METHOD DEVELOPMENT FOR PROTEIN IDENTIFICATION WITH MALDI-TOF/TOF BY USING ON-SURFACE DIGESTION

A Thesis Submitted to
the Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in Chemistry

by Melike DİNÇ

December 2012
İZMİR

We approve the thesis of **Melike DİNÇ**

Examining Committee Members

Prof. Dr. Talat YALÇIN
Department of Chemistry, Izmir Institute of Technology

Assist. Prof. Dr. Gülşah ŞANLI-MOHAMED
Department of Chemistry, Izmir Institute of Technology

Assoc. Prof. Dr. Ahmet KOÇ
Department of Molecular Biology and Genetic, Izmir Institute of Technology

10 December 2012

Prof. Dr. Talat YALÇIN
Supervisor, Department of Chemistry
Izmir Institute of Technology

Prof. Dr. Durmuş ÖZDEMİR
Head of the Department of Chemistry

Prof. Dr. R. Tuğrul SENGER
Dean of the Graduate School of
Engineering and Sciences

ACKNOWLEDGEMENTS

Completion of this dissertation was possible with the support of several people. I would like to express my sincere gratitude to all of them.

First and foremost, I would like to thank to my supervisor Prof. Dr. Talat YALÇIN for his continuous guidance, understanding and encouragement throughout the course of this thesis. I always consider myself very lucky to have such a good mentor by whom I am inspired and motivated in the scientific field with his academic success and life experiences. I am extremely grateful to him for giving me the chance of working together.

I would also like to thank to my committee members Assoc. Prof. Dr. Ahmet KOÇ and Assist. Prof. Dr. Gülşah ŞANLI-MOHAMED, for their suggestions and contributions.

I am also thankful to members of Molecular Genetic Laboratory in the Molecular Biology and Genetics department for their helps and kindness.

My special thanks go to my dear friends and co-workers Melda Z. GÜRAY, Ahmet E. ATİK, Çağdaş TAŞOĞLU and Dr. Filiz YEŞİLIRMAK for their sincere, helps, friendships and moral supports. It has been a pleasure working with you.

Last but not least, I would like to send my deep appreciation to my family. I owe a lot to my parents and my sister, who encouraged and supported me at every stage of my personal and academic life, and longed to see this achievement come true.

Finally, I wish to acknowledge the TUBITAK funding (Project No. 109T430) which was granted during my graduate education.

ABSTRACT

METHOD DEVELOPMENT FOR PROTEIN IDENTIFICATION WITH MALDI-TOF/TOF BY USING ON-SURFACE DIGESTION

Protein identification is predominantly carried out by searching tandem mass spectrometric data of peptides in a protein database. For this reason, proteins are converted to peptides through a digestion process by using some certain endoproteinases. Trypsin is mostly preferred in this sample preparation step due to its high activity and products having appropriate mass range. Whereas in-solution digestion method is applied for the proteins in solution, proteins trapped in the gel can be digested by using in-gel digestion technique. Alternative to these traditional digestion methods, it has been reported that proteins can be digested too while they were adsorbed onto solid surfaces.

In this study, digestion process of the adsorbed proteins, namely on-surface digestion is examined widely by using both hydrophobic and ionic adsorbents on different proteins. Results of the on-surface digestion were compared with in-solution digestion and in-gel digestion methods. As a conclusion, on-surface digestion is applicable for the protein identification by mass spectrometry; however, its yield may change from one experiment to another, depending on two separate but related processes: protein adsorption before the digestion and peptide recovery after the digestion. Nevertheless on-surface digestion has the advantages of protein enrichment and protein purification prior to mass spectrometry. These processes are necessary and significant especially for the samples containing minute amounts of protein and an effective enzymatic activity. Last but not least, this method may be performed complementarily to other digestion methods since new and different peptides may be acquired from the same sample source.

ÖZET

MALDI-TOF/TOF İLE PROTEİN TANIMLAMASI İÇİN YÜZEYDE PARÇALAMA YÖNTEMİNİ KULLANARAK METOT GELİŞTİRME

Protein tanımlaması ağırlıklı olarak, peptitlerin sıralı kütle spektrometrik verisinin bir protein veri tabanında taranmasıyla gerçekleştirilir. Bu sebeple, proteinler bir parçalama işlemi üzerinden belli bazı endoproteinazlar kullanılarak peptitlere dönüştürülürler. Bu örnek hazırlama basamağında, yüksek aktivitesi ve uygun kütle aralığına sahip ürünlerinden dolayı çoğunlukla tripsin tercih edilir. Solüsyondaki proteinler için solüsyon-içinde parçalama metodu uygulanırken, jele hapsolmuş proteinler jel-içinde parçalama tekniği ile parçalanabilir. Bu geleneksel parçalama metotlarına alternatif olarak proteinlerin, katı yüzeyler üzerine tutturulmuşken de parçalanabildiği gösterilmiştir.

Bu çalışmada, yüzeyde parçalama olarak adlandırılan adsorplanmış proteinlerin parçalanma işlemi, hidrofobik ve iyonik adsorbentler kullanılarak farklı proteinler üzerinde kapsamlı olarak incelenmiştir. Ayrıca yüzeyde parçalama sonuçları solüsyon-içinde ve jel-içinde parçalama metodları ile karşılaştırılmıştır. Sonuç olarak yüzeyde parçalama metodu kütle spektrometresi ile protein tanımlamasında uygulanabilir ancak verimi, parçalama öncesi protein tutturulması ve parçalama sonrası peptit geri kazanımı şeklinde ayrı fakat birbiriyle ilişkili iki sürece bağlı olarak bir denemeden diğerine değişiklik gösterebilir. Buna rağmen, yüzeyde parçalama kütle spektrometresi öncesi protein zenginleştirme ve saflaştırma avantajlarına sahiptir. Bu işlemler özellikle çok az miktarda protein içeren örnekler ve etkili bir enzimatik aktivite için gerekli ve mühimdir. Son ve bir o kadar önemli olarak, aynı örnek kaynağından yeni ve farklı peptit sinyalleri elde edilebildiğinden, bu metot diğer parçalama metotlarına tamamlayıcı olarak uygulanabilir.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

Table **Page**

LIST OF ABBREVIATIONS

2D-GE	Two Dimensional-Gel Electrophoresis
ABC	Ammonium Bicarbonate
BSA	Bovine Serum Albumin
CHCA	α-Cyano-4-Hydroxycinnamic Acid
CID	Collision-Induced Dissociation
ECD	Electron Capture Dissociation
ESI	Electrospray Ionization
ETD	Electron Transfer Dissociation
DHB	2,5-Dihydroxybenzoic Acid
DTT	1,4-Dithiothreitol
FT-ICR	Fourier Transform-Ion Cyclotron Resonance
HPLC	High-Performance Liquid Chromatography
IAM	Iodoacetamide
IPG	Immobilized pH Gradient
IT	Ion Trap
LC	Liquid Chromatography
LC/MS	Liquid Chromatography/Mass Spectrometry
MALDI	Matrix-Assisted Laser Desorption Ionization
MS	Mass Spectrometry
m/z	Mass-to-Charge
MS/MS	Tandem Mass Spectrometry
PMF	Peptide Mass Fingerprinting
PTM	Post-Translational Modification
rf	Radio Frequency
RP	Reverse Phase
SA	Sinapinic Acid; 3,5-Dimethoxy-4-hydroxycinnamic acid
SAX	Strong Anion Exchange
SCX	Strong Cation Exchange
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
% SC	% Sequence Coverage

CHAPTER 1

INTRODUCTION TO MASS SPECTROMETRY-BASED PROTEOMICS

1.1. Introduction to Proteomics

Since proteins govern the function in the cells or body fluids, towards the end of Human Genome Project it has been realized that a comprehensive understanding of biological activities can be attained through proteins rather than genes[1]. In addition, studies on the correlation of mRNA to protein have proved that mRNA is insufficient to predict the expressions of all the proteins[2]. Thus, post-genomic era started with an extensive interest in the direct analysis of proteins. However, proteins have extremely dynamic nature and complex structure. Moreover, they carry out their function through interactions with other proteins and molecules. Proteomics, the global scale analysis of proteins has enabled scientist to study complex protein mixtures without the need of complete amino acid sequence of a protein. In conclusion, proteomics, which was coined in analogy to genomics by Marc Wilkins in the early 1990s, has been widely adopted by the biological community in a short time[3].

Numerous functional diversity of proteins arises from the linear arrangement of specific twenty amino acids in different composition. Once the primary structure is formed as an amino acid chain, local conformation of the peptide sequence generates the secondary structure. After that, prevailing interactions between stabilizing forces such as hydrophobic effects and hydrogen bonds, lead protein to fold creating the tertiary structure. Together with protein folding, diversifications on the polypeptide chain, which occurs after the translation by covalent modifications (post-translational modifications, PTM) play a key role in several biological processes[4]. As the links between proteins and the metabolic pathways have been uncovered, scientists started to discuss the problem of protein complexity systematically by drawing protein interaction maps and protein networks via wires and nodes[5] (Figure 1.1. c). Proteins tend to respond to changing stimuli, therefore, some certain predeterminations as the environmental condition, status of the protein resource and the methodology used

should be thoroughly stated according to the purpose of the research. At this point, the original scope of proteome, which is defined as *entire protein complement expressed by a genome or by a cell or tissue type*[6], may be narrowed defining the material or working area like plant proteomics, structural proteomics, targeted proteomics and so forth.

A typical proteomic analysis primarily starts with expression proteomics in which the sample preparation, analyte detection and monitoring are performed. Next, the bioinformatic analysis takes over the task and provides information about the given protein. Ultimate and most importantly, in functional proteomics, role of the targeted protein is tried to find in the biological sense[7]. Following the innovations in ionization techniques, analytes of biological macromolecules, which are relatively large and fragile to ionize, have become eligible to be measured by mass spectrometers. In time, mass spectrometry (MS) became a method of choice for proteomics, and proteomics researches involving mass spectrometer is named mass spectrometry-based proteomics.

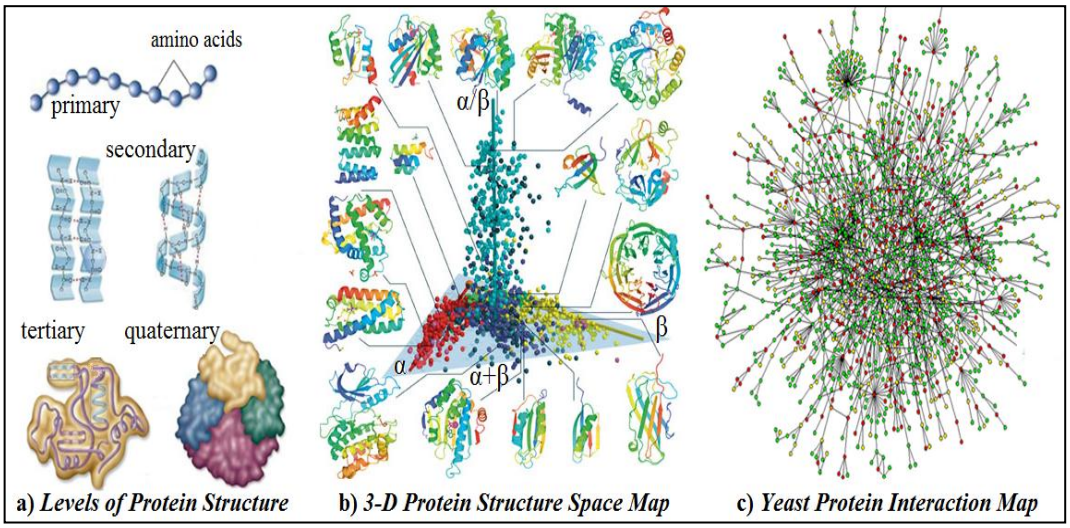


Figure 1.1. Complexity of proteins [8]

Unlike the traditional protein sequencing method Edman degradation[9], in mass spectrometry-based proteomics, it is not anticipated to have the complete sequence of the protein of interest because cell events associated with proteins can be described without 100% sequence coverage[10]. Although the whole proteome analysis of some model organisms, especially the proteomes of the yeast *Saccharomyces cerevisiae* and well-known bacteria *Escherichia coli*, have been attained to some extent, in the case of proteome analysis of rather complex organisms, this attempt get more formidable.

As an example, the number of proteins cataloged (over thousands) is estimated to correspond to half of the predicted genome roughly in the reports of the well-studied organisms; the nematode *Caenorhabditis elegans* and the fruit fly *Drosophila melanogaster* [11]. When it comes to human proteome, the total number of proteins is expected to reach millions in the recently established Human Proteome Project[12].

Abilities of mass spectrometry are not limited to protein identification, characterization of PTMs is also being conducted by mass spectrometry for a long time[13]. In addition to this, discrepancy between normal organism and pathological one, can be revealed by comparing their proteomes. Appearance of new protein fragments, which is regarded as the signature of the disease, appears as a result of protein breakdown, modification process, change in protein concentration beyond the standard deviation or occurrence of protein aggregation. So far, many biological connections and physiological processes, which used to be thought unrelated have been enlightened by mass spectrometry [14]. In time, applications of proteomics have been developed from bench-to-bedside investigations, especially in the field of drug discovery[15] and biomarker design[16].

Last but not least, protein quantities reflect dynamics of biological system and responses to a changing environment. Protein quantification can be carried out by mass spectrometry using one of two different ways: differential stable isotope-labelled and label-free. In label-free methods, either signal intensity of the peptides or number of acquired spectra are used. On the other hand, labelled methods are performed considering the mass shift between *heavy* and *light* peptides, which is produced by adding stable isotopes to the sample beforehand.[17].

Unfortunately in mass spectrometry-based proteomics, there is no single method addressing to all problems, therefore, scientists have to select the most appropriate methods and the most advanced techniques for specific aims and problems.

1.2. Mass Spectrometry

Mass spectrometers separate the charged analytes according to their mass-to-charge (m/z) ratios. Mass spectrometry is the name of this analytical technique in which the ions are detected in proportion to their abundances. A typical mass spectrometer is comprised of three main parts: ion source, mass analyzer and detector.

In order to mobilize and manipulate the molecules under the impact of electric or magnetic field, analytes are first ionized in the ion source. Since internal energy is transferred during this process, physicochemical properties of the ionic compound are of considerable importance. Conventional energetic ionization techniques, which cause extensive fragmentation, were not suitable for large, nonvolatile and thermally unstable species; proteins, oligosaccharides, oligonucleotides etc. For this reason, new techniques were necessary to introduce biological macromolecules to the system by extracting them directly from condensed form to gas phase without degradation. Matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) have been undertaking this mission since late 1980s[18] (Figure 1.2.).

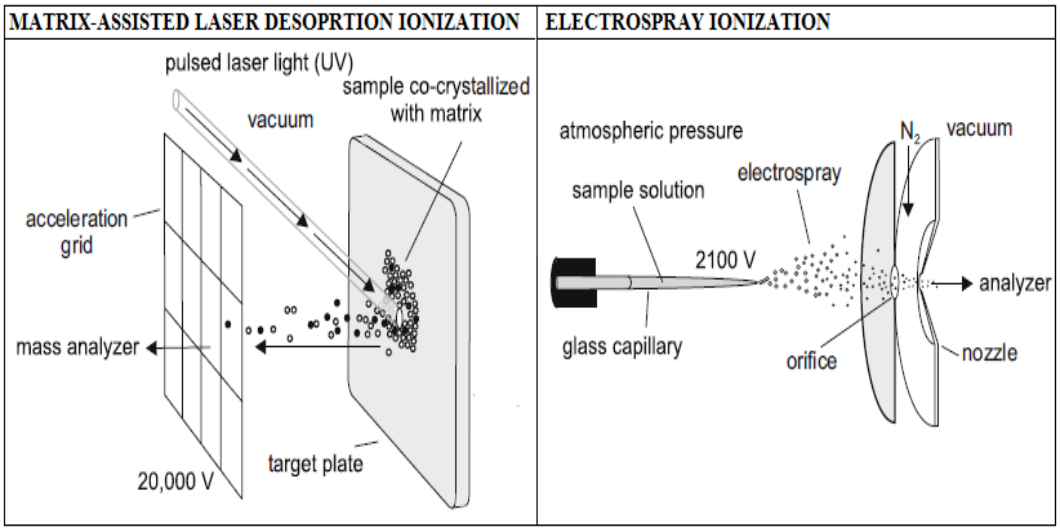


Figure 1.2. Soft ionization techniques [19]

In MALDI, both desorption and ionization processes are carried out by matrix compounds, which are mostly chosen from ultraviolet absorbing organic molecules. Matrix also serves as a proton donor or acceptor according to the positive or negative ionization mode while minimizing the sample damage from the laser energy for a nondestructive vaporization. Furthermore, matrix-solvent composition and sample-matrix preparation have considerable importance on the quality of spectrum[20]. Therefore, both the choice of matrix and preparation procedure are critical for a favorable MALDI analyses.

Lack of chemical reactivity, solubility in different solvents, ability to promote analyte ionization, strong absorbance at the laser wavelength and low mass can be listed the features of a good MALDI matrix. Although there is no universal sample

preparation method for peptides and proteins, α-cyano-4-hydroxycinnamic acid (CHCA), 2,5- dihydroxybenzoic acid (DHB), 3,5-dimethoxy-4-hydroxycinnamic acid (Sinapinic acid, SA) are the most common matrices in use. Sample-matrix mixture is mostly deposited onto target either *dried-droplet* or *thin layer* methods. In both methods, analytes are embedded throughout excess quantity of matrix and then left to dry to co-crystallization with matrix[18].

Ionization in bundles through an intermittent process rather than a continuous ion beam makes this technique more appropriate to time-of-flight (TOF) analyzers. Today, MALDI-TOF is an essential and widespread technique due to its favorable features; singly charged ion acquisition, high mass range, user-friendliness and partial tolerance to contaminants. Although a number of pathways including ion-molecule reaction, excited-state proton transfer, thermal ionization, energy pooling, disproportionation were proposed, there is no single mechanism which can precisely explain the ionization process in MALDI. Considering the broad analyte range (from biological macromolecules to synthetic polymers, synthetically prepared dendrimers and fullerenes) and different sample preparation methods, it is not surprising why one definite mechanism addressing to all these analytes, could not be found[21].

Whereas MALDI introduces the solid phase analytes to the analyzer; other soft ionization technique ESI produces ions from bulk solution via capillary tube. This structure of ESI enables scientist to combine mass spectrometry with liquid chromatography. In this LC-MS/MS system, analytes can be separated, purified and enriched at the same time prior to mass analysis. Under atmospheric pressure, a high voltage about 2-6 kV is applied to the tip of the capillary creating highly charged droplets. To direct the spray towards the mass analyzer and to disperse spray for better nebulization, a gas, dry N_2 is injected coaxially. Solvent evaporation occurs when the charged droplets pass through either the heated inert gas or heated capillary at high vacuum. Nature of the solvent, flow rate, size of the capillary, potential applied, surface tension, nature of the analyte and electrolytes are the factors which affect the electrochemical process and the ion formation in ESI. For the explanation of ionization process in electrospray, two mechanisms are proposed: charged residue model and ion evaporation model. Whereas ion evaporation model suggests that removal of charge is replaced by ion evaporation mostly as Coulomb fission at larger droplet radii, charge residue model explains the phenomenon through jet-fission or evaporation cycles, at the end leaving a ‘residue’ of charge on the analyte[22]. In contrast to MALDI, multiply

charged ions are mostly occurred in ESI. Later, they are reduced to reach the monoisotopic mass by a mathematical charge deconvolution method. This multiply charge formation gives better results in ECD/ETD fragmentations and provides further analyses for protein three-dimensional structure and noncovalent interaction.

Mass analyzer is the part in which ions are trapped or transmitted according to their m/z values. Today several different mass analyzers and their sequentially positioned same or hybrid assemblies are being utilized in proteomics especially for their unique properties: resolution, mass range, sensitivity, ion transmission, dynamic range and analysis speed (Figure 1.3.).

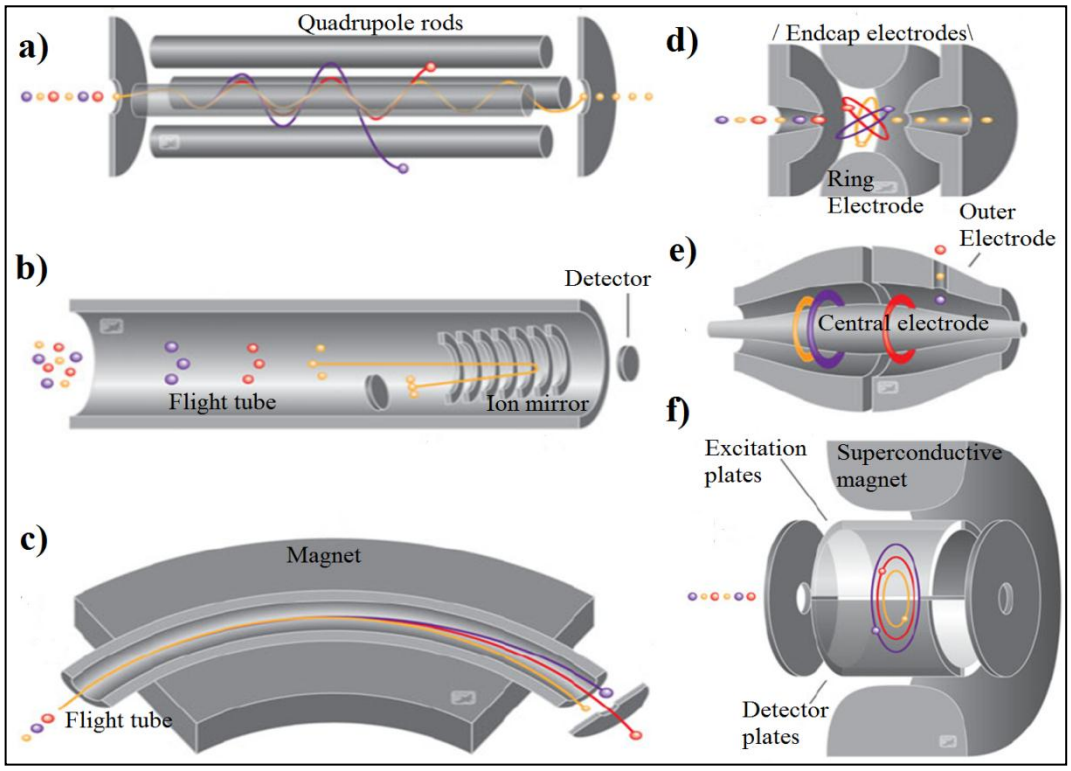


Figure 1.3. Mass analyzers [23] a) Quadrupole b) TOF/TOF c) Magnetic sector instrument d) Quadrupole ion trap e) Orbitrap f) FT-ICR

Electric or magnetic fields are used in mass analyzers by using different ion manipulation principles. For example, while ions are separated according to their flight time in a time-of-flight (TOF) analyzers, ion stability is used for quadrupole analyzers and resonance frequency of a m/z value is utilized for trapping the ions in an ion trap, orbitrap and cyclotron resonance mass analyzers[24].

