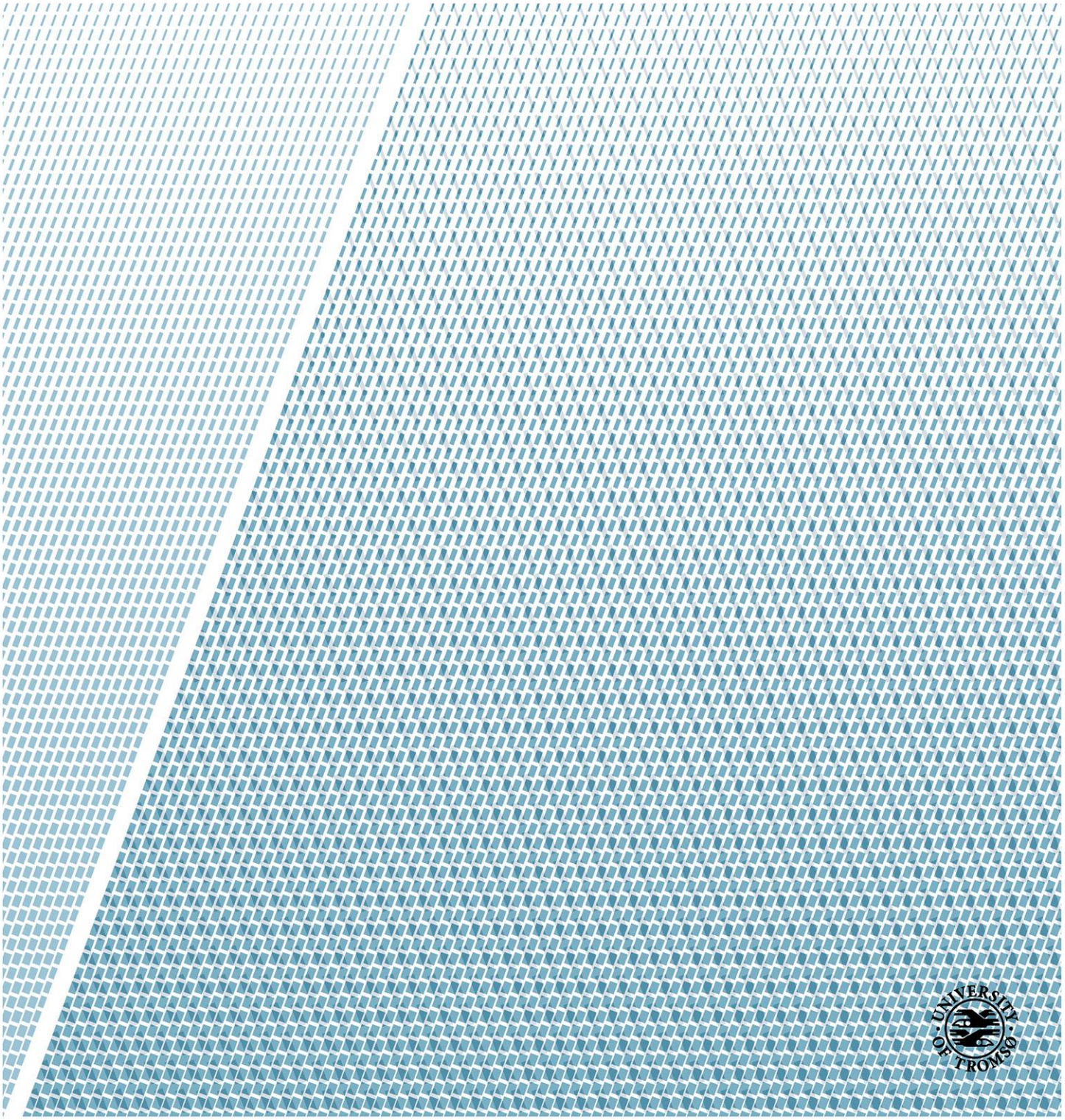
Natural Products and Medicinal Chemistry Research Group
Department of Pharmacy,
Faculty of Health Sciences
The Arctic University of Norway

Development of COFRADIC HPLC system for proteomics

Sorting of N-terminal peptides

Maryam Ghazi Ali Banki

Master thesis in Medicinal Biology / Analytical Chemistry … May 2014



PREFACE

This master`s thesis was completed at the Natural Products and Medicinal Chemistry Research Group, Department of Pharmacy, The Arctic University of Norway from August 2013 to June 2014. The supervisor for this thesis was Dr. Terkel Hansen and the co-supervisor was Dr. Jack Ansgar Bruun.

First, I would like to express my sincere gratitude to my supervisor Terkel Hansen for his support and guidance during the whole project. I very much appreciate your helpful advices and your support. I have to say that I learned many things from you and you were always there when I needed your help. It have been a year full of challenges, new experiences, new knowledge and a lot of joy. I am grateful to you, because without your guidance and support this work would not have been completed.

A special thanks to Toril Anne Grønset and Jack Ansgar Bruun for their hard work with the management of the data from mass spectrometric analysis.

Secondly, a huge and special thanks to my dear father and mother, and my siblings. Thank you for being there for me all the time and for believing in me and my skills. I would never have made it without your love and support.

Furthermore, I would like to thank my dear friends Limou and Gurjit for their support during difficult times and thank you for the laughter's we shared. It will be a memory for life.

Finally, I dedicate this master thesis to my dear brother. Even though you are gone, you will never be forgotten and thank you for giving me the strength to finish what I started.

ABSTRACT

Proteomics, the study of the proteins constituting the proteome, has featured in the past few years as an essential tool in a number of diverse areas of therapeutic research for detection of the initial stage of a disease, for consideration of therapy response and for disintegrating fundamental pathophysiological mechanisms. Proteomics has a wide range of applications such as diabetes research, neurology, nutrition research, diagnosis of renal disease, tumor metastasis, urological cancer research, cardiovascular disease, etc. This context makes proteomics a link between genomics and biology. Proteins can be subjected to posttranslational modifications (PTMs) processes. The results are that the kinds of measured proteins have inter – and intra individual variability. PTMs have a great influence on proteins in the cause of its activity dynamics regulation.

In recent years, COFRADIC has become a particular technique designed for peptide selection from complex mixtures. This method comprises of two successive equal chromatographic separations, including an alteration stage aimed to a subgroup of peptides between the two isolations. The altered peptides achieves various chromatographic characteristics and isolates from the unmodified peptides in the second run. The ultimate investigation depends on the isolated peptides, therefore it requires a decreased sample intricacy but preserved properties of the proteome.

In this study, we developed a RP-HPLC COFRADIC method in contemplation of identification of peptides from large-scale proteins. Our goal was to reach peptide separation by the HPLC system. Optimization of the method was challenging due to its implemented instrument method for both primary and secondary run in order to isolate N-terminal peptides by either acetylation or di-methylation modifications strategies. The Escherichia coli proteome was studied and we extracted proteins by several sample preparation procedures. The enrichment of identified N-terminal peptides extracted from *E. coli* bacteria through this study was 333 protein families identified by acetylation and 443 protein families identified by di-methylation modifications method. These results were obtained by utilizing a Thermo Scientific Q-Exactive mass spectrometer.

LIST OF ABBREVIATIONS

- AA – Amino Acid
- ACN - Acetonitrile
- BCA – Bicinchoninic Acid
- BSA – Bovine Serum Albumin
- CID – Collision Induced Dissociation
- COFRADIC – Combined Fractional Diagonal Chromatography
- DTT – Dithiothreitol
- ESI – Electrospray Ionization
- FA – Formic acid
- FASP – Filter Aided Sample Preparation
- HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
- IAA – Iodoacetamide
- MALDI – Matrix-Assisted Laser Desorption Ionization
- MeCN - Acetonitrile
- MS-MS – Tandem mass spectrometry
- PDA – Photodiode Array Detector
- PMF – Peptide Mass Fingerprinting
- PTM – Posttranslational modification
- Q-TOF – Quadropole-Time-Of-Flight hybrid
- RP-LC – Reversed-Phase Liquid Chromatography
- RT – Room Temperature
- SDS – Sodium dodecylsulfate
- SCX – Strong Cation-exchange Chromatography
- TFA – Trifluoroacetic Acid
- TNBS – 2,4,6-Trinitrobenzenesulfonic Acid
- UV – Ultraviolet detection
- UPLC – Ultra Performance Liquid Chromatography
- VIS – Visible Infrared Spectrum
- WR – Working Reagent

Table of Contents

INTRODUCTION

In this study, the main goal is to develop a working protocol for HPLC-UV analysis and fractionation of peptides. This project will be mainly based on establishing COFRADIC technology at the Department of Pharmacy, which will then be utilized to identify proteins. In this thesis, state of the art mass spectrometers will be utilized and the developed technology will be at the forefront of one of today`s rapidly developing research fields.

Since the accomplishment of genome sequencing of various organisms containing the human genome, the focus has changed from genome to analysis of proteome. Wilkins et al. was the pioneer who coined the term proteome in 1995 and it designates the total number of proteins articulated by a genome at a particular time (1). Gene activities in living cells are the functional aspects constituted by proteins. The investigation of proteome or proteomics is affected with protein recognition, determination of the function or assignment networks of proteins and creation of databases saving the gained information.

Proteomics is the investigation of quantitative progressions of protein articulation levels and their operation to medicine revelation, diagnostics and treatment.(2) Moreover, proteomics denotes large-scale studies of protein structure and function. Proteins can be studied in various ways, but in recent years, proteomics based on mass spectrometers have become the most important technique due to the produced data's detailed level and the high number of samples that can be processed per unit of time. Mass spectrometric proteomics are used to both identify and quantitate proteins. Because there are about 100,000 different proteins in humans (3), much research has been done to identify more and more proteins using different techniques. In mass spectrometric techniques, the biggest challenge is that the samples are enormously complex, which makes great demands on analytical instruments.

One of the chromatography methods, COFRADIC, used in recent years reduces the complexity of the samples and at the same time extracts as much information as is needed. This is based on chemical reactions for all parts of the sample that are not of interest (derivatization) so that they can be distinguished from the rest of the sample by using chromatography.

AIM OF THE THESIS

The overall aim of this dissertation was to develop a COFRADIC HPLC-system for identification of proteins. The N-terminal peptides was altered by either acetylation and/or di-methylation procedure in order to determine which of these posttranslational modification methods is most efficient for protein identification. A reversed-phase C18 column was utilized for separation of peptides.

1. THEORY

Proteins are recognized as one of the most prominent components in a living organism. The study of proteins (proteomics) requires the scientist to focus on various lifespan events of these proteins. The creation of a protein is a product of transcription (4), translation processes (5) and post-translational modifications (6). Translation process involves the formation of a native three-dimensional protein structure after that a primary strand of amino acids undergoes a complex folding process. Proteins are involved in a wide range of functions in the body such as receptor activity, formation of the complex and transduction of signals. These activities are central to the formation of a cellular phenotype. Therefore, protein analysis is a contributory process necessary for understanding of molecular biology. They are analyzed through marking their presence and positioning, as well as indicating their alteration state and biochemical framework. (7)

1. 1 Post-Translational Modification

Posttranslational modification is a venture in the biosynthesis of proteins. Ribosomes translates mRNA into polypeptide chains, creating proteins. These polypeptide chains experience PTMs, e.g. cutting, folding and other different methods, prior to turning into the sophisticated protein output.

Protein posttranslational modification (PTM) increases the functional variety of the proteome by the covalent expansion of proteins or functional groups, proteolytic cleavage of administrative subassemblies or deterioration of the complete protein. These alterations

incorporate acetylation, glycosylation, lipidation, methylation, nitrosylation, phosphorylation, proteolysis and ubiquitination. These PTMs affect roughly all parts of typical cell biology and pathogenesis. Subsequently, recognition and comprehension of PTMs is basic in the investigation of cell science, ailment medication and prevention.(6)

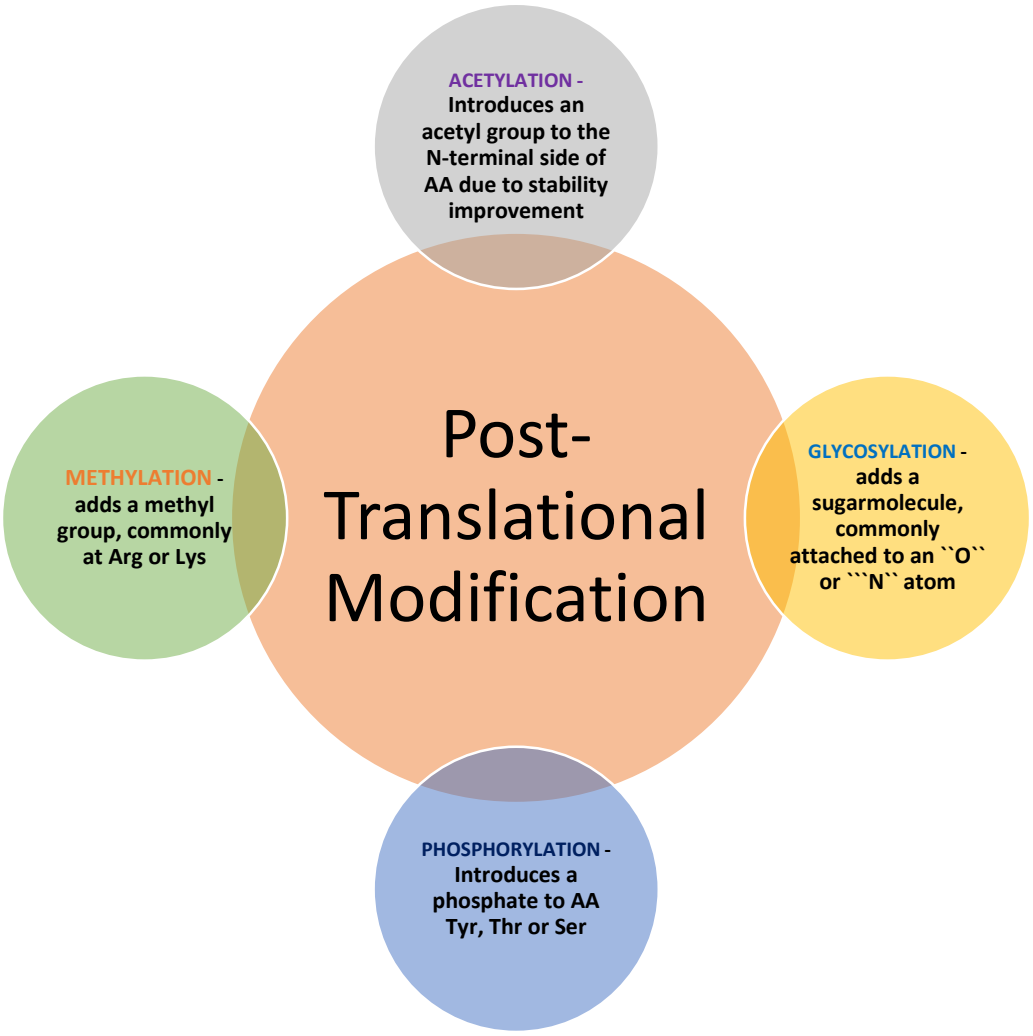


Figure 1: Graphic overview for diverse PTMs processes.

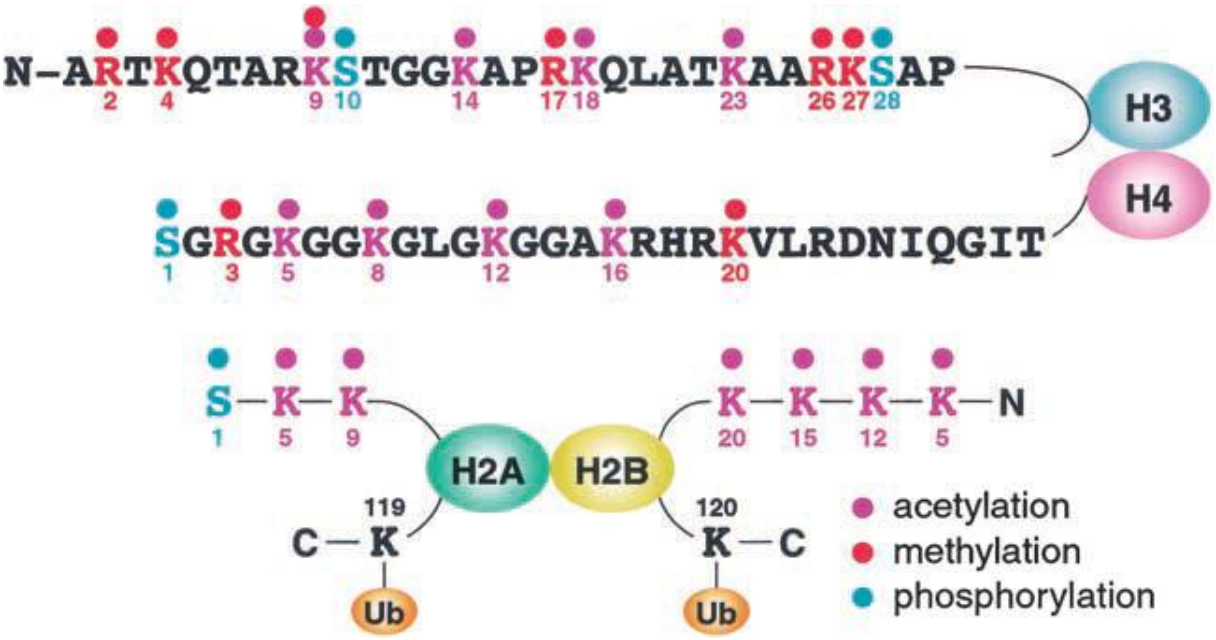


Figure 2: The areas of PTMs on the tails of histones.(8)

Acetylation

Lately, acetylation of proteins has appeared as crucial PTMs, opposing phosphorylation in its significance to the organization of biological procedures. Acetylation of proteins may occur at the alpha-amino group of the N-terminus side or the epsilon-amino group of lysine residues. When an acetyl group is introduced to these amines, then it causes de-protonation of these primary amines, hence their distinctive positive charge eliminates. The reversible character of lysine acetylation takes a central role when it comes to association between proteins and other biomolecules, especially with DNA. However, other targets are also involved (it is a part of gene regulation). This methodology exist roughly in all eukaryotic proteins.(9)

In this thesis, Sulfo-NHS-acetate, which is a water-soluble agent that irreversibly blocks primary amines like lysine at pH value above 7 to yield the acetylated structure, is applied during the acetylation procedure. Utilizing sulfo-NHS-acetate avoids polymerization in the case of cross-linking reactions of proteins.(10) Citraconic Anhydride is utilized in favor of reversible amine blocking.

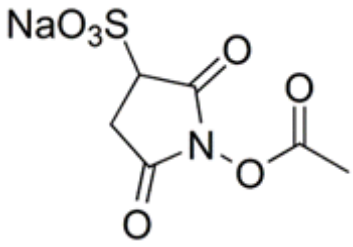


Figure 3: Chemical structure for Sulfo-NHS-acetate with a molecular weight of 259.17.(10)

Di-methylation (Reductive Amination)

A widespread post-translational alteration in eukaryotic cells is protein methylation at lysine or arginine residues. Reductive amination is a useful synthetic process of amines. Condensation of a carbonyl derivative (an aldehyde or a ketone) with an amine forms an imine (a Schiff base), followed by reduction to the equivalent amine, commonly by utilizing a borohydride reagent. Although, this reaction consist of two stages, such as formation of imine subsequent by reduction of borohydride, it is typical to accomplish the sequence in ``one pot``. Nevertheless, this reaction mechanism presents a possible complexity within the reaction condition. Creation of imine is a reversible procedure, and aldehydes and ketones as well as imine became reduced by reducing agents such as sodium borohydride. Sodium cyanoborohydride is usually applied in order to prevent this reduction of the initial carbonyl compound to the equivalent alcohol. This is because sodium cyanoborohydride ($NaBH_3CN$) chemoselectively reduces iminium ions over ketones and aldehydes. Sodium cyanoborohydride ($NaBH_3CN$) is a moderate reducing agent (transforms imines to amines), which achieves this reaction efficiently without any reduction of other chemical groups in samples of a biological nature (11). Nevertheless, hydrogen cyanide is released when the reaction is extinguished, which is extremely lethal. Sodium triacetoxyborohydride ($NaBH(OAc)_3$) is an another chemoselective reducing agent that has been applied, but this one is not stable in the presence of methanol and water (hydroxylic/ protic solvents).(12)

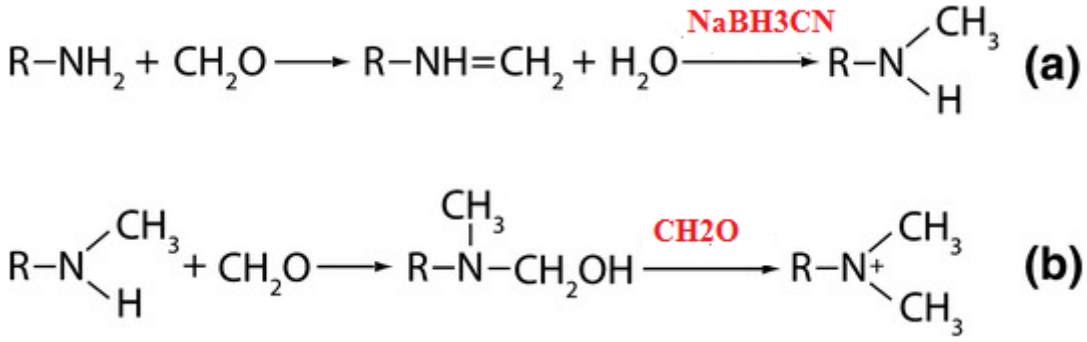


Figure 4: Reductive methylation reaction mechanism shows the addition of one methyl group to an epsilon-amine group of a lysine residue or the N-terminal amino group by creating a Schiff base intermediate by utilizing i.e. sodium cyanoborohydride (a). This process proceeds quickly to provide the di-methylated product by using formaldehyde (b).(13)

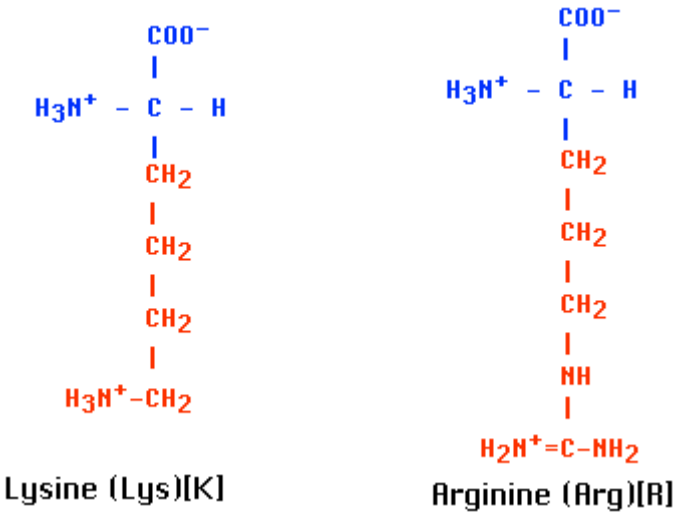


Figure 5: Chemical structures for amino acids Lysine and Arginine, which are classified as charged amino acids.(14)

1.2 Classic protein chemistry

The most basic knowledge of protein can be derived through the process of amino acid sequencing. The approach of Edman sequencing, first published in the 1950s, is a landmark breakthrough which for the first time made it possible for amino acids to be arranged in a particular order (15). This method demands first the cleansing of protein before undergoing different modification stages. In the first stage, the amino-terminal residue is altered to form a cyclic phenylthiocarbamyl. Under acidic condition, this cyclic phenylthiocarbamyl can eventually be liberated from the protein in the form of thiazolinone amino acid derivative. The altered amino acid is transformed into phenylthiohydantoin (PTH) amino acid. Once transformed, the phenylthiohydantoin (PTH) amino acid can be recognized through a process known as chromatographic separation since every PTH amino acid has a time-based graphic output of the chromatograph (various elution profiles), which slightly differs from each other. The repetition of this process will force a series of amino acid to be released in rapid succession resulting in the formation of chronological ladder, beginning with the protein N-terminus. The major technical disadvantages found in Edman sequencing is the non-compatibility of N-terminally blocked proteins (acetylation) with the agreement. As a general rule, when the Edman sequencing is utilized, the reaction is indeed not complete since only 30 amino acids can actually be arranged in a particular order. However, because the genome sequencing was still in the making stage and many years away before it could be considered, the use of Edman sequencing was most of the time not placed in the proper framework. The advancement in the automated Edman sequencing during the 1970s significantly increased the yield of the approach, which produced an avenue to arrange multiple proteins in a particular order in one single project (16).

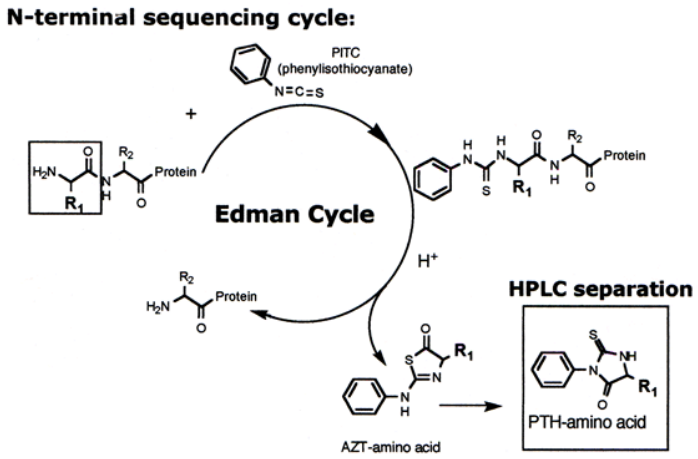


Figure 6: The chemical reaction of Edman Degradation for N-terminal protein sequencing.(17)

However, automated Edman sequencing has shown to be inadequate and insensitive when utilized on a set of protein sequences obtained by translation of all protein coding genes of a completely sequenced genome containing thousands of proteins and covering several systems of classification in a very large quantity. (18) (7)

The presentation of a new standard proteome investigation was conducted by peptide-centric proteomics, where the center of attraction is converted from protein separation to peptide segregation. Chromatography is a method that can segregate peptides, which can be connected directly to tandem mass spectrometry (MS/MS). The determination of peptide sequences is possible from the outcomes of MS/MS spectra, which could be utilized to identify the parent proteins. Because of the change in epitome for analysis of peptides, there is an increase in the probability of recognizing a protein seeing as various peptides can be used to determine the parent protein. In addition, peptides are smaller in physiochemical aspects and size in comparison to proteins. As a result, the susceptibility of peptide-centric proteome analysis increases. (7)

Data processing from MS/MS spectra

Mascot server is an effective search engine, which utilizes information from mass spectrometry to distinguish proteins from essential sequence databases.

While various comparative projects are accessible, Mascot is extraordinary in that it reconciles the greater part of the demonstrated techniques for seeking. These diverse search routines may be categorized as follows:

- Peptide Mass Fingerprint: In this search method, the only experimental data are peptide mass values.
- Sequence Query is another search method, where the peptide mass information is combined with composition data and amino acid sequence.
- MS/MS Ion Search is the search method, which is used in this project to extract data of peptides.

The sequence databases that can be applied on the Matrix Science are free, or public. Mascot servers are as follows:

- **SwissProt** is an extremely curated and good quality protein database. Normally the Sequences are non-redundant, instead of non-similar; hence, you may gain more matches for an MS/MS exploration than you might achieve from a widespread database, for example, NCBInr. SwissProt is perfect for searching peptide mass fingerprint and MS/MS explorations of well-defined organisms where it is not vital to fit every single spectrum.
- **NCBInr** is a wide-ranging, dissimilar protein database stated by NCBI for the application with their finding tools for example BLAST and Entrez. The accesses have been accumulated from GenBank CDS transformations, PIR, SWISS-PROT, PRF, and PDB.(19)

From classic protein chemistry to high-throughput proteomics

In the field of proteomics, which is gradually emerging, the quick segregation, characterization and identification of proteins from complex samples is a demanding task. The challenges includes guaranteeing investigation of how alterations in protein expression and their PTMs can be associated to a conversion at the genomic level, a specific illness/development stage of a disease or a particular signaling pathway.(20) (21) (22) Proteomics has subsequently

developed as one of the most significant “post-genomics” methodologies to better comprehend genes and the function of proteins since the accomplishment of the human genome sequencing. (23) (24)

How can proteomics be applied to medicine?

