

**MASS SPECTROMETRIC ANALYSIS OF PROTEIN  
CARBONYLATION AND GLYCOSYLATION**

BY

YIYING ZHU

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE  
DEPARTMENT OF CHEMISTRY AT BROWN UNIVERSITY

PROVIDENCE, RHODE ISLAND

MAY 2014

© Copyright 2014 by Yiying Zhu

All Rights Reserved

This dissertation by Yiying Zhu is accepted in its present form by  
the Department of Chemistry as satisfying the dissertation requirement for  
the degree of Doctor of Philosophy

Date \_\_\_\_\_  
\_\_\_\_\_  
Professor Carthene R. Bazemore-Walker

Recommended to the Graduate Council

Date \_\_\_\_\_  
\_\_\_\_\_  
Professor Arthur Salomon, Reader

Date \_\_\_\_\_  
\_\_\_\_\_  
Professor J. William Suggs, Reader

Approved by the Graduate Council

Date \_\_\_\_\_  
\_\_\_\_\_  
Professor Peter M. Weber,  
Dean of the Graduate School

## **CURRICULUM VITAE**

Yiying Zhu was born in Quzhou, Zhejiang Province of P.R. China on January 21, 1988. She received her Bachelor's degree in Pharmacy at Fudan University in Shanghai in 2008. Her thesis research evaluated drug release properties of a thermo-sensitive gel. After graduation, she came to the U.S.A to pursue her doctoral work in the Department of Chemistry at Brown University in Providence, RI. She joined Carthene R. Bazemore-Walker's proteomics laboratory and started using mass spectrometry to analyze proteins responsible for diseases and disorders. Her Ph.D. research has mainly focused on developing and validating LC-MS/MS methods to identify protein carbonylation in protein mixtures and quantify N-linked glycosylation in purified proteins.

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Prof. Carthene R. Bazemore-Walker. She has been a strong adviser to me based on her patience, motivation and encouragement. Her insightful guidance helped me throughout my doctoral studies. She is a professor that truly cares about her students. I would also like to thank my thesis committee members, who were also members of my ORP and RPD committee, Prof. J. William Suggs and Prof. Arthur Salomon, for their time, encouragement, and suggestions.

I would like to show my great appreciation to Prof. Matthew B. Zimmt for his advice and help, Prof. Gaurav Choudhary for his initiation and support of my protein carbonylation project, Dr. Tun-Li Shen for assistance with instruments, and Prof. Shouheng Sun for the usage of bio-facility.

I thank my fellow labmates (Yuan Cao, Chao Gong, Hongbo Gu, Michael Ellisor, Chloe Poston, Zhuo Chen, Shumin Yao, Anthony Bui, Ellen Duong, Michael Hogan, Damilola Idowu, Helen Johnson, and Seth Levin) for all the fun we have had over the last five years. Thanks to all of my friends in and out of this department, related or not related to my thesis work. Each of you has made my life at Brown very enjoyable. A special thanks goes to Fei Guo who continues to inspire me in research and life. I want all of you to know that I really appreciate your support!

Last but not least, I would like to send my deepest gratitude to my father Xiaomin Zhu and mother Meijiao Zhou for their tremendous love and support throughout my life. Your love has always been my driving force to success. I dedicate this dissertation work to you and I hope that you are proud of me.

## **ABSTRACT**

Post-translational modifications (PTMs) of proteins can have significant effects on different biological processes in which the proteins are involved. Analysis of protein post-translational modifications (PTMs) provides crucial information that elucidates molecular mechanisms of diseases and leads to new diagnosis paradigms and treatment strategies. This dissertation serves as a record of our efforts to identify and quantify two types of protein PTMs by mass spectrometry: carbonylation and glycosylation.

Carbonylation of proteins induced by acrolein is a non-enzymatic chemical modification. Through Michael addition, acrolein adds a carbonyl group to the side chain of cysteine, lysine, or arginine. Acrolein can be generated in the liver as a metabolite of chemicals such as ally alcohol and cyclophosphamide; and as a result, may be responsible for each molecule's demonstrated liver toxicity. To determine the identity of the proteins selectively adducted by acrolein, rat liver microsomes were treated with acrolein under non-denature condition, then an analysis strategy combining biotin tagging of carbonylated proteins, avidin enrichment of biotinylated peptides and LC-MS/MS characterization was utilized. Proteins potentially involved in liver toxicity were characterized including enzymes involved in metabolism of acrolein. This study validates the peptide-centric approach and should facilitate an understanding of the role that protein carbonylation plays in acrolein's toxicity within the liver.

The glycosylation profile of prostate specific antigen (PSA) is altered in patients suffering from prostate cancer. Characterizing the glycoforms of PSA in serum that are specific to the disease state could potentially be a diagnostic method of prostate cancer with high specificity. As proof of principle, two different commercially-available PSA

samples with different glycosylation profiles were analyzed. A bottom-up approach was applied for glycol-quantification of a protein, which included protein digestion, HILIC SPE purification of glycopeptides, and detection of glycopeptides *via* reversed-phase (RP) LC-MS/MS. By revealing relative abundances of twenty-eight differential glyco-isoforms, this pilot study confirms the potential utility of a peptide-centered approach for detecting and quantifying different glycoforms of a specific protein.

## TABLE OF CONTENT

CHAPTER 1 MASS SPECTROMETRY CHARACTERIZATION OF PROTEIN CARBOXYLATION INDUCED BY ACROLEIN IN THE LIVER.....	1
1.1 INTRODUCTION .....	2
1.1.1 Acrolein and carbonylation .....	2
1.1.2 Analytical method .....	6
1.2 MATERIALS AND METHODS.....	15
1.2.1 Materials .....	15
1.2.2 Preparation of liver microsomes .....	16
1.2.3 Treatment of liver microsomes with acrolein .....	16
1.2.4 Biotin labeling and western blotting.....	17
1.2.5 Synthesis of acrolein/biotin-labeled BSA and control BSA .....	17
1.2.6 Avidin column enrichment .....	18
1.2.7 Diethyl labeling.....	20
1.2.8 LC-MS/MS analysis .....	20
1.2.9 Bioinformatic analysis .....	21
1.3 RESULTS AND DISCUSSION.....	23
1.3.1 Method optimization .....	23
1.3.2 Results .....	27
1.3.3 Effects of acrolein on proteins.....	33
1.3.4 Diethylation quantification .....	38
1.3.5 Accomplishments.....	42
1.4 IDENTIFICATION TABLES .....	43
1.5 REFERENCES .....	59
CHAPTER 2 MASS SPECTROMETRY QUANTIFICATION OF DIFFERENTIAL GLYCOFORMS IN PROSTATE-SPECIFIC ANTIGEN.....	69
2.1 INTRODUCTION .....	70
2.1.1 Background on glycosylation.....	70
2.1.2 Prostate specific antigen.....	74



2.1.3 Protein glycosylation analysis by mass spectrometry .....	78
2.1.4 Purification of glycopeptides .....	82
2.1.5 Background of the interlaboratory study.....	85
2.2 EXPERIMENTAL PROCEDURES .....	86
2.2.1 Trypsin Digestion.....	86
2.2.2 Purification of glycopeptides by HILIC SPE.....	87
2.2.3 Mass spectrometric analysis.....	87
2.2.4 Identification and quantification of glycopeptides .....	88
2.3 DISCUSSION .....	89
2.3.1 Method development.....	89
2.3.2 Data analysis and result discussion.....	96
2.3.3 Accomplishments.....	102
2.4 REFERENCES .....	104

## LIST OF TABLES

Table 1.1. Examples of identified acrolein-targeted peptides under different concentrations of acrolein. Identified peptides with modification sites marked by red text and corresponding proteins were listed.....	29
Table 1.2. Acrolein-targeted proteins that might be involved in the metabolism of acrolein. The list included identified isotypes of aldehyde dehydrogenase, glutathione S-transferase, and cytochrome P450 which were the enzymes in the metabolism pathways of acrolein. ....	36
Table 1.3. Results of diethylation quantification for avidin purification of acrolein/biotin labeled BSA (ACR BSA) and control BSA. The BSA peptides were analyzed by Qstar Elite. Peak areas were integrated for each peptide in the extracted ion chromatographs. Ratios were calculated based on peak areas.....	41
Table 1.4. Identified modified peptides and corresponding proteins at 2 $\mu$ M acrolein.....	43
Table 1.5. Identified modified peptides and corresponding proteins at 20 $\mu$ M acrolein.....	45
Table 1.6. Identified modified peptides and corresponding proteins at 200 $\mu$ M acrolein.....	49
Table 2.1. Relative abundances of selected glycans attached to Asn158 of glycoprotein fetuin in our data and reference data [35]. In our study, a glycopeptide approach was utilized, while in reference they used both peptide and glycan analysis. The percentage was calculated by dividing the peak area of target glycoform by the summed peak area of four selected glycoforms. ....	93
Table 2.2. The percentages and ranks of selected glycoforms in PSA before and after PolyHEA purification. Quantification was based on glycans attached to peptide NKSVILLGR. The percentage was calculated by dividing the peak area of target glycoform by the sum of peak area of seven selected glycoforms. They were ranked from the most abundant (1) to least abundant (7).....	94
Table 2.3. Relative abundances of identified 28 glycoforms in PSA and PSA high isoform. Glycoforms composed with different numbers of Hex, HexNAc, dHex, and Neu5Ac were identified in two different PSA samples. The relative intensities were calculated by comparing the peak areas of glycopeptides to the total peak area of all glycopeptides (n=2).....	97
Table 2.4. Percentages of different types of N-glycans in major and intermediate glycans.....	100
Table 2.5. Percentages of different types of N-glycans in significantly changed glycans..	100

## LIST OF FIGURES

- Figure 1.1. The formation of acrolein-adducts with amino acids [11, 12]. Acrolein reacts with the sulfhydryl group, amino group and imidazole group *via* 1, 4-Michael addition. Lysine not only forms the cyclic bi-adduct with two acrolein molecules *via* 1, 4-Michael addition [5] but also Schiff base product *via* 1, 2-addition [6]..... 3
- Figure 1.2. Proposed metabolism of acrolein [16]. 1. Transformed into acrylic acid by aldehyde dehydrogenase and NAD<sup>+</sup> in cytosol and microsomes of livers 2. Reacted with glutathione (GSH) with or without catalyzation by glutathione S-transferase 3. Oxidized to glycidaldehyde in the present of epoxidase, cytochrome P450 and NADPH in liver and lung microsomes..... 6
- Figure 1.3. The method (peptide level) used to enrich acrolein protein targets. (Biotin labeling): Tagging carbonylated proteins with biotin groups (Colored lines represent proteins). A hydrazone bond is formed between biotin hydrazide and a representative acrolein-induced carbonyl group at cysteine residue and then stabilized by sodium cyanoborohydride. Protein labeled with an acrolein/biotin tag has a mass shift of 298.15 Da. (Trypsin): Digestion of proteins into peptides with trypsin which is a protease cleaving at N-terminal of lysine and arginine. (Avidin purification): Purification of acrolein/biotin labeled peptides by monometric avidin column. .... 9
- Figure 1.4. Nomenclature of peptide fragment ions [74]. The names of the ions are based on the cleaved bonds and location of charges. Fragments contain N-terminal amines are classified as a, b, and c ions. Fragments contain C-terminal carboxylic groups are named x, y, and z ions. Collision induced dissociation (CID) usually cleaves amide bonds and generates b ions and y ions. .... 13
- Figure 1.5. Immunoblotting of biotinylated proteins in acrolein-exposed (0, 2, 20, 200  $\mu$ M) rat liver microsomes. After acrolein modification, microsomal proteins were exposed to biotin hydrazide, by which carbonyl groups were labeled with biotin. Proteins were then separated by SDS-PAGE and transferred to a PVDF membrane where the antibody HRP-conjugated streptavidin was added to detect the biotinylated proteins. The result showed increased protein carbonylation level with increased acrolein concentration in a dose-dependent manner..... 28
- Figure 1.6. Preferred location of Michael-type acrolein addition. Most peptides (81 %) were modified on the cysteine, while histine and lysine only took 8 % and 11 % respectively. This result was consistent with the literature that indicated the sulfhydryl group was the most reactive group in proteins [9, 65, 82, 83]. .... 30
- Figure 1.7. Characteristic cleavage fragments of the acrolein-biotin label. (A) Structure of an acrolein-biotin labeled N-terminal Cys residue with proposed cleavage sites indicated in red. (B) MS/MS spectrum of the CLVEELR peptide (from cytochrome P450) with

acrolein-modified Cys. Proposed structures for the observed fragment ions indicated in red were shown in (C). . . . .	32
Figure 1.8. The crystal structure of human alcohol dehydrogenase 1 (PDB:1HTB). Two modified sites were detected: Cys 46 and Cys 97. Derivatization of these residues likely led to protein deactivation since Cys 46 coordinated one zinc atom (shown as a light blue ball) responsible for substrate binding and Cys 97 coordinated a second zinc that is crucial for protein stability [85, 86]. . . . .	35
Figure 1.9. Classification of identified proteins according to Gene Ontology (GO) molecular function ( $p < 0.05$ ). The bar showed the number of proteins corresponding to a category of molecular function. . . . .	37
Figure 1.10. The strategy for determining the reproducibility and background binding for avidin purification using diethylation quantification. Acrolein/biotin labeled BSA were separated into two aliquots and purified by an avidin cartridge. After labeled with $^{13}\text{C}/^{12}\text{C}$ acetaldehyde reagents, an equal amount of tryptic peptides from carbonylated BSA were mixed and analyzed by LC-MS/MS to evaluate the reproducibility of the quantification method. Unmodified control BSA was purified, digested and labeled with light acetaldehyde in the same way as acrolein/biotin labeled BSA. An equal amount of modified BSA peptides were mixed with control BSA peptides eluted from avidin and analyzed by LC-MS/MS to determine the background binding. . . . .	40
Figure 2.1. Haworth projections of naturally occurring monosaccharides. They are classified into groups by the molecular weight or acidity (sialic acid). D-glucose (Glu/Glc), D-galactose (Gal), and D-mannose (Man) belong to the category of Hexose (Hex). L-fucose (Fuc) and L-Rhamnose (Rham) are deoxyhexose (dHex). N-acetylglucosamine (GlcNAc) and N-acetylgalactosamine (GalNAc) are N-acetylhexosamine (HexNAc). N-acetylneuraminic acid (Neu5Ac/NeuAc) and N-Glycolylneuraminic acid (Neu5Gc) are both called sialic acid. . . . .	72
Figure 2.2. Structures of three N-glycan subtypes: high-mannose, complex, and hybrid. High mannose type of glycans has only mannose residues besides core structure. Complex type of glycans contains a variety of the other types of saccharides. Hybrid glycan is characterized as one mannose in core structure substituted with mannose as in high mannose type and the other mannose with other types of monosacchrides as in the complex glycan. . . . .	74
Figure 2.3. Sequence of prostate specific antigen (PSA). The amino acid Asparagine (N) 69 which carries N-glycans is marked in red. . . . .	77
Figure 2.4. Schematic illustration of polyhydroxyethyl aspartamide stationary phase. . . . .	85
Figure 2.5. Differential profiles of seven major N-glycans derived from the consensus data of 15 labs (A) and my data (B). Top blue color columns represented relative percentages of glycans in PSA. Bottom red colored columns represented relative abundances of glycoforms in PSA high isoform. . . . .	99

## **CHAPTER 1**

---

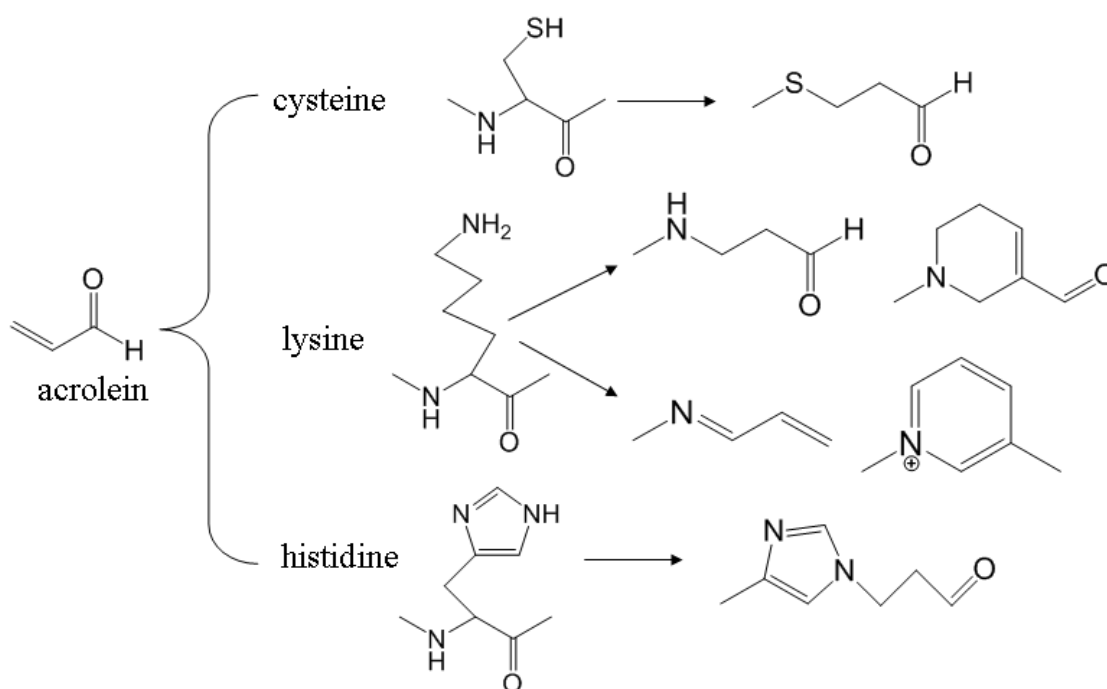
# **MASS SPECTROMETRY CHARACTERIZATION OF PROTEIN CARBOXYLATION INDUCED BY ACROLEIN IN THE LIVER: FOCUS ON SITE-SPECIFIC ANALYSIS**

## 1.1 INTRODUCTION

### 1.1.1 Acrolein and carbonylation

#### Reactivity of acrolein with proteins

Acrolein (systematic name: 2-propenal) is the smallest and most reactive  $\alpha$ ,  $\beta$ -unsaturated aldehyde. The reactivity may be responsible for the toxicity of acrolein within the cells. The compound is capable of modifying cellular molecules with nucleophilic groups such as protein, DNA and glutathione [1-3]. When reacting with glutathione, acrolein forms conjugates more rapidly than other aldehydes such as crotonal, pentenal, hexenal, 3-methyl-butenone and the form adducts are very stable (half-lives for reverse reaction 4.6 days) [1, 4]. In proteins, it reacts readily with sides chains of cysteine, lysine or histidine in proteins (Figure 1.1). Specifically, it introduces a carbonyl group to the sulfhydryl group in cysteine, and imidazole group in histidine *via* 1, 4-Michael addition. The amino group in lysine or protein N-terminus could not only react with one or two acrolein molecule through 1, 4-Michael addition [5], but also create a Schiff base product through 1, 2-addition [6]. Only carbonylated proteins after acrolein treatment have been profiled in this study, since the formation of Schiff base products is much slower and take less responsibility on the cytotoxicity [7]. Carbonyl adducts could not only directly influence protein functions or change structures of protein but also induce further protein cross-linking through Schiff base formation with amines [8-10]. In addition, Schiff base type of adducts are difficult to enrich for identification since they do not possess specific reactivity towards any functional group as carbonyls.



**Figure 1.1. The formation of acrolein-adducts with amino acids** [11, 12]. Acrolein reacts with the sulphydryl group, amino group and imidazole group *via* 1, 4-Michael addition. Lysine not only forms the cyclic bi-adduct with two acrolein molecules *via* 1, 4-Michael addition [5] but also Schiff base product *via* 1, 2-addition [6].

#### Sources and toxicity of acrolein

Acrolein could be generated both endogenously and exogenously. Acrolein is a ubiquitous environmental pollutant, which was produced in incomplete combustion of organic materials such as cigarette smoke (3-220 mg/cigarette [13]), heated cooking oil (11.9 to 38.1  $\mu\text{g/g}$  [14]), vehicle exhaustion, and smoke of wild fire [15-17]. Humans could be exposed to acrolein by inhalation, ingestion, or dermal contact. Acrolein has a strong piercing smell. It causes severe irritation on eyes, skin, nasal passages, and larynx [15]. Historically, it was used as a chemical weapon during World War I. Acrolein is proposed to be the major factor of tobacco induced chronic obstructive pulmonary disease (COPD) and lung cancer [18-20]. The cardiotoxicity of acrolein has been widely

studied. Acrolein intravenous administrated close-chest mice (0.5 mg/kg) had left ventricular dilatation and dysfunction compared to vehicle treated mice [21]. Acrolein oral exposed mice developed dilated cardiomyopathy [22]. Acrolein perfusing caused rat hearts beat irregularly or even arrest [23]. There was a fatal human intoxication report described a man died from orally ingested allyl alcohol, in which metabolite acrolein in blood might be the reason to cease the heart beating [24]. Acrolein gavage fed mice prior to myocardial ischemic injury had increased myocardial infarct size [25]. In chick embryonic myocardial myocyte reaggregate, neonatal rat cardiac fibroblasts and myocytes, acrolein decreased spontaneous beating activity, reduced cellular ATP and caused cell lysis [26, 27]. Acrolein could be generated by lipid peroxidation, an oxidative degradation process of lipids [5] in cells. The proposed mechanism is that acrolein is originated from beta carbon-carbon bond cleavage in the center of the aliphatic chain [11]. Acrolein could be converted from amino acids and polyamine under oxidative stress [28-30]. The conversion of threonine or spermine to acrolein is mediated by myeloperoxidase or amine peroxidase [28]. The level of acrolein is found elevated in patients of myocardial infarction, Alzheimer's disease, renal failure, and type 2 diabetes [30-33]. Increased acrolein could also be due to metabolite processes of xenobiotics. For example, allyl alcohol, which is a chemical used in the manufacture of drugs, organic chemicals, plastics, herbicides and pesticides, is transformed into acrolein by alcohol dehydrogenase in the liver [34]. Allyl alcohol is a well-known hepatotoxicant and it causes occupational liver diseases. Since allyl alcohol does not have toxic effects itself, it has been documented that acrolein may be responsible for allyl alcohol induced hepatic necrosis by direct alkylating proteins [35, 36]. Another example is cyclophosphamide, which is a

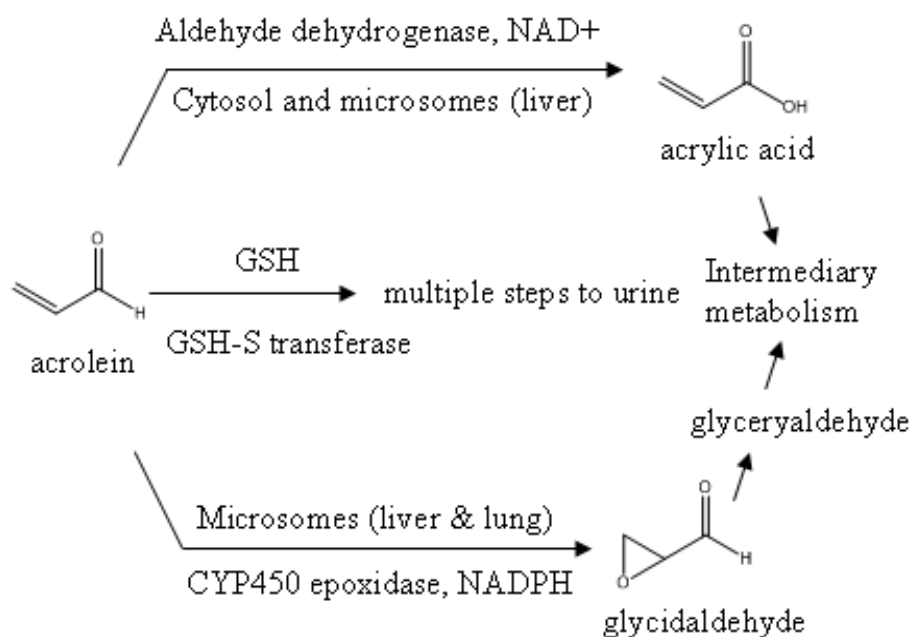


chemotherapy drug used in blood and marrow transplantation [37, 38]. Cyclophosphamide is oxidized into 4-hydroxycyclophosphamide by cytochrome P450 in the liver and then changed to aldophosphamide which finally decomposed into phosphoramidate mustard and acrolein [38, 39]. It has been found that cyclophosphamide depresses hepatic drug metabolism at high dose, possibly caused by acrolein [40, 41]. According to previous reports, the molecular mechanism of cyclophosphamide liver toxicity may be alkylation of sulfhydryls in the active site of cytochrome P450 by acrolein [42, 43].

*Reasons to study the toxicity of acrolein in liver microsomes*

This study has focused on identifying of protein carbonylation induced by acrolein, specifically in liver microsomes for the following reasons. First, liver microsomes are the primary region for metabolism of xenobiotics such as allyl alcohol and cyclophosphamide, which are metabolized into acrolein as introduced in the previous paragraph. Secondly, acrolein is metabolized mainly in the liver (Figure 1.2). It is transformed into acrylic acid by aldehyde dehydrogenase and NAD<sup>+</sup> or oxidized to glycidaldehyde in the present of epoxidase, cytochrome P450 and NADPH in liver microsomes [44]. Thirdly, hepatic effects of acrolein have been observed in both acute and chronic inhalation toxicology studies including hepatic hyperemia, perivascular edema, nonspecific inflammation, and focal hepatic necrosis [15]. In addition, endoplasmic reticulum (ER) plays an important role in the toxicity of acrolein. ER is the place where the proteins are synthesized, modified, folded and delivered. Studies indicate that exposure of acrolein leads to problems of protein folding and triggers the unfolded protein response (UPR) and ER stress in hepatocytes [45], human umbilical vein

endothelial cells [46], and rat lung tissue [47]. The contribution of protein-acrolein adducts to those cellular effects needs further investigation [48]. Thus, it is very beneficial to identify the carbonylated proteins induced by acrolein with the exact modification sites in liver microsomes for elucidating its toxicity mechanism.



**Figure 1.2. Proposed metabolism of acrolein [16].** 1. Transformed into acrylic acid by aldehyde dehydrogenase and NAD<sup>+</sup> in cytosol and microsomes of livers 2. Reacted with glutathione (GSH) with or without catalyzation by glutathione S-transferase 3. Oxidized to glycidaldehyde in the present of epoxidase, cytochrome P450 and NADPH in liver and lung microsomes

### 1.1.2 Analytical method

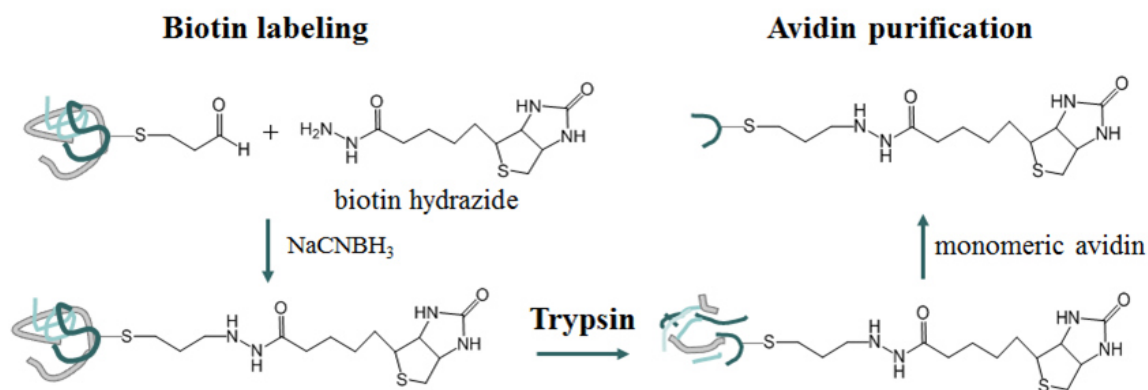
#### Commonly used analytical methods

Various techniques have been developed for detection and isolation of carbonylated proteins induced by acrolein. For evaluation of global protein carbonylation level, a

commonly used method is first derivatizing carbonyl groups with reagents such as dinitrophenylhydrazine (DNPH) [49-52] or biotin hydrazide [53, 54] and then detecting by antibodies (anti-DNP or -biotin) after gel electrophoresis separation. This method is capable of identifying all carbonylated proteins but could not distinguish types of carbonylation and carbonylation sites. Protein carbonylation could not only come from reaction with acrolein but also from other sources. For example, protein carbonyls could be formed by reacting with other lipid peroxidation products such as 4-hydroxynonenal (4-HNE) and malondialdehyde (MDA), which are generally called reactive carbonyl species (RCS) [55]. Carbonyls are also present in advanced glycation end products (AGE) which are results of non-enzymatic glycosylation reactions. Oxidative cleavage of protein backbones or side chains of amino acids is another source of protein carbonylation [56]. In order to specifically target one type of acrolein-modified proteins, an antibody has been created to recognize cyclic bi-acrolein-lysine type of adducts [57]. However, until now there are no specific antibodies to detect mono-acrolein type of adducts with cysteine, histidine or lysine. Compared to traditional immuno-based methods, recent development on tandem mass spectrometry coupled with high performance liquid chromatography (LC-MS/MS) provides a powerful tool to analyze post-translation modifications of proteins. Not only LC-MS/MS is a large scale method which could sequence hundreds of proteins in one run, but also it profiles all types of carbonylation adducts including lysine, arginine and cysteine. LC-MS/MS has already been used to identify acrolein-protein adducts in the literature [53, 54, 58]. The details of the characterization method will be introduced in the following paragraphs.

### *Biotin-avidin enrichment before LC-MS/MS analysis*

Since carbonylated proteins are usually low-abundant compared to non-carbonylated proteins, enrichment is a necessary process before mass spectrometry analysis [56]. Almost all enrichment methods utilize the unique reactivity property of carbonyl groups. Biotin hydrazide types of compounds are commonly used reagents to derivatize carbonyls. After biotin labeling, an avidin column could be used to isolate carbonylated proteins. Avidin is a protein which has specific affinity towards the small molecule biotin. The biotin-avidin binding is by far the strongest known non-covalent biological interaction. The dissociation constant of a typical tetramer avidin with biotin is  $10^{-15}$  M [59]. Because of the specificity and strength of biotin and avidin interaction, avidin has already been widely used as a valuable tool in various researches as an affinity probe or matrix [60-62]. Due to the strong binding, denaturing conditions such as boiling in SDS buffer would be utilized to elute biotinylated molecules. In order to make the binding reversible, monomeric avidin has been developed to reduce the binding affinity to  $10^{-7}$  M [63]. Elution could be completed by changing pH or adding d-biotin. Monomeric avidin columns could be regenerated many times without significant loss of binding capability.



**Figure 1.3. The method (peptide level) to enrich acrolein protein targets.** (Biotin labeling): Tagging carbonylated proteins with biotin groups (Colored lines represent proteins). A hydrazone bond is formed between biotin hydrazide and a representative acrolein-induced carbonyl group at cysteine residue and then stabilized by sodium cyanoborohydride. Protein labeled with an acrolein/biotin tag has a mass shift of 298.15 Da. (Trypsin): Digestion of proteins into peptides with trypsin which is a protease cleaving at N-terminal of lysine and arginine. (Avidin purification): Purification of acrolein/biotin labeled peptides by a monomeric avidin column.

