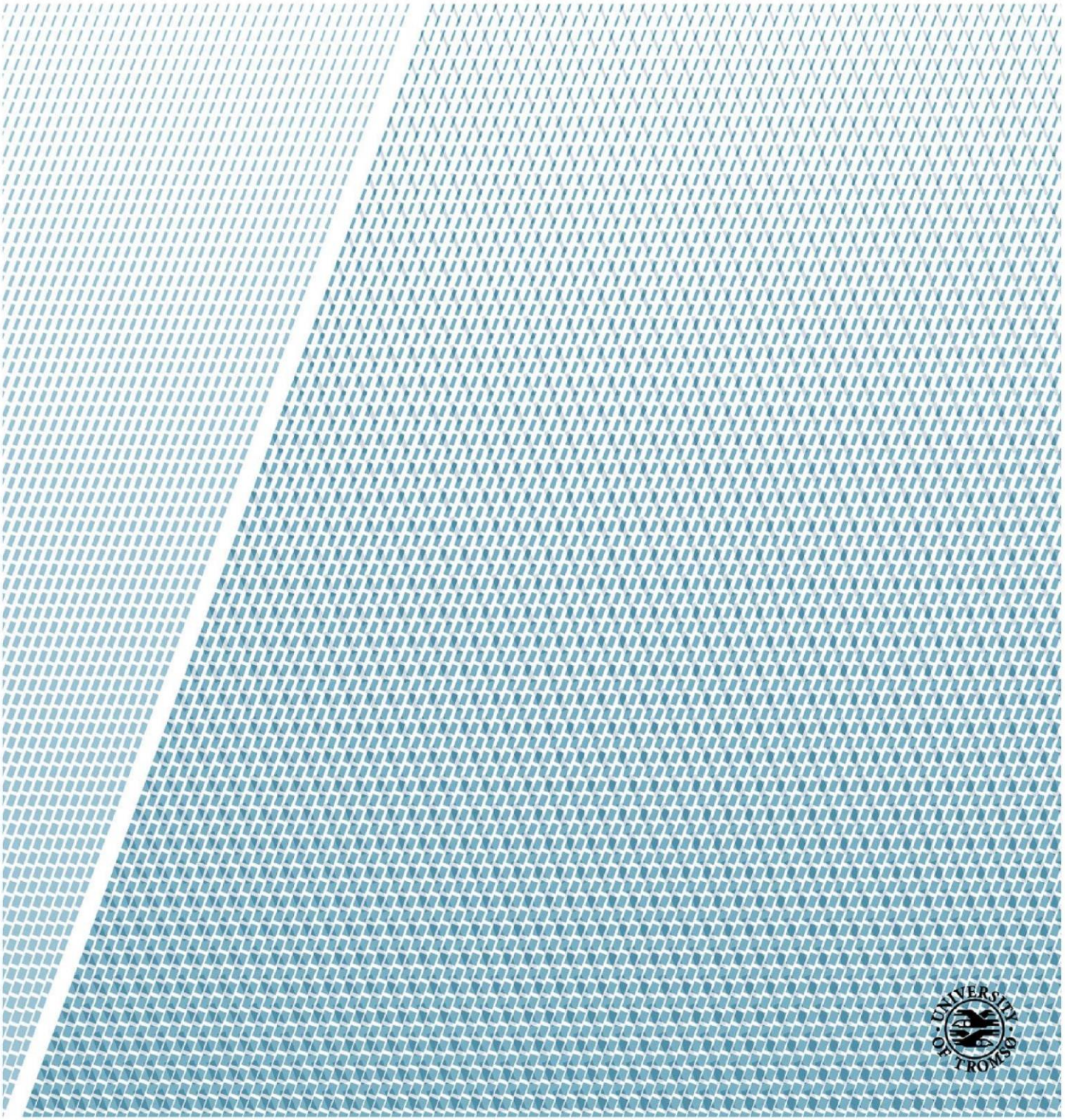
Department of Pharmacy, Faculty of Health Sciences
UiT - The Arctic University of Norway

Quantitative Phosphoprotein Analysis in Adipose Tissue

—

Edmund Theodore Fiifi Assignon

Thesis for the degree Master of Pharmacy, May 2016



MASTER THESIS FOR THE DEGREE MASTER OF PHARMACY

QUANTITATIVE PHOSPHOPROTEIN ANALYSIS IN ADIPOSE TISSUE

BY

EDMUND THEODORE FIIFI ASSIGNON

SUPERVISORS

Associate professor Terkel Hansen

Ph.D. candidate Yvonne Pasing, UNN

Natural Products and Medicinal Chemistry

Department of Pharmacy

Faculty of Health Sciences

UiT – The Arctic University of Norway

In loving memory of Jimmy Assignon and Rose Abblode

Acknowledgements

The work accomplished in this thesis was carried out at Natural Products and Medicinal Chemistry Research Group, Department of Pharmacy, UiT –The Arctic University of Norway during the period September 2015 to May 2016. During this period, I have received a lot of support from people around me.