The technology of proteomics acts as a liaison between genes, proteins and disease; therefore, they play a crucial role in diagnostics, the discovery of new medications and molecular medicine. The findings of the growing research on the causation of diseases by defective proteins will help in developing drugs suitable for changing the form of the defective protein or replacing it. Currently, huge parts of the globally used drugs either perform by protein targeting or are proteins. Progress in proteomics may eventually assist the researchers to develop drugs that are ``personalized`` in order to provide individual responses in each patient and thus become more effective with minimal side effects. Existing research on proteomics has focused on the protein families connected to chronic diseases such as diabetes, cancer and heart disease. (25)

As stated earlier, analyzing a disease and profiling a proteome has been applied at the protein level. This implementation symbolizes a strong method to examine and better comprehend the molecular etiology and phenotype of the disorders and associate the proteome to the genome. Latest technologies such as mass spectrometry based proteomics permit the direct assessment of proteome profiles and in this way, these are applied as a vital apparatus for recognizing proteins and the biomarkers of proteins that are essential for research, characterization and diagnosis of ailments. (26)

Several areas of proteomics research include the study of highly complex protein blends like tissue homogenates, serum, plasma, or cellular lysates. (27)(28) Such biological samples might comprise plentiful amounts of different kinds of proteins expressed at varied concentrations, such as plasma proteins which may extend up to 10 orders of magnitude.(27) Whereas these kinds of samples consist of an abundance of details at the protein level, their intricacy entails a considerable quantity of sample treating, along with fractionation, refinement and multi-dimensional segregation before the MS and relative analysis of proteomics. (29) (30)

Proteomics strategies

Peptide-centric proteomics analysis is classified into two types: quantitative proteomics, aimed to quantify shifts in protein abundance in samples, and qualitative proteomics, designed to expansively identify the existence of all types of proteins in a sample. (7)

Qualitative Proteomics and Proteome Coverage

Three key factors affect proteome coverage. The initial factor is the susceptibility of the mass spectrometer, which indicates the minimum quantity of detectable analyte. Modern instruments are usually capable to carry out measurements in the order of attomoles or femtomoles. The instrument`s dynamic range is the second crucial factor, which denotes the intensity range of the signal where two distinct analytes can be revealed. This signal intensity range usually extends two to three orders of magnitude. Finally, the third influencing factor of proteome coverage is the duty cycle of the mass spectrometer, which involves the amount of fragmentation spectra (with average quality) that the apparatus can yield within a specific period of time. This differs from one spectrum per second for conventional apparatus to ten spectra per second for modern apparatus. The number of peptides detected is affected by this limitation when combined with chromatographic resolution. It is evident that even with high-speed spectrometers not every single peptide is detectable due to the elevated intricacy of proteome samples. This observable fact is referred to as random sampling. (7)

Quantitative Proteomics

Quantitative proteomics broadens the compositional map yielded by qualitative proteomics with total or comparative abundance data. Performance of quantitative proteomics is done on samples with diverse cellular phenotype like malignant versus benign cancer. The sample are exposed to singular stimuli such as growth versus control factor, and then monitored over time. An example of this monitoring would be cell cycle checkpoints or other cellular states where there is probability of an unusual proteome composition by hypothesis-driven investigation. Regarding to the different methods tested to perform quantitative proteomics, two are worth mentioning. The first set of methods inaugurates mass tags that make it possible to distinguish peptides from individual samples through MS examination (labeled samples). The second set of methods incorporates associated concentration profiles from MS/MS or LC-MS analyses to identify variances between samples (label free quantification).(7)

2. TECHNOLOGICAL DEMANDS

2.1 Peptide fragmentation in mass spectrometry

Several factors influence the sorts of peptide fragmentation noticed in mass spectrometry, such as primary sequence, internal energy quantity, introduction of the energy, state of charges and so forth. Roepstorff and Fohlman were the pioneers who suggested the fragment ions terminology (see figure 7) in 1984, which is popular among scientists today (31). Three years later, Johnson et. Al came with some alterations to the existed fragment ion nomenclature (named Collision-Induced Decomposition – CID) (see figure 5) (32).

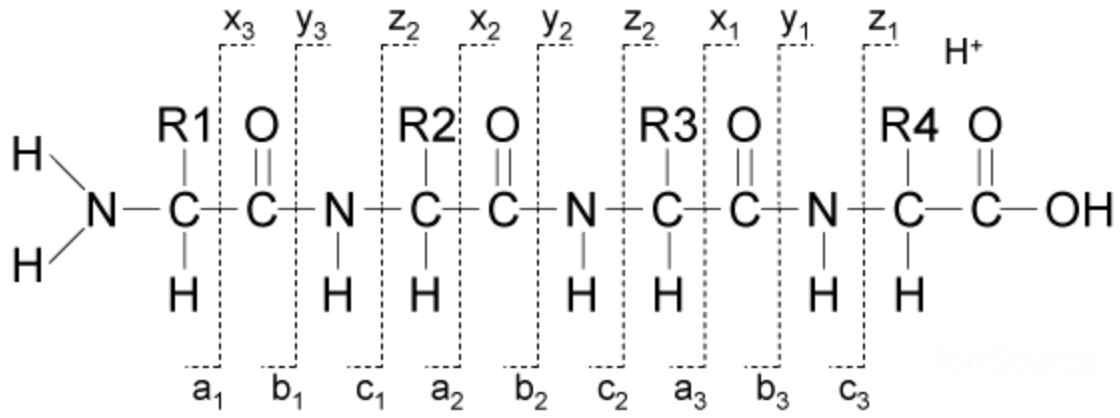


Figure 7: Peptide fragmentation terminology proposed by Roepstorff and Fohlman (33).

Peptide fragments must convey a minimum of one charge in order to be observed in MS. In accordance with Roepstorff nomenclature, the ion is categorized either as a, b or c, if the charge is placed on the N-terminus fragment (amino side). In contrary, if the ion is classified as either x, y or z, the charge is kept on C-terminal fragment (carboxyl side). As sketched in the figure 8, the most widely recognized peptide pieces in low energy collisions are a, b and y ions. The occurrence of b ions stretches from the N-terminal fragment while y ions are shown to reach out from the C-terminal. B ions are easily recognized and characterized compared with a ions, because of that b ions appear at a higher frequency and abundance than a ions. The a ions are frequently utilized as a diagnostic for b. For example, a-b sets are regularly seen in fragment spectra. The a-b sets are differentiated by 28u, which indicates the mass for the carbonyl, C=O (33).

Supplementary to the positively charged fragment ions (protons), c and y ions eliminate an extra proton from the precursor peptide. In this manner, the six individually charged sequence ions have the subsequent structures:

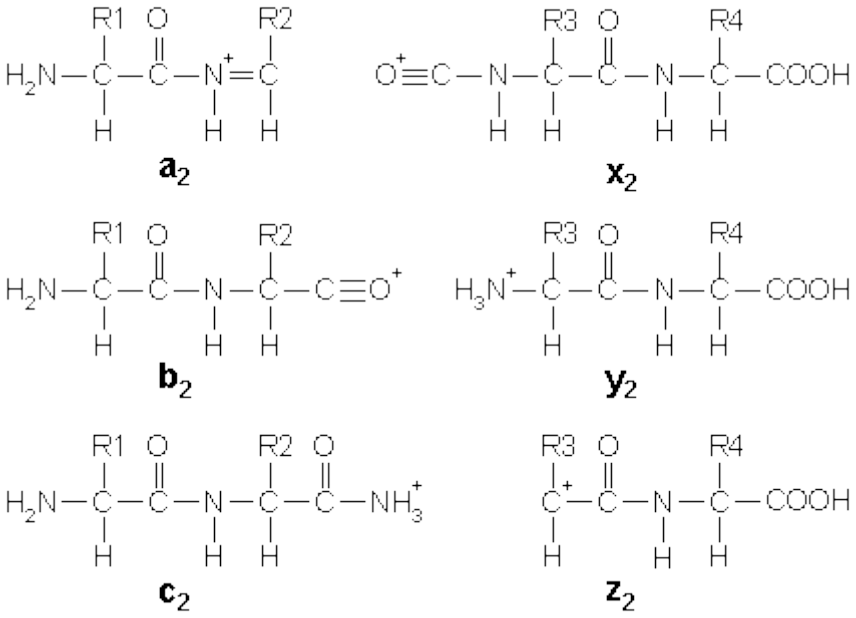


Figure 8: Six separately charged sequence ions. A single positive charge is the common factor between all six sequences. Tryptic peptides in electrospray ionization normally brings more than two charges, thus fragment ions are able to bring more than one proton.

A consecutive fragmentation of peptides, does not occur. This means that the initial fragmentation occurrence does not begin at the N-terminus end and continue successively one residue at a time down the amino acid chain. The fragmentation occurrences are to some extent irregular and most likely not consecutive. Moreover, a few fragmentations are favored over others as shown by the diversity in the abundance of recognized peaks in the spectrum beneath (figure 9). It is somewhat easy to perceive a peptide fragment spectrum by looking at it. The peaks will seem to be dissimilar by the approximate mass of an amino acid residue (see figure 7) (33).

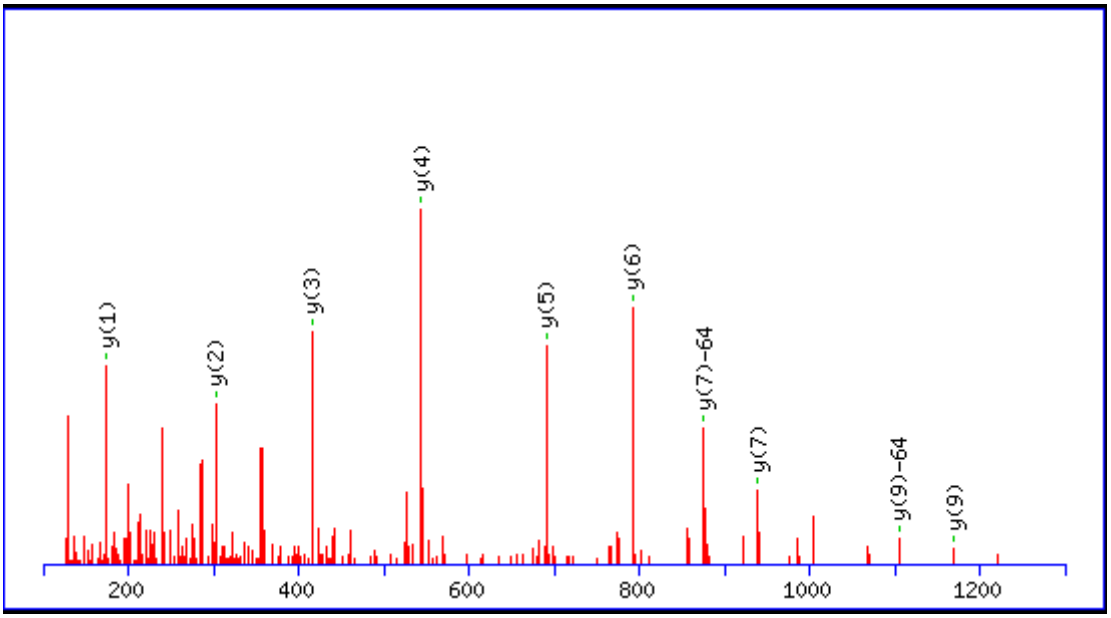


Figure 9: The MS/MS spectrum of the tryptic peptide SEQNNTEMTFQIQR from acetylation modification on the protein level.

2.2 Peptide separation

Peptides are dissolved via a column having a synthetic resin in a mobile phase flow in liquid chromatography. In the stationary phase, peptides bond with it and could be washed progressively by changing water or organic solvents in the mobile phase in order to look like the properties in the stationary phase. Several numbers of mobile and stationary phases have proven their worth in dividing peptides, and they were found using mass spectrometry to be suitable. Thus, they were ready for use in peptide-centric proteomics. RP-HPLC (reversed-phase high performance liquid chromatography) is one of the most general techniques, which use an extremely hydrophobic stationary phase. This is usually a column containing C-18-coated beads. The peptides bind themselves on these beads through hydrophobic interactions, and through raising the organic solvents concentration (usually acetonitrile) in the mobile phase. This way the hydrophobic peptides are increasingly liberated and thus eluted over a period. (7)

2.3 Chromatography

The physical process through which components are distributed selectively between two immiscible phases is known as chromatography. It is characterized by a mobile phase flowing over a stationary phase bed where the component moves along a definite direction (definition provided by IUPC)(34). The mobile phase determines the name given to the technique e.g. liquid chromatography (LC), gas chromatography (GC), or supercritical fluid chromatography (SFC). This process occurs due to repeated adsorption/desorption steps as the analytes move along the stationary phase. The separation is caused by the disparity in the distribution coefficient of the analytes present in the sample. (35)

High Performance Liquid Chromatography (HPLC)

In 1970, Prof. Csaba Horváth was the first one who invented the HPLC acronym (High Pressure Liquid Chromatography), which he used in the Pittcon work he presented. He pointed out that high pressure could be utilized in the generation of flow which liquid chromatography needs in packed columns. At its initial stage, the capabilities of pump pressures were only placed at 500 psi. (36)(37)

The early part of the 1970s saw growth of the technology with new instruments (HPLC) having the capacity to develop pressures to a point of 6,000 psi, and, at the same time had improved injectors and detectors as well as columns incorporated into the instruments. HPLC truly started to take hold from the middle of 1970s to the later part of the decade. Even in the light of enhanced performance, the acronym HPLC continued. The *high performance* liquid chromatography name came later.

Today, high performance liquid chromatography is a tool recognized as one of the most versatile when it comes to analytical chemistry. It is able to isolate, identify (by utilizing various detectors couplet to HPLC), and simultaneously quantitate different compounds present in any sample that can be dissolved in a fluid. Detection limit is highly dependent on the detector. Nowadays, it is simple to determine complexes in trace concentrations even as low as parts per trillion [ppt]. HPLC is being used today in any sample like pharmaceuticals, food, and nutraceuticals. HPLC can also be used in cosmetics and environmental matrices as well as in forensic samples and industrial chemicals.(36) A more convenient method than HPLC, named Ultra Performance Liquid Chromatography has been developed in the 20th century.

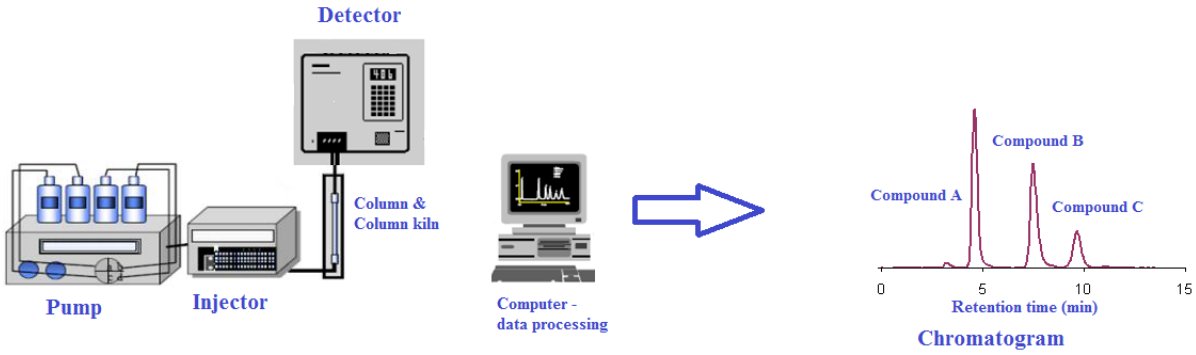


Figure 10: Schematic overview for how HPLC system operates (38).

Ultra Performance Liquid Chromatography (UPLC)

The year 2004 saw instrumentation together with column technology undergoing many enhancements. This has remarkably increased liquid chromatography's resolution, velocity and sensitivity. To reach a new level of performance, columns with smaller particles were needed [1.7 micron]. The instrumentation design possesses some qualities that aids it in the sending of mobile phase, which happens at the rate of 15,000 psi, thus creating a completely holistic new system. This technology can exhibit Ultra-Performance Liquid Chromatography. (36)

Instrumentation for Liquid Chromatography

In this section, a brief description about the essential elements used in HPLC will be provided.

Pump

The pump consist of mobile phase(s), which at a specific rate of stream makes the mobile phase pass through the liquid chromatograph. The regular stream rates employed in HPLC are between 1-2 mL/min. Conventional pumps are capable to extert pressures in the 6000-9000 psi range. Throughout the chromatographic analysis, two types of elution patterns are possible to apply. Initially, the isocratic elution implies that the composition of mobile phase is unchanged during the whole elution process. Secondly, the gradient elution means that elution process begins with a mobile phase that is of low strength (weak) and the strength of the solvent gradually increases throughout the elution operation. The pump is an important part of the tools for a well performing HPLC (38).

Injector

The LC method involves the use of an injection port to inject a sample into the mobile-phase stream that is availed by the high-pressure pump and transported for separation through the column. The column is the most important component of the LC system; thus, it should be handled with absolute caution. The flow-through detector is required at this stage for monitoring the separation (39).

Column

The column is recognized as ``the core of the chromatograph``. By applying different physical and chemical parameters, the separation of desired compounds happens in the stationary phase. A suitable selection of column is essential for successful separation in HPLC. The column is packet with small porous particles that lead to the high backpressure at regular stream rates. The commonly used columns have an inner diameter that is 1.8µm, 3.5µm and 5µm. A widely used bonded phase is C18 (reversed phase), where the stationary phase is comprised of non-polar saturated hydrocarbon chains. The hydrocarbon chains are fastened to silanol groups on the column, which collaborates with the analytes in order to distinguish them from each other.

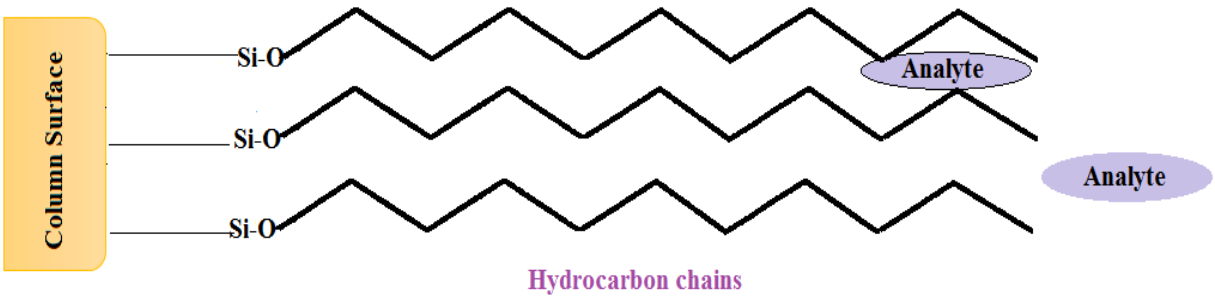


Figure 11: The structure of a conventional reversed phase column and tying of analytes.

To better, comprehend the retention time regularities and the compounds separation mechanism in a chromatographic procedure, it is important to study how mobile phase modifier concentration impacts on the retention in HPLC. Duple eluents that contain a weak and a strong solvent are broadly utilized within HPLC. In the reversed-phase liquid chromatography (RP) approach, for instance, the main components of the blended eluent are an exceedingly polar solvent such as water, while an organic non-polar solvent such as methanol, acetonitrile, etc. is supplied as a modifier in order to control the elution process.(40)

Reversed-Phase Chromatography is the most widely used LC method suitable for the analysis of polar and ionogenic analytes. As such, it is ideally suitable for application in LC-MS. Most analytical methods are performed by RP-LC. In the stationary phases, nonpolar, chemically modified silica or other nonpolar packing materials, e.g. hybrid silicon-carbon particles, are used in collaboration with the aqueous-organic solvent mixtures. Polymeric materials are less frequently used for packing compared to silica based ones. (39) (41)

Table 1: LC columns characteristics showing with different internal diameters. (39) (41)

Type	ID (mm)	F (µl/min)	V_{inj} (µl)	C_{max} at detector	Relative loading capacity
Nano-LC	0,05	0,120	0,012	8464	1
Micro capillary	0,32	4,9	0,49	207	41
Micro bore	1,0	47	4.7	21,2	392
Narrow bore	2,0	200	19	5,3	1583
Conventional	4,6	1000	100	1	8333

The maximum concentration at detector is based on the columns internal diameter, while the relative loading capacity is based on the given injection volume.

The final component of HPLC is a computer connected to the system in order to process data achieved from the analysis such as a chromatogram.

Detectors used in HPLC

Mass spectrometer is considered to be one of the most important LC detectors, but other detectors are also used for various applications. The most alternatively used detector in LC is the UV-absorbance type that has a specific detector characterized by a broad applicability range. Its detection capability is a product of absorption of photons by a chromophore such as aromatic rings or double bonds.(41)

Determination of proteins and peptides by UV absorption

A plain spectrometer is capable of quantifying a soluble protein amount. The radiation absorption close to the UV region relies upon a high content of aromatic amino acids (chromophores), such as Tyrosine (Tyr) and Tryptophan (Trp), along with a small degree of disulfide bridges and phenylalanine (Phe). Due to this fact, the absorbance at 280 nm (A_{280}) differs a lot among various peptides. In a protein mixture solution with a concentration of 1 mg/mL, this difference extend from 0 up to 4 (for some proteins with high content of tyrosine). However, the most values for absorption of radiation are in the range 0.5-1.5. The benefits of this approach are that it is easy to perform and the sample is reparable. On the other hand, the drawbacks of this method involves disturbance from other chromophores, and the particular absorption rate for a certain protein must be decided. (42)

The peptide bond assimilates powerfully in the far UV region with a maximum at approximately 190 nm. This exceptionally solid assimilation of proteins at these wavelengths has been utilized as a part of protein determination. Due to the challenges this method brings, such as absorption by oxygen and the low yield of traditional spectrophotometers at this wavelength, estimations are more practically made at 205 nm, where the absorbance is circa half that at 190 nm. The plurality of proteins have extinction coefficients at 205 nm for a 1 mg/mL solution of 30-35 L mol^{-1} cm^{-1}, additionally at 210 nm of 20-24 L mol^{-1} cm^{-1}. Different amino acid side chains, containing those of Trp, Phe, Tyr, His, Cys, Met and Arg (in that downhill order), provide inputs to the absorbance at 205 nm (A_{205}). (42)

UV, VIS and PDA Detectors

At pictograms level, the ultraviolet (UV), visible infrared spectrum (VIS) and photodiode array (PDA), which are in the category of absorbance detectors, are supplying sensitivity for light-absorbing compounds. The detectors offers excellent stability since they are simple to operate. The most generally applied detector for HPLC analysis is the UV detector.(43)

Throughout the analysis, the samples moves over a visibly uncolored glass cell, named the flow cell. A piece of UV light becomes absorbed by samples during the irradiation of UV light on the flow cell. This way, the UV light intensity is seen for the mobile phase without the sample becoming dissimilar from the eluent that includes the sample. By weighing this variance, the quantity of the sample can be decided. Selecting the correct wavelength is very important, and such selection will rely upon the type of analyte since the difference in UV absorbance depends on the wavelength applied.(43)

The consumer can utilize a standard UV detector to select the correct wavelength from a range between 195 and 370 nm. A VIS detector normally utilizes more extended wavelengths. It ranges between 400 to 700 nm in contrast to the UV detector. Both UV and VIS detectors allows for choosing a wider wavelength that ranges from 195 to 700 nm. The PDA detector reveals the whole spectrum at the same time. The UV and VIS detectors in two separate dimensions visualize the results derived namely light intensity and time. However, PDA includes wavelength, which is the third dimension. This has a great advantage as this does not require frequent analysis.(43)

2.4 Mass Spectrometry (MS)

Brief historical background for MS

In 1906, Thompson was the first one who developed the concept of mass spectrometry, where he termed the detachment of mass as canal rays. Since then, the significance of mass spectrometry has increased. It is evident from the Nobel prizes awarded in this field over the years:

- In 1906, Joseph John Johnson won the Nobel Prize for electricity conduction via gas.
- Francis William Aston received the Nobel Prize in 1922 for discovering isotopes.
- Ernest Orlando Lawrence was awarded the Prize for developing the cyclotron in 1939.
- Wolfgang Paul in 1989 received the award for the development of the ion trap.
- Koichi Tanaka and John Bennet Fenn were awarded the Nobel in 2002 for discovering the methods of soft ionization. (7)

How mass spectrometry works?

Mass spectrometry results from the production of ions that are separated based on their mass-to-charge ratio (m/z) and later detected. The plot of the (relative) abundance of the ions produced as a function of this ratio gives the resulting mass spectrum. Calculation of this ratio should be conducted with accuracy, as the results are crucial in quantitative trace analysis. The mass spectrometer is a complex computerized instrument that consists of five parts, namely the introduction of the sample, ionization, analyzing the mass, ion detection, and data handling (figure 12). The initial component is the ion source that serves to charge the analytes that will be measured. The analytes then enter the mass analyzer, which is the second component. In this stage, a magnetic or electrical field directly affects their trajectories leading to separation or selection of ions with diverse ratios of mass to change (*m/z*). The detection device is the third component, which precisely captures their particular *m/z* ratios and reads them out. Although most users view a mass spectrometer as a simple detector for LC, online chromatography-MS systems have offered additional value, e.g. additional selectivity. These techniques were widely used for the characterization of peptides and proteins following the introduction of electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) for mass spectrometry in 1988. (41) (39)(7)

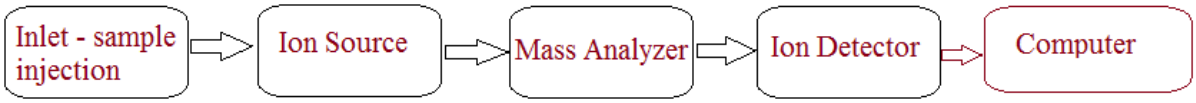


Figure 12: Schematic overview of the fundamental components of a mass spectrometer.

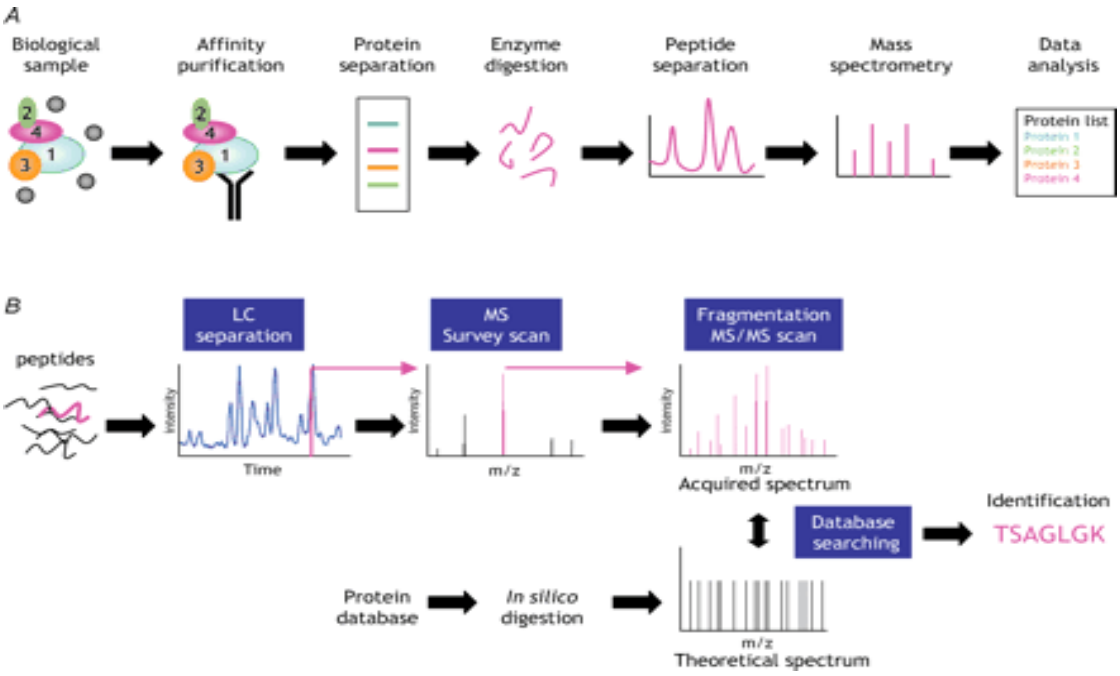


Figure 13: General methodology for protein complex identification utilizing mass spectrometry. A – Purifying and separation of a biological sample into its constituents, which are then proteolysed and dissected by LC-MS. B – Mass spectrometry based protein recognizable proof. A mixture of peptides (the peptide interest is labeled in pink) is separated by RP-HPLC. (44)

Electrospray Ionization (ESI)

The development of electrospray ionization has demonstrated to be an incredible device in biochemistry, permitting determination, characterization and sequencing of peptides, proteins and other essential biopolymers by utilizing mass spectrometry couplet to ESI mode. Electrospray ionization is considered a fundamental tool with highly documented capabilities. In an ESI source, the column flow is administered through a conductive tube, such as stainless steel capillary. A high voltage is connected to the capillary (ca. 3-5 Kv), which is retained in a coaxial stream of nitrogen nebulizing gas, making a fine spray of little droplets, each of which

conveys numerous surplus charges at its surface. The subsequent spray (the consequence of a columbic explosion) then provide rise to ions confined in the aerosol drops as they isolated to around a 10-micron radius. The ions are naturally protonated and identified in the form M+H+ in the positive ionization approach or M-H- in the negative ion method.