Avidin column could be utilized to enrich both biotinylated peptides and biotinylated proteins, designated as “peptide-level” and “protein-level” purification. In peptide-level isolation, carbonylated proteins are labeled with biotin-affinity tags, digested into peptides and then biotinylated peptides are enriched by an avidin column (Figure 1.3) [64, 65]. The protein-level method directly enriches biotin labeled proteins *via* avidin column, and then isolated proteins are digested into peptides to be analyzed in LC-MS/MS [66, 67]. The advantage of peptide-level for LC-MS/MS characterization is having more identification of acrolein/biotin-modified peptides. Peptide-level purification removes more non-carbonylated peptides compared to protein-level purification. In bottom-up proteomics which is a method analyzing peptides by mass spectrometry after proteolytic digestion of proteins, only relative high abundant peptides could be sequenced by

MS/MS. Excluding non-modified peptides increases the chance to characterize low abundant acrolein-modified peptides. The disadvantage of peptide-level purification is that when an acrolein-modified peptide is too long or short for LC-MS/MS analysis, the protein will be missed. Protein-level purification in avidin column could maintain the information of the carbonylated protein even when the carbonylated peptide is missed. The peptide mixtures after protein-level purification are from intact acrolein protein adducts which include modified and unmodified peptides. When the acrolein-modified peptide could not be sequenced, identification of unmodified peptides could still provide the ID of the protein. In addition, in protein-level purification, protein isotype could be distinguished since there are multiple peptides from one protein. The disadvantage of protein-level enrichment is that there are unmodified protein contaminants in the eluate from the avidin column. These proteins include naturally biotinylated proteins and proteins interacting with avidin through electrostatic effects. It is necessary to determine the background-binding proteins by conducting a control experiment. Peptide-level enrichment is preferred when research is focusing on identification of acrolein-targeted amino acid sites. Ugur *et al.* evaluated three commercial available biotin-and-hydrazide based reagents with different space arms between biotin and hydrazide groups in the identification of acrolein induced protein carbonylation using peptide-level avidin enrichment [68]. Three tested biotin reagents perform equally well in identification numbers of acrolein-modified peptides from a standard protein human bovine serum. In this study, we utilized the simplest molecule biotin hydrazide to label the carbonyls. The procedures for enrichment prior to LC-MS/MS analysis are shown in Figure 1.3.

### Mass spectrometer

Mass spectrometry (MS) is an analytical technique widely applied in high-throughput proteomics analysis. A mass spectrometer is made of an ion source, analyzer(s) and a detector. Analytes are converted into gas phase molecules by ion source, separated based on their  $m/z$  in mass analyzer, and then signals are recorded by a detector. Electrospray ionization (ESI) is a widely used ionization approach, which could be coupled to online liquid chromatography (LC) for molecule separation before MS analysis [69]. Three MS instruments were employed in this study: Qstar Elite, LTQ-FT, and LTQ-Orbitrap Velos, all of which are equipped with ESI. A high voltage is applied to the tip of analytical column and electric-charge droplets are formed. As traveling from the emitter to the inlet of mass spectrometer, droplets reduce their sizes and finally ions are desolvated. Mass analyzers utilized in Qstar Elite, LTQ-FT, and LTQ-Orbitrap Velos are quadrupole (Q)/time of flight (TOF), linear ion trap (LTQ)/Fourier transform ion cyclotron resonance (FT-ICR), and linear ion trap (LTQ)/Orbitrap respectively. Quadrupole (Q) selectively stabilizes or destabilizes ions in oscillating electric fields with applied radio frequency and/or direct current voltages. In linear ion trap (LTQ), ions are trapped radially by radio frequency only electric field and axially by a static direct current potential applied to end electrodes [70]. TOF mass analyzer measures the time that an ion travels a defined distance after accelerating by a known electric field [71]. The velocity of an ion is depending on its mass-to-charge ratio. FT-ICR determines ion's mass-to-charge ratios based on their cyclotron frequency in a fixed magnetic field [72]. Ions are excited by an oscillating electric field, and then the currents of coherence cyclotron motion of ions are recorded after removal of the excitation field. In Orbitrap, ions are attracted by electric

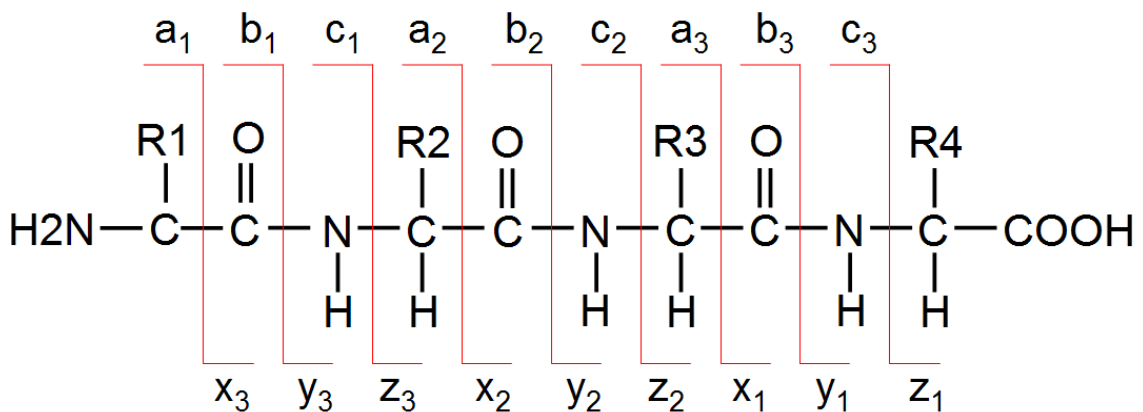
field balanced with centrifugal forces [73]. Coherent axial oscillations of ions start with or without applying radio frequency waveforms when their image currents are detected by sensors. The feature of Q and LTQ only selectively allowing ions to pass through makes them capable to be exploited as mass filters in tandem mass spectrometers. TOF, FT-ICR and Orbitrap are high mass accuracy and resolution mass analyzers utilized in tandem mass spectrometry.

#### Tandem mass spectrometry (MS/MS)

Tandem mass spectrometry, also known as MS/MS, is a powerful tool to sequence peptides and detect post-translation modifications of proteins. In bottom-up proteomics, peptides after enzyme digestion are ionized and their mass-to-charge ratios are measured. The ions of intact peptides are called precursor ions. The objective peptide selected by a mass filter is broken down into fragments and then a mass analyzer scans ions of fragments, the so-called product ions. From the spectrum of MS/MS, the information from peptide fragments could be pieced together to obtain the structure arrangement and modification of amino acids. When analytes are separated by LC gradient prior to MS/MS analysis, identification number of peptides could be largely increased. It is because the instrument only selects the most abundant ions to conduct fragmentation at one specific time. The types of fragment ions are diverse when applying different dissociation methods. The nomenclature for MS/MS ions is demonstrated in Fig. 1.4 [74]. Names of the ions are based on the cleaved bonds and location of charges. Fragments contain N-terminal amines are classified as a, b, and c ions. Fragments contain C-terminal carboxylic groups are named x, y, and z ions. Collision induced dissociation (CID), also



called collision activated dissociation (CAD), is a commonly used dissociation method, in which peptides are colliding with inert neutral gas molecules such as nitrogen, helium, and argon [75]. Almost all tandem mass spectrometers are equipped with CID capabilities. The mechanism of CID is that the kinetic energy of gas molecules is turned into internal energy of peptides during collision and it leads to dissociation of peptide when there is enough internal energy. According to the “mobile proton model”, the mobile proton in charged peptides stays on the amide nitrogen in CID and leads to break of amide bonds to form b and y ions [20–22].



**Figure 1.4. Nomenclature of peptide fragment ions [74].** The names of the ions are based on the cleaved bonds and location of charges. Fragments contain N-terminal amines are classified as a, b, and c ions. Fragments contain C-terminal carboxylic groups are named x, y, and z ions. Collision induced dissociation (CID) usually cleaves amide bonds and generates b ions and y ions.

### Diethylation

Diethylation is a stable-isotope labeling technique for quantitative proteomics analysis. It is used to quantify relative abundances of proteins in different biological samples based on mass spectrometry. The technique utilizes a reagent acetaldehyde which carries heavy

or light isotopes. In this study, a pair of  $^{13}\text{C}/^{12}\text{C}$  acetaldehyde has been applied to label carbonylated and control BSA. Acetaldehyde reacts with amino groups at the protein N-terminus and lysine side chains. Light or heavy acetaldehyde is added to different biological samples after proteolytic digestion, and imine bonds are formed between amines and carbonyls. Formed imine bonds are reduced by sodium cyanoborohydride to be stabilized. Two molecules of acetaldehyde add two ethyl groups in the amine group, which is the reason that the quantification method is named “diethylation”. Peptides with an amino group obtain a mass shift of 56 Da or 60 Da for light or heavy reagents respectively. When two different samples are combined and analyzed in LC-MS/MS, the same peptides with different isotope labeling have the same retention time in HPLC and could be distinguished by 4 Da mass differences in MS. Their relative abundances could be determined by comparing peak areas in MS spectra.

#### Searching software Mascot

Search algorithms such as Mascot, SEQUEST, X!Tandem, OMSSA, and ProteinPilot are employed in automatic identification of peptides from tandem mass spectra. The basic principles of these algorithms are generally the same. Proteins from database are digested by indicated enzyme theoretically. Masses of peptides and fragmentations of peptides are generated *in silico*. The obtained experimental spectra are compared with theoretical data. The absolute probability (P) that the observed match is a random event is calculated. Since the probability P of a good match is usually very small and includes a wide range of magnitudes, a Mascot score of  $-10\log_{10}(P)$  is reported for convenience. A Mascot score of 30 is a probability P of  $10^{-3}$ . Depending on absolute probability P and the size of

the database, the significance of protein identification could be measured. A commonly accepted threshold of  $p=0.05$  indicates a 5% chance that the protein identified is a random match. The threshold of Mascot score is usually around 30 ( $p<0.05$ ). Spectra with Mascot scores higher than the threshold are considered significant identification. However fragmentations and neutral losses from peptide modifications may reduce Mascot scores of corresponding peptides since they could not be matched to theoretical peaks. A spectrum obtaining a low Mascot score may still be a true identification. This is the case in Mascot search of acrolein/biotin modified peptides which will be further discussed in the part of result and discussion.

## **1.2 MATERIALS AND METHODS**

### **1.2.1 Materials**

Protease inhibitor cocktail was purchased from Roche Applied Science (Indianapolis, IN). Sequencing grade trypsin was ordered from Promega (Madison, WI). PVDF membranes (0.2  $\mu\text{m}$ ) and Amicon ultra centrifugal filters were from Millipore (Billerica, MA). Acrolein and sodium cyanoborohydride were purchased from Sigma-Aldrich (St. Louis, MO). Microcentrifugal spin columns, biotin hydrazide, SuperSignal chemiluminescent substrate, and BCA assay kits were from Thermo Fisher Scientific (Rockford, IL). NuPAGE Novex 3-8% Tris-Acetate gel was from Invitrogen (Calsbad, CA). All other standard laboratory chemicals were obtained from Thermo Fisher Scientific (Rockford, IL) or Sigma-Aldrich (St. Louis, MO). Rat livers were obtained either from Pel-Freez Biologicals (Rogers, AR; referred to as liver 1 and originating from an adult Sprague

Dawley rat with unknown sex) or Charles River Laboratories (Wilmington, MA; referred to as liver 2 and originating from an adult female CD rat).

### **1.2.2 Preparation of liver microsomes**

Rat livers were washed using an ice-cold buffer (225 mM mannitol, 75 mM sucrose, 0.5% BSA, 0.5 mM EGTA and 30 mM Tris-HCl, pH 7.4) containing protease inhibitor cocktail. They were then homogenized using a motor driven Potter-Elvehjm glass homogenizer on ice. All centrifugation steps were performed at 4 °C. The suspensions were centrifuged at 740×g for 5 min followed by 20,000×g for 30 min. Obtained supernatants were further centrifuged at 100,000×g for 1 h. The pellets were washed in PBS buffer (10 mM sodium phosphate, 150 mM NaCl, pH 7.2) and re-pelleted (twice) at 100,000×g for 1 h. Finally, the pellets were stored in -80 °C until needed.

### **1.2.3 Treatment of liver microsomes with acrolein**

Microsomal pellets were thawed and resuspended in PBS buffer *via* two strokes of a hand-held glass-teflon homogenizer. A BCA protein assay kit was used to determine protein concentrations and then the microsomal suspensions were distributed into 0.5-mg aliquots. Acrolein was added to each tube to a final concentration as follows: 200 μM for liver 1; 0, 2, 20, and 200 μM for liver 2. One replicate was done for each biological sample. The acrolein treatment was conducted at room temperature for 30 min with constant rotation. At the end of the reaction time, all suspensions were pelleted at 100,000×g to remove acrolein.

#### **1.2.4 Biotin labeling and western blotting**

Acrolein-modified microsomal proteins were resuspended in 1 mL PBS (pH adjusted to 5.5) and reacted with 5 mM biotin hydrazide for 2 h at room temperature with end-over-end rotation. Sodium cyanoborohydride was added to a final concentration of 15 mM and the mixture was incubated on ice for an additional 1 h with occasional vortex agitation. Excess reagent was removed by pelleting microsomal proteins at 100,000×g. Aliquots (5 µg) of samples from liver 2 were separated on a NuPAGE Novex 3-8% tris-acetate gel according to manufacturer's instructions and then electrophoretically transferred onto a PVDF membrane at 30 V for 90 min using the Invitrogen (Calsbad, CA) XCell II blot module. The membrane was washed in PBS with 0.2% Tween-20 for 30 min and incubated with HRP-conjugated streptavidin (diluted 1:500) for 60 min. The blot was exposed to SuperSignal chemiluminescent substrate for 5 min and visualized using a Syngene GeneGnome (Frederick, MD) bio-imaging system.

#### **1.2.5 Synthesis of acrolein/biotin-labeled BSA and control BSA**

BSA was dialyzed against MilliQ water overnight to remove impurities that could interfere with the reaction. Dialyzed BSA was diluted to a concentration of 0.25 mg/ml in 2× PBS buffer (pH=8.3). Then, 80 mL of BSA solution was treated sequentially with 250 µl DTT (from a 60 mM stock) for 30 min at 56 °C, 1.2 ml acrolein (from a 80 mM stock) for 30 min at room temperature, and a final 800 µl DTT (from a 60 mM stock) to quench the reaction. The reaction mixture was concentrated using a 30 kDa cut-off centrifugal

filter. After the pH of the reaction mixture was adjusted to 5.5, biotin hydrazide was added to a final concentration of 5 mM, and the mixture was incubated at room temperature for 2 h. Sodium cyanoborohydride was added to a final concentration of 15 mM and incubated for 1 h on ice. Control BSA was made by the exact same procedures, except adding distilled water instead of acrolein. Excess reactants were removed by dialysis against 1 L PBS (×3).

### **1.2.6 Avidin column enrichment**

#### *Peptide-level*

After acrolein/biotin labeling, pelleted microsomal proteins were resuspended in 100 mM ammonium bicarbonate (AMBIC, pH 8.2) buffer. BSA sample were aliquoted and PBS was removed by a 10 kDa centrifugal filter before resuspension. Each sample (0.5 mg) was reduced using a total of 2.25  $\mu$ mol dithiothreitol (DTT) (1 h, 56 °C) and alkylated using a total of 4.75  $\mu$ mol iodoacetic acid (IAA) (30 min, room temperature). Excess IAA was quenched by an additional aliquot of 2.25  $\mu$ mol DTT and then each sample was incubated with 10  $\mu$ g trypsin (overnight, 37 °C). Digestion reactions were terminated by passing solutions through 10 kDa MW cutoff centrifuge filters that remove undigested proteins including trypsin. Collected flow-through containing derivatized peptides was ready for avidin purification.

Two types of ready-to-use cartridges containing monomeric avidin were used: Pierce® Monomeric Avidin UltraLink® from Thermo Fisher Scientific (Rockford, IL) and monomeric avidin found in the ICAT reagent kit from ABSciex (Foster City, CA). For

convenience, the Thermo product is referred to as cartridge 1 (or C1) and the ABSciex product is called cartridge 2 (C2). C1 contained 2 ml of avidin while C2 had 200 µl of avidin. A third cartridge 3, referred to as C3, was self-packed with 200 µl of the Thermo avidin product. All avidin cartridges were conditioned with d-biotin before use.

The purification procedure varied depending on whether C1 or C2/C3 was used. C1 was only used to purify a 1-mg aliquot of liver 1. Tryptic peptides were loaded and non-labeled peptides were removed using 5-6 column volumes (CV) of PBS and eluted using 5-6 CV of 2 mM d-biotin in PBS. The eluate was desalted using a C18 cartridge. C2 was used to enrich a second aliquot (0.5 mg) of liver 1 and C3 was used to purify samples from liver 2 (each 0.5 mg). Tryptic peptides were loaded onto either C2 or C3 and washed with 1 ml of washing buffer (100% PBS, 80% (50 mM AMBIC)/20% methanol pH 8.3, and 100% MilliQ water). Biotinylated peptides were eluted using 1 mL of elution buffer (0.4% trifluoroacetic acid in 30% acetonitrile). Elutions were reduced to near dryness by vacuum centrifugation and reconstituted in aqueous 0.1 M acetic acid.

### Protein-level

The Acrolein/biotin modified BSA and control BSA were centrifuged at 16,000 g for 3 min to remove precipitate. Protein concentration was determined by BCA assay. An aliquot of 2.5 mg BSA was loaded to the avidin cartridge C1. After washing C1 by 12 ml PBS, the biotinylated peptides were eluted by 12 ml of 2 mM d-biotin in PBS and 12 ml glycine (0.1 M glycine, pH=2.8) sequentially. The pH of glycine elute was neutralized with 1 M Tris-HCl. The eluates were combined and concentrated by a centrifugal filter

with a 3 kDa molecular weight cut-off. It was finally dialyzed in a 10 kDa dialysis cassette against 1 L PBS (×3).

### **1.2.7 Diethyl labeling**

After trypsin digestion of acrolein/biotin modified and control BSA, peptides were desalted by a C18 spin tip. Each sample was reconstituted with 200 µl of 100 mM sodium acetate buffer (pH=5.5). Eight microliter of 40% (wt/wt) light <sup>12</sup>C and heavy <sup>13</sup>C acetaldehyde were added and incubated at room temperature for 2 h. Then 8 µl of 0.6 M sodium cyanoborohydride was added and reacted at room temperature for 1 h. Four more additions of 8 µl sodium cyanoborohydride were conducted with an interval of 20 min and finally the solution was kept at room temp for overnight. Ammonia solution was used to stop the reaction. The samples were acidified by formic acid. Same amount of light or heavy labeled BSA digests were combined. Finally the samples were centrifuged to near dry and resuspended in solvent A.

### **1.2.8 LC-MS/MS analysis**

For all experiments on liver microsomes, modified peptides were eluted into a mass spectrometer using a gradient from 5 to 50% solvent B over 60 min. All BSA tryptic digest was analyzed using a shorter gradient from 5 to 50% B over 17 min. The solvent systems varied slightly depending on the MS instrument utilized. For the QSTAR Elite (ABSciex, Foster City, CA), solvent A was 98% water, 2% ACN, 0.1% FA and solvent B was 98% ACN, 2% water, 0.1% formic acid. For the LTQ-FT (Thermo Scientific, San



Jose, CA) and the LTQ-Orbitrap Velos (Thermo Scientific), solvent A was 0.1 M acetic acid in water and solvent B was 0.1 M acetic acid in ACN.

Acrolein/biotin-labeled peptides from liver 1 were detected using the QSTAR Elite and LTQ-FT mass spectrometers while samples derived from liver 2 were analyzed on an LTQ-Orbitrap Velos. On all systems, a dynamic exclusion time of 30 seconds and a collision energy setting of 35 was used. On the QSTAR Elite, a full MS scan was followed by three data-dependent MS/MS scans with a 2 sec precursor ion accumulation time. On the LTQ-FT, a full MS scan was followed by nine data-dependent 250 msec MS/MS acquisitions. On the LTQ-Orbitrap, a full MS scan was followed by ten 100 msec data-dependent acquisition of MS/MS spectra.

### **1.2.9 Bioinformatic analysis**

#### *Qualitative analysis for liver samples*

Data files were search against a rat UniprotKB database (downloaded 3/2012) using Mascot version 2.2 (Matrix Science, Boston, MA) [76]. Mascot parameters were as follows: enzyme set as trypsin with maximum of 2 missed cleavages; precursor ion error tolerance was set at 50 ppm for QSTAR, 20 ppm for LTQ-FT, and 7 ppm for Orbitrap data; fragment ion tolerance was set at 0.2 Da; static modification of  $C_{13}H_{22}N_4O_2S$  (+298.1463) on lysine, protein N-terminus, histidine, and cysteine; mass addition of  $C_{16}H_{24}N_4O_2S$  (+336.1620), a bi-acrolein cyclic adduct, on lysine; ignore mass additions of  $C_{10}H_{15}N_2O_2S$  (+227.0854 Da),  $C_{15}H_{25}N_4O_2S_2$  (+357.1419 Da),  $C_{10}H_{18}N_3O_2S$

(+244.1120 Da),  $C_{13}H_{25}N_4O_2S_2$  (+333.1419 Da), and  $C_{13}H_{25}N_5O_2S_2$  (+347.1450 Da) on cysteine; variable modification of acrolein/biotin label (+298.1463) on cysteine, lysine, histidine and protein N-terminus (the cyclic lysine adduct was not found in preliminary studies); variable oxidation of methioinine; and variable carboamidomethyl on cysteine. Identified peptides with a Mascot score of 25 or greater were segregated into nonredundant protein groups using ProteoIQ version 2.1 (NuSep, Bogart, GA). Peptides with Mascot scores of 25 - 30 were manually validated. Theoretical MS/MS fragments derived from acrolein/biotin-labeled peptides were generated using ProteinProspector, a publicly-accessible web-based program that allows users to define new and unusual amino acids. For example, carboamidomethylation of cysteine results in a new amino acid that is defined as  $C_5H_8N_2O_2S_1$ . Of relevance to this work, amino acids with acrolein modification *via* 1, 4-Michael addition as defined as follows:  $C_{19}H_{34}N_6O_3S_1$  for lysine,  $C_{19}H_{29}N_7O_3S_1$  for histidine, and  $C_{16}H_{27}N_5O_3S_2$  for cysteine.

#### Quantitative analysis for diethylated BSA

Data files were searched by Mascot against Swiss-Prot bovine database (released in 03/2012). Peptide error tolerance was set at 0.1 Da and MS/MS error tolerance was set at 0.2 Da. Enzyme was set as trypsin with maximum 2 miss cleavages. Variable modifications were set at +56 (light) and +60 (heavy) for lysine and N-terminal amine; +298.1463 for cysteine, lysine, and histidine; oxidation for methioinine, and carboamidomethyl for cysteine. Carbonylated peptides detected before diethyl labeling were excluded from the list for quantification. Peak areas of identified peptides were manually integrated in extracted ion chromatography using Analyst QS 2.0.

## 1.3 RESULTS AND DISCUSSION

### 1.3.1 Method optimization

#### Method applied in this study

Recent developments in bottom-up proteomics have enabled the research community to identify post-translation modifications of proteins directly *via* sequencing peptides. Biotin affinity-tagging provides an opportunity to enrich low abundant carbonylated proteins allowing for their efficient detection by tandem mass spectrometry. A peptide-centered enrichment method was employed in this study as shown in Fig.1.3. Specifically, carbonylated proteins were reacted with biotin hydrazide. Formed hydrazone bonds were reduced by sodium cyanoborohydride. After trypsin digestion, the complex mixture of peptides was loaded onto a monometric avidin cartridge. Acrolein/biotin modified peptides were specifically selected by avidin while non-biotinlyated peptides were removed by washing buffers. Finally, bound peptides were eluted and analyzed by LC-MS/MS. Spiess *et al.* demonstrated the successful application of this method in lung epithelial cells by identifying relevant protein targets of acrolein [65].

#### Optimization of buffer conditions in avidin enrichment

At first, buffer conditions from a paper were exactly copied in this study since Spiess *et al* successfully identified protein-acrolein adducts in lung epithelial cell [65]. Basically, the avidin cartridge was washed by PBS and peptides were eluted by d-biotin in PBS.

After being concentrated and desalted by a C18 cartridge, the eluate was analyzed by LC-MS/MS. By applying these purification steps, in 200  $\mu$ M acrolein-exposed liver 1, 22 acrolein-modified peptides corresponding to 19 proteins were identified by Qstar Elite. Although carbonylated peptides/proteins were characterized, high abundant singly charged molecules were observed in MS spectrum. These contaminants which should come from the avidin column could not be removed by C18 column. Those compounds suppressed signals of carbonylated peptides, so it is necessary to carry out a purification step to remove the contaminants. Because of this problem, another combination of washing and eluent buffers were utilized to enrich acrolein/biotin labeled peptides. Basically the avidin cartridge was washed by a series of PBS, 80%AMBIC/20% MeOH, water, and finally biotinylated peptides were eluted by 0.4% TFA/30% ACN. After concentration and regeneration in an aqueous solvent, enriched peptides were directly analyzed by LC-MS/MS. With half sample loading amount (0.5 mg compared to 1 mg) in avidin cartridge, 47 acrolein-modified peptides corresponding to 32 proteins were found in the same protein sample in Qstar. MS spectrum showed no trace of high abundant contaminants any more. The result suggested that optimized solvents increased identification numbers of acrolein-adducted peptides. It also simplified the procedures since generated eluate did not need further purification. In addition, up to 80% peptides detected in the eluate were acrolein/biotin modified which indicated the high enrichment efficiency of avidin.

Acrolein-modified and biotin-tagged BSA was synthesized to be used as a standard biotinylated protein. Sometimes avidin reduced or even totally lost binding activity

towards biotin, so it was helpful to evaluate the performance of the cartridge before purifying real biological samples. Besides avidin activity testing, biotinylated BSA was used to confirm the elution buffer 0.4% TFA/30% ACN was better than aqueous 0.4% TFA. Acidic water without organic phase was also used to elute biotinylated peptides from avidin columns in the literature [77]. The elution ability of aqueous 0.4% TFA was compared to 0.4% TFA/30% ACN by analyzing the same acrolein/biotin labeled BSA tryptic peptides by Qstar. There were 10 acrolein/biotin labeled BSA peptides identified in 0.4% TFA/30% ACN and only 7 peptides in 0.4% TFA. The result illustrated ACN increased the elution efficiency, thus 0.4% TFA/30% ACN was used in future experiments.

#### Choice of mass spectrometer

A mass spectrometer with faster sequencing speed and sensitivity is expected to increase peptide identification number. Peptide ID capabilities of Qstar Elite and LTQ-FT were evaluated by running the same sample. In 200  $\mu$ M acrolein-exposed liver 1, 127 peptides/81 proteins were detected in LTQ-FT compared to 47 peptides/32 proteins in Qstar Elite. LTQ-FT demonstrated its advantage of rapid duty cycle time, which was the time of an ion produced in the source being effectively analyzed. In Qstar Elite, the accumulating time was 2 secs and there were 3 MS/MS analysis (the upmost was 4 MS/MS scans) following a MS scan. In LTQ-FT, it took only 250 ms to analyze one particular ion and it sequenced 9 ions (upmost 10) following a MS screening. LTQ-Orbitrap was the state-of-the-art mass spectrometer with high sensitivity and fast duty

cycle, so it was expected to sequence more peptides than Qstar Elite and LTQ-FT. Final protein samples from acrolein-exposed liver 2 were analyzed in LTQ-Orbitrap.

#### Sample loading amount on LC-MS/MS

Due to the detection limitation of a mass spectrometer, increasing sample amount may be critical for detecting more carbonylated peptides. Increasing sample loading amount could increase peak intensities in MS, so those ions previously below detection limit could be selected for MS/MS sequencing. Using 200  $\mu$ M acrolein-exposed liver 1 sample, the difference of identification numbers were compared between 0.1 mg and 0.5 mg sample loading in avidin cartridge followed by LTQ-FT analysis. As a result, 25 peptides/23 proteins were detected in 0.1 mg sample and 127 peptides/81 proteins in 0.5 mg sample. Thus 0.5 mg liver protein samples were used for LC-MS/MS characterization in this study.

#### Other valuable points

There are several points that needed to pay attention to avoid mistakes in the biotin labeling experiment. First, the buffer in which biotin hydrazide is added to derivatize carbonylated proteins should not contain any component that has reactivity with carbonyls. The commonly used ammonium bicarbonate buffer and Tris buffer in biological samples are not suitable for the biotin labeling reaction. Secondly, biotin hydrazide has to be removed completely in the sample before avidin cartridge isolation of acrolein/biotin modified peptides, since biotin hydrazide will compete with biotinylated

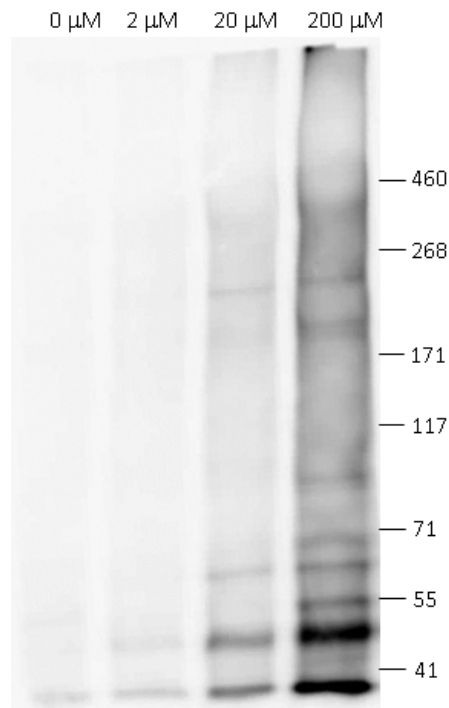
peptides binding to avidin and reduce their retention. There are multiple ways to eliminate biotin hydrazide from soluble proteins, such as gel filtration, dialysis, centrifugal filtering, and TCA precipitation. Acetone precipitation is a widely accepted method to remove non-protein component such as detergents from protein samples. Be aware that acetone precipitation is not a good way to remove biotin hydrazide because biotin hydrazide could not be well solubilized in acetone.

### **1.3.2 Results**

#### *Confirmation of protein carbonylation by western blotting*

Proteins were exposed to several concentrations of acrolein (2, 20, 200  $\mu\text{M}$ ) since the exact concentration of acrolein in the liver under normal physiological conditions or disease states was unknown. It was proposed that acrolein was in micromolar level. A study showed that the maximal blood concentration of acrolein was 10.2  $\mu\text{M}$  on days 2 in patients receiving cyclophosphamide, by i.v. infusion (60 mg/kg over 1 h) for 2 consecutive days [78]. Other *in-vitro* studies of acrolein's toxicity within the liver were also constantly in micromolar level [79-81]. The western blotting of biotinylated proteins illustrated increasing protein carbonylation with rising acrolein concentrations in a dose-dependent manner (Fig. 1.5). Carbonylated liver proteins induced by acrolein were tagged by biotin-hydrazide. The antibody HRP-conjugated streptavidin recognized biotinylated proteins, so the blot evaluated the level of total protein biotinylation. Although signals in western blot were not only from acrolein-adducts but also other carbonylated proteins and naturally biotinylated proteins, increased signals compared to

control were surely from acrolein adducts because acrolein-acute exposure was the only perturbed condition within those samples.



**Figure 1.5. Immunoblotting of biotinylated proteins in acrolein-exposed (0, 2, 20, 200 μM) rat liver microsomes.** After acrolein modification, microsomal proteins were exposed to biotin hydrazide, by which carbonyl groups were labeled with biotin. Proteins were then separated by SDS-PAGE and transferred to a PVDF membrane where the antibody HRP-conjugated streptavidin was added to detect the biotinylated proteins. The result showed increased protein carbonylation level with increased acrolein concentration in a dose-dependent manner.