First and foremost, I would like to show my deepest gratitude to my thesis supervisor Terkel Hansen of the Department of Pharmacy at University of Tromsø. I want to thank you for the continuous support throughout the entire project. The door to Associate professor Terkel Hansen's office was always open whenever I ran into obstacles, had a question about my laboratory work or writing. He consistently supported my decisions in order to allow this paper to be my own work, and eased the pressure off me.

I would also like to thank my second supervisor Ph.D. candidate Yvonne Pasing at Universty Hospital of North Norway for introducing me to the lab and the field of proteomics. I am very grateful for everything you have thought me about proteomics, and enjoyed learning it

I would further like to thank the Natural Products and Medicinal Chemistry Research Group for taking me in with open arms, and for having a very friendly and quite stimulating research environment. Special thanks to Ph.D. candidate Armin Schniers, Senior engineers Martina Havelkova, Trude Anderssen, Jack-Ansgar Bruun and Toril Anne Grønset for sacrificing your valuable time in order to help me in the laboratory and answering my questions.

I would like to deepen my gratitude to Ph.D. candidate Armin Schniers for your readily availability when I needed help, and Senior Engineer Trude Anderssen for input to improve my work. Thank you very much.

I want to thank my fellow master students at the Department of Pharmacy as well as other students at the University of Tromsø for making my period in Tromsø very memorable, and initiating what I truly believe is going to be life-long friendships.

Thank you Lloyd Mbugua Kangu, Damir Dugalic, Kristian M. Sørensen, Abel Teckle, Nikolai Ciric, Peter Moen, Awfa Al-Shamkawy, Sabah Bashir and Lars-Arne Johansen

Finally, I must express my very profound gratitude to my family and my friend for providing mw with endless support and continuous encouragement throughout my years of study and through process of researching and writing this thesis.

Abstract

Background: Mass spectrometry-based proteomics has increasingly been the choice of method for the global analysis of the composition, modification and dynamics of proteins. Quantitative analysis of the proteome with tandem mass tag is a technique for calculating the relative abundance of the proteome in tissues and organelles. In this thesis, present a novel bottom-up method for the quantitative analysis of the phospho proteome in adipose tissue.

Materials and methods: The approach employed in this thesis is a bottom-up based method. The samples were enriched with TiO2 using the enrichment protocols published by Dickhut et al. to determine which protocol yielded the highest number of phosphopeptides. The next step was to validate the procedure by running the procedure on HeLa cells. The last step in procedure was to determine the labeling of the samples (before or after enrichment). The post-processing of data output was done on PD, Peaks and PeptideShaker to determine the most compatible search engine.

Results and discussion: Mascot in PD identified the highest number of phosphopeptides with 252 confident identifications compared to Sequest HT (146), Peaks (239) and PeptideShaker (197). Mascot was unique in 51 phosphopeptides compared to Peaks' 33 unique phosphopeptide identifications. PD was also successful in interpreting the fragment spectra giving relatively good information about the fragments compared to Peaks and PeptideShaker. Micro column protocol was more successful in terms of number of identification (252) compared to the batch mode protocol (147 identified). PhosSTOP had a negative effect with the results and was therefore excluded from further experiments, as the phosphopeptides identified with PhosSTOP were lower and less confident. The procedure was validated using HeLa cell with 1068 phosphopeptides confidently identified. The labeling step was determined to be after enrichment based on the number of confident identification (147 phosphopeptides labeled before enrichment versus 3 phosphopeptides labeled after enrichment).

Conclusion: In this this study, a method for quantifying the phospho proteome in adipose tissue has been presented. The method conducts the micro column protocol, which has proven to be more successful than the batch mode protocol. The procedure was validated using HeLa cells, which worked successfully. The labeling of samples was determined to be conducted before enrichment due to low yield when labeling after enrichment. Mascot search engine in PD was conducted due to the higher compatibility with the data compared to PeptideShaker and Peaks.