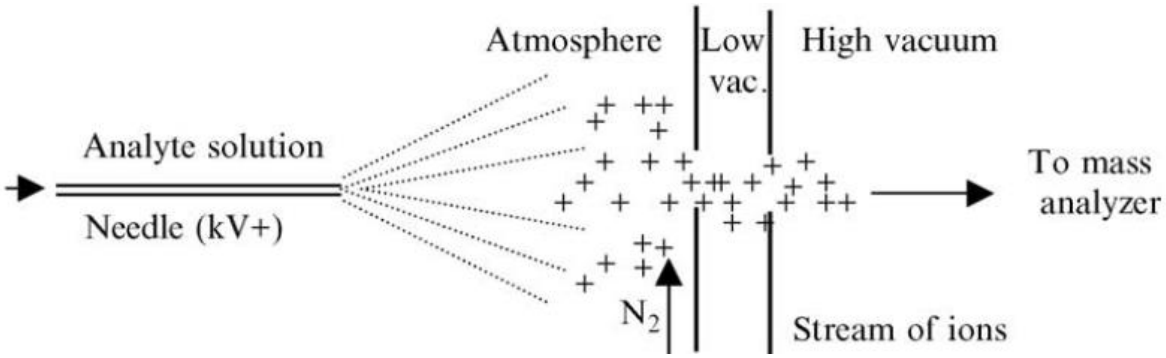


Figure 14: ESI – multiple charge states.(45)

2.1 Time-Of-Flight Mass Spectrometry

A mass analyzer separates ions of different mass-to-charge ratios according to their time of travel through a field-free vacuum area after being provided with similar kinetic energy. Precisely, the speed of the ions is reliant on their mass-to-charge ratio in addition to the distance travelled, for the time observed to reach the detector permits the mass-to-charge ratios to be determined with the comparatively heavier ions consuming extended time.(46)

A method of mass measurement to use despite complexities that come with higher resolutions is TOF-MS (Time-of-flight mass Spectrometry). In the late 1950s, Bendix Corporation marketed the first commercial TOF apparatus. The Corporation derived the basis of their design from an apparatus published by Wiley & McLaren in 1955. Recently, TOF-MS has emerged as an indispensable apparatus for biological analysis applications, particularly with the connection of TOF-MS to MALDI and ESI ionization techniques and the improvement of high-resolution and hybrid apparatus such as Q-TOF and TOF-TOF forms. Distinctiveness of TOF-MS includes high sensitivity that detects all ions, unrestricted mass range and rapidity of analysis, with recent apparatus attaining full spectra in seconds. This ranks TOF-MS as the most advantageous technique of mass analysis. (47) (48)

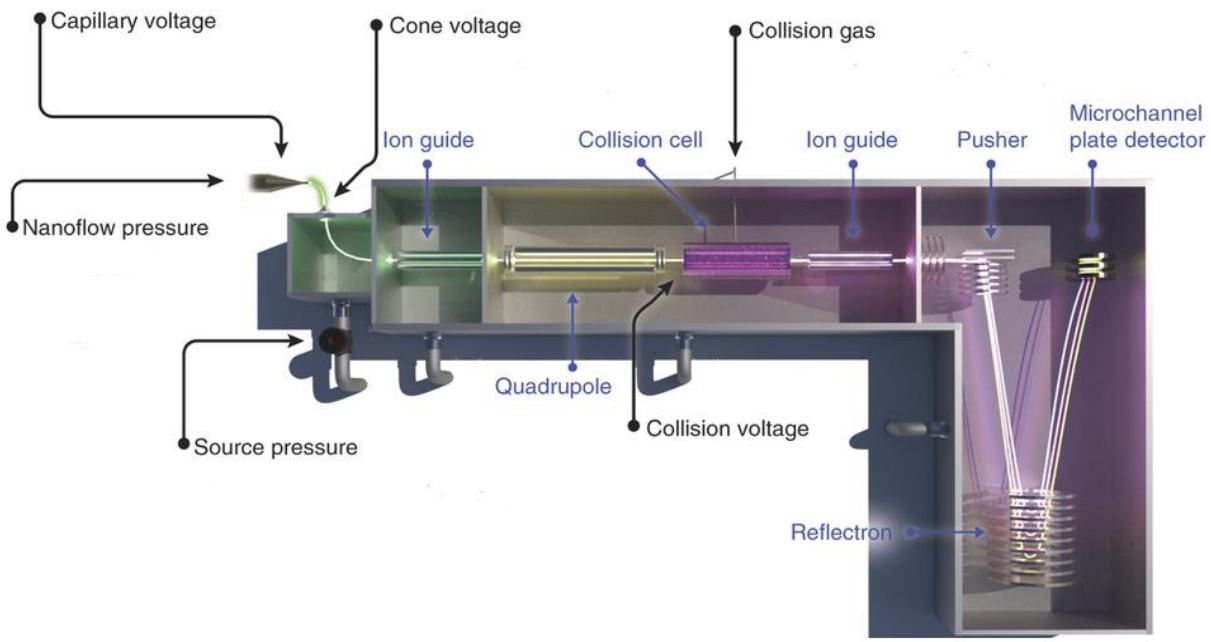


Figure 15: This figure shows TOF-MS working in Reflectron Mode (49).

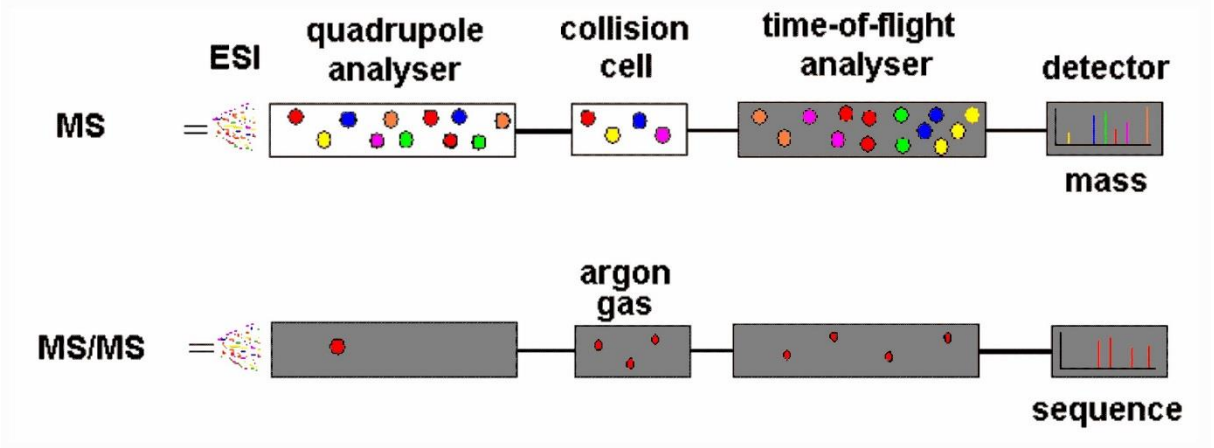


Figure 16: Q-TOF mass spectrometer working in tandem MS mode (lower) and MS mode (upper).(46)

2.2 Q-Exactive

In current years, from massive developments in high-resolution instrumentation, mass spectrometry-based proteomics has significantly advanced. More precisely, the consolidation of a linear ion trap with the Orbitrap analyzer has been demonstrated to be a popular and standard instrument formation. The stand-alone Orbitrap analyzer coined as Exactive supplements is the hybrid trap-trap instrument. A. Michalski et al. presented a combination of quadrupole mass filter to an Orbitrap analyzer in their study that was performed in 2011(50). This "Q Exactive" instrument characterizes high ion flows because of an S-lens, and quick high-energy collision-induced dissociation peptide fragmentation because of booth parallel filling and recognition methods. The image current from the detector is treated by an "enhanced Fourier Transformation" algorithm, amplifying mass spectrometric resolution. The instrument meets the time of the overall cycle of 1 s for a top10 higher energy collisional dissociation process along with instant segregation and fragmentation.

Over 2500 proteins could be recognized in standard 90-min gradients of tryptic digestion of the mammalian cell lysate— an improvement over earlier Orbitrap mass spectrometers. Additionally, the quadrupole Orbitrap analyzer consolidation allows multiplexed operation at both MS and tandem MS levels. This is established in a multiplexed single ion-observing mode, in which the quadrupole quickly shifts among diverse narrow mass ranges that are studied in a single compound MS spectrum. Likewise, the quadrupole permits fragmentation of diverse precursor masses in quick progression, ensued by the combined analysis of the higher energy collisional dissociation fragment ions in the Orbitrap analyzer. High performance in a healthy benchtop format along with the capability to perform compound multiplexed scan modes make the Q Exactive a stimulating new instrument for the proteomics and general analytical communities. (50)

2.3 Enzymatic digestion of proteins

Trypsin is a pancreatic serine endoprotease enzyme with an average molecular weight of 23.29 kDa and a pH optimum around 8.0**.** Trypsin has always been used according to established practice in conducting proteomics research for the mapping of peptide and the sequencing work of protein. This is because it has a highly specific cleavage resulting only in a few number of tryptic peptides. This enzyme breaks down peptide bonds through chemical reactions. This happens through trypsin forming a cleavage at the C-terminal (carboxyl) side of Lysine or Arginine if the adjacent C-terminal amino acid is not proline. The chemical symbol for Tryptic peptides is Tx, where x represents the number of the trypticpeptide from the N-terminal side of the protein, for example GLSDGEWQQVLNVWGK. Two methods that can be applied in this reaction include the use of trypsin immobilized enzyme reaction (IMER) and homogenous digestion through the use of trypsin immobilized in beads. Successful homogenous digestion requires the use of phosphate or ammonium bicarbonate buffer in a solution at pH 7.5 and 37^{o}C and within 6-24 hours. The trypsin-to-substrate ratio should be kept minimal to limit the auto digestion of trypsin. Many studies propose that proteolytic digestion should precede the reduction and alkylation of the protein. Better accessibility to protease is achieved through the reduction of Cys-Cys disulfide bridges in the protein by the use of β-mercaptoethanol or dithiothreitol (DTT). Protection of the reactive Cys-SH groups is achieved through subsequent alkylation using iodoacetamide. (39)

2.4 COFRADIC^{TM} – recently developed method for proteomics

One of the chromatographic methods intended to pick peptides from complex mixtures is COmbined FRActional DIagonal Chromatography (COFRADIC^{TM}). It is composed of two sequential similar chromatographic segregation with an adjustment step targeted on a division of peptides between the two separations. The changed peptides gain different chromatographic characteristics and separate from the majority of unchanged peptides in the subsequent run. Final analysis of the sorted peptides minimizes the intricacy while retaining all the properties of the proteome.

COFRADIC is a mixture of three successive phases: preliminary split-up peptides, a chemical or enzymatic alteration of preferred peptides, and ultimately an isolation of (mixed) primary

peptide segments. Theoretically, each class of peptide bearing an amino acid or, more usually, a functional group that can be particularly altered, can be isolated by COFRADIC. Presently, COFRADIC arranging processes allow routine separation of methionyl, cysteinyl, N-terminal, and phosphorylated peptides respectively. These processes are basically diverse both in the methods by which proteins are arranged before the actual COFRADIC categorization phase and in the chemicals or enzymes that are utilized in the middle of the two chromatographic isolations.(51)

The only requirement necessary to facilitate the isolation of peptides in the COFRADIC^{TM} technique is that these peptides have a minimum of a single functional group that differentiates them from all the other peptides. These peptides should be capable of quantitative and specific modification through an enzymatic or chemical reaction undertaken linking the two sequential analogous chromatographic separations. If there is sufficient alteration of chromatographic properties, the bespoke peptides will exhibit different preservation periods when re-separated, whereas unaffected peptides will exhibit the same characteristics in the subsequent separation. Therefore, the altered peptides can be isolated for further investigation. (52)

Table 2: This table represents various announced COFRADIC approaches from 2002 until 2009.(53)

Peptides aimed by COFRADIC	Sorting reaction	Citation/Bookmark
Tyrosine Nitrated	Dithionite reduction	Ghesquière et al. 2009
Sialylated	Neuraminidase	Ghesquière et al. 2007
ATP binding	Alkaline hydrolises	Hanouille et al. 2006
N-Glycosylated	PNGaseF	Ghesquière et al. 2006
Phosporylated	Phosphatases	Gevaert et al. 2005
Cysteinyl	Ellman reduction by TCEP	Gevaert et al. 2004
N-terminal	TNBS	Gevaert et al. 2003
Methionyl	H_2O_2	Gevaert et al. 2002

An overall pattern of peptide separation by COFRADIC is demonstrated in figure 1 (below), and it is comprised of three vital proceedings (54):

- A peptide mixture, which undergoes primary RP-HPLC COFRADIC fractionations.
- A single or various primary fractions became either chemical or enzymatic modified.
- A set of secondary, exact RP-HPLC fractionations of the altered primary fraction(s)

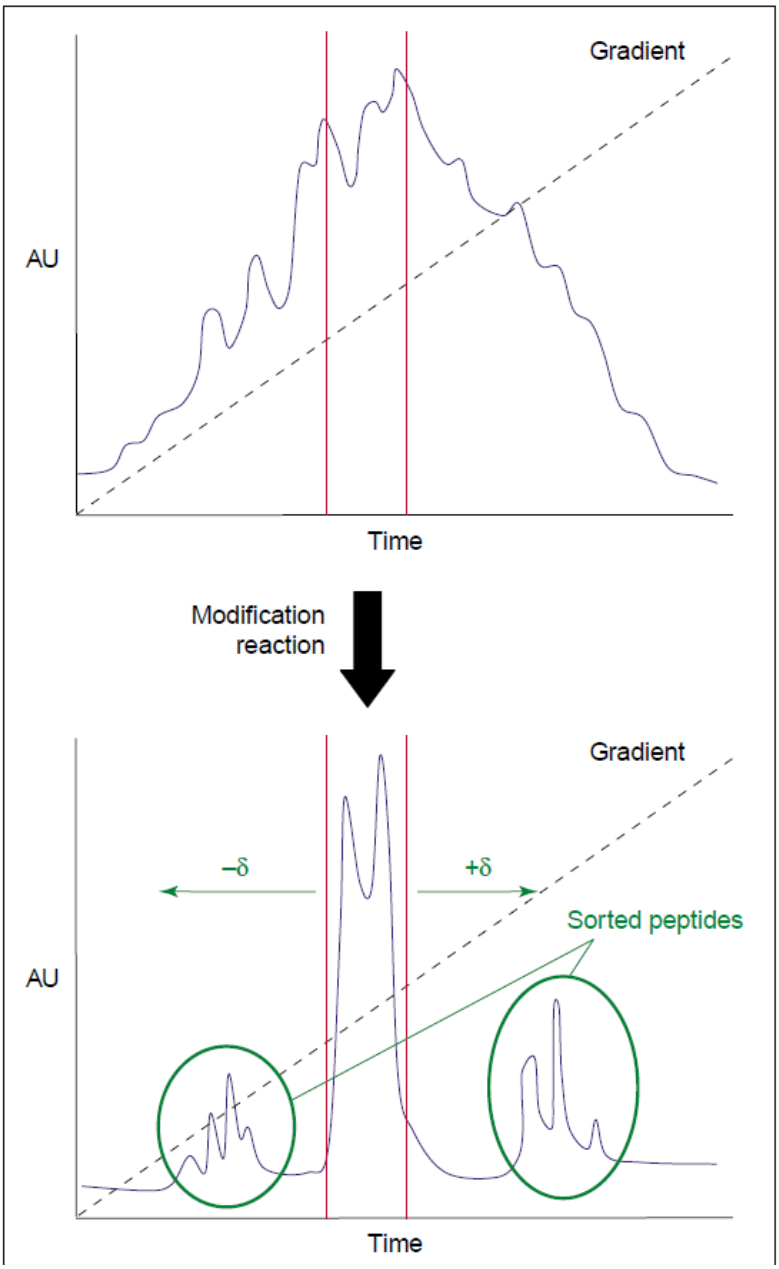


Figure 17: COFRADIC™ principle for separating representative peptides from an intricate peptide mixture. In the primary run (top), separation of a peptide concoction occurs on a reserved-phase high performance liquid chromatography column (RP-HPLC). Exposure of the fractions to enzymatic or chemical alteration modifies the column retention characteristics of a group of peptides in each primary fraction. In the secondary run (bottom), reloading of the altered primary fractions occurs on the same column and isolation of peptides is carried out under similar conditions. Altered peptides move out of the initial collection interval and can be exclusively secluded for in depth investigations. These altered peptides can either exhibit the following characteristics; early elution with hydrophilic shift (−δ) or later with hydrophobic shift (+δ) compared to the unaltered peptides. (52)

How COFRADIC™ works?

The fundamental standard of COFRADIC stems from the diagonal electrophoresis method that was initially emitted in 1966 (55). The phrase ``diagonal chromatography`` was coined, consequently when the process was undertaken with reversed-phase or ion exchange chromatography as an alternative to paper electrophoresis. Compared to usual diagonal chromatography, the design of COFRADIC makes it capable of dealing with complex mixtures. In the primary run, the peptides are separated and a number of key fractions are pooled to minimize the number of secondary runs. Fractions are pooled in order that altered peptides from one fraction will not interfere with unmodified peptides from other primary fraction. Thus, a minimizing of the total number of secondary runs will occur by a determinant of approximately 4. Recovery of a peptide occurs in 1-2 secondary fractions under a vigilantly controlled chromatographic environment, guaranteeing high sensitivity. (52)

N-terminal COFRADIC

The recent technological development has encouraged the development of procedures suitablefor selecting the N-terminal peptides of proteins while utilizing the analysis by mass spectrometry. These are necessary for characterizing protease-mediated cleavage and protein α-*N*-acetylation on a proteomic level. There has been a rise in the application of N-terminal combined with the fractional diagonal chromatography (COFRADIC) in the numerous studies dealing with the application of protein modifications. Trypsin digestion occurs after the derivatization of primary amines to allow the appearance of cleavage after arginine residues. (56)

Separation of N-terminal peptides

Methionine and cysteine are sulfur-consisting amino acids, which are also considered as hydrophobic and non-polar. As mentioned earlier, the focus within the COFRADIC strategy is to reduce a sample`s complexity prior to analysis. Despite the fact that separating of cysteinyl and methionyl peptides can decrease the intricacy of a sample by a determinant of five, this lessening in complexity is inadequate for cell lysates from higher organisms. Therefore, a plentiful amount of segregated peptides elopes from analysis due to the great peptide flow against the mass spectrometer. For that reason, Gevaert and Vandekerckhove in 2004 have proceeded the COFRADIC ranking chemistry that includes separation of N-terminal peptides (57)(52).

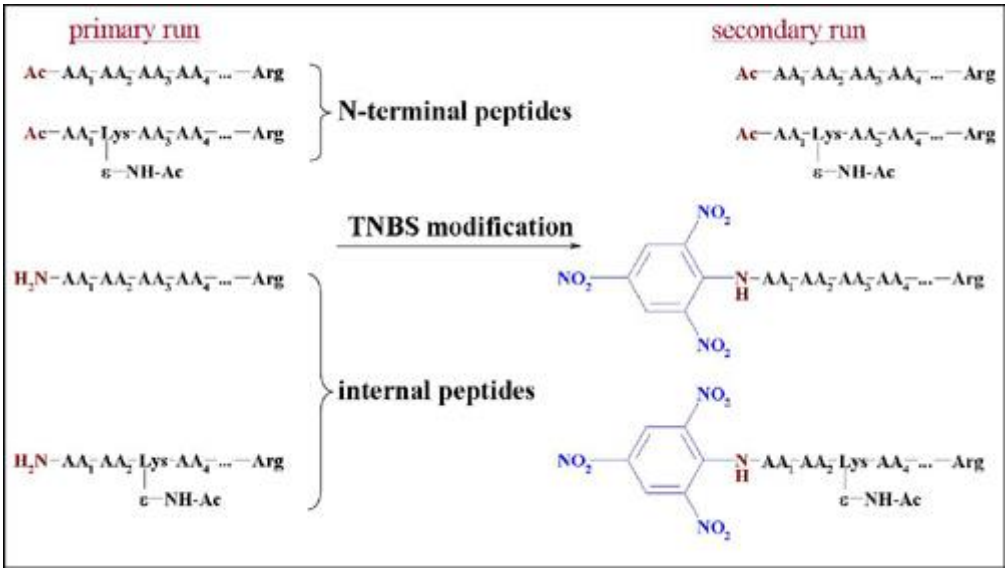


Figure 18: Separation of N-terminal peptides by TNBS modification. A hydrophobic shift appears between primary to secondary run.(58)

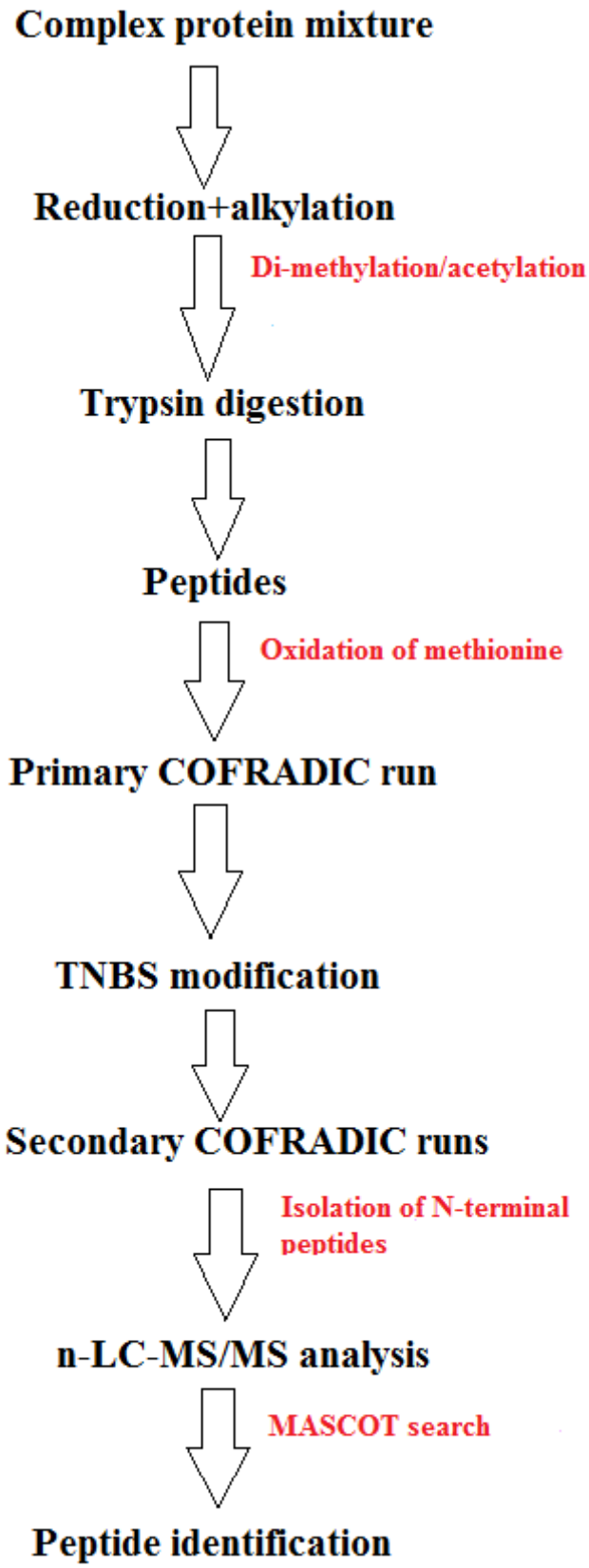


Figure 19: General workflow of the N-terminal COFRADIC method.

3 MATERIALS

All equipment and chemicals with details can be found in Appendix 1 and 2, respectively.

4 METHODS

This chapter will present the methods that have been applied during this thesis as well as detailed procedures. The sample preparation and detection methods used in this thesis will be discussed. The project started with a qualitative analysis to determine which peptides the Mass spectrometry system is capable to identify.

4.1 UV-HPLC instrumentation

Analytical HPLC was performed on a Waters 2695 HPLC provided with an RP-HPLC XBridge^{TM} C18, 5 µm, 4.6 x 250 mm column (made in Ireland) and analyzed at wavelengths 214 nm with a PDA detector that extends from wavelengths 210 to 310 nm.

4.2 UPLC-MS/MS instrumentation

Mixtures of peptide that contained 0.1% triflouroacetic acid were loaded onto a nanoAcquity^{TM} Ultra Performance LC (purchased from Waters, MA, USA). This instrumentation consisted of a 3-µm Symmetry® C18 Trap column i.e. 75 µm × 100 mm (from Waters) placed before a 1.7-µm Atlantis^{TM} C18 analytical column. This analytical pre-column had an inner diameter of 100 µm × 100 mm (Waters). By utilizing a gradient of 5-95% acetonitrile and 0.1% formic acid, separation of peptides were obtained. The separation of peptides was achieved with a flow rate of 0.4 µl/min that was eluted onto a Q-TOF Ultima mass spectrometer, i.e. Micromass/Waters and exposed to tandem that is data dependent using mass spectrometry analysis. The generation of peak lists was done by a Global software server (version 2.1) known as ProteinLynx. Exploration for the developed pkl files were performed towards the protein sequence database named SwissProt 2013_03, which is operating with an in-house Mascot server (Matrix Sciences, UK). In the seeking process, 100 ppm was utilized as peptide mass tolerance and 0.1 Da was used as fragment mass tolerance.

Table 3: Gradient program used for LC-MS/MS

Time (min)	Flow (µl/min)	%A (Water)	%B (ACN)
0	0.400	95	5
5	0.400	85	15
25	0.400	50	50
30	0.400	20	80
48	0.400	5	95
60	0.400	95	5

4.3 Q-Exactive instrumentation

A Thermo Fisher Scientific EASY-nLC1000 system and an EASY-Spray column (C18, 2µm, 100 Å, 50µm, 15 cm) were loaded with peptide mixtures that comprised of 0.1% triflouroacetic acid. Fractionation of peptides was carried out with a gradient of acetonitrile in range 2-45% containing 0.1% formic acid. A flow rate of 250 nl/min was used for more than 50 minutes. Analysis of separated peptides was conducted utilizing a Thermo Scientific
Q-Exactive mass spectrometer. Data obtained from this analysis were collected by using a data dependant mode such as Top10. Proteome Discoverer version 1.4 software was used to process the raw data. The seeking process for the fragmentation spectra was performed towards a database named Swissprot SwissProt_2011_12, utilizing an in-house Mascot server (Matrix Sciences, UK). The Mascot search was accomplished with peptide mass tolerance of 10 ppm while the fragment mass tolerance was 0.02 Da. A false discovery rate (FDR) set to 1% for peptide identifications was adapted to filter the peptide ions.

The Thermo Scientific Q Exacative mass spectrometer helps in identifying, quantifying and confirming compounds. It incorporates quadruple ion selection and high resolution mass detection resulting into high quality performance.

Table 4: Gradient program used for Q-Exactive.

Time (min)	Flow (nl/min)	%B (ACN)
0	250	0
10	250	3
60	250	45
68	250	100
78	250	100

4.4 Sample preparation according to Novagen protocol

The lysis of *Escherichia coli* bacteria (our protein source) and preparation of BugBuster Master Mix sample were conducted in accordance to the protocol supplied by Novagen™ (purchased from USA). Since a 2D-gel electrophoresis technique was not carried out, step 3 and 6 in the BugBuster Master Mix protocol were excluded (59).

4.5 Pierce™ BCA Protein Assay Kit

The determination of protein concentration was carried out in accordance with the protocol from Pierce supplied with the kit (Microplate procedure).(60)

Known concentration dilutions were prepared from Bovine Serum Albumin (BSA). The dilutions were then assayed together with the unknown(s). The standard curve helped in determining the concentration of each unknown. BSA (Bicinchoninic acid) assay that is obtained from Thermo Scientific helped in determining the protein concentration, which was established to be 4.30 mg/ml. The Thermo Scientific Pierce BCA Protein Assay Kit is regarded to be very efficient. This is because it is highly accurate and the detergent-compatible assay reagent has been established to determine total protein concentration at 562 nm compared to a protein standard.(60)

In this thesis, the concentration of protein in the initial BugBuster Master Mix sample was determined to be 4.30 mg/mL by using Spectra Max Microplate Spectrophotometer (made in USA).

Later on, during the project, two other BugBuster Master Mix samples were made in a similar manner as described previously and their concentrations were respectively 7.0 mg/mL and 16.0 mg/mL (no guarantee that this BCA was valid).

4.6 Procedure for Filter Aided Sample Preparation

The accompanying protocol depicts a technique for creation of tryptic peptides from unrefined lysates for LC-MS/MS analysis. Thus, sample preparation was performed in accordance with the FASP protocol.