#### Overview of results from LC-MS/MS analysis

By LC-MS/MS analysis followed by Mascot searching, peptides with acrolein 1,4-Michael addition at cysteine, histidine and lysine were identified with a mass shift of 298.15 Da due to the acrolein/biotin label. Bi-acrolein lysine adducts were not found in



this study. All identified peptides had only one single modification site. Consistent with western blotting result, obtained LC-MS/MS data demonstrated increased acrolein-protein adducts with increased dose of acrolein. There were 0 peptides/0 proteins, 29 peptides/18 proteins, 107 peptides/70 proteins, and 271 peptides/159 proteins identified respectively in 0, 2, 20, and 200  $\mu\text{M}$  acrolein-exposed microsomes in liver 2 by LTQ-Orbitrap analysis. Combined with 155 peptides screened by LTQ-FT in liver 1, totally 303 peptides/173 proteins in 200  $\mu\text{M}$  acrolein-exposed microsomes were detected.

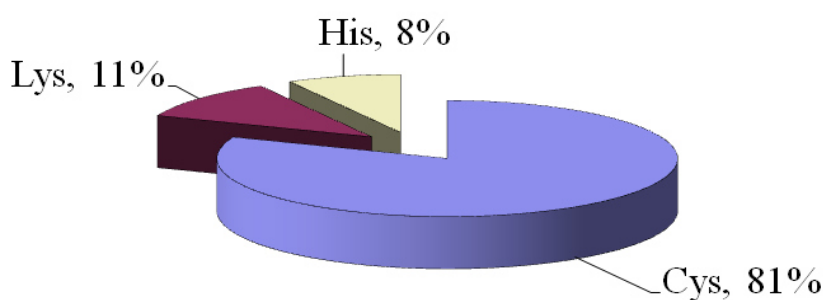
**Table 1.1.** Examples of identified acrolein-targeted peptides under different concentrations of acrolein. Identified peptides with modification sited marked by red text and corresponding proteins were listed.

Acrolein	Identified peptides in selected proteins		
	60S ribosomal protein L11	Cytochrome P450 2E1	B-cell receptor-associated protein 31
2 $\mu\text{M}$	TG <b>C</b> IGAK	N/A	N/A
20 $\mu\text{M}$	TG <b>C</b> IGAK	EHLQSLDIN <b>C</b> AR	N/A
200 $\mu\text{M}$	TG <b>C</b> IGAK	EHLQSLDIN <b>C</b> AR	LE <b>K</b> AENEALAMQK
		AKEHLQSLDIN <b>C</b> AR	GTAEDGG <b>K</b> LVDVGSP <b>E</b> MK
		DVTD <b>C</b> LLIEM <b>E</b> K	<b>K</b> GTAEDGG <b>K</b>
		<b>L</b> CVIPR	<b>L</b> KDELAST <b>K</b>
			<b>K</b> QAESASEA <b>A</b> K

This study provided a chance to determine the dynamic change in acrolein-targeted proteins and sites under different concentrations of acrolein. Proteins identified in the low concentration of acrolein might be more biologically relevant, while those in high concentration indicated more possible modification sites. Table 1.1 gave some example proteins that were modified in various concentrations of acrolein. Only one cysteine site

in 60S ribosomal protein L11 was targeted by acrolein starting from the low concentration of 2  $\mu\text{M}$  to high concentration of 200  $\mu\text{M}$ . Increased modification sites were identified from 20  $\mu\text{M}$  to 200  $\mu\text{M}$  in Cytochrome P450 2E1. B-cell receptor-associated 31 was not identified in 2  $\mu\text{M}$  and 20  $\mu\text{M}$  acrolein, but started to be targeted in multiple sites by acrolein at 200  $\mu\text{M}$ . The details of targeted proteins and peptides in different concentrations of acrolein were listed in section 1.4-identification tables.

Relative reactivity of amino acids to acrolein



**Figure 1.6. Preferred location of Michael-type acrolein addition.** Most peptides (81 %) were modified on the cysteine, while histidine and lysine only took 8 % and 11 % respectively. This result was consistent with the literature that indicated the sulfhydryl group was the most reactive group in proteins [9, 65, 82, 83].

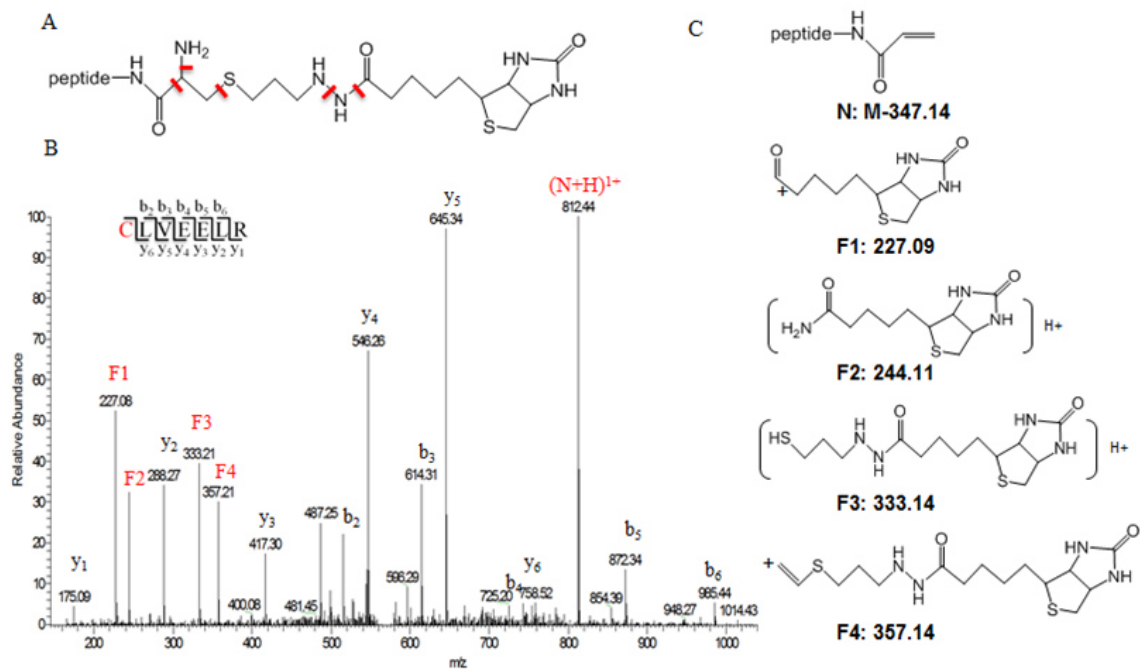
Among 303 identified acrolein-targeted peptides, the relative percentages of three types of residues with Michael-type addition were plotted as a pie chart shown in Figure 1.6. It demonstrated that most peptides (81%) were modified on cysteine residue, while histidine and lysine only took the percentages of 8 % and 11 % respectively. This result was consistent with the conclusions from previous literatures. Not only quantum mechanical and kinetic data indicated the sulfhydryl group in cysteine residue was the most reactive nucleophile in proteins [9, 82], but also proteomic data in human lung epithelial cells and

rat cardiac mitochondrial supported the theory that cysteine was the major modification site of acrolein [65, 83]. However, reactivity of residue did not directly relate to the significant consequence of adduction. The impact of acrolein was dependent on the role of the alkylated residue in the protein's function and structure [12].

#### Fragmentation of the acrolein/biotin label in MS/MS

Mascot searching allowed accurate identification of acrolein-modified peptides. The acrolein/biotin label showed in Figure 1.3 gave the peptide an increased mass of 298.15 Da. Besides the significant mass shift in precursor ions, the acrolein/biotin label created specific fragments in MS/MS spectrum of a peptide. Figure 1.7B presented a representative MS/MS spectrum from a peptide with N-terminal cysteine modified by acrolein/biotin. The peptide sequence and observed b and y ions were indicated in Figure 1.7A. Besides the peaks from b and y ions (marked in black) that were generated by cleavage of amide bonds of the peptide, there were several abundant peaks that were not assigned to any type of common peptide fragments. These ions were constantly appeared in spectra of carbonylated peptides. They were from the acrolein/biotin label. The structures matching the m/z for fragments were proposed which were indicated in Figure 1.7C. This was the first time that the fragmentation of the acrolein/biotin label was comprehensively annotated. Generally acrolein-cysteine adducted peptides had fragments of 227.09, 244.11, 333.14, and 357.14 Da. Specifically in N-terminal cysteine, the molecular ion reduced 374.14 Da (labeled as N) due to loss of the terminal amine and break of the thioether bond. Mascot score of this spectrum matching to the peptide was only 26 although almost all b and y ions were found. The score was low because

abundant fragment ions from the acrolein/biotin label were not in the database of theoretically generated peaks. Although in Mascot configuration there were options to “ignore ions” and edit “neutral losses”, it did not work well. Take the scoring of shown



**Figure 1.7. Characteristic cleavage fragments of the acrolein-biotin label.** (A) Structure of an acrolein-biotin labeled N-terminal Cys residue with proposed cleavage sites indicated in red. (B) MS/MS spectrum of the CLVEELR peptide (from cytochrome P450) with acrolein-modified Cys. Proposed structures for the observed fragment ions indicated in red were shown in (C).

spectrum from peptide CLVEELR for example. When “ignore mass” of F1 to F4 were configured, the Mascot score was remaining at 26. When neutral loss of 347.14 Da at cysteine was added into Mascot, the score even dropped to 20. These observations were consistent with a published paper where they claimed that Mascot scores were remaining low for biotin labeled peptides after including ignored masses and neutral losses [84].

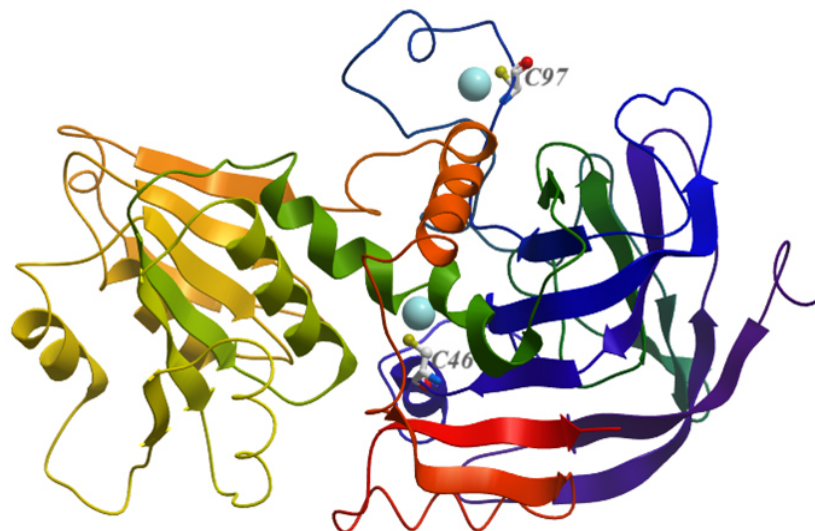
The high abundant characteristic ions were constantly observed in low-score spectra because the unassigned peaks brought down the Mascot score. They were not present in all spectra of acrolein/biotin labeled spectra because bonds in the acrolein/biotin label were not preferentially cleaved compared to amide bonds. In this study, in order to ensure the accurate identification of acrolein-modified peptides, all spectra which were given Mascot scores from 25 to 30 were manually validated. In future, developing a searching algorithm capable of taking specific fragment ions into scoring is very necessary and important for protein carbonylation analysis.

### **1.3.3 Effects of acrolein on proteins**

#### *Site-specific effects*

This study concentrated on identification of acrolein modification sites, which provided the opportunity to evaluate effects of acrolein on proteins. Not all modifications by acrolein changed functions and structures of proteins. Only modification on amino acids that played important roles in protein's stability and function would lead to significant consequence. The following were two examples of distinguished acrolein-protein adducts in important modification sites. Two acrolein-modified sites, Cys 46 and Cys 97, were identified in alcohol dehydrogenase. Figure 1.8 showed the crystal structure of human alcohol dehydrogenase 1 (PDB: 1HTB) with two cysteine sites marked. Cys 46 coordinated one zinc atom responsible for substrate binding and Cys 97 coordinated second zinc that was crucial for protein stability [85, 86]. Derivatization of these residues by acrolein likely led to protein deactivation. Previous studies suggested administration of allyl alcohol in rat decreased the activity of alcohol dehydrogenase in

the liver [87, 88]. Our result suggested that it might be due to alkylation on those Cys sites by its metabolite acrolein. The other example was microsomal glutathione S-transferase 1 (MGST1), an oxidation and chemical stress sensor. A study demonstrated *in-vitro* incubation of liver microsomes with acrolein activated MGST [89]. The other study showed MGST increased its activity when Cys 49 was alkylated [90]. Our result suggested that this increase activity of MGST by acrolein was highly likely to be caused by alkylation on Cys 49 since carbonylation of this site was detected starting from 2  $\mu$ M to 200  $\mu$ M of acrolein. In all, whether the effect of acrolein carbonylation to a protein was significant or not was largely depending on the fact that if the amino acid residue was crucial to the function or structure of the protein. This study profiled targeted proteins and modified amino acid residues to evaluate the consequence of acrolein modification. The results benefited the understanding of the impact of acrolein on liver microsomal proteins. In future, if researchers are interested in a particular protein from our list, studies such as enzyme activity assay could be conducted to confirm the effects of acrolein.



**Figure 1.8. The crystal structure of human alcohol dehydrogenase 1 (PDB:1HTB).** Two modified sites were detected: Cys 46 and Cys 97. Derivatization of these residues likely led to protein deactivation since Cys 46 coordinated one zinc atom (shown as a light blue ball) responsible for substrate binding and Cys 97 coordinated a second zinc that was crucial for protein stability [85, 86].

#### Identification of functionally relevant protein targets

The aim of this study was to elucidate the molecular mechanism of the toxicity of acrolein within the liver by characterizing the protein targets in the liver microsomes. In the list of protein targets identified at 200  $\mu$ M acrolein, carbonylated enzymes which might be involved in metabolic pathways of acrolein were present. There were three major metabolic pathways of acrolein reported (Fig. 1.3) [16]: transferred to acrylic acid in cytosol and microsomes in the liver by aldehyde dehydrogenase; metabolized to glycidaldehyde by epoxidase and cytochrome P450 in liver microsomes; conjugated with glutathione directly or catalyzed by glutathione S-transferase. Characterized isotypes of carbonylated aldehyde dehydrogenase, glutathione S-transferase and cytochrome P450

were listed in Table 1.2. The details of modification sites of these proteins were shown in section 1.4-identification tables.

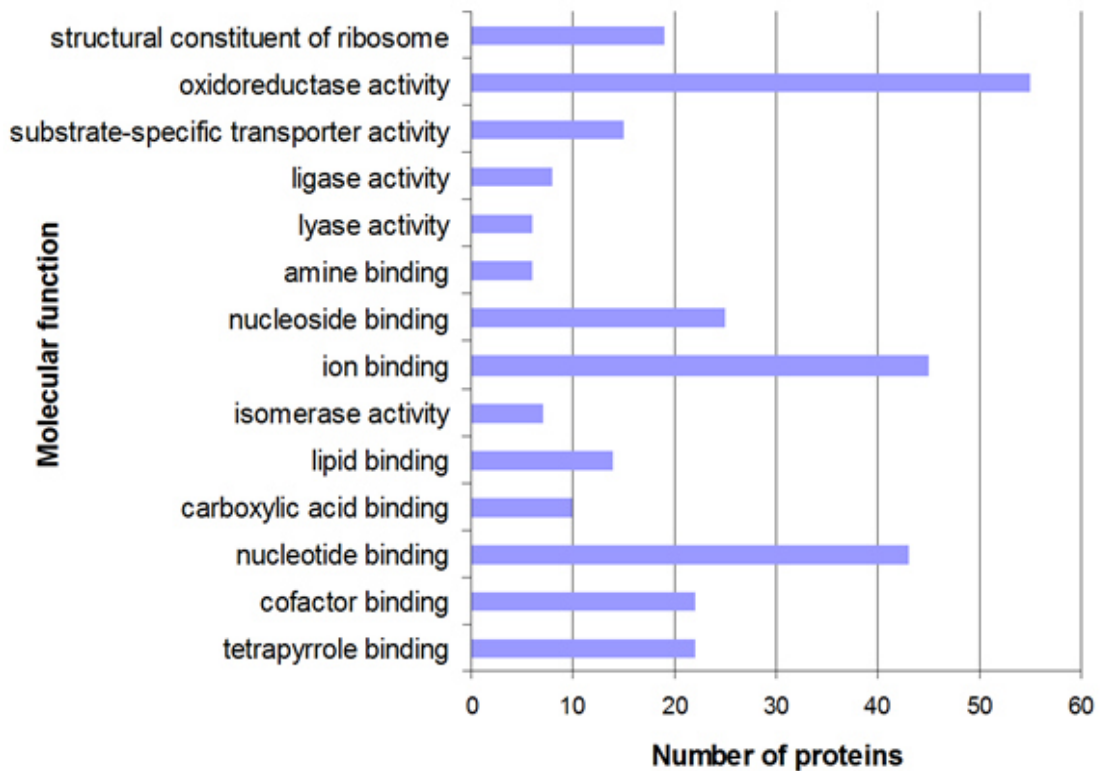
**Table 1.2.** Acrolein-targeted proteins that might be involved in the metabolism of acrolein. The list included identified isotypes of aldehyde dehydrogenase, glutathione S-transferase, and cytochrome P450 which were the enzymes in the metabolism pathways of acrolein.

Protein names	Identified protein isotypes
Aldehyde dehydrogenase	ALDH3A2
Glutathione S-transferase	MGST1, GSTM2
Cytochrome P450	CYP1A2, CYP2A1, CYP2A2, CYP2B3, CYP2C11, CYP2C12, CYP2C23, CYP2C6, CYP2C7, CYP2D10, CYP2E1, CYP3A18, CYP3A2, CYP4A2, CYP4F1, CYP4F6, CYP4V2

Acrolein-modified proteins were classified according to their molecular functions based on gene ontology annotation (Fig. 1.9). Identified acrolein-carbonylated proteins included a wide range of molecular functions such as transporter activity, ligase activity, and ion binding activity. It indicated that acrolein affected multiple protein cellular activities and cellular responses, which possibly were results of a combination of cellular molecular mechanisms. In the categorization chart, 54 proteins out of identified 173 proteins contained oxidoreductase activity. The published literature suggested that both allyl alcohol [91] and cyclophosphamide [40] reduced oxidase activity in the liver. Our data suggested that the liver toxicity induced by these compounds might be due to malfunction of oxidoreductase caused by acrolein, an intermediate in the metabolism of cyclophosphamide. The figure showed 19 ribosomal proteins were modified by acrolein. Ribosomes were in charge of biological protein synthesis within cells. Modification of



ribosomal proteins by acrolein might jeopardize the proper functioning of ribosomes. In fact, allyl alcohol was reported to reduce protein synthesis in an *in-vitro* study using rat liver slices [92]. Allyl alcohol was not toxic itself since it did not have reactivity towards cellular molecules. Our data indicated that the metabolite acrolein might take responsibility for the liver toxicity of allyl alcohol, partially by reducing protein synthesis *via* carbonylation of ribosomal proteins. Comparing acrolein accessible proteins



**Figure 1.9. Classification of identified proteins according to Gene Ontology (GO) molecular function ( $p < 0.05$ ).** The bar showed the number of proteins corresponding to a category of molecular function. For example, fifty-four proteins possessed oxidoreductase activity and nineteen proteins were structural constituent of ribosome.

identified in liver microsomes in this study to other previous studies, various protein targets were in common. For example, 50 ribosomal proteins and 46 proteins possessed

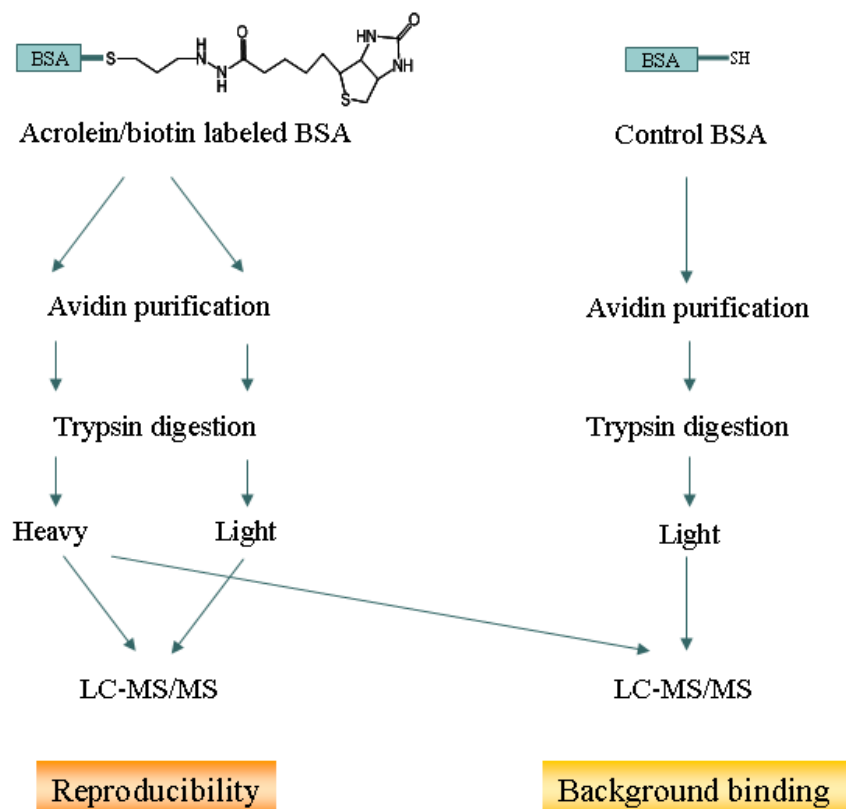
oxidoreductase activity were carbonylated by acrolein in human lung epithelial cells [65]. Voltage-dependent anion channel protein (VDMA), which was an ion channel protein that facilitated exchange of ions and molecules, was alkylated by acrolein in liver microsomes in this study, while it was also found carbonylated by acrolein in gerbil synaptosomes [50] and in aged-rat cardiac tissues at the same modification site (Cys 232) [83]. There were other types of toxic reactive electrophiles which targeted same protein sites as acrolein. Guo *et al.* profiled carbonylated proteins in rat liver mitochondria induced by *in-vitro* exposure of a reactive carbonyl species 4-HNE [93]. Carbamoyl-phosphate synthase was modified by 4-HNE at five sites, of which His 47, His 817, and His 1162 were also modified by acrolein in our study. Shin *et al.* profiled targeted proteins to reactive electrophiles in human liver microsomes using a biotin tagged standard electrophile [94]. Carbamoyl-phosphate synthase which was present in our table was also in their list of electrophile accessible proteins. Nowadays, there is growing interests in characterization of protein carbonylation in the proteomic field. With more studies conducted, the understanding of molecular mechanisms of acrolein in cells would be more comprehensive in future.

#### **1.3.4 Diethylation quantification**

A quantification method named diethylation was utilized to determine the background-binding proteins from avidin column in protein-level purification. The idea was, if a protein was non-specifically attaching to the avidin column, the binding amount of the protein should be the same in acrolein treated and non-treated (control) samples. If a

protein was a real acrolein target, the amount of protein in acrolein treated sample after avidin purification should be much more than the protein amount in control sample.

Figure 1.10 represented the proteomic strategy for identifying real carbonylated proteins by diethylation. Bovine serum albumin (BSA) was carbonylated by acrolein and tagged with biotin to be used as a standard carbonylated protein. To test the reproducibility of avidin purification and diethyl labeling method, two aliquots of acrolein/biotin labeled BSA were labeled with light and heavy acetaldehyde respectively after avidin purification and trypsin digestion. Table 1.3 indicated that quantification on multiple BSA peptides gave an average quantification ratio of 1 : 1.3 for two aliquots of carbonylated BSA, which was close to the expected ratio of 1 : 1. To determine the background binding ratio, Acrolein/biotin labeled BSA was mixed with control BSA after diethylation (Fig. 1.10). Since control BSA did not contain biotin group, theoretically it should not be retained in avidin cartridge. In reality, it still existed in avidin eluate due to electrostatic interaction. The result showed that the amount of carbonylated BSA versus the amount of control BSA was 1 : 0.2 based on diethylation quantification (Table 1.3). The experiments showed that background binding proteins could be distinguished from carbonylated proteins by having a lower binding ratio to avidin column.



**Figure 1.10. The strategy for determining the reproducibility and background binding for avidin purification using diethylation quantification.** Acrolein/biotin labeled BSA were separated into two aliquots and purified by an avidin cartridge. After labeled with  $^{13}\text{C}/^{12}\text{C}$  acetaldehyde reagents, an equal amount of tryptic peptides from carbonylated BSA were mixed and analyzed by LC-MS/MS to evaluate the reproducibility of the quantification method. Unmodified control BSA was purified, digested and labeled with light acetaldehyde in the same way as acrolein/biotin labeled BSA. An equal amount of modified BSA peptides were mixed with control BSA peptides eluted from avidin and analyzed by LC-MS/MS to determine the background binding.

Although the experiment with standard BSA showed that diethylation quantification could distinguish non-carbonylated proteins from real acrolein-modified proteins, there was a problem using diethyl labeling on carbonylated proteins: the acrolein/biotin labeled peptides were missing after reacting with acetaldehyde. The mass-to-charge ratios of acrolein/biotin labeled peptides with and without diethyl groups were manually searched in obtained data. There was no trace of either. It was very possible that acrolein/biotin

labeled peptides had reacted with acetaldehyde but the reaction did not lead to a mass shift of 56 Da or 60 Da as other non-modified peptides. In the literature, Meany *et al.* [67] used another quantitative labeling technique named Isobaric Tag for Relative and Absolute Quantitation (ITRAQ) to characterize carbonylated proteins. The reagent of ITRAQ reacted with amino groups of peptides which was similar to diethylation labeling. Their experimental results showed barely any carbonylated peptide directly identified after ITRAQ labeling. It was possible that ITRAQ reagent caused their carbonylated peptides missing.

**Table 1.3.** Results of diethylation quantification for avidin purification of acrolein/biotin labeled BSA (ACR BSA) and control BSA. The BSA peptides were analyzed by Qstar Elite. Peak areas were integrated for each peptide in the extracted ion chromatographs. Ratios were calculated based on peak areas.

Peptides	ACR BSA Heavy	ACR BSA Light	ACR BSA H: ACR BSA L	ACR BSA Heavy	control BSA Light	ACR BSA: control BSA
LVTDLTK	3.30E+06	3.11E+06	1: 0.9	2.85E+06	6.77E+05	1: 0.2
DDSPDLPK	5.13E+05	9.98E+05	1: 1.9	8.98E+05	1.16E+05	1: 0.1
AEFVEVTK	1.46E+06	2.50E+06	1: 1.7	3.29E+06	7.59E+05	1: 0.2
LVNELTEFAK	5.02E+06	5.79E+06	1: 1.2	1.39E+06	2.34E+05	1: 0.2
KVPQVSTPTLVEVSR	2.28E+06	4.07E+06	1: 1.8	1.18E+06	2.49E+05	1: 0.2
RHPYFYAPELLYYANK	2.03E+06	3.69E+06	1: 1.8	1.09E+04	0.19E+04	1: 0.2
YLYEIAR	4.70E+06	5.20E+06	1: 1.1	4.15E+06	7.69E+05	1: 0.2
LVVSTQTALA	3.90E+06	4.20E+06	1: 1.1	2.42E+05	2.32E+04	1: 0.1
QTALVELLK	5.68E+06	5.23E+06	1: 0.9	7.20E+05	9.70E+04	1: 0.1
RHPEYAVSVLLR	2.03E+06	3.69E+06	1: 1.8	3.74E+05	1.68E+05	1: 0.4
DAFLGSFLYEYSR	1.10E+06	9.28E+05	1: 0.8	7.68E+04	1.86E+04	1: 0.2
KQTALVELLK	5.90E+06	5.66E+06	1: 1.0	3.00E+05	7.40E+04	1: 0.2
Average±SD			1: 1.3±0.4			1: 0.2±0.1

### **1.3.5 Accomplishments**

We conducted a peptide-centered analysis of protein carbonylation induced by acrolein. An analysis strategy combining biotin tagging of carbonylated proteins, avidin enrichment of biotinylated peptides and LC-MS/MS characterization of acrolein/biotin labeled peptides was utilized. In this study, the buffer conditions for avidin isolation of acrolein/biotin tagged peptides were optimized. The most relative reactive amino acid to acrolein was verified to be cysteine. Fragments specifically from the acrolein/biotin label in tandem mass spectrometry were determined, which facilitated the identification of peptides in MS/MS spectrum. Functionally relevant proteins were characterized including enzymes involved in acrolein's metabolism and important structurally functional proteins that might be responsible for acrolein's liver toxicity. In conclusion, our study confirmed the well application of an analysis method for characterizing carbonylated proteins with modification sites induced by acrolein. Our results also provided a molecular level explanation for the potential role of protein carbonylation in acrolein's toxicity in the liver.

## 1.4 IDENTIFICATION TABLES

Three tables (Table 1.4, 1.5, 1.6) showed the identified carbonylated peptides/proteins in rat liver microsomes exposed to three concentrations (2  $\mu$ M, 20  $\mu$ M, 200  $\mu$ M) of acrolein respectively.