One of the earliest analyzer, sector instruments functions in scanning mode by focusing the ions to a magnetic and electrostatic field. Accelerated ions passing through

a magnetic sector are deflected to a circular motion of a unique radius. Angle and radius of the deflection can be derived from magnetic field strength, accelerating voltage and m/z. In double focusing magnetic sectors, ions are first focused according to their kinetic energy by an electrostatic sector and then separated according to their momentum by a magnetic sector[25].

The most commonly used mass analyzer, quadrupole was first invented as a mass filter. Quadrupole mass filter consists of four circular symmetrically arranged rods to which rf and dc voltages are supplied. Ion oscillates in the x, y-plane with a frequency depending on its m/z value. If the oscillations of the ion in this plane are stable, the ion will continue to drift down the rods and reach the detector. Later, quadrupole ion trap is modified from linear quadrupole mass filter. In this device, ions are subjected to forces applied by a rf field too. Ions are trapped within the system of three electrodes-a ring electrode and two end-cap electrodes in a hyperbolic cross-section. The motion of the ions lasts if they never hit to the electrodes. Elevating the voltage to the stability limit causes ions to have unstable trajectory and expels ions to be analyzed from the IT[26].

In a linear TOF analyzer, ions in bundles expelled from the source are accelerated by an electric field towards the flight tube. In this field free region, they are separated according to their velocities and from the relationship between the mass of an ion and its kinetic energy, m/z value can be deduced by comparing their time of flights. Simply put, TOF analyzers measure the flight time of the ions in a tube which have been acquired the same kinetic energy with different flight times due to the mass difference. Sensitivity and upper mass range are the most striking features of these instruments that from femtomole to attomole levels, and even masses over 100 kDa can be detected. Since the flight time is proportional to the resolution, length of flight tube is critical for high resolution, however, too long flight path results with the loss of ions. Hence flight tube with a length of 1 to 2 m and an acceleration voltage of at least 20 kV keep both the sensitivity and resolution at a reasonably high values. Apart from this, delayed pulsed extraction and reflectron are two techniques for the improvement of resolution. Ions having same m/z ratio with different kinetic energy cause peak broadening because they reach the detector slightly different times. Delayed pulsed extraction technique enables analyzer to correct this energy dispersion by transmitting more energy to the ions which spend more time in the source, after a certain delay. On the other hand, reflectrons increase the resolution via a second transmission of the ions

in the reverse direction by deflecting them with an electrostatic reflector. Equally spaced ring electrodes in a reflectron act as an ion mirror, therefore, speedy ions penetrate into the reflectron more deeply than the ions having lower kinetic energy. Next, they are repelled outside of the reflectron with the same absolute velocity to the opposite direction to reach the detector.

Among the present MS technologies, Fourier Transform Ion Cyclotron Resonance (FTICR) analyzers exhibit the best mass resolving power, mass resolution, mass accuracy and sensitivity. High cost of the system and necessity of expertise, however, render their usage limited compared to the others. In these instruments, ions trapped in the center of the cell are then excited to a larger radius by the excitation plates which are perpendicular to the magnetic field to create cyclotron frequency. Since frequency is measured more accurately than the other experimental parameters, higher resolution can be acquired. After that, the frequency and time domains are converted to mass spectrum by a mathematical transform based on Fourier inversion theorem[27].

Orbitrap, a lately invented (2005) analyzer is the most noticeable indication of the explicit turn towards proteomics because it was developed from Knight *modified Kingston trap* as a result of extensive demand to a higher performance instrument with low cost and size. In the orbitrap analyzers, once the ions are injected tangentially, they orbit around the central spindle-like electrode, and their electrostatic attraction is prevented by centrifugal force balancing. Thus, ions are trapped around an electrode under the influence of the electrostatic fields rather than magnetic fields and radio frequency. In conclusion, mass spectrum is generated from axial oscillation frequencies of the rotating ion rings by using Fourier transformation[28].

In tandem mass spectrometers, multiple stages of mass analysis and fragmentation between the stages can be carried out. Tandem-in-space analysis requires at least two independent mass analyzers sequentially positioned such as TOF/TOF and triple quadrupole. For tandem-in-time approach, ion trapping mass analyzers, in which the ion of interest is isolated before the fragmentation, are used[29].

1.3. Separation Techniques Before the Mass Analysis

Since high concentration and sample purity are critical for sensitive and accurate mass measurement, some physicochemical properties of proteins, which can be listed as

size, charge, shape, isoelectric point, hydrophobicity, solubility, ligand/metal affinity and structure, are used to reduce the complexity and dynamic range of proteomes. For this purpose, either prefractionation techniques or enrichments strategies are performed before the mass analysis[30]. Although there have been innovative techniques utilizing biochips and nanoparticles, proteins and peptides are separated and isolated by the two techniques: chromatography and electrophoresis.

1.3.1. Chromatographic Techniques

General principle underlying all types of chromatography is the interaction of components in the sample with stationary and mobile phase. While molecules are reversibly detained according to their affinity to the stationary phase, they are dragged to move via stationary phase by the flow of mobile phase. As a result, in a typical liquid chromatography, separation relies on retarding time of the compounds in the colum[31]. High performance liquid chromatography (HPLC), which has a high recovery, reproducibility, speed and particularly superior resolution, became an essential method for analytical separation of proteins and peptides. Furthermore, HPLC provides various separation modes such as hydrophobic interaction, reverse phase (RP), hydrophilic interaction, ion-exchange, gel filtration, immobilized metal ion affinity and immunoaffinity[32].

In addition to these benefits, combining HPLC to electrospray mass spectrometers generates an excellent on-line procedure, in which the ion-suppression effect is reduced, low abundance peptides are enriched, and salts are removed. Mostly RP mode is coupled to MS owing to its appropriate mobile phase content. Moreover, after the production of smaller particles as packing material, ultra high pressure capillary LC systems, which work at high pressure up to 7000 bar, has taken place in the market. These systems exhibit considerable rapidity and high sensitivity particularly for the limited amount samples; however, they require nano-ESI interface, specific pumping equipment and proper detector. Despite its efficiency and practicality, one dimensional liquid chromatography seems insufficient considering the complexity of the proteomes and a vast amount of resultant peptides from the digestion. Therefore, multi dimensional or orthogonal chromatographic techniques are developed for better resolution. In these techniques, RP-LC is always placed prior to mass analyzer, and the

preceding separation is primarily carried out by ion-exchange chromatography. On the other hand, affinity chromatography has been used efficiently for targeted proteins like phosphorylated or glycosylated ones; however, size exclusion chromatography is occasionally utilized because of its low resolution and limited loading capacity. Apart from these, LC is an inevitable segment for quantitative proteomics [33].

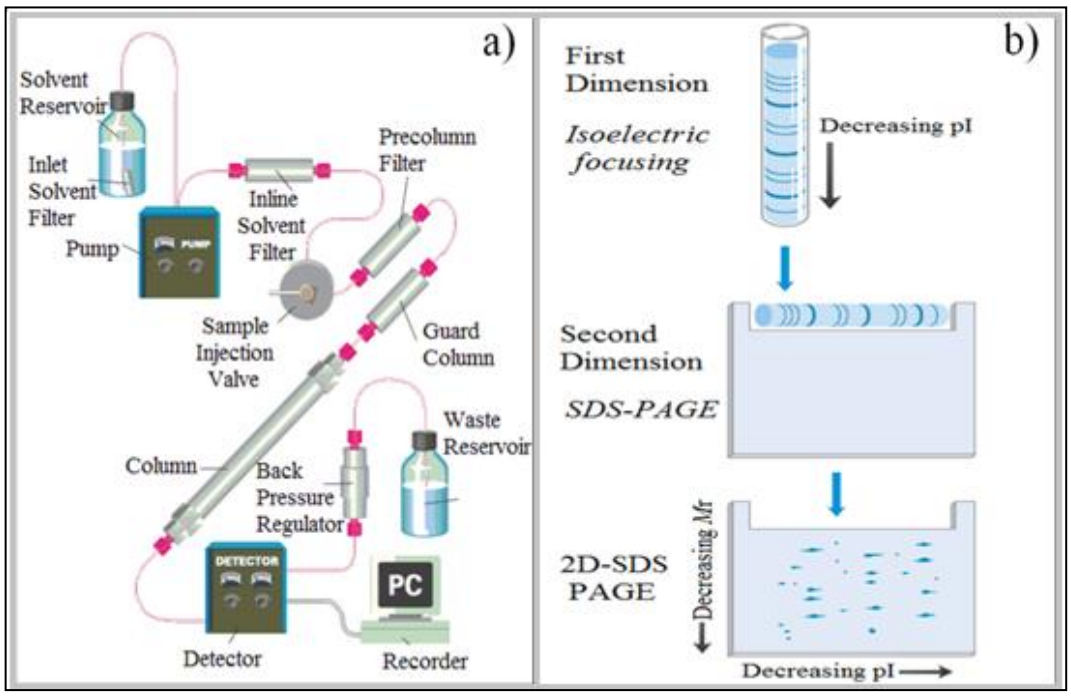


Figure 1.4.Separation techniques [34] a) HPLC b) 2-D SDS PAGE

1.3.2. Electrophoretic Techniques

Electrophoresis is one of the most widely used analytical tool in which the charged molecules are migrated in an electrical field. Electrophoretic methods are usually carried out in the aim of separation rather than purification because proteins' structure and function are affected adversely from the technique. Polyacrylamide gels are prepared from free radical polymerization of acrylamide and cross-linking agent N,N'- methylene-bis-acrylamide forming a physically stable matrix with high resolving power for proteins under the control of the initiator-catalyst couple, ammonium persulphate-N,N,N',N'-tetramethylethylenediamine (TEMED) [31]. To reduce the adverse effect of diffusion, two-phased (stacking phase and resolving phase) discontinuous electrophoresis method is used. In zone electrophoresis, samples are visualized as a spot

or thin band in the medium as a result of speed race depending on the size and charge of proteins. On the other hand, isoelectric focusing separates proteins by concentrating them at their isoelectric points on a pH gradient medium[35].

These two modes of electrophoresis, zone electrophoresis and isoelectric focusing can be applied sequentially creating a remarkably effective and easy technique called two dimensional gel electrophoresis (2D-GE). 2D-GE has gained much more reproducibility after the advent of immobilized pH gradient (IPG) that in the place of carrier ampholytes. Moreover, today's routine IEF gel strips has fairly facilitate the association of IEF method to SDS-GE in comparison to cumbersome and unsuccessful tube gel mode.

2D-GE technique provides a readout to visualize hundreds to thousand of proteins. This unique property enables scientist to differentiate the expressed proteins qualitatively and to some degree quantitatively. Disease state differences, toxic influences and stress impacts have been revealed by comparing readouts of two samples in different states. Despite the increment in resolution obtained by the narrow range pH strips and sensitive staining methods (with silver and fluorescent staining reach up to nanogram level), 2D-GE is not sufficient to cope with the dynamic range of the complex biological samples and fluids at once. One way to solve this problem is lowering the complexity of the sample by handling subcellular proteomes or domains[35].

1.4. Protein Identification By Mass Spectrometry

Present day mass spectrometry-based proteomic studies are being conducted with different methods by using different instruments. In addition, today one can choose one of the bottom-up and top-down approaches. Whereas the former analyze the intact proteins or protein domains, in bottom-up approach proteins are digested to peptides prior to MS analysis. Bottom-up proteomics has been widely used due to the solubility and mass range suitability of the peptides. Nevertheless after the advent of new collision techniques and improvements in resolution, today informative fragment ions are being gained from top-down analysis too [36]. Both approaches have limitations and drawbacks therefore, integration of the two approaches will provide complementary results giving the best yield that can be achieve from mass spectrometry if there is no sample shortage problem[37 38].

Top-down proteomics initiates the analysis with intact protein providing the molecular weight information at the beginning of the measurement. However, the key advantage of the approach is the possible acquisition of the complete sequence coverage because peptides can be lost during chromatography. In addition, too large or too small peptides may have to be ignored during mass analyses. Apart from this, peptides are usually assigned to more than one protein, which generates a new challenge protein inference. However expensive instrumental setup, necessity of a large amount sample and the limited number of bioinformatic tools are the leading downsides of this method[39].

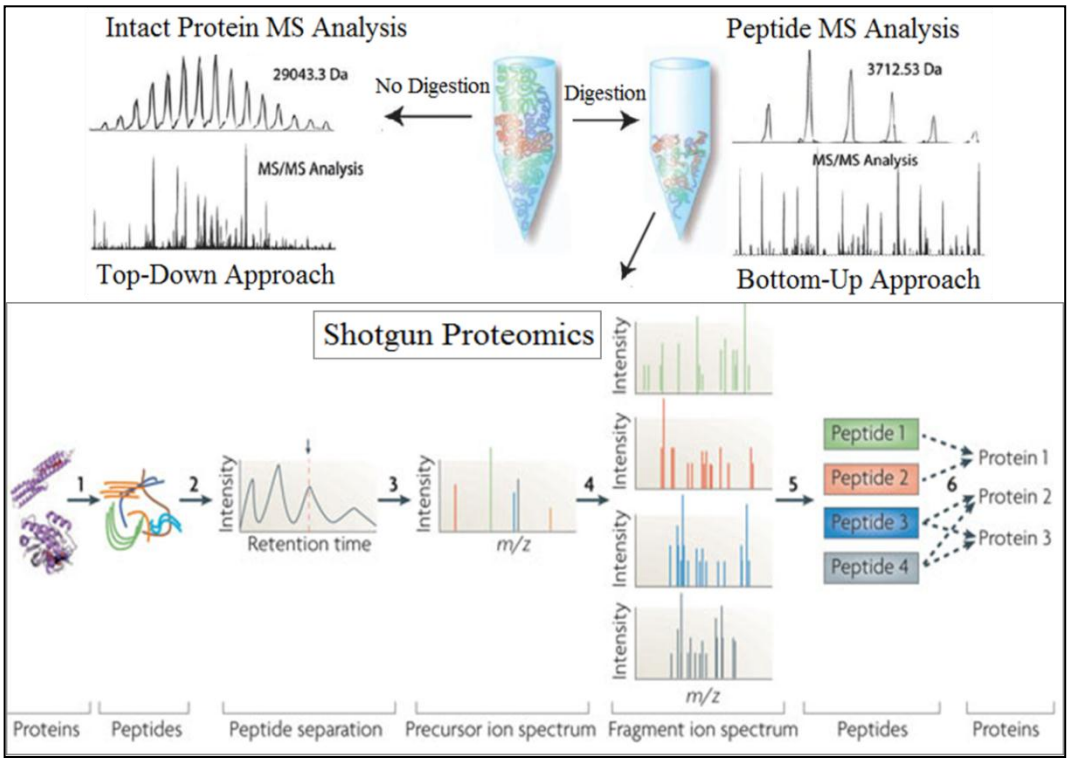


Figure 1.5. Proteomics approaches[40]

Since the information is obtained by putting individual peptides together, bottom-up proteomics is frequently likened to jigsaw puzzles with missing pieces. A well-known digestion protease, trypsin is mostly used for the cleavage due to its high activity and stability. More importantly, trypsin specifically breaks down proteins on the carboxy-terminal side of arginine and lysine residues[41] generating peptides in the effectual mass range (500-2000 Da). Other sequence specific enzymes: endoproteinase Asp-N[42], endoproteinase Glu-C[43], endoproteinase Lys[44] and the less-sequence specific enzymes: chymotrypsin[45] and pepsin [46] have been alternatively used to improve protein

identification and characterization, but they are mostly not preferred. When the sample of protein mixture is not pretreated for the separation, digesting more than one protein together in a sample solution is called shotgun proteomics[47] (Figure 1.5). This high-throughput method, however, has to be coupled with high performance liquid chromatography, favorably with a multidimensional nanoLC, and necessitate a reliable search engine.

1.4.1. Sample Preparation for Bottom-Up Proteomics

Sample preparation and data analysis are the two bottlenecks of MS-based proteomics. Since sample preparation is the first step before the MS analysis, high sequence coverage and protein recovery can not be gained without convenient and reproducible sample handling. Conversion of proteins to peptides is one of the most critical process in the sample preparation procedure. Unfortunately, all of the detected peptides are not informative for the identification, therefore, successful protein digestion, and utmost peptide recovery would increase the chance of protein identification. In addition, artificial modifications and contamination, which might be occurred during the sample preparation protocols, should be avoided as far as possible.

There are many factors influencing the proteolytic result. First of all, enzyme should be added sufficient amount to perform a good digestion, at the same time low enough to eliminate the autolysis products of trypsin. It is not recommended to use less than 1:100 enzyme to protein ratio, mainly 1:50 or 1:25 ratios are preferred. To ensure the best overall digestion efficiency, incubation time is kept long from 9 hours to overnight at 37 °C. In addition, proteins in most samples need to be denatured either by using chaotropic agents or increasing the temperature. However, diluting the sample up to 8-fold, which is necessary for the following tryptic digestion, causes volume elevation; an unfavorable condition for an effective digestion. Apart from this, disulphide bonds between cysteine residues are reduced with one of the reducing reagents; 1,4-dithiothreitol (DTT), β-mercaptoethanol or tris (2-carboxyethyl) phosphine (TCEP). Subsequently, proteins are alkylated with iodoacetamide (IAM) or iodoacetic acid to prevent the potential renaturation[48]. Furthermore, several digestion methods; including ultrasonic-assisted, infrared radiation-assisted, microwave-assisted, pressure-assisted, vortex-assisted, have been proposed to accelerate the digestion

process or to change the resultant peptides[49]. In conclusion, optimal conditions for tryptic digestion increase the digestion efficiency, however, concentrations of protein and other contaminants in the sample also affect the result of the enzymatic reaction.

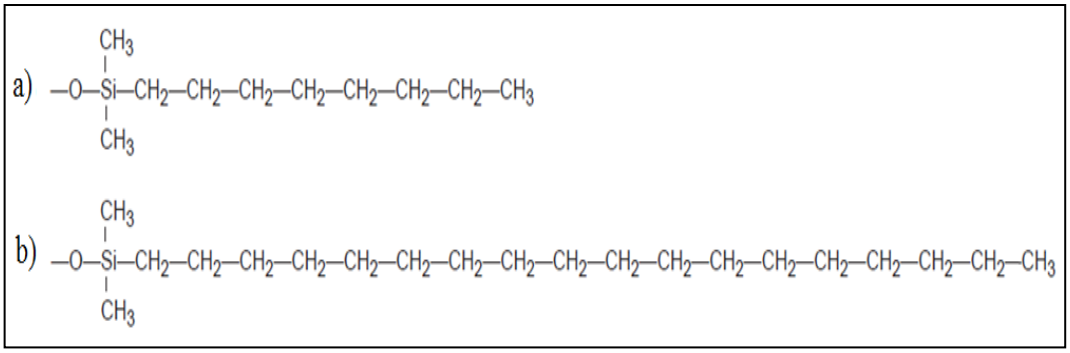


Figure 1.6. N-alkyl hydrocarbon ligands [50] a) Octyl (C8) b) Octadecyl (C18)

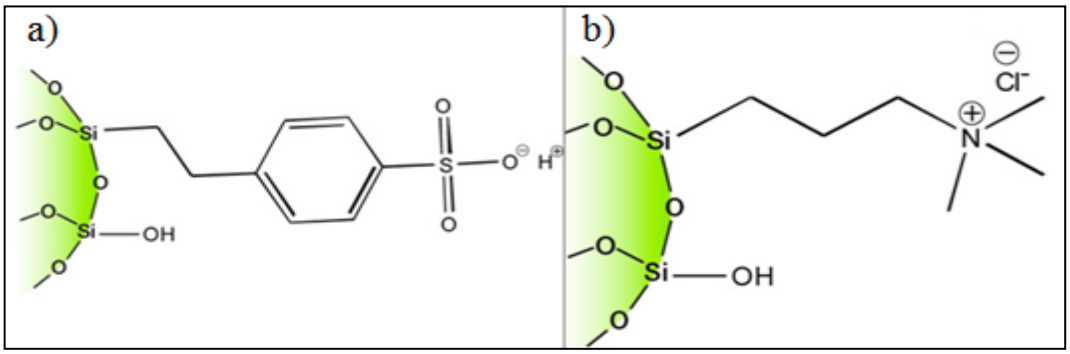


Figure 1.7. Ionic adsorbents a) SCX b) SAX[51]

1.4.2. Peptide Mass Fingerprinting

Peptide mass fingerprinting (PMF) or peptide mass mapping is a protein identification method, which uses masses of protein products after a predictable cleavage, to match with *in silico* digested proteins' peptide masses, in the database. Simply put, the sizes of the pieces are thought as the fingerprint for that protein so in this sense experimentally obtained peptide mass data can be searched in the theoretically constructed peptide mass list of the databases.

In fact, this method come into being after the frequent appearance of some proteins namely contaminant proteins like serum albumin, during sequencing with

Edman degradation thereafter sequences of these proteins started to take place in the databases. Method is rather fast and easy and also effective especially for 2D gel-separated proteins. Advent of MALDI-TOF instruments took away the need for early used less sensitive fast atom bombardment ionization and relatively expensive sector instruments. Moreover, immense accumulation of proteins and DNA sequences in the databases after the advent of proteomics rendered this approach as a simple, efficient and widespread protein identification method[52].