Chemicals required

- Urea [MW: 60.06]
- Tris-HCl (Trizma® HCl) [MW: 157.60]
- Ammonium Bicarbonate – ABC (NH_4HCO_3) [MW: 79.06]
- Calcium Chloride ($CaCl_2$) [MW: 110.98]
- Trypsin
- Dithiothreitol (DTT) [MW: 154.24]
- Iodoacetoamine (IAA) [MW: 185]
- Acetonitrile
- 10% and 0.1% Triflouroacetic acid (TFA) solution

Buffer preparation

Prior to conducting the FASP protocol workflow, various buffers were made with subsequent concentrations:

- **Buffer 1**: 8 M Urea, 100 mM Tris-HCl; pH 8,48
- **Buffer 2**: 50 mM Ammonium Bicarbonate (ABC) (Digesting buffer)
- **Buffer 3**: 50 mM ABC, 0,1 M Urea, 1 mM $CaCl_2$
- **Trypsin stock**: 20 µg lyophilized Trypsin powder was dissolved in 20 µl 1 mM HCl solution to obtain a final concentration of 1µg/µl. This solution was stored in the freezer at -20^{o}C.
- **Reducing Buffer:** This buffer solution had a concentration of 200 mM DTT in 50mM ABC. Reducing buffer was stored at -20^{o}C. Later on, this solution was prepared in the same way, but dissolved in water, in absence of ammonium bicarbonate.

NOTE: During di-methylation and acetylation procedure, ABC buffer should not be used prior due to interference with acetylation and methylation of amines.

- **Alkylation buffer:** A stock solution of 1M was used. This buffer was prepared prior to usage because of its light sensitive nature simultaneously it was dispensed in a foil-wrapped tube to avoid light exposure.

NOTE: High care must be taken while utilizing TFA as it is a tremendously lethal chemical component. The concentrated acid is required to be applied in a fume hood, and suitable protecting clothing should be worn.

Protocol for FASP workflow

The samples were prepared several times in various ways before adapting FASP protocol in order to enhance the analysis outcomes. The amount of protein at the start of the process was different from that used in subsequent runs. In addition, the alkylation process was performed later on in the first run. It is important to note that the reduction and alkylation steps were not carried out subsequently in the first run. This explains why no results were obtained in the initial LC-MS/MS analysis.

In this section, there will be a general description of FASP protocol and how it was implemented.

The description and implementation of the FASP protocol – Sample preparation

First, a desired initial amount of protein was chosen and calculated according to the BugBuster Master Mix sample concentration[1]. The sample was then either reduced with DTT and alkylated with IAA directly, or prior to reduction and alkylation initially diluted in an appropriate buffer solution (e.g. ABC buffer), and the volume was then adjusted to
100 µl.

The proteins were reduced by adding 100 µL 200 mM DTT in 50 mM ABC, shaken gently and then incubated at 37 ºC for 1 hour shaking at 200 rpm. This was followed by

[1] In the first run, the desired volume of BugBuster Master Mix sample was directly added to the filter after equilibration (began with 1mg protein and added 232 µL sample (protein cons. was 4.30 mg/mL)).

carbamidomethylation of cysteine by applying 100 µL of freshly prepared alkylation buffer 10 mM IAA in 50 mM ABC. The sample was then incubated darkly for 30 minutes at room temperature.

NOTE: Reduction and alkylation steps should be carried out directly after each other to prevent reformation of the reduced thiol groups.

A 30 kDa Pall spin-filter was equilibrated with 100 µL Tris-HCl buffer (buffer # 1), then it was centrifuged for 10 minutes at 13000 rpm at 20°C. Thereupon the sample was placed on the filter and centrifugation at 13000 rpm for 30 minutes at 20°C was performed. In order to conserve the structure and/or function of the contents of the lysate, 100µl of lysis buffer # 1 (Tris-HCl buffer) was added on the filter to mimic the internal circumstances of the cell or cellular compartment. Thereafter, the sample was shaken gently at 200 rpm for about
5 minutes and then was centrifuged for 15 minutes at 13000 rpm and 20°C. This process was repeated three times.

NOTE: Between every centrifugation, the filter was checked for protein solubility. The centrifugation can be prolonged if necessary due to remaining liquid on the filter.

Consequently, 100µl of digesting buffer 50mM ABC was added on the filter. The pall spin-filter was placed in a rotating mixer for 5 minutes at 200 rpm. Then, the sample was centrifuged for 15 minutes at 13000 rpm and 20°C. This step was also conducted three times.

Ultimately, the spin filter was positioned in a new original pall-spin tube so that it could collect peptides overnight[2]. 50µl of buffer 3 was added on the filter and vortex gently at 200 rpm for about 5 minutes. Thereafter, trypsin was added to the sample in a ratio of 1:50 (used trypsin stock 1µg/µl)[3] combined with 100 µl 50mM ABC (digesting buffer). The sample was then placed in an incubator at 200 rpm, and digested overnight at a temperature of 37°C.

Elution:

The digested sample was taken out from the incubator. Filter leakage was checked before proceeding. If no leakage was found, the sample was then centrifuged at 13000 rpm for

[2] To avoid filter leakage, a new original pall-spin tube was used instead of a regular Eppendorf tube.
[3] The first run failed. In the second run – started with 100 µg protein and used 2µl trypsin solution (ratio 1:50).

20 minutes at RT and the peptides that had been created overnight by the digesting process were collected. In case of leakage, the leaked sample was added on the filter again and sat to further digest for an additional 1-3 hours.

Thereupon, 100 µl 50 mM ABC buffer was added on the filter, vortexed shortly at 400 rpm and for further peptide collection, the sample was centrifuged at 13000 rpm for 20 minutes at RT. Eventually, for additional collection of peptides, 100 µl purified water was added on the filter, vortexed shortly at 400 rpm, and then centrifuged at 13000 rpm for 20 minutes at RT.

Later, the filter was removed from the tube and the eluate was acidified by using 10% TFA stock solution to reach a final concentration of 1% TFA[4].

Protein sample desalting:

Eluate containing 10 µg peptide was desalted by C18 tip according to the protocol submitted by Vendor, dried under nitrogen gas. Thereafter, the dried sample was re-suspended in 10-12 µl 0.1% TFA solution, shaken gently and the sample was quick spun by utilizing a table centrifuge. The re-suspended sample was then transferred to an auto sampler vial for n-UPLC-MS/MS analysis.

NOTE: Only 2µl of the re-suspended sample can be injected and analyzed on the n-UPLC-MS/MS instrument. Thus, if any troubleshooting appears during the analysis, the sample can be re-analyzed.

FASP protocol was implemented several times with modifications in order to improve and increase the amount of identified peptides. In addition, the next FASP run contained a protein methylation process.

[4] This step was performed with some discrepancy. For instance, 20 µl 10% TFA solution was added to acidify several samples, even if the total sample volume was not 200 µl.

4.7 Adapted protocol for protein methylation

Required buffer solutions/chemicals

- 100 mM Sodium Acetate (NaOAC) (pH 5-6) [MW: 82.03]
- Formaldehyde 37% solution
- 520 mM Sodium Cyanoborohydride ($NaBH_3CN$) [MW: 62.84] – was added as freshly prepared in order to obtain optimal quantification of amines.

Procedure for Di-methyl labeling on protein level

The sample was prepared in a very similar manner to that described previously in section 4.6. 100 µg protein from the BugBuster Master Mix sample was used to run this procedure. The adapted volume of BugBuster Master Mix sample was diluted with 100 mM Sodium Acetate buffer solution (pH 6) and the volume adjusted to 100 µL. New reducing buffer solution (200 mM DTT stock) was prepared in water and in absence of ABC[5]. The disulfide bonds were reduced with 100 µl 5 mM DTT. The sample was incubated at 37 degrees and 200 rpm for 1 hour. Then residues of cysteine were alkylated with 100 µl of freshly prepared 10 mM IAA before incubating in a dark place for 30 minutes at room temperature.

Later on, the FASP procedure was conducted as usually by adding the whole sample on an equilibrated pall spin-filter and beginning centrifugation at 13000 rpm for 30 minutes at 20^{o}C to spin out eventual ABC buffer in the sample.

The centrifugation procedure was continued with Tris-HCl buffer # 1 as explained earlier in section 4.6. After that 100 µl 100 mM NaOAC buffer (pH 6) was added on the filter, it was shaken at 200 rpm for about 5 minutes and then centrifuged at 13000 rpm for 15 minutes at RT. This step was carried out three times.

NOTE: Protein solubility was controlled between each centrifugation[6]. If needed, the centrifugation can be prolonged if liquid remains on the filter.

Reductive Amination:

[5] 200 mM DTT in 50 mM ABC solution was used several times during the methylation process, but for optimal peptide quantification, this should not be utilized during this procedure because of interference.
[6] The protein solubility proved to be slightly low, as some insoluble protein was left on the filter.

The protein methylation procedure was introduced with 100 µl 100 mM NaOAC buffer added on the filter, combined with 10 µl 8% formaldehyde. The sample was shaken briefly at 400 rpm, and then was instantly mixed with freshly prepared sodium cyanoborohydride solution (520 mM, 10 µL)[7]. The sample was subjected to additional shaking at 400 rpm for a while, and then the sample was allowed to react for 5 minutes at RT. Consequently, the centrifugation was performed at 13000 rpm for 15 minutes at RT. This procedure was conducted twice with freshly prepared sodium cyanoborohydride solution.

After that protein methylation process was implemented, the FASP protocol was continued as usual (see section 4.6) by washing the sample three times with digesting buffer, and then replacing the filter on a new original tube in order to collect peptides overnight.
In brief, the sample for digestion consisted of 50 µL buffer # 3 plus 100 µL digesting buffer plus 2 µL 1µg/µL trypsin solution. Ultimately, the sample was incubated and digested under the same conditions as described previously in section 4.6.
In addition, the elution and desalting steps were performed in a same manner as stated earlier.

NOTE: According to previous research by Jue-Liang Hsu et al. described in ``Stable –Isotope Dimethyl Labeling for Quantitative Proteomics.`` 1 µl 4% formaldehyde and 1 µl 260 mM sodium cyanoborohydride were utilized for reductive amination. The amount of formaldehyde and cyanoborohydride was increased here due to the unsuccessful first methylation run. It was not possible to identify enough peptides since the amount of methylated peptides were insufficient.(61)

[7] In the original protocol for methylation, it said to use 1µL 4% CH_2O and 1µL 260 mM $NaBH_3CN$, but these amounts seemed to be inadequate.

4.8 Determination of protein solubility

In this section, the protein solubility was evaluated by utilizing three different buffer solutions working on FASP in order to determine the cause for inadequate number of protein family hits. The following buffer solutions were used:

- 50 mM ABC (100 µL) → freshly prepared!
- 0.1% SDS in 50 mM ABC (100 µL)
- 4M Guanidine HCl in 50 mM ABC (100 µL)

Six samples were run side-by-side or in parallel, where three samples underwent di-methylation labeling and the rest were not modified. The samples were spiked with buffer solutions in the order listed above, respectively.

All six samples containing 100 µg protein were exposed to reduction with DTT (10 mM, 100 µL) and were incubated at 37ºC for 1 hour, shaking at 200 rpm. Subsequently, proteins were carbamidomethylated with IAA (20 mM, 100 µL) followed by incubation darkly for 30 minutes at room temperature.

NOTE: DTT and IAA were made in each of the listed buffer solutions, such as
10 mM DTT/20 mM IAA in 50 mM ABC, 10 mM DTT/20 mM IAA in 0.1% SDS and
10 mM DTT/20 mM IAA in 4 M Guanidine HCl.

Six 30-kDa pall spin-filters were equilibrated with 100µL of each buffer solution, and were thereafter centrifuged for 10 minutes at 13000 rpm and 20ºC. Each pall spin-filter was labeled according to buffer solution and methylation versus no-methylation. Next, the samples were added to the equilibrated pall spin-filters, correspondingly in the same order as mentioned above. The following step was that all six pall-spin filters were centrifuged to spin down the samples under the same centrifugation conditions as explained previously in section 4.6.

The samples were then washed with the proper buffer solution (100 µL of each one) according to filter labeling. All centrifugation steps were conducted under the similar terms as described earlier in section 4.6.
Each filter was checked for protein solubility between every centrifugation, and the conclusion was that no optimal solubility appeared.

While three samples were prepared for the di-methyl procedure, the rest of them were placed in the refrigerator. The di-methylation procedure was carried out with NaOAC (pH 6) buffer solution in the same manner as mentioned previously in section 4.7. Consequently, the FASP procedure was continued by adding 50 µl of each buffer solution to both methylated and un-methylated samples followed by shaking at 400 rpm for ca. 4 minutes (see section 4.7). This addition happened after each filter was replaced into a new original tube in order to collect peptides by digesting overnight.

Proteolytic digestion of the samples was conducted by adding trypsin 1 µg/µL in a ratio of 1:50, thus 100 µL of each buffer solution was added to the samples, followed by incubation at 37ºC and 200 rpm overnight. Total volume on each pall spin-filter during digestion procedure was 152 µL.

NOTE: During the proteolytic digestion step, the last buffer solution, which was 4 M GuHCl, 50 mM ABC, were changed to 2 M GuHCl, 50 mM ABC in order to preserve the activity of trypsin which terminates in the presence of 4 M GuHCl (62).

On the next day, the samples were controlled for filter leakage. The elution and first desalting processes were implemented as outlined earlier in section 4.6. The exception here was that the samples, which contained sodium dodecyl sulfate (SDS) were desalted using the SCX (strong cation exchange) method by using 10 µL tip in order to get rid of SDS residuals before Q-TOF analysis.

Specific method for SCX (Bond Elut OMIX SCX) sample desalting

Interfering detergents from SDS samples were removed in accordance to Agilent Bond Elut OMIX SCX 10µl tip desalting's procedure (63).

Prior to MALDI analysis, the SDS samples were prepared this way:

After desalting of SDS containing samples, they were run with MALDI in order to check after SDS residues before analysis with n-LC-MS/MS. In theory, SDS residues should be removed during the desalting procedure, but in order to make certain that there was no SDS left, the samples underwent analysis with MALDI.

Alpha-cyano-4-hydroxycinnamic acid (CHCA) was the crystallized molecules containing matrix solution utilized for the UV MALDI instrument (wavelength 337 nm). The matrix solution consisted of water and organic solvent (acetonitrile, acetone, ethanol), which in reaction with the analyte permitted both hydrophobic and hydrophilic molecules to dissolve into the solution. First, 1µL of each SDS sample was speckled onto a MALDI target plate (a metal plate) at two specific positions, and then allowed to vaporize for about one minute at RT. The sample: matrix ratio was 1:1000 (w/w) or 1:1 (v/v). Thereafter 1µL of the matrix solution was spotted onto the same position as the samples. The matrix solvents allowed evaporation, hence only the recrystallized matrix was left on the plate. In this way, both the matrix and analyte molecules were inserted into MALDI crystals, hence the matrix and analyte were co-crystallized. Co-crystallization is said to be the major factor in choosing a suitable matrix solution. Therefore, an excellent crystal is important in order to acquire a great quality mass spectrum of the analyte of interest.(64)

Consequently, the MALDI target plate with embedded sample and matrix was placed into the MALDI instrument and started the analysis with a mass range between 50-800 m/z. A software named MassLynx version 4.1 was used to generate the results as mass spectra.

Thereafter, all the desalted samples were evaporated completely by heated nitrogen gas at 40 degrees for about 1 hour. Next, the dried samples were re-suspended in 12 µL 0.1% TFA solution, spun down and transferred to 6 respectively auto sampler vials for analysis with LC-MS/MS

4.9 COFRADIC™

In this section, both acetylation and TNBS modification processes will be described in details. Afterwards UV-HPLC instrumentation will be used to initially run a primary COFRADIC by collecting 20 fractions followed by a secondary COFRADIC of preferred fractions from primary run.

Chemicals required

- Ice-cold Ethanol
- 600 mM Sulfo-NHS –acetate (stored in the freezer) [MW: 259.17]
- Urea [MW: 60.06]
- 200 mM Sodium hydrogen phosphate anhydrous [MW: 358.14]
- 1 M Hydroxylamine [MW: 69.49]
- 6 M Sodium hydroxide [MW: 40.00]

In Solution Sample Preparation

Prior to primary COFRADIC, the sample was prepared this way:

This procedure was initiated with 600 µg protein. The calculated volume from the BugBuster Master Mix sample was transferred to an Eppendorf tube. The protein precipitation was carried out with nine volume units of ice-cold ethanol[8]. The sample was shaken to get a homogenous solution. Consequently, the sample was inserted in the freezer at -20°C for a minimum of 4 hours so protein precipitation could occur.

NOTE: For larger sample volume, a 15 mL Falcon tube was used instead of a regular Eppendorf tube.

Thereafter, to gain the pellet, the sample was centrifuged at 13000 rpm for 15 minutes at 4 degrees. After centrifugation, the excess ethanol was poured gently and then the pellet was washed with one volume unit of ice-cold ethanol. Again, the surplus of ethanol was slowly drained and the pellet allowed drying completely at RT. If necessary, nitrogen-gas was used to accelerate the drying process (the pellet was not subjected to heat).

[8] In the 1th COFRADIC run, the volume of BugBuster Master Mix sample (BCA 4.30 mg/mL) used was 139 µl (equal to 600 µg protein), thus 9x139 µl = 1255 µl ice-cold ethanol was added.

Subsequently, the dried pellet was dissolved in 85 µl 6M Guanidine HCl in 50 mM ABC buffer (protein to urea ratio 1:7 µg/µl). Thereafter, the proteins were oxidized with 5 mM DTT (used 200 mM stock) as a final concentration and then the reduction of proteins allowed to arise by incubation for 45 minutes at room temperature. The subsequent step was alkylation of proteins by adding 15 mM IAA as a final concentration and thus the sample was incubated darkly for 30 minutes at room temperature[9].

NOTE: In this procedure, the IAA buffer was made in water and without ammonium bicarbonate.

Thereupon, to reduce the Urea concentration from 6 M to 0.85 M, the sample was diluted in an adequate volume of 50 mM ABC buffer in order to gain a final concentration of 1 µg/µL for the whole sample. This means that since the start amount of protein was 600 µg, thus the total volume ended with 600 µL[10]. Trypsin 1ug/µL solution was added to the sample in a ratio of 1:75 (protein to trypsin ratio 1:75). In this case, the sample was spiked with 8µL 1ug/µL trypsin solution (1:75x600µg = 8µg trypsin). The sample was digested overnight in an incubator at 37 degrees and 200 rpm.

Primary COFRADIC

Before primary collection of peptides, the pH value of the digested sample was adjusted to below four (pH 2 was achieved) (acidic solution). Prior to adjusting the sample was around pH 8 (alkaline solution).

Consequently, 100µL of the sample was transferred to a UV-HPLC vial. The protein amount for analysis was around 100µg. Thus, primary COFRADIC was run by loading the sample on a RP-HPLC column connected to a PDA detector and 20 fractions was collected within 40 minutes.

The fraction collection procedure was conducted by supplying Fraction Collector III device that was tuned on time mode and parameter number one in order to collect 1 mL/min fractions (2

[9] In the 1th COFRADIC run, 2.18µL 200 mM DTT + 1.32µL 1M IAA were added to the sample.
[10] 512µL 50 mM ABC buffer was added to the sample to achieve a total volume of 600µL (600µL – 85 - 2.18 – 1.32 ≈ 512µL).

mL/tube) during 40 minutes. Initially, the UV-HPLC apparatus was equilibrated for 15 minutes before sample run. The utilized Instrument method was C 70 40 min (the software used here was EmPower version 2).

Thereafter, the chromatogram was evaluated and in light of those results, fraction F5, F6 and F7 were selected for further treatment with sulfo-NHS-acetate to make a secondary COFRADIC run. In addition, F18, F19 and F20 were applied to test the pH adjustment of the treated fractions by using 6 M NaOH.

Table 5: Primary COFRADIC overview for 20 fractions and elution time.

Fraction Number	Elution time (minutes)
1	0 – 2
2	2 – 4
3	4 – 6
4	6 – 8
5	8 – 10
6	10 – 12
7	12 – 14
8	14 – 16
9	16 – 18
10	18 – 20
11	20 – 22
12	22 – 24
13	24 – 26
14	26 – 28
15	28 – 30
16	30 – 32
17	32 – 34
18	34 – 36
19	36 – 38
20	38 – 40

The highlighted rows represents elution time for the selected fractions, which F6 & F7 were treated with sulfo-NHS-acetate.

Table 6: Gradient of elution program for RP-HPLC-UV, used for the initial COFRADIC runs.

#	Time (min)	Flow (ml/min)	%A (water)	%B (ACN)	Curve
1	Initial	1.00	98.0	2.0	Initial
2	3.00	1.00	98.0	2.0	6
3	30.00	1.00	30.0 (60)	70.0 (40)	6
4	31.00	1.00	5.0	95.0	6
5	38.00	1.00	5.0	95.0	6
6	39.00	1.00	98.0	2.0	6
7	40.00	1.00	98.0	2.0	6

The highlighted row is linked to the main gradient change that occurred between the initial COFRADIC procedure and the subsequent ones.

Table 7: Mobile phase composition used for UV-HPLC apparatus.

Mobile phase A	Mobile phase B
95% H_2O	95% MeCN
5% MeCN	5% H_2O
0.1% TFA	0.1% TFA

Enrichment of N-terminal Peptides Using COFRADIC according to Venne et al: (65)

Fractions F5, F6 and F6 were chosen to undergo a secondary COFRADIC run because these fractions might comprise an adequate amount of peptides. As previously mentioned, peptide separations were conducted utilizing a Waters 2695 HPLC system. The separation of peptides occurred with an enhanced gradient to easily separate peptides, thus fractions were collected using a fraction collector instrument, which was couplet to the HPLC system.

The collected fractions from the primary run were reduced to approximately 40 µl under heated nitrogen gas at 60 degrees for about 2 hours. Next, fraction 6 and 7 were chosen to undergo derivatization with sulfo-NHS-acetate. After volume reduction, 200 mM Na_2HPO_4 (pH 8) was used to bring the total volume to 300µl and to obtain a final pH value of about 7. This procedure was also performed for fraction 5, even though this fraction was not modified. All three

fractions were pH adjusted with 6 M NaOH solution (6 µL). Next, derivatization with sulfo-NHS-acetate of free N-termini of internal peptides were achieved in two steps according to Staes et al.(56)

At the beginning, F6 and F7 (consisting of around 300 µL) were spiked with freshly prepared sulfo-NHS-acetate to a final concentration of 20 mM, and the sample was incubated for 1 hour at 37 ºC and 200 rpm. Consequently, additional 10 mM sulfo-NHS-acetate was added under the identical reaction conditions. After 2 hours of total incubation time, the excess of sulfo-NHS-acetate was quenched by adding Hydroxylamine to a final concentration of 100 mM (used 1 M stock, 30 µL[11]).

Thereafter, volume reduction was conducted for the fractions with N_2-gas at 60 ºC for few minutes and brought it to ca. 100 µL. Next, pH value was reduced to 1-2 by using pure TFA. Consequently, all 100 µL of the sample was transferred to a UV-HPLC vial and re-loaded onto the RP-HPLC column, then secondary COFRADIC was started by collecting 20 fractions for every derivatized and underivatized primary fractions.

[11] Actually, 35 µL 1M hydroxylamine should be added to obtain a final cons. of 100 mM.

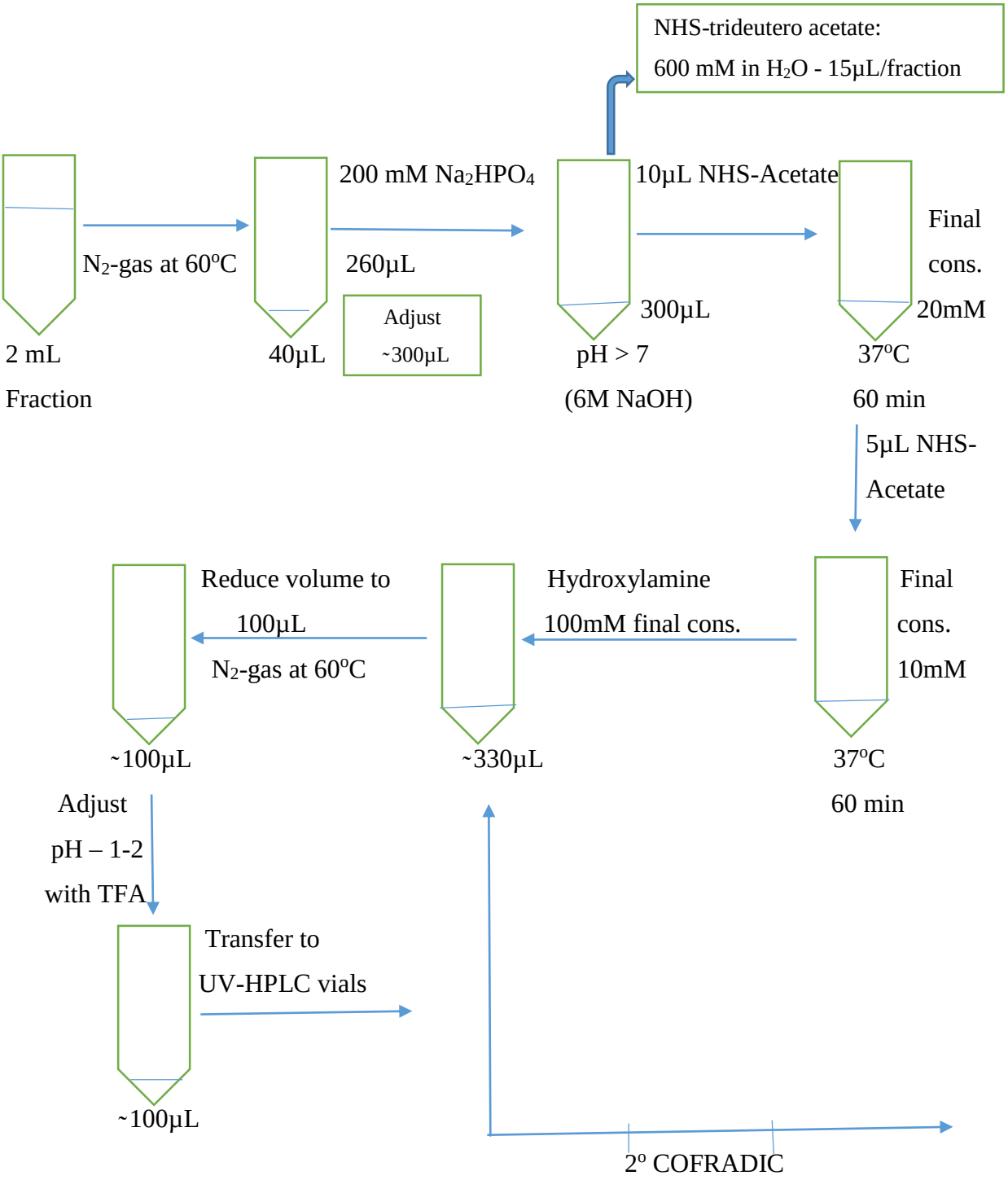


Figure 20: General workflow for secondary COFRADIC sample derivatization.

Secondary COFRADIC

The secondary run started with equilibration of the instrument for 15 minutes, therefore the sample ran with an injection delay time set to 15 minutes between each run. The fractions were collected at equal times to the primary fractions but with elution time of plus/minus 0.5 minutes.

The fraction collector instrument was still set to time mode, but parameter 2 was used for collecting peptides in 4 mL/tube. This method led to the subsequent collections pattern as showed in table 7.

Table 8: Collection pattern for secondary COFRADIC in the initial stage.

Fraction 5	**Fraction 6**	**Fraction 7**
1 → 7.5 – 10.5 min (3 mL)	1 → 9.5 – 13.5 min (4 mL)	1 → 11.5 – 15.5 min (4 mL)
	2 → 13.5 – 17.5 min (4 mL)	2 → 15.5 – 19.5 min (4 mL)
	3 → 17.5 – 21.5 min (4 mL)	3 → 19.5 – 23.5 min (4 mL)
	4 → 21.5 – 25.5 min (4 mL)	4 → 23.5 – 27.5 min (4 mL)
	5 → 25.5 – 29.5 min (4 mL)	5 → 27.5 – 31.5 min (4 mL)
	6 → 29.5 – 33.5 min (4 mL)	6 → 31.5 – 35.5 min (4 mL)
	7 →33.5 – 37.5 min (4 mL)	7 → 35.5 – 38.5 min (3 mL)
	8 → 37.5 – 38.5 min (1 mL)	

After secondary run collection of the fractions listed in table 7, they were completely evaporated under the same conditions as described for primary COFRADIC. F5-1, F6-1, F6-2, F6-3, F7-1, F7-2 and F7-3 were chosen to be analyzed with n-LC-MS/MS. The basis for this choice was that these fractions might include an appropriate quantity of peptides. Thus, each dried fraction was re-suspended in 15 µL 0.1% TFA solution and then transferred to auto sampler vials for analysis on Q-TOF instrument.

When it comes to the subsequent COFRADIC runs, so the instrument method was changed to a more efficient gradient. Mainly, the change was that the new instrument method ``C 2 40 40 min`` was applied, which the acetonitrile gradient reduced from 70% to 40% in order to make peptide separation more powerful. In addition, during both primary and secondary COFRADIC run, the fraction collector apparatus was tuned to time mode and parameter number 1 (1 mL/min → 2 mL/tube → 40 minutes run). Thus, primary and secondary

run were performed with the same instrument method in order to collect 20 fractionations for each derivatized or underivatized fraction and then the most enriched fractions were selected to undergo analysis on the Q-TOF system. Otherwise, the sample preparation was done in a similar manner as described earlier in this section. The main difference between the COFRADIC runs was that the started protein amount analyzed gradually increased from 100 µg at the beginning to about 600 µg later on.