List of abbreviations

Abbreviation	Explanation
MS	Mass spectrometer/spectrometry
MS^2	Tandem mass spectrometer/spectrometry
PTM	Posttranslational modifications
PPAR	Peroxisome proliferator-activated receptor
C/EBP	CCAAT enhancer-binding proteins
WAT	White adipose tissue
BAT	Brown adipose tissue
MSC	Multipotent mesenchymal stem cells
hMADS	Human multipotent adipose tissue derived stem cells
UCP-1	Uncoupling protein-1
PMF	Peptide Mass Fingerprinting
ESI	Electrospray Ionization
MALDI	Matrix-Assisted Laser Desorption/Ionization
LC	Liquid chromatography
IMAC	Immobilized metal affinity chromatography
IDA	Iminodoacetate
NTA	Nitrilotriacetate
MOAC	Metallic oxide affinity chromatography
TFA	Trifluoroacetic acid
TMT	Tandem mass tag
SILAC	Stable isotope labelling by amino acids in cell culture
ICAT	Isotope-coded affinity tag
ISCL	Isotope-coded protein labelling
iTRAQ	Isotope tag for relative and absolute quantification
HPLC	High pressure liquid chromatography
DDA	Data dependent analysis
FFT	Fast Fourier Transform
CID	Collision-induced dissociation
CAD	Collision-activated dissociation
HCD	Higher collision dissociation
ECD	Electron capture dissociation
ETD	Electron transfer dissociation
PD	Proteome Discoverer
fXCorr	Fast cross-correlation
PSM	Peptide spectrum matches
SDC	Sodium deoxy cholate
AmBic	Ammonium bicarbonate
TEAB	Triethylammonium bicarbonate
DTT	Dithiotreitol
IAA	Iodoacetic acid
ACN	Acetonitrile
FA	Formic acid
BCA Assay	Bicinchoninic acid assay
PS	PhosSTOP
NPS	No PhosSTOP
WR	Working reagent
RT	Room temperature
HFBA	Heptafluorobutyric acid
FDR	False discovery rate

List of figure and tables

Table of contents

1 Introduction

1.1 Background

Proteomics is the understanding of proteins in cell and organism. From quantitative proteomics, information is obtained about the abundancy of the proteome in a tissue and how alterations of proteins can be modified prior to, e.g. ligand binding. Mass spectrometry (MS) has increasingly become the choice of method for analysis of complex protein mixtures (1, 2). In brief, a mass spectrometer consist of a ion source that ionizes the samples, a mass analyzer that measures the mass-to-charge ration (m/z) of ionized analytes and a detector that registers the intensity of ions at each m/z value (2)

Quantitative proteomics may involve the labeling of proteins or peptides. It is a powerful approach for the discovery and targeting of the proteome in order to understand the global proteomic dynamics in a cell, tissue or an organism. In the clinical aspect of proteomics, label-free quantification of large samples analysis is usually conducted for identifying biomarkers or clinical screening. This is a rapid and a low-cost alternative to labeling of samples, where large changes in the proteome is observed (3, 4). Relative quantitation in proteomics compares the abundancy of peptides between samples (5). The quantitative method for profiling the phospho proteome in this thesis conducts of a bottom-up approach, followed by analysis on MS or tandem mass spectrometry (MS^2) to obtain the sequential information about the peptide. The data is further processed by searching against a protein database using search engines. The data obtained from the search engines are used to do a relative quantitation on the samples (6).

Phosphoproteomics is the main target in this project and human adipose tissue is the source used to extract protein. The human adipose tissue because it is of interest to analyze the phospho proteome upon exposition to Vitamin D, as human adipose tissue also functions as the storage for vitamin D (7, 8).

In this master thesis, a novel method will be presented for the quantification of the phospho proteome in adipose tissue

1.2 Biology of proteins

1.2.1 The Central dogma of molecular biology

"The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred from protein to either protein or nucleic acid"

By Francis Crick(9)

The central dogma of molecular biology describes the general pathways of information. Briefly, DNA replicates itself and thereafter DNA is transcripted to RNA. Finally, RNA is translated to protein (fig. 1.2-1) (10)

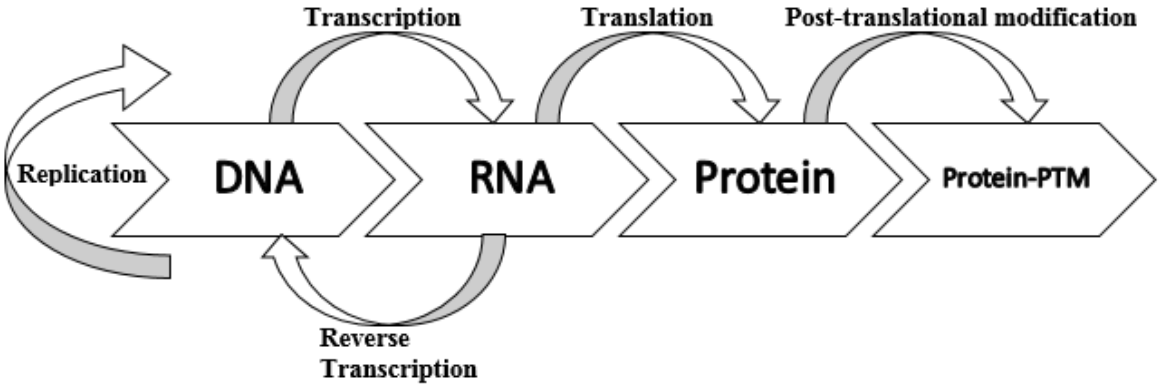


Figure 1.2-1: Illustration of the central dogma in molecular biology.