Table 1.4. Identified modified peptides and corresponding proteins at 2  $\mu$ M acrolein

No	Sequence Id	Protein Name	Score <sup>#</sup>	Peptide Sequence	Modification*
1	splQ62812IMYH9_RAT	Myosin-9 GN=Myh9 PE=1 SV=3			
1			35.4	KLEEDQIIMEDQNCK	+298.15 (C14)
1			37.4	LEEDQIIMEDQNCK	+15.99 (M8); +298.15 (K14)
1			27.6	KMEDGVGCLETAEAAK	+15.99 (M2); +298.15 (C8)
1			33.8	MEDGVGCLETAEAAK	+298.15 (C7)
1			31.8	KGTGDCSDEEVDGK	+298.15 (C6)
1			41.1	MEDGVGCLETAEAAK	+15.99 (M1); +298.15 (C7)
1			36.0	KMEDGVGCLETAEAAK	+298.15 (C8)
1			51.0	LEEDQIIMEDQNCK	+298.15 (K14)
2	splP05179ICP2C7_RAT	Cytochrome P450 2C7 GN=Cyp2c7 PE=1 SV=2			
2			39.1	ACVGEGLAR	+298.15 (C2)
2			72.8	VQEEAQCLVEELR	+298.15 (C7)
2			78.4	VQEEAQCLVEELRK	+298.15 (C7)
3	splQ5XIU9IPGRC2_RAT	Membrane-associated progesterone receptor component 2 GN=Pgrmc2 PE=1 SV=1			
3			42.8	GLATFCLDK	+298.15 (C6)
3			37.6	GLCSGPGAGEESPAATLPR	+298.15 (C3)
4	splO88813IACSL5_RAT	Long-chain-fatty-acid--CoA ligase 5 GN=Acs15 PE=2 SV=1			
4			73.0	GSFEELCQNQCVK	+57.02 (C7); +298.15 (C11)
5	splP24470ICP2CN_RAT	Cytochrome P450 2C23 GN=Cyp2c23 PE=2 SV=2			
5			25.3	NGCFK	+298.15 (C3)
5			36.8	DYIDCFLSK	+298.15 (C5)
6	splO54753IH17B6_RAT	17-beta-hydroxysteroid dehydrogenase type 6 GN=Hsd17b6 PE=1 SV=2			
6			58.0	ESYQQQFFDDFCNTTR	+298.15 (C12)
7	splP08011IMGST1_RAT	Microsomal glutathione S-transferase 1 GN=Mgst1 PE=1 SV=3			
7			56.6	VFANPEDCAGFGK	+298.15 (C8)
8	splP07153IRPN1_RAT	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1 GN=Rpn1 PE=2 SV=1			
8			51.2	TEGSDLCDR	+298.15 (C7)
9	splQ9JHZ9IS38A3_RAT	Sodium-coupled neutral amino acid transporter 3 GN=Slc38a3 PE=2 SV=1			
9			41.9	AEDAQHCGEK	+298.15 (C7)
10	splP11507-2IAT2A2_RAT	Isoform SERCA2A of Sarcoplasmic/endoplasmic reticulum calcium ATPase 2 GN=Atp2a2			
10			40.0	VGEATETALTCLVEK	+298.15 (C11)
11	splQ3B7U9IFKBP8_RAT	Peptidyl-prolyl cis-trans isomerase FKBP8 GN=Fkbp8 PE=2 SV=1			
11			35.4	SCSQVLEHQPDNIK	+298.15 (C2)
12	splQ7TQM4ISOAT2_RAT	Sterol O-acyltransferase 2 GN=Soat2 PE=2 SV=1			
12			34.4	QGEEQENGACGEGNTR	+298.15 (C10)
13	splP27952IRS2_RAT	40S ribosomal protein S2 GN=Rps2 PE=1 SV=1			
13			33.9	GCTATLGNFAK	+298.15 (C2)
14	splP62914IRL11_RAT	60S ribosomal protein L11 GN=Rpl11 PE=1 SV=2			
14			29.6	TGCIGAK	+298.15 (C3)

15	splP04799 CP1A2_RAT	Cytochrome P450 1A2 GN=Cyp1a2 PE=1 SV=2		
15		29.4	TCEHVQAWPR	+298.15 (C2)
16	tr D3ZEA0 D3ZEA0_RAT	Fibronectin type III domain-containing 3a GN=Fndc3a PE=4 SV=1		
16		29.0	GCTQVDQEIEEK	+298.15 (C2)
17	tr F1LW74 F1LW74_RAT	IQ motif-containing GTPase-activating protein 2 (Fragment) GN=LOC100360623 PE=4 SV=1		
17		26.5	NSCISEEER	+298.15 (C3)
18	splP18163 ACSL1_RAT	Long-chain-fatty-acid--CoA ligase 1 GN=Acs1l PE=1 SV=1		
18		26.4	CGVEIIGLK	+298.15 (C1)

# Spectra with Mascot score below 30 were manually validated

\* +298.15 is carbonylation by acrolein *via* 1,4-Michael addition on Cys (C), Lys (K) and His (H); +15.99 is oxidation at methionine (M); +57.02 is carboamidomethylation by IAA at Cys (C)



Table 1.5. Identified modified peptides and corresponding proteins at 20  $\mu$ M acrolein

	Protein Id	Name	Score <sup>#</sup>	Peptide Sequence	Modification*
1	splQ62812IMYH9_RAT	Myosin-9 GN=Myh9 PE=1 SV=3			
1			31.6	KMEDGVGCLETAEAAK	+298.15 (C8)
1			61.9	MEDGVGCLETAEAAK	+298.15 (C7)
1			30.5	VEDMAELTCLNEASVLHNLK	+298.15 (C9)
1			39.1	KMEDGVGCLETAEAAK	+15.99 (M2); +298.15 (C8)
1			34.4	KLEEDQIIMEDQNCK	+298.15 (K15)
1			43.6	MEDGVGCLETAEAAK	+15.99 (M1); +298.15 (C7)
1			81.2	LEEDQIIMEDQNCK	+298.15 (K14)
1			48.6	LEEDQIIMEDQNCK	+15.99 (M8); +298.15 (K14)
1			47.4	KLEEDQIIMEDQNCK	+15.99 (M9); +298.15 (K15)
1			34.5	KGTGDCSDEEVDGK	+298.15 (C6)
2	splP05179ICP2C7_RAT	Cytochrome P450 2C7 GN=Cyp2c7 PE=1 SV=2			
2			46.7	VQEAAQCLVEELRK	+298.15 (C7)
2			89.1	VQEAAQCLVEELR	+298.15 (C7)
2			42.0	ACVGEGLAR	+298.15 (C2)
3	splQ9EQ76IFMO3_RAT	Dimethylaniline monooxygenase [N-oxide-forming] 3 GN=Fmo3 PE=1 SV=1			
3			50.6	GTCILPSVNDMMDDIDEK	+298.15 (C3); +15.99 (M11)
3			41.6	GTCILPSVNDMMDDIDEK	+298.15 (C3); +15.99 (M12)
3			28.0	CPDFSTTGK	+298.15 (C1)
3			59.3	GTCILPSVNDMMDDIDEK	+298.15 (C3)
3			26.7	ILCGTFSIKPNVK	+298.15 (C3)
4	splP11507-2IAT2A2_RAT	Isoform SERCA2A of Sarcoplasmic/endoplasmic reticulum calcium ATPase 2 GN=Atp2a2			
4			29.2	TGTLTTNQMSVCR	+15.99 (M9); +298.15 (C12)
4			51.1	VGEATETALTCLVEK	+298.15 (C11)
4			31.7	ANACNSVIK	+298.15 (C4)
4			52.3	TGTLTTNQMSVCR	+298.15 (C12)
5	splP05178ICP2C6_RAT	Cytochrome P450 2C6 GN=Cyp2c6 PE=2 SV=2			
5			25.8	CLVEELR	+298.15 (C1)
5			27.1	MCAGEGLAR	+15.99 (M1); +298.15 (C2)
5			31.7	MCAGEGLAR	+298.15 (C2)
6	splP07153IRPN1_RAT	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1 GN=Rpn1 PE=2 SV=1			
6			30.2	LKTEGSDLCDR	+298.15 (C9)
6			37.8	VACITEQVLTLVNK	+298.15 (C3)
6			55.2	TEGSDLCDR	+298.15 (C7)
7	splP18163IACSL1_RAT	Long-chain-fatty-acid--CoA ligase 1 GN=Acs11 PE=1 SV=1			
7			25.4	GEGEVCVK	+298.15 (C6)
7			45.0	GIQVSNDBGPLGSR	+298.15 (C10)
7			25.5	TKPKPPEPEDLAICFTSGTTGNPK	+298.15 (C15)
7			36.6	CGVEIIGLK	+298.15 (C1)
8	splO88813IACSL5_RAT	Long-chain-fatty-acid--CoA ligase 5 GN=Acs15 PE=2 SV=1			
8			37.3	GLAVSDNGPCLGYR	+298.15 (C10)
8			47.0	GSFEELCQNQCVK	+57.02 (C7); +298.15 (C11)
8			36.2	KPMPPNPEDLSVICFTSGTTGDPK	+298.15 (C14)
9	splO54753IH17B6_RAT	17-beta-hydroxysteroid dehydrogenase type 6 GN=Hsd17b6 PE=1 SV=2			
9			54.6	CSTNLSLVTDCMEHALTSK	+298.15 (C1); +57.02 (C11)
9			39.0	ESYGQQFFDDFCNTR	+298.15 (C12)
10	splP24470ICP2CN_RAT	Cytochrome P450 2C23 GN=Cyp2c23 PE=2 SV=2			
10			27.3	ACVGESLAR	+298.15 (C2)

10			26.0	NGCFK	+298.15 (C3)
10			38.2	DYIDCFLSK	+298.15 (C5)
11	splQ5XIU9IPGRC2_RAT	Membrane-associated progesterone receptor component 2 GN=Pgrmc2 PE=1 SV=1			
11			39.4	GLATFCLDK	+298.15 (C6)
11			48.1	GLCSGPGAGEESPAATLPR	+298.15 (C3)
12	splP12939ICP2DA_RAT	Cytochrome P450 2D10 GN=Cyp2d10 PE=2 SV=1			
12			35.0	ITSCDIEVQDFVIPK	+298.15 (C4)
12			25.9	CLGVKPR	+298.15 (C1)
12			25.5	CPEMTDQAHMPYTNAVIHEVQR	+298.15 (C1)
13	splP16391IHA12_RAT	RT1 class I histocompatibility antigen, AA alpha chain PE=1 SV=2			
13			46.8	DSSQSSDVSLPDCK	+298.15 (C13)
14	trIF1LW74IF1LW74_RAT	IQ motif-containing GTPase-activating protein 2 (Fragment) GN=LOC100360623 PE=4 SV=1			
14			37.2	NPNAVLTCVDDSLSQEYQK	+298.15 (C8)
14			41.9	NSCISEEER	+298.15 (C3)
15	splP08011IMGST1_RAT	Microsomal glutathione S-transferase 1 GN=Mgst1 PE=1 SV=3			
15			63.9	VFANPEDCAGFGK	+298.15 (C8)
16	splQ6RUV5IRAC1_RAT	Ras-related C3 botulinum toxin substrate 1 GN=Rac1 PE=1 SV=1			
16			28.8	HHCPNTPHILVGTK	+298.15 (C3)
17	splQ0ZHH6-2IATLA3_RAT	Isoform 2 of Atlastin-3 GN=Atl3			
17			61.3	YQQELEEEITELYENFCK	+298.15 (K18)
18	trID3ZEA0ID3ZEA0_RAT	Fibronectin type III domain-containing 3a GN=Fndc3a PE=4 SV=1			
18			33.0	GCTQVDQEIEEK	+298.15 (C2)
18			27.8	GCTQVDQEIEEKDEETK	+298.15 (C2)
19	splP11510ICP2CC_RAT	Cytochrome P450 2C12, female-specific GN=Cyp2c12 PE=2 SV=1			
19			31.1	HRSPCMLDR	+298.15 (C5)
19			29.6	SPCMLDR	+298.15 (C3)
20	trIF1M5X1IF1M5X1_RAT	Ribosome-binding protein 1 (Fragment) GN=Rrbp1 PE=4 SV=1			
20			30.4	LTAEFEEAQSTACR	+298.15 (C13)
20			28.6	LKELESQVSCLEK	+298.15 (C10)
21	splQ9JHZ9IS38A3_RAT	Sodium-coupled neutral amino acid transporter 3 GN=Slc38a3 PE=2 SV=1			
21			54.0	AEDAQHCGEK	+298.15 (C7)
22	trIB0BN81IB0BN81_RAT	Ribosomal protein S5, isoform CRA_b GN=Rps5 PE=2 SV=1			
22			52.3	TIAECLADELINAALK	+298.15 (C5)
23	trIB5DF48IB5DF48_RAT	Radical S-adenosyl methionine and flavodoxin domains 1 (Predicted), isoform CRA_a GN=Tyw1 PE=2 SV=1			
23			51.7	GEGDCNAVQSK	+298.15 (C5)
24	splQ9Z2Z8IDHCR7_RAT	7-dehydrocholesterol reductase GN=Dhcr7 PE=2 SV=1			
24			51.2	AIECSYTSADGLK	+298.15 (C4)
25	splP21531IRL3_RAT	60S ribosomal protein L3 GN=Rpl3 PE=1 SV=3			
25			51.1	TVFAEHISDECK	+298.15 (K12)
26	splB0BNG2ITM6S2_RAT	Transmembrane 6 superfamily member 2 GN=Tm6sf2 PE=2 SV=1			
26			49.6	IFNQSPAPSSCTCDVVQEEQK	+57.02 (C11); +298.15 (C13)
27	splP05182ICP2E1_RAT	Cytochrome P450 2E1 GN=Cyp2e1 PE=1 SV=4			
27			48.8	EHLQSLDINCAR	+298.15 (C10)
28	splB5DEH2IERLN2_RAT	Erlin-2 GN=Erlin2 PE=1 SV=1			
28			46.8	ADAECYTALK	+298.15 (C5)
29	trID4A9Y2ID4A9Y2_RAT	RCG29101 GN=LOC691931 PE=4 SV=1			
29			42.0	GGTGGGEDDEDGAAPAGR	+298.15 (C8)
30	splQ5EB77IRAB18_RAT	Ras-related protein Rab-18 GN=Rab18 PE=2 SV=1			
30			41.8	TCDGVQCAFEEELVEK	+57.02 (C2); +298.15 (C7)
31	splP19225ICP270_RAT	Cytochrome P450 2C70 GN=Cyp2c70 PE=2 SV=1			
31			41.8	ACIGEGLAR	+298.15 (C2)

32	splA2RR29ICP4V2_RAT	Cytochrome P450 4V2 GN=Cyp4v2 PE=2 SV=1			
32			41.5	AEQDCIGAGR	+298.15 (C5)
33	splP02770IALBU_RAT	Serum albumin GN=Alb PE=1 SV=2			
33			41.1	ECCHGDLLECADDR	+57.02 (C2); +57.02 (C10); +298.15 (H4)
34	splQ3MIB4ILONP2_RAT	Lon protease homolog 2, peroxisomal GN=Lonp2 PE=1 SV=2			
34			40.5	SDVADGEGCK	+298.15 (K10)
35	splP17077IRL9_RAT	60S ribosomal protein L9 GN=Rpl9 PE=1 SV=1			
35			40.2	TGVACSVSQAQK	+298.15 (C5)
36	splP97612IFAAH1_RAT	Fatty-acid amide hydrolase 1 GN=Faah PE=1 SV=1			
36			38.3	GTNCVTSYLTDCETQLSQAPR	+57.02 (C4); +298.15 (C12)
37	splP97524IS27A2_RAT	Very long-chain acyl-CoA synthetase GN=Slc27a2 PE=1 SV=1			
37			38.1	GEVGLLICK	+298.15 (C8)
38	splQ5RJR8ILRC59_RAT	Leucine-rich repeat-containing protein 59 GN=Lrrc59 PE=1 SV=1			
38			37.9	ATVLDLSCNK	+298.15 (C8)
39	splQ9JJ2IERAP1_RAT	Endoplasmic reticulum aminopeptidase 1 GN=Erap1 PE=2 SV=2			
39			37.7	CVQQTIETIEENIR	+298.15 (C1)
40	splP27952IRS2_RAT	40S ribosomal protein S2 GN=Rps2 PE=1 SV=1			
40			36.9	GCTATLGNFAK	+298.15 (C2)
41	trID3ZIM4ID3ZIM4_RAT	Aldehyde dehydrogenase GN=Aldh3a2 PE=3 SV=1			
41			36.6	YMNCGQTCIAPDYILCEASLQDQIVQK	+57.02 (C4); +57.02 (C16); +298.15 (C8)
42	splQ3B7U9IFKBP8_RAT	Peptidyl-prolyl cis-trans isomerase FKBP8 GN=Fkbp8 PE=2 SV=1			
42			36.5	SCSQVLEHQPDNIK	+298.15 (C2)
43	splP04905GSTM1_RAT	Glutathione S-transferase Mu 1 GN=Gstm1 PE=1 SV=2			
43			35.9	KHHLCGETEER	+298.15 (C5)
44	splQ5BK32IFAF2_RAT	FAS-associated factor 2 GN=Faf2 PE=2 SV=1			
44			35.3	KLECLPPEPSPDDPDSVK	+298.15 (C4)
45	splP01026ICO3_RAT	Complement C3 GN=C3 PE=1 SV=3			
45			35.2	DSCVGTLVVK	+298.15 (C3)
46	splP08753IGNA13_RAT	Guanine nucleotide-binding protein G(k) subunit alpha GN=Gnai3 PE=1 SV=3			
46			35.0	IIHEDGYSEDECK	+298.15 (K13)
47	splP04799ICP1A2_RAT	Cytochrome P450 1A2 GN=Cyp1a2 PE=1 SV=2			
47			34.8	SFSIASDPTSVSSCYLEEHVSK	+298.15 (K22)
48	splQ498D5IRMD2_RAT	Regulator of microtubule dynamics protein 2 GN=Fam82a1 PE=2 SV=1			
48			34.7	LEECIQDELGVR	+298.15 (C4)
49	splP62909IRS3_RAT	40S ribosomal protein S3 GN=Rps3 PE=2 SV=1			
49			33.1	GLCAIAQAESLR	+298.15 (C3)
50	splP18445IRL27A_RAT	60S ribosomal protein L27a GN=Rpl27a PE=1 SV=3			
50			32.8	NQSFCTVNLDK	+298.15 (C5)
51	splQ6AYS8IDHB11_RAT	Estradiol 17-beta-dehydrogenase 11 GN=Hsd17b11 PE=2 SV=1			
51			32.4	ALTDELAALGCTGVR	+298.15 (C11)
52	splQ05962IADT1_RAT	ADP/ATP translocase 1 GN=Slc25a4 PE=1 SV=3			
52			31.6	GIIDCVVR	+298.15 (C5)
53	trID3ZJ32ID3ZJ32_RAT	RCG21039, isoform CRA_a GN=RGD1565705 PE=4 SV=1			
53			31.6	ACDLPAVWHFPDTER	+298.15 (C2)
54	splQ9JJ40INHRF3_RAT	Na(+)/H(+) exchange regulatory cofactor NHE-RF3 GN=Pdzk1 PE=1 SV=2			
54			31.5	FSPLLYCQSQELPNGSVK	+298.15 (C7)
55	splP11711ICP2A1_RAT	Cytochrome P450 2A1 GN=Cyp2a1 PE=1 SV=2			
55			30.8	MLQGTCGAPIDPTIYLSK	+298.15 (C6)
56	splQ5M875IDHB13_RAT	17-beta-hydroxysteroid dehydrogenase 13 GN=Hsd17b13 PE=2 SV=1			
56			30.1	NSGHIVTVASVCGHR	+298.15 (C12)
57	splP24090IFETUA_RAT	Alpha-2-HS-glycoprotein GN=AhsG PE=1 SV=2			

57			30.0	VGQPGDAGAAGPVAPLCPGR	+298.15 (C17)
58	splP09606 GLNA_RAT	Glutamine synthetase GN=Glul PE=1 SV=3			
58			29.1	CIEEAIDK	+298.15 (C1)
59	splQ62730 DHB2_RAT	Estradiol 17-beta-dehydrogenase 2 GN=Hsd17b2 PE=2 SV=1			
59			29.1	DIQHAICAK	+298.15 (C7)
60	splP62914 RL11_RAT	60S ribosomal protein L11 GN=Rpl11 PE=1 SV=2			
60			29.0	TGCIGAK	+298.15 (C3)
61	splP32089 TXTP_RAT	Tricarboxylate transport protein, mitochondrial GN=Slc25a1 PE=1 SV=1			
61			28.9	NLDCGVQILK	+298.15 (C5)
62	splQ7TQM4 SOAT2_RAT	Sterol O-acyltransferase 2 GN=Soat2 PE=2 SV=1			
62			28.7	QGEEQENGACGEGNTR	+298.15 (C10)
63	splQ2V057 PROD2_RAT	Probable proline dehydrogenase 2 GN=Prodh2 PE=2 SV=1			
63			28.4	EDCTQPDYEATSR	+298.15 (C3)
64	splP04797 G3P_RAT	Glyceraldehyde-3-phosphate dehydrogenase GN=Gapdh PE=1 SV=3			
64			27.2	IVSNASCTTNCLAPLAK	+298.15 (C7); +57.02 (C11)
65	splP00176 CP2B1_RAT	Cytochrome P450 2B1 GN=Cyp2b1 PE=1 SV=1			
65			26.5	IQEEAQLVEELR	+298.15 (C7)
66	splP62836 RAP1A_RAT	Ras-related protein Rap-1A GN=Rap1a PE=1 SV=1			
66			26.5	CDLEDER	+298.15 (C1)
67	splP35435 ATPG_RAT	ATP synthase subunit gamma, mitochondrial GN=Atp5c1 PE=1 SV=2			
67			26.3	GLCGAIHSSVAK	+298.15 (C3)
68	splP11442 CLH_RAT	Clathrin heavy chain 1 GN=Cltc PE=1 SV=3			
68			25.8	IHEGCEEPATHNALAK	+298.15 (C5)
69	trlG3V6I4 G3V6I4_RAT	RCG20363, isoform CRA_a GN=rCG_20363 PE=4 SV=1			
69			25.7	LCEPSEQALCGK	+57.02 (C2); +298.15 (C10)
70	splQ562C4 MET7B_RAT	Methyltransferase-like protein 7B GN=Mettl7b PE=1 SV=1			
70			25.4	VTCVDPNPNFEK	+298.15 (C3)

# Spectra with Mascot score below 30 were manually validated

\* +298.15 is carbonylation by acrolein *via* 1,4-Michael addition on Cys (C), Lys (K) and His (H); +15.99 is oxidation at methionine (M); +57.02 is carboamidomethylation by IAA at Cys (C)

Table 1.6. Identified modified peptides and corresponding proteins at 200  $\mu$ M acrolein

	Sequence Id	Name	Score <sup>#</sup>	Peptide Sequence	Modification*
1	splP07756 CPSM_RAT	Carbamoyl-phosphate synthase [ammonia], mitochondrial GN=Cps1 PE=1 SV=1			
1			32.0	MCHPSVDGFTPR	+15.99 (M1); +298.15 (C2)
1			32.2	VSQEHPVVLTK	+298.15 (H5)
1			34.8	MCHPSVDGFTPR	+298.15 (H3)
1			35.6	AQTAHIVLEDGTK	+298.15 (H5)
1			40.7	VVAVDCGIK	+298.15 (C6)
1			43.9	TVVVNCNPETVSTDFDECDK	+57.02 (C6); +298.15 (K20)
1			48.9	CLGLTEAQTR	+298.15 (C1)
1			68.5	TSACFEPGLDYMVTK	+298.15 (C4)
1			91.8	SAYALGGLGSGICPNK	+298.15 (C13)
2	splQ62812 MYH9_RAT	Myosin-9 GN=Myh9 PE=1 SV=3			
2			27.8	KMEDGVGCLETAEEAK	+298.15 (C8)
2			30.1	KMEDGVGCLETAEEAK	+298.15 (K1); +57.02 (C8)
2			30.8	HEDELLAK	+298.15 (H1)
2			31.0	KQELEEICHDLER	+298.15 (C8)
2			37.1	CQYLQAEK	+298.15 (C1)
2			38.5	LEEDQIIMEDQNCK	+15.99 (M8); +298.15 (K14)
2			52.1	LEEDQIIMEDQNCK	+298.15 (K14)
2			53.0	NKHEAMITDLEER	+298.15 (H3)
2			66.1	MEDGVGCLETAEEAK	+298.15 (C7)
3	splP05179 CP2C7_RAT	Cytochrome P450 2C7 GN=Cyp2c7 PE=1 SV=2			
3			30.9	FINFVPTNLPHAVTCDIK	+298.15 (C15)
3			40.4	ACVGEGLAR	+298.15 (C2)
3			47.8	IEEHQESLDVTNPR	+298.15 (H4)
3			48.5	KIEEHQESLDVTNPR	+298.15 (K1)
3			69.1	VQEEAQCLVEELRK	+298.15 (C7)
3			81.2	VQEEAQCLVEELR	+298.15 (C7)
4	splP18163 ACSL1_RAT	Long-chain-fatty-acid--CoA ligase 1 GN=Acs11 PE=1 SV=1			
4			28.1	LGKNAGLKPFEQVK	+298.15 (K3)
4			30.7	ALKPPCDLSMQSVEVTGTTEGVR	+298.15 (C6)
4			37.0	CGVEIIGLK	+298.15 (C1)
4			40.8	GFQGSFEELCR	+298.15 (C10)
4			47.6	TKPKPPEPEDLAHCFTSGTTGNPK	+298.15 (C15)
5	splP05182 CP2E1_RAT	Cytochrome P450 2E1 GN=Cyp2e1 PE=1 SV=4			
5			43.4	AKEHLQSLDINCAR	+298.15 (C12)
5			52.3	LCVIPR	+298.15 (C2)
5			60.5	EHLQSLDINCAR	+298.15 (C10)
5			63.1	DVTDCLLIEMEK	+298.15 (C5); +15.99 (M10)
5			66.6	DVTDCLLIEMEK	+298.15 (C5)
6	splP08683 CP2CB_RAT	Cytochrome P450 2C11 GN=Cyp2c11 PE=1 SV=1			
6			38.8	VKEHQESLDKDNPR	+298.15 (H4)
6			42.3	ICAGEALAR	+298.15 (C2)
6			53.7	DFIDCFLNK	+298.15 (C5)
6			62.2	IQEEAQCLVEELRK	+298.15 (C7)
6			84.7	IQEEAQCLVEELR	+298.15 (C7)
7	splO88813 ACSL5_RAT	Long-chain-fatty-acid--CoA ligase 5 GN=Acs15 PE=2 SV=1			
7			30.5	CGIEMLSLHDAENLKG	+298.15 (C1)
7			45.5	KPMPPNPEDLSVICFTSGTTGDPK	+298.15 (C14)

7			67.8	GSFBEELCQNQCVK	+298.15 (C7); +57.02 (C11)
7			72.9	GSFBEELCQNQCVK	+57.02 (C7); +298.15 (C11)
8	splP06757IADH1_RAT	Alcohol dehydrogenase 1 GN=Adh1 PE=1 SV=3			
8			37.0	MVATGVCR	+298.15 (C7)
8			41.3	VIPLFSPQCGK	+298.15 (C9)
9	splP05178ICP2C6_RAT	Cytochrome P450 2C6 GN=Cyp2c6 PE=2 SV=2			
9			25.8	CLVEELRK	+298.15 (C1)
9			27.1	IKEHQESLDVTNPR	+298.15 (H4)
9			31.1	FIDLIPNLPHAVTCDIK	+298.15 (C15)
9			31.4	CLVEELR	+298.15 (C1)
9			34.8	MCAGEGLAR	+298.15 (C2)
9			27.3	CPEVTAK	+298.15 (C1)
10	trIQ6AY58IQ6AY58_RAT	B-cell receptor-associated protein 31 GN=Bcap31 PE=2 SV=1			
10			28.6	LEKAENEALAMQK	+298.15 (K3)
10			34.3	GTAEDGGKLDVGSPEMK	+298.15 (K8)
10			38.9	KGTAEDGGK	+298.15 (K1)
10			48.4	LKDELASTK	+298.15 (K2)
10			65.0	KQAESASEAAK	+298.15 (K1)
11	splQ9EQ76FMO3_RAT	Dimethylaniline monooxygenase [N-oxide-forming] 3 GN=Fmo3 PE=1 SV=1			
11			33.8	GTCILPSVNDMMDDIDEK	+298.15 (C3); +15.99 (M11)
11			34.0	CPDFSTTGK	+298.15 (C1)
11			39.8	ILCGTVSIKPNVK	+298.15 (C3)
11			43.8	GTCILPSVNDMMDDIDEK	+298.15 (C3); +15.99 (M12)
11			50.4	GTCILPSVNDMMDDIDEK	+298.15 (C3)
12	splQ9ES38IS27A5_RAT	Bile acyl-CoA synthetase GN=Slc27a5 PE=1 SV=1			
12			42.8	YLCNVPGQPEDK	+298.15 (C3)
12			43.3	SLMPDVYQAVCEGTWK	+15.99 (M3); +298.15 (C11)
12			51.5	LKEATIQEDK	+298.15 (K2)
13	splP11507-2IAT2A2_RAT	Isoform SERCA2A of Sarcoplasmic/endoplasmic reticulum calcium ATPase 2 GN=Atp2a2			
13			38.9	ANACNSVIK	+298.15 (C4)
13			74.8	TGTLTTNQMSVCR	+298.15 (C12)
13			76.0	VGEATETALTCLVEK	+298.15 (C11)
13			28.9	DACLNAR	+298.15 (C3)
14	trIF1M5X1IF1M5X1_RAT	Ribosome-binding protein 1 (Fragment) GN=Rrbp1 PE=4 SV=1			
14			30.9	KGEGAQNQGK	+298.15 (K1)
14			37.3	KGEGAQNQAK	+298.15 (K1)
14			41.6	EAEETQNSLQAECQYR	+298.15 (C13)
14			44.8	LTAEFEEAQSTACR	+298.15 (C13)
15	splP07153IRPN1_RAT	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1 GN=Rpn1 PE=2 SV=1			
15			31.5	HFDETVNR	+298.15 (H1)
15			41.7	VACITEQVLTLVNK	+298.15 (C3)
15			50.5	LKTEGSDLCDR	+298.15 (C9)
15			55.0	TEGSDLCDR	+298.15 (C7)
16	splP24470ICP2CN_RAT	Cytochrome P450 2C23 GN=Cyp2c23 PE=2 SV=2			
16			31.4	ACVGESLAR	+298.15 (C2)
16			32.7	NGCFK	+298.15 (C3)
16			42.8	DYIDCFLSK	+298.15 (C5)
16			44.3	CLVEELQK	+298.15 (C1)
16			46.8	LCLVPR	+298.15 (C2)
17	splQ64428IECHA_RAT	Trifunctional enzyme subunit alpha, mitochondrial GN=Hadha PE=1 SV=2			
17			28.6	CLAPMMSEVIR	+298.15 (C1)
17			35.7	YESAYGTQFTPCQLLR	+298.15 (C12)

17			77.9	ALMGLYNGQVLCK	+298.15 (C12)
18	splP04797IG3P_RAT	Glyceraldehyde-3-phosphate dehydrogenase GN=Gapdh PE=1 SV=3			
18			31.9	AG AHLK	+298.15 (H4)
18			37.4	VPTPNVSVVDLTCR	+298.15 (C13)
18			58.8	IVSNASCTTNCLAPLAK	+298.15 (C7); +57.02 (C11)
18			63.4	IVSNASCTTNCLAPLAK	+57.02 (C7); +298.15 (C11)
19	splP05183ICP3A2_RAT	Cytochrome P450 3A2 GN=Cyp3a2 PE=1 SV=2			
19			37.3	FDMECYK	+298.15 (C5)
19			59.7	VLQNFSPQCK	+298.15 (C10)
20	splQ9Z270IVAPA_RAT	Vesicle-associated membrane protein-associated protein A GN=Vapa PE=1 SV=3			
20			43.9	CVFEMPENENDK	+298.15 (C1)
21	splQ5M875IDHB13_RAT	17-beta-hydroxysteroid dehydrogenase 13 GN=Hsd17b13 PE=2 SV=1			
21			35.2	HGVEETAACKR	+298.15 (K9); +57.02 (C10)
21			36.6	NSGHIVTVASVCGHR	+298.15 (C12)
21			45.7	HGVEETA AK	+298.15 (H1)
21			60.0	TSCLCPVFNVTGFTK	+57.02 (C3); +298.15 (C5)
22	splP97521IMCAT_RAT	Mitochondrial carnitine/acylcarnitine carrier protein GN=Slc25a20 PE=1 SV=1			
22			39.5	YSGTLDCAK	+298.15 (C7)
22			61.2	CLLQIQASSGK	+298.15 (C1)
23	splP088011MGST1_RAT	Microsomal glutathione S-transferase 1 GN=Mgst1 PE=1 SV=3			
23			41.2	LTNKVFANPEDCAGFGK	+298.15 (K4); +57.02 (C12)
23			77.0	VFANPEDCAGFGK	+298.15 (C8)
24	splP69897ITBB5_RAT	Tubulin beta-5 chain GN=Tubb5 PE=1 SV=1			
24			30.8	GHYTEGAEVDSVLDVVR	+298.15 (H2)
24			30.9	EIVHIQAGQCGNQIGAK	+298.15 (C10)
24			34.4	TAVCDIPPR	+298.15 (C4)
24			35.1	EIVHIQAGQCGNQIGAK	+298.15 (H4); +57.02 (C10)
24			37.6	NMMAACDPR	+298.15 (C6)
25	splO88618IFTCD_RAT	Formimidoyltransferase-cyclodeaminase GN=Ftd PE=1 SV=4			
25			31.8	TVYTFVQPECVVEGALSAAR	+298.15 (C11)
25			46.9	TCALQEGLR	+298.15 (C2)
25			67.2	AFAACLGAIK	+298.15 (C5)
26	splP29147IBDH_RAT	D-beta-hydroxybutyrate dehydrogenase, mitochondrial GN=Bdh1 PE=1 SV=2			
26			33.7	YEMHPLGVK	+298.15 (H4)
26			33.7	FGVEAFSDCLR	+298.15 (C9)
26			77.5	TIQLNVCNSEEVEK	+298.15 (C7)
27	splP13107ICP2B3_RAT	Cytochrome P450 2B3 GN=Cyp2b3 PE=2 SV=1			
27			26.7	CLVEELK	+298.15 (C1)
27			26.9	ICFVAR	+298.15 (C2)
27			39.1	CEAFMPFSIGK	+298.15 (C1)
27			41.5	FSDVSPMGLPCR	+298.15 (C11)
28	splO54753IH17B6_RAT	17-beta-hydroxysteroid dehydrogenase type 6 GN=Hsd17b6 PE=1 SV=2			
28			28.1	CSTNLSLVTDCEHALTSK	+298.15 (C1); +57.02 (C11)
28			30.3	ESYQQFFDDFCNTTR	+298.15 (C12)
28			36.5	CSTNLSLVTDCEHALTSK	+57.02 (C1); +298.15 (C11)
28			38.8	GAEELKSK	+298.15 (K6)
29	splP55006IRDH7_RAT	Retinol dehydrogenase 7 GN=Rdh7 PE=2 SV=1			
29			64.3	MSLLGGGYCISK	+298.15 (C9)
29			67.0	MSLLGGGYCISK	+15.99 (M1); +298.15 (C9)
30	splQ6RUV5IRAC1_RAT	Ras-related C3 botulinum toxin substrate 1 GN=Rac1 PE=1 SV=1			
30			31.3	CVVVGDGAVGK	+298.15 (C1)
30			36.3	HHCPNTPILVGTK	+298.15 (C3)