4.10 In Solution Sample Preparation according to Venne et. al.

Required Solutions:

- 6 M Guanidine HCl, 50 mM Na_2HPO_4 (Lysis buffer) (pH 7.8)
- Methylation Buffer composition: 20 mM CH_2O (formaldehyde), 40mM $NaBH_3CN$ (sodium cyanoborohydride), 200mM HEPES (pH 8.0)
- 1 M Glycine [MW: 75.07]
- 1 M Hydroxylamine [MW: 69.49]
- 50 mM ABC, 5% ACN, 1 mM $CaCl_2$
- 2 M Guanidine HCl, 50 mM Na_2HPO_4

This procedure was carried out as described in the Venne et al. research.(65)
Protein quantity that has been used here was 100 µg. Two samples were prepared: methylation versus no-methylation.

Sample preparation:

The samples underwent dilution in 100 µl 6 M GuHCl, 50 mM Na_2HPO_4 (pH 7.8). The proteins went through further reduction with 15 mM DTT for 30 min at a temperature of 56°C. Subsequently, the free thiol groups were carbamidomethylated with 30 mM IAA and incubated for 30 min at room temperature in a dark place.

Through the alkylation reaction, it is crucial that the pH is between 7 and 8. Otherwise, Cys-comprising peptides will be lost and then the alkylation reaction will become inadequate.

Dimethyl Labeling on the Protein Level:

The sample was incubated in di-methyl labeling buffer for a period of 2 hours at 37°C for a particular di-methylation of free N-termini and lysine residues which adheres in accordance with Jentoft et al (66). The buffer solution consisted of 20 mM CH_2O, 40 mM $NaBH_3CN$, and 200 mM HEPES[12] and was adjusted to pH 8.0 before being added to the sample. The excess of formaldehyde was quenched by adding glycine to a final concentration of 60 mM (w/v) for a period of 10 minutes at RT. Later, the sample was treated with 130 mM (w/v) hydroxylamine for a period of 15 minutes at room temperature.

NOTE: Both samples were treated with Glycine (60 mM as a final concentration) and Hydroxylamine (130 mM), even if one sample did not undergo the di-methylation procedure, in the interest of comparison between the two samples.

Proteolytic Digestion:

The samples were made to undergo ethanol precipitation before proteolytic digestion. The samples were diluted 10 times with ice-cold ethanol and stored in the freezer overnight (-20^{o}C). On the next day, the samples were centrifuged for 30 minutes at 16000 rpm and a temperature of 4°C. After the centrifugation of both samples, the supernatant was removed thoroughly and the pellet was allowed to dry with nitrogen gas. Consequently, the pellet was re-solubilized in 10µL 2 M GuHCl, 50 mM Na_2HPO_4, (pH 7.8). Thus, the samples were diluted with 100µL 50 mM ABC, 5% ACN, 1 mM $CaCl_2$. Both samples were gently shaken in order to gain a homogeneous solution. The samples underwent proteolytic digestion with trypsin in a ratio of 1:20 (w/w) (1µg/µL, 5µL) and sat to incubation overnight at a temperature of 37°C. The absolute volume of each sample was 115µL, thus a concentration of 0.87 µg/µL.

On the next day (after 16 hours of incubation), the samples were desalted with the C18 tip method similarly as described previously. In contrary to the previous samples, the peptide amount used for LC-MS/MS analysis in this procedure was 34.8 µg (40 µL x 0.87 µg/µL). Consequently, two peptide mixtures containing 12 µL 0.1 % TFA were transferred into two auto sampler vials and run analysis with LC-MS/MS.

[12] Volume of light buffer that was added to the methylated sample was equivalent to the existing sample volume (127µL). The absolute volume of the di-methylated sample was larger than the other one.

4.11 Q-Exactive versus Q-TOF sample selection

The sample used for the initial Q-Exactive run was prepared almost similarly to the one described in section 4.9.2. Unlike the previously prepared sample, this procedure consisted of two samples, where the first one contained regular final concentrations of reduction and alkylation buffers (5 mM DTT, 15 mM IAA) and the other one included 2-folded concentrations (10 mM DTT, 30 mM IAA). Otherwise, both samples were prepared in the same manner as the one in section 4.9 (600 µg protein with a concentration of~1µg/µL).

First, 25 µL of the regular and the 2-folded sample was transferred into two respectively Eppendorf tube, then each one of them was diluted in 75 µL 0.1% TFA. The protein amount in each tube was 25 µg (Q-TOF samples). Subsequently, the samples were desalted according to the C18 tip method.

The dried samples were then re-suspended in 14 µL 0.1% TFA solution. From the regular sample, 2 µL was transferred to a new Eppendorf tube and was spiked with additional 10 µL 0.1% TFA (Q-Exactive sample). Next, the same was done for the 2-folded sample. Thus, four samples (two Q-TOF versus two Q-Exactive) were transferred to four auto sampler vials for analysis with Q-TOF and Q-Exactive. Protein concentration for the Q-TOF samples were 1.78 µg/µL (25 µg / 14 µL). The amount of protein used for Q-Exactive was 3,56 µg, thus a concentration of 0.3µg/µL. Each vial contained a 12µL sample.

NOTE: Underivatized samples underwent analysis with Q-TOF and Q-Exactive.

4.12 Sorting of N-terminal peptides with TNBS on peptide level

The required solutions:

- 5% Picrylsulfonic acid (TNBS) solution (equal to 170mM) [MW: 293.17] (stored cold)
- 50mM Boric acid, 25 mM NaOH (pH 9.5) [MW: 61.83]
- 30% Hydrogen peroxide solution (stored cold)

Prior to primary and secondary COFRADIC runs, the sample was prepared in the same way explained earlier in section 4.9. Two samples were prepared in parallel. One sample contained a normal concentration of DTT and IAA (5 mM DTT and 15 mM IAA), while the other one

consisted of two-folded concentrations of DTT and IAA. Protein precipitation occurred with ice-cold ethanol. The introduced protein quantity during this process was 600 µg.
The sample, which contained normal concentrations of DTT and IAA, was chosen to undergo primary and subsequently secondary COFRADIC. This sample was used earlier for assessment of Q-Exactive and Q-TOF analysis, so the remaining protein quantity was 575µg with a concentration of~1µg/µL.

The methionine`s were oxidized with hydrogen peroxide solution in order to reach a final concentration of 0.5% and then the sample was incubated for 30 minutes at 30ºC. After half hour of incubation, the volume was reduced to 100 µL by utilizing heated nitrogen gas at 60ºC for a couple of minutes. The color changed from white to yellow after addition of H_2O_2. For primary COFRADIC separation, the sample (100 µL) was loaded onto the reverse-phase column and fractionated into 20 consecutive fractions of 2 minutes using instrument method C 2 40 for 40 min.

After collection of 20 fractions by the primary run, F11, F12, F13 and F14 were selected to be subjected to derivatization with TNBS (2,4,6-trinitrobenzenesulfonic acid) and then the secondary run. These fractions were initially completely dried by heated nitrogen gas at 60ºC for ca. 2 hours. Thus, each primary fraction was re-suspended in 50µL 50 mM sodium borate, 25 mM NaOH buffer (pH 9.5).

The selected fractions were treated with freshly prepared TNBS as follows:

- F11 – No TNBS (Loaded volume on RP-HPLC column was 50 µl)
- F12 – TNBS (15 mM, 10 µl) (Loaded volume on RP-HPLC column was 80 µl)
- F13 – partially modified with TNBS (F13A, 15 mM, 10 µl)
  (Loaded volume on RP-HPLC column was 80 µl)
- F14 – two-folded volume and concentration of TNBS (30 mM, 20 µl).
  (Loaded volume on RP-HPLC column was ca. 100 µl)

Consequently, the TNBS modified samples were incubated at a temperature of 37 ºC for 1 hour. To ensure close quantitative TNBS modification of free alpha-amino groups, this step was done three times. After three hours with incubation of fraction F12, F13A and F14, both F13A and B were pooled and all fractions were re-loaded onto the RP-HPLC column one by one. Each fraction was then fractionated with the similar solvent gradient used during the

primary run. The N-terminal peptides were gathered in 20 equivalent-volume secondary fractions, each in a 40-minute-long time intermission.

The subsequent step was to make the correct selection of several secondary fractions for further LC-MS/MS analysis. The selection pattern was as follows:

Table 9: Selections pattern for secondary COFRADIC fractions.

F11 (1º run)	**F12 (1º run)**	**F13 (1º run)**	**F14 (1º run)**
F11-10 (2º run: n-1)	F12-11 (2º run: n-1)	F13-12 (2º run: n-1)	F14-13 (2º run: n-1)
F11-11 (2º run: n)	F12-12 (2º run: n)	F13-13 (2º run: n)	F14-14 (2º run: n)
F11-12 (2º run: n+1)	F12-13 (2º run: n+1)	F13-14 (2º run: n+1)	F14-15 (2º run: n+1)
F11-13 (2º run: n+2)	F12-14 (2º run: n+2)	F13-15 (2º run: n+2)	F14-16 (2º run: n+2)
F11-14 (2º run: n+3)	F12-15 (2º run: n+3)	F13-16 (2º run: n+3)	F14-17 (2º run: n+3)
F11-15 (2º run: n+4)	F12-16 (2º run: n+4)	F13-17 (2º run: n+4)	F14-18 (2º run: n+4)
F11-16 (2º run: n+5)	F12-17 (2º run: n+5)	F13-18 (2º run: n+5)	F14-19 (2º run: n+5)
F11-17 (2º run: n+6)	F12-18 (2º run: n+6)	F13-19 (2º run: n+6)	F14-20 (2º run: n+6)

These fractions from secondary run were evaporated and dried for approximately 2 hours with heated nitrogen gas at 60ºC. Next, each fraction was re-suspended in 15µL 0.1% TFA solution and then transferred to auto sampler vials for LC-MS/MS analysis.

4.13 Sulfo-NHS-acetate versus Di-methyl labeling on the protein level and TNBS on the peptide level

The adapted protocol described in section 5.9, was implemented during this procedure to prepare two consecutive samples by treating with di-methylation and acetylation on the protein level, then modifying them with TNBS on the peptide level in order to determine the enrichment of N-terminal peptides on both samples, and then to compare both modification methods.

The quantity of protein applied in this process was 500 µg. Protein precipitation occurred according to the protocol on section 4.9. One sample was di-methylated while the other one was acetylated. In the first round, 2 M Urea (pH~ 8.5) was used as lysis buffer because ammonium bicarbonate could not be applied during di-methylation modification step. The reason for this buffer choice was that the physiochemical properties of urea makes it capable of

creating more efficient peptide separation and improves detection of modified peptides. Thus, after protein precipitation on ice with ethanol, the pellet was re-solubilized in 2 M Urea buffer (500 µL). Proteins were then reduced and subsequently carbamidomethylated with
10 mM DTT and 20 mM IAA added respectively as final concentrations. This happened under the same reaction conditions as described previously according to Venne et al. The next round consisted of two samples spiked with 2 M Guanidine HCl (pH~8.5) instead of 2 M Urea. This procedure was performed in almost exactly the same manner as the previous one.

Urea treated samples were di-methylated with freshly prepared light-buffer composition, while Guanidine HCl samples were modified with reductive amination buffer composition.

NOTE: DTT and IAA were prepared in water during this process. Protein precipitation was conducted twice with ice-cold ethanol in order to enhance protein solubility.

A summary of sample preparation

500 µg protein was distributed into four tubes. Nine volume units of ice-cold ethanol were added, and the samples sat on ice until the next day. On the next day, all four samples were centrifuged at 13000 for 15 minutes and 4^{o}C in order to gain the protein pellet. The supernatant was removed by a gentle pipette technique and the pellet was carefully vaporized with nitrogen gas. The pellet was redissolved in 500 µL 2 M Urea (pH~8.5) / 500 µL 2 M Guanidine HCl (pH~8.5). Next, the proteins were reduced and alkylated in the same manner as mentioned above.

The acetylation modification step on the protein level was conducted in the method described by Venne et al. In addition, di-methyl labeling on the protein level was performed according to Venne et al for to the sample treated with Urea. In contrast, the sample that was spiked with Guanidine HCl underwent reductive amination on the protein level in congruence with the theory of Venne et al. The methylated samples were furthermore treated with
60 mM Glycine for 10 minutes at RT followed by 130 mM Hydroxylamine for 15 minutes at RT. The acetylated samples were treated only with 100 mM Hydroxylamine in order to quench sulfo-NHS-acetate residues.

Before the second protein precipitation round, pH value of the samples was measured, which was near 8 for Guanidine HCl samples and between 5-6 for Urea. It was therefore adjusted to 8 with 6 M NaOH. The samples were subjected to a second protein precipitation as described above. The pellet was then re-solubilized in 500 µL freshly prepared 50 mM ABC buffer (pH~ 8). Finally, the samples were spiked with trypsin in a ratio of 1:75 and then digested on an incubator at 37ºC overnight (200 rpm).

On the next day, the digested samples were pH adjusted to acidic conditions. The pH value of the samples were around 8 before adjustment, this was because trypsin is not active in an acidic medium (pH below 7). Prior to the primary COFRADIC run, the samples were pH adjusted to 1-2 by addition of pure TFA solution. Next, methionine`s were oxidized with H_2O_2 to gain a final concentration of 0.5%, and all samples were incubated beneath the same reaction conditions as explained previously in section 6.13[13]. Afterwards, primary COFRADIC was run with the same instrument method as used earlier in order to collect 20 equivalent-volume fractions. After sorting of N-terminal peptides onto 20 primary run fractions, all of them were treated further with TNBS derivatization on the peptide level in order to make the peptides more hydrophobic and thus provide a longer retention time as previously described (see section 4.12).

NOTE: An error occurred during this step, which was that the selection pattern was initially similar to the one described before. The correct selection pattern actually was that all 20 fractions were intended to undergo derivatization by TNBS.

In regard to the acetylation fractions (comprised Urea buffer), F13 and F14 were treated by partial modification with TNBS (F13) and 2-folded concentration (F14 – 30 mM TNBS). Otherwise, the rest of the fractions were derivatized with freshly prepared 10µL
15 mM TNBS and then were incubated for 1 hour at 37 degrees. This was done three times in order to obtain optimal peptide quantification.

Each derivatized fraction from either sulfo-NHS-acetate or di-methyl (Urea and Guanidine HCl containing samples) was re-loaded onto the RP-HPLC column for supplementary fraction collection by a secondary COFRADIC run. Every primary run fraction was moreover

[13] Unlike the previous sample, no color change occurred by addition of H_2O_2 during this procedure.

fractionated into 20 secondary run fractions, and hence from each secondary fraction, n-1, n and n+1 were selected for analysis with LC-MS/MS.

Since only 20 fractions were obtained from the secondary run, so from F1 and F20 only n and n+1 (F1), n-1 and n (F20) were chosen for analysis with Q-TOF. After the secondary run fraction collection, the fractions were dried entirely with heated nitrogen gas at 60^{o}C for about 1-2 hours, then each fraction n-1, n and n+1 was re-suspended in 15 µL 0.1% TFA solution (freshly prepared). Ultimately, all three fractions were pooled to one single sample (total sample volume was 45 µL). Next, each sample was transferred to an auto sampler vial and subjected to LC-MS/MS analysis.

Fraction selection for Q-Exactive analysis:

From previously made di-methyl and acetyl samples in the urea buffer solution, a variety of fractions were selected from each sample separately as the odds and evens, then they were pooled together.

5. RESULTS

The purpose of proteomics is to recognize, distinguish and sketch gene functions at the protein level for entire cells, tissues or whole organism. A conventional empirical plan for large-scale proteomics analysis includes fractionation of a complex protein mixture by chromatographic approaches succeeded by consequent components identification in the distinct fractions by mass spectrometry. Due to persistent and quick improvement in instrument susceptibility, throughput capacity, programming flexibility, and strategies of statistical verification of the data, mass spectrometry based methodologies are becoming the prevailing approaches in analysis of a proteome. Owing to this technology, proteomics have become a vital research field for investigation of disease, drug discovery and prevention of ailments.

This dissertation is generally based on qualitative proteomics, and our main goal was to develop a COFRADIC method for identification of N-terminal peptides by the RP-HPLC system.

It is not an easy task to develop a COFRADIC RP-HPLC system for proteomics, but it is a powerful tool to overcome one of the main obstacles in qualitative proteomics - to reduce the intricacy of a complex peptide mixture.

The initial part of this thesis focuses on achievement of adequate protein solubility in order to obtain a sufficient degree of modifications, then to create better separation in the HPLC system and better identification with LC-MS/MS analysis.

The second part of this dissertation focuses on a comparison between acetylation and di-methylation as two widely utilized modifications methods in order to determine the efficiency of each of them when it comes to identification and separation of proteins and peptides. In addition, we aimed to investigate a number of identified N-terminal peptides within a sample through a primary and secondary COFRADIC run.

5.1 Results from FASP sample runs

The FASP protocol was initially implemented to determine the abundance of identified peptides in differently prepared samples in order to evaluate the quality of samples based on di-methyl modification versus no modification. The overall outcomes are summarized in graphic representation as shown below.

The initial FASP sample was not altered, while the other three samples were di-methylated on the protein level. Moreover, each samples was analyzed directly in n-LC-MS/MS apparatus with the similar elution gradient mode. Filter leakage occurred during the unaltered sample and within the di-methylated samples containing 25 mM and 520 mM sodium cyanoborohydride. Reducing and alkylation agents (DTT and IAA) were added in a concentration of 5 mM DTT and 10 mM IAA in each of the mentioned samples. In addition, 100 µg protein was applied within each analysis. Trypsin was added to the samples in a ratio of 1:50.

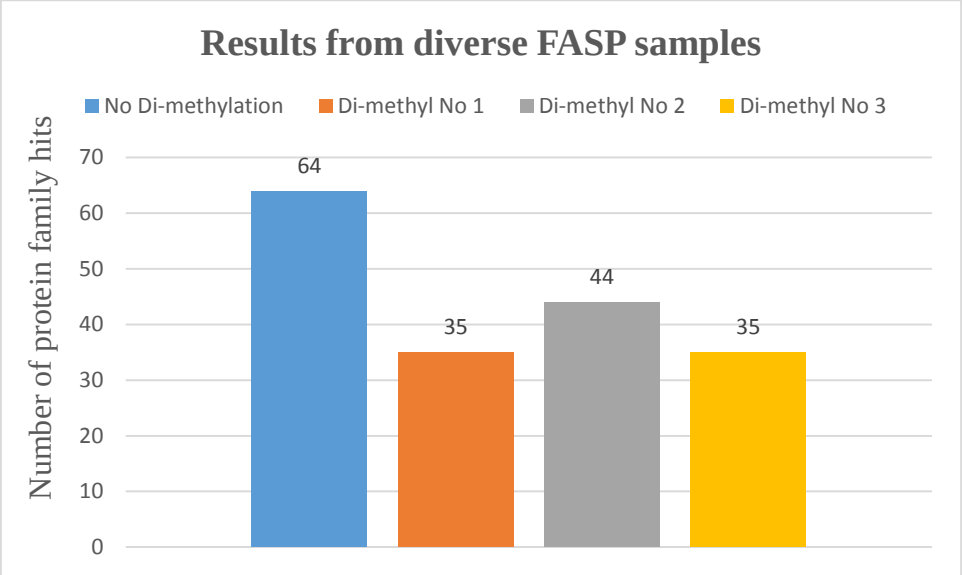


Figure 21: This figure shows four FASP samples, that were prepared differently. The first sample (labeled blue) did not contain any modification agent but was oxidized and alkylated as usual. The second sample (labeled orange) was di-methylated, but by adding the wrong amount of cyano (25 mM was added instead of 260 Mm – Not freshly prepared and filter leakage occured). The grey bar represents the di-methylated sample with 260 mM Cyano and 4% formaldehyde that was stored in the freezer for a long period of time before analysis. Yellow bar represents the di-methylated sample with 520 mM Cyano and 8% formaldehyde (Filter leakage occurred).

All samples in figure 21 were run on a n-LC-MS/MS system with the same elution gradient, and the results clearly shows that the occurrence of protein family hits was higher in samples that is not di-methylated than the modified samples.

MASCOT searches were carried out by Mascot Daemon v2.3.2 from Matrix science. The search parameters utilized are summarized in the following table. Di-methylated samples were searched with 4 missed cleavages due to the fact that trypsin will not cleave after di-methylated lysine`s.

Table 10: The MASCOT search was conducted under the following parameters:

Parameter	**FASP No modification**	**Di-methyl Protein level**	**Di-methyl Peptide level**
Type of search	MS/MS Ion search	MS/MS Ion search	MS/MS Ion search
Database	SwissProt	SwissProt	SwissProt
Enzyme	Trypsin	Trypsin	Trypsin
Taxonomy	Escherichia coli	Escherichia coli	Escherichia coli
Max. Missed cleavages	2	4	4
Fixed modifications	Carbamidomethyl (C) (IAA)	Carbamidomethyl (C) + Dimethyl (K)	Carbamidomethyl (C) + Dimethyl (K)
Variable modifications	Oxidation (M) (DTT)	Oxidation (M) + Dimethyl (Protein N-term)	Oxidation (M) + Dimethyl (N-term)
Mass values	Monoisotopic	Monoisotopic	Monoisotopic
Peptide mass tolerance	100 ppm	100 ppm	100 ppm
Fragment mass tolerance	0.1 Da	0.1 Da	0.1 Da
Peptide charge	2+, 3+, 4+	2+, 3+, 4+	2+, 3+, 4+
Instrument type	ESI-QUAD-TOF	ESI-QUAD-TOF	ESI-QUAD-TOF

Determination of protein solubility

Throughout this experiment, three buffer solutions were studied in regard to assessment of protein solubility. Six samples were run in parallel: three were di-methylated and the rest were not altered. The protein quantity was equal in each sample (100 µg), and both reductions and alkylations reagents were prepared in the respective buffer solutions. This process was carried out on a pall spin-filter device in the same manner as described previously in chapter 4. Every sample was incubated under the same reactions condition, and the only differences between them were buffer solution type and modification or non-modification. Otherwise, they were equal, but within proteolytic digestion with trypsin, the guanidine concentration was decreased from 4 M to 2 M in the interest of preservation of trypsin activity. Trypsin was added in a ratio of 1:50.

Ammonium bicarbonate and guanidine HCl samples were desalted with the C18 tip method, while SDS containing samples were desalted with the SCX 10 µl tip prior to analysis in MALDI in order to ensure elimination of SDS residues before MS analysis.

This process was conducted twice due to filter leakage during the first run, as shown in figure 22.

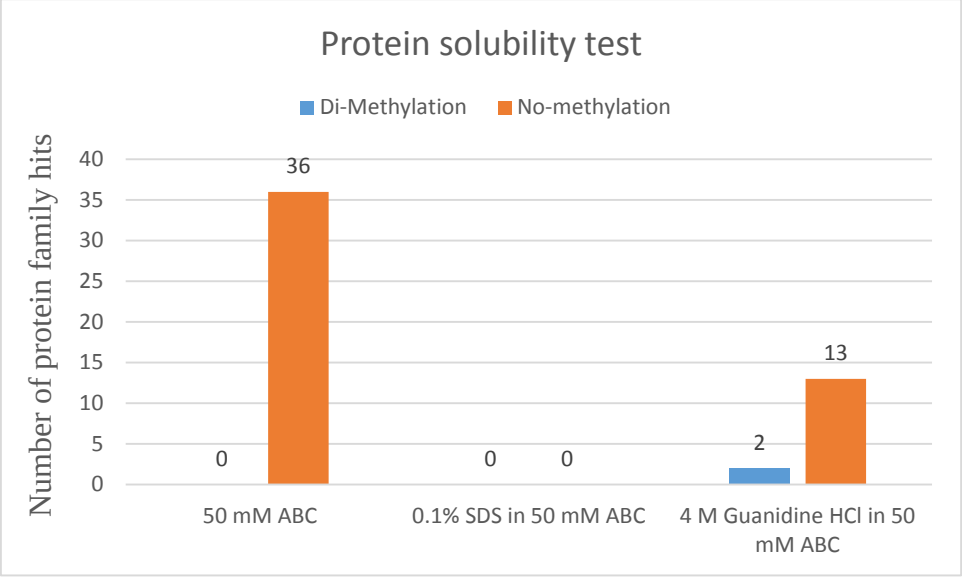


Figure 22: Protein solubility test performed with 3 various buffers. Six samples divided into di-methylation and no-methylation. Unsatisfactory results due to filter leakage.

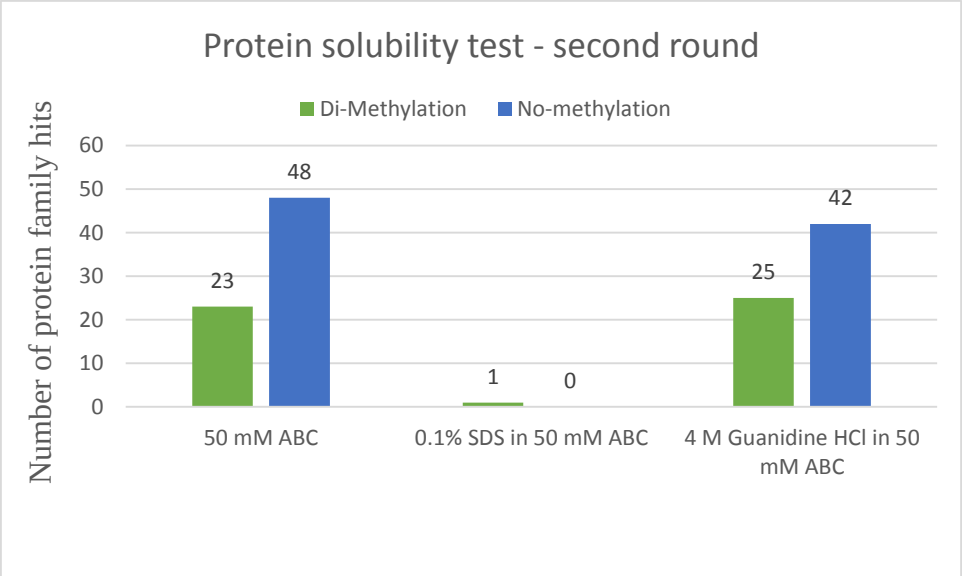


Figure 23: The second round of protein solubility test avoiding filter leakage. A noticeable number of protein family hits has occurred.

This sample set was also prepared in the same manner as explained above, but during the second round, the leakage problem was fortunately solved, and this was clearly showed by the results in graph 23. Protein mixture samples containing 0.1% TFA were loaded into nano-Aquity column for analysis with a n-LC-MS/MS instrument. Mascot searches were conducted with the parameters specified in table 10.

5.2 In Solution Sample Preparation

The di-methylation was conducted in accordance with Venne et. al (65), with the differences that Venne et al used *S. cerevisiae* as a protein source and the study was quantitative. Protein precipitation was carried out with ice-cold ethanol after di-methyl labeling and prior to proteolytic digestion. Two samples were prepared, one di-methylated and one unmodified. In regard to figure 24, the protein amount was 100 µg as usual. The proteins were dissolved in 6 M Guanidine HCl, 50 mM Na_2HPO_4 (pH 7.8). The proteins were reduced and alkylated with 15 mM DTT and 30 mM IAA, respectively. Reduction of proteins was performed by incubation for 30 minutes at 56ºC followed by alkylation for 30 minutes at room temperature. For particular di-methylation of free N-termini and lysine residues, the sample was incubated in a buffer composition of 20 mM CH_2O, 40 mM $NaBH_3CN$, and 200 mM HEPES (p H 8.0) for 2 hours at 37 ºC. During proteolytic digestion, trypsin was applied in a ratio of 1:20.

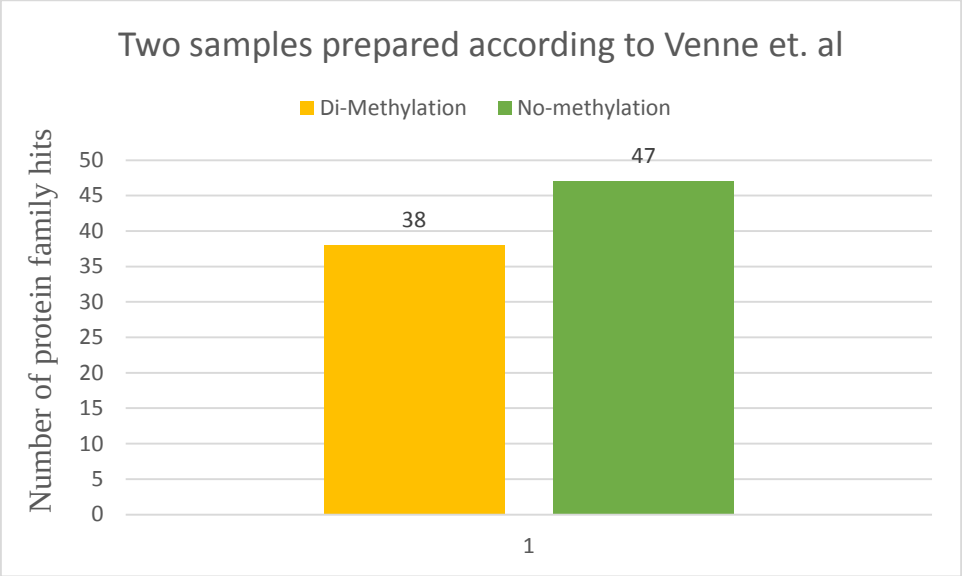


Figure 24: These two samples were prepared according to Venne et al. Di-methyl labeling is performed on the protein level.