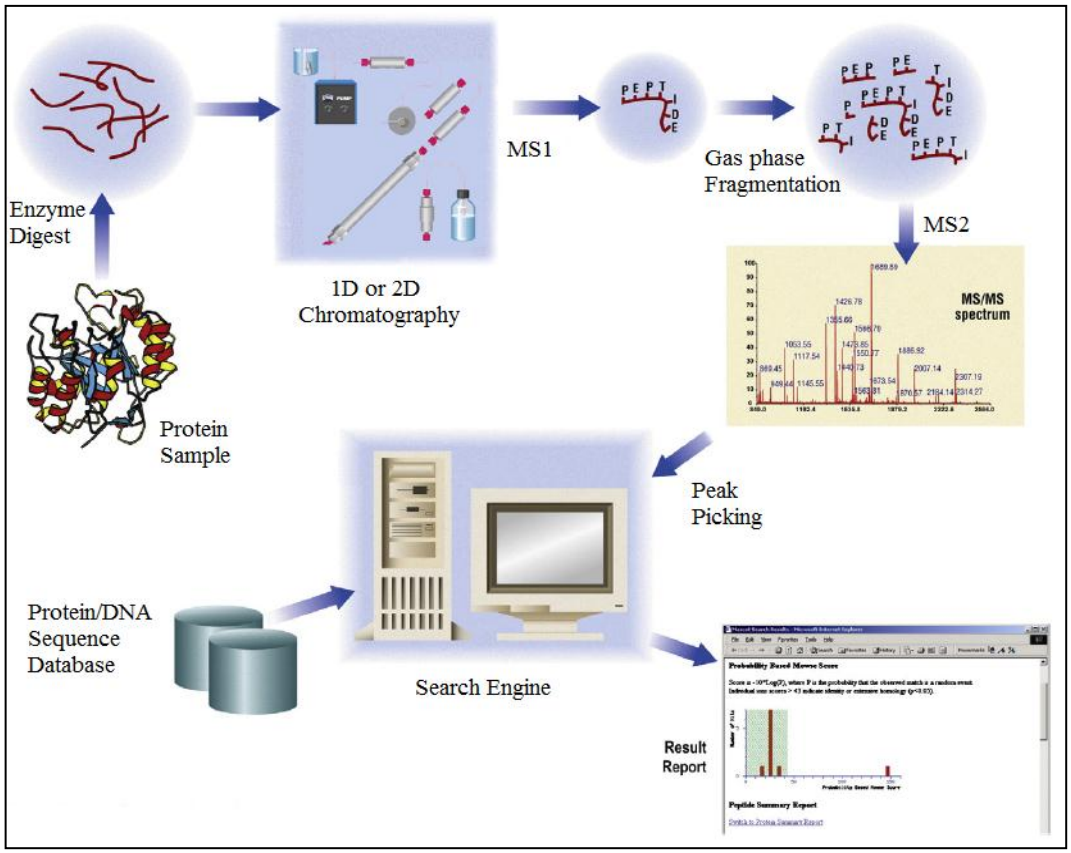


Figure 1.8. Experimental workflow for protein identification [53]

The general strategy in PMF is starting with the elimination of contaminant masses such as trypsin autodigestion peptides, keratin peptides and peaks arising from matrices or dye. After the determination of the parameters; enzyme name, mass tolerance, molecular weight and isoelectric point of the protein, taxonomy, database and allowed potential modifications are determined, and then PMF search is initiated[54].

Despite its simplicity, today PMF is not a reliable or preferable peptide identification method owing to its limitation, among the most striking is it is being relied on only the peptide masses, which can easily correspond to more than one

sequence combination. Consequently to add specification to those masses, further examination for the sequence analysis is necessary which is now being implemented as peptide fragmentation. Apart from that, protein has to be fairly pure; therefore protein mixtures can not be studied with this method. In addition, the protein of interest has to be included in the database; otherwise there will not be any match except the false positives. Post-translationally modified proteins or other modifications occurred during the sample handling, unpredictable adducts and protein splicing variants are the other problematic issues for the identification with PMF[52].

1.4.3. Tandem Mass Spectrometry

In tandem mass spectrometry, a specific ion (precursor ion) is selected and induced to fragmentation. After that, the m/z values of the fragment ions are measured. Tandem MS is either, performed by combining mass analyzers in a tandem configuration or isolating and activating the selected ion (only by ion traps) by the multiple isolation and fragmentation stages, which is abbreviated as MS^n. Collision-induced dissociation (CID), electron capture (ECD) and electron transfer dissociation (ETD) are the commonly used MS/MS fragmentation techniques in proteomics.

CID fragmentation occurs by colliding the neutral gas atoms (helium, nitrogen, argon) with accelerated ions. In this process, the internal vibrational energy is converted to bond cleavage. This widespread, fast and efficient fragmentation method creates b and y fragment ions by breaking the peptide bonds. On the other hand, in ECD multiply charged ions trapped in FT-ICR cell are irradiated with a beam of low energy electrons (< 0,2 eV). Likewise in ETD, the reaction between multiply protonated peptide cations and small molecule anions are proceeded, through the electron transfer from anion to peptide cation. This ECD-like fragmentation also has the advantage of mass analyzer selection other than FT-ICR MS. While y and b ions originate from the dissociation of the amide bond, c and z fragments, which are generated from both ECD and ETD techniques, are gained from the cleavage of N-C_α amine backbone. Whereas protein identification is employed better with CID due to its high performance on the protein coverage, ECD is more suitable for the protein characterization owing to its high performance on the peptide coverage and its success on the detection of PTMs. Nevertheless complementary studies will give the best result since the performance of

the techniques depend on to composition, the length of the peptides and charge state of the ion[55].

Under low-energy collision conditions, fragmentation preferentially occurs along the peptide backbone forming the most abundant fragments, b- and y- ions. Fragmented ions are designated according to the position of charge, and they are labelled consecutively from the original amino terminus of the peptide. According to this nomenclature, fragment ions retaining the positive charge on the amino terminus are called a-, b-, or c ion and likewise x-, y- or z- ion if the charge stays on the C-terminal.

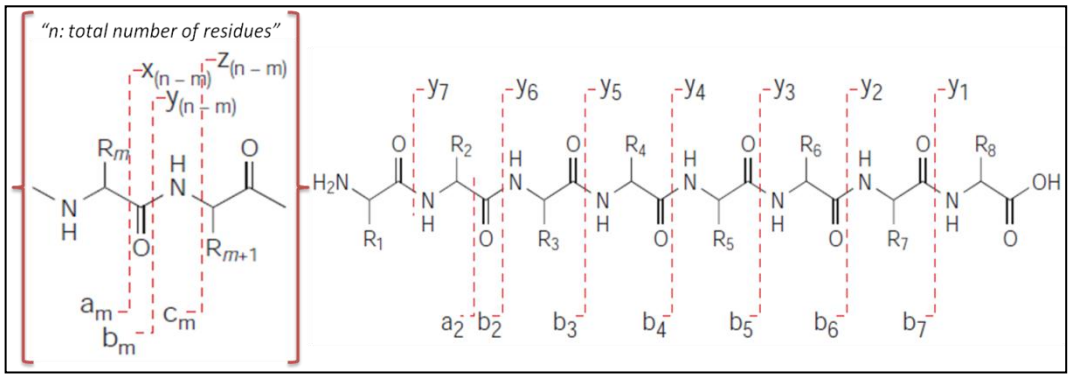


Figure 1.9. Nomenclature of peptide fragment ions [56]

Furthermore, fragmentation on both amino- and carboxy- terminal of the same amino acid produces immonium ions, which appear among the low masses in the spectrum yielding information about the amino acid composition of the sample. Peptide fragmentation is much useful when it produces a sequence ladder, in which the mass difference between the fragments can be correlated to a certain residue. From this ladder, a partial sequence of a peptide can be read forward via from b ions or backward via y ions. Since the labile PTMs are retained on the fragments of the peptide backbone, both the ECD and ETD dissociation techniques are implemented mostly in the protein characterization studies.

1.5. Analysis of Proteomic Data

In mass spectrometry based proteomics, protein identification relies on partial sequence analysis. Although this approach offers a high-throughput platform for the

intricate and formidable proteome researches, reliability of these results are notably questionable. Correct assignment of MS/MS spectrum to a peptide sequence and the succeeding peptide association of these data with the correct protein require high computational capacity and software tools. Database searching, *de novo* sequencing and hybrid approach sequence tags are strategies employed[57].

De novo sequencing is more inductive method than the others since it does not use information beforehand coming from genomic or peptidomic databases. Therefore, it is more applicable for organisms with unknown genome sequence, protein splice isoforms and amino acids modified. Early *de novo* sequencing method, Edman degradation is based on chemical derivatization of amino acids in a peptide adjacently and subsequent separation of these products. However limited study region, (only the isolated, unmodified proteins with accessible N terminus), insensitivity and low sample throughput render this method laborious and less effective. Thus, mass spectrometry has become the method of choice in protein sequencing. *De novo* sequencing with mass spectrometry deduce the sequence from tandem MS spectrum, therefore, completeness of ion series and mass accuracy are of great importance [58]. Type of the ion series can be predicted from the dissociation technique to some extent (e.g. b/y ions from low energy CID) but these C and N terminal fragment series generating a sequence ladder are hardly seen together. Apart from that, neutral losses, internal fragment ions and unpredictable noise peaks cause further complications for the interpretation of the spectrum or less likely provide extra information about the sequence. While sequence ladder is much more necessary for the amino acid order, information about the composition can be inferred through the immonium ions at the low mass region. Nevertheless assigning a peak to an ion type is a formidable task. Although manual interpretation has been carried out to reveal the fragmentation mechanisms, for the peptide identification, more speedy and unbiased solutions are required. For this purpose, several computer algorithms have been devised their success, however, depends on the quality of the spectra and knowledge about peptide fragmentation [59].

Even though, all ion series are not detected after the collision, it has proposed that easily identified partial sequences even containing only two to three amino acids can be useful. In this hybrid approach, short piece identified, and the molecular weight of its preceding and trailing region are together introduced as a sequence tag. This sequence tags are then used to locate the peptide in the database by a search engine[60].

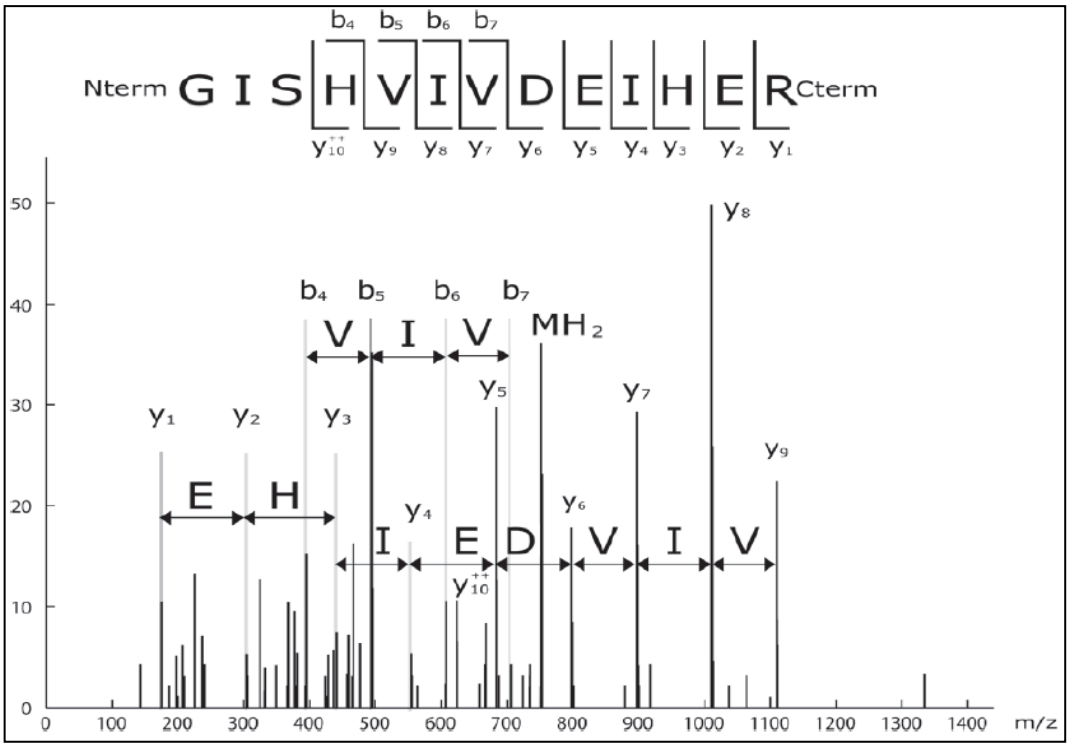


Figure 1.10. An example to sequence ladder [61]

A number of database search algorithms using different models have been developed. MASCOT, SEQUEST, OMSSA, X!Tandem are some of these programs widely use today. Generally they all read the experimental spectra as input, query a sequence database then score them against the theoretical fragmentation patterns and finally provide an output list of the matches that are ranked according to the similarities between experimental spectrum and theoretical spectrum. Common search parameters, database and taxonomy selection, mass tolerance, enzymatic constraint, and modifications are set before the run[62]. The performances of these algorithms, however, mostly differ in terms of sensitivity and selectivity, therefore, credibility of the protein identification considerably depends on the strengths and weaknesses of the MS/MS search algorithm[63]. Interpretation of the data especially gets difficult in the event of shotgun proteomic studies. Although excluding the separation at the protein level facilitates the sample handling and increases the output, losing the connectivity between peptides and protein renders the computational analysis exceedingly intricate. Assembling the identified peptides into proteins, which is also known as protein inference problem, is much more complex in the case of higher eukaryotic organisms due to the sequence redundancy arising from splice forms of the same gene and abundance of proteins with sequence homology. Despite the fact that there is no

established valid way of calling a protein "identified", this problem can be overcame more or less by increasing the protein sequence coverage thus proteins corresponding to only one peptide usually are not accepted as identified[64].

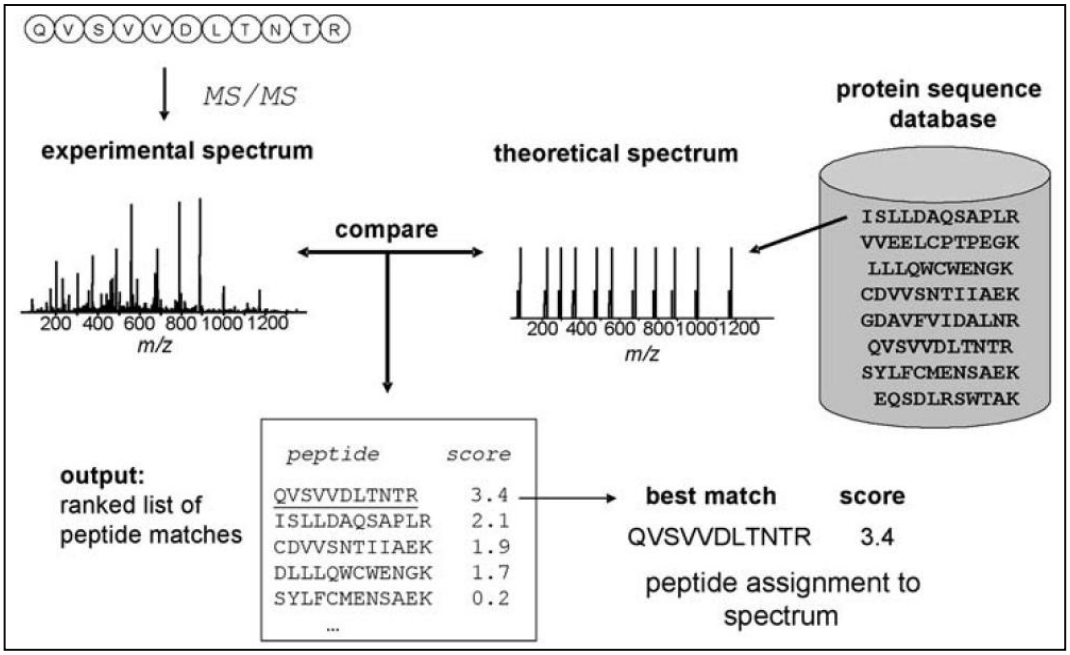


Figure1.11. Tandem mass spectrometry database searching [65]

Mascot is a search engine which uses a probability based scoring. According to this model, MS data are introduced to the system in the form of peak list. From the frequency of matched ions, score is calculated hypothesizing the observed match as a random event. As a result, low probability indicates high score and the default value of significance threshold is set to less than five per cent taking into account commonly accepted threshold probability of an event occurring by chance[66].

1.6. Aim of the Study

When the standard free energy of an interface between two different phases is higher than the bulk phase, protein adsorption to the surface is expected. This principle is used to enrich the protein from dilute or contaminated solution by the help of various sorbent materials. Those adsorbents are able to adsorb the proteins selectively according to the physicochemical properties of the protein, sorbent and bulk solution. However adsorption of proteins on solid surfaces is common but extremely complicated

phenomenon because protein-surface interactions are highly dependent on the individual properties of the system[67].

One of the earliest study in this subject, which was carried out by using the reverse-phase column packing materials C4 and C18 hydrophobic surfaces, reported that proteins could be digested while they were bound to sorbents. In this work, differences between the digestion patterns of in solution digestion and on-surface digestion are related to the partial hinderance for proteolytic cleavage due to the adsorption[68]. Doucette et al. reported a similar study to highly contaminated dilute solutions of model proteins.[69] While the former one used HPLC with UV detection for the detection of eluted peptides, Doucette et al. developed their method for the protein identification by PMF; therefore, they used MALDI-TOF instrument by direct deposition of the beads on the target without peptide elution. Nevertheless, they obtained MS spectrum for peptide mass mapping of dilute and contaminated solutions by the enrichment of proteins on microbead with subsequent rapid cleaning. In a following study, they also tested the effect of the type of the bead support. They compared polymeric poros R2 with conventional C4, C8 and C18 and observed minor differences in terms of protein sequence coverage thus concluded that pore size of the beads does not have a significant effect on the digestion characteristic of an adsorbed protein[70].

Those aforementioned studies, however, only include the hydrophobic surfaces and unfavorable protein identification technique, PMF. On the other hand, Figeys et al. have used ionic adsorbent materials, strong cation exchanger (SCX) and strong anion exchanger in a microfluidic device form, namely proteomic reactor[71]. In this reactor system, a capillary tubing was filled with a slurry of SCX/SAX by applying pressure. Then protein and trypsin were bound to sorbent at adjusted pH. After the reduction and alkylation steps, tryptic reaction was started by increasing the pH up to 8. Finally, peptide elution was carried out. Figeys et al. compared their proteomic reactors (SAX & SCX) and traditional in-solution digestion methods on the proteome of yeast by using LC-MS/MS [72]. They concluded that those complementary reactors together outperform the conventional in-solution digestion for peptide and protein identification. Although proteins were adsorped to SCX and SAX by pH adjustment at the beginning of the experiment, in their study they facilitated those SCX and SAX sorbent materials to enrich the protein rather than performing the digestion on the surface. Nevertheless,

their efforts and success on various proteomic reactors are quite promising for the usage of sorbent materials in the field of protein identification.

Objective of this study was mainly to investigate this occasionally studied alternative digestion method, on-surface digestion by using both two ionic and three hydrophobic adsorbent materials on the different standard protein and their mixture. MALDI-TOF/TOF mass spectrometry was the main instrument for the protein identification throughout the study therefore it was aimed to adapt the on-surface digestion protocol for the tandem mass analysis with MALDI-TOF/TOF MS. We intentially preferred rather practical and applicable work steps to be able to address the method to any proteomic laboratory without need to serios expertise and money. As distinct from the aforementioned studies, various type of adsorbents were involved to the work together, which are strong cation exchange, strong anion exchange, hdyrophobic C8, C18 and rather smaller one, poros R2 micro beads. Despite its longer sample preparation steps comparing to in-solution digestion, we believe that on-surface digestion has a potential in proteomics for the improvement of protein digestion results.

CHAPTER 2

MATERIALS AND METHODS

2.1. Verification of Protein Adsorption

A simple verification of protein adsorption to different adsorbents (C18, C8, SCX, SAX) was performed by using Bradford protein assay. This method relies on the absorbance shift of the dye Coomassie Brilliant Blue at 595nm due to the protein binding. Coomassie Brilliant Blue dye binds arginine, lysine and histidine residues of proteins. Absorption decrease in protein solutions after sorbent addition was considered as protein adsorption onto beads.

Benzenesulfonic SCX, quaternary amino SAX (Spe-ed SPE cartridges; Applied Separations) and hydrophobic adsorbent materials C18, C8 (Finisterre C18/C8 SPE column; Teknokroma) were obtained by breaking the disposable solid phase extraction columns. 3 mg from each microbead was weighed and washed 3 times with 250ul wash solution before the addition of proteins. As the wash solution, organic solvents are used for hydrophobic sorbents, C8 and C18; acidic and basic solutions are chosen respectively for SCX and SAX. C18 and C8 microbeads were washed with 100% methanol (Sigma-Aldrich) twice and once with 50% methanol. SAX and SCX were washed three times with 1% ammonium hydroxide (26% NH_3 Riedel-deHaen) and 1% trifluoroacetic acid (Merck) respectively.

To compare the proteins' adsorption behaviour, 100ul 0,5 mg/ml protein solutions of myoglobin, cytochrome c, bovine serum albumin and lysozyme (Sigma) were added on the sorbents, then tubes were adhered onto vortex shaker with sticky tape and next they were left to agitation for 4-5 hours. 150 μL of coomassie plus^{TM} protein assay reagent (Thermo Scientific) was mixed with 90 μL distilled water and 60 μL protein sample in a well (96 well F-bottom plate; greiner bio-one). 5 minutes later absorbance values at 595nm were measured by spectrophotometer (Multiskan Spectrum; Thermo Electron Corporation). Preparation and the composition of Bradford reagent alternative to commercial one and the standard curve are given in Appendix A.