1.2.2 Protein architecture

A peptide is formed when two or more amino acids are covalently joined together, with the elimination of water through the formation of the peptide bond. As the peptide chain elongates, domains and subunits are formed, which completes the complex structure of a protein (11).

Proteins may consist of 20 different amino acids. Each amino acid has different side groups that can influence the properties of proteins. For example, a protein that mainly consists of polar amino acids will gain hydrophilic properties (10, 11).

The protein structure is defined on four levels; Primary structure, secondary structure, tertiary structure and quaternary structure. The primary structure is a description of amino acids linked together through peptide bonds. It also carries information about possible posttranslational modification sites, e.g. cysteine bridges or possible phosphorylation sites. The secondary structure describes the arrangement of peptides, such as in this case an α-helix

or a β-sheet. Several arranged polypeptide units linked together creates a complex domain or subunits, called the tertiary structure. All domains or subunits together completes the complex molecule, described as the quaternary structure. An illustration of the protein structure is shown in fig. 1.2-2 (11).

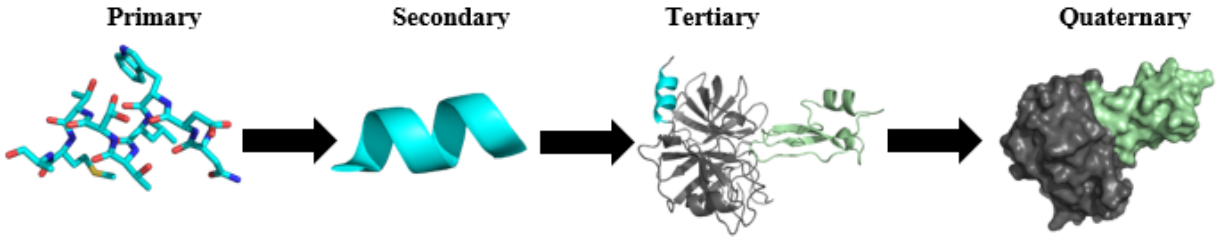


Figure 1.2-2: The levels of protein structure from Anionic Salmon Trypsin (grey) bound to Bovine Pancreatic Trypsin Inhibitor (pale green).

1.2.3 Phosphorylation

After translation, proteins can go through posttranslational modification (PTM) required to differentiate their functions as well as stabilization, degradation and control of various biological processes (1).

Some of the modifications proteins can go through after translation are for example phosphorylation; the attachment of phosphor group, glycosylation; the attachment of a glycosyl group and ubiquitination; the attachment of small proteins called ubiquitin. These modifications along with sulfation, methylation, prenylation, lipidation, hydroxylation and carboxylation are the most common posttranslational modification of proteins (1, 12). An example of the phosphorylation/dephosphorylation of a serine residue is shown in fig. 1.2-3.

In mammalians, approximately 30% of the PTMs are phosphorylation. Phosphorylation of proteins implies attaching at least one phosphor group to a serine, threonine or tyrosine residues in proteins. Phosphorylation is often referred to as a molecular switch because it can alter the activity of a protein, The protein will then functions as an activated or deactivated protein until enzymes that add (kinases) or removes (phosphatases) phosphor groups catalyzes the reaction. (1, 12, 13). In terms of kinases, it has been observed that serine and threonine residues in protein are more likely to be phosphorylated than tyrosine residue, with a phosphorylation ratio of 90:10 (13).

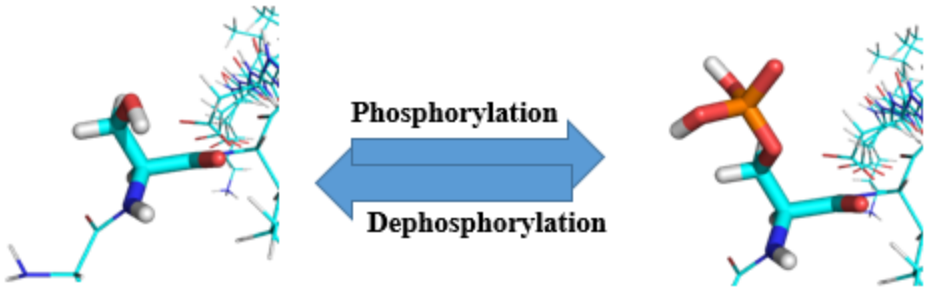


Figure 1.2-3: Phosphorylation of serine residue from the peptide sequence MSLPDVDLDLK.