30			58.6	YLECSALTQR	+298.15 (C4)
31	splQ6AYS8IDHB11_RAT	Estradiol 17-beta-dehydrogenase 11 GN=Hsd17b11 PE=2 SV=1			
31			32.7	KLGAQVHPFVVDCSQR	+298.15 (C13)
31			39.7	LGAQVHPFVVDCSQR	+298.15 (C12)
31			50.7	ALTDELAALGCTGVR	+298.15 (C11)
32	splP04785IPDIA1_RAT	Protein disulfide-isomerase GN=P4hb PE=1 SV=2			
32			28.5	KEECPAVR	+298.15 (C4)
32			34.5	EADDIVNWLKK	+298.15 (K11)
33	splP97524IS27A2_RAT	Very long-chain acyl-CoA synthetase GN=Slc27a2 PE=1 SV=1			
33			33.0	GEVGLLICK	+298.15 (K9)
33			39.1	SLLHCFQCCGAK	+57.02 (C5); +57.02 (C9); +298.15 (C8)
33			42.6	GEVGLLICK	+298.15 (C8)
34	splP12939ICP2DA_RAT	Cytochrome P450 2D10 GN=Cyp2d10 PE=2 SV=1			
34			26.8	CPEMTDQAHMPYTNNAVIHEVQR	+298.15 (C1)
34			61.5	ITSCDIEVQDFVPIK	+298.15 (C4)
35	trIG3V6I4IG3V6I4_RAT	RCG20363, isoform CRA_a GN=rCG_20363 PE=4 SV=1			
35			33.8	DLLLPITPPATNPLLQCR	+298.15 (C17)
35			52.6	CLLTVDPDPTGIMDK	+298.15 (C1)
36	splP23358IRL12_RAT	60S ribosomal protein L12 GN=Rpl12 PE=2 SV=1			
36			35.5	EILGTAQSVGCNVDGR	+298.15 (C11)
36			75.6	CTGGEVGATSALAPK	+298.15 (C1)
37	splP13437ITHIM_RAT	3-ketoacyl-CoA thiolase, mitochondrial GN=Acaa2 PE=1 SV=1			
37			33.4	VVGYFVSGCDPAIMGIGPVPAITGALK	+298.15 (C9); +15.99 (M14)
37			33.6	LCGSGFQSIVSGCQEICK	+57.02 (C2); +57.02 (C13); +298.15 (C17)
37			42.9	LCGSGFQSIVSGCQEICK	+57.02 (C2); +57.02 (C17); +298.15 (C13)
38	splQ562C4MET7B_RAT	Methyltransferase-like protein 7B GN=Mettl7b PE=1 SV=1			
38			32.6	HIGDGCHLTR	+298.15 (H1); +57.02 (C6)
38			36.8	VTCVDPNPNFEK	+298.15 (C3)
38			37.2	HIGDGCHLTR	+298.15 (C6)
39	splP33274ICP4F1_RAT	Cytochrome P450 4F1 GN=Cyp4f1 PE=2 SV=1			
39			27.1	ACNLVHEFTDAVIR	+298.15 (C2)
39			49.1	CCTQDILLPDGR	+298.15 (C1); +57.02 (C2)
40	splO09171IBHMT1_RAT	Betaine--homocysteine S-methyltransferase 1 GN=Bhmt PE=1 SV=1			
40			50.5	QVADEGDALVAGGVSTPSYLSCK	+298.15 (K24)
40			52.5	VNEAACDIAR	+298.15 (C6)
41	splP32089ITXTP_RAT	Tricarboxylate transport protein, mitochondrial GN=Slc25a1 PE=1 SV=1			
41			38.9	GIGDCVR	+298.15 (C5)
41			63.8	NTLDCGVQILK	+298.15 (C5)
42	trIB0BN81IB0BN81_RAT	Ribosomal protein S5, isoform CRA_b GN=Rps5 PE=2 SV=1			
42			42.3	VNQAIWLLCTGAR	+298.15 (C9)
42			57.3	TIAECLADELINAACK	+298.15 (C5)
43	trIG3V9Y7IG3V9Y7_RAT	RCG24095, isoform CRA_b GN=Faf2 PE=4 SV=1			
43			38.4	KLECLPPEPSPDDPDVSK	+298.15 (C4)
43			60.3	LLQFQDLTGIDSMEQCR	+298.15 (C16)
44	splA2RR9ICP4V2_RAT	Cytochrome P450 4V2 GN=Cyp4v2 PE=2 SV=1			
44			43.9	AEQDCIGAGR	+298.15 (C5)
44			52.4	SLSEDCEVAGYK	+298.15 (C6)
45	trIF1LW74IF1LW74_RAT	IQ motif-containing GTPase-activating protein 2 (Fragment) GN=LOC100360623 PE=4 SV=1			
45			42.4	NSCISEEER	+298.15 (C3)
45			53.4	NPNAVLTCVDDSLSQEYQK	+298.15 (C8)



46	spIP005071AATM_RAT	Aspartate aminotransferase, mitochondrial GN=Got2 PE=1 SV=2			
46			32.9	NLDKEYLPIGGLADFCK	+298.15 (K17)
46			35.4	VGAFTVVCK	+298.15 (C8)
47	spIQ3B7U9IFKBP8_RAT	Peptidyl-prolyl cis-trans isomerase FKBP8 GN=Fkbp8 PE=2 SV=1			
47			29.7	SCSQVLEHQPDNIK	+298.15 (C2)
47			30.2	CLNNLAASQLK	+298.15 (C1)
47			33.1	VDMTCEEEEEELLQLK	+298.15 (C5)
48	spIP11510ICP2CC_RAT	Cytochrome P450 2C12, female-specific GN=Cyp2c12 PE=2 SV=1			
48			39.3	IKEHEESLDVSNPR	+298.15 (H4)
49	spIP11711ICP2A1_RAT	Cytochrome P450 2A1 GN=Cyp2a1 PE=1 SV=2			
49			26.4	MLQGTGCGAPIDPTIYLSK	+298.15 (C6)
49			29.1	MLQGTGCGAPIDPTIYLSK	+15.99 (M1); +298.15 (C6)
49			34.0	FCLGDGLAK	+298.15 (C2)
50	spIP62755IRS6_RAT	40S ribosomal protein S6 GN=Rps6 PE=1 SV=1			
50			40.1	LNISFPATGCQK	+298.15 (C10)
50			49.2	LNISFPATGCQK	+298.15 (K12)
51	spIQ5XIU9IPGRC2_RAT	Membrane-associated progesterone receptor component 2 GN=Pgrmc2 PE=1 SV=1			
51			34.5	GLCSGPGAGEESPAATLPR	+298.15 (C3)
51			54.1	GLATFCLDK	+298.15 (C6)
52	spIP08010IGSTM2_RAT	Glutathione S-transferase Mu 2 GN=Gstm2 PE=1 SV=2			
52			26.5	KHNLCGETEER	+298.15 (C5)
52			33.8	CLDAFPNLK	+298.15 (C1)
53	spIQ9JJ22IERAP1_RAT	Endoplasmic reticulum aminopeptidase 1 GN=Erap1 PE=2 SV=2			
53			87.0	CVQQTITIEENIR	+298.15 (C1)
54	spIP68370TBA1A_RAT	Tubulin alpha-1A chain GN=Tuba1a PE=1 SV=1			
54			56.9	TIQFVDWCPTGFK	+298.15 (C8)
55	spIP07896IECHP_RAT	Peroxisomal bifunctional enzyme GN=Ehhadh PE=1 SV=2			
55			60.7	LCNPPVNAVSPVIR	+298.15 (C2)
56	spIP10867IGGLO_RAT	L-gulonolactone oxidase GN=Gulo PE=1 SV=3			
56			26.2	IFTYECR	+298.15 (C6)
57	spIQ5XIF6ITBA4A_RAT	Tubulin alpha-4A chain GN=Tuba4a PE=2 SV=1			
57			52.1	SIQFVDWCPTGFK	+298.15 (C8)
58	spIQ5XIB4UFSP2_RAT	Ufm1-specific protease 2 GN=Ufsp2 PE=2 SV=1			
58			81.6	ICNSSVYLWPNSDANTGELTDSSACK	+298.15 (C2); +57.02 (C25)
59	spIQ02769IFDFT_RAT	Squalene synthase GN=Fdft1 PE=2 SV=1			
59			50.7	TQSLPNCQLISR	+298.15 (C7)
60	spIQ8CFN2-2ICDC42_RAT	Isoform 2 of Cell division control protein 42 homolog GN=Cdc42			
60			31.3	CVVVGDGAVGK	+298.15 (C1)
60			47.0	YVECSALTQK	+298.15 (C4)
61	spIP21531IRL3_RAT	60S ribosomal protein L3 GN=Rpl3 PE=1 SV=3			
61			25.9	YCQVIR	+298.15 (C2)
61			52.4	TVFAEHISDECK	+298.15 (C11)
62	spIP62703IRS4X_RAT	40S ribosomal protein S4, X isoform GN=Rps4x PE=2 SV=2			
62			76.9	FDTGNLCMVTGGANLGR	+298.15 (C7)
63	trIF1MA49IF1MA49_RAT	Methyltransferase-like 7A-like GN=LOC100362962 PE=4 SV=1			
63			75.2	LSLLEVGCGTGANFK	+298.15 (C8)
64	spIO88994IMOSC2_RAT	MOSC domain-containing protein 2, mitochondrial GN=Marc2 PE=2 SV=1			
64			32.1	CVLTTVDPDTGIIDR	+298.15 (C1)
64			43.0	LCDPSVK	+298.15 (C2)
65	spIP04799ICP1A2_RAT	Cytochrome P450 1A2 GN=Cyp1a2 PE=1 SV=2			
65			48.7	TCEHVQAWPR	+298.15 (C2)

66	splQ66HD0IENPL_RAT	Endoplasmic reticulum chaperone protein			
66			44.5	LTESPCALVASQYGWSGNMER	+298.15 (C6)
67	splP27952IRS2_RAT	40S ribosomal protein S2			
67			30.2	CGSVLVR	+298.15 (C1)
67			41.3	GCTATLGNFAK	+298.15 (C2)
68	splQ09073IADT2_RAT	ADP/ATP translocase 2			
68			30.6	GLGDCLVK	+298.15 (C5)
68			40.1	GIIDCVVR	+298.15 (C5)
69	splQ5RJR8ILRC59_RAT	Leucine-rich repeat-containing protein 59			
69			45.4	ATVLDLSCNK	+298.15 (C8)
70	splP16232-2IDHI1_RAT	Isoform 11-HSD1B of Corticosteroid 11-beta-dehydrogenase isozyme 1			
70			40.6	EECALEIHK	+298.15 (C3)
71	splP00173-2ICYB5_RAT	Isoform Short of Cytochrome b5			
71			29.5	STWVILHHK	+298.15 (H7)
72	splQ75Q39ITOM70_RAT	Mitochondrial import receptor subunit TOM70			
72			29.5	CIDLEPDNATTVVHK	+298.15 (C1)
72			39.3	CAEGYALYAQALTDQQQFGK	+298.15 (C1)
73	trID3ZEA0ID3ZEA0_RAT	Fibronectin type III domain-containing 3a			
73			25.8	GCTQVDQEIEEK	+298.15 (C2)
73			42.8	GCTQVDQEIEEKDEETK	+298.15 (C2)
74	splQ8K4C0FMO5_RAT	Dimethylaniline monooxygenase [N-oxide-forming] 5			
74			40.0	CCLEEGLEPVCFER	+298.15 (C1); +57.02 (C2); +57.02 (C11)
75	splP97612IFAAH1_RAT	Fatty-acid amide hydrolase 1			
75			65.7	GTNCVTSYLTDCETQLSQAPR	+57.02 (C4); +298.15 (C12)
76	splP20816ICP4A2_RAT	Cytochrome P450 4A2			
76			64.6	ACQIAHEHTDGVK	+298.15 (C2)
77	splP62250IRS16_RAT	40S ribosomal protein S16			
77			33.9	TATAVAHCK	+298.15 (C8)
78	trIB2RZD1IB2RZD1_RAT	Sec61 beta subunit			
78			33.5	NASCGTR	+298.15 (C4)
79	splQ9Z2L0IVDAC1_RAT	Voltage-dependent anion-selective channel protein 1			
79			62.0	YQVDPDACFSAK	+298.15 (C8)
80	trIO89035IO89035_RAT	Mitochondrial dicarboxylate carrier			
80			61.9	GEYQGVFHCAVETAK	+298.15 (C9)
81	splP19511IAT5F1_RAT	ATP synthase subunit b, mitochondrial			
81			61.4	HVIQSISAQQEK	+298.15 (H1)
82	splQ5XI60IREEP6_RAT	Receptor expression-enhancing protein 6			
82			61.4	HHVALDSAASQLSGR	+298.15 (H2)
83	splP50878IRL4_RAT	60S ribosomal protein L4			
83			28.0	RGPCIHYNEDNGIHK	+298.15 (C4)
83			31.2	SGQGAFGNMCR	+298.15 (C10)
84	splQ7TQM4ISOAT2_RAT	Sterol O-acyltransferase 2			
84			28.7	QGEEQENGACGEGNTR	+298.15 (C10)
84			30.1	TQCLEQAQR	+298.15 (C3)
85	splQ0ZHH6-2IATLA3_RAT	Isoform 2 of Atlastin-3			
85			58.6	YQEELEEEITELYENFCK	+298.15 (K18)
86	splQ66HF1INDUS1_RAT	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial			
86			57.7	AVTEGAQAVEEPSIC	+298.15 (C15)
87	splP62718IRL18A_RAT	60S ribosomal protein L18a			
87			56.9	DLTTAGAVTQCYR	+298.15 (C11)
88	splQ63357IMYO1D_RAT	Unconventional myosin-IId			

88			28.3	SNCVLEAFGNAK	+298.15 (C3)
88			28.4	TCASDK	+298.15 (C2)
89	spIP31399 ATP5H_RAT	ATP synthase subunit d, mitochondrial GN=Atp5h PE=1 SV=3			
89			56.6	NCAQFVTGSQAR	+298.15 (C2)
90	spIP54313 GBB2_RAT	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2 GN=Gnb2 PE=1 SV=4			
90			56.3	TFVSGACDASIK	+298.15 (C7)
91	spIB5DEH2 ERLN2_RAT	Erlin-2 GN=Erlin2 PE=1 SV=1			
91			54.5	ADAECYTALK	+298.15 (C5)
92	spIQ9Z2Z8 DHCR7_RAT	7-dehydrocholesterol reductase GN=Dhcr7 PE=2 SV=1			
92			54.3	AIECSYTSADGLK	+298.15 (C4)
93	spIQ63081 PDIA6_RAT	Protein disulfide-isomerase A6 GN=Pdia6 PE=1 SV=2			
93			25.8	KDVVELTDDTFDK	+298.15 (K1)
93			27.7	NKPEDYQGR	+298.15 (K2)
94	trIB5DF48 B5DF48_RAT	Radical S-adenosyl methionine and flavodoxin domains 1 (Predicted), isoform CRA_a GN=Tyw1 PE=2 SV=1			
94			53.1	GEGDCNAVQSK	+298.15 (C5)
95	spIQ9R1Z0-2 VDAC3_RAT	Isoform 2 of Voltage-dependent anion-selective channel protein 3 GN=Vdac3			
95			53.0	LCQNNFALGYK	+298.15 (C2)
96	spIP00884 ALDOB_RAT	Fructose-bisphosphate aldolase B GN=Aldob PE=1 SV=2			
96			52.6	ISDQCPSSLAIQENANALAR	+298.15 (C5)
97	spIQ64581 CP3A1_RAT	Cytochrome P450 3A18 GN=Cyp3a18 PE=2 SV=1			
97			26.2	LAVIGVLQNFNIQPCEK	+298.15 (K17)
98	spIP02706 ASGR1_RAT	Asialoglycoprotein receptor 1 GN=Asgr1 PE=1 SV=2			
98			49.8	SLSCQMAALR	+298.15 (C4)
99	spIP16391 HA12_RAT	RT1 class I histocompatibility antigen, AA alpha chain PE=1 SV=2			
99			47.4	DSSQSSDVSLPDCK	+298.15 (C13)
100	trID3ZIM4 D3ZIM4_RAT	Aldehyde dehydrogenase GN=Aldh3a2 PE=3 SV=1			
100			46.6	IKSLLEGQK	+298.15 (K2)
101	spIP00388 NCPR_RAT	NADPH--cytochrome P450 reductase GN=Por PE=1 SV=3			
101			45.4	EVGETLLYYGCR	+298.15 (C11)
102	spIQ9JHZ9 S38A3_RAT	Sodium-coupled neutral amino acid transporter 3 GN=Slc38a3 PE=2 SV=1			
102			45.2	AEDAQHCGEGK	+298.15 (C7)
103	spIQ9JIY6 CML06_RAT	Probable N-acetyltransferase CML6 GN=Cml6 PE=2 SV=1			
103			44.2	CLHTDMADITK	+298.15 (C1)
104	spIQ5I0H9 PDIA5_RAT	Protein disulfide-isomerase A5 GN=Pdia5 PE=2 SV=1			
104			43.8	DKNQDLCQQESVK	+298.15 (K2); +57.02 (C7)
105	spIP80432 COX7C_RAT	Cytochrome c oxidase subunit 7C, mitochondrial GN=Cox7c PE=1 SV=2			
105			43.0	SHYEEGPGK	+298.15 (H2)
106	spIP04644 IRS17_RAT	40S ribosomal protein S17 GN=Rps17 PE=2 SV=3			
106			42.9	VCEEIAIIPSK	+298.15 (C2)
107	spIQ2V057 PROD2_RAT	Probable proline dehydrogenase 2 GN=Prodh2 PE=2 SV=1			
107			42.7	EDCTQPDYEATSR	+298.15 (C3)
108	spIQ66X93 SND1_RAT	Staphylococcal nuclease domain-containing protein 1 GN=Snd1 PE=2 SV=1			
108			42.6	LSECEEQAK	+298.15 (C4)
109	trIA1L1K2 A1L1K2_RAT	Igtp protein GN=Igtp PE=2 SV=1			
109			42.1	YRDPLETSLSQVCDK	+298.15 (C12)
110	trIG3V9N7 G3V9N7_RAT	RCG27172, isoform CRA_a GN=Pacsin3 PE=4 SV=1			
110			42.0	YMEDMEQAFESCQAAER	+298.15 (C12)
111	trIQ7TP42 Q7TP42_RAT	Ab2-292 GN=Sec62 PE=2 SV=1			
111			41.7	AVECLLDSK	+298.15 (C4)
112	spIP35435 ATPG_RAT	ATP synthase subunit gamma, mitochondrial GN=Atp5c1 PE=1 SV=2			
112			41.1	GLCGAIHSSVAK	+298.15 (C3)

113	splP62832 RL23_RAT	60S ribosomal protein L23 GN=Rpl23 PE=2 SV=1			
113			41.1	ISLGLPVGAVINCADNTGAK	+298.15 (C13)
114	splP27653 C1TC_RAT	C-1-tetrahydrofolate synthase, cytoplasmic GN=Mthfd1 PE=1 SV=3			
114			41.1	CTHWAEGGQALALAAQAVQR	+298.15 (C1)
115	splQ4V8C2 ZW10_RAT	Centromere/kinetochore protein zw10 homolog GN=Zw10 PE=2 SV=3			
115			40.9	GEVCNMISK	+298.15 (C4)
116	splP29995 ITPR2_RAT	Inositol 1,4,5-trisphosphate receptor type 2 GN=Itp2 PE=1 SV=1			
116			40.9	TGISMSDIQCLLDK	+298.15 (C10)
117	splQ53UA7 TAOK3_RAT	Serine/threonine-protein kinase TAO3 GN=Taok3 PE=2 SV=1			
117			40.9	LDEAQEAECQALR	+298.15 (C9)
118	splP17077 RL9_RAT	60S ribosomal protein L9 GN=Rpl9 PE=1 SV=1			
118			40.7	TGVACSVSQAQK	+298.15 (C5)
119	splQ63355 MYO1C_RAT	Unconventional myosin-Ic GN=Myo1c PE=2 SV=2			
119			40.7	ETMCSSTNPMAQCQCFDK	+298.15 (C4); +57.02 (C14)
120	triQ6UPE0 Q6UPE0_RAT	Choline dehydrogenase GN=Chdh PE=2 SV=1			
120			40.6	GCPALGDENVVYKQTLDTQR	+298.15 (C2)
121	splP10860 DHE3_RAT	Glutamate dehydrogenase 1, mitochondrial GN=Glud1 PE=1 SV=2			
121			40.1	CAVVDVPFGGAK	+298.15 (C1)
122	splP97562 ACOX2_RAT	Peroxisomal acyl-coenzyme A oxidase 2 GN=Acox2 PE=1 SV=1			
122			40.0	CSAQTAAEFR	+298.15 (C1)
123	splP17988 IST1A1_RAT	Sulfotransferase 1A1 GN=Sult1a1 PE=1 SV=1			
123			39.8	CPGVPSGLETLEETPAPR	+298.15 (C1)
124	splP23562-2 B3AT_RAT	Isoform Kidney of Band 3 anion transport protein GN=Slc4a1			
124			39.3	LYCAQAEGGSEEPSGILK	+298.15 (C3)
125	splP36365 FMO1_RAT	Dimethylaniline monooxygenase [N-oxide-forming] 1 GN=Fmo1 PE=1 SV=2			
125			38.9	SCDLGGLWR	+298.15 (C2)
126	splQ811A2 BST2_RAT	Bone marrow stromal antigen 2 GN=Bst2 PE=1 SV=1			
126			38.3	KVSQTQEQQAR	+298.15 (K1)
127	splP84100 RL19_RAT	60S ribosomal protein L19 GN=Rpl19 PE=1 SV=1			
127			38.3	KPVTVHSR	+298.15 (H6)
128	splQ4V8K1 STEAP4_RAT	Metalloreductase STEAP4 GN=Steap4 PE=2 SV=1			
128			38.2	QVFVCGNDSK	+298.15 (C5)
129	splP62909 RS3_RAT	40S ribosomal protein S3 GN=Rps3 PE=2 SV=1			
129			38.0	GCEVVVSGK	+298.15 (C2)
130	splP62912 RL32_RAT	60S ribosomal protein L32 GN=Rpl32 PE=1 SV=2			
130			37.0	SYCAEIAHNVSSK	+298.15 (C3)
131	splQ9Z1X1 ESYT1_RAT	Extended synaptotagmin-1 GN=Esyt1 PE=2 SV=1			
131			36.9	ATYSTNCPVWEEAFR	+298.15 (C7)
132	splP62914 RL11_RAT	60S ribosomal protein L11 GN=Rpl11 PE=1 SV=2			
132			36.8	TGCGIAK	+298.15 (C3)
133	splQ62730 DHB2_RAT	Estradiol 17-beta-dehydrogenase 2 GN=Hsd17b2 PE=2 SV=1			
133			36.8	DIQHAICAK	+298.15 (C7)
134	splP01026 CO3_RAT	Complement C3 GN=C3 PE=1 SV=3			
134			35.6	DSCVGLTVVK	+298.15 (C3)
135	splQ07116 SUOX_RAT	Sulfite oxidase, mitochondrial GN=Suox PE=1 SV=2			
135			35.6	LCDVLAQAGHR	+298.15 (C2)
136	splP28037 AL1L1_RAT	Cytosolic 10-formyltetrahydrofolate dehydrogenase GN=Aldh1l1 PE=1 SV=3			
136			35.5	GENCIAAGR	+298.15 (C4)
137	splP60868 RS20_RAT	40S ribosomal protein S20 GN=Rps20 PE=3 SV=1			
137			35.4	VCADLIR	+298.15 (C2)
138	triD3ZRF5 D3ZRF5_RAT	RCG47744, isoform CRA_c GN=Slc25a22 PE=3 SV=1			
138			35.1	NHGIAGLYK	+298.15 (H2)

139	splP09606 GLNA_RAT	Glutamine synthetase GN=Glul PE=1 SV=3			
139			35.1	CIEEAIDK	+298.15 (C1)
140	splO35132 CP27B_RAT	25-hydroxyvitamin D-1 alpha hydroxylase, mitochondrial GN=Cyp27b1 PE=2 SV=2			
140			34.9	SCIGR	+298.15 (C2)
141	splP11598 PDIA3_RAT	Protein disulfide-isomerase A3 GN=Pdia3 PE=1 SV=2			
141			34.8	YKELGEK	+298.15 (K2)
142	splP08753 GNAI3_RAT	Guanine nucleotide-binding protein G(k) subunit alpha GN=Gnai3 PE=1 SV=3			
142			34.4	IHHEDGYSEDECK	+298.15 (K13)
143	splP84092 AP2M1_RAT	AP-2 complex subunit mu GN=Ap2m1 PE=1 SV=1			
143			34.3	IPPLNTSGVQVICMK	+298.15 (C14)
144	splP62630 EF1A1_RAT	Elongation factor 1-alpha 1 GN=Eef1a1 PE=1 SV=1			
144			33.4	STTTGHLIYK	+298.15 (H6)
145	splP11442 CLH_RAT	Clathrin heavy chain 1 GN=Cltc PE=1 SV=3			
145			33.4	IHEGCEEPATHNALAK	+298.15 (C5)
146	trID3ZUC2 ID3ZUC2_RAT	RCG51996 GN=Mov10 PE=4 SV=1			
146			33.4	YCITK	+298.15 (C2)
147	splQ4KLZ6 DHAK_RAT	Bifunctional ATP-dependent dihydroxyacetone kinase/FAD-AMP lyase (cyclizing) GN=Dak PE=1 SV=1			
147			33.3	AAGDGDGCGSTHSR	+298.15 (C7)
148	splP17178 CP27A_RAT	Sterol 26-hydroxylase, mitochondrial GN=Cyp27a1 PE=1 SV=1			
148			32.9	NTQFVLCHYVVSRR	+298.15 (C7)
149	splP15149 CP2A2_RAT	Cytochrome P450 2A2 GN=Cyp2a2 PE=1 SV=1			
149			32.9	TLQGTGCGAPIDPSIYLSK	+298.15 (C6)
150	splQ64232 TECR_RAT	Trans-2,3-enoyl-CoA reductase GN=Tecr PE=2 SV=1			
150			32.9	LCFLDKVEPQATISEIK	+298.15 (C2)
151	splB0BNG2 TM6S2_RAT	Transmembrane 6 superfamily member 2 GN=Tm6sf2 PE=2 SV=1			
151			32.9	IFNQSPAPSSCTCDVQQEEQK	+57.02 (C11); +298.15 (C13)
152	splQ63507 RL14_RAT	60S ribosomal protein L14 GN=Rpl14 PE=1 SV=3			
152			32.7	CMQLTDFILK	+298.15 (C1)
153	splQ6J1Y9 UBP19_RAT	Ubiquitin carboxyl-terminal hydrolase 19 GN=Usp19 PE=1 SV=1			
153			32.5	ELECAEDPGSAGEAAR	+298.15 (C4)
154	splP17625 GYS2_RAT	Glycogen [starch] synthase, liver GN=Gys2 PE=2 SV=2			
154			32.4	TQVEPCEPANDAVR	+298.15 (C6)
155	splP11915 NLTP_RAT	Non-specific lipid-transfer protein GN=Scp2 PE=1 SV=3			
155			32.2	YGMSACPFAPQLFGSAGK	+298.15 (C6)
156	splQ6UPE1 ETFD_RAT	Electron transfer flavoprotein-ubiquinone oxidoreductase, mitochondrial GN=Etfdh PE=2 SV=1			
156			32.0	AAQIGAHTLSGACLDPAAFK	+298.15 (C13)
157	trIQ4QQS6 Q4QQS6_RAT	Asparagine-linked glycosylation 5 homolog (Yeast, dolichyl-phosphate beta-glucosyltransferase), isoform CRA_a GN=Alg5 PE=2 SV=1			
157			31.9	GLSDLQPWPEQMAIACGSR	+298.15 (C16)
158	splQ4KM77 EI24_RAT	Etoposide-induced protein 2.4 homolog GN=Ei24 PE=2 SV=1			
158			31.7	DSIWGICTISK	+298.15 (C7)
159	splP18445 RL27A_RAT	60S ribosomal protein L27a GN=Rpl27a PE=1 SV=3			
159			31.6	NQSFCTVNLDK	+298.15 (C5)
160	splQ9JJ40 NHRF3_RAT	Na(+)/H(+) exchange regulatory cofactor NHE-RF3 GN=Pdzk1 PE=1 SV=2			
160			30.9	FSPLLYCQSQELPNGSVK	+298.15 (C7)
161	splP22071 3BHS1_RAT	3 beta-hydroxysteroid dehydrogenase/Delta 5->4-isomerase type 1 GN=Hsd3b1 PE=2 SV=3			
161			30.9	TSEWIGTLVEQHR	+298.15 (H12)
162	splQ5M7T2 SPRY7_RAT	SPRY domain-containing protein 7 GN=Spryd7 PE=2 SV=1			
162			30.7	ICGTGGCLASAPLHQNK	+298.15 (C2); +57.02 (C7)
163	trID4A9Y2 ID4A9Y2_RAT	RCG29101 GN=LOC691931 PE=4 SV=1			
163			30.5	GGTGGGECDGEDGAAPAGR	+298.15 (C8)
164	splP15999 ATPA_RAT	ATP synthase subunit alpha, mitochondrial GN=Atp5a1 PE=1 SV=2			

164			29.6	SDGKISEQSDAK	+298.15 (K4)
165	spIP04897IGNAI2_RAT	Guanine nucleotide-binding protein G(i) subunit alpha-2 GN=Gnai2 PE=1 SV=3			
165			28.7	IIHEDGYSEEECR	+298.15 (C12)
166	spIP83732IRL24_RAT	60S ribosomal protein L24 GN=Rpl24 PE=1 SV=1			
166			28.1	CESAFLSK	+298.15 (C1)
167	spIP51871ICP4F6_RAT	Cytochrome P450 4F6 GN=Cyp4f6 PE=2 SV=1			
167			28.0	ACDVVHNFTDAVIR	+298.15 (C2)
168	spIP07687IHYP_RAT	Epoxide hydrolase 1 GN=Ephx1 PE=1 SV=1			
168			27.9	VETSDEEIKDLHQR	+298.15 (H12)
169	spIQ5EB77IRAB18_RAT	Ras-related protein Rab-18 GN=Rab18 PE=2 SV=1			
169			27.3	TCDGVQCAFEELVEK	+57.02 (C2); +298.15 (C7)
170	spIP62836IRAP1A_RAT	Ras-related protein Rap-1A GN=Rap1a PE=1 SV=1			
170			27.0	CDLEDER	+298.15 (C1)
171	spIP19643IAOFB_RAT	Amine oxidase [flavin-containing] B GN=Maob PE=1 SV=3			
171			26.9	TLNHEIYEAK	+298.15 (H4)
172	spIP35280IRAB8A_RAT	Ras-related protein Rab-8A GN=Rab8a PE=2 SV=2			
172			26.4	CDVNDKR	+298.15 (C1)
173	spIQ5FVL2ICX4NB_RAT	Neighbor of COX4 GN=Cox4nb PE=2 SV=1			
173			26.1	FTMDCAAPTIVYEQHENR	+298.15 (C5)