Table 2.1 Specifications for standard proteins

Accession number	Index number	Name of the Protein	pI	Protein Molecular Weight (Da)	Organism
P00004	81522	Cytochrome c	9.6	11833	Equus caballus (Horse)
P00698	213825	Lysozyme c	9.4	16239	Gallus gallus (Chicken)
P68082	245046	Myoglobin	7.2	17083	Equus caballus (Horse)
P02662	48707	Alpha-S1-casein	5.0	24529	Bos taurus (Bovine)
P02663	48720	Alpha-S2-casein	8.5	26019	Bos taurus (Bovine)
P00921	46499	Carbonic anhydrase 2	6.4	29114	Bos taurus (Bovine)
P02769	12073	Serum albumin	5.8	69294	Bos taurus (Bovine)

Table 2.2. Specifications for adsorbents

Sorbent	Functional group	Size
SCX	Benzenesulfonic Acid	40µm irregularly-shaped silica, 60 Å mean porosity
SAX	Quaternary Amino	40µm irregularly-shaped silica, 60 Å mean porosity
C8	Polymerically bonded octadecyl C8	50 µm irregular-shaped silica, 60 Å mean porosity
C18	Polymerically bonded octacyl C18	50 µm irregular-shaped silica, 60 Å mean porosity
Poros R2	Reverse phase	2000 angstrom pore-size Poly (Styrene-Divinylbenzene)

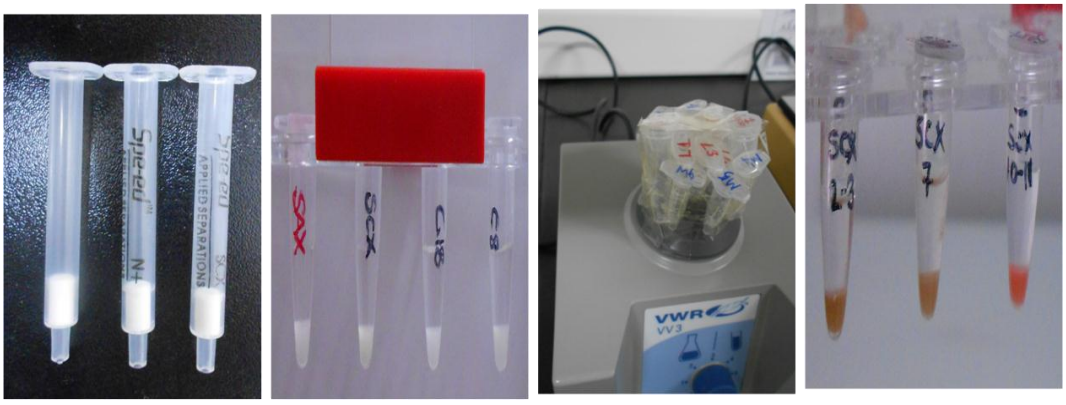


Figure 2.1. Workflow of protein adsorption onto adsorbents

2.2. Protein Digestion Methods

For the protein tryptic digestion, three different digestion methods were carried out to compare the digestion efficiency through the signal quality and resultant peptides.

2.2.1 In-Gel Digestion

This protocol takes two days or more depending on the removal of the dye. Reagent are as follows:

- Wash solution: 50% (v/v) methanol and 5% (v/v) acetic acid.
- 100 mM ammonium bicarbonate in water:
- 50 mM ammonium bicarbonate in water
- 10 mM DTT in 100 mM ammonium bicarbonate
- 100 mM iodoacetamide in 100 mM ammonium bicarbonate
- Extraction buffer: 50% (v/v) acetonitrile and 5% (v/v) formic acid.
- Trypsin solution is prepared by adding 1.0 mL of ice cold 50 mM ammonium bicarbonate to 20 μg of sequencing-grade modified trypsin. The final concentration is 20 ng/μL trypsin.

First protein bands are cut from the gel as closely as possible with a sharp scalpel, and divided into smaller pieces that are approximately 1mm^3 to 2 mm^3. Gel pieces are placed in a new plastic microcentrifuge tubes and 200 μL of the wash solution is added and they are rinsed overnight at room temperature. If desired, this washing step can be carried out over the weekend or, alternatively, for 4 h. On the second day, the wash solution is removed from the sample with a plastic pipette and discarded. Then 200 μL of acetonitrile was added to dehydrate the gel pieces for ~5 min at room temperature. When dehydrated, the gel pieces will have an opaque white color and will be significantly smaller in size. Carefully the acetonitrile is removed from the sample with a plastic pipette. Completely the gel pieces are dried at ambient temperature in a vacuum centrifuge for 2 to 3 minutes. Then 30 μL of 10 mM DTT is added and proteins are reduced for 0.5 hour at room temperature. DTT solution is removed from the sample carefully and 30 μL of 100 mM iodoacetamide is added to alkylate the protein at room temperature for 0.5 hour. After 30 minutes the iodoacetamide solution is removed from the sample with a plastic pipette carefully. Next 200 μL of acetonitrile is added to dehydrate the gel pieces for ~5 min at room temperature. Acetonitrile is removed from the sample with a plastic pipette again. Once more the gel pieces are rehydrated in 200 μL of 100 mM ammonium bicarbonate for 10 minutes at room temperature. After that the gel pieces are dehydrated last time with acetonitrile and then they are dried at ambient temperature in a vacuum centrifuge for 2

to 3 minutes. Finally 30 μL of the trypsin solution is added to the sample and the gel pieces are allowed to rehydrate for 10 minutes with occasional vortex mixing. 5 μL of 50 mM ammonium bicarbonate is added to the sample to cover the gels when necessary. The sample is drove to the bottom of the tube by centrifuging for 30 sec, and the digestion is carried out overnight at 37 °C

On the third day peptides produced by the digestion are extracted in three steps. First 30 μL of 50 mM ammonium bicarbonate is added to the digest and incubated fr 10 minutes with occasional gentle vortex mixing. The digest is drove to the bottom of the tube by centrifuging the sample for 30 second. Then 30 μL of the extraction buffer is added to the tube containing the gel pieces and incubated in the sample for 10 minutes with occasional gentle vortex mixing. The extract is drove to the bottom of the tube by centrifuging the sample for 30 seconds. Supernate is collected carefully with a plastic pipette and combined in the 0.5-mL plastic microcentrifuge tube. A second 30-μL aliquot of the extraction buffer is added to the sample and last step is repeated. The volume of the extract is reduced to < 20 μL by evaporation in a vacuum centrifuge at ambient temperature. The volume of the digest is adjusted ~ 20 μL, as need, with 1% acetic acid. At this point, the sample is ready for analysis however a final step for the salt remove is needed and. C18 ZipTip™ from Millipore Corporation, a 10 µL pipette tip with a bed of chromatography media fixed at its end is mostly used for this purpose. To equilibrate the ZipTip pipette tip for sample binding, maximum volume setting of 10 µL 0,1% TFA in water is aspirated as wetting solution and dispensed to waste three times. After equilibrating the tip, peptides are bound by fully depressing the pipette plunger to a dead stop. The sample is aspirated and dispensed 7-10 cycles without dropping the sample to the waste. Then pipette is again washed with wash solution at least twice. 1 to 4 µL of elution solution which contains 50% acetonitrile in 0,1% TFA is dispensed into a clean vial.

2.2.2. In-Solution Digestion

Trypsin works best in a pH range of 7,5-8,5 and it is resistant to mild denaturing conditions such as 0,1% SDS, 1M urea or 10% acetonitrile therefore procedure are designed considering working conditions of trypsin. However protein folding can protect the amino acid chain from enzymatic cleavage so denaturation may be necessary

for efficient cleavage. 6-8 M urea is mostly used as common denaturant but this time sample volume needs to be increased to reduce the concentration of urea to 1 M before trypsin addition. According to the preliminary results we observed that urea did not affect the results to a great extent so we excluded urea addition in our experiment since overnight incubation would provide a gradual digestion. Generally digestion is carried out in 100 mM ammonium bicarbonate or Tris-HCl buffer but our experiments showed that 20 mM concentration is enough for less amount of protein sample therefore to eliminate the salt remove step or need to ziptip, 20 mM ammonium bicarbonate can be used. For in solution digestion lower sample volumes/higher protein concentration work best therefore protein enrichment by evaporation is recommended. Ratio of enzyme to trypsin changes according to incubation time but no less than 1:100 should be used. We mostly used 1:25 or 1:50 for in solution digestion. To cleave the disulphide bonds, reduction and alkylation buffers are used. To 1 mg of total protein 5 μL of 200 mM reducing reagent DTT is added to before starting the digestion. After 45 minutes incubation of reducing reagent, 20 μL of 200 mM alkylating reagent iodoacetamide is added and incubated for 45 minutes. 20 μL of the reducing agent is added again to consume any unreacted iodoacetamide. If one is sure about the content of cysteines and disulphide bonds, these alkylation and reduction steps may not be necessary.

2.2.3. On-Surface Digestion

Protein adsorption was carried out by adding 100 μL of 0,1 mg/ml protein solution and 0,22 mg/ml protein mixture solution onto the washed microbeads as described at 2.1 then tubes were adhered onto vortex shaker with sticky tape and shaked for 4-5 hours. Before adding alkylation and reduction buffer, beads are washed with water three times to remove any unbound proteins. 5 μL of 100 mM DTT was added and incubated for 45 minutes. After that 5 μL of 200 mM iodoacetamid was added and incubated for 45 minutes. 5 μL DTT solution was added again to consume any unreacted iodoacetamide. Before trypsin addition beads are pulled down by centrifugation and liquid part was discarded. 10 μL of trypsin solution with 1:10 ratio was added and proteins bound to microbead are incubated at 37 C overnight.

2.3. One Dimensional SDS-Polyacrylamide Gel Electrophoresis

- Preparation of 12% resolving gel for tris-glycine SDS-Polyacrylamide gel electrophoresis:

All the chemical used in SDS-PAGE were purchased from applichem. For mini gel of Mini-PROTEAN® Tetra Cell (Bio-rad) 5 ml volume of total resolving gel is sufficient. It contains 1.6 ml H2O, 2.0 ml 30% acrylamide solution, 1.5 M Tris-HCl (pH 8.8), 0.05 ml 10% SDS, 0.05 ml freshly prepared 10% ammonium persulfate and 0,002 ml tetramethylethylene diamine (TEMED).

- Preparation of 5% stacking gel for tris-glycine SDS-polyacrylamide gel electrophoresis:

For mini gel of Mini-PROTEAN® Tetra Cell (Bio-rad) 3 ml stacking gel is sufficient. The solution components are as follows :2.1 ml H2O, 0.5 ml 30% acrylamide mixture in water, 0.38 1.0 M Tris-HCl pH(6,8), 0.03 ml 10% SDS, 0.03 ml freshly prepared 10% ammonium persulphate and 0,003 ml TEMED.

- Pouring SDS-polyacrylamide gels: The glass pates are assembled according to the manufacturer's instructions.

First resolving gel is poured into the gap between the glass plates. Sufficient space is left for the stacking gel. Acylamide solution is overlayed with water or isobutanol to prevent oxygen diffusion which inhibits the polymerization. After about thirty minutes when the polymerization is completed, overlay is poured off and top of gel is washed with deionized water to remove any unpolymerized acrylamide. Then remaining water is removed with the edge of a paper towel. After that stacking gel solution is poured directly onto the surface of the polymerized resolving gel. Comb is immediately inserted avoiding to trap any air bubbles. The gel is left vertical position at room temperature to polymerize.

2.4. Protein Extraction By Passive Elution From Polyacrylamaide Gel

Proteins are extracted from polyacrylamide gel into the solvent (formic acid, water, isopropanol 1:3:2 v/v/v) by passive elution. Protein bands are cut from the gel, crushed into small pieces with micro pestle. 50 μL of extracting solvent is added and the tube is adhered onto vortex by sticky tape for overnight. Next day supernatant is placed

into a clean tube after centrifugation. For MALDI-TOF analysis volume is reduced to ~5 μL with vacuum concentrator (Christ RVC 2-33).

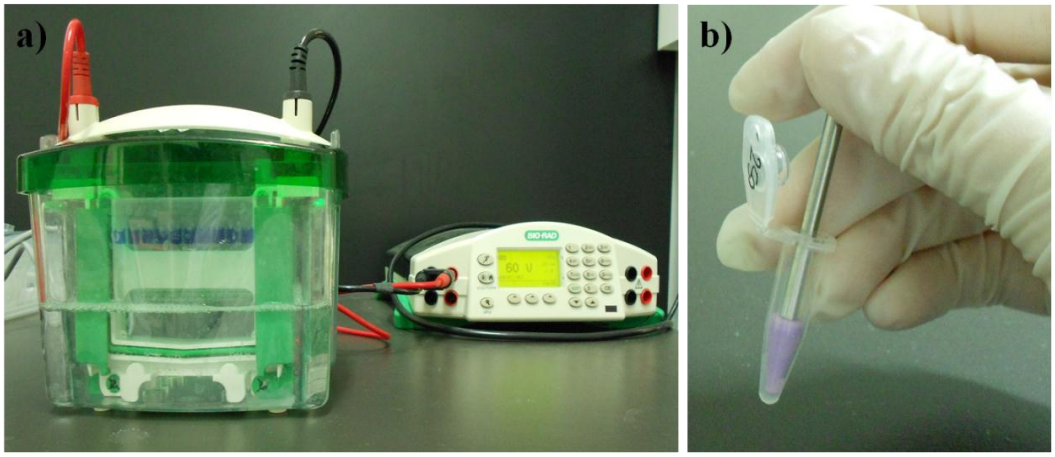


Figure 2.2. a)Mini SDS-PAGE system b) Micropestle: A gel crushing tool

2.5. Sample-Matrix Deposition Onto MALDI-TOF/TOF Target

The matrix used for the peptides was alpha cyano-4-hydroxycinnamic acid (CHCA) with a two-layer MALDI deposition method. This involves the deposition on the MALDI target of a microcrystalline matrix layer via fast evaporation from 0,6-1 μL solution of CHCA (10 mg/ml) dissolved in 20% methanol in acetone. A 1 μL aliquot of the digested protein sample was mixed with either 2 or 4 of saturated CHCA solution in 40% methanol in 0,1% TFA-water. The peptide-matrix solution was vortexed and 1 μL portion was deposited on top of the first matrix layer.

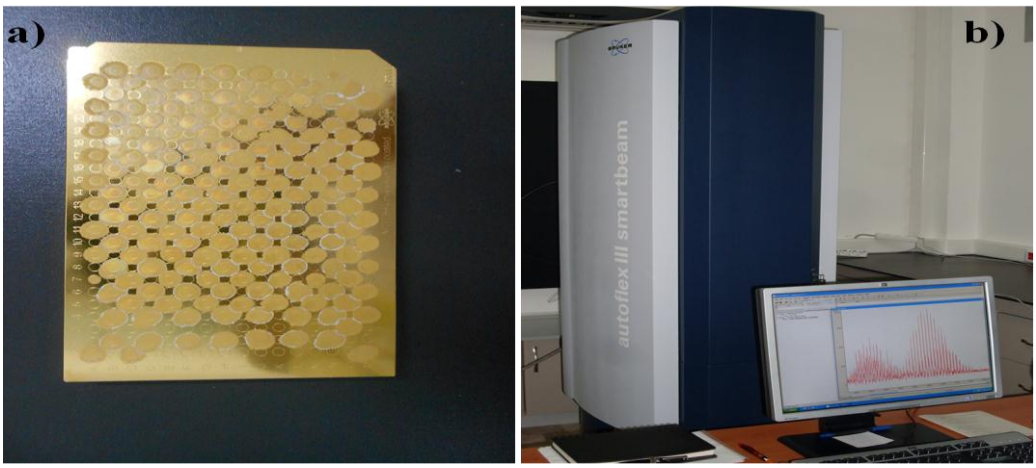


Figure 2.3. a) Gold MALDI target plate b) MALDI-TOF/TOF MS system

For the protein analyses, a mixture of α-cyano-4-hydroxycinnamic acid and 2,5-dihydroxybenzoic acid was used. 5 mg of CHCA was dissolved in 1ml of 20% methanol/acetone and 20 mg of DHB was dissolved in 1 ml of 20% acetonitrile in 0,1 TFA-water. A 1 μL of the protein sample was mixed with 3 or 5 μL of 1:2 CHCA/DHB mixture. The protein-matrix solution was vortexed and c portion was deposited onto the target.

2.6 Protein Identification by Mascot Search Engine

Spectra were processed and analyzed by Autoflex III Smartbeam (Bruker) which uses internal MASCOT software (Matrix Science, London,UK) for searching the MS/MS data. This type of MALDI TOF/TOF works with programs flexControl 3,0 and flexAnalysis 3,0. Processed data by flexControl is transferred to MASCOT software by another licensed program biotools 3.1. NCBI nonredundant and Swiss-Prot protein sequence databases were used for searches under metazoa (animals) taxonomy . Other database search parameters were as follows: carbamidomethylation (C) as fixed global modification allowance for up to one (increased up to four when necessary) missed tryptic cleavage. The peptide mass tolerance was 1,2 Da and fragment ion mass tolerance was 0,6 Da. Charge state was 1+ and monoisotopic mass was considered. Mass range of the analyses was set to 700-3500 Da. Protein Prospector v 5.10.6. is used to check manually the assigned peaks to the calculated peptide masses.

CHAPTER 3

RESULTS AND DISCUSSION

3.1. Protein Adsorption onto Sorbents in Solution

When an amino acid is dissolved in water, it can act as either an acid or a base owing to its dipolar ion (zwitterion) nature. Proteins exhibit acidic or basic characteristic too depending on their isoelectric point (pI) which is defined as the characteristic pH at which the net electric charge is zero. Simply isoelectric points reflect the nature of the ionizing R groups present in the protein sequence. This feature helps proteins and peptides to be separated by cation and anion exchanger adsorption materials. In addition, clustered hydrophobic amino acid residues in the proteins lead to hydrophobic interactions with hydrophobic ligands; typically linear hydrocarbon chains. As a matter of fact, a well-accepted and very successful analytic separation technique, High Performance Liquid Chromatography (HPLC) bases on this adsorption ability of sorbent materials by using continuous flow under high pressure.

In this study, we first tested whether proteins were adsorbed on sorbents while they were mixed with sorbents in a solution environment rather than tightly packed column without the driving force of pressure. In protein/peptide HPLC columns, silica-coated octyl (C8) and octadecyl (C18) alkyl hydrocarbon ligands are mostly used as the packing material. In this part of the study, for the hydrophobic interaction both C8 and C18 and for ionic interaction bead-coated with functional groups of benzenesulphonic (SCX) and quaternary amino (SAX) were utilized. Although these materials may address to all proteins, their adsorption efficiency changes according to the protein and its physicochemical properties, therefore, examination of proteins with various characteristics will strengthen the validation of the method being developed. For this reason, four different proteins; myoglobin, lysozyme c, bovine serum albumin and cytochrome c were used. In order to verify and observe the adsorption behaviour of protein in solution, a quite uncomplicated experiment with 4 hours vortex agitation and Bradford assay were used. Decrease in protein concentration after the sorbent addition was shown through the absorbance decrease at 595 nm.

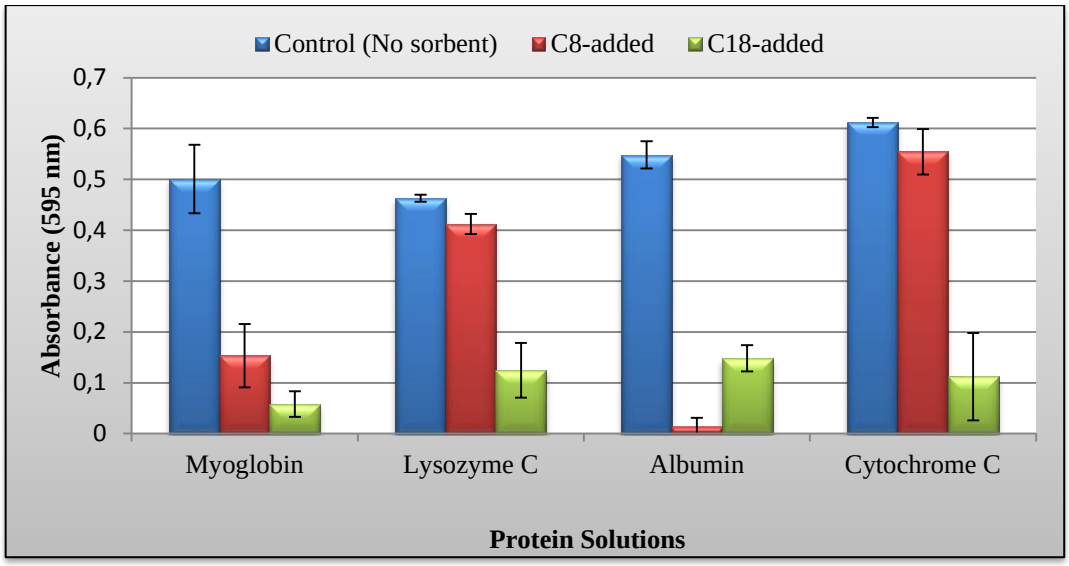


Figure 3.1. Absorbance decrease after hydrophobic sorbents addition

Among the hydrophobic surfaces, absorbance decrease after C18 addition was greater than C8 except albumin (Figure 3.1.). Since C8 columns are preferred for large proteins and likewise C18 column for rather small proteins and peptides, this result is consistent with general HPLC rules by reason of the mass difference between albumin and other proteins. On the other hand, adsorption behavior of ion exchanger sorbents were more selective however, this selectivity was less predictable. As a rule, proteins with p*I* < 6 (i.e., acidic proteins) are chromatographed on an anion-exchange column, while proteins of p*I* > 8 (basic proteins) are chromatographed on a cation-exchange column, and proteins with p*I* between 6 and 8 can be chromatographed on either type.