This is due to the abundancy of serine/threonine kinases compared to tyrosine-specific kinase; of the discovered 500 kinases approximately 100 kinases are tyrosine-specific kinases

Protein kinases and phosphatases are dynamic, a mechanism that is well-known in signal transduction, where a ligand binds to a receptor, and a cascade of protein activation occurs as the result of phosphorylation (12).

1.2.4 Other post-translational modifications

Glycosylation

Glycosylation is the addition of a carbohydrate molecule primarily on asparagine, serine and threonine residues, but could also appear on hydroxylated proline and lysine residues (12). The carbohydrate component of glycosylated proteins may vary in size. Glycosylated proteins where the carbohydrate component is a major portion are called proteoglycans, whereas glycosylated proteins where the carbohydrate component is a smaller portion are called glycoproteins (1).

Ubiquitination

Ubiquitin is a small protein of 8 kD, with a functionality as a protein marker. The knowledge about ubiquitinated proteins are limited but so far it is understood to be readily for degradation by proteasome, or other functions as DNA repair, cell cycle and transcription. It attaches to a lysine residue on a protein from its glycine residue. Ubiquitin can build continuously on another ubiquitin (poly-ubiquitin) creating a long ubiquitin-chain or attached to several lysine residue (multi-ubiquitin) on a protein (1).

Sulfation

Some proteins undergo sulfation, which is the attachment of sulfate group to a tyrosine residue. Although the knowledge about the biological function of tyrosine O-sulfation is limited, it is believed to be involved in hemostasis regulation, a modulator for protein-protein interactions of secreted and membrane-bound proteins (14, 15).

Methylation

There are three types of methylation on proteins; N-Methylations which is understood to play an important role in gene transcription and a minor role in signal transduction, O-Methylation (poorly understood) and S-Methylation (also poorly understood) (12).

Lipidation

Some proteins may undergo lipid modifications which is the attachment of cholesterol, isoprenoids, fatty acids etc. to proteins, increasing their lipophilicity and giving them transmembrane domains (1).

Hydroxylation

Hydroxylation is the addition of a hydroxyl-group on proline, lysine or asparagine. Hydroxyproline and hydroxylysine are important components for collagen fiber, while hydroxyasparagine can be found in antifungals and antibiotics (12)

1.3 Adipose Tissue

1.3.1 Anatomy and physiology

Previously adipose tissue was not considered interestingly to study. It was considered an organ or tissue that stored triacylglycerol. Adipose tissue in specific regions of the body are referred to as depots, and are currently found subcutaneous, visceral, in bone marrow, intermuscular and in the mammary glands. With increasing interest, scientists have revealed that adipose tissue plays a bigger role than to store fat, e.g. energy storage and secreting biological compounds in order to regulate metabolic homeostasis in addition responding to physiological and environmental stimuli due to its connection to the nervous system (16, 17).

To understand the generation of adipocytes, there are two master regulators of adipogenesis: Peroxisome proliferator-activated receptor (PPARγ) and CCAAT-enhancer-binding proteins (C/EBPα). Briefly, C/EBPα derives from the C/EBP family of transcription factors expressed in adipocytes. The other two family members are C/EBPβ and C/EBPδ. C/EBPα is the isoform proven to be crucial in regulating gene expression in adipogenesis. It is understood that the α-isoform is activated by the β- and δ-isoforms, and the transcriptions of the isoforms are regulated by insulin. (18). PPARγ, the other master regulator of adipogenesis, regulates the differentiation of pre-adipocytes into mature adipocytes. PPARγ is one of three isoforms in the PPAR family. The other two isoforms are PPARα and PPARβ/δ. The α-isoform is mostly involved in fatty acid catabolism, and exists in a lesser extent in white adipose tissue (WAT) compared to brown adipose tissue (BAT). PPARγ, existent in γ1- and γ2-isoform, is expressed in the later stages of adipogenesis. PPARγ are upregulated by hormones (i.e. insulin and glucocorticoid) and downregulated by cytokines (18). A microscopically view of the adipocyte development is shown in fig. 1.3-1.

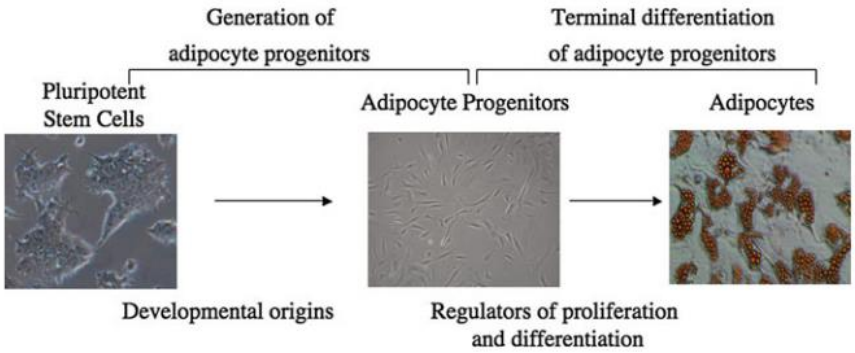


Figure 1.3-1: Development of adipocytes (Mohsen-Kansen et. al (2013)).