# Spectra with Mascot score below 30 were manually validated

\* +298.15 is carbonylation by acrolein *via* 1,4-Michael addition on Cys (C), Lys (K) and His (H); +15.99 is oxidation at methionine (M); +57.02 is carboamidomethylation by IAA at Cys (C)

## 1.5 REFERENCES

1. Witz, G., *Biological interactions of alpha,beta-unsaturated aldehydes*. Free Radic Biol Med, 1989. **7**(3): p. 333-49.
2. Chung, F.L., R. Young, and S.S. Hecht, *A Study of Chemical Carcinogenesis .61. Formation of Cyclic 1,N2-Propanodeoxyguanosine Adducts in DNA Upon Reaction with Acrolein or Crotonaldehyde*. Cancer Research, 1984. **44**(3): p. 990-995.
3. Kehrer, J.P. and S.S. Biswal, *The molecular effects of acrolein*. Toxicological Sciences, 2000. **57**(1): p. 6-15.
4. Esterbauer, H., H. Zollner, and N. Scholz, *Reaction of glutathione with conjugated carbonyls*. Z Naturforsch C, 1975. **30**(4): p. 466-73.
5. Uchida, K., et al., *Acrolein is a product of lipid peroxidation reaction - Formation of free acrolein and its conjugate with lysine residues in oxidized low density lipoproteins*. Journal of Biological Chemistry, 1998. **273**(26): p. 16058-16066.
6. Furuhashi, A., et al., *N(epsilon)-(3-methylpyridinium)lysine, a major antigenic adduct generated in acrolein-modified protein*. Journal of Biological Chemistry, 2003. **278**(49): p. 48658-65.
7. Kaminskas, L.M., S.M. Pyke, and P.C. Burcham, *Differences in lysine adduction by acrolein and methyl vinyl ketone: implications for cytotoxicity in cultured hepatocytes*. Chemical Research in Toxicology, 2005. **18**(11): p. 1627-33.
8. Burcham, P.C. and S.M. Pyke, *Hydralazine inhibits rapid acrolein-induced protein oligomerization: role of aldehyde scavenging and adduct trapping in cross-link blocking and cytoprotection*. Mol Pharmacol, 2006. **69**(3): p. 1056-65.
9. Cai, J., A. Bhatnagar, and W.M. Pierce, Jr., *Protein modification by acrolein: formation and stability of cysteine adducts*. Chem Res Toxicol, 2009. **22**(4): p. 708-16.
10. Burcham, P.C., et al., *Intermolecular protein cross-linking during acrolein toxicity: efficacy of carbonyl scavengers as inhibitors of heat shock protein-90*

- cross-linking in A549 cells*. Chemical Research in Toxicology, 2007. **20**(11): p. 1629-37.
11. Esterbauer, H., R.J. Schaur, and H. Zollner, *Chemistry and biochemistry of 4-hydroxynonenal, malonaldehyde and related aldehydes*. Free Radic Biol Med, 1991. **11**(1): p. 81-128.
  12. LoPachin, R.M., et al., *Molecular mechanisms of 4-hydroxy-2-nonenal and acrolein toxicity: nucleophilic targets and adduct formation*. Chemical Research in Toxicology, 2009. **22**(9): p. 1499-508.
  13. IARC, *Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans. Dry Cleaning, Some Chlorinated Solvents and Other Industrial Chemicals*. International Agency for Research on Cancer, 1995. **63**.
  14. Abraham, K., et al., *Toxicology and risk assessment of acrolein in food*. Mol Nutr Food Res, 2011. **55**(9): p. 1277-90.
  15. Beauchamp, R.O., Jr., et al., *A critical review of the literature on acrolein toxicity*. Crit Rev Toxicol, 1985. **14**(4): p. 309-80.
  16. Gomes, R., M.E. Meek, and M. Eggleton, *Acrolein*. World Health Organization, 2002: p. 14.
  17. Kuwata, K., M. Uebori, and Y. Yamasaki, *Determination of Aliphatic and Aromatic-Aldehydes in Polluted Airs as Their 2,4-Dinitrophenylhydrazones by High-Performance Liquid-Chromatography*. Journal of Chromatographic Science, 1979. **17**(5): p. 264-268.
  18. Stevens, J.F. and C.S. Maier, *Acrolein: sources, metabolism, and biomolecular interactions relevant to human health and disease*. Mol Nutr Food Res, 2008. **52**(1): p. 7-25.
  19. Moretto, N., et al., *Acrolein effects in pulmonary cells: relevance to chronic obstructive pulmonary disease*. Ann N Y Acad Sci, 2012. **1259**: p. 39-46.
  20. Feng, Z.H., et al., *Acrolein is a major cigarette-related lung cancer agent: Preferential binding at p53 mutational hotspots and inhibition of DNA repair*.



Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(42): p. 15404-15409.

21. Luo, J., et al., *Mechanisms of acrolein-induced myocardial dysfunction: implications for environmental and endogenous aldehyde exposure*. Am J Physiol Heart Circ Physiol, 2007. **293**(6): p. H3673-84.
22. Ismahil, M.A., et al., *Chronic oral exposure to the aldehyde pollutant acrolein induces dilated cardiomyopathy*. Am J Physiol Heart Circ Physiol, 2011. **301**(5): p. H2050-60.
23. Sklar, J.L., P.G. Anderson, and P.J. Boor, *Allylamine and acrolein toxicity in perfused rat hearts*. Toxicol Appl Pharmacol, 1991. **107**(3): p. 535-44.
24. Toennes, S.W., et al., *A fatal human intoxication with the herbicide allyl alcohol (2-propen-1-ol)*. J Anal Toxicol, 2002. **26**(1): p. 55-7.
25. Wang, G.W., et al., *Acrolein consumption exacerbates myocardial ischemic injury and blocks nitric oxide-induced PKCepsilon signaling and cardioprotection*. J Mol Cell Cardiol, 2008. **44**(6): p. 1016-22.
26. Kesingland, K., et al., *Allylamine toxicity in embryonic myocardial myocyte reaggregate cultures*. Toxicol In Vitro, 1991. **5**(2): p. 145-56.
27. Toraason, M., et al., *Comparative toxicity of allylamine and acrolein in cultured myocytes and fibroblasts from neonatal rat heart*. Toxicology, 1989. **56**(1): p. 107-17.
28. O'Brien, P.J., A.G. Siraki, and N. Shangari, *Aldehyde sources, metabolism, molecular toxicity mechanisms, and possible effects on human health*. Critical Reviews in Toxicology, 2005. **35**(7): p. 609-662.
29. Anderson, M.M., et al., *Human neutrophils employ the myeloperoxidase-hydrogen peroxide-chloride system to convert hydroxy-amino acids into glycolaldehyde, 2-hydroxypropanal, and acrolein - A mechanism for the generation of highly reactive alpha-hydroxy and alpha,beta-unsaturated aldehydes by phagocytes at sites of inflammation*. Journal of Clinical Investigation, 1997. **99**(3): p. 424-432.

30. Vasilyev, N., et al., *Myeloperoxidase-generated oxidants modulate left ventricular remodeling but not infarct size after myocardial infarction*. *Circulation*, 2005. **112**(18): p. 2812-2820.
31. Williams, T.I., et al., *Increased levels of 4-hydroxynonenal and acrolein, neurotoxic markers of lipid peroxidation, in the brain in Mild Cognitive Impairment and early Alzheimer's disease*. *Neurobiol Aging*, 2006. **27**(8): p. 1094-9.
32. Sakata, K., et al., *Increase in putrescine, amine oxidase, and acrolein in plasma of renal failure patients*. *Biochem Biophys Res Commun*, 2003. **305**(1): p. 143-9.
33. Daimon, M., et al., *Increased urinary levels of pentosidine, pyrroline and acrolein adduct in type 2 diabetes*. *Endocr J*, 2003. **50**(1): p. 61-7.
34. Serafini, F., *Conversion of Allyl Alcohol into Acrolein by Rat-Liver*. *Biochemical Journal*, 1972. **128**(5): p. 1103-&.
35. Kaminskas, L.M., S.M. Pyke, and P.C. Burcham, *Strong protein adduct trapping accompanies abolition of acrolein-mediated hepatotoxicity by hydralazine in mice*. *J Pharmacol Exp Ther*, 2004. **310**(3): p. 1003-10.
36. Burcham, P.C. and F. Fontaine, *Extensive protein carbonylation precedes acrolein-mediated cell death in mouse hepatocytes*. *J Biochem Mol Toxicol*, 2001. **15**(6): p. 309-16.
37. Emadi, A., R.J. Jones, and R.A. Brodsky, *Cyclophosphamide and cancer: golden anniversary*. *Nat Rev Clin Oncol*, 2009. **6**(11): p. 638-47.
38. Murray, M., A.M. Butler, and I. Stupans, *Competitive-Inhibition of Human Liver Microsomal Cytochrome-P450 3a-Dependent Steroid 6-Beta-Hydroxylation Activity by Cyclophosphamide and Ifosfamide in-Vitro*. *Journal of Pharmacology and Experimental Therapeutics*, 1994. **270**(2): p. 645-649.
39. McDonald, G.B., et al., *Cyclophosphamide metabolism, liver toxicity, and mortality following hematopoietic stem cell transplantation*. *Blood*, 2003. **101**(5): p. 2043-8.

40. Gurtoo, H.L., et al., *Effects of the induction of hepatic microsomal metabolism on the toxicity of cyclophosphamide*. Br J Cancer, 1985. **51**(1): p. 67-75.
41. Berrigan, M.J., et al., *Protective role of thiols in cyclophosphamide-induced urotoxicity and depression of hepatic drug metabolism*. Cancer Res, 1982. **42**(9): p. 3688-95.
42. Gurtoo, H.L., et al., *Studies on the mechanism of denaturation of cytochrome P-450 by cyclophosphamide and its metabolites*. J Biol Chem, 1981. **256**(22): p. 11691-701.
43. Marinello, A.J., et al., *Metabolism and binding of cyclophosphamide and its metabolite acrolein to rat hepatic microsomal cytochrome P-450*. Cancer Res, 1984. **44**(10): p. 4615-21.
44. Patel, J.M., J.C. Wood, and K.C. Leibman, *The biotransformation of allyl alcohol and acrolein in rat liver and lung preparations*. Drug Metab Dispos, 1980. **8**(5): p. 305-8.
45. Mohammad, M., et al., *Multiple cell-death mechanisms are triggered in hepatotoxicity induced by acrolein, an environmental pollutant and lipid peroxidation product*. Faseb Journal, 2012. **26**.
46. Haberzettl, P., et al., *Role of endoplasmic reticulum stress in acrolein-induced endothelial activation*. Toxicology and Applied Pharmacology, 2009. **234**(1): p. 14-24.
47. Kitaguchi, Y., et al., *Acrolein induces endoplasmic reticulum stress and causes airspace enlargement*. PLoS One, 2012. **7**(5): p. e38038.
48. Bhatnagar, A., et al., *Environmental cardiology: pollution and heart disease*. 2011: p. 336-338.
49. Zheng, J. and O.A. Bizzozero, *Traditional reactive carbonyl scavengers do not prevent the carbonylation of brain proteins induced by acute glutathione depletion*. Free Radical Research, 2010. **44**(3): p. 258-266.
50. Mello, C.F., et al., *Acrolein induces selective protein carbonylation in synaptosomes*. Neuroscience, 2007. **147**(3): p. 674-9.

51. Thompson, C.A. and P.C. Burcham, *Protein alkylation, transcriptional responses and cytochrome c release during acrolein toxicity in A549 cells: influence of nucleophilic culture media constituents*. Toxicol In Vitro, 2008. **22**(4): p. 844-53.
52. Ansari, M.A., J.N. Keller, and S.W. Scheff, *Protective effect of Pycnogenol in human neuroblastoma SH-SY5Y cells following acrolein-induced cytotoxicity*. Free Radic Biol Med, 2008. **45**(11): p. 1510-9.
53. Hristova, M., et al., *The tobacco smoke component, acrolein, suppresses innate macrophage responses by direct alkylation of c-Jun N-terminal kinase*. Am J Respir Cell Mol Biol, 2012. **46**(1): p. 23-33.
54. Spiess, P.C., et al., *Proteomic profiling of acrolein adducts in human lung epithelial cells*. J Proteomics, 2011. **74**(11): p. 2380-94.
55. Mano, J., *Reactive carbonyl species: Their production from lipid peroxides, action in environmental stress, and the detoxification mechanism*. Plant Physiology and Biochemistry, 2012. **59**: p. 90-97.
56. Madian, A.G. and F.E. Regnier, *Proteomic identification of carbonylated proteins and their oxidation sites*. J Proteome Res, 2010. **9**(8): p. 3766-80.
57. Uchida, K., et al., *Protein-bound acrolein: potential markers for oxidative stress*. Proc Natl Acad Sci U S A, 1998. **95**(9): p. 4882-7.
58. Wu, J., J.F. Stevens, and C.S. Maier, *Mass spectrometry-based quantification of myocardial protein adducts with acrolein in an in vivo model of oxidative stress*. Mol Nutr Food Res, 2011. **55**(9): p. 1401-10.
59. Green, N.M., *Avidin .I. Use of [14c]Biotin for Kinetic Studies and for Assay*. Biochemical Journal, 1963. **89**(3): p. 585-&.
60. Hofmann, K. and Y. Kiso, *An approach to the targeted attachment of peptides and proteins to solid supports*. Proc Natl Acad Sci U S A, 1976. **73**(10): p. 3516-8.
61. Bayer, E.A., et al., *Preparation of Ferritin-Avidin Conjugates by Reductive Alkylation for Use in Electron-Microscopic Cytochemistry*. Journal of Histochemistry & Cytochemistry, 1976. **24**(8): p. 933-939.

62. Angerer, L., et al., *An electron microscope study of the relative positions of the 4S and ribosomal RNA genes in HeLa cells mitochondrial DNA*. Cell, 1976. **9**(1): p. 81-90.
63. Kohanski, R.A. and M.D. Lane, *Monovalent avidin affinity columns*. Methods Enzymol, 1990. **184**: p. 194-200.
64. Wu, J., J.F. Stevens, and C.S. Maier, *Mass spectrometry-based quantification of myocardial protein adducts with acrolein in an in vivo model of oxidative stress*. Molecular Nutrition & Food Research, 2011. **55**(9): p. 1401-10.
65. Spiess, P.C., et al., *Proteomic profiling of acrolein adducts in human lung epithelial cells*. Journal of Proteomics, 2011. **74**(11): p. 2380-2394.
66. Madian, A.G., et al., *Determining the effects of antioxidants on oxidative stress induced carbonylation of proteins*. Analytical Chemistry, 2011. **83**(24): p. 9328-36.
67. Meany, D.L., et al., *Identification of carbonylated proteins from enriched rat skeletal muscle mitochondria using affinity chromatography-stable isotope labeling and tandem mass spectrometry*. Proteomics, 2007. **7**(7): p. 1150-63.
68. Ugur, Z., C.M. Coffey, and S. Gronert, *Comparing the efficiencies of hydrazide labels in the study of protein carbonylation in human serum albumin*. Analytical and Bioanalytical Chemistry, 2012. **404**(5): p. 1399-1411.
69. Cech, N.B. and C.G. Enke, *Practical implications of some recent studies in electrospray ionization fundamentals*. Mass Spectrometry Reviews, 2001. **20**(6): p. 362-387.
70. Douglas, D.J., A.J. Frank, and D.M. Mao, *Linear ion traps in mass spectrometry*. Mass Spectrometry Reviews, 2005. **24**(1): p. 1-29.
71. Guilhaus, M., *Principles and Instrumentation in Time-of-Flight Mass-Spectrometry - Physical and Instrumental Concepts*. Journal of Mass Spectrometry, 1995. **30**(11): p. 1519-1532.

72. Marshall, A.G., C.L. Hendrickson, and G.S. Jackson, *Fourier transform ion cyclotron resonance mass spectrometry: A primer*. Mass Spectrometry Reviews, 1998. **17**(1): p. 1-35.
73. Perry, R.H., R.G. Cooks, and R.J. Noll, *Orbitrap Mass Spectrometry: Instrumentation, Ion Motion and Applications*. Mass Spectrometry Reviews, 2008. **27**(6): p. 661-699.
74. Johnson, R.S., et al., *Novel Fragmentation Process of Peptides by Collision-Induced Decomposition in a Tandem Mass-Spectrometer - Differentiation of Leucine and Isoleucine*. Analytical Chemistry, 1987. **59**(21): p. 2621-2625.
75. Cooks, R.G., *Collision-Induced Dissociation - Readings and Commentary*. Journal of Mass Spectrometry, 1995. **30**(9): p. 1215-1221.
76. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data*. Electrophoresis, 1999. **20**(18): p. 3551-67.
77. Roe, M.R., et al., *Targeted 18O-labeling for improved proteomic analysis of carbonylated peptides by mass spectrometry*. J Am Soc Mass Spectrom, 2010. **21**(7): p. 1190-203.
78. Ren, S., T.F. Kalhorn, and J.T. Slattery, *Inhibition of human aldehyde dehydrogenase 1 by the 4-hydroxycyclophosphamide degradation product acrolein*. Drug Metab Dispos, 1999. **27**(1): p. 133-7.
79. Srivastava, S.C., R.K. Upreti, and A.M. Kidwai, *Action of acrolein on rat liver membrane proteins and enzymes*. Bull Environ Contam Toxicol, 1992. **49**(1): p. 98-104.
80. Sun, L.J., et al., *Acrolein is a mitochondrial toxin: Effects on respiratory function and enzyme activities in isolated rat liver mitochondria*. Mitochondrion, 2006. **6**(3): p. 136-142.
81. Patel, J.M., et al., *Selective inactivation of rat lung and liver microsomal NADPH-cytochrome c reductase by acrolein*. Drug Metab Dispos, 1984. **12**(4): p. 460-3.

82. LoPachin, R.M., et al., *Molecular mechanisms of 4-hydroxy-2-nonenal and acrolein toxicity: nucleophilic targets and adduct formation*. Chem Res Toxicol, 2009. **22**(9): p. 1499-508.
83. Chavez, J.D., et al., *Site-specific proteomic analysis of lipoxidation adducts in cardiac mitochondria reveals chemical diversity of 2-alkenal adduction*. J Proteomics, 2011. **74**(11): p. 2417-29.
84. Slade, P.G., et al., *A filtered database search algorithm for endogenous serum protein carbonyl modifications in a mouse model of inflammation*. Mol Cell Proteomics, 2011. **10**(10): p. M111 007658.
85. Hammes-Schiffer, S. and S.J. Benkovic, *Relating protein motion to catalysis*. Annual Review of Biochemistry, 2006. **75**: p. 519-541.
86. Jelokova, J., et al., *Features of structural zinc in mammalian alcohol dehydrogenase. Site-directed mutagenesis of the zinc ligands*. Eur J Biochem, 1994. **225**(3): p. 1015-9.
87. Lake, B.G., et al., *The effect of repeated administration on allyl alcohol-induced hepatotoxicity in the rat [proceedings]*. Biochem Soc Trans, 1978. **6**(1): p. 145-7.
88. Jaeschke, H., C. Kleinwaechter, and A. Wendel, *The role of acrolein in allyl alcohol-induced lipid peroxidation and liver cell damage in mice*. Biochem Pharmacol, 1987. **36**(1): p. 51-7.
89. Haenen, G.R., et al., *Activation of the microsomal glutathione-S-transferase and reduction of the glutathione dependent protection against lipid peroxidation by acrolein*. Biochem Pharmacol, 1988. **37**(10): p. 1933-8.
90. Busenlehner, L.S., et al., *Location of substrate binding sites within the integral membrane protein microsomal glutathione transferase-1*. Biochemistry, 2007. **46**(10): p. 2812-22.
91. Gregus, Z., et al., *Resistance of Some Phase Ii Biotransformation Pathways to Hepatotoxins*. Journal of Pharmacology and Experimental Therapeutics, 1982. **222**(2): p. 471-479.

92. Smith, P.F., et al., *In vitro* cytotoxicity of allyl alcohol and bromobenzene in a novel organ culture system. *Toxicol Appl Pharmacol*, 1987. **87**(3): p. 509-22.
93. Guo, J., et al., *Protein targets for carbonylation by 4-hydroxy-2-nonenal in rat liver mitochondria*. *Journal of Proteomics*, 2011. **74**(11): p. 2370-2379.
94. Shin, N.Y., et al., *Protein targets of reactive electrophiles in human liver microsomes*. *Chemical Research in Toxicology*, 2007. **20**(6): p. 859-867.
95. Madian, A.G., et al., *Differential Carbonylation of Proteins as a Function of in vivo Oxidative Stress*. *Journal of Proteome Research*, 2011. **10**(9): p. 3959-3972.



## **CHAPTER 2**

---

### **MASS SPECTROMETRY QUANTIFICATION OF DIFFERENTIAL GLYCOFORMS IN PROSTATE-SPECIFIC ANTIGEN: APPLICATION OF HILIC SOLID-PHASE EXTRACTION**

## 2.1 INTRODUCTION

### 2.1.1 Background on glycosylation

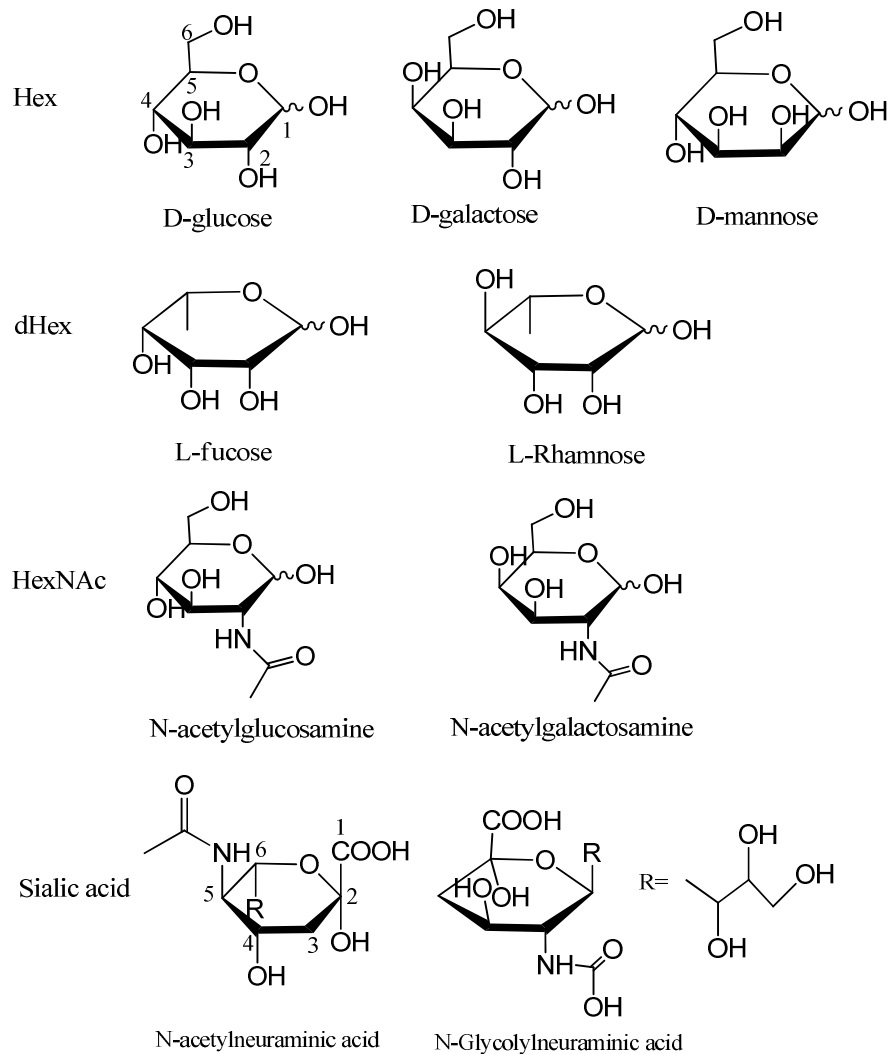
#### *Protein glycosylation*

Protein glycosylation is one of the common post-translational modifications which play important roles in the structure and function of proteins. Almost all studied living organisms including eukaryotes, eubacteria and archae express glycoproteins [1, 2]. Around 50-70% of human proteins are glycosylated, including surface receptors, organelle-resident proteins, secreted proteins, and trafficking proteins. Protein glycosylation is involved in many biological processes such as mediating cell attachment, monitoring the status of protein folding and facilitating the delivery of proteins, stimulating signal transduction pathways, affecting protein-protein interactions and altering the solubility of proteins [3, 4]. Glycosylation could be classified into several subtypes by different glycosidic linkages: N-linked glycosylation, O-linked glycosylation, glypiation, C-linked glycosylation, and phosphoglycosylation. N-linked oligosaccharide attaches to a nitrogen atom of asparagine (Asn) [5]. O-linked glycosylation is attaching sugars to an oxygen atom in serine, threonine or tyrosine [6]. C-linked glycans, which is not commonly observed, link to a carbon atom of a tryptophan side-chain [7]. Glypiation, which is also named GPI anchor attachment, links proteins to lipids through glycans at C-terminus of target proteins [8]. Phosphoglycosylation attaches glycans to serine *via* phosphodiester bond which is observed in parasites and slime molds [9, 10]. More emphasis in the introduction of this chapter will be placed on N-glycosylation since this study is focusing on relative quantification of N-glycans.

### Glycans

Glycans are composed of basic structural unit monosaccharides. The glycosidic bond could be formed between the intramolecular hemiacetal group of one monosaccharide and the hydroxyl group of the other monosaccharide. Monosaccharides that are commonly occurring in nature are shown in Figure 2.1. Glucose (Glu/Glc), Galactose (Gal), and Mannose (Man) are stereoisomers which are called Hexose (Hex), since they all have six carbon atoms with a molecular formula of  $C_6H_{12}O_6$  and a molar mass of 180.16 g/mol. Only D-configured enantiomer is present in biological system. Deoxyhexose (dHex) is hexose without an oxygen atom due to replacement of the hydroxyl group to the hydrogen atom. Fucose (Fuc) and Rhamnose (Rham) are classified as dHex and have a molecular formula  $C_6H_{12}O_5$  and molecular weight at 164.16 g/mol. Rham does not exist in mammalian cells but in Buckthorn, poison sumac, and plants. Fuc is commonly observed in humans. Besides lack of a hydroxyl group, Fucose is distinguished from other six-carbon sugars by L-configuration in biological system. N-acetylglucosamine (GlcNAc) and N-acetylgalactosamine (GalNAc) are both N-acetylhexosamine (HexNAc) which has a molecular formula  $C_8H_{15}NO_6$  and a molar mass of 221.21 g/mol. It is formed by attaching acetic acid to hexosamine by amide bond. Sialic acid is a general term for substituted nine-carbon neuraminic acid. Only alpha anomer is found in protein glycans. Common members found in mammals are N-acetylneuraminic acid (Neu5Ac/NeuAc) and N-glycolylneuraminic acid (Neu5Gc). The chemical formulas and molecular weight of Neu5Ac and Neu5Gc are  $C_{11}H_{19}NO_9$  (309.27 g/mol) and  $C_{11}H_{19}NO_{10}$  (325.27 g/mol) respectively. Neu5Ac is widely present in human

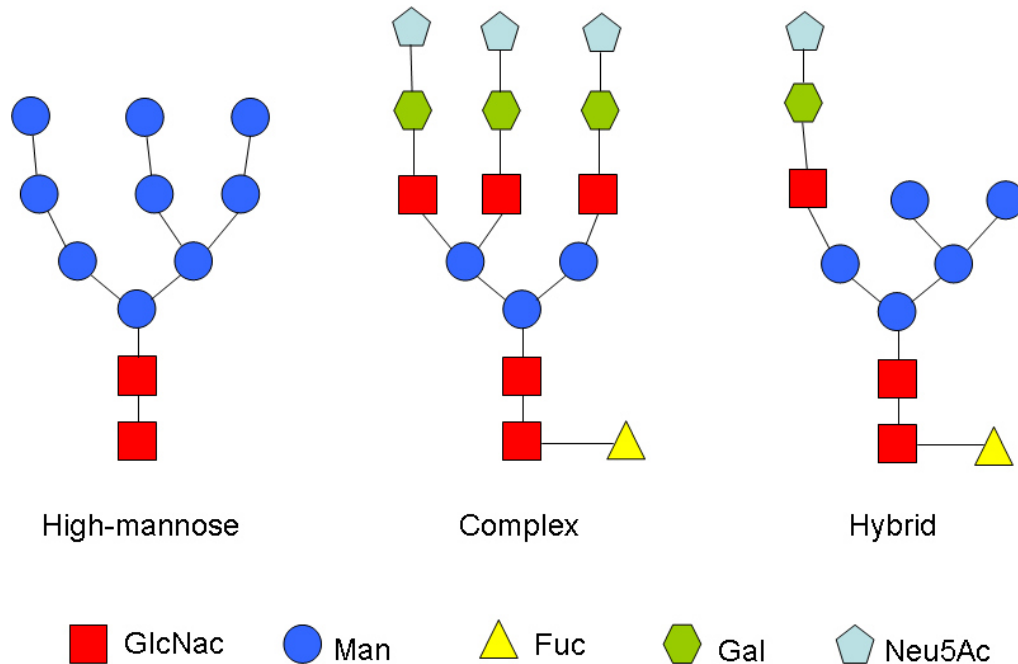
proteins, while Neu5Gc is non-human sialic acid but has been found in apes. Because of the carboxylic group, sialic acid could change the isoelectric point (pI) of proteins.



**Figure 2.1. Haworth projections of naturally occurring monosaccharides.** They are classified into groups by the molecular weight or acidity (sialic acid). D-glucose (Glu/Glc), D-galactose (Gal), and D-mannose (Man) belong to the category of Hexose (Hex). L-fucose (Fuc) and L-Rhamnose (Rham) are deoxyhexose (dHex). N-acetylglucosamine (GlcNAc) and N-acetylgalactosamine (GalNAc) are N-acetylhexosamine (HexNAc). N-acetylneuraminic acid (Neu5Ac/NeuAc) and N-Glycolylneuraminic acid (Neu5Gc) are both called sialic acid.

### N-glycans

N-glycosylation is the most common type of glycosylation. N-linked oligosaccharide attaches to a nitrogen atom of asparagine (Asn) in the sequence of Asn-X-Ser or Asn-X-Thr, where X is any amino acid except Proline [11]. N-glycans have a common pentasaccharide core structure which contains two GlcNAc and three Man. It is because that all N-glycans are originally assembled in endoplasmic reticulum (ER) with an identical 14 sugar unit of  $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$  [12]. A diversity of N-linked glycans is created by further sugar removal in ER and modification in golgi apparatus [13]. N-glycans could generally be classified into three different subtypes: high-mannose, complex, and hybrid [14] as shown in Figure 2.2. High mannose type of glycans has only mannose residues besides core structure. The name for high mannose glycan such as  $\text{Man}_5$  is abbreviation of totally five mannose sugars including two in the core structure. Complex type of glycans contains a variety of the other types of saccharides but not mannose except two in the core structure. It is commonly terminated by sialic acid, and sometimes has fucose residue attached to GlcNAc on the core structure. Complex glycan is predominant type of oligosaccharides on both cell surface and secreted glycoproteins. Hybrid glycan is a hybrid of high mannose and complex types. It is characterized as attaching one mannose in core structure with terminal Man as in high mannose type and the other mannose with other types of monosacchrides as in complex type of glycans.