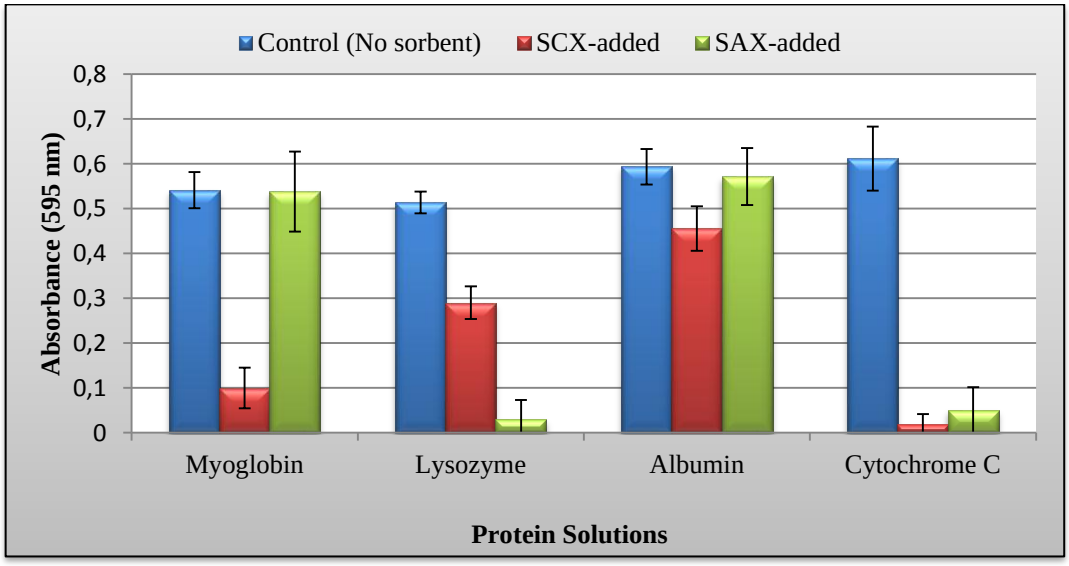


Figure 3.2. Absorbance decrease after ionic sorbent addition

Thus, basic proteins lysozyme c and cytochrome c in the water medium at pH 7 were being expected to be adsorbed strongly by SCX and slightly by SAX. Although they were both adsorbed more or less by SCX, in contrast to expectations SAX strongly adsorbed both basic proteins, which was concluded from the concentration decrease after SAX beads addition (Figure 3.2.). Myoglobin might be misleading for the SCX and SAX comparison by reason of being a neutral protein. In addition, proteins precipitate when the pH of the solution is close to pI of the protein therefore, any precipitation arising from this principle might deceive and give false positive result. Nevertheless control samples, which did not contain sorbents, were controlled beforehand for any precipitation arising from vortex agitation or pH. Other proteins; carbonic anhydrase and casein were excluded because of this false positive effect of precipitation. As mentioned earlier in the introduction part, protein-surface interactions are highly dependent on the individual properties of the system. Therefore, specificity of each protein in terms of molecular weight, amino acid composition, localization of amino acids, protein three-dimensional structure and other prevailing interactions have contributions to the protein adsorption.

Apart from that, some enzymes require additional chemical component to carry out their function. They are called prosthetic group when they tightly or even covalently bound to the protein. As an example, both myoglobin and cytochrome c proteins contain heme prosthetic group in their structure so the iron atom in the structure of heme might affect the prevailing forces governing the ion exchange adsorption phenomenon. Proteins were not denatured before the sorbent addition therefore, structural formation and the locations of acidic, basic and hydrophobic residues of great importance for the adsorption.

Last but not least, adsorption relies on energetically favorable and collectively dominant noncovalent interactions therefore changes in the environment and the external parameters influence the adsorption phenomenon, especially the hydrophobic interactions. For example, better protein adsorption to hydrophobic surfaces was observed a at higher temperature and at higher ionic strength however, in these cases protein aggregation again might be encountered as a false positive for this measurement. A better quantification method is necessary to detect only the proteins bound to the sorbent beads. In short, this experiment was performed to have an opinion about the adsorption behaviours of different proteins in the solution. After that, additional Bradford assay were carried to determine the sufficient bead amount by using two

proteins; bovine serum albumin (BSA) and cytochrome c (cyt c). Consequently, less than 2 mg sorbent amount was sufficient for 1µg protein therefore 2-5 mg mass range was fixed for all experiment in this study (Figure 3.3.).

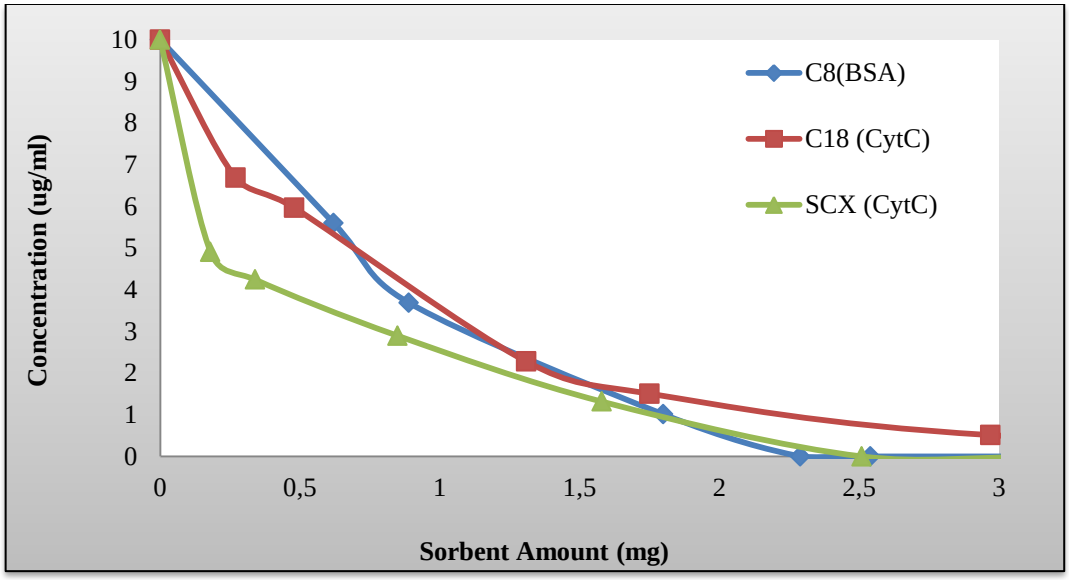


Figure 3.3. Effect of sorbent amount on protein concentration

3.2 Comparison of Protein Digestion Methods on Individual Proteins

For the MALDI-TOF analysis, 1µl or even less volume of the sample is sufficient however concentration of contaminants and salt coming from the buffer, increase too during the protein concentration step, solvent evaporation. Therefore peptide preconcentration by using Ziptip may be necessary prior to the deposition of the sample onto MALDI target plate. Instead of using ziptip to remove the salt, urea, detergent etc., decreasing the concentration of the ammonium bicarbonate salt is preferred to the degree of MALDI tolerance. For this reason, effect of the buffer concentration to the tryptic digestion was first examined. In APPENDIX B, it can be seen that proteins can be digested in medium containing lower buffer concentration too. In conclusion, trypsin can perform its function in 1-25 mM ammonium bicarbonate solution keeping the pH around 7 so we compromised the slight effect of buffer concentration and did not use 100 mM ammonium bicarbonate solution, which is generally preferred. In addition, proteins were not denatured by the addition of urea. Probably long duration of the overnight digestion causes the gradual digestion of

proteins from the outer side to the inner side thus it might not require the usage of urea as denaturing reagent. Using urea was avoided to eliminate the negative effects of urea to trypsin thus no volume increment was necessary.

The conventional tryptic digestion method, in-solution digestion was compared with on-surface digestion through the MS spectrum of each method. These MS spectra were used to identify the proteins by peptide mass fingerprinting and to reveal the differences in digestion patterns of the two methods. Although same dominant peptide peaks were mostly observed on the spectrum of in-solution digestion and on-surface digestion, there were still new or missing peptide peaks leading to the difference. Appearance of new signals, which are assigned to a peptide, helps to increase the sequence coverage. As proposed in the literature, these new peptide signals are generated from the miscleaved peptides. However when these new peaks do not correspond to a peptide from the protein of interest, they might mislead the researcher and the search engine thus cause to confusion and false results. Especially noise peaks which were arisen from the matrix, Na/K adducts and chemical modifications were often encountered during the experiments. To make an unbiased comparison, protein sequence coverage and sequence list of the peptides, which were identified by Mascot PMF search, were put on the spectra. All the spectra shown in figures were drawn by Igor Pro program according to the 100% relative intensity. In addition, some particular peptides identified by Mascot MS/MS search, are listed in APPENDIX C

In conclusion, we demonstrated that proteins can be digested while bound to the microbeads. Efficiency of on-surface digestion depends on two uncertain experimental processes: adsorption of the protein and elution of the peptides. Nevertheless in this method, proteins can be selectively concentrated from the solution and removed from the contaminants. Unlike the studies in the literature, peptides were retrieved from the ion exchanger surfaces by pH alteration rather than salt addition. This enables one to concentrate the sample by evaporating the solvent; thus ziptip usage became optional. However it is observed that duration of the elution is significant for the peptide recovery therefore overnight elutions, and frequent vortex shaking should be performed to release the peptides from the sorbent surfaces.

Table 3.1. Amino acid sequence of Alpha-S1-casein protein from *Bos Taurus*

1	MKLLILTCLVAVALARPK**HPIKHQGLPQEVLNENLLRFFVAPFPEVFGK**EKVNELSKDIG
61	SESTEDQAMEDIKQMEAESISSSEEIVPNSVEQK**HIQKEDVPSERYLGYLEQLLR**LKK**YK**
121	**VPQLEIVPNSAEER**LHSMKEGIHAQQK**EPMIGVNQELAYFYPELFR**QFYQLDAYPSGAWY
181	YVPLGTQYTDAPSFSDIPNPIGSENSEKTTMPLW

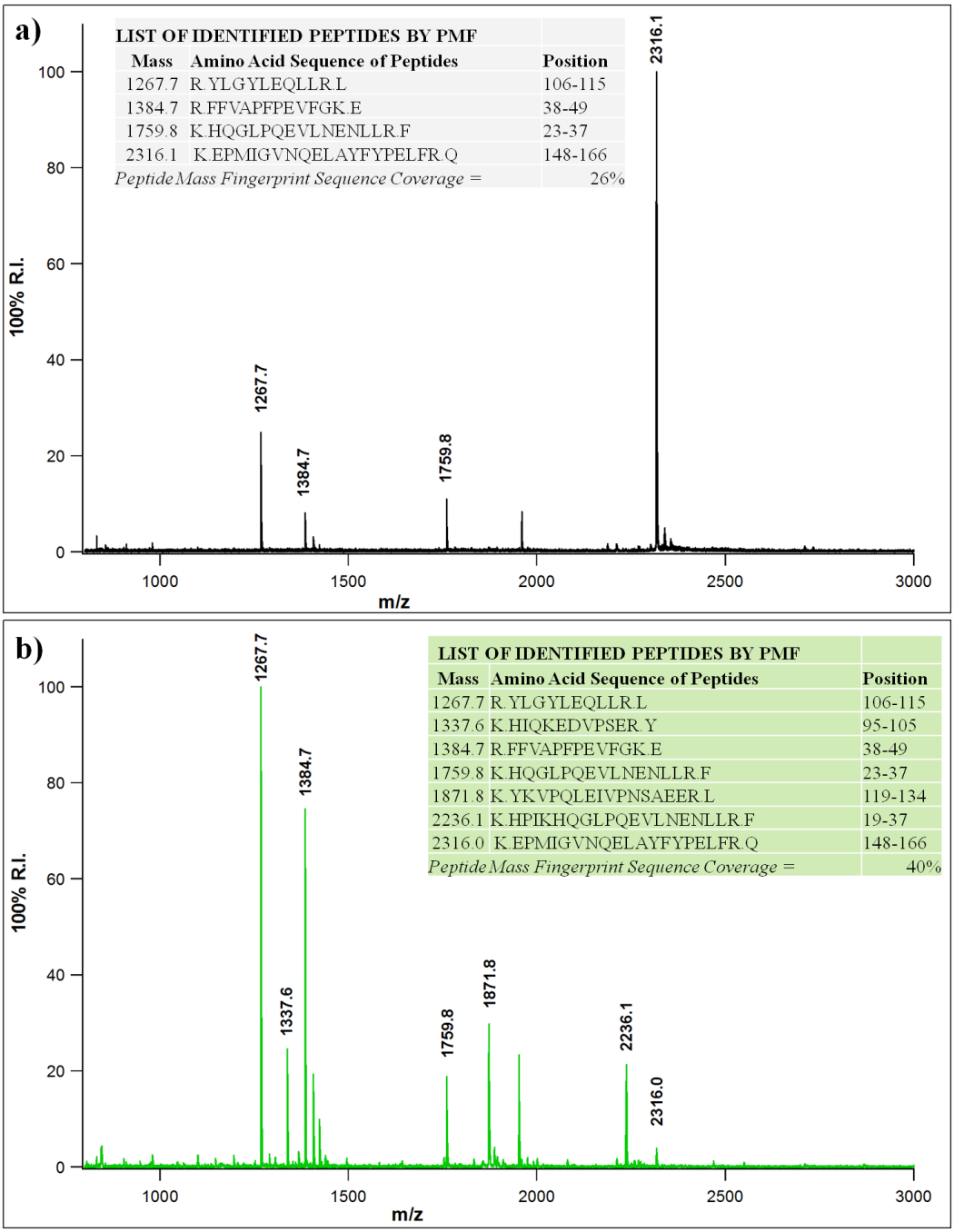


Figure 3.4. Mass spectra of Alpha-S1-casein digestions a) in-solution b) on-SCX

Table 3.2. Amino acid sequence of *Bos Taurus* Serum albumin protein

1	MKWVTFISLLLLFSSAYSRGVFRR**DTHKSEIAHRFKDLGEEHFKGLVLIAFSQYLQQCPF**
61	**DEHVKLVNELTEFAK**TCVADESHAGCEKSLHTLFGDELCKVASLRETYGDMADCCEKQEP
121	ERNECFLSHKDDSPDLPKLKPDPNTLCDEFKADEKKFWGK**YLYEIARRHPYFYAPELLYY**
181	**ANK**YNGVFQECCQAEDKGACLLPKIETMREKVLASSARQRLRCASIQKFGER**ALKAWSVA**
241	**RLSQK**FPKAEFVEVTKLVTDLTKVHKECCHGDLLECADDRADLAKYICDNQDTISSKLKE
301	CCDKPLLEKSHCIAEVEKDAIPENLPPLTADFAEDKDVCK**NYQEAKDAFLGSFLYEYSRR**
361	**HPEYAVSVLLR**LAKEYEATLEECCAKDDPHACYSTVFDKLKHLVDEPQNLIKQNCDQFEK
421	**LGEYGFQNALIVR**YTRK**VPQVSTPTLVEVSR**SLGKVGTRCCTKPESERMPCTEDYLSLIL
481	NRLCVLHEKTPVSEKVTKCCTESLVNRRPCFSALTPDETYVPKAFDEKLFTFHADICTLP
541	DTEKQIKKQTALVELLKHKPKATEEQLKTVMENFVAFVDKCCAADDKEACFAVEGPKLVV
601	STQTALA

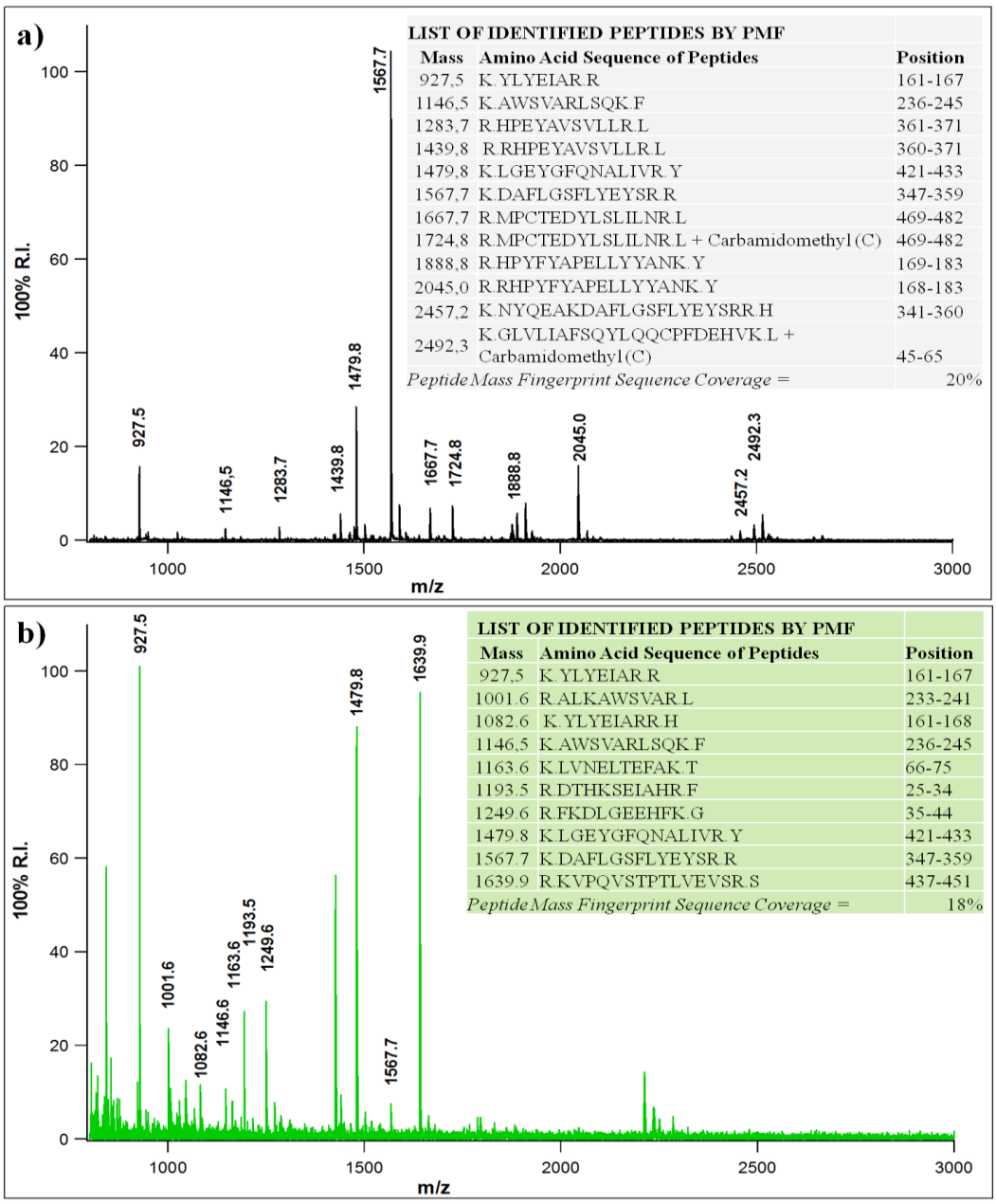


Figure 3.5. Mass spectra of albumin digestions a) in-solution b) on-SCX

Table 3.3. Amino acid sequence of *Gallus gallus* Lysozyme C protein

1	MRSLLILVLCFLPLAALGKVFGRCELAAAMKR**HGLDNYR**GYSLGNWVCAAK**FESNFNTQA**
61	**TNRNTDGSTDYGILQINSRWWCNDGR**TPGSRNLCNIPCSALLSSDITASVNCAK**KIVSDG**
121	**NGMNAWVAWRNRCKGTDVQAWIRGCR**L

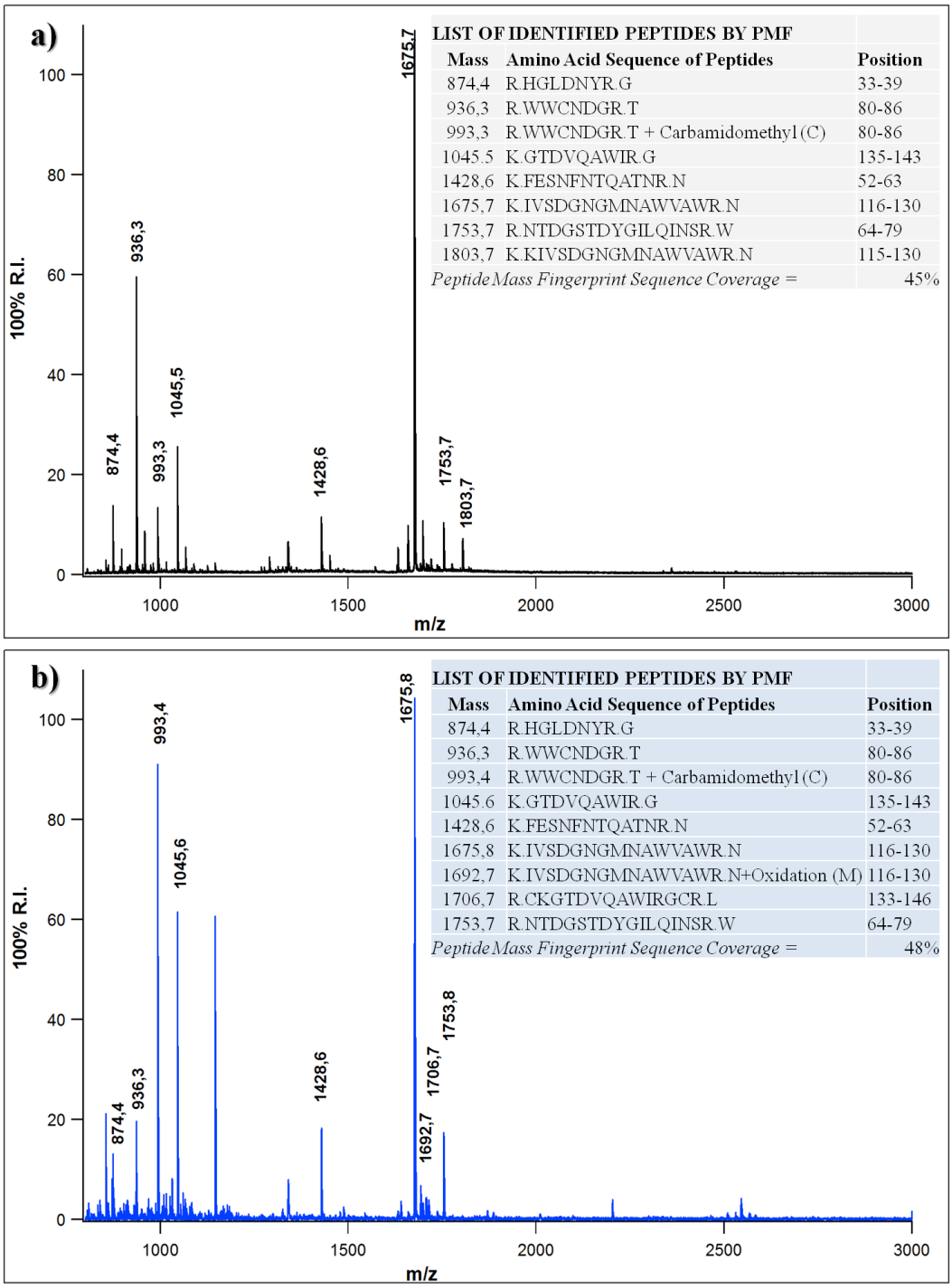


Figure 3.6. Mass spectra of lysozyme c digestions a) in-solution b) on-SAX

Table 3.4. Amino acid sequence of *Bos Taurus* Carbonic anhydrase protein

1	MSHHWGYGK**HNGPEHWHKDFPIANGER**QSPVDIDTK**AVVQDPALKPLALVYGEATSR**RMV
61	NNGHSFNVEYDDSQDKAVLK**DGPLTGTYRLVQFHFHWGSSDDQGSEHTVDR**KKYAAELHL
121	VHWNTKYGDFGTAAQQPDGLAVVGVFLK**VGDANPALQKVLDALDSIK**TKGKSTDFPNFDP
181	GSLLPNVLDYWTYPGSLTTPPLLESVTWIVLKEPISVSSQQMLKFR**TLNFNAEGEPELLM**
241	**LANWRPAQPLK**NR**QVRGFPK**

