.7923	0.29
dLFLkSGVR	P0A7V8from18to26	0.045	2.25	67.65	1220.725	-1.59

		8			13	
dAAIkPGNTLPMR	P60422from121to133	0.0193	1.54	66.88	1569.86991	0.52
dNEIVAkLFNELGPR	P0AG44from72to86	0.0357	2.17	39.9	1901.0427	1.38
eAIALLk	P31828from234to258	0.0209	1.85	94.97	943.61058	1.07
eVPVEVkPEVR	P68066from56to66	0.0209	2.37	160.19	1466.84836	-0.21
dAGLPIPk	P04982from28to39	0.0431	1.51	86.07	996.59935	-0.4
dAGFAITk	P0C0L2from90to143	0.0357	1.7	85.23	1008.56389	0.53
dALEFSEk	P37769from100to154	0.0209	1.61	86.12	1124.5706	-3.3

VITA

Yanjie Jiang was born in Heilongjiang, China in 1974. She graduated with her Bachelor degree from Nankai University with a major in Chemistry in 1996. Later she obtained her Master degree from Dalian Institute of Chemical Physics, Chinese Academy of Sciences with a major in Analytical Chemistry in 1999. She received her Master degree from University of New Orleans in Analytical Chemistry in 2004. She then moved to Boston and has been working at Momenta Pharmaceuticals Inc. ever since. She conducted research for this thesis to pursue her PhD degree since 2011.