In order to certify that the samples are of adequate quality when it comes to degree of purification and absence of detergents, four samples were prepared to undergo analysis in Q-TOF and Q-Exactive. Q-Exactive operates with high resolution and sensitivity compared to Q-TOF. Two unmodified samples containing 600 µg protein were prepared and involved various concentrations of DTT and IAA. The first one contained 5 mM DTT and 15 mM IAA, while the other one consisted of 10 mM and 30 mM IAA. The proteins were dissolved in 6 M Urea, 50 mM ABC prior to reducing of proteins and carbamidomethylating of free thiol groups. Trypsin was applied in a ratio of 1:75. Each sample was divided in two, where approximately 21 µg protein was extracted for Q-TOF analysis and approximately 3.75 µg for Q-Exactive investigation. According to figure 24, alphabet I indicates the sample with regular amount of DTT and IAA, while II contains a 2-fold amount.

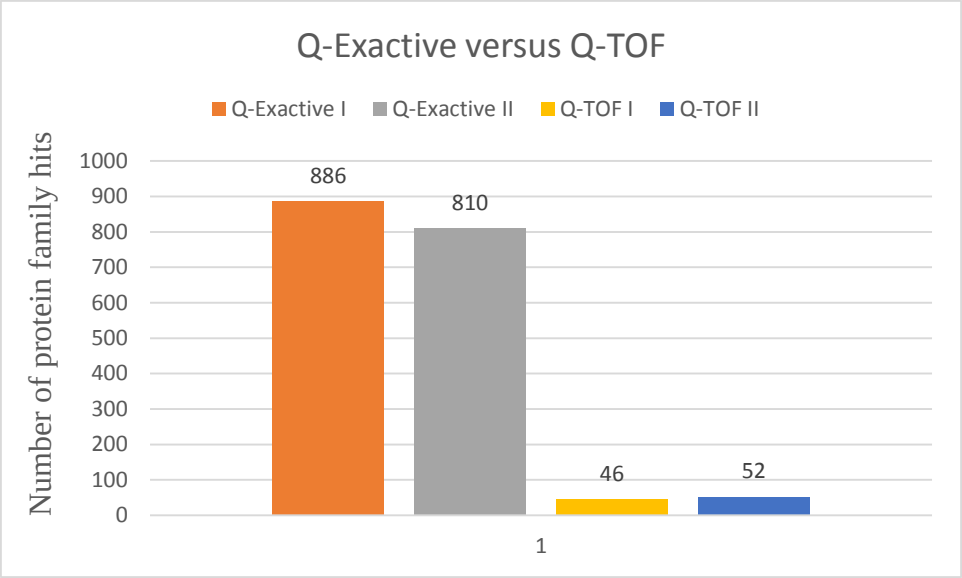


Figure 25: Comparison between Q-Exactive and Q-TOF instrumentation.

Raw data were searched against Mascot software; Q-Exactive samples were conducted with 10 ppm peptide mass tolerance and 0.02 Da fragment mass tolerance. Otherwise, the Q-TOF samples were searched with the commonly utilized parameters for an unaltered sample.

5.3 Results from COFRADIC^{TM} runs

In this section, the results from the COFRADIC process will be presented. Initially, the RP-HPLC instrument differentiated the entire proteome digest into consecutive 20 fractions. A combination or every primary fraction was thereafter handled through chemical (Sulfo-NHS-acetate / Di-methylation / TNBS) alteration of peptides. The background for this alteration was that chromatographic separation should occur for these modified peptides with different retention times. Subsequently, when such altered primary fractions are segregated by a second run under similar chromatographic conditions as within the primary run, they become distinguished from non-altered peptides and then are ready to be isolated by LC-MS/MS investigation. Gevaert et al. first created the COFRADIC method due to segregation of peptides that comprise the uncommon amino acid methionine (sulfur containing).(67)

Isolation of N-terminal peptides with Sulfo-NHS-Acetate by COFRADIC™

The initial COFRADIC run was modified at the peptide level with Sulfo-NHS-acetate after the primary run and prior to the secondary run. The protein quantity used was 100 µg, and divided into 20 fractions, hence each fraction contained about 5 µg protein. According to figure 26, the blue-labeled bar shows outcomes for the unaltered sample (56 hits). Otherwise, the secondary run subfractions were derived from 20 primary fractions. The primary fractions were collected by instrument method C 2 70 40 min, which means that 20 fractions were collected during 40 minutes and eluted with 70% acetonitrile. In addition, the utilized parameter for fraction collection was 2 mL/min. Fraction 5 was not acetylated, while F6 and F7 were acetylated with 15 mM Sulfo-NHS-acetate. All three fractions were re-loaded on the RP-HPLC column for a secondary run. The secondary fractions were collected at equal time intervals as primary fractions eluted +/- 0.5 minute. Thus, from the secondary run, the subfractions F5-1, F6-1, F6-2, F6-3, F7-1, F7-2 and F7-3 were selected to be analyzed with LC-MS/MS.

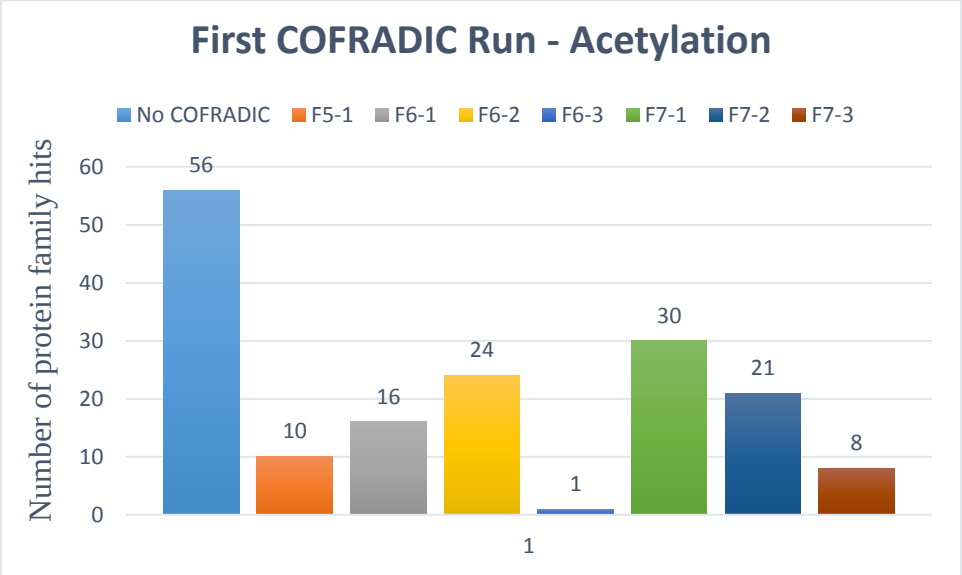


Figure 26: This figure shows MASCOT search results for the first COFRADIC run. The first sample that is labeled blue present the desalted sample and was loaded directly into the LC-MS/MS instrument. The other bars indicate samples that were not desalted with the C18 tip, but had underwent primary and secondary COFRADIC runs.

Figure 27 shows results from subsequent COFRADIC analyses, but with different elution gradient. The acetonitrile end percentage was reduced from 70% to 40% in order to obtain better separation of peptides. Otherwise, from 20 primary fractions, fraction F5-F13 were chosen to

be treated differently before a secondary run. Fractions 10 and 13 were first split in two, where the first part was acetylated as described earlier. The other part was set aside until end of the acetylation procedure. Both parts of F10 and F13 were pooled together prior to the secondary run. 100 µl from each modified/unmodified fraction was re-loaded separately into the RP-HPLC column for further collection of 20 fractions. The selection pattern was n-1, n, n+1, n+2 and n+3, where n indicates the number of primary fraction and 1, 2, 3 indicates the number of secondary fractions.

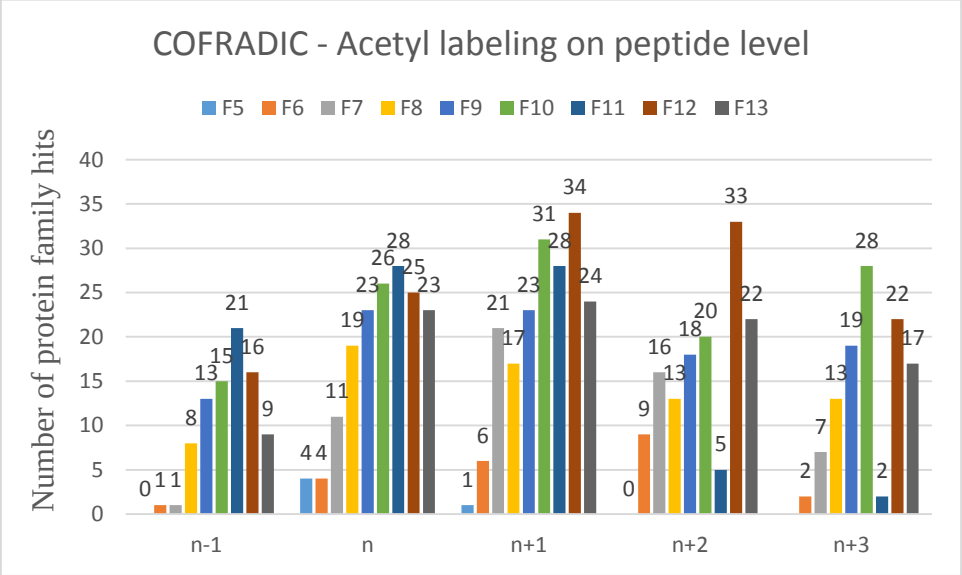


Figure 27: Nine primary run fractions were chosen to label the N-terminal peptides with NHS-Acetat. They were prepared differently before re-loading on the RP-HPLC column (see table 11).

Table 11: General procedure for differentiation of diverse COFRADIC fractions plus their outcomes.

Fraction No.	Treatment	N-1	N	N+1	N+2	N+3	Sum
F5	No acetylation	0	4	1	0	0	5
F6	Acetylation	1	4	6	9	2	22
F7	Acetylation	1	11	21	16	7	56
F8	No acetylation	8	19	17	13	13	70
F9	Acetylation	13	23	23	18	19	96
F10	Partially acetylated	15	26	31	20	28	120
F11	No acetylation	21	28	28	5	2	84
F12	Acetylation	16	25	34	33	22	130
F13	Partially acetylated	9	23	24	22	17	95
Sum		84	163	185	136	110	

The resulting pkl file from the raw data were searched against the Mascot server with the same parameters as usual, but in order to count the number of identified lysine`s and furthermore the quality of acetylation by counting acetylated lysine`s, acetyl (K) was set as a variable modification. In the same way, the quality of di-methylation procedure was evaluated.

Derivatization of peptides by TNBS

Within this experiment, TNBS modification at peptide level was used as an alternative for acetylation. The main reason for this modification step was to obtain a hydrophobic shift where the peptides elutes later and are retained for a longer time on the RP-column. This modifications procedure was conducted in accordance with Van Damme et al.(68)

The fraction pattern selection was performed in a similar manner as described above: however, here it was fewer primary fractions but more secondary subfractions, as is shown in figure 28. Prior to primary run, the methionine were oxidized with 0.5% hydrogen peroxide in favor of to convert methionine to its sulfoxide derivative without any influence of other exposed amino acids such as tryptophan or cysteine. Since methionine has a large degree of hydrophobicity, while methionine sulfoxide is hydrophilic, then this will provide a quite predictable hydrophilic shift for the majority of the influenced peptides. This event will lead to combination of numerous peptides subsequent to the primary run.

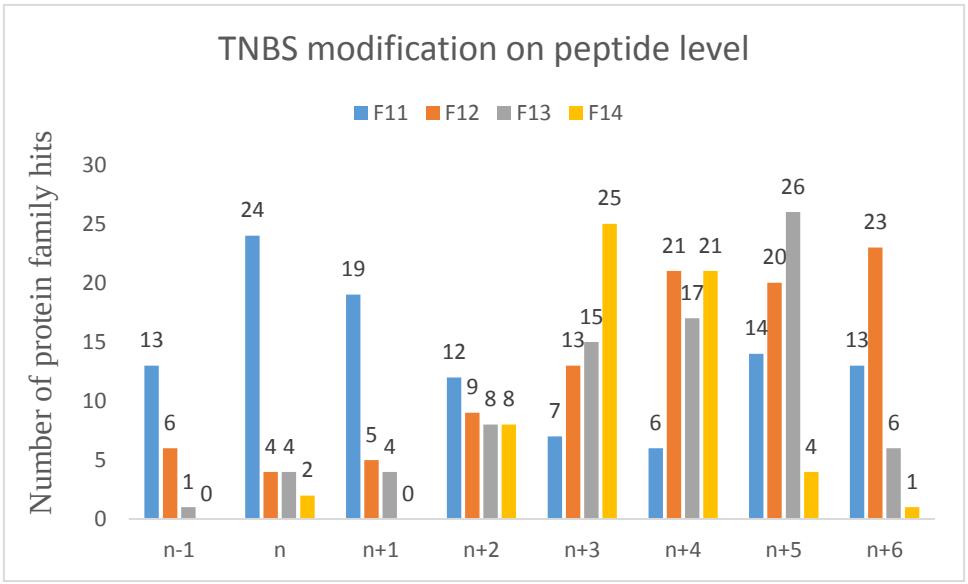


Figure 28: The first COFRADIC run with TNBS modification on the peptide level (see table 12).

Table 12: Outcomes referred to figure 27.

Fraction No.	F11 Unmodified	F12 Modified	F13 Partially modified	F14 2-folded modified	Sum
N-1	13	6	1	0	20
N	24	4	4	2	34
N+1	19	5	4	0	28
N+2	12	9	8	8	37
N+3	7	13	15	25	60
N+4	6	21	17	21	65
N+5	14	20	26	4	64
N+6	13	23	6	1	43
Sum	108	101	81	61	

The MASCOT search was performed with TNBS (K) and TNBS (N-term) set as variable modifications, otherwise the remainder parameters were as usual.

Abundance of N-terminus peptides by utilizing Di-methyl and acetyl labeling on the protein level versus TNBS on the peptide level

This experiment consisted of five COFRADIC samples, which differs by the modification stage, selection of buffer solution and FASP. First, acetylation and di-methylation were performed separately with 6 M urea buffer in the initial run followed by a subsequent run with 6 M guanidine HCl, under the same analysis conditions. This way, acetylation was compared with di-methylation and urea buffer with guanidine HCl through four COFRADIC runs. Protein quantity applied here was 500 µg in each sample. Every primary fraction was modified either with di-methyl labeling or Sulfo-NHS—acetate, thus they were re-loaded into RP-HPLC column for a secondary run. Thereafter, from each secondary run subfraction n-1, n and n+1 (for example F2-1 + F2-2 +F2-3) were chosen for further LC-MS/MS analysis. Before MS analysis, each subfraction was first evaporated, then were separately re-solubilized in 15 µL 0.1% TFA, thus they were pooled together. Hence, 15 µL of peptide mixture from every pooled sample was analyzed with LC-MS/MS instrument. Figure 29 shows outcomes for the entire 20 secondary run subfractions for each sample.

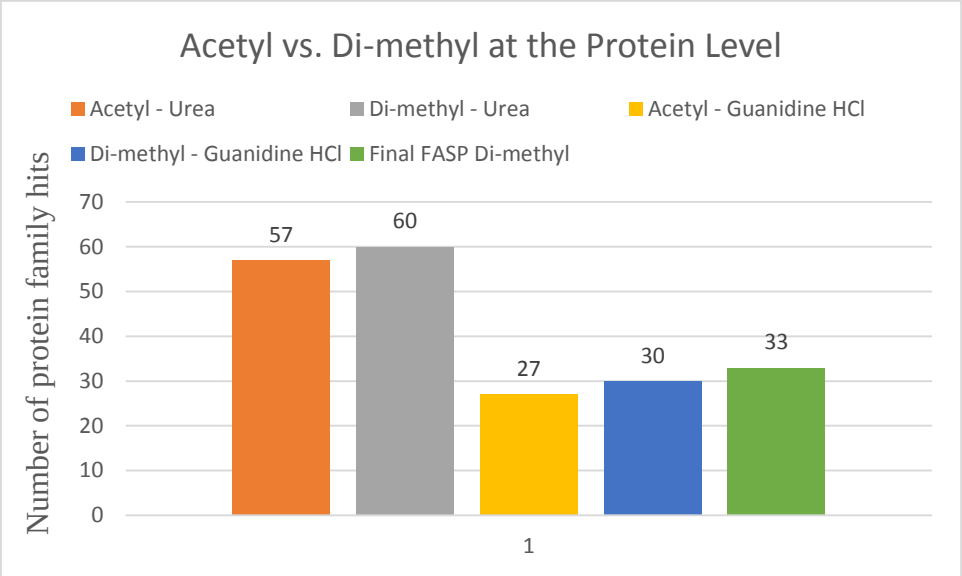


Figure 29: COFRADIC runs for comparison and differentiation of two modifications procedures and buffer solutions. The modifications steps were done on the protein level followed by TNBS derivatization on the peptide level in order to increase peptides' hydrophobicity and retention time. It was obvious that the di-methyl modifications reaction had a better performance than acetyl and at the same time urea buffer provided better protein solubility. Finally, the only FASP sample (containing di-methyl labeling) that was followed by COFRADIC runs, provided some poor identification of peptides.

Some of the previously mentioned samples were subsequently analyzed in Q-Exactive for detection of more peptide sequences, since Q-Exactive operates with high resolution and better accuracy than the Q-TOF available. Di-methylated and acetylated samples from urea were determined to undergo further analysis by selecting the odds and even fractions from the secondary subfractions, such as F1-1+F1-3+F1-5 etc. (odds) and F1-2+F1-4+F1-6 etc. (even). These odds and even fractions were pooled together followed by Q-Exactive analysis.

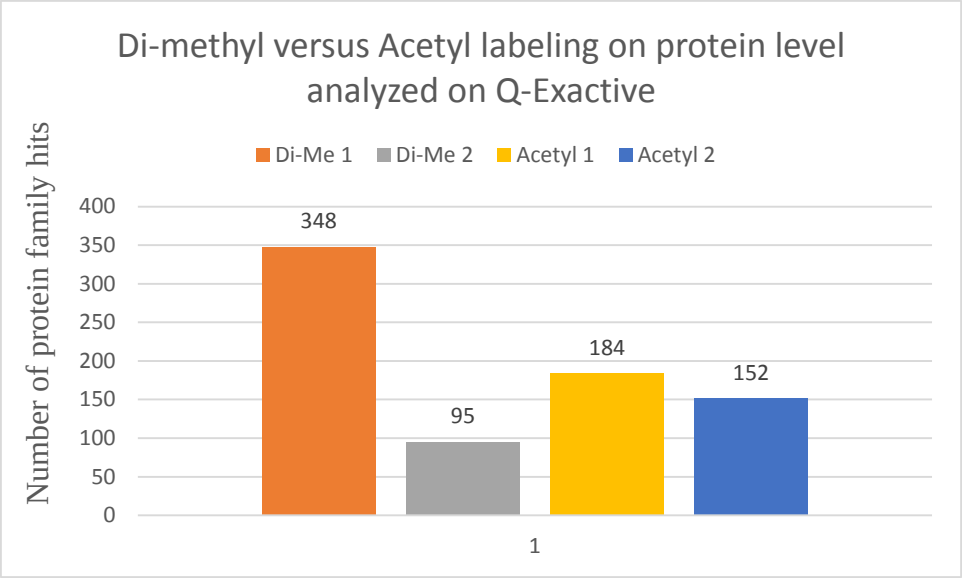


Figure 30: Enrichment of N-terminal peptides by modification with either Di-methyl labeling or acetylation on the protein level followed by TNBS derivatization on the peptide level in order to achieve a hydrophobic shift. The final peptide mixture was analyzed with the Q-Exactive instrument to determine which modifications method is most efficient for peptide identification.

Table 13: The quality and success evaluation of di-methyl and acetyl by counting number of identified lysine's.

Modification type	# Identified Lysine`s	# Labeling at Lysine`s	Percentage of successful labeling
Di-methyl (FASP)[1]	218	41	18.8
Di-methyl (FASP)[2]	32	29	90.62
Di-methyl (Venne et. al – in solution)	66	61	92.4
Acetyl	93	83	89.2

1 Regular amount di-methyl reagents.
2 20-folded increased amount di-methyl reagents.

5.4 The assessment of COFRADIC chromatograms

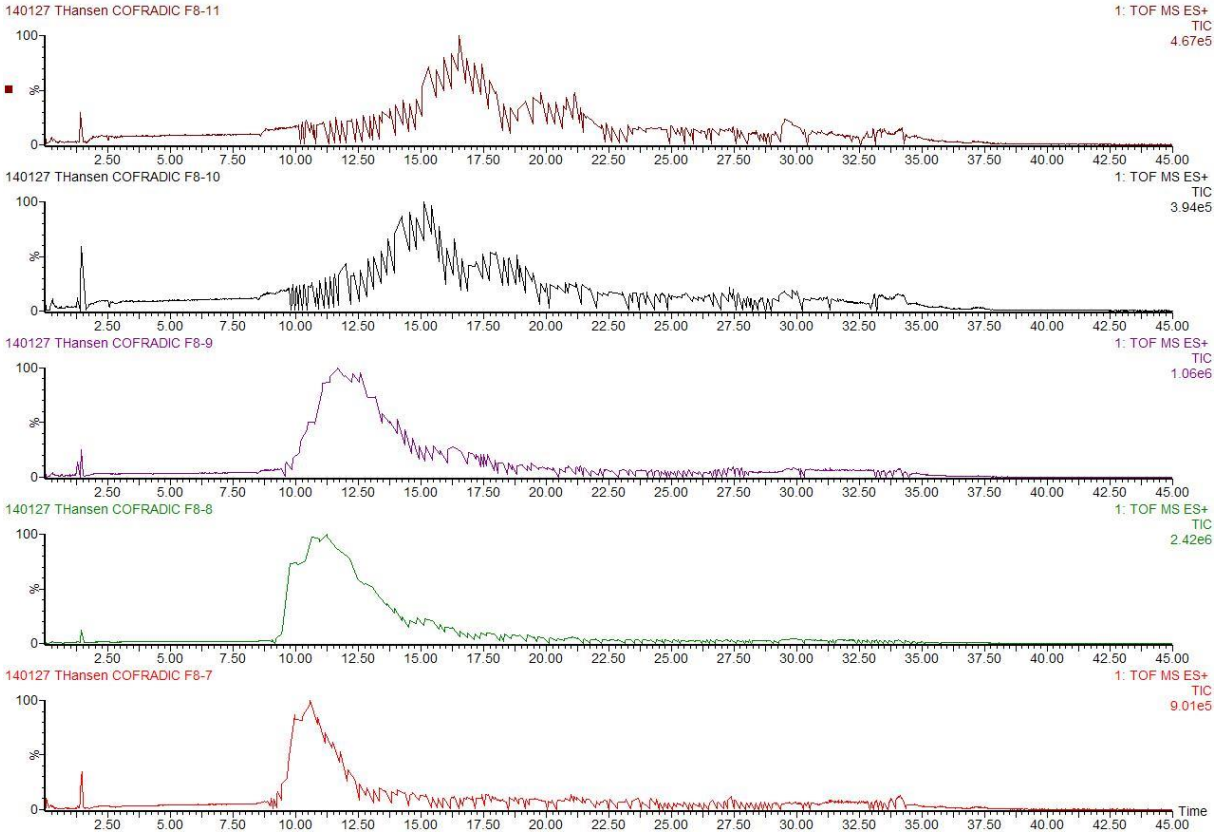


Figure 31: The chromatogram for COFRADIC run modified with Sulfo-NHS-acetate in absence of TNBS (UV absorbance at 254 nm).

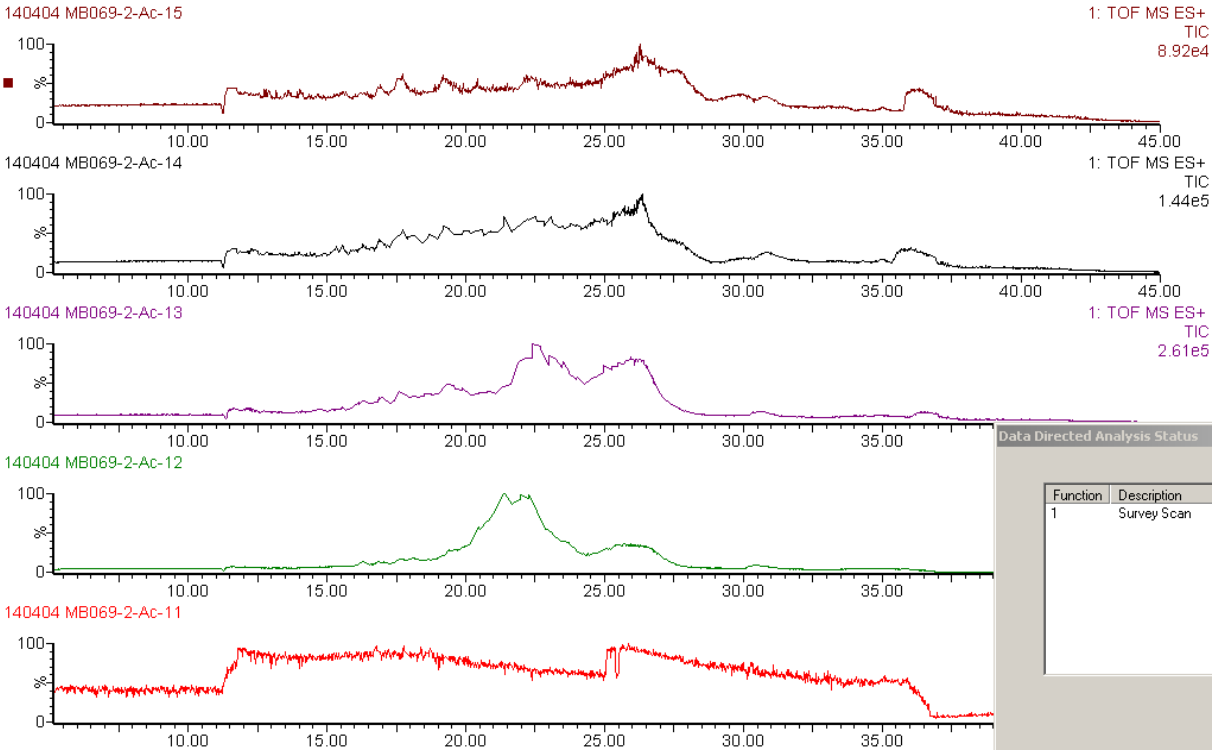


Figure 32: COFRADIC chromatogram of the acetylated fractions 11-15 from secondary run (proteins denatured in 2 M Guanidine HCl) derivatized with TNBS between 1º and 2º run. (UV absorbance at 254 nm)