**Figure 2.2. Structures of three N-glycan subtypes: high-mannose, complex, and hybrid.** High mannose type of glycans has only mannose residues besides core structure. Complex type of glycans contains a variety of the other types of saccharides. Hybrid glycan is characterized as one mannose in core structure substituted with mannose as in high mannose type and the other mannose with other types of monosacchrides as in the complex glycan.

## 2.1.2 Prostate specific antigen

### Prostate cancer

Prostate cancer is malignant tumor developed in the prostate gland of males. Most prostate cancers are growing very slowly and show no symptom in early stage patients. With the development of the cancer, patients will have pain, problems of urinating and erection. Cancer may spread from the prostate to bones and lymph nodes and finally cause death. World-widely prostate cancer resulted in 256,000 deaths in 2010. It is the 2<sup>nd</sup> common cancer and the 6<sup>th</sup> leading cause of cancer death in males [15]. It is most

frequently diagnosed in the United States and least common in South and East Asia. From the statistics of Surveillance, Epidemiology and End Results (SEER), a premier source for cancer statistics in the United States, from 2006 to 2010, the occurrence rate of prostate cancer was 152.0 per 100,000 men per year and the death rate was 23.0 per 100,000 men per year [16]. By estimation, in 2013, 238,590 American men will be diagnosed with cancer of prostate and 29,720 men will die because of it. Early diagnosis and treatment are very important for survival of the patients. When the cancer is confined to primary site, the five year relative survival ratio is 100%. When cancer has spread, the five year relative survival ratio drops to only 27.9% [16]. People have been looking for an accurate diagnostic method to improve the detection rate of prostate cancer in an early stage. An approach named PSA screen has been traditionally applied in testing prostate cancer which will be introduced in the following paragraph.

#### *Prostate specific antigen (PSA) test*

Prostate-specific antigen (PSA) is a glycoprotein (molecular weight around 30 kDa) secreted into seminal fluid by epithelial cells of prostates. This protein is a peptidase which cleaves semenogelins in the seminal coagulum, liquefies semen, and makes sperm swim freely [17]. Normally PSA is restricted within the prostate gland, so the level of PSA in serum is very low. In patients of prostate cancer, the serum concentration of PSA is increasing because PSA is secreted into serum due to disruption of basement membrane. Screening the concentration of PSA in blood, named PSA test, is a widely used method to diagnose and monitor the progress of prostate cancer. Generally, if the PSA level of a man is below 4.0 ng/ml, he is diagnosed as cancer-free [18]. The

concentration range from 4 ng/ml to 10 ng/ml of PSA is categorized as grey zone. A man in grey zone has around 25 % chance to have prostate cancer [19]. If the PSA concentration of patient is higher than 10 ng/ml, there is 60 % chance that he has prostate cancer. However, it is still controversial to utilize PSA test as an indicator of prostate cancer. High PSA level in serum could also be caused by prostatitis or benign prostatic hyperplasia. Patients diagnosed having large chance of prostate cancer by PSA test need to go through further invasive examination such as core needle prostate biopsy and even receive false harmful treatment. Thus, The United States Preventive Services Task Force (USPSTF), a government advisory body, does not recommend the PSA screen for prostate cancer test in healthy men because “the potential benefit does not outweigh the expected harms” as stated [20, 21]. Thus, the specificity of PSA as a cancer biomarker needs to be improved to determine benign from malignant prostate states. Researchers have already made efforts on developing better approaches. Two new tests focusing on gene screen called Prolaris (Myriad Genetics, Salt lake city, UT) and Oncotype DX Prostate Cancer Test (Genomic Health, Redwood city, CA) are recently marketed to help predicting if prostate cancers are aggressive and need immediate treatment. On protein level, the efforts to increase the specificity of prostate cancer diagnosis are still continuing. For example, PSA forms complexes with other proteins such as alpha-1-antichymotrypsin in blood. Researches have tried to compare the free PSA with total PSA since the ratios are different between cancer patients and controls [22]. The specific isoelectric point (pI) subform of PSA could be employed as a new way to improve PSA test [23]. It has been found that the percentage of high pI PSA increases in cancer patients. Change of pI may due to differential expression of glycans, especially sialic acid. It leads



to another promising approach, which is screening glycoforms of PSA. An encouraging finding is oligosaccharides in PSA are different in patients of prostate cancer compared to ones of benign prostate disease. Monitoring alteration of glycosylation may have great potential to increase the specificity of PSA test.

MWVPVVFLTLSVTWIGAAPLILSRIVGGWECEKHSQPWQVLVASRGRAVCGGV  
HPQWVLTAAHCIRNKS VILLGRHSLFHPEDTGQVFQVSHSFPHPLYDMSLLKRFL  
RPGDDSSHDLMMLRLSEPAELTDAVKVMDLPTQEPALGTTTCYASGWGSIEPEEFL  
TPKKLQCVDLHVISNDVCAQVHPQKVTKFMLCAGRWTGGKSTCSGDSGGPLVC  
NGVLQGITSWGSEPCALPERPSLYTKVVHYRKWIKDTIVANP

**Figure 2.3. Sequence of prostate specific antigen (PSA).** The amino acid Asparagine (N) 69 which carries N-glycans is marked in red.

### Glycosylation of PSA

PSA is a glycoprotein containing one glycosylation site which is at Asn 69 (Fig. 2.3). Studies have shown that glycosylation profile of PSA alters in prostate cancer compared to controls. Ariadna Sarrats *et al.* using two-dimensional electrophoresis demonstrated that the pI of PSA was modified in cancer patient serum due to different sialic acid content [23]. Toyohiro Okada *et al.* used high-performance liquid chromatography (HPLC) to analyze different PSA pI subforms from human seminal fluid suggesting that PSA high isoform had higher contents of disialylated glycans [24]. By lectin examination and surface plasmon resonance analysis, Ohyama *et al.* found alpha 2-3 sialic acid increased in prostate cancer patients compared to benign prostate hypertrophy patients [25]. By mass spectrometry analysis, Peracaula *et al.* characterized that glycans of PSA from seminal fluid were mostly sialylated in the low pI fraction and disialylation in the

high pI fraction. Glycans of PSA from prostate cancer cell line LNCaP were neutral with increased fucose and GalNAc [26]. Tajiri *et al.* suggested that PSA in serum of prostate cancer patients mostly fucosylated and sialylated and high mannose and hybrid types were predominant in the seminal plasma by mass spectrometry analysis [27]. Sarrats *et al.* sequenced glycans by LC/MS combined with exoglycosidase array digestions and reported increase of alpha 2-3 sialic acid together with decrease of core fucosylation and total sialylation in sera of prostate cancer patients compared to benign prostatic hyperplasia [28]. Although multiple studies have been conducted in profiling glycosylation of PSA in prostate cancer, the results are not conclusive and validated. The significant changes of major glycoforms from prostate cancer to control are not consistent and some results are even controversial. The differences may due to various samples, handling procedures and analytical methods used in different studies. It is very essential to develop an analytical method that allows robustly and reproducibly identification and quantification of *N*-glycans.

### **2.1.3 Protein glycosylation analysis by mass spectrometry**

#### *Mass spectrometry based glycoproteomics*

There has been growing interest in applying mass spectrometry on analysis of glycosylation nowadays. Mass spectrometry has obvious advantages over traditional techniques such as nuclear magnetic resonance (NMR). The requirement of sample purity in mass spectrometry is much less than other techniques. Mass spectrometry could identify and quantify several glycoforms at one time. The detection limit of MS is as low

as femtomole, so it requires less sample amount. There are generally three ways to identify and quantify N-glycosylation of proteins by mass spectrometry [29, 30]. The first way is to analyze N-glycans after deglycosylation by PNGaseF, which is an enzyme that cleaves the link between asparagine and N-acetylglucosamines [31]. The second way is to analyze glycopeptides after protease digestion, the so-called bottom-up approach [32, 33]. The third way is to analyze glycoproteins directly, the so-called top-down approach. Released glycans generally have good fragmentation in MS/MS, so glycan analysis is better for structural analysis of glycans compared to bottom-up and top-down analysis. The disadvantage of glycan analysis is that it could not determine the glycosylation site due to loss of connection between glycan and the protein, thus it could not tell if the glycan is from the target protein. Glycopeptide and glycoprotein analysis are able to rule out the false positives because they could discriminate contaminants by measuring mass from target peptide or protein. Usually, glycan analysis provides all possible glycan structures and then glycopeptide or protein analysis is utilized to confirm the site of modification or to quantify glycoforms [34]. Bottom-up method is more commonly used in proteomics field than top-down method because intact-protein analysis is more complicated than digested-peptide analysis, especially when the glycoprotein is very large and/or carries multiple glycosylation sites. Top-down analysis also has more requirements on instruments. In this study, glycopeptide approach has been applied to quantify different glycoforms of PSA, so the following introductions focus on bottom-up method.

### Mass spectrometer

Different types of mass spectrometers have been employed in characterization and quantification of glycopeptides. Ionization methods of matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI, introduced in Chapter 1 section 1.1.2) have been commonly utilized in quantitative glycopeptides analysis [34-36]. In MALDI, the energy from the laser is absorbed by a matrix and transferred to analytes. Molecules are then protonated in positive-mode and desorbed becoming gas-phase ions [37]. Because of different ionization mechanisms, generally ESI generates multiple-charged ions and MALDI forms single-charged peptides. Loss of sialic acid has been observed in glycosylated peptides because of metastable decay in MALDI/TOF [38]. To reduce the loss of sialic acid during MALDI analysis, methods have been developed such as protecting sialic acid by an amidation reaction [39], changing the matrix [40-42], and using IR MALDI instead of UV MALDI [43]. A variety of MS/MS dissociation techniques have been applied to fragment glycopeptides. Collision-induced dissociation (CID, introduced in Chapter 1, section 1.12) cleaves glycosidic bonds in glycans. High-energy collisional dissociation (HCD) is a type of collision dissociation which has high-energy from high voltage or pressure. Equipped in Orbitrap mass spectrometer, it has been applied to determine amino acid sequences of glycopeptides since it not only dissociates glycosidic bonds but further induces fragmentation of peptide backbone. Electron capture dissociation (ECD) and electron transfer dissociation (ETD) are dissociation methods based on energy from electron. Ions in ECD directly capture free electrons, while ions in ETD contact radical anions such as azobenzene to get internal energy. The energy induces fragmentation of peptides to generate c and z ions (nomenclatures of ions are shown in Figure 1.4 in Chapter 1), while the structures of

glycans are kept intact. ECD and ETD are suitable for sequencing peptide and determining glycosylation site. Complimentary data could be produced by using CID, ETD/ECD and HCD to characterize both oligosaccharide and peptide structures [44].

#### Quantification of glycopeptides by mass spectrometry

The relative quantities of different glycoforms could be determined by labeling or label-free strategies in mass spectrometry. Stable isotope labeling and isobaric labeling which have been utilized in glycan quantification [45-47] are potentially applicable in glycopeptide quantification. Label-free quantification methods are preferred when analyzing multiple samples since most labeling quantification methods are limited to two biological samples. Normalization of data is necessary when using label-free methods due to variation in different MS runs of samples. Previously, researchers normalized their data by dividing the target ion abundance by the total ion abundance [48-50]. Since glycopeptides are ionized poorer than non-glycopeptides, changes in abundance of non-glycopeptides could largely affect the quantification on glycopeptides by this normalization technique. To solve the problem, the normalization method is adjusted to divide the abundance of target glycopeptides by the summed abundance of all glycopeptides present in a MS spectrum. Applying this normalization method, Rebecchi *et al.* demonstrated quantification of glycopeptides based on peak intensities in ESI-LTQ-FTICR produce robust and reproducible quantitative data by using asialofetuin and ribonucleaseB as standards [34]. Thaysen-Andersen *et al.* proved that MALDI-TOF signal strength of glycopeptides accurately reflected the relative abundances of glycoforms in standard glycoproteins ribonucleaseB, IgG, and ovalbumin [35].

## 2.1.4 Purification of glycopeptides

### *Necessity of glycopeptides purification before LC-MS/MS analysis*

Before MS analysis, glycopeptides need to be well separated from each other and more importantly from abundant non-glycosylated peptides. Abundant signals from non-glycosylated peptides not only suppress signals of glycopeptides, but also affect the quantification if the signals are saturated in the detector. Most of proteomics labs including ours couple mass spectrometers with C18 reverse-phase liquid chromatography (RPLC), except some labs focusing on characterization of glycoproteins have other LC separation options such as HILIC [51] and porous graphitized carbon (PGE) [52]. RPLC is usually not capable of separating glycopeptides well as observed in our preliminary data. When online separation is not enough for removing non-glycosylated peptides, a purification step for glycopeptides before LC-MS/MS is necessary for detection and accurate quantification.

### *Techniques for glycopeptide purification*

There are multiple techniques developed for purifying and enriching glycopeptides. The following summarized reported methods based on different properties between glycopeptides and non-glycopeptides. 1) Hydrophilic interaction liquid chromatography (HILIC). It is based on the fact that usually glycopeptides are more hydrophilic than non-glycopeptide due to large glycan moiety. 2) Lectin. Lectins are a series of proteins that have very specific affinity to a certain type of sugar moiety. For example, Concanavalin A (conA) binds  $\alpha$ -D-mannosyl and  $\alpha$ -D-glucosyl residues and Wheat Germ Agglutinin

(WGA) has affinity to N-acetyl-D-glucosamine and sialic acid. A lot of studies have used a combination of several lectins to enlarge the isolated types of glycans [53-55]. 3) Hydrazide resin. After oxidizing the diols of saccharides by periodate, formed aldehydes are covalently attached to hydrazide resin, and then immobilized glycopeptides could be released by enzyme [56-58]. However, this method suffers from loss of carbohydrate structures during deglycosylation. 4) Anion exchange chromatography. It is employed to separate sialylated glycopeptides based on acidity. 5) Size exclusion chromatography. The mechanism is that most N-linked glycopeptides are significantly larger than the non-glycosylated peptides [59]. 6) Boronic acid. A strategy has been developed that utilizes boronic acid beads to capture cis-diols of glycans by forming heterocyclic diesters at high pH and release at low pH [60]. In this study, HILIC was used to purify variety types of N-glycopeptides from PSA.

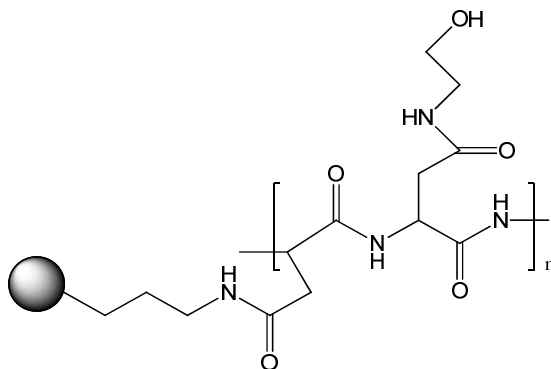
*Hydrophilic interaction liquid chromatography (HILIC) solid-phase extraction (SPE) to purify glycopeptides*

The name of hydrophilic interaction liquid chromatography (HILIC) was proposed by Dr. Andrew Alpert in 1990 [61]. It was reported to be a variant of normal phase liquid chromatography (NPLC), which employed polar stationary phases. However, the mobile phase components of HILIC and NPLC are different. NPLC uses 100% organic solvent or mixable organic solvents and HILIC uses organic solvent with a proportion of water in eluent solvent which is similar to reverse-phase. There are various types of HILIC, including carbohydrate gel matrices such as cellulose or sepharose, charged type such as ones with amino or cyano groups, zwitterionic type such as ZIC-HILIC (Sequant/Merck,

Uppsala, Sweden) with inner quaternary amines and outer sulfonic acids, and neutral type such as polyhydroxyethyl aspartamide (PolyHEA) (PolyLC, Columbia, MD) which is an amide-based material. Solid-phase extraction (SPE) is a technique to separate targeted analytes from other compounds. A cartridge/column packed with stationary phase is loaded with sample and then washed with buffer. Since generally glycopeptides are more hydrophilic than non-glycopeptides because of the large glycan moieties, glycopeptides are adsorbed to HILIC while non-glycopeptides are removed. Finally an eluent buffer will be added to recover glycopeptides. HILIC has been used as SPE materials for purification of glycopeptides from both complex and simple protein samples in several studies [34, 35, 51, 62-64]. Complex sample studies are usually large-scale characterization, such as studying a whole cell lysate, in which the research aims at identification of glycosylated proteins rather than determining oligosaccharides for each protein. Simple sample is usually from a purified protein with multiple glycoforms at one or several glycosylation sites. Studies for simple samples are aiming at characterizing or quantifying glycans of a specific protein. HILIC purification has been applied in a few quantification study of a single glycoprotein. Rebecchi *et al.* used sepharose to purify glycopeptides from standard proteins and quantify them by MALDI-TOF. They proved the HILIC purification step did not affect accurate quantification of glycopeptides [34]. In the product introduction of ZIC HILIC, it indicated purification of glycopeptides could be performed without loss of quantitative information. Thaysen-Andersen *et al.* proved that PolyHEA was suitable for quantitative purification of high-mannose N-glycopeptides in a standard glycoprotein Ribonuclease (RNase B) [35]. In this study, HILIC material



PolyHEA has been proved to be applicable in purification of glycopeptides with different sizes and types of glycans.



**Figure 2.4.** Schematic illustration of polyhydroxyethyl aspartamide (PolyHEA) stationary phase.

### 2.1.5 Background of the interlaboratory study

Glycoprotein research group of Association of Biomolecular Resource Facilities (ABRF) organized a worldwide inter-laboratory study in 2012. ABRF is a society dedicated to advancing core and research biotechnology laboratories through research, communication, and education throughout the world (<http://www.abrf.org/>). The goal was to determine the capability of the glycoproteomics community to do comparative analysis of protein N-glycosylation, since glycosylation analysis by mass spectrometry was still a great challenge in this field. They sent out two commercial available PSA purified from seminal fluid to participated labs. Totally 35 labs from North America (20), Europe (12), Australia (1), Japan (1), and China (1) initially joined and finally 24 labs were capable of finishing analysis and sent back 26 data sets. Final data analysis showed that 4 labs had obvious analytical problems and 5 labs did not detect most of major glycans or their

quantifications were very different from others. Quantification data from other 15 labs including ours was generally consistent. A consensus data file included 3 datasets from glycan analysis, 4 from glycoprotein analysis and 10 from glycopeptide analysis was generated. The combined results were published at the journal of Molecular & Cellular Proteomics [58].

## **2.2 EXPERIMENTAL PROCEDURES**

### **2.2.1 Trypsin Digestion**

PSA (lot No. M02015, 100 µg) and PSA high isoform (lot No. M15097, 20 µg) from Lee Biosolutions (St. Louis, MO) were received in a buffer containing 0.05 M phosphate, 0.15 M NaCl, 0.09% NaN<sub>3</sub>, pH=7.5. Bovine fetuin (Sigma-Aldrich, St. Louis, MO) was solubilized in 100 mM ammonium bicarbonate (AMBIC) buffer (pH=8.2). Samples were stored in 4 °C as indicated. Twenty µg of PSA samples were diluted in 50 µl of 100 mM AMBIC buffer. Each sample (20 µg) was reduced by 90 nmol dithiothreitol (DTT) at 56 °C for 1 h and alkylated by 190 nmol iodoacetic acid (IAA) at room temperature for 30 min. Excess IAA was quenched by an additional 90 nmol DTT. Each sample was incubated with 0.4 µg trypsin at 37 °C overnight. The reaction was quenched by 5 µl formic acid and saved in -80 °C.

### **2.2.2 Purification of glycopeptides by HILIC SPE**

SPE column was made by 200  $\mu\text{m}$  i.d. x 365  $\mu\text{m}$  o.d. fused silica capillary tubing (Polymicro Technologies, Phoenix, AZ). A frit was made by drawing silicate solution into the column *via* capillary effect and polymerizing by applying heat (375 °F) with a solder iron at about 3 cm from top of column. HILIC polyhydroxyethyl aspartamide (PolyHEA) (12  $\mu\text{m}$  particle size, 300 Å pore diameter) was from PolyLC (Columbia, MD). The packing material was suspended in 95% acetonitrile (ACN) with 5% isopropanol and loaded to the column by pressure bomb. In the following steps, the flow rate was kept at 6-7  $\mu\text{l}/\text{min}$ . The SPE column was washed with milliQ water, followed by 0.2 M sodium phosphate and 0.1 M ammonium acetate. Then, the column was equilibrating with buffer A (80% ACN, 19% H<sub>2</sub>O, 1% TFA, pH=1.6) before loading the sample. Each HILIC SPE column was only used once. Tryptic digest was reduced to near dry and reconstituted in buffer A. Ten microgram peptides were loaded into HILIC SPE column by a pressure bomb. Non-glycosylated peptides were removed by washing with 60  $\mu\text{l}$  buffer A. Elution from the HILIC SPE column was performed by 30  $\mu\text{l}$  5% formic acid (FA). Eluent was further dried in a vacuum centrifuge and reconstituted in solvent A (2% ACN, 0.1 % FA). Each PSA peptide samples were separated by different HILIC columns and analyzed by LC-MS/MS twice.

### **2.2.3 Mass spectrometric analysis**

LC-MS/MS analysis was performed using Qstar Elite (AB Sciex, Foster City, CA). Enriched glycopeptides from 2-4  $\mu\text{g}$  PSA or fetuin tryptic digest were loaded onto a pre-column (75  $\mu\text{m}$  x 4 cm) packed with 5  $\mu\text{m}$  Monitor C18 particles (Column Engineering.

Ontario, CA). A linear gradient from 0% to 50% solvent B (98% ACN, 0.1% FA) was ran over 50 minutes on Tempo nanoLC system coupled to Qstar at a flow rate of 100 nL/min. The analytical column was 75  $\mu\text{m}$   $\times$  7 cm packed with 3  $\mu\text{m}$  C18 Monitor particles (Column Engineering, Ontario, CA). Precursor ions were scanned over the range of m/z 500-2500. MS/MS spectra were acquired by selecting three most abundant peaks within 800-1800 and generating fragments with m/z range 100-2000 by auto-collision energy. Dynamic exclusion was set at 60 seconds.

#### **2.2.4 Identification and quantification of glycopeptides**

Glycopeptides were identified by matching actual peaks acquired in the MS scan to theoretical glycopeptide masses of known glycoforms. Totally 83 PSA glycans identified in published papers were used as a database for calculating theoretical mass-to-charge ratios of glycopeptides [24-28, 65-69]. The table of theoretical glycopeptide m/z was constructed by the following steps. (1) Theoretical tryptic fragments were calculated by a web-based program ProteinProspector (<http://prospector.ucsf.edu/>). (2) Masses of glycans were calculated by excel with mono masses of saccharides: 291.0954 Da for NeuAc, 203.0794 Da for HexNAc, 162.0528 for Hex, and 146.0579 for dHex. (3) After combining peptide masses with known glycan masses, the calculated glycopeptide masses were converted to m/z in different charge states. Glycopeptides were identified by comparing theoretical masses and peaks in MS when the mass error was within 20 ppm. MS/MS data were interpreted manually to determine the presence of at least two N-glycosylation diagnostic ions: m/z=204.1(HexNAc<sub>1</sub>), 274.1(Neu5Ac-H<sub>2</sub>O), 292.1(Neu5Ac<sub>1</sub>), 366.1(Hex<sub>1</sub>HexNAc<sub>1</sub>), 454.2(Hex<sub>1</sub>Neu5Ac<sub>1</sub>), 512.2(Fuc<sub>1</sub>Hex<sub>1</sub>HexNAc<sub>1</sub>

), 528.2(Hex<sub>2</sub>HexNAc<sub>1</sub>), 657.2(Hex<sub>1</sub>HexNAc<sub>1</sub>Neu5Ac<sub>1</sub>), and 674.3(Fuc<sub>1</sub>Hex<sub>2</sub>HexNAc<sub>1</sub>). The peptide NKSIVLLGR with one tryptic miss cleavage was used to quantify PSA glycans. The peptide LCPDCPLLAPLNSDR was used to quantify fetuin glycans. Quantification was done by manual inspection of MS data using Analyst QS 2.0 with the following three steps. (1) Peak areas of first four isotope peaks from extracted ion chromatography for each glycopeptide ion were integrated by IntelliQuan. (2) The peak areas for all charge states (2+/3+ for PSA and 3+ /4+ for fetuin) of the same glycopeptide were summed together. (3) The peak area sum for each glycopeptide was divided by the total area for all identified glyco-isoforms containing the peptide.

## **2.3 DISCUSSION**

### **2.3.1 Method development**

#### *Enzyme digestion*

Trypsin was selected for PSA glycosylation analysis, since there were lysine and arginine close to Asn (N) 69 which was attached with glycans. In this interlaboratory study, most participated labs (12 out of 17) utilizing glycopeptide approach chose enzyme trypsin. Most labs (8 out of 10) whose results were included in the consensus data used trypsin while the other two labs chose Lysine C and chymotrypsin. Trypsin was expected to generate glycopeptides NK or NKSIVLLGR (one missed cleavage) to be detected by mass spectrometry. In our data files, peptide NK was not observed but peptide NKSIVLLGR was detected after trypsin digestion. It was thought due to bulky glycans

which blocked the enzyme cleavage at K between N and S. In the consensus data, three labs including ours used peptide NKSVILLGR for quantification, while three labs used peptide NK. The phenomenon that peptide NK without miss cleavage was not observed in our data, but was detected by other labs when using the same enzyme trypsin, might be because of different digestion efficiency. The efficiency of trypsin digestion might be affected by conditions such as pH, time, and ratio. Besides NKSVILLGR, AVCGGVLVHPQWVLTAAH CIRNK and AVCGGVLVHPQWVLTAAH CIRNKSVIL LGR which contained N69 with one or two missed tryptic cleavage (s) were also observed in our MS spectra. There were two ways to conduct the quantification based on the detected peptides. One way was to quantify solely on the most abundant peptide NKSVILLGR. The other way was to sum the peak areas of three peptides carrying same type of glycans and quantify based on summed peak areas. It was found that the relative ratio of the three peptides carrying a same glycan was not consistent in different replicates, so we quantified only based on one peptide NKSVILLGR. In this interlaboratory study, a similar conclusion was drawn in regard of the number of peptides chosen to conduct quantification. Most of labs (7 out of 10) in consensus data used a single peptide sequence for glycan composition quantification. Those excluded from consensus data all used more than one peptide for quantification. Different ionization efficiency of glycopeptides and false identification might be responsible for the inaccuracy quantification on multiple peptides.

Evaluation of PolyHEA SPE on quantitative purification of glycopeptides

HILIC material polyhydroxyethyl aspartamide (PolyHEA) which was available in our lab, had not been comprehensively evaluated on application of quantitative purification of complex and hybrid types of N-glycans [35]. It was known that PSA contained a variety of sizes and types of N-glycans, so it was very necessary to make sure that purification of glycopeptides by PolyHEA did not loss quantitative information under proper procedures.

Morten Thaysen-Andersen *et al.* concluded that sample overloading in HILIC affected relative quantification of glycopeptides [35]. They used Ribonuclease B as a model glycoprotein, which was occupied with a series of high-mannose glycans at one N-glycosylation site. Loading 50 fmol RNase B digest per nL HILIC material provided accurate glycoprofiles. Increasing the ratio to 125 fmol/nL or even as high as 600 fmol/nL made the glycan profile bias in higher mannosylated glycopeptides. For example, the percentage of Man5 after HILIC purification at 50 fmol/nL was 48%, but it was decreased to 4% at 300 fmol/nL. The percentage of Man9 rose from 5% to 25 % when the ratio of sample to HILIC material increased from 50 to 600 fmol/nL. It could be explained that the retention of more hydrophilic glycopeptides increased after applying excess samples because of competitive binding. Thus, the relative loading amount in this study was kept at 1 µg of PSA digest per µl of HILIC material, which was 30-40 fmol/nl.

There were studies that showed adding a component of trifluoroacetic acid (TFA) in loading and washing solvents could assist the separation of non-glycopeptides from glycopeptides in PolyHEA [56, 70]. The mechanism was that increase of hydrophobicity

of non-glycopeptides after TFA electrostatic attachment was higher than glycopeptides. Glycopeptides were still able to adsorb to PolyHEA when non-glycopeptides were less capable of binding to HILIC. The addition of TFA in HILIC buffers had never been used in quantitative analysis in the literature, thus it was evaluated in this study in purification of glycopeptides.

At first, a commercial available glycoprotein fetuin was used to evaluate whether PolyHEA purification would affect the quantitation of glycopeptides by comparing our results to the literature. Fetuin was trypsin digested, purified by PolyHEA and then quantified by RPLC-MS/MS. The relative abundances of four glycoforms in Asn 138 were calculated based on the glycopeptide LCPDCPLLAPLNSR. In the reference paper, they quantified glycoforms in fetuin using two approaches: glycopeptide analysis and glycan analysis [35]. The comparison of results was shown in Table 2.1. Most abundant glycan Hex<sub>6</sub>HexNAc<sub>5</sub>NeuAc<sub>3</sub> and second most abundant Hex<sub>5</sub>HexNAc<sub>4</sub>NeuAc<sub>2</sub> were consistent between our data and reference data. However, the least abundant glycans were inconsistent. The inconsistency might be because of different sources of fetuin used in their study and our study. It might also be due to different analytical methods, since even in the same group they produced different results using two differential approaches: peptide and glycan analysis (Table 2.1). The error incorporated the differences of instruments, human operations, and quantification methods. It might be inappropriate to directly compare the glycan quantification data to the reference since there were a lot of variations between their and our evaluations. To control the variation to be just the HILIC purification step, the glycan profiles with or without PolyHEA SPE were compared. Since the glycans in fetuin were almost all sialylated, in order to be more comprehensive



in the evaluation study which should include diverse types of glycans, another standard glycoprotein was utilized, which would be introduced in the following paragraph.

**Table 2.1.** Relative abundances of selected glycans attached to Asn 158 of glycoprotein fetuin in our data and reference data [35]. In our study, a glycopeptide approach was utilized, while in reference they used both peptide and glycan analysis. The percentage was calculated by dividing the peak area of target glycoform by the sum peak area of four selected glycoforms.

Glycoform	Our data	Reference data	
	Peptide (n=1)	Peptide (n=5)	Glycan (n=5)
Hex <sub>6</sub> HexNAc <sub>5</sub> NeuAc <sub>2</sub>	0.3%	15.6±0.91%	11.8±0.33%
Hex <sub>6</sub> HexNAc <sub>5</sub> NeuAc <sub>3</sub>	63.3%	51.0±1.53%	51.9±0.33%
Hex <sub>6</sub> HexNAc <sub>5</sub> NeuAc <sub>4</sub>	9.2%	6.8±0.86%	14.1±0.94%
Hex <sub>5</sub> HexNAc <sub>4</sub> NeuAc <sub>2</sub>	27.3%	26.6±0.67%	22.2±0.84%

Two different sources of PSA were distributed, designated as “PSA” and “PSA high form”. From preliminary study, “PSA” was found to be purer than “PSA high form”. The interference from non-glycopeptides was little in “PSA”, but much more in “PSA high form”. Pure “PSA” could be directly analyzed by LC-MS/MS without HILIC purification, thus it could also be utilized as a standard glycoprotein to evaluate PolyHEA on quantitative purification of glycopeptides. In “PSA high form”, the nonglycosylated peptides suppressed the signals from glycopeptides, thus HILIC purification was still necessary for quantification. This work on validating the method of PolyHEA SPE purification of glycopeptides facilitated studies on the real biological samples which usually were contaminated with multiple proteins.