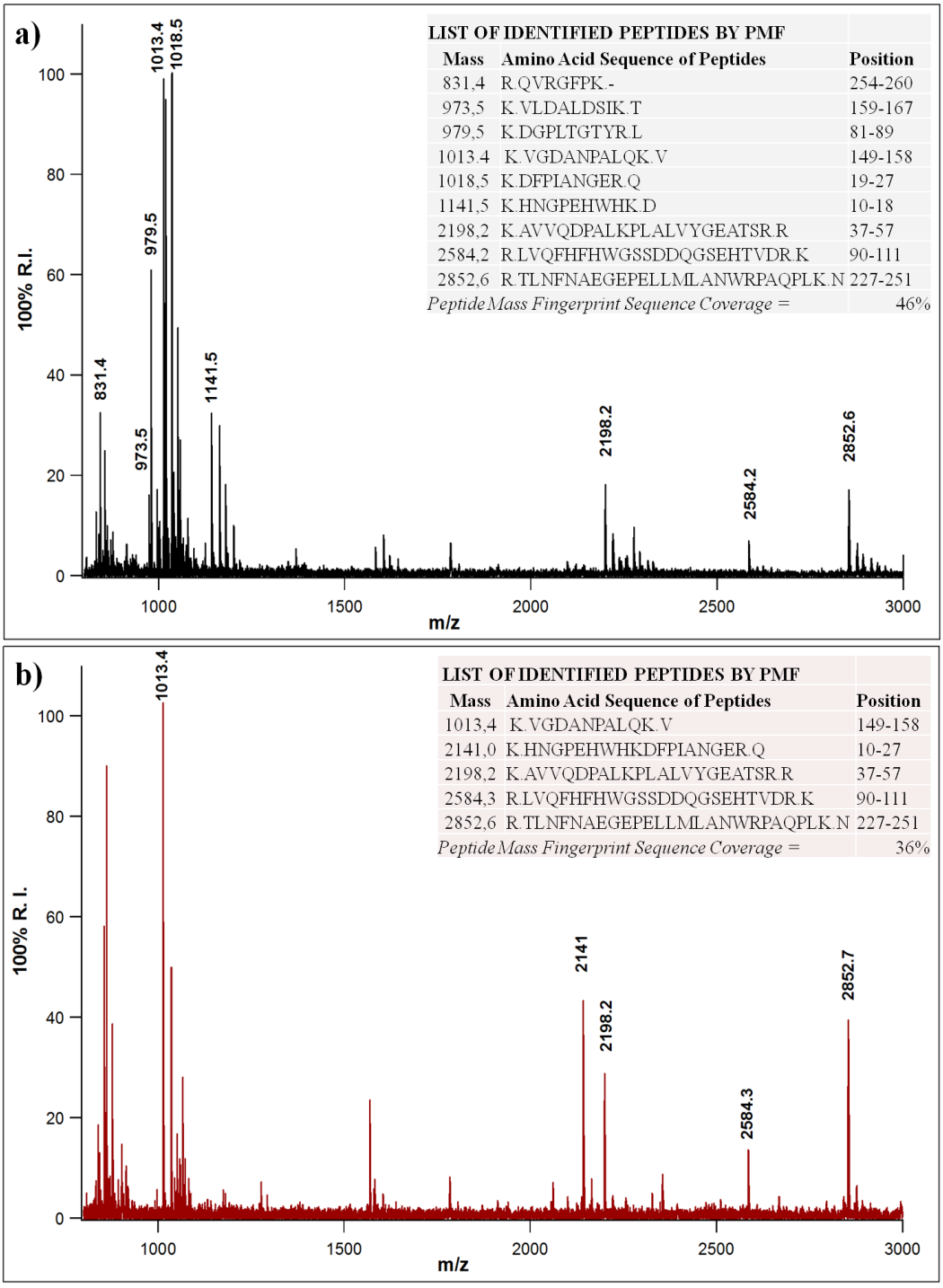


Figure 3.7. Mass spectra of carbonic anhydrase digestions a) in-solution b) on-C8

3.3. Comparison of On-Surface Digestion with In-Gel Digestion

Although SDS-PAGE electrophoresis is an indispensable method for protein separation, yield of protein recovery from gel is generally too low for the further examinations. Nevertheless proteins are being extracted from the gel either by electroelution or passive elution methods. On the other hand, proteins can also be digested while they were trapped in the gel by in-gel digestion method. This digestion method has gained massive success and widespread usage especially for the identification of proteins separated with 2D-SDS PAGE. However trypsin penetration into the gel to reach the protein and the pore size of the gel to retrieve the peptides are considered as the two main limitations of this technique.

In this section, a considerably small protein cytochrome c, which was extracted from polyacrylamide gel to fairly acidic solution by passive elution, was used to compare in-gel digestion with on-surface digestion. Firstly proteins were analyzed while they were in water by MALDI-TOF MS to observe the signals of proteins at a much higher mass range. On Figure 3.8., mass analysis of some model proteins are demonstrated. Although gel-extracted proteins, alpha-S1-casein and myoglobin were analyzed barely by MALDI-TOF MS at high mass range, only the peptides of cytochrome c protein were detected as adsorbed on sorbent beads before the tryptic digestion. MS result of on-surface digestion of gel-extracted cytochrome c was compared with the MS result of in-gel digestion. In this example, protein was extracted from gel to a present solvent mixture: formic acid-isopropanol-water. Protein in this highly acidic solution can not be digested directly without increasing the pH because trypsin does not active at low pH. However increasing the pH by adding basic solution would increase the volume and salt concentration. In conclusion cytochrome c was concentrated on the beads and at the same time removed from acidic solution by washing step. Apart from this, new peaks corresponding to cytochrome c peptides were observed at rather high mass range, over 1400 Da as can be seen on the Figure 3.11. These new peptides, which help to strengthen the identification of the protein by increasing the sequence coverage, are generally resulted from the peptides miscleaved by trypsin.

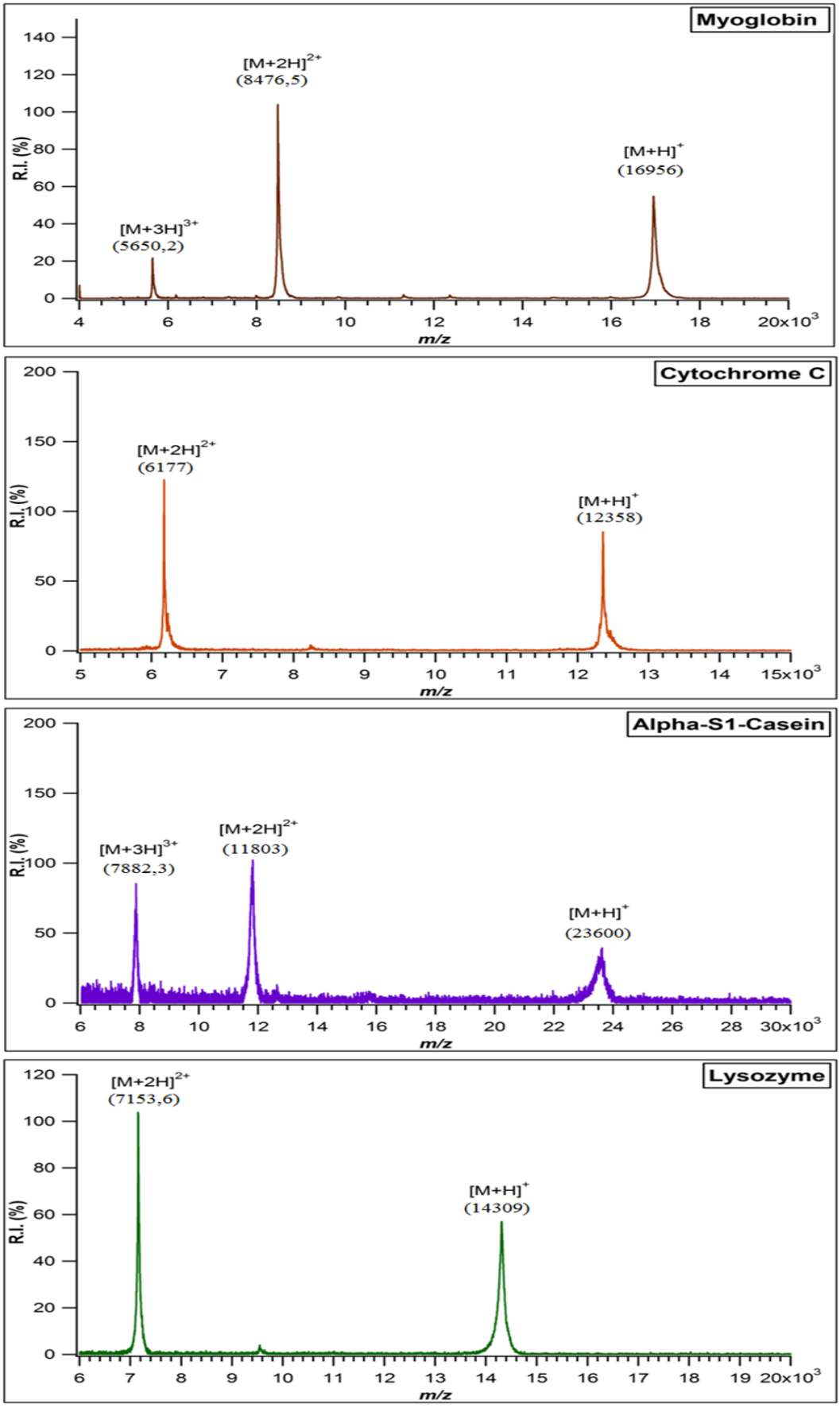


Figure 3.8. Mass spectrum of some model proteins

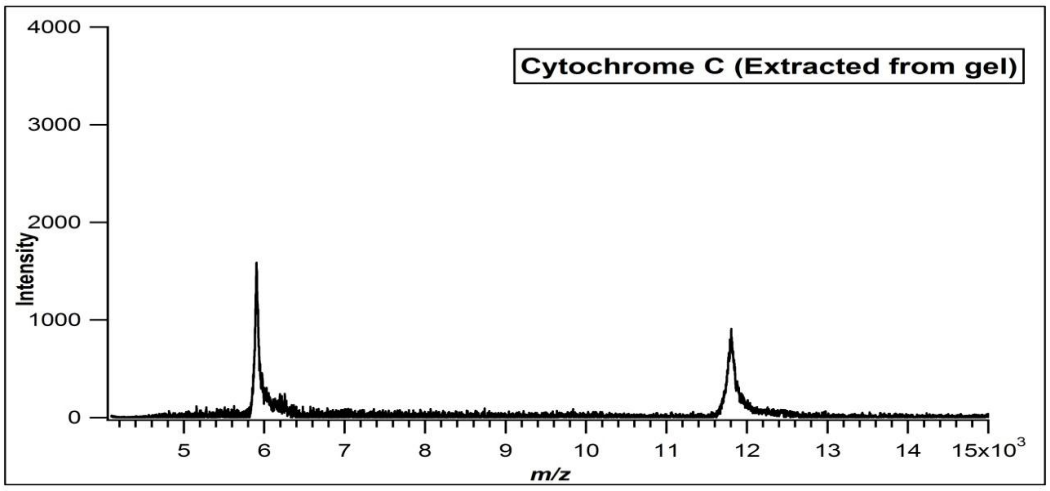


Figure 3.9. Mass spectrum of cytochrome c protein extracted from PA gel

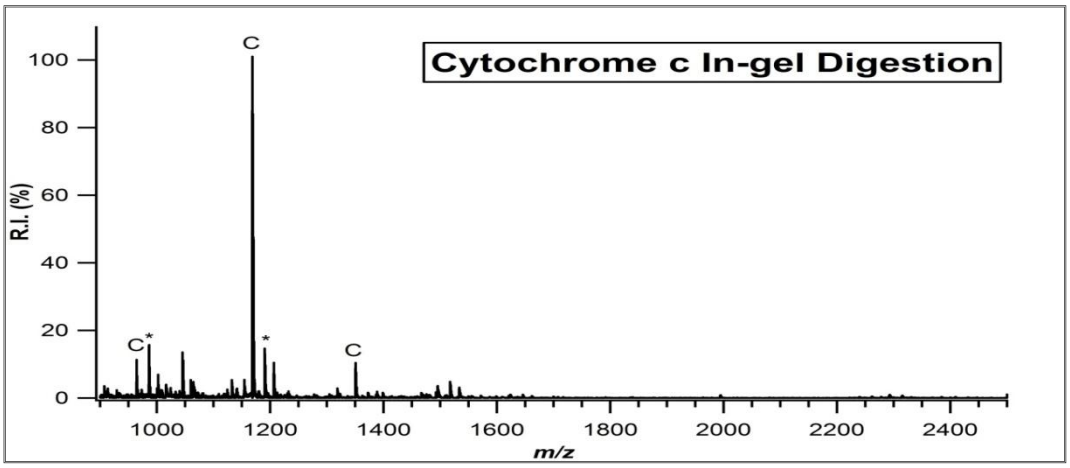


Figure 3.10. Mass spectrum of in-gel digested cytochrome c

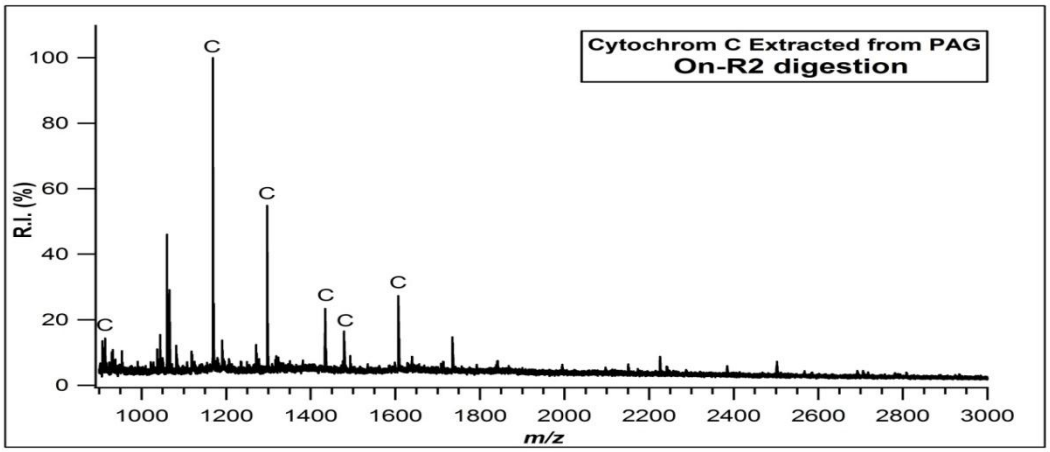


Figure 3.11. Mass spectrum of on-surface digested cytochrome c extracted from PA gel

3.4. On-Surface Digestion of Protein Mixture

On-surface digestion was also tested on protein mixture. Protein mixture containing six protein was digested on five different sorbents and compared with in-solution digestion. Peptides assigned to a protein by Mascot search are differentiated by different colours on the spectra. According to the general acceptance, at least two peptides are necessary to declare a protein identified. When we look at the spectrum of in-solution digestion of protein solution, we see that only one peptide was assigned to myoglobin and cytochrome c. Therefore these proteins' identifications are less reliable than the others (BSA by six, casein by four, lysozyme by three and carbonic anhydrase by two peptides). In terms of protein coverage, in-solution digestion alone gave the best result because none of the sorbents alone sufficed to identify all the proteins in the mixture however on-surface digestion totally increased the sequence coverage of some proteins and protein coverage. Besides one peptide was assigned to a new protein; alpha-S2-casein. Generally this low abundant protein is suppressed by alpha-S1 casein.

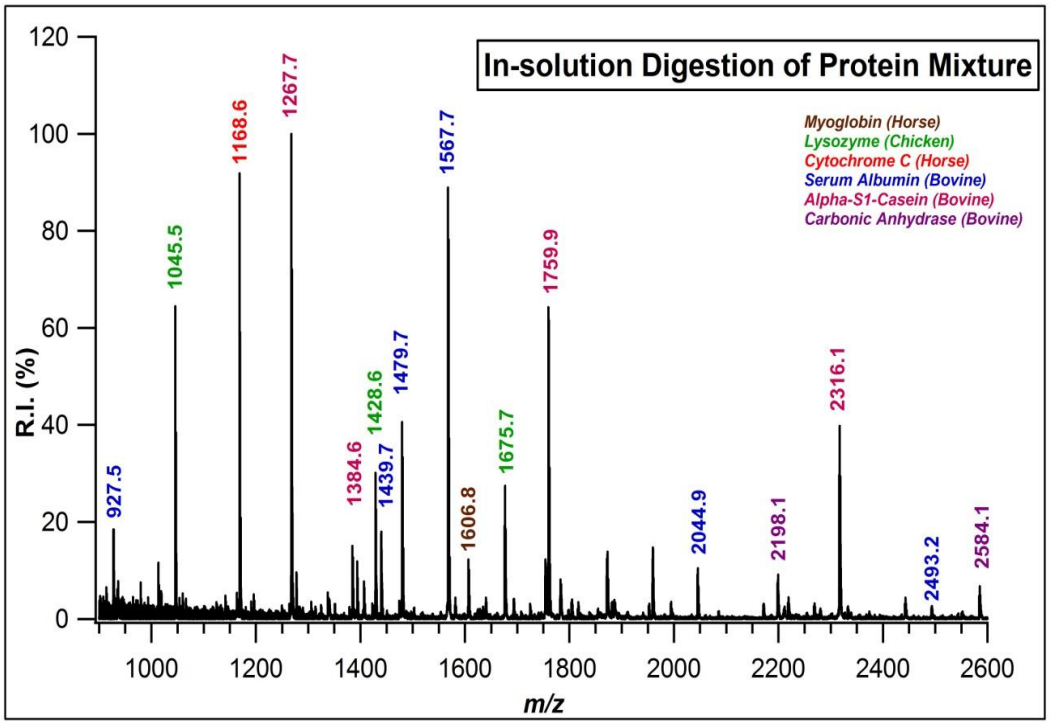


Figure 3.12. In-solution digestion of protein mixture

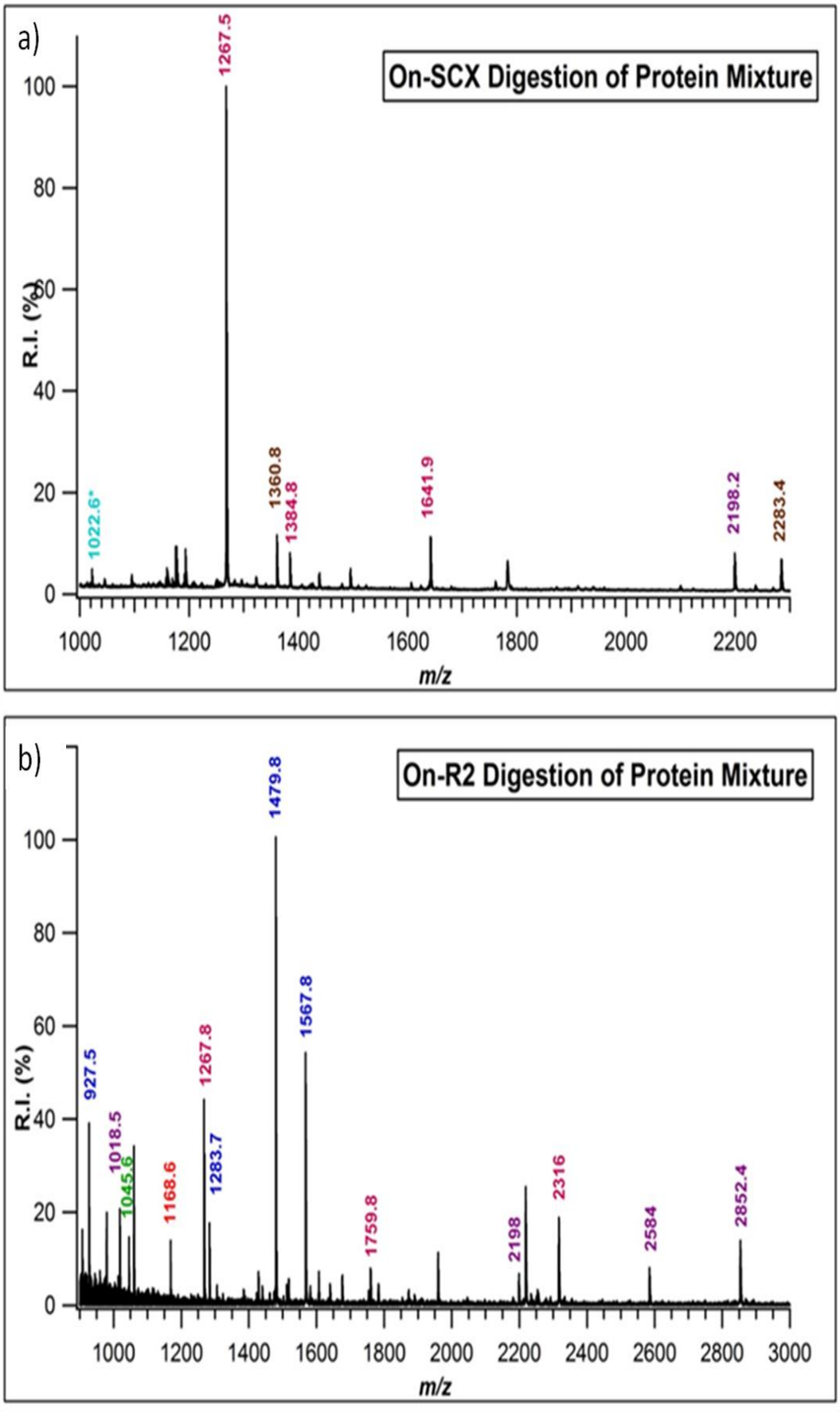


Figure 3.13. Digestion of protein mixture on different sorbents a) on-SCX b) on R2