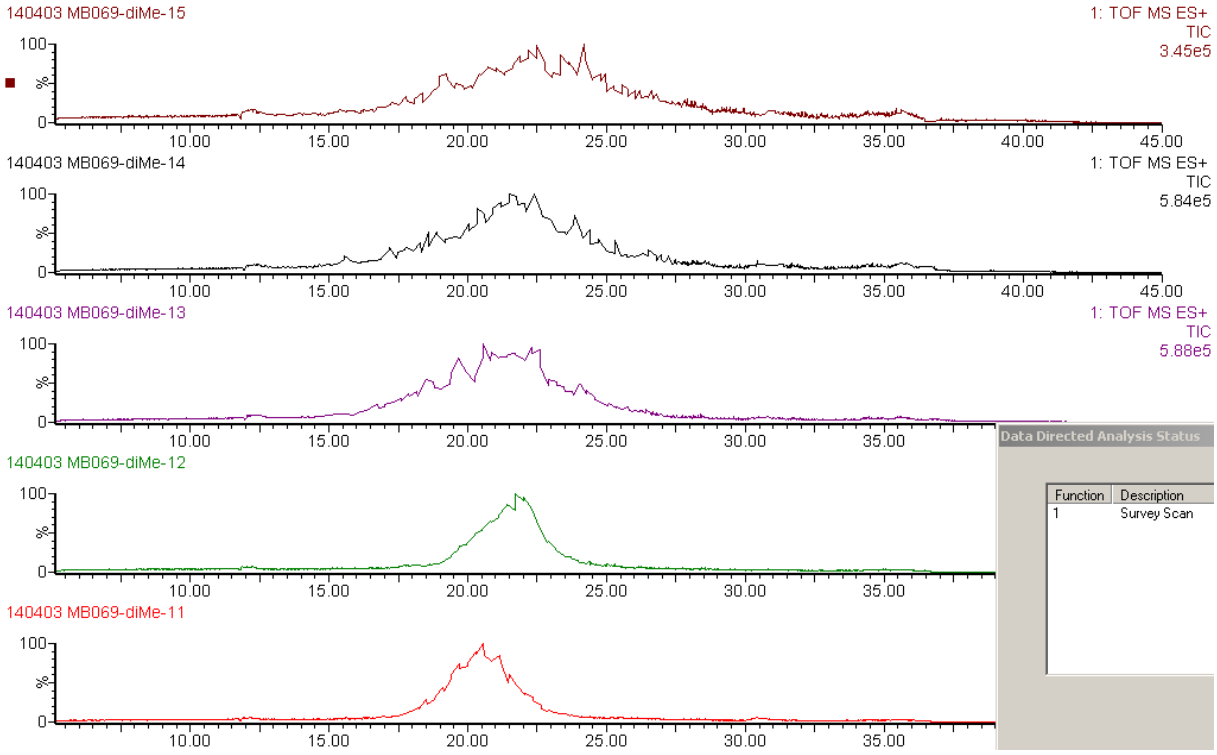


Figure 33: COFRADIC chromatogram of the di-methylated fractions 11-15 from 2º run (2 M Guanidine HCl). (UV absorbance at 254 nm)

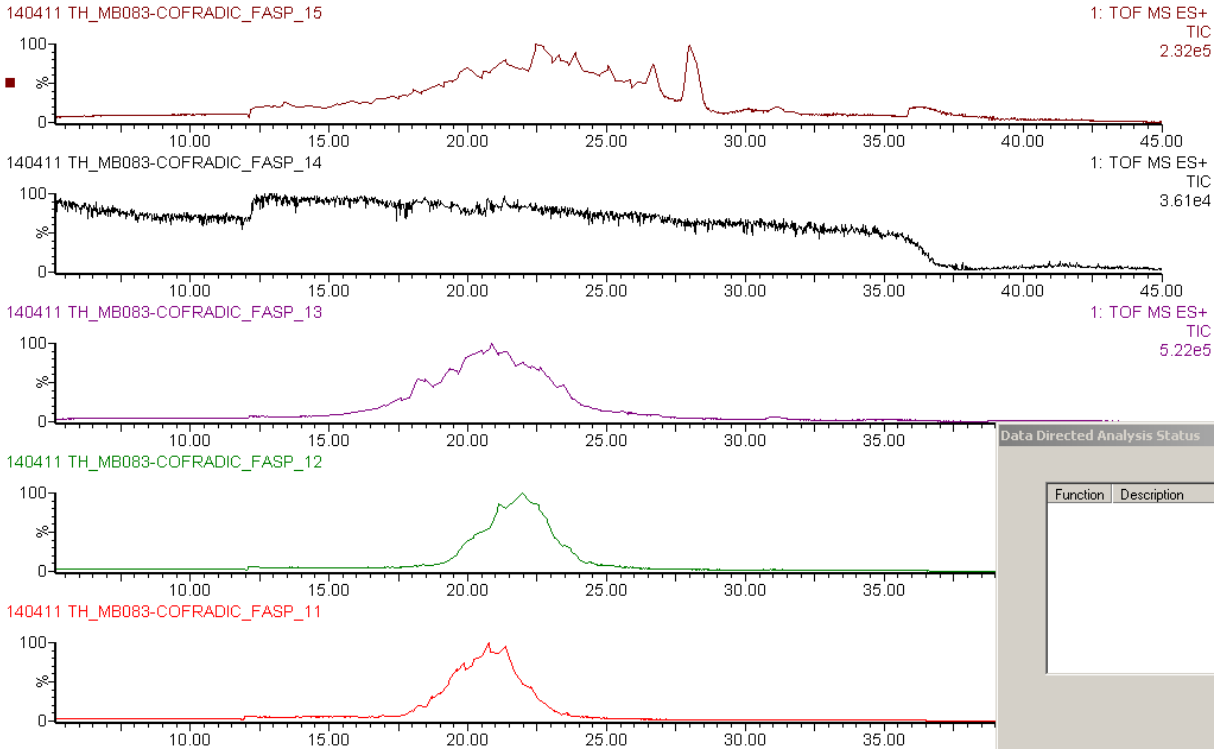


Figure 34: COFRADIC chromatogram of FASP fractions 11-15 for 2º run. (UV absorbance at 254 nm).

5.5 MS/MS Fragmentation Presentation

The MS/MS spectrum decides the identity of a peptide and the fragmentation information can be translated by different programming systems, all utilizing the sequence data intrinsic in a fragmentation spectrum.

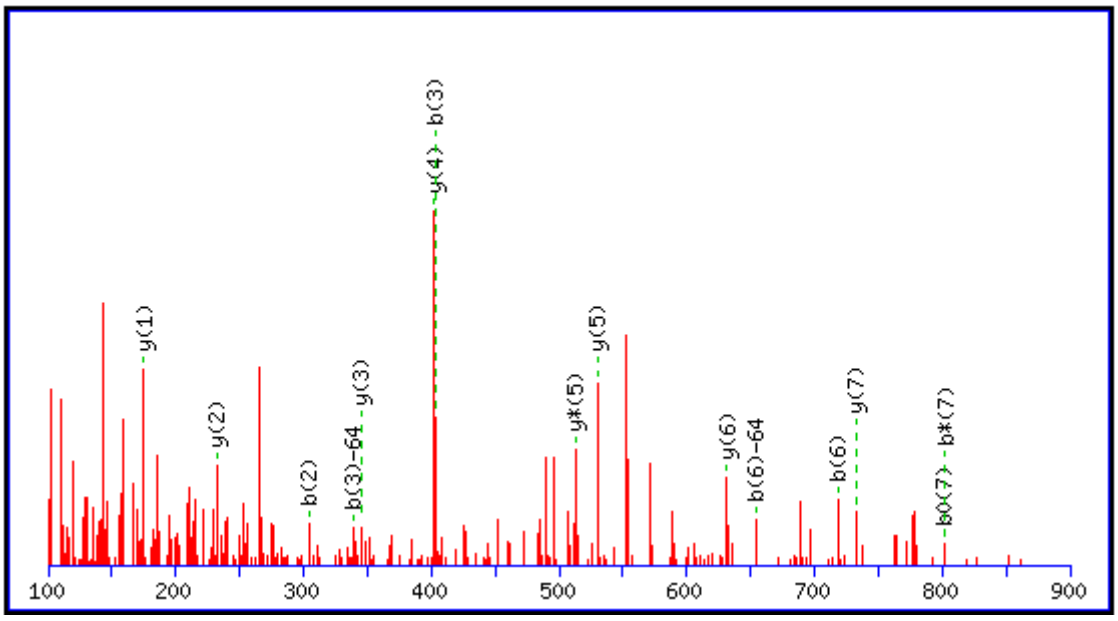


Figure 35: MS spectrum of fragmentation MQVSVETTQGLGR (N-terminal peptide) (*E.coli* strain K12 - P0A850). This MS spectrum stems from the di-methylated sample that was analyzed with Q-Exactive.

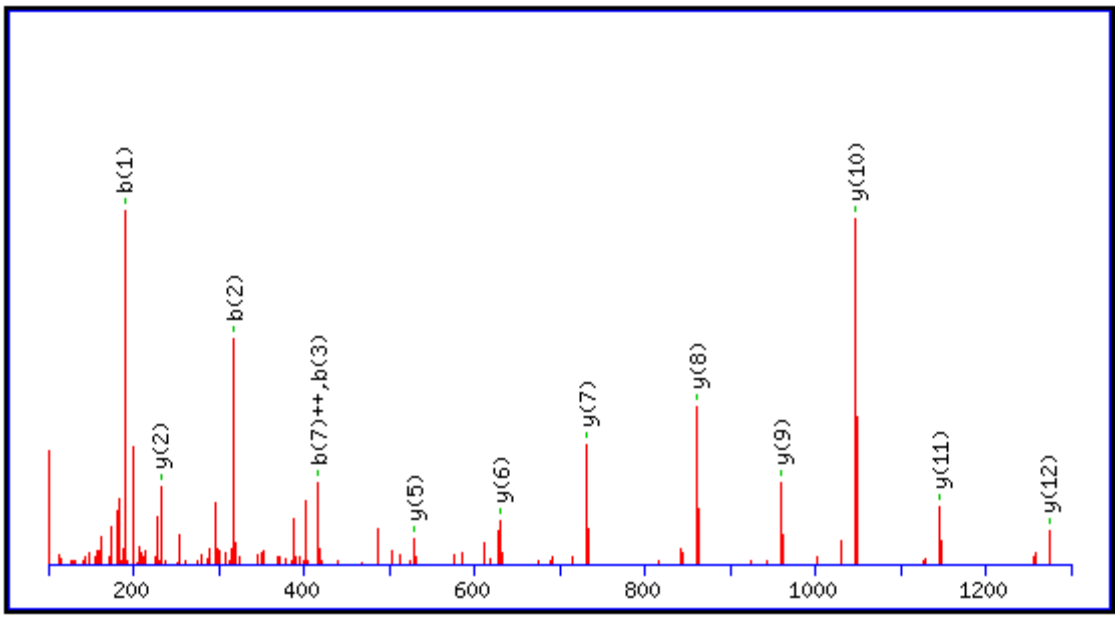


Figure 36: MS spectrum of fragmentation MQVSVETTQGLGR (N-terminal peptide) (E.coli strain K12 - P0A850 protein family). This MS spectrum stems from the acetylated sample that was analyzed with Q-Exactive.

6. DISCUSSION

The following steps characterize a mass spectrometry based proteomic study: lysis of the cells, extracting the proteins, truncating the proteins into peptides, segregating the peptides, and ultimately studying the peptides by mass spectrometry. This concept is deduced from a reductionist point of view. The output of the mass spectrometer marks a drastic change. This is because mass spectrometry is purely numerical. The data should be employed primarily for purposes of reconstructing information about the test samples, the domain of (bio-) informatics.(7)

To decrease sample intricacy, single samples of proteins should be selected to represent each peptide. This can be achieved by choosing either its N- or C-terminal peptide. This was the basis for encouragement to develop the N-terminal COFRADIC (57). The real alteration reaction utilizes 2, 4, 6-Trinitrobenzene sulfonic acid (TNBS) that makes the non-N-terminal peptides more hydrophobic so that N-terminal peptides are easily separated.(51)

The flexible Combined FRActional DIagonal Chromatography (COFRADIC) methodology offers an approach to decreasing sample intricacy. This finding is capable of selecting cysteinyl and methionyl peptides according to Gevaert et al. research (2002 and 2007). This method reduces sample complexity of the peptide mixture through two identical and consecutive RP-HPLC separations. A modification reaction is then carried out in between these two separations to transform the physicochemical properties of an aimed group of peptides. The second RP-HPLC separation is executed in order to distinguish the non-altered peptides from the altered peptides. The modified peptides develop a dissimilar elution pattern during this process. Different sets of peptides can be targeted and separated by altering the actual modification reactions.

The overall outcomes were of average quality, since the initial number of identified peptides in LC-MS/MS were low, but with ability for enhancement. Based on inadequate peptide identification within FASP runs, thus several methods were implied in order to improve the results. Modification of the FASP protocol was not easy to perform, therefore it was repeated several times. The main problem that occurred during the FASP run was filter leakage after the trypsinization procedure. This problem was solved later on by utilizing an original pall spin

tube in order to collect peptides during the proteome digestion step. It is important to emphasize that the introducing FASP analysis were not carried out by the COFRADIC method.

6.1 Results from FASP analysis

Three key factors are considered essential for proteomic coverage. These key factors include the methods used for protein extraction, solubilization and digestion. However, a complex protein mixture contains a large amount of detergents that need to be removed prior to MS analysis in order to yield successful peptide identification. Additionally, factors such as chaotropes content in a sample and mechanical disruption can affect the analysis results. Detergents are able to contaminate the chromatography and generates noise in the mass spectrometer, i.e. noise-to-signal ratio increase in presence of detergents, thus a minimum quantity of peptides may be identified. A novel research done by Erde et al.(69) introduce a FASP (Filter Aided Sample Preparation) approach for proteomic sample preparation. FASP certainly prevent the majority of the challenges associated with conventional protein purification methods. Nevertheless, it leads to considerable loss of the sample. The method introduced by Erde et al. involved an enhanced FASP with improved susceptibility, recovery and proteomic coverage for a specific sample. The FASP procedure was utilized for removing of contaminants or detergents, included repeated washing processes that made the sample ready for protease digestion with trypsin.

The LC-MS/MS outcomes of FASP samples were compared to in-solution digestion. The primary obstacles here were that the protein solution on the filter was too low and also as stated earlier, filter leakage occurred several times during the trypsin digestion step. This protocol was implemented in order to compare the protein solubility on the filter against that in solution.

In general, the maximum number of protein family hits during this project was between 40-50, which indicates that this quantity of identified peptides was difficult to enhance or increase, unless a new separation method was implemented.

The initial underivatized FASP samples and the subsequent di-methylated samples provided different LC-MS/MS outcomes. A hypothetical reason could be that eventually error occurred during the modification step. Since trypsin did not cleave at di-methylated lysine’s, this was a substantial consideration with the di-methylation of intact proteins. Miss cleavage results in

longer peptides, which could be considered as both advantage and disadvantage. Sometimes, regions of proteins that became digested into very short peptides and that cannot provide confident identifications will become accessible. Nevertheless, in case of obtaining a large quantity of long peptides will not be suitable for the standard LC-MS/MS proteomics method.

The FASP protocol was utilized as a refinement device for proteins in order to provide a sample, which does not contain detergents. This method was suitable as it allows for purification of samples that contain high detergents concentrations.

Preceding research done by Jentoft et al in 1979, have shown that reductive methylation with sodium cyanoborohydride and labeled formaldehyde gives a straightforward technique for explicitly marking the amino acids of proteins utilizing intensely softened reaction conditions. In this study, they utilized radioactive formaldehyde, in order to run separately prepared samples in parallel where one of them has incorporated dimethyl reagents in ^{14}C-formaldehyde while the other one did not. Consequently, both samples were compared with each other to detect the differences between them (66).

Reductive Amination of Primary Amines

Reductive amination is also called alkylation, where this modifications method of proteins and peptides has gained success as long as it has been performed (70). In accordance with Hsu et al the methylation of peptides with formaldehyde act to be unchallenging mechanism that during short time (5 minutes) provides fully transformation of free amines to their mono- and di-methylated forms (61). This di-methylation strategy was utilized during this thesis, but the amount of methylations reagents had to be increased 20-fold in order to obtain successful quantification of proteins. Through reductive amination, lysine's N-terminal side and ε-amino group were universally labeled with formaldehyde reagent (37% solution in H_2O). In our study we reached a level of successful di-methyl labeling of lysine's by over 90%, which was as expected.

A challenge occurred during di-methylation process was that ammonium bicarbonate buffer could not be applied during this process through its interference with sodium cyanoborohydride. It was not easy to confirm if the obtained results come from di-methyl labeling or ABC.

Ammonium bicarbonate could be used after the di-methyl modification step as a digesting buffer through trypzination reaction in order to enhance the activity of trypsin within pH above 7.0. This limitation is because during the reaction mechanism of reductive amination, when the secondary amine is combined with formaldehyde it will end up making an ammonium cation that it is not necessary to be protonated. The result will be an ammonium cation that will have a methyl group with a carbon-nitrogen double bond from formaldehyde with a positive charge. Therefore, when sodium cyanoborohydride is combined with ammonium bicarbonate, this ammonium cation will be confused with ABC, thus uncertain MS results may appear.

Di-methylation is widely utilized due to its cost effectiveness, safety and simple implementation. A MASCOT search was conducted in accordance with the parameters specified in table 11, and the difference in the number of missed cleavages was due to the fact that trypsin will not cleave after methylated lysine's. The quality of methylation procedure was evaluated in the same manner as for acetylation. After several run, the methylation modification method provided some interesting results by increased peptide identification. The reproducibility of this method could not be determined due to preliminary data. This obstacle refers to a small set of outcomes that have not been duplicated or sufficiently validated.

Di-methylation combined with mass spectrometry has been considered as a robust process for determination of the N-terminus residue of proteins and peptides due to rare observation of a1 ions in the MS/MS spectra of unchanged peptides and proteins. This is because a1 fragment ions of di-methylated peptides and proteins have improved the MS/MS spectrum's ion intensities. This modifications method is widely employed in quantitative proteomics (61). Fragmentation of di-methylated proteins was a proper method created in order to investigate proteolytic procedures. Nevertheless, incomplete protein modification has been announced due to occurrence of side products when utilizing sodium cyanoborohydride as catalyzing reagent in reductive methylation (61).

According to figure 21, three various FASP samples were di-methylated. The number of protein family hits for the first sample was due to failed modifications procedure and in addition filter leakage issue during proteolytic digestion. However, the second sample contained the correct amount of di-methylation reagents in accordance with Hsu et. al. In opposite to the first sample, filter leakage did not occur during this procedure. Affection of filter leakage on the results yield

was a considerable factor. Although, the third sample showed fewer results, even if the di-methylated reagent amount was 20-fold increased. Most likely, this was because the double concentrated sample had more charge states than the normal one, thereby poor fragmentation of these ions occurred. The MASCOT searche were conducted with +2 and +3 peptide charges during the initial analysis, but later on the searches were performed with +2,+3 and +4 peptide charges.

Proteolytic digestion denotes the breakdown of proteins into peptides by utilizing trypsin enzyme. This process can be incomplete when proteins goes directly through the filter without being digested. This way, only a minimum amount of proteins transforms into peptides, thus the MS analysis will provide insufficient results. This was how filter leakage affected on the MS outcomes.

The unaltered samples yield more protein family hits than the modified ones. In theory, when a sample is di-methylated it indicates that primary amines became labeled with two methyl groups, thus during analysis in LC-MS/MS, this peptides will differs from the unaltered ones. Therefore, only the altered fraction of peptide mixture is detectable in MS. In case of failed or incomplete modifications procedure, there will not be 100% identification of the altered peptides.

Determination of protein solubility

The protein solubility was the main challenge in this study, since during the FASP procedure, it showed poor solubility on the pall spin-filter. For that reason, a number of buffer solutions were tested in order to determine which buffer is applicable and provides greatest solubility. This process was conducted twice due to filter leakage during the first run.

It was obvious that ammonium bicarbonate buffer solution provided highest amount of identified peptides. In contrast, sodium dodecyl sulphate (SDS) buffer showed significant ineffectiveness. SDS is an organic compound that consist of a tail with 12-carbon atoms attached to a sulfate group, which provides amphiphilic properties to the compound. This kind of properties is required of a detergent (71). Due to the extremely negative charge of SDS, it was desalted with SCX OMIX 10 µl tip prior to MALDI analysis. One hundred microliter eluat

was desalted with a 10 µl SCX tip that indicate incomplete desalting procedure and loss of peptides. After desalting, the SDS samples were run in the MALDI instrument in favor of to control eliminated SDS residues before the MS/MS analysis. The MALDI spectrum showed that there was a minimal content of SDS residues in both samples. However, the LC-MS/MS showed that SDS samples were almost free for peptides to be identified. This was probably due to the application of 10 µl tip instead of 100 µl during the desalting process.

The Guanidine HCl buffer solution showed almost same degree of solubility as the ABC buffer. Guanidine HCl is a strong denaturants and is universally used in the field of proteomics for denaturation of proteins. The positive effect of guanidine HCl on protein denaturation was remarkable during this process.

6.2 In Solution Sample Preparation

The protein precipitation occurred using ice-cold ethanol, which indicated an improvement in the protein solubility compared to the FASP procedure. As stated earlier, the sample purification or reduction of sample complexity was the general challenge that resulted in insufficient peptide identification in LC-MS/MS analysis. For this reason, protein precipitation was carried out in a conventional way. This method was time consuming since the protein precipitation occurred on ice for more than 4 hours followed by modification steps.

As shown in figure 24, the number of identified protein families did not vary much from FASP outcomes. Di-methylated procedure was conducted on protein level and in contrast to the previous modified samples this one was carried out with a buffer comparison of HEPES in addition to sodium cyanoborohydride and formaldehyde and it was incubated for 2 hours instead for 5-10 minutes on bench reaction. However, the outcomes are almost similar even if the samples were prepared differently. HEPES is a globally used buffer reagent in cell culture and its characteristics properties includes high solubility, enzymatically and chemically stable and restricted effect on biochemical reactions.(72) In favor of optimal di-methyl labeling, this buffer was pH adjusted to 8.0 just before application. The MASCOT search defined that di-methyl labeling performed with FASP and in solution digestion yield various success in sufficient lysine's labeling. Table 14 presents that di-methyl labeling in solution in accordance with Venne et al. was more successful than the initial one with FASP.

In this project, our target was the N-terminal peptides, where the central point was directed to primary amines as well as lysine's and arginine's residues. We cannot expect 100% peptide identification because the fragmentation of very long peptides will not allow well detection of these peptides. This problem may refers to the trypsin usage and peptides with a high content of lysine's. This event can explain why the altered samples yield less than the unaltered.

Another aspect that is worth mentioning, is the reduction and alkylation of proteins in the initial stage of the procedure. The proteins were reduced with the strong reducing agent DTT under several concentrations in order to determine the most correct and efficient DTT amount to reduce disulfide bonds of proteins in absence of any side reactions. The results showed that there was no need to 2-fold increase the amount of DTT in order to achieve better reduction of proteins. Additionally, IAA was used to bind covalently to the thiol groups of cysteine in order to prevent reformation of disulfide bonds. Various IAA concentrations were applied in order to enhance the abundance of free thiol groups. However, overalkylation of a protein seemed to be negative in the cause of reduction in the number of identified proteins compared with the outcomes from a normal sample. This was a substantial invent to determine that a concentration of 5 mM DTT and 10 mM IAA were enough to yield satisfactory reduction and alkylation of proteins preventing possible side reactions.

Q-TOF versus Q-Exactive instrumentation

High resolution and sensitivity seems to be crucial in peptide identification on Q-Exactive instrumentation, which provides high scan-speeds. The abundance of identified peptides seemed to be significant by using the orbitrap mass analyzer compared to the outcome provided by Q-TOF. This analysis was performed to prove the content of peptides in the sample. The outcomes showed how much the high resolution and accuracy can influence on proteomics. This is a very expensive instrument, which makes it not frequently used.

Both Q-TOF and Q-Exactive are valuable systems. Q-Exactive is considered as an excellent system for analysis of complex samples and in general proteomics applications. A drawback of Q-Exactive is that it requires frequently cleaning of the system, almost once a month, but it also depends on the type of the sample. This cleaning procedure involves both the S-lens and the

quadrupol. Since our work was based on proteomics, which consist of very complex samples, so this instrument could be valuable to provide satisfactory results of peptide identifications. Briefly, Q-Exactive operates with fast scan rates that leads to better sampling and then the result is a great content of proteins/peptides identifications.(73)

6.3 Outcomes from COFRADIC analysis

Isolation of N-terminal peptides with Sulfo-NHS-Acetate by COFRADIC™

Acetylation of proteins and peptides in the framework of MS/MS peptide sequencing is carried out for several reasons. The acetylation of N-terminal peptides reduces the total charge of a peptide thereby the solubility may be reduced. However, the stability and the biological activity of the peptides could be increased during this modifications process.

So far, the discussion was focused on di-methylation modifications process on protein level, but acetylation process was also important since the initial acetylation's procedure was conducted at the peptide level.

The initial acetylation procedure showed that subset F7-1 contained the highest number of protein identifications after the secondary run, while subset F6-3 was almost free for peptides. This difference can refers to the secondary run of subset F6-3 including variances in elution profile for peptides. According to figure 26, the first 4 minutes of peptide collection (F6-1) gained an adequate amounts of ID's, after 8 minutes collection the amount increased, and after 12 minutes an extremely decreasing in ID's appeared. This implies that peptide separation during the second run was most efficient after 8 minutes collection. The peptide separation depends on different factors such as the mobile phase composition and the general elution gradient varying between water (strong mobile phase) and acetonitrile (weak mobile phase). At the beginning procedure the COFRADIC runs were carried out with a elution gradient that extend to 70% acetonitrile after 30 minutes of peptide collection, which seemed to be somehow high, therefore this was reduced to 40% in order to obtain a flatter gradient that provides better peptide separation. The same isocratic elution gradient was applied within primary and secondary COFRADIC runs.

As given in figure 27, several subset from secondary run with various chemical treatment were selected for LC-MS/MS analysis. Fraction F12 showed the highest number of identifications

(120 ID's), and subset n+1 had a large abundance of ID's (185). F12 was acetylated and n+1 was collected 2 minutes after each altered or unaltered secondary fractions. This event designated that the elution time in the column was short, and required to be prolonged in order to achieve a better peptide isolation. For that reason TNBS was applied to make the peptides more hydrophobic and retain for a longer time in the RP column.

Before trypsin digestion, the protein natural quality was altered in a high concentration of chaotropes such as 6 M urea. Protein reduction of disulfide bridges occurred so that the free thiol groups became alkylated with the use of iodoacetamide. The next process involved acetylation of all free primary amines like α- and ε-amines to be blocked; hence if trypsin was added it could only cleave after arginines. This was because trypsin cannot recognize acetylated lysine's. Trypsin performed in the form of endoproteinase Arg-C, thereby produced peptides ending on arginine. Protein N-terminal peptides are quite different from any other peptides for the fact that they carry an acetylated α-amino group whether *in vivo* or acquired *in vitro.* The chemical difference was taken advantage of in the initial setup when TNBS targeted the internal peptides carrying a free α-amino group. TNBS, the actual sorting reagent, reacts only with primary amines. This means that it does not react with peptides such as proline residue.

TNBS derivatization on the peptide level

A hydrophobic shift occurred when 2,4,6-trinitrobenzenesulfonic acid (TNBS) was applied on the peptide level. This was because TNBS (K) could not be recognized in any peptide sequence within the MASCOT search. Hence, acetylated lysine's had retained longer on the reversed phase column and the selected fractions had an inadequate amount of TNBS modified peptides.

Within the primary COFRADIC modified with sulfo-NHS-acetate, the peptides were separated into fractions, and each composed peptide fraction comprised of two sorts of peptides: internal peptides carrying a free alpha-amino terminus and N-terminal peptides with a blocked amino terminal. Both alpha and epsilon amines became acetylated with NHS-acetate.

Prior to TNBS modifications step, the methionine's were oxidized with 0.5% hydrogen peroxide in order to sort methionyl peptides. The sorting of methionyl was carried out under reaction conditions by 30 minutes at 30 ^{o}C. In case of prolonged reaction or increased temperature, then undesired oxidation reactions on methionine and other amino acids may occur. Additionally, methionine sulfoxide transforms to methionine sulfone, which involve that

peptides bearing methionine sulfone has a larger degree of hydrophilicity that methionine sulfoxide. This results in different retention time in the RP-HPLC column, where methionine sulfone has much lower retention and thereby insufficient sorting of methionyl peptides takes place. (74)

When TNBS was added, a reaction arose between TNBS and the internal peptides, where a hydrophobic trinitrophenyl group was introduced to their N-terminus side causing them to be more hydrophobic. Within a set of secondary COFRADIC, the internal peptides were separated from the primary collection intervals and the N-terminal peptides that were not altered became eluted with a elution gradient at the same time, thereby were gathered for LC-MS/MS analysis.(75)

Primarily, the TNBS modification process was occurred in the absence of acetylation in order to estimate the degree of hydrophobic shift TNBS provides. As stated in figure 28 and table 12, still the modified fraction F12 has the largest abundance of identifications, while subset n+4 and n+5 had adequate number of identifications. This observation was as anticipated, since the altered peptides created stronger hydrophobic bonds with the C18 beads. However, the partially modified fraction F13 contained an average number of protein family identifications, while the 2-fold increased volume and concentration of TNBS seemed to be ineffectual.

Comparison between di-methylation and acetylation

During our gel-free proteomic study, we compared di-methylation and acetylation on the protein level. The aim of this study was to prove which of these posttranslational modifications processes is most eligible in the field of proteomics. The yield of protein family identifications were assessed between these two methods. The advantage of COFRADIC is that a selected category of peptides can be sorted.

The final process composed of acetylation and di-methylation performed on the protein level, while TNBS modification was on the peptide level. During this process, the sorting of N-terminal peptides by acetylation became more effective compared with the previous procedure conducted without the application of TNBS. Figure 29 demonstrate the diverse COFRADIC runs, where di-methylation with urea buffer solution showed desirable results. Additionally, there was no discernible difference between di-methylation and acetylation combined with urea.