**Table 2.2.** The percentages and ranks of selected glycoforms in PSA before and after PolyHEA purification. Quantification was based on glycans attached to peptide NKSVILLGR. The percentage was calculated by dividing the peak area of target glycoform by the sum of peak area of seven selected glycoforms. They were ranked from the most abundant (1) to least abundant (7).

Glycoforms	Before HILIC		After HILIC	
	Percentage	Rank	Percentage	Rank
Hex <sub>4</sub> HexNac <sub>5</sub> dHex <sub>1</sub> NeuAc <sub>2</sub>	2.7±0.3%	6	1.5±0.2%	6
Hex <sub>5</sub> HexNac <sub>4</sub> NeuAc <sub>1</sub>	3.3±0.2%	5	3.9±0.3%	5
Hex <sub>5</sub> HexNac <sub>4</sub> dHex <sub>1</sub> NeuAc <sub>1</sub>	31.8±0.5%	1	32.6±0.7%	1
Hex <sub>5</sub> HexNac <sub>4</sub> dHex <sub>1</sub> NeuAc <sub>2</sub>	20.4±0.2%	3	18.1±1.4%	3
Hex <sub>5</sub> HexNac <sub>4</sub> dHex <sub>1</sub>	1.3±0.3%	7	1.1±0.3%	7
Hex <sub>4</sub> HexNac <sub>4</sub> dHex <sub>1</sub> NeuAc <sub>1</sub>	27.5±0.4%	2	29.1±0.4%	2
Hex <sub>4</sub> HexNac <sub>5</sub> dHex <sub>1</sub> NeuAc <sub>1</sub>	8.3±0.1%	4	10.3±0.4%	4

Seven representative glycan compositions including different numbers of Hex, HexNac, dHex, and NeuAc were selected from PSA based on peptide NKSVILLGR. Table 2.2 showed the percentages of these glycopeptides were generally consistent comparing the data from before and after PolyHEA purification. Rebecchi *et al.* indicated that the percentages of differential glycoforms might be slightly different when analyzed by different techniques, but the ranks should stay the same [34]. Thus, the ranks of these glycoforms were listed based on their relative percentages (Table 2.2). The consistent ranking indicated that the glycan profile did not change after PolyHEA purification. The glycan moieties were sufficient for the retention of glycopeptides, regardless of the sizes and types of sugars. In conclusion, PolyHEA SPE could be applied in quantitative analysis of glycopeptides. In addition, almost all obtained MS/MS contained glycan signature ions which indicated that all purified peptides were glycopeptides.

### LC-MS/MS analysis

From previous experience, it was known that quantification of extra high abundant peptides was not accurate in Qstar Elite. Shapes of ion peaks which should be gaussian distribution were distorted at high intensity due to saturation of signals in detector. Normally, peak distortion was observed at signal above 5000 counts. To ensure accuracy of glycopeptide quantification which was based on calculations of peak areas, we managed to maintain the most abundant ions below 1500 counts by loading fewer samples in LC-MS/MS. Reducing the amount of sample decreased the identification number of glycopeptides due to the limitation of instrument sensitivity. However, the most important task in this study was accurate measurement of high abundant glycol-isoforms, rather than the number of identification.

After peaks in MS data were matched to the mass-to-charge ratios calculated by known glycan compositions and peptide sequence, MS/MS spectra were used to confirm the assignment. Collision induced dissociation (CID) usually cleaved at the glycosidic bonds of glycans. In this study, CID did not generate fragments from PSA glycopeptides well enough to extract glycan structure information. In other words, it was not possible to indicate the glycan compositions from MS/MS spectra created by CID. Adjusting the conditions such as collision energy and accumulation time did not improve the spectra. However, the characteristic ions of glycans were very abundant which could confirm that the selected ions were glycopeptides. Identification of PSA glycopeptides within 20 ppm

mass error in MS spectra was validated by characteristic ions of glycans in MS/MS spectra.

### **2.3.2 Data analysis and result discussion**

#### *Identification and quantitation*

Twenty-eight different glycoforms were identified and quantified in PSA and PSA high isoform. Their relative percentages were shown in Table 2.3. The details to quantify glycoforms were described in experimental section. Generally, peaks in MS data were matched with calculated theoretical masses of glycopeptides in PSA, and then their areas were further calculated from extracted ion chromatography. The lists of N-glycan composition from 24 participating laboratory ranged from 8 to 58 glycan compositions. In final statistical analysis, 61 glycoforms were higher than 0.1% and observed by more than one lab. Glycans were classified into major, intermediate, and minor groups. Seven major N-glycans were detected by more than 65% of labs. Eleven intermediate N-glycans were detected by 30-65% of participated labs. The other 43 glycans were in minor group which were observed by less than 30% of the participants. The classification was based on detection frequency but it was also related to the relative intensity since more abundant glycoforms had more chance to be observed. The identification number of glycoforms was related to instrument sensitivity, loading amount and the analysis method. For example, the state-of-art mass spectrometer Orbitrap usually provided more peptide identification number than Qstar Elite from a same sample because of greater sensitivity. Glycan analysis was expected to identify more glycan compositions than glycopeptide

and glycoprotein analysis because the generated fragment spectra could be searched in a database to identify new compositions which had not been reported by previous researches.

**Table 2.3.** Relative abundances of identified 28 glycoforms in PSA and PSA high isoform. Glycoforms composed with different numbers of Hex, HexNAc, dHex, and Neu5Ac were identified in two different PSA samples. The relative intensities were calculated by comparing the peak areas of glycopeptides to the total peak area of all glycopeptides (n=2).

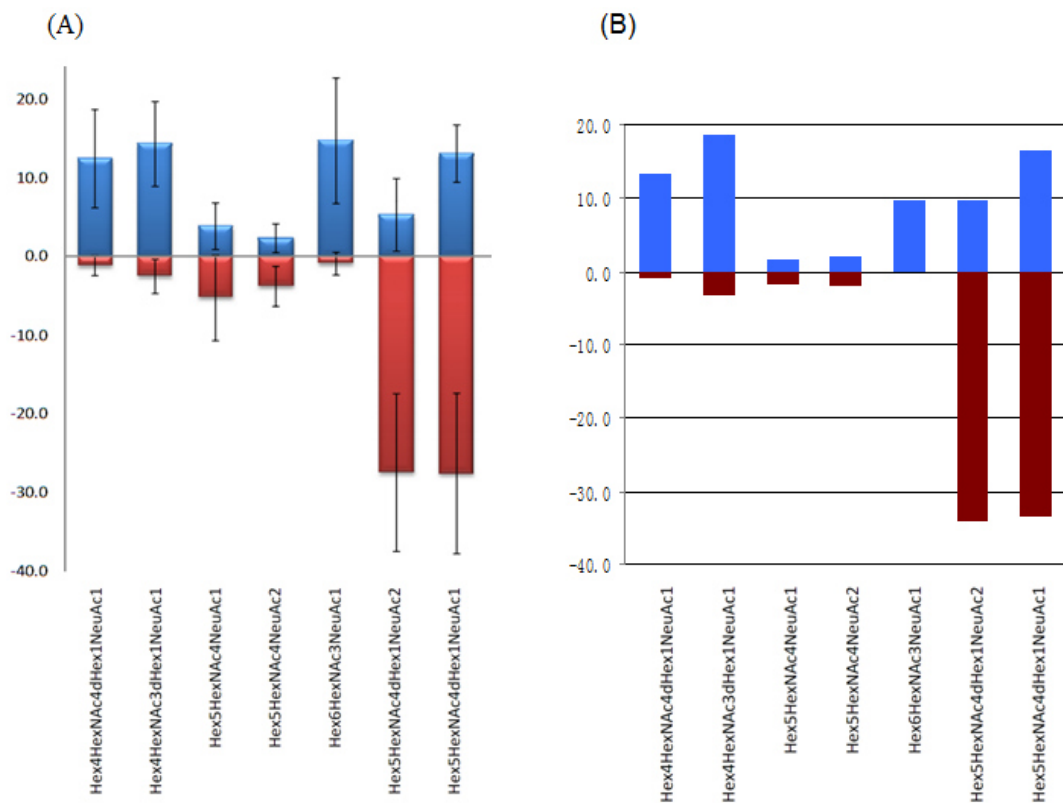
Glycoforms				Percent abundance	
Hex	HexNAc	dHex	Neu5Ac	PSA	PSA high isoform
5	4	1	2	9.9±2.3%	34.0±1.6%
5	4	1	1	16.6±0.3%	33.3±0.9%
4	5	1	2	1.0±0.2%	6.0±0.3%
5	5	0	1	0.0±0.0%	6.0±0.3%
4	5	1	1	4.7±0.3%	4.8±0.6%
4	3	1	0	4.0±0.5%	3.7±0.1%
4	3	1	1	18.6±1.6%	3.1±0.2%
5	4	1	0	0.8±0.2%	2.7±0.4%
5	4	0	2	2.2±0.4%	1.8±0.1%
5	4	0	1	1.8±0.4%	1.6±0.1%
4	4	1	1	13.4±1.5%	1.0±0.0%
3	3	1	1	0.7±0.2%	0.7±0.0%
4	5	0	1	0.9±0.1%	0.6±0.1%
5	5	0	0	0.0±0.0%	0.5±0.0%
3	5	1	1	4.0±0.3%	0.2±0.0%
4	4	1	0	1.1±0.2%	0.1±0.1%
6	3	0	1	9.8±0.9%	0.0±0.0%
5	2	0	0	2.8±0.8%	0.0±0.0%

4	3	0	1	2.2±0.4%	0.0±0.0%
4	2	0	0	0.9±0.2%	0.0±0.0%
6	3	0	0	0.8±0.2%	0.0±0.0%
3	4	1	0	0.6±0.2%	0.0±0.0%
5	5	1	1	0.2±0.0%	0.0±0.0%
5	4	2	1	1.4±0.1%	0.0±0.0%
5	5	2	1	1.0±0.0%	0.0±0.0%
2	2	0	0	0.6±0.0%	0.0±0.0%
5	3	0	1	0.4±0.0%	0.0±0.0%
5	3	1	1	0.2±0.0%	0.0±0.0%

### Comparison of results

In consensus data composed from 15 labs, seven major glycans detected by more than 65% of labs were listed as follows: Hex<sub>5</sub>HexNac<sub>4</sub>dHex<sub>1</sub>NeuAc<sub>1</sub> (95%/93%), Hex<sub>4</sub>HexNac<sub>4</sub>dHex<sub>1</sub>NeuAc<sub>1</sub> (80%/71%), Hex<sub>5</sub>HexNac<sub>4</sub>NeuAc<sub>1</sub> (82%/71%), Hex<sub>5</sub>HexNac<sub>4</sub>NeuAc<sub>2</sub> (84%/75%), Hex<sub>6</sub>HexNac<sub>3</sub>NeuAc<sub>1</sub> (68%/64%), Hex<sub>5</sub>HexNac<sub>4</sub>dHex<sub>1</sub>NeuAc<sub>2</sub> (93%/89%), Hex<sub>4</sub>HexNac<sub>3</sub>dHex<sub>1</sub>NeuAc<sub>1</sub> (80%/75%). Two percentages in parentheses meant the percentage of participating laboratories detecting the N-glycan and the percentage of participating labs using bottom-up method detecting the N-glycan. For example, 82% (18 out of 22) participating labs and 71% (12 out of 17) labs using bottom-up method detected Hex<sub>5</sub>HexNac<sub>4</sub>NeuAc<sub>1</sub>. We had detected all seven major glycans and their relative abundances were consistent with averages in consensus data (Fig. 1.4). From the consensus data, one conclusion was eight N-glycan compositions showed significant changes in average abundances between “PSA” and “PSA high form” when applied a statistical significance test named Wilcoxon signed-

rank test. Specifically, four compositions ( $\text{Hex}_5\text{HexNAc}_4\text{dHex}_1\text{NeuAc}_2$ ,  $\text{Hex}_5\text{HexNAc}_4\text{dHex}_1\text{NeuAc}_1$ ,  $\text{Hex}_4\text{HexNAc}_5\text{dHex}_1\text{NeuAc}_2$ ,  $\text{Hex}_5\text{HexNAc}_4\text{dHex}_1$ ) were increased in “PSA high form” compared to “PSA”, while four glycans ( $\text{Hex}_4\text{HexNAc}_4\text{dHex}_1\text{NeuAc}_1$ ,  $\text{Hex}_4\text{HexNAc}_3\text{dHex}_1\text{NeuAc}_1$ ,  $\text{Hex}_6\text{HexNAc}_3\text{NeuAc}_1$ ,  $\text{Hex}_4\text{HexNAc}_3\text{NeuAc}_1$ ) had higher abundances in “PSA” than “PSA high form”. The increase and decrease of those eight glycans in our data were the same as in the consensus data.



**Figure 2.5. Differential profiles of seven major N-glycans derived from the consensus data of 15 labs (A) and our data (B).** Top blue color columns represented relative percentages of glycans in “PSA”. Bottom red colored columns represented relative abundances of glycoforms in “PSA high isoform”.

The other conclusion was “PSA” had fewer percentages of disialylated and fucosylated glycans, more monosialylated glycans, and same abundance of nonsialylated glycans comparing to “PSA high isoform”. Table 2.4 and 2.5 showed the relative abundances of different types of glycans according to the number of sialic acids and the presence of fucose from average of the consensus data and our data. In Table 2.4, the percentages were based on calculation on 18 major and intermediate glycans, while in Table 2.5 only eight significant changed N-glycans were considered. These results indicated that our final conclusions were the same as the ones drawn from the consensus data.

**Table 2.4.** Percentages of different types of N-glycans in major and intermediate glycans

N-glycans	Intensities of major and intermediate glycans			
	Average in consensus data		Our data	
	PSA	PSA high isoform	PSA	PSA high isoform
Unisialylated	4.4%	4.5%	5.9%	6.5%
Monosialylated	71.2%	43.5%	72.0%	44.6%
Disialylated	8.3%	37.9%	13.1%	41.8%
Fucosylated	53.4%	72.9%	74.1%	88.5%

**Table 2.5.** Percentages of different types of N-glycans in significantly changed glycans

N-glycans	Intensities of eight significantly changed glycans			
	Average in consensus data		Our data	
	PSA	PSA high isoform	PSA	PSA high isoform
Unisialylated	0.3%	2.8%	0.8%	2.7%
Monosialylated	57.4%	32.5%	61.4%	40.1%
Disialylated	5.7%	33.0%	10.9%	40.0%
Fucosylated	45.9%	67.1%	60.3%	80.1%



*Other points from the interlaboratory study*

Based on the results from this interlaboratory study, the consistency of results among the methods ranked as follows: top-down (4 datasets)  $\geq$  glycan analysis (5 datasets) > bottom-up (17 datasets). However, a final conclusion could not be drawn because of limited numbers of labs using top down and glycan approaches. The better consistency of results from top-down method might be due to less sample handling steps which introduced less error, but it also might be because that labs performed top-down were usually skilled groups and with dedicated instruments. Groups applied glycan analysis might routinely perform glycan analysis and hence have more experience than bottom-up labs. The successful analysis by top-down and glycan approaches in this study was largely dependent on selection of the sample. If the protein was large and/or had multiple glycosylation sites, it would be too complicated to perform top-down analysis. If the PSA sample was contaminated with other glycoproteins or it included multiple glycosylation sites, glycan analysis could not be used on quantification because it could not discriminate the amino acid site attaching glycans. To analyze complex samples, bottom-up method was convenient and applicable. This study proved that bottom-up was a good approach since results from most of labs quantifying on glycopeptides were consistent with the ones on glycans and glycoprotein in the consensus data. Labs which had obvious analytical problems were the ones that were not very skilled in glycoproteomics field. In addition, by using the same enzyme for digestion and choosing the same peptide for quantification, the reproducibility in different labs was expected to be improved in bottom-up approach.

In this interlaboratory study, 18 out of 24 labs manually interpreted their mass spectrometry data, indicating the great needs of informatics tools in glycoproteomics field. Specifically, in labs using bottom-up approach, 14 out of 18 labs interpreted data manually and only 4 labs used the following softwares: FindPept [71], Glycopep [72], SimGlycan [73], and GlypID [74]. SimGlycan and GlypID were used to assign MS/MS fragments. FindPept identified peptides resulting from unspecific cleavage. Glycopep compared measured masses to calculated masses. Concluded from our experience, a very time-consuming part of data interpretation was manually assigning MS peaks to the mass-to-charge ratios of calculated glycopeptides, so software Glycopep could be helpful in glycopeptide identification.

### **2.3.3 Accomplishments**

In this study, we first validated the application of HILIC material PolyHEA on quantitative purification of glycopeptides, and then utilized a bottom-up approach to quantify glycol-isoforms in glycoprotein PSA, which included the following steps: Protein digestion into glycopeptides, glycopeptides purification by PolyHEA SPE, and analysis of glycopeptides by RPLC-MS/MS. We completed the comparison of glycan compositions from two different sources of PSA, “PSA” and “PSA high isoform”, and furnished a list including 28 glycans. Eight N-glycan compositions showed significant changes between PSA and PSA high isoform. PSA high isoform had more disialylated, fucosylated, less monosialylated, and same asialylated glycans compared to PSA. By comparing the results from other labs in a multi-institute study, we confirmed that our

peptide-centric approach of identifying and quantifying different glycoforms in a specific protein provided robust and accurate results.

## 2.4 REFERENCES

1. Lechner, J. and F. Wieland, *Structure and Biosynthesis of Prokaryotic Glycoproteins*. Annual Review of Biochemistry, 1989. **58**: p. 173-194.
2. Messner, P., *Bacterial glycoproteins*. Glycoconjugate Journal, 1997. **14**(1): p. 3-11.
3. Wormald, M.R., et al., *Conformational studies of oligosaccharides and glycopeptides: Complementarity of NMR, X-ray crystallography, and molecular modelling*. Chemical Reviews, 2002. **102**(2): p. 371-386.
4. Rudd, P.M. and R.A. Dwek, *Glycosylation: Heterogeneity and the 3D structure of proteins*. Critical Reviews in Biochemistry and Molecular Biology, 1997. **32**(1): p. 1-100.
5. Debeer, T., et al., *The Hexopyranosyl Residue That Is C-Glycosidically Linked to the Side-Chain of Tryptophan-7 in Human Rnase U-S Is Alpha-Marmopyranose*. Biochemistry, 1995. **34**(37): p. 11785-11789.
6. Gemmill, T.R. and R.B. Trimble, *Overview of N- and O-linked oligosaccharide structures found in various yeast species*. Biochimica Et Biophysica Acta-General Subjects, 1999. **1426**(2): p. 227-237.
7. Furmanek, A. and J. Hofsteenge, *Protein C-mannosylation: facts and questions*. Acta Biochim Pol, 2000. **47**(3): p. 781-9.
8. Kobayashi, T., R. Nishizaki, and H. Ikezawa, *The presence of GPI-linked protein(s) in an archaeobacterium, Sulfolobus acidocaldarius, closely related to eukaryotes*. Biochimica Et Biophysica Acta-General Subjects, 1997. **1334**(1): p. 1-4.
9. Mehta, D.P., et al., *A lysosomal cysteine proteinase from Dictyostelium discoideum contains N-acetylglucosamine-1-phosphate bound to serine but not mannose-6-phosphate on N-linked oligosaccharides*. Journal of Biological Chemistry, 1996. **271**(18): p. 10897-10903.

10. Haynes, P.A., *Phosphoglycosylation: a new structural class of glycosylation?* *Glycobiology*, 1998. **8**(1): p. 1-5.
11. An, H.J., J.W. Froehlich, and C.B. Lebrilla, *Determination of glycosylation sites and site-specific heterogeneity in glycoproteins*. *Current Opinion in Chemical Biology*, 2009. **13**(4): p. 421-426.
12. Burda, P. and M. Aebi, *The dolichol pathway of N-linked glycosylation*. *Biochimica Et Biophysica Acta-General Subjects*, 1999. **1426**(2): p. 239-257.
13. Trombetta, E.S., *The contribution of N-glycans and their processing in the endoplasmic reticulum to glycoprotein biosynthesis*. *Glycobiology*, 2003. **13**(9): p. 77r-91r.
14. Berninsone, P.M., *Carbohydrates and glycosylation*. *WormBook*, 2006: p. 1-22.
15. Jemal, A., et al., *Global cancer statistics*. *CA Cancer J Clin*, 2011. **61**(2): p. 69-90.
16. SEER Program (National Cancer Institute (U.S.)), et al., *SEER cancer statistics review*, in *NIH publication*. 1993, U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, National Cancer Institute: Bethesda, Md. p. v.
17. Balk, S.P., Y.J. Ko, and G.J. Bubley, *Biology of prostate-specific antigen*. *J Clin Oncol*, 2003. **21**(2): p. 383-91.
18. Catalona, W.J., et al., *Selection of optimal prostate specific antigen cutoffs for early detection of prostate cancer: receiver operating characteristic curves*. *J Urol*, 1994. **152**(6 Pt 1): p. 2037-42.
19. Catalona, W.J., et al., *Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men*. *J Urol*, 1994. **151**(5): p. 1283-90.
20. Moyer, V.A., *Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement*. *Ann Intern Med*, 2012. **157**(2): p. 120-34.

21. Chou, R., et al., *Screening for prostate cancer: a review of the evidence for the U.S. Preventive Services Task Force*. *Ann Intern Med*, 2011. **155**(11): p. 762-71.
22. Sokoll, L.J., et al., *A prospective, multicenter, National Cancer Institute Early Detection Research Network study of [-2]proPSA: improving prostate cancer detection and correlating with cancer aggressiveness*. *Cancer Epidemiol Biomarkers Prev*, 2010. **19**(5): p. 1193-200.
23. Sarrats, A., et al., *Differential percentage of serum prostate-specific antigen subforms suggests a new way to improve prostate cancer diagnosis*. *Prostate*, 2010. **70**(1): p. 1-9.
24. Okada, T., et al., *Structural characteristics of the N-glycans of two isoforms of prostate-specific antigens purified from human seminal fluid*. *Biochim Biophys Acta*, 2001. **1525**(1-2): p. 149-60.
25. Ohyama, C., et al., *Carbohydrate structure and differential binding of prostate specific antigen to Maackia amurensis lectin between prostate cancer and benign prostate hypertrophy*. *Glycobiology*, 2004. **14**(8): p. 671-679.
26. Peracaula, R., et al., *Altered glycosylation pattern allows the distinction between prostate-specific antigen (PSA) from normal and tumor origins*. *Glycobiology*, 2003. **13**(6): p. 457-470.
27. Tajiri, M., C. Ohyama, and Y. Wada, *Oligosaccharide profiles of the prostate specific antigen in free and complexed forms from the prostate cancer patient serum and in seminal plasma: A glycopeptide approach*. *Glycobiology*, 2008. **18**(1): p. 2-8.
28. Sarrats, A., et al., *Glycan Characterization of PSA 2-DE Subforms from Serum and Seminal Plasma*. *Omics-a Journal of Integrative Biology*, 2010. **14**(4): p. 465-474.
29. Pan, S., et al., *Mass spectrometry based glycoproteomics--from a proteomics perspective*. *Mol Cell Proteomics*, 2011. **10**(1): p. R110 003251.
30. Mechref, Y. and M.V. Novotny, *Structural investigations of glycoconjugates at high sensitivity*. *Chemical Reviews*, 2002. **102**(2): p. 321-369.

31. Zhao, J., et al., *N-linked glycosylation profiling of pancreatic cancer serum using capillary liquid phase separation coupled with mass spectrometric analysis*. Journal of Proteome Research, 2007. **6**(3): p. 1126-1138.
32. Wuhrer, M., et al., *Glycoproteomics based on tandem mass spectrometry of glycopeptides*. J Chromatogr B Analyt Technol Biomed Life Sci, 2007. **849**(1-2): p. 115-28.
33. Dalpathado, D.S. and H. Desaire, *Glycopeptide analysis by mass spectrometry*. Analyst, 2008. **133**(6): p. 731-738.
34. Rebecchi, K.R., et al., *Label-Free Quantitation: A New Glycoproteomics Approach*. Journal of the American Society for Mass Spectrometry, 2009. **20**(6): p. 1048-1059.
35. Thaysen-Andersen, M., S. Mysling, and P. Hojrup, *Site-Specific Glycoprofiling of N-Linked Glycopeptides Using MALDI-TOF MS: Strong Correlation between Signal Strength and Glycoform Quantities*. Analytical Chemistry, 2009. **81**(10): p. 3933-3943.
36. Wuhrer, M., et al., *Glycosylation profiling of immunoglobulin G (IgG) subclasses from human serum*. Proteomics, 2007. **7**(22): p. 4070-81.
37. Chang, W.C., et al., *Matrix-assisted laser desorption/ionization (MALDI) mechanism revisited*. Anal Chim Acta, 2007. **582**(1): p. 1-9.
38. Huberty, M.C., et al., *Site-Specific Carbohydrate Identification in Recombinant Proteins Using Mald-Tof Ms*. Analytical Chemistry, 1993. **65**(20): p. 2791-2800.
39. Sekiya, S., Y. Wada, and K. Tanaka, *Derivatization for stabilizing sialic acids in MALDI-MS*. Analytical Chemistry, 2005. **77**(15): p. 4962-4968.
40. Papac, D.I., A. Wong, and A.J.S. Jones, *Analysis of acidic oligosaccharides and glycopeptides by matrix assisted laser desorption ionization time-of-flight mass spectrometry*. Analytical Chemistry, 1996. **68**(18): p. 3215-3223.
41. Harvey, D.J., *Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates*. Mass Spectrom Rev, 1999. **18**(6): p. 349-450.

42. Lattova, E., et al., *Matrix-assisted laser desorption/ionization on-target method for the investigation of oligosaccharides and glycosylation sites in glycopeptides and glycoproteins*. Rapid Commun Mass Spectrom, 2007. **21**(10): p. 1644-50.
43. Tajiri, M., T. Takeuchi, and Y. Wada, *Distinct features of matrix-assisted 6 microm infrared laser desorption/ionization mass spectrometry in biomolecular analysis*. Anal Chem, 2009. **81**(16): p. 6750-5.
44. Scott, N.E., et al., *Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of Campylobacter jejuni*. Mol Cell Proteomics, 2011. **10**(2): p. M000031-MCP201.
45. Zhang, H., et al., *Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry*. Nature Biotechnology, 2003. **21**(6): p. 660-666.
46. Atwood, J.A., 3rd, et al., *Quantitation by isobaric labeling: applications to glycomics*. J Proteome Res, 2008. **7**(1): p. 367-74.
47. Alvarez-Manilla, G., et al., *Tools for glycomics: relative quantitation of glycans by isotopic permethylation using <sup>13</sup>CH<sub>3</sub>I*. Glycobiology, 2007. **17**(7): p. 677-87.
48. Old, W.M., et al., *Comparison of label-free methods for quantifying human proteins by shotgun proteomics*. Molecular & Cellular Proteomics, 2005. **4**(10): p. 1487-1502.
49. Callister, S.J., et al., *Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics*. Journal of Proteome Research, 2006. **5**(2): p. 277-286.
50. Ono, M., et al., *Label-free quantitative proteomics using large peptide data sets generated by nanoflow liquid chromatography and mass spectrometry*. Molecular & Cellular Proteomics, 2006. **5**(7): p. 1338-1347.
51. Zauner, G., A.M. Deelder, and M. Wührer, *Recent advances in hydrophilic interaction liquid chromatography (HILIC) for structural glycomics*. Electrophoresis, 2011. **32**(24): p. 3456-66.



52. Zhang, T., et al., *Effective analysis of branched glycopeptides using porous graphite chromatography*. *The column*, 2011. **7**(8).
53. Bunkenborg, J., et al., *Screening for N-glycosylated proteins by liquid chromatography mass spectrometry*. *Proteomics*, 2004. **4**(2): p. 454-465.
54. Yang, Z. and W.S. Hancock, *Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column*. *J Chromatogr A*, 2004. **1053**(1-2): p. 79-88.
55. Wang, Y., S.L. Wu, and W.S. Hancock, *Approaches to the study of N-linked glycoproteins in human plasma using lectin affinity chromatography and nano-HPLC coupled to electrospray linear ion trap--Fourier transform mass spectrometry*. *Glycobiology*, 2006. **16**(6): p. 514-23.
56. Ding, W., et al., *Identification and Quantification of Glycoproteins Using Ion-Pairing Normal-phase Liquid Chromatography and Mass Spectrometry*. *Molecular & Cellular Proteomics*, 2009. **8**(9): p. 2170-2185.
57. Liu, T., et al., *Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry*. *J Proteome Res*, 2005. **4**(6): p. 2070-80.
58. Lewandrowski, U., et al., *Elucidation of N-glycosylation sites on human platelet proteins: a glycoproteomic approach*. *Mol Cell Proteomics*, 2006. **5**(2): p. 226-33.
59. Alvarez-Manilla, G., et al., *Tools for glycoproteomic analysis: Size exclusion chromatography facilitates identification of tryptic glycopeptides with N-linked glycosylation sites*. *Journal of Proteome Research*, 2006. **5**(3): p. 701-708.
60. Sparbier, K., et al., *Selective isolation of glycoproteins and glycopeptides for MALDI-TOF MS detection supported by magnetic particles*. *J Biomol Tech*, 2005. **16**(4): p. 407-13.
61. Alpert, A.J., *Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds*. *J Chromatogr*, 1990. **499**: p. 177-96.

62. Parker, B.L., et al., *Quantitative N-linked glycoproteomics of myocardial ischemia and reperfusion injury reveals early remodeling in the extracellular environment*. Mol Cell Proteomics, 2011. **10**(8): p. M110 006833.
63. Wada, Y., M. Tajiri, and S. Yoshida, *Hydrophilic affinity isolation and MALDI multiple-stage tandem mass spectrometry of glycopeptides for glycoproteomics*. Anal Chem, 2004. **76**(22): p. 6560-5.
64. Selman, M.H., et al., *Cotton HILIC SPE microtips for microscale purification and enrichment of glycans and glycopeptides*. Anal Chem, 2011. **83**(7): p. 2492-9.
65. Jeong, H.J., et al., *High-throughput quantitative analysis of total N-glycans by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry*. Anal Chem, 2012. **84**(7): p. 3453-60.
66. Mattsson, J.M., et al., *Structural characterization and anti-angiogenic properties of prostate-specific antigen isoforms in seminal fluid*. Prostate, 2008. **68**(9): p. 945-954.
67. Saldoval, R., et al., *Core fucosylation and alpha 2-3 sialylation in serum N-glycome is significantly increased in prostate cancer comparing to benign prostate hyperplasia*. Glycobiology, 2011. **21**(2): p. 195-205.
68. Tabares, G., et al., *Different glycan structures in prostate-specific antigen from prostate cancer sera in relation to seminal plasma PSA*. Glycobiology, 2006. **16**(2): p. 132-45.
69. White, K.Y., et al., *Glycomic characterization of prostate-specific antigen and prostatic acid phosphatase in prostate cancer and benign disease seminal plasma fluids*. J Proteome Res, 2009. **8**(2): p. 620-30.
70. Mysling, S., et al., *Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics*. Anal Chem, 2010. **82**(13): p. 5598-609.
71. Gattiker, A., et al., *FindPept, a tool to identify unmatched masses in peptide mass fingerprinting protein identification*. Proteomics, 2002. **2**(10): p. 1435-44.

72. Woodin, C.L., et al., *GlycoPep grader: a web-based utility for assigning the composition of N-linked glycopeptides*. *Anal Chem*, 2012. **84**(11): p. 4821-9.
73. Apte, A. and N.S. Meitei, *Bioinformatics in glycomics: glycan characterization with mass spectrometric data using SimGlycan*. *Methods Mol Biol*, 2010. **600**: p. 269-81.
74. Mayampurath, A.M., et al., *Improving confidence in detection and characterization of protein N-glycosylation sites and microheterogeneity*. *Rapid Commun Mass Spectrom*, 2011. **25**(14): p. 2007-19.