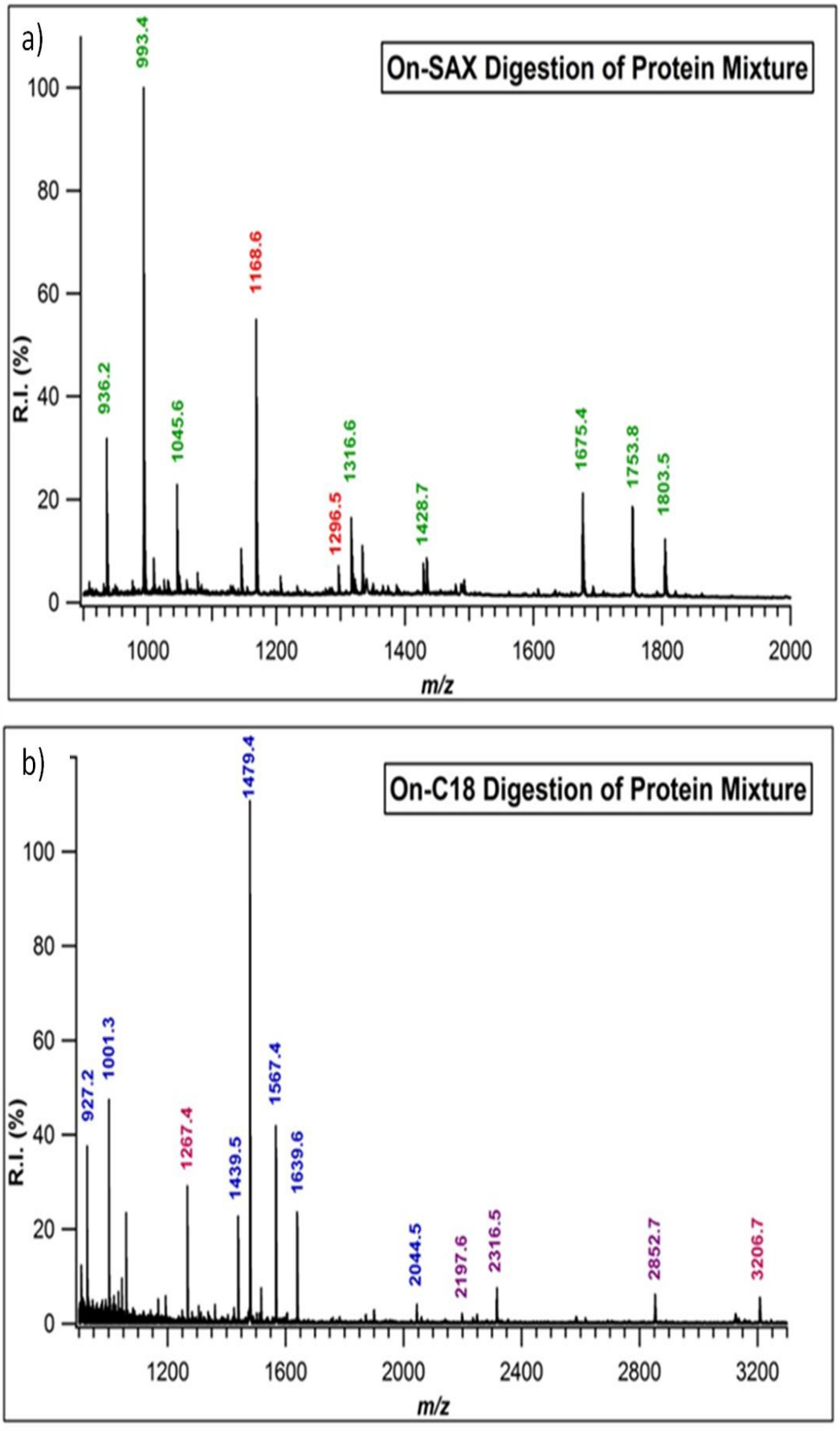


Figure 3.14. Digestion of protein mixture on different sorbents a) on-SAX b) on C18

Table 3.5. Identified peptides by MASCOT from in-solution digestion

m/z	Protein	Peptide Sequence
927.4730	Serum Albumin	K-YLYEIAR-R
1045.5350	Lysozyme	K-GTDVQAWIR-G
1168.5950	Cytochrome c	K-TGPNLHGLFGR-K
1267.6750	Alpha-S1-casein	R-YLGYLEQLLR-L
1384.6520	Alpha-S1-casein	R-FFVAPFPEVFGK-E
1439.7620	Serum Albumin	R-RHPEYAVSVLLR-L
1479.7320	Serum Albumin	K-LGEYGFQNALIVR-Y
1567.6790	Serum Albumin	K-DAFLGSFLYEYSR-R
1606.7910	Myoglobin	K-VEADIAGHGQEVLIR-L
1675.7130	Lysozyme	K-IVSNGNGMNAWVAWR-N
1759.8740	Alpha-S1-casein	K-HQGLPQEVLNENLLR-F
2044.9500	Serum Albumin	R-RHPYFYAPELLYYANK-Y
2198.1250	Carbonic Anhydrase 2	K-AVVQDPALKPLALVYGEATSR-R
2316.0681	Alpha-S1-casein	K-EPMIGVNQELAYFYPELFR-Q
2492.2129	Serum Albumin	K-GLVLIAFSQYLQQCPFDEHVK-L
2584.1260	Carbonic Anhydrase 2	R-LVQFHFHWGSSDDQGSEHTVDR-K

Table 3.6. Identified new peptides by MASCOT from on-sorbent digestion

m/z	Protein	Surface	Sequence
936,406	Lysozyme	SAX	R.WWCNDGR.T
993,434	Lysozyme	SAX	R.WWCNDGR.T + Carbamidomethyl (C)
1001,387	Serum albumin	C18	R.ALKAWSVAR.L
1018,502	Carbonic anhydrase 2	R2	K.DFPIANGER.Q
1022,625	Alpha-S2-casein	SCX	K.VIPYVRYLK.
1283,700	Serum albumin	R2	R.HPEYAVSVLLR.L
1296,757	Cytochrome c	SAX	K.TGPNLHGLFGRK.T
1333,709	Lysozyme	SAX	R.CKGTDVQAWIR.G + Carbamidomethyl
1360,777	Myoglobin	SCX	K.ALELFR NDIAAK.Y
1639,567	Serum albumin	C18	R.KVPQVSTPTLVEVSR.S
1640,888	Alpha-S1-casein	SCX	R.FFVAPFPEVFGKEK.V
1753,860	Lysozyme	SAX	R.NTDGSTDYGILQINSR.W
1803,900	Lysozyme	SAX	K.KIVSDGNGMNAWVAWR.N
2284,163	Myoglobin	SCX	K.ALELFR NDIAAKYKELGFQG
2851,729	Carbonic anhydrase 2	C18	R.TLNFNAEGEPELLMLANWRPAQPLK.N
2852,400	Carbonic anhydrase 2	R2	R.TLNFNAEGEPELLMLANWRPAQPLK.N
3206,800	Alpha-S1-casein	C18	K.EGIHAQQKEPMIGVNQELAYFYPELFR.Q

CHAPTER 4

CONCLUSION

In this study, an alternative protein digestion method namely on-surface digestion or on-bead digestion, which is already proposed but occasionally in use, was examined comprehensively by using different type of sorbent materials on various proteins with different physicochemical properties. Firstly, protein adsorption to the sorbents in the solution environment was verified with Bradford assay by considering the absorbance decrease after sorbent addition. Adsorption performance of the sorbents was pretty changeable from protein to protein. As a rough conclusion, hydrophobic sorbent having rather short hydrocarbon chain tends to adsorb the larger proteins whereas longed-chain hydrophobic sorbents like C18 is able to adsorb smaller proteins and larger ones as well. This result is consistent with general principles of HPLC since C18 columns are particularly preferred for the peptide and small proteins analysis. Although the selectivity in ionic sorbents was greater than hydrophobic surfaces, their adsorption was quite unpredictable in contrast to the expectations from acidic or basic proteins. It is probably arising from the localization of acidic and basic residues on the tertiary structure of the protein.

On-surface digestion was first tried on particular proteins individually and results were compared with traditional in-solution digestion. Although differences were observed on digestion patterns, some certain peptides were generally detected on the spectra of both methods. Similar experiments were repeated for the protein mixture including cytochrome c, lysozyme, myoglobin, carbonic anhydrase, bovine serum albumin and casein by using SAX, SCX, poros R2 and C18 sorbents. None of the sorbents alone sufficed to identify all the proteins as the in-solution digestion performed alone. On the other hand, on-surface digestion totally increased the sequence coverage of some proteins, especially the lysozyme's. In addition to this, detection of a peptide identifying the Alpha-S2-casein (one of the low abundant content of casein) only by on-SCX digestion, implies that protein-level separations before the digestion process may increase more or less both the sequence coverage of protein and protein coverage of the sample. Moreover, repeating peaks, which were obtained from different sorbents of the

on-surface digestion, enable scientist a second chance of the mass analysis for that ion. This feature is especially important for the MALDI because tandem mass spectrometric analysis of close signals is considerably difficult. Besides, matrix-sample co-crystallization might not form favorable enough for a successful analysis. Since protein adsorption is the first and key step for on-surface digestion, using more than one sorbent on the same sample consecutively will increase the chance of the adsorption especially for low abundant unknown proteins. Downsides of this method may be counted as the low protein adsorption and peptide recovery.

Last but not least, experimental setup and the devices were kept as simple as possible throughout this study by excluding the sophisticated and expensive HPLC system in order to address to any proteomics laboratory. In addition to this, need of ZipTip cleaning for the peptide recovery and desalting was lessened by using solvents easily evaporate prior to mass analysis.

REFERENCES

1. Montpetit, B., Proteomics in the Post-Genomics Era: Weighing in the Data. *BioTeach Journal* **2002,** *Vol. 1*.

2. Gygi, S. P.; Rochon, Y.; Franza, B. R.; Aebersold, R., Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **1999,** *19* (3), 1720-1730.

3. Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J. C.; Yan, J. X.; Gooley, A. A.; Hughes, G.; Humpery-Smith, I.; Williams, K. L.; Hochstrasser, D. F., From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Nature Biotechnology* **1996**.

4. Whitford, D., *PROTEINS Structure and Function*. John Wiley & Sons, Ltd: West Sussex, 2005; p 528.

5. Enrico Pieroni, S. d. l. F. v. B., Gianmaria Mancosu,; Enrico Capobianco, H. H., Alberto de la Fuente, Protein networking: insights into global functional organization of proteomes. *Proteomics* **2008**.

6. Wilkins, M. R.; Sanchez, J. C.; Gooley, A. A.; Appel, R. D.; HumpherySmith, I.; Hochstrasser, D. F.; Williams, K. L., Progress with proteome projects: Why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng* **1996,** *13*, 19-50.

7. Klein, J. B.; Thongboonkerd, V., Overview of proteomics. *Contrib.Nephrol.* **2004,** *141*, 1-10.

8. (a) Proteinstructure http://www.macalester.edu/psychology/whathap/UBNRP/tse10/prionproteins.html ; (b) Folded-proteins http://images.nigms.nih.gov/index.cfm?event=viewDetail&imageID=2367&add=2367 ; (c) Proteininteractionsmap http://www.sciencephoto.com/media/210725/view.

9. Pehr, E., Method for determination of the amino acid sequence in peptides. *acta chemica scandinavica* **1950**.

10. Meyer, B.; Papasotiriou, D.; Karas, M., 100% protein sequence coverage: a modern form of surrealism in proteomics. *Amino Acids* **2011,** *41* (2), 291-310.

11. Khan, M. F.; Bennett, M. J.; Jumper, C. C.; Percy, A. J.; Silva, L. P.; Schriemer, D. C., Proteomics by mass spectrometry—Go big or go home? *Journal of Pharmaceutical and Biomedical Analysis* **2011,** *55* (4), 832-841.

12. Pierre Legrain, R. A., Alexander Archakov, Amos Bairoch, Kumar Bala, Laura

Beretta, John Bergeron, Christoph H. Borchers, Garry L. Corthals, Catherine E. Costello, Eric W. Deutsch, Bruno Domon, William Hancock, Fuchu He, Denis Hochstrasser, György Marko-Varga, Ghasem Hosseini Salekdeh, Salvatore Sechi, Michael Snyder, Sudhir Srivastavak, Mathias Uhlen, Cathy H. Wu, Tadashi Yamamoto, Young-Ki Paiko and Gilbert S. Omenn, The Human Proteome Project: Current State and Future Direction. *Molecular & Cellular Proteomics* **2011,** *10.7*.

13. Witze, E. S.; Old, W. M.; Resing, K. A.; Ahn, N. G., Mapping protein post-translational modifications with mass spectrometry. *Nat Methods* **2007,** *4* (10), 798-806.

14. Cravatt Benjamin F., S. G. M., Yates III John R., The biological impact of mass-spectrometry-based proteomics. *Nature* **2007,** *450*.

15. Douglas A Jeffery, M. B., Chemical proteomics and its application to drug discovery. *Curretn Opinion in Biotechnology* **2003**.

16. Gramolini AO, P. S., Kislinger T, Mass Spectrometry-Based Proteomics: A Useful Tool for Biomarker Discovery? *Clinical Pharmacology & Therapeutics, Nature* **2008**.

17. Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* **2007,** *389* (4), 1017-1031.

18. Edmond de Hoffmann, V. S., *Mass Spectrometry: Principles and Applications*. 3rd Edition ed.; WILEY: 2007.

19. Sven Schuchardt, A. S., Protein identification using mass spectrometry:A method overview. *Plant Systems Biology* **2007**.

20. Steven L. Cohen, B. T. C., Influence of Matrix Solution Conditions on the MALDI-MS Analysis of Peptides and Proteins. *Anal. Chem* **1996**.

21. Zenobi, R.; Knochenmuss, R., Ion formation in MALDI mass spectrometry . *Mass Spectrom. Rev.* **1998,** *17* (5), 337-366.

22. Liuni, P.; Wilson, D. J., Understanding and optimizing electrospray ionization techniques for proteomic analysis. *Expert Review of Proteomics* **2011,** *8* (2), 197-209.

23. Pol, J.; Strohalm, M.; Havlicek, V.; Volny, M., Molecular mass spectrometry imaging in biomedical and life science research. *Histochem. Cell Biol.* **2010,** *134* (5), 423-443.

24. Yates, J. R.; Ruse, C. I.; Nakorchevsky, A., Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. In *Annual Review of Biomedical Engineering*, Annual Reviews: Palo Alto, 2009; Vol. 11, pp 49-79.

25. Badman, E. R.; Cooks, R. G., Special feature: Perspective - Miniature mass

analyzers. *J. Mass Spectrom.* **2000,** *35* (6), 659-671.

26. Jonscher, K. R.; Yates, J. R., The quadrupole ion trap mass spectrometer - A small solution to a big challenge. *Anal. Biochem.* **1997,** *244* (1), 1-15.

27. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S., Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* **1998,** *17* (1), 1-35.

28. Perry, R. H. C., R. G. Noll, R. J., Orbitrap Mass Spectrometry: Instrumentation, Ion Motion and Applications. *Mass Spectrom. Rev.* **2008,** *27* (6), 661-699.

29. Hager, J. W., Recent trends in mass spectrometer development. *Anal. Bioanal. Chem.* **2004,** *378* (4), 845-850.

30. Taras Stasyk, L. A. H., Zooming in: Fractionation strategies in proteomics. *Proteomics, WILEY-VCH Verlag GmbH & Co. KGaA* **2004**.

31. Boyer, R., *Modern Experimental Biochemistry*. 3. ed.; Addison Wesley Longman: 2000.

32. Aguilar, M. I., *HPLC of Peptides and Proteins Methods and Protocols*. Humana Press: 2004; Vol. 251.

33. Yang Shi, R. X., Csaba Horvath, James A. Wilkins, The role of liquid chromatography in proteomics. *J. Chromatogr. A* **2004**.

34. (a) SCIENCE, I. H. a. http://www.idex-hs.com/support/upchurch/hplc_center.aspx; (b) David L. Nelson, M. M. C., *Lehninger PRINCIPLES OF BIOCHEMISTRY*. Fourth ed.

35. Thierry Rabilloud, A. R. V., Noelle Potier,Ce´cile Lelong, Emmanuelle Leize-Wagner, Mireille Chevallet, Power and limitations of electrophoretic separations in proteomics *Mass Spectrom. Rev.* **2008**.

36. Chait, B. T., Mass Spectrometry: Bottom-Up or Top-Down? *Science* **2006,** *314* (5796), 65-66.

37. VerBerkmoes, N. C.; Bundy, J. L.; Hauser, L.; Asano, K. G.; Razumovskaya, J.; Larimer, F.; Hettich, R. L.; Stephenson, J. L., Integrating "top-down" and "bottom-up" mass spectrometric approaches for proteomic analysis of Shewanella oneidensis. *J. Proteome Res.* **2002,** *1* (3), 239-252.

38. Running, W. E.; Ravipaty, S.; Karty, J. A.; Reilly, J. P., A top-down/bottom-up study of the ribosomal proteins of Caulobacter crescentus. *J. Proteome Res.* **2007,** *6* (1), 337-347.

39. Armirotti, A.; Damonte, G., Achievements and perspectives of top-down proteomics. *Proteomics* **2010,** *10* (20), 3566-3576.

40. (a) Christian H. Ahrens, E. B., Ermir Oeli, Konrad Basler, Ruedi Aebersold, Protein identification by shotgun proteomics and targeted proteomics approaches. *Nat. Rev. Mol. Cell Biol.* **2010**; (b) John F. Kellie, J. C. T., Ji Eun Lee, Dorothy R. Ahlf, Haylee M. Thomas, Ioanna Ntai, Adam D. Catherman, Kenneth R. Durbin, Leonid Zamdborg, Adaikkalam Vellaichamy, Paul M. Thomas and Neil L. Kelleher, The emerging process of Top Down mass spectrometry for protein analysis: biomarkers, protein-therapeutics, and achieving high throughput. *Molecular Biosystems* **2010**.

41. Trypsin Swiss Institute of Bioinformatics. http://enzyme.expasy.org/EC/3.4.21.4.

42. Asp-N Swiss Institute of Bioinformatics. http://enzyme.expasy.org/EC/3.4.24.33.

43. Glu-C Swiss Institute of Bioinformatics. http://enzyme.expasy.org/EC/3.4.21.19.

44. Lys-C Swiss Institute of Bioinformatics. http://enzyme.expasy.org/EC/3.4.21.50.

45. Chymotry Swiss Institute of Bioinformatics . http://enzyme.expasy.org/EC/3.4.21.1.

46. Pepsin Swiss Institute of Bioinformatics. http://enzyme.expasy.org/EC/3.4.23.1.

47. Wolters, D. A.; Washburn, M. P.; Yates, J. R., An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **2001,** *73* (23), 5683-5690.

48. Hanne Kolsrud Hustoft, H. M., Steven Ray Wilson, Leon Reubsaet, Elsa Lundanes, Tyge Greibrokk, *A Critical Review of Trypsin Digestion for LC-MS Based Proteomics*. InTech: Shanghai, China, 2012.

49. Hanne Kolsrud Hustoft, L. R., Tyge Greibrokk, Elsa Lundanes, Helle Malerod, Critical assessment of accelerating trypsination methods. *Journal of Pharmaceutical and Biomedical Analysis* **2011**.

50. Biosciences, A., Reversed Phase Chromatography Principles and Methods. **1999**.

51. (a) phenomenex http://www.phenomenex.com/Products/SPDetail/Strata/SCX; (b) phenomenex. http://www.phenomenex.com/Products/SPDetail/Strata/SAX.

52. Thiede, B.; Höhenwarter, W.; Krah, A.; Mattow, J.; Schmid, M.; Schmidt, F.; Jungblut, P. R., Peptide mass fingerprinting. *Methods* **2005,** *35* (3), 237-247.

53. Cottrell, J. S., Protein identification using MS/MS data. *Journal of Proteomics* **2011,** *74* (10), 1842-1851.

54. Matthiesen, R., *Mass Spectrometry Data Analysis in Proteomics*. Humana Press

Inc.: 2007.