Adipose tissue is mostly composed of adipocytes. The differentiation of pre-adipocytes in to adipocytes, also known as adipogenesis, is described in two phases. The determination phase is where pluripotent stem cell are converted to multipotent mesenchymal stem cell (MSC), also termed human Multipotent Adipose Tissue-Derived Stem Cells (hMADS). hMADS are converted to pre-adipocytes. In terminal differentiation phase, the pre-adipocytes gains characteristics of mature adipocytes by acquiring the necessary machinery (19, 20). Besides adipocytes and pre-adipocytes, the tissue also contains fibroblast, endothelial cells and multipotent stem cells that can further differentiate into several cell types (21). There are two types of adipose tissues: WAT and BAT. WAT and BAT have distinct functions and cell morphology. WAT is mainly known for its energy storage and mobilization of lipids. This is the reason for the wide spread distribution of WAT throughout the body. On the other hand, BAT is known as a thermogenic organ burning stored fat in form of energy or heat through expression of uncoupling protein-1 (UCP-1) in mitochondria (20, 21). BAT appears brown due to its abundancy of mitochondria compared to WAT, which has larger lipid droplets than BAT (22). BAT is mainly existent in neonates, but recent studies have shown evidence of increased BAT in adults through β_3-stimuli or cold exposure. Yet, the mechanism for "white adipocyte-turn-brown adipocytes" is still unknown (20). The possible factors involved in browning of white adipose tissue is shown in fig. 1.3-2.

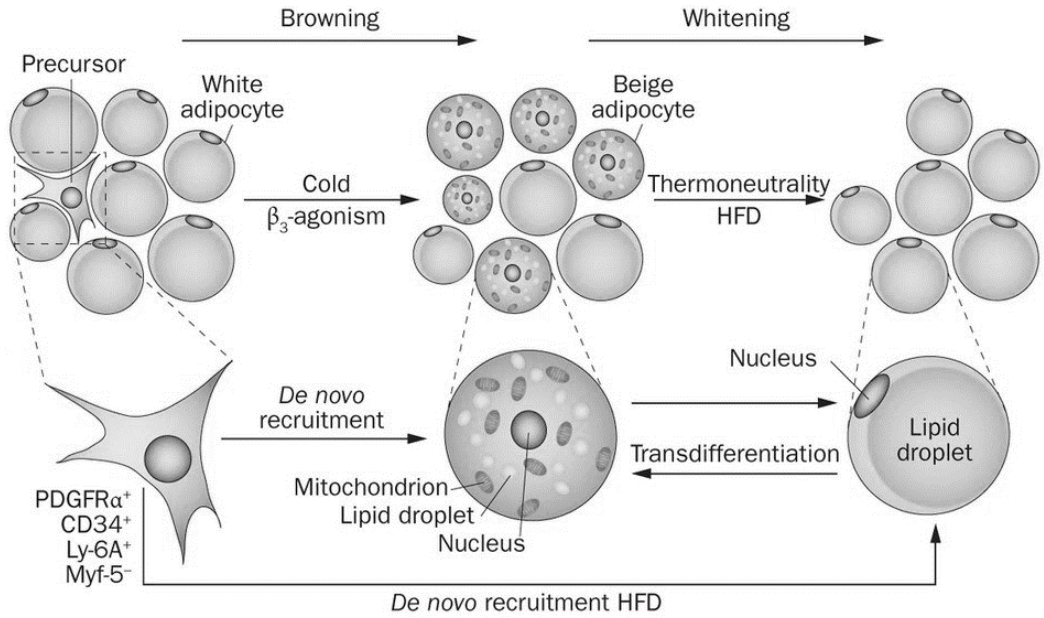


Figure 1.3-2: Adipose tissue browning (Bartelt & Heeren (2013)).

1.3.2 The proteome in adipose tissue

Adipocytes primarily store and mobilize fat depending on the expenditure rate of energy. The expenditure of fat, lipolysis and lipogenesis, is carried out by the enzymes and regulatory proteins that the adipocytes possess (23). This process is also affected by proteins secreted in adipocytes, also known as adipokines or adipocytokines, which can accelerate or slow down the metabolism of fat. Adipokines function either distant (endocrine) or locally (autocrine/paracrine). The knowledge about adipose tissue and adipocytokines has increased since leptin, the first adipocytokine discovered in 1994 (24). Adipocytokines are understood to regulate other biological processes than fat storage. These processes include blood clothing, insulin sensitivity regulation, and uptake in skeletal muscle. The proteins secreted in adipose tissue and their functions are shown in table 1.3-1.