Separation of N-terminal peptides was not a simple task to perform because each protein contains only one amino terminus side represented by only one peptide, thus one by one peptide was separated in the UV-HPLC instrument, where the amino acid sequence was detected by the time the peak was captured (chromatogram).

Theoretically, the N-terminal COFRADIC method ought to exclusively separate the N-terminal peptides of proteins. Nevertheless, practically, numerous other kinds of peptides are inevitably co-sorted, for example peptides bearing a blocked, non-acetylated N-terminal amino acid such as pyrrolidone carboxylic acid, proline or cyclic S-carbamoylmethylcysteine. Due to the chemical properties of their N-terminus side , these peptides are then easily co-sorted. This event seems to be beneficial in the field of proteomics for different purposes since certain peptides in every protein are monitored, possibly arising the precision of deciding the ratio of enriched proteins. Still, the number of isolated peptides bearing free alpha-amines can be very low, and most likely, this can refers to an extreme abundance of structural proteins and enzymes. If the prevalence of these structural proteins and enzymes became decreased, thus it will result in enhanced detection of real N-terminal peptides and thereby a wider and more susceptible proteome coverage. Still, the expectation for proteome coverage can never reach 100% quantitation. This is because especially abundant and simply ionizable peptides may emerge in their free alpha-amine form and subsequently be separated.(74)

Both di-methylation and acetylation showed remarkable abundance of N-terminal peptide separation from a complex peptide mixture. COFRADIC methodology is vast and it decrease the complexity of protein and peptide mixtures or a general complex mixture. In the research of biomarkers, COFRADIC has a leading role for characterizing and identification of specific aimed peptides. Di-methylation proved to be a successful modification method (443 ID's), while acetylation of N-terminal peptides (333 ID's) was demanding, but it is very powerful PTMs within the field of proteomics and for targeted drug discovery, where N-terminal peptides are unique.

In this project, we identified more than 300/400 Escherichia coli proteins by Q-Exactive and more than 50 *E.coli* proteins by tandem mass spectrometric analysis (n-LC-MS/MS) of segregated methionine-containing peptides. A diagonal method is described to isolate N-

terminal peptides in order to decrease the complexity of the peptide mixture, because as stated earlier each protein has one N-terminus that is represented by just one peptide.

The UV-HPLC chromatograms 31-33 shows how TNBS influenced the retention time for a subset of secondary run fractions. A hydrophobic shift occurred from 10-15 minutes in the absence of TNBS and over 25 minutes in the presence of TNBS.

7. CONCLUSION

Our study was aimed to develop a COFRADIC HPLC system for proteomics, where the focus was directed to the separation of N-terminal peptides by utilizing two widely used modification processes: di-methylation and acetylation. These two methods were compared with one another in order to detect the robustness, complexity, sensitivity and reliability when combined with COFRADIC methodology. COFRADIC is a vital methodology that allows routine separation of methionyl, cysteinyl, N-terminal peptides and phosphorylated peptides respectively.
The acetylation of N-terminal side of amino acids was carried out with sulfo-NHS-acetate, while di-methylation procedure was conducted with sodium cyanoborohydride combined with formaldehyde. However, the di-methylation modification method proved to be somehow more effective and sensitive than the acetylation method. These outcomes were due to analysis with Q-Exactive that was more accurate and operated with higher resolution than the conventional Q-TOF instrumentation. This observation was as expected, but there is always ability for enhancement within the area of proteomics and particularly COFRADIC. Implementation of a COFRADIC HPLC system for proteomics showed to be challenging and time-consuming due to utilization of various protocols in order to reach adequate and successful results that are robust. Di-methylation and acetylation alteration on the protein level and TNBS derivatization on the peptide level provided satisfactory abundance of N-terminal peptides.

The enrichment of N-terminal peptides by di-methylation and acetylation have an obvious opportunity for enhancement. The classification of PTMs and a comprehension of their functional importance is crucial to all areas of proteomics.

8. FUTURE PERSPECTIVES

In current research, N-terminus COFRADIC investigation is one of the most interesting research methodologies within the field of proteomics. The field of proteomics is in frequently developing specially including new COFRADIC methodology for improvement of this precious research area. Many of nowadays disease stems from either defection or mutation related to proteins and its dynamics activities. In order to improve the abundance of N-terminal peptides by the COFRADIC methodology, it will be useful to analyze the purified samples in the Q-Exactive instrumentation or to obtain an adequate identifications amount of those peptides, the sample has to be free for salts and detergents (requirement for extreme purification). This way, when a highly purified sample with optimal modification is subjected to the Q-TOF analysis, then most likely the results will be valid and useful.

9. REFERENCE LIST

1. Al WV et. Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium [Internet]. p. 16(7):1090–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7498152

2. Proteomics Workflows & Technologies. Available from: http://proteomics.missouri.edu/Tutorials/Proteomics workflows & technologies2.pdf

3. Frauenfelder H, Mcmahon B. Commentary Dynamics and function of proteins : The search for general concepts. 1998;95(April):4795–7. Available from: http://www.pnas.org/content/95/9/4795.full.pdf

4. Williams SK, Tyler JK. Transcriptional regulation by chromatin disassembly and reassembly. Department of Biochemistry and Molecular Genetics, University of Colorado at Denver and Health Sciences Center, Aurora, CO 80045, USA. jessica.tyler@uchsc.edu; 2007 Apr;17(2):88–93. Available from: http://europepmc.org/abstract/MED/17307351

5. Kelen K Van Der, Beyaert R, Inzé D, Veylder L De. Translational control of eukaryotic gene expression. Crit Rev Biochem Mol Biol [Internet]. Informa Healthcare; 2009 Jul 16;44(4):143–68. Available from: http://informahealthcare.com/doi/abs/10.1080/10409230902882090

6. Snider J. Overview of Post-Translational Modifications (PTMs) [Internet]. Available from: http://www.piercenet.com/method/overview-post-translational-modification

7. Helsens K, Martens L, Vandekerckhove J, Gevaert K. Mass spectrometry-driven proteomics: an introduction. Methods Mol Biol. 2011;753:1–27.

8. Zhang Y, Reinberg D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev [Internet]. 2001 Sep 15 [cited 2014 Apr 29];15(18):2343–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11562345

9. Smith KT, Workman JL. Introducing the acetylome. Nat Biotech [Internet]. Nature Publishing Group; 2009 Oct;27(10):917–9. Available from: http://dx.doi.org/10.1038/nbt1009-917

10. Na O. Sulfo-NHS Acetate. 0747(26777):3–4. Available from: http://www.piercenet.com/instructions/2161503.pdf

11. Hayworth D. carbonyl-reactive-crosslinker-chemistry @ www.piercenet.com [Internet]. Available from: http://www.piercenet.com/method/carbonyl-reactive-crosslinker-chemistry

12. Reductive B. Reductive Amination and Acylation. :1–5. Available from: http://www.udel.edu/chem/dawatson/classes/Chem334_S14/labs/ReductiveAmination2014.pdf

13. Sherwin J. Abraham SH and VG. Detection of protein–ligand interactions by NMR using reductive methylation of lysine residues [Internet]. 2008. Available from: http://link.springer.com/article/10.1007/s10858-008-9274-y/fulltext.html

14. Research. Amino Acids [Internet]. 2010. Available from: http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/A/AminoAcids.html

15. Edman P. Method for Determination of Amino Acid Sequence in peptides [Internet]. 1950. p. (1950) 283—293. Available from: http://actachemscand.org/pdf/acta_vol_04_p0283-0293.pdf

16. Niall HD. [36] Automated edman degradation: The protein sequenator. In: C. H. W. Hirs SNTBT-M in E, editor. Part D: Enzyme Structure [Internet]. Academic Press; 1973. p. 942–1010. Available from: http://www.sciencedirect.com/science/article/pii/S0076687973270398

17. Biotechnologies G. Edman-Degradation [Internet]. Available from: http://www.genosphere-biotech.com/Edman-Degradation.html

18. De Godoy LMF, Olsen J V, de Souza G a, Li G, Mortensen P, Mann M. Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. Genome Biol [Internet]. 2006 Jan [cited 2014 Mar 25];7(6):R50. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1779535&tool=pmcentrez&rendertype=abstract

19. Science M. Mascot search overview [Internet]. 2013. Available from: http://www.matrixscience.com/search_intro.html

20. Dauly C, Perlman DH, Costello CE, Mccomb ME. mapping analyses. 2008;5(7):1688–700. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2590652/pdf/nihms61185.pdf

21. Tyers M, Mann M. From genomics to proteomics. Nature [Internet]. 2003 Mar 13;422(6928):193–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12634792

22. Pandey A, Mann M, Interaction P, Denmark S, M D-O, Protana AS. genes and genomes. 2000;405(June). Available from: http://www.nature.com/nature/journal/v405/n6788/pdf/405837a0.pdf

23. Abbott a. A post-genomic challenge: learning to read patterns of protein synthesis. Nature [Internet]. 1999 Dec 16;402(6763):715–20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10617183

24. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science [Internet]. 2001 Feb 16 [cited 2014 Mar 19];291(5507):1304–51. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11181995

25. Science A, U.S. Department of Energy Office of Science, Office of Biological and Environmental Research HGP, JainPharmaBiotech. Proteomics [Internet]. Available from: http://www.ama-assn.org//ama/pub/physician-resources/medical-science/genetics-molecular-medicine/current-topics/proteomics.page

26. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature [Internet]. 2003 Mar 13;422(6928):198–207. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12634793

27. Anderson NL. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. Mol Cell Proteomics [Internet]. 2002 Oct 29 [cited 2014 Mar 20];1(11):845–67. Available from: http://www.mcponline.org/cgi/doi/10.1074/mcp.R200007-MCP200

28. Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H, et al. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. Proteomics [Internet]. 2005 Aug [cited 2014 May 8];5(13):3226–45. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16104056

29. Al XL et. Comparison of alternative analytical techniques for the characterisation of the human serum proteome in HUPO Plasma Proteome Project [Internet]. 2005. Available from: http://onlinelibrary.wiley.com/doi/10.1002/pmic.200401226/pdf

30. Al SYC et. Efficient prefractionation of low-abundance proteins in human plasma and construction of a two-dimensional map [Internet]. 2005. Available from: http://onlinelibrary.wiley.com/doi/10.1002/pmic.200401310/pdf

31. Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biol Mass Spectrom [Internet]. Wiley Subscription Services, Inc., A Wiley Company; 1984 Nov 1;11(11):601. Available from: http://dx.doi.org/10.1002/bms.1200111109

32. Johnson RS, Martin SA, Biemann K, Stults JT, Watson JT. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. Anal Chem [Internet]. American Chemical Society; 1987 Nov 1;59(21):2621–5. Available from: http://dx.doi.org/10.1021/ac00148a019

33. Source I. De Novo Peptide Sequencing Tutorial. Available from: http://ionsource.com/tutorial/DeNovo/introduction.htm

34. Mobile H, Considerations P. Instrumentation of HPLC Mobile Phase Considerations. Available from: http://forensicscienceeducation.org/wp-content/uploads/2013/04/Instrumentation_Of_HPLC_Mobile_Phase_Considerations.pdf

35. Chromatography I. Gas-Liquid Chromatography. 2011;3–22. Available from: http://web.vscht.cz/~cajkat/doc/ABAPL/ABAPL_07.pdf

36. Waters. What Is HPLC? [Internet]. 2014. Available from: http://www.waters.com/waters/en_US/HPLC---High-Performance-Liquid-Chromatography/nav.htm?locale=en_us&cid=10048919

37. Molnár I, Horváth C, Goizueta RC, Haven N. Csaba Horváth. 2004; Available from: http://www.lcgceurope.com/lcgceurope/data/articlestandard/lcgceurope/282004/102973/article.pdf

38. Hplc C. HPLC Basics Fundamentals of Liquid Fundamentals of High Performance Liquid Chromatography (HPLC) This course will enable you to : 1–37. Available from: http://polymer.ustc.edu.cn/xwxx_20/xw/201109/P020110906263097048536.pdf

39. Niessen WMA, Tinke AP. Liquid chromatography-mass spectrometry General principles and instrumentation. J Chromatogr A [Internet]. 1995;703:37–57. Available from: http://www.sciencedirect.com/science/article/pii/002196739401198N

40. Zapala W. Influence of Mobile Phase Composition on Retention Factors in Different HPLC Systems with Chemically. 2003;41(July):289–94. Available from: http://chromsci.oxfordjournals.org/content/41/6/289.full.pdf?origin=publication_detail

41. Niessen wilfried MA. Liquid Chromatography - Mass Spectrometry [Internet]. Available from: http://www.crcnetbase.com/doi/pdfplusdirect/10.1201/9781420014549.fmatt

42. Aitken A, Learmonth MP, Absorbance FU V. Protein Determination by UV Absorption. :3–6. Available from: http://download.bioon.com.cn/upload/month_0810/20081028_b89a5bd42d711d37fb77m6aaG8vGhQjS.attach.pdf

43. 2006 EKD (in J by HN. Detectors for HPLC: UV, VIS, and PDA Detectors [Internet]. Showa Denko America, Inc., 420 Lexington Ave, Suite 2335A New York, NY 10170; 2006. Available from: http://www.shodex.net/index.php?seitenid=1&applic=1485

44. Gingras A-C, Aebersold R, Raught B. Advances in protein complex analysis using mass spectrometry. J Physiol [Internet]. 2005 Feb 15 [cited 2014 Mar 19];563(Pt 1):11–21. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1665575&tool=pmcentrez&rendertype=abstract

45. Mobley JA. Qualitative Proteomics [Internet]. Available from: https://www.uab.edu/proteomics/workshops/2006/pdf/Qualitative Proteomics.pdf

46. Ashcroft AE. An Introduction to Mass Spectrometry [Internet]. Available from: http://www.astbury.leeds.ac.uk/facil/MStut/mstutorial.htm

47. McLaren WCW and IH. Time-of-Flight Mass Spectrometer with Improved Resolution [Internet]. Available from: http://masspec.scripps.edu/mshistory/timeline/time_pdf/1955_WileyWC.pdf

48. W.C. Wiley & I.H. MacLaren, R.J. Cotter, Mamyrin et al MG. Time-of-Flight (TOF) Mass Analysis [Internet]. Available from: http://www.chm.bris.ac.uk/ms/newversion/tof-ms.htm

49. Al AL et. Mass spectrometry of intact membrane protein complexes [Internet]. p. 639–51. Available from: http://www.nature.com/nprot/journal/v8/n4/fig_tab/nprot.2013.024_F2.html

50. Michalski A, Damoc E, Hauschild J-P, Lange O, Wieghaus A, Makarov A, et al. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. Mol Cell Proteomics [Internet]. 2011 Sep;10(9):M111.011015. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3284220&tool=pmcentrez&rendertype=abstract

51. Staes A, Van Damme P, Helsens K, Demol H, Vandekerckhove J, Gevaert K. Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). Proteomics [Internet]. 2008 Apr [cited 2014 Mar 29];8(7):1362–70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18318009

52. Gevaert K, Vandekerckhove J. COFRADIC™: the Hubble telescope of proteomics. Drug Discovery Today: TARGETS. 2004. p. 16–22.

53. Helsens K. Increasing the specificity and the sensitivity in peptide centric proteomics Increasing the specificity and the sensitivity. 2010;

54. Department of Medical Protein Research Ug. COFRADIC: COmbined FRActional DIagonal Chromatography [Internet]. Available from: http://penyfan.ugent.be/labo/joelv/Cofradic.html

55. Brown JR, Hartley BS. Location of disulphide bridges by diagonal paper electrophoresis. The disulphide bridges of bovine chymotrypsinogen A. Biochem J. Nature Publishing Group; 1966;101:214–28.

56. Staes A, Impens F, Van Damme P, Ruttens B, Goethals M, Demol H, et al. Selecting protein N-terminal peptides by combined fractional diagonal chromatography. Nat Protoc [Internet]. Nature Publishing Group; 2011 Aug [cited 2014 Mar 20];6(8):1130–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21799483

57. Gevaert K, Goethals M, Martens L, Van Damme J, Staes A, Thomas GR, et al. Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. Nat Biotech [Internet]. Nature Publishing Group; 2003 May;21(5):566–9. Available from: http://dx.doi.org/10.1038/nbt810

58. Staes A, Timmerman E. Enrich for your kind of peptides : COFRADIC , get the. Nat Biotechnol. 2003;21:566–9.

59. Protein B, Reagent E. Novagen - BugBuster® Protein Extraction Reagent. User Protoc TB245 Rev E 0304 [Internet]. 1800;1–8. Available from: http://pef.aibn.uq.edu.au/wordpress/wp-content/blogs.dir/1/files/Support/Bacteria/Manuals/Bugbuster_Protein_Extraction_Reagent.pdf

60. Ampules AS. Pierce ™ BCA Protein Assay Kit. 2000;0747(23225). Available from: http://www.piercenet.com/instructions/2161296.pdf

61. Hsu J, Huang S, Chow N, Chen S. Stable-Isotope Dimethyl Labeling for Quantitative. 2003;75(24):6843–52.

62. Sigma-Aldrich. Trypsin [Internet]. Available from: http://www.sigmaaldrich.com/life-science/metabolomics/enzyme-explorer/analytical-enzymes/trypsin.html

63. Tips P, Extractions M. OMIX Pipette Tips for Micro Extractions Agilent Bond Elut OMIX. Available from: https://www.chem.agilent.com/Library/primers/Public/5990-9049EN-Omix-Sept11-lo.pdf

64. Watson JT, Sparkman OD. Introduction to Mass Spectrometry. Chichester, UK: John Wiley & Sons, Ltd; 2007 Oct 5; Available from: http://doi.wiley.com/10.1002/9780470516898

65. Venne a S, Vögtle F-N, Meisinger C, Sickmann A, Zahedi RP. Novel highly sensitive, specific, and straightforward strategy for comprehensive N-terminal proteomics reveals unknown substrates of the mitochondrial peptidase Icp55. J Proteome Res [Internet]. 2013 Sep 6;12(9):3823–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23964590

66. Dearborn NJ and DG. Labeling of Proteins by Reductive Methylation Using Sodium Cyanoborohydride*. J Biol Chem Vol 254, No 11, Issue June 10, pp 4359-4365, 1979 Print USA. 1979;254, NO 11.

67. Gevaert K. Chromatographic Isolation of Methionine-containing Peptides for Gel-free Proteome Analysis: Identification Of More Than 800 Escherichia Coli Proteins. Mol Cell Proteomics [Internet]. 2002 Oct 24 [cited 2014 Apr 26];1(11):896–903. Available from: http://www.mcponline.org/cgi/doi/10.1074/mcp.M200061-MCP200

68. Damme P Van, Impens F, Gevaert K. Functional Proteomics. Thompson JD, Ueffing M, Schaeffer-Reiss C, editors. Totowa, NJ: Humana Press; 2008 [cited 2014 May 3];484. Available from: http://link.springer.com/10.1007/978-1-59745-398-1

69. Erde J, Loo RR LJ. Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments [Internet]. p. 13(4):1885–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24552128

70. Roger L. Lundblad. Boca Raton F. The Evolution from Protein Chemistry to Proteomics: Basic Science to Clinical Application [Internet]. 2006. p. 295 pp, chapter 2. Available from: http://www.clinchem.org/content/52/4/789.1.full

71. Scientific T. Sodium Dodecyl Sulfate (Lauryl SDS) [Internet]. Available from: http://www.piercenet.com/product/sds-sodium-dodecyl-sulfate-lauryl

72. HEPES Use in Cell Culture [Internet]. Life Technologies; Available from: http://www.lifetechnologies.com/no/en/home/life-science/cell-culture/mammalian-cell-culture/reagents/hepes.html

73. Ph SS, Manager MD. The Q Exactive – A Benchtop Orbitrap Mass Spectrometer for DMPK Presented to: 5. 2012; Available from: http://bcn2012.europeanbioanalysisforum.eu/slides/day 2/6 technology updates ii/2_szwandt.pdf

74. Gevaert K, Vandekerckhove J. CHAPTER 4 Peptide sorting by reverse-phase diagonal chromatography. 2008;61–82. Available from: http://www.scionpublishing.com/shop/ProductImages/Proteomics Chapter 4.pdf

75. Van Damme P, Van Damme J, Demol H, Staes A, Vandekerckhove J, Gevaert K. A review of COFRADIC techniques targeting protein N-terminal acetylation. BMC Proc. 2009;3 Suppl 6:S6.

10. APPENDIX 1

The analytical tools utilized in the laboratory work is summarized in the subsequent tables.

Table 13: List over analytical equipment used in this thesis.

Equipment	Model & Producer
Pipettes	
0.1 – 2.5 µl	Eppendorf Research plus (ADJ, Germany)
0.5 – 10 µl	Eppendorf Research plus (ADJ, Germany)
10 – 100 µl	Eppendorf Research plus (ADJ, Germany)
100 – 1000 µl	Eppendorf Research plus (ADJ, Germany)
0.5 – 5 mL	Eppendorf Research plus (ADJ, Germany)
Shaker	
Model 1: Vortex mixer	Heidolph Reax (Germany)
Model 2: MS1 Minishaker	IKA-works INC (USA)
Centrifuge	
Model 1	Biofuge fresco, Heraeus instruments (made in Germany)
Model 2: Mini Star Silverline (Galaxy mini centrifuge)	VWR international (made in Korea)
Model 3	Eppendorf AG Centrifuge 5810 R (Hamburg, Germany)
Heating	
Model 1: Heating cabinet	Termaks AS (Bergen, Norway)
Block heater	Stuart sample concentrator, SBHCONC/1, SBH200D/3 (Stone Staffordshire, UK)
Drying	
Automatic Environmental SpeedVac® System AES1010 with VaporNet®	Savant Instruments, INC. Holbrook, NY (US)
Nitrogen gas	-
pH – meter	
Manual 744 pH meter	Metrohm ion analysis (Switzerland)
pH-indicator paper	Universal indicator paper from Merck

Table 14: The following equipment / instrumentation were used during UV-HPLC analysis:

Equipment/instrumentation	Type	Manufacturer	Batch No
UV-HPLC Vial	9 mm Screw top vial, 12x32mm with cap & PTFE/Silicone septa	Waters (The science of what`s possible™) (Made in USA)	P/N: 186000272C L/N: 0272540300
UV-HPLC Vial	Polypropylene plastic, snap-top vial, 300 µl, with cap and bonded, PTFE/Silicone septa	Waters (product of USA)	P/N: 186002642 L/N: 2642511990
Insert*	Insert 150 µl with preinstalled plastic spring	Waters (made in USA)	P/N: WAT094171 Lot: 4171633570
Screw cap	Screw cap with bonded, pre-slit PTFE/Silicone septa for 12x32 mm vial	Waters (product of USA)	P/N: 186000305 L/N: 1522520110
Fraction collector III		Waters (made in Japan)	
Tubes for fraction collection	Test Tube Soda glass, w/o rim, round bottom (70x10x0.8-1.0mm)	VWR International (made in Germany)	European catalog No. 212-0012

*The inserts were utilized prior to primary and secondary run so that the vials can be used several times.

Table 15: Additional analytical equipment applied throughout this thesis.

Equipment	Type	Lot #	Manufacturer
Eppendorf tupe	Safe-Lock Tubes 1.5 mL	C152631M	Eppendorf AG, Germany, Hamburg
96-well plate	MICROTEST Tissue cellulaire plate, 96 well, Flat Bottom with Low Evaporation	REF 353072	Falcon, Becton Dickinson, France
30 kDa Pall Spin-Filter	VWR Centrifugal Filter, 500 µl	88147-352C Cat. No. 8231-352	VWR North America (Mexico)
Bond Elut, Agilent Sample Prep Solutions	OMIX, 96 C18, 100 µl	6180803-03	Agilent Technologies (USA)
Pipette tips for Micro Extractions	OMIX, SCX MB 96/PK (10 µl)	41201-5356	Varian 96 Tip Tray (USA)
Atlantis™ dC 18, 3µm Column	Waters, 75 µm X 100 mm nanoAcquity™ Column	M052791	Waters, Made in USA.
Peptide separation Column	Waters, nanoAcquity UPLC® Column, 100 µm x 100 mm, 1,7 µm, BEH130 C18	0193111941	Waters, Made in USA

Table 16: Software that have been used

Instrument	Software type
n-HPLC-MS/MS	MassLynx 4.1 SCN744
UV-HPLC	Empower - Alliance 2695
MALDI micro MX	MassLynx V4.1

11. APPENDIX 2

The chemicals used for the work within this project are listed below.

Table 17: Chemicals purchased from J.T. Baker (Deventer, Holland)

Chemical	CAS No	Purity (%)	Product No	Batch No
Ammonium Hydroxide 25 % solution	-	25.6	6051	B 4670
Urea	-	-	0346	83972

Table 18: Chemicals purchased from Sigma-Aldrich (Germany)

Chemical	CAS No	Purity (%)	Product No	Batch No
Acetonitrile CHROMASOLV®	75-05-8	≥99.9	11143996	SZBD230AV
Guanidine hydrochloride	50-01-1	≥98	101282312	BCBJ7434V
Methanol	67-56-1	≥99.8	603-001-00-x	SZBC2220V
Methanol CHROMASOLY®, for HPLC	67-56-1	≥99.9	603-001-00-x	SZBC272MV
Hydrochloric acid ≥37% solution	7647-01-0	≥37	017-002-01-X	SZBB0500V

Table 19: Chemicals purchased from Sigma-Aldrich (USA)

Chemical	CAS	Purity (%)	Product No	Batch No
4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid (HEPES)	7365-45-9	≥99.5	1001554967	SLBB6189V
HEPES	7365-45-9	≥99.5	101097713	041M54001V
Boric Acid	10043-35-3	≥99.5	101405919	BCBK5744V
Iodacetamide (IAA)	144-48-9	≥99	1001494931	SLBC7561V
Trizma® Hydrochloride	1185-53-1	≥99.0	1001385504	SLBC7057V
Triflouroacetic Acid	76-05-1	99	101157433	BCBH2182V

Table 20: Chemical purchased from Sigma-Aldrich (Spain)

Chemical	CAS No	Purity (%)	Product No	Batch No
Ammonium bicarbonate (ABC)	1066-33-7	≥99.0	101297402	BCBJ0213V

Table 21: Chemical purchased from Riedel – de – Haen (Seelze, Germany)

Chemical	CAS No	Purity (%)	Product No	Batch No
Acetonitril CHROMASOLV®, for HPLC	75-05-8	99.9	R34851	3341A

Table 22: Chemicals purchased from Merck (Germany)

Chemical	CAS No	Purity (%)	Product No	Batch No
Di-sodium hydrogen phosphate	-	≥99.9	1.06579.1000	A992379 845
Di-sodium hydrogen phosphate anhydrous	1.4561-500	99.0	1.06586	751 509 1105
Sodium hydroxide	215-185-5	>99.98	1.06469.100	-

Table 23: Chemicals purchased from Sigma-Aldrich (Canada)

Chemical	CAS No	Purity (%)	Product No	Batch No
DL-Dithiothreitol (DTT)	3483-12-3	≥99.0	1001437609	110M1424V
Calcium chloride (granules)	10043-52-4	≥93.0	C1016	017-013-00-2

Table 24: Chemical purchased from Fluka Analytical (China)

Chemical	CAS No	Purity (%)	Product No	Batch No
Hydroxylamine hydrochloride	5470-11-1	≥99.0	101228717	BCBJ5684V

Table 25: Chemical purchased from Sigma-Aldrich (Austria)

Chemical	CAS No	Purity (%)	Product No	Batch No
Formaldehyde 37% solution	-	37	UN 2209	SZBC1430V

Table 26: Chemical purchased from VWR International, BDH, PROLABO (CE - France)

Chemical	CAS No	Purity (%)	Product No	Batch No
Sodium Acetate	-	>98	27 650.292	0502209

Table 27: Chemical purchased from Fluka Chemika (Switzerland)

Chemical	CAS No	Purity (%)	Product No	Batch No
Sodium cyanoborohydride	25895-60-7	~90 (RT)	GA12473	Analysis #: 381960/241298

Table 28: Chemical purchased from Thermo Scientific (Rockford, USA)

Chemical	CAS No	Purity (%)	Product No	Batch No
Sulfo-NHS-Acetate (Sulfosuccinimidyl Acetate)	-	-	26777	OC184750

Table 29: Chemical purchased from AppliChem (Darmstadt, Germany)

Chemical	CAS No	Purity (%)	Product No	Batch No
Glycine	56-40-6	99.8	A1067.1000	R7654

Table 30: Chemical purchased from VWR International PROLABO (Made in EC-European Community)

Chemical	EC No	Purity (%)	Product No	Batch No
Hydrogen peroxide	2317650	30	23612.294	06J270022

Table 31: Trypsin from porcine pancreas was used as proteolytic enzyme in this study:

Substance	CAS No	Product No	Batch No	Producer
Trypsin from porcine pancreas	9002-07-7	1001334850	SLBC8958V	Sigma-Aldrich (Purchased from USA)
Trypsin from porcine pancreas	9002-07-7	1001618886	SLBH4422V	Sigma-Aldrich (product of USA)

*Trypsin purchased from USA was used in general, but two different orders had various product and batch numbers.