55. Jones, A. W.; Cooper, H. J., Dissociation techniques in mass spectrometry-based proteomics. *Analyst* **2011,** *136* (17), 3419-3429.

56. Steen, H.; Mann, M., The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* **2004,** *5* (9), 699-711.

57. Nesvizhskii, A. I.; Vitek, O.; Aebersold, R., Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat. Methods* **2007,** *4* (10), 787-797.

58. Standing, K. G., Peptide and protein de novo sequencing by mass spectrometry . *Curr. Opin. Struct. Biol.* **2003,** *13* (5), 595-601.

59. Allmer, J., Algorithms for the de novo sequencing of peptides from tandem mass spectra. *Expert Review of Proteomics* **2011,** *8* (5), 645-657.

60. Mann, M. W., M., Error Tolerant Identification of Peptides in Sequence Databases by Peptide Sequence Tags. *Anal. Chem.* **1994,** *66* (24), 4390-4399.

61. Marc R. Wilkins, K. L. W., Ron D. Appel, Denis F. Hochstrasser, *Proteome Research Concepts, Technology and Application*. Second ed.; Springer: 2007.

62. Eng, J. K.; Searle, B. C.; Clauser, K. R.; Tabb, D. L., A Face in the Crowd: Recognizing Peptides Through Database Search. *Molecular & Cellular Proteomics* **2011,** *10* (11).

63. Kapp, E. A.; Schutz, F.; Connolly, L. M.; Chakel, J. A.; Meza, J. E.; Miller, C . A.; Fenyo, D.; Eng, J. K.; Adkins, J. N.; Omenn, G. S.; Simpson, R. J., An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: Sensitivity and specificity analysis. *Proteomics* **2005,** *5* (13), 3475-3490.

64. Nesvizhskii, A. I.; Aebersold, R., Interpretation of shotgun proteomic data - The protein inference problem. *Molecular & Cellular Proteomics* **2005,** *4* (10), 1419-1440.

65. Nesvizhskii, A. I., Protein Identification by Tandem Mass Spectrometry and Sequence Database Searching. In *Mass Spectrometry Data Analysis in Proteomics*, Matthiesen, R., Ed. Humana Press: 2007.

66. Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999,** *20* (18), 3551-3567.

67. Kazuhiro Nakanishi, T. S., Koreyoshi Imamura, On the Adsorption of Proteins on Solid Surfaces, a Common but Very Complicated Phenomenon. *Journal of Bioscience and Bioengineering* **2001**.

68. Marie-Isabel Aguilar, D. J. C., Phillip Holt, Veronica Kronina, Reinhard I.

Boysen, Anthony W. Purcell, Milton T. W. Hearn, RP-HPLC Binding Domains of Proteins. *Anal. Chem.* **1998**.

69. Alan Doucette, D. C., Liang Li, Protein Concentration and Enzyme Digestion on Microbeads for MALDI-TOF Peptide Mass Mapping of Proteins from Dilute Solutions. *Anal. Chem.* **2000,** *72*.

70. Alan Doucette, D. C., Liang Li, Mass Spectrometric Study of the Effects of Hydrophobic Surface Chemistry and Morphology on the Digestion of Surface-Bound Proteins. *American Society for Mass Spectrometry* **2003**.

71. Martin Ethier, W. H., Henry S. Duewel, Daniel Figeys, The Proteomic Reactor : A Microfluidic Device for Processing Minute Amounts of Protein Prior to Mass Spectrometry Analysis. *J. Proteome Res.* **2006**.

72. Hu Zhou, W. H., Jean-Philippe Lambert, Daniel Figeys, New ammunition forthe proteomic reactor: strong anion exchange beads and multiple enzymes enhance protein identification and sequence coverage. *Anal Bioanal Chem* **2010**.

APPENDIX A

PROTEIN STANDART CURVE AND PREPARATION OF BRADFORD REAGENT

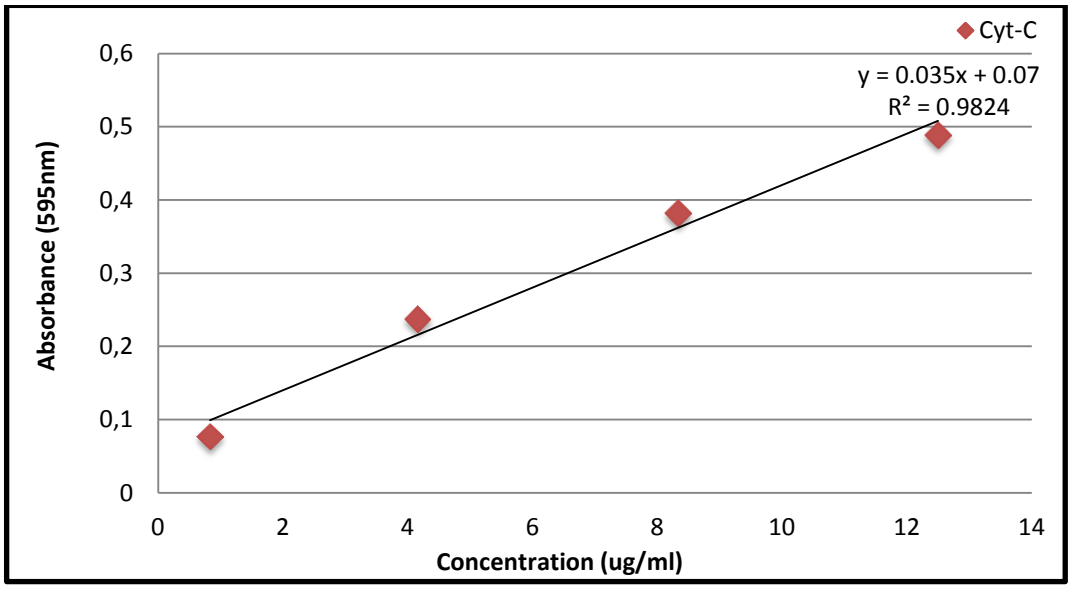


Figure1 : Standard curve for Bradford Assay

Preparation of Bradford Reagent

• 10.0 mg Coomassie Brilliant Blue G-250 (CBB G-250)

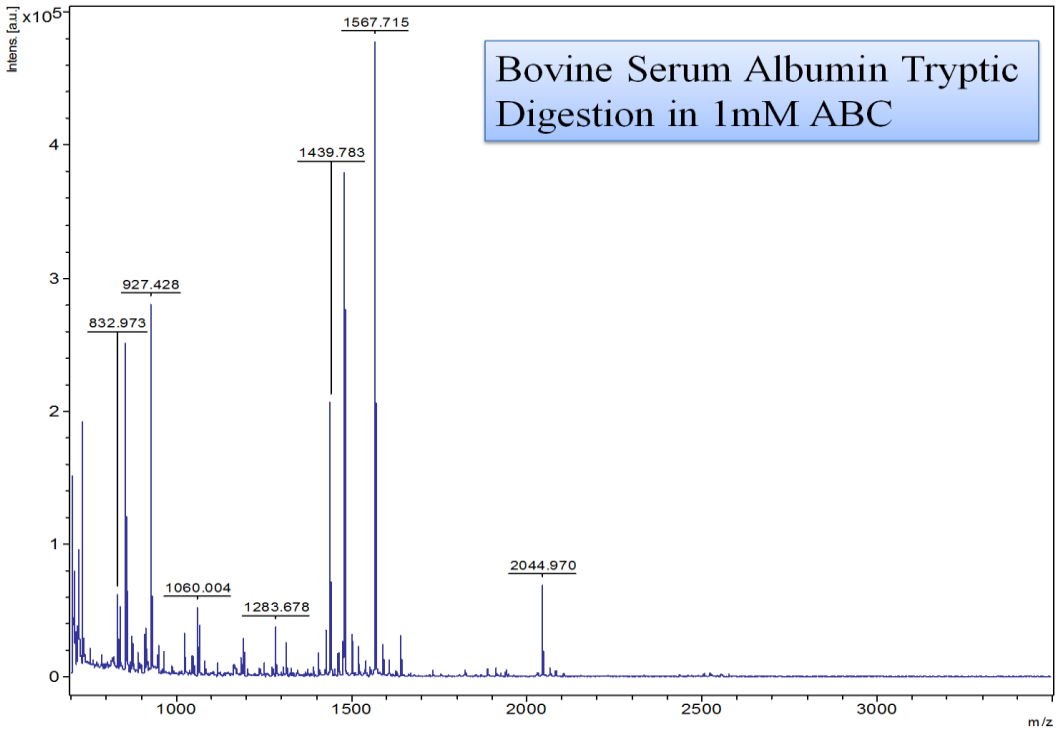
• 5ml 95% ethanol

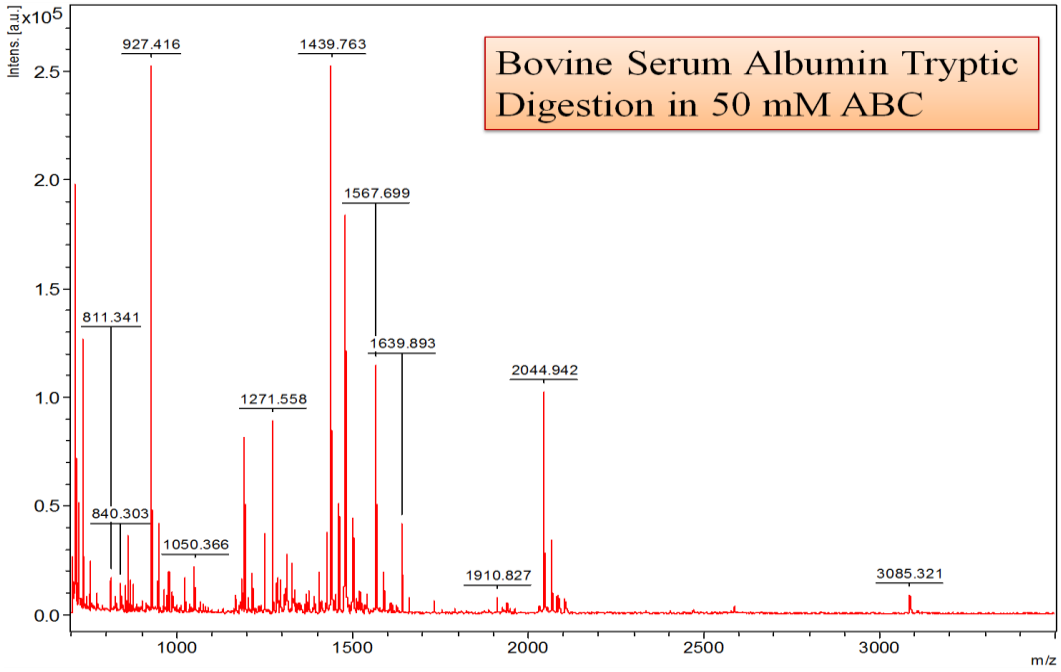
• 10ml 85% phosphoric acid

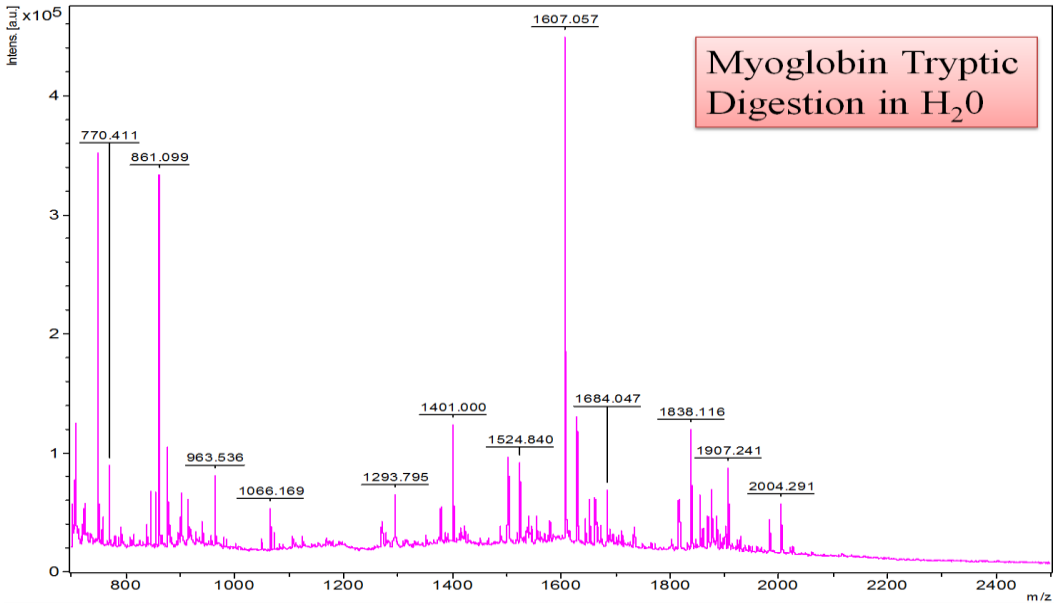
10.0 mg of CBBG-250 is dissolved in ethanol. Then 10 ml ortho phosphoric acid is added. Volume is adjusted to 100 ml with ultra pure water. When the dye was completely dissolved, solution is filtered with Whatman No. 1 paper. It is stored at 4 °C.

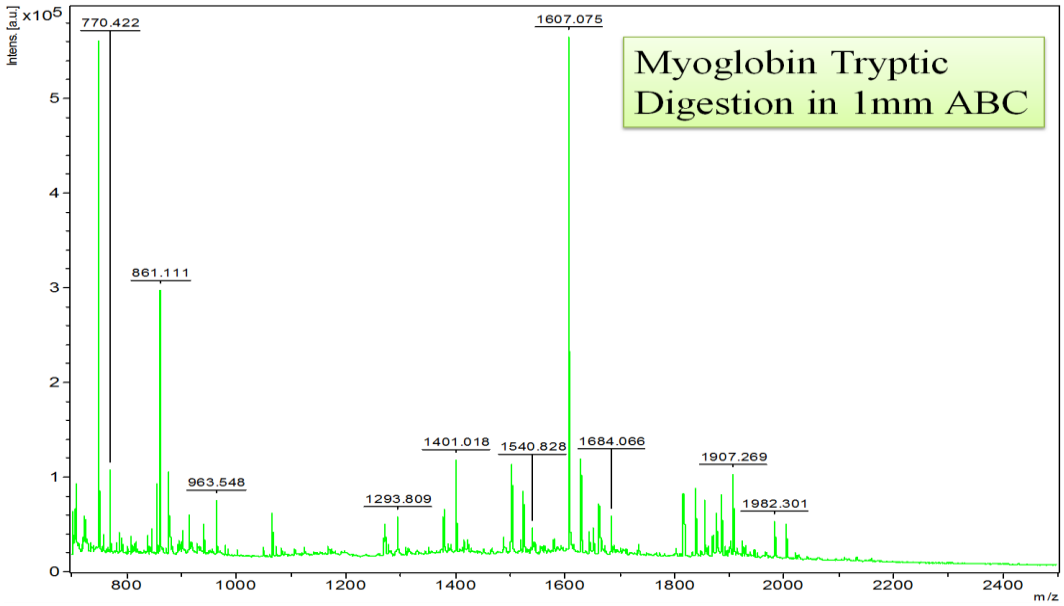
APPENDIX B

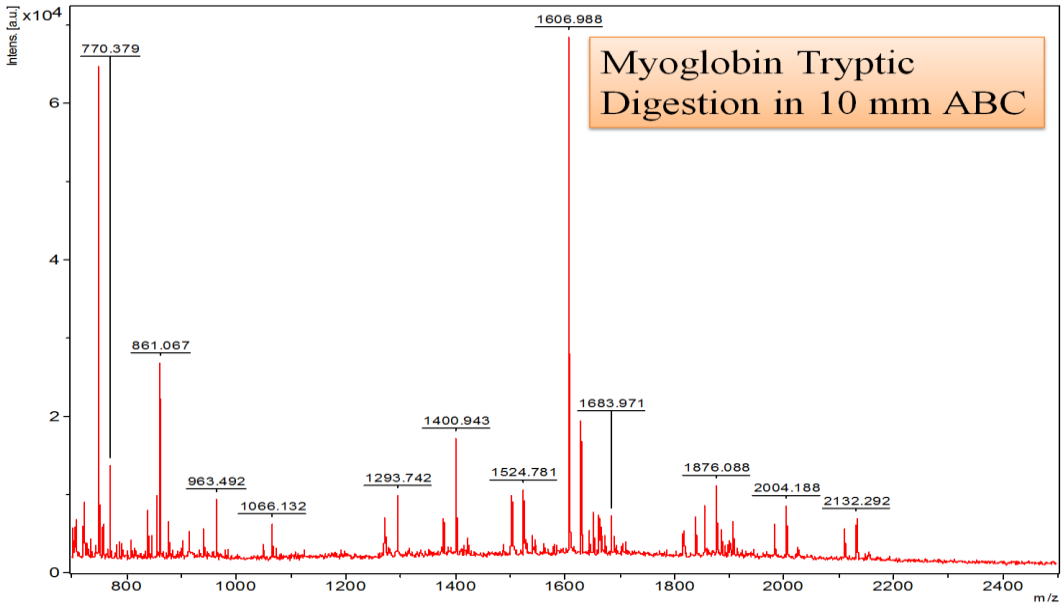
BUFFER EFFECT TO TRYPTIC DIGESTION






Intens. [a.u.]
x10^5
Myoglobin Tryptic Digestion in H20
1607.057
770.411
861.099
1684.047
1838.116
1401.000
1524.840
1907.241
963.536
1293.795
1066.169
2004.291
m/z


Intens. [a.u.]
x10^5
Myoglobin Tryptic Digestion in 1mm ABC
770.422
1607.075
861.111
1401.018
1540.828
1684.066
1907.269
963.548
1293.809
1982.301
m/z


Intens. [a.u.]
x10^4
Myoglobin Tryptic Digestion in 10 mm ABC
1606.988
770.379
861.067
1683.971
1400.943
1876.088
963.492
1293.742
1524.781
2004.188
2132.292
1066.132
m/z

APPENDIX C

MS/MS RESULTS OF ON-SURFACE DIGESTION

Mass	Met	Protein	Taxonomy	Sequence	% SC
927,2	C18	Serum albumin	Bos taurus	K.YLYEIAR.R	1
936,3	SAX	Lysozyme C	Gallus gallus	R.WWCNDGR.T	5
1001,4	C18	Serum albumin	Bos taurus	R.ALKAWSVAR.L	1
1018,5	R2	Carbonic anhydrase	Bos taurus	K.DFPIANGER.Q	3
1045,5	R2	Lysozyme C	Gallus gallus	K.GTDVQAWIR.G	6
1163,6	SAX	Albumin	Bos taurus	K.LVNELTEFAK.T	1
1168,6	R2	Cytochrome c	Equus caballus	K.TGPNLHGLFGR.K	10
1193,6	SAX	Albumin	Bos taurus	R.DTHKSEIAHR.F	1
1267,7	R2	Alpha-S1-casein	Bos taurus	R.YLGYLEQLLR.L	4
1283,7	R2	Serum albumin	Bos taurus	R.HPEYAVSVLLR.L	1
1337,6	SCX	Alpha-S1-casein	Bos taurus	K.HIQKEDVPSER.Y	5
1350,7	SCX	Cytochrome c	Equus caballus	K.TEREDLIAYLK.K	10
1384,6	SCX	Alpha-S1-casein	Bos taurus	R.FFVAPFPEVFGK.E	5
1428,6	SCX	Lysozyme	Gallus gallus	K.FESNFNTQATNR.N	8
1439	C8	Serum albumin	Bos taurus	R.RHPEYAVSVLLR.L	1
1439,4	C18	Serum albumin	Bos taurus	R.RHPEYAVSVLLR.L	1
1470,6	SCX	Cytochrome C	Equus caballus	K.TGQAPGFTYTDANK.N	13
1479	C8	Serum albumin	Bos taurus	K.LGEYGFQNALIVR.Y	2
1479,4	C18	Serum albumin	Bos taurus	K.LGEYGFQNALIVR.Y	2
1491,6	C8	Lysozyme	Gallus gallus	R.WWCNDGRTPGSR.N+ Carbamidomethyl (C)	8
1567,2	C8	Serum albumin	Bos taurus	K.DAFLGSFLYEYSR.R	2
1567,3	C18	Serum albumin	Bos taurus	K.DAFLGSFLYEYSR.R	2
1639,5	C18	Serum albumin	Bos taurus	R.KVPQVSTPTLVEVSR.S	2
1639,9	SCX	Albumin	Bos taurus	R.KVPQVSTPTLVEVSR.S	2
1675,7	SAX	Lysozyme C	Gallus gallus	R.IVSDGNGMNAWVAWR.N	10
1675,8	C8	Lysozyme	Gallus gallus	R.IVSDGNGMNAWVAWR.N	10
1753,8	C8	Lysozyme	Gallus gallus	R.NTDGSTDYGILQINSR.W	10
1759,8	SCX	Alpha-S1-casein	Bos taurus	K.HQGLPQEVLNENLLR.F	7
1803,8	C8	Lysozyme	Gallus gallus	K.KIVSDGNGMNAWVAWR.N	10
2044,4	C18	Serum albumin	Bos taurus	R.RHPYFYAPELLYYANK.Y	2
2197,9	C18	Carbonic anhydrase	Bos taurus	K.AVVQDPALKPLALVYGEAT SR.R	8
2199	C8	Carbonic anhydrase	Bos taurus	K.AVVQDPALKPLALVYGEAT SR.R	8
2315,5	C18	Alpha-S1-casein	Bos taurus	K.EPMIGVNQELAYFYPELFR.Q	8
2583,3	C8	Carbonic anhydrase	Bos taurus	R.LVQFHFHWGSSDDQGSEHTVDR.K	8
2851,5	C8	Carbonic anhydrase	Bos taurus	R.TLNFNAEGEPELLMLANWRPAQPLK.N	9
2851,8	C18	Carbonic anhydrase	Bos taurus	R.TLNFNAEGEPELLMLANWRPAQPLK.N	9
3206,8	C18	Alpha-S1-casein	Bos taurus	K.EPLPYLYRKPTVELLDLNTMEESSEIK.V	12