Table 1.3-1: Adipokines and their effects (23, 25, 26).

Function/Effect	Adipocytokines
Cytokines and cytokine-related proteins	Leptin TNFα IL-6
Other immune-related proteins	MCP-1 MIF
Proteins involved in the fibrinolytic system	PAI-1 Tissue factor
Complement and complement-related proteins	Adipsin Complement factor D ASP Adiponectin
Lipids and proteins for lipid metabolism or transport	LPL PPARγ CETP AdipoQ Apo E
Proteins of the RAS	Angiotensinogen
Other proteins	Resistin Monobutyrin Visfatin

1.4 Bottom-Up proteomics

In the field of MS-based proteomics, there are two strategies for profiling the proteome. The two strategies are “Top-Down proteomics” and “Bottom-Up proteomics”. In this thesis, the Bottom-Up proteomics approach is conducted.

Bottom-Up is the commonly used approach for profiling proteins - protein identification, understanding tryptic peptide sequences and PTMs. There are two ways of protein profiling in Bottom-Up: “Sort-then-break” and “break-then-sort”. Briefly explained, the sort-then-break method fractionates proteins on an offline chromatography system prior to proteolytic digestion of proteins into peptides. The process is followed by analysis using peptide mass fingerprint (PMF), electrospray (ESI) or matrix-assisted laser desorption/ionization (MALDI) (27).

The Break-then-sort is the other method that digests a complex mixture of proteins into peptides. The peptides, depending on analysis of interest, can be enriched to increase concentration of phosphopeptide, glycopeptide etc. The sample is injected into the mobile phase in liquid chromatography (LC) and analysis using electrospray ionization mass spectrometry (ESI-MS) or matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) (28, 29). An illustration of the Bottom-up workflow is shown in fig. 1.4-1.

The Bottom-Up approach is mostly suited for proteins over 30 kD, due to the fact that smaller proteins and peptides generate less trustworthy peptides for identifications. The reason for this is that smaller proteins have fewer sites for proteolytic cleavage. Therefore, the sufficient information needed might be lost if this method is used for proteins below 30 kD. The main advantage using this approach is the possibility to achieve high-resolution separations and working with the small amount of sample (30)

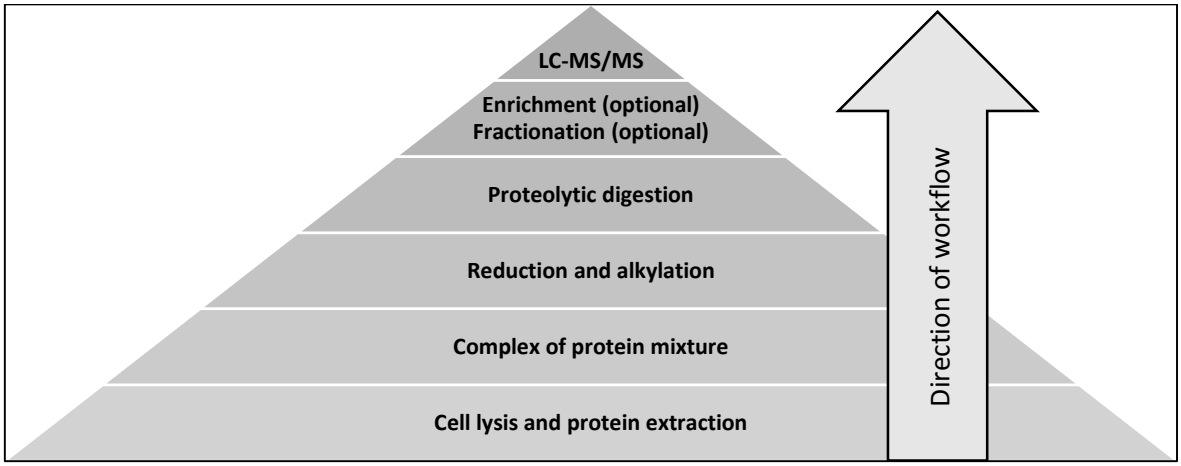


Figure 1.4-1: Workflow in Bottom-Up proteomics.

1.4.1 Enrichment techniques proteomics on a peptide level

Enrichment in proteomics is done to increase the concentration of certain modifications in digested proteins and simultaneously clean samples for impurities prior to analysis in LC-MS. An example of increasing phosphopeptides through enrichment is shown the chart (fig. 1.4-2)

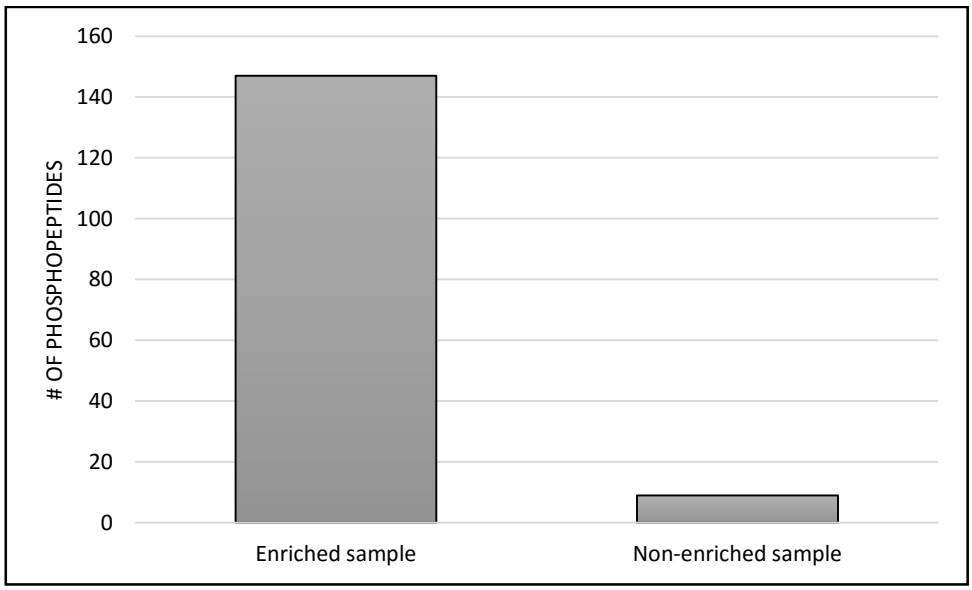


Figure 1.4-2: Phosphopeptides identified - enrichment versus no enrichment. Chart taken form adipose tissue result file.

There are different types of enrichment techniques used on a peptide level. These are illustrated in fig. 1.4-3.

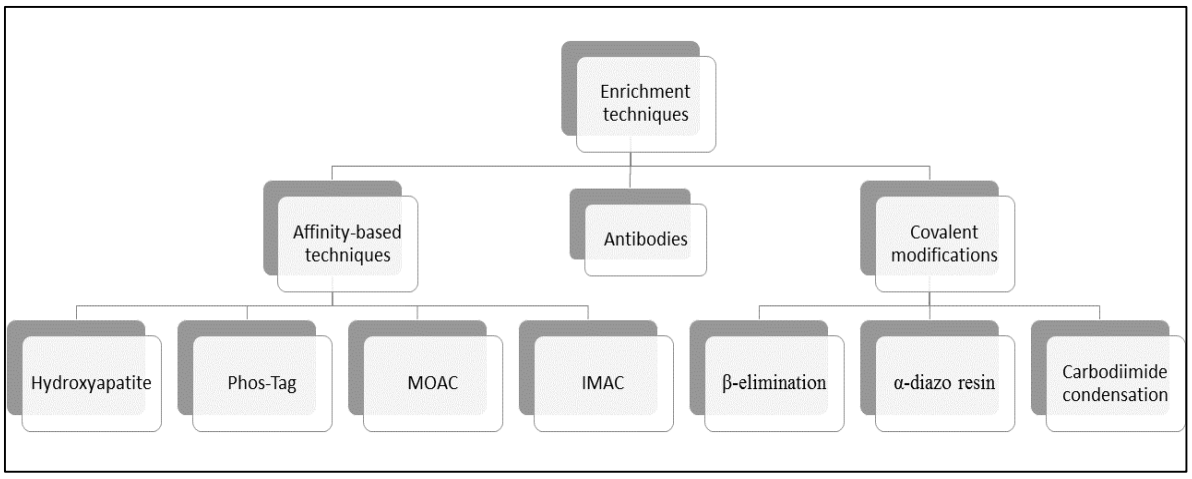


Figure 1.4-3: Enrichment techniques. The enrichment-tree is based on the publication by Fila & Honys (2012).

Immobilized Metal Affinity Chromatography (IMAC) involves interaction between metal-ligand and phosphate-groups on peptides. The metal-binding ligands employed as stationary phase in IMAC are iminodoacetate (IDA) and nitrilotriacetate (NTA) bound to metals like Fe(III), Ga(III), Zr(IV) and Al(III). The most frequently used is Fe(III). Metal Oxide Affinity Chromatography (MOAC) employs titanium dioxide (TiO_2), zirconium dioxide (ZrO_2), aluminum hydroxide ($Al(OH)_3$), alumina and niobium (V) oxide (N_2O_5). The most frequently used in MOAC, also conducted in this thesis, is enrichment with TiO_2. Both IMAC and MOAC share the similarity where the phosphopeptide binds noncovalent to the stationary phase during the enrichment, in order to wash out impurities. The enrichment is followed by elution to yield a higher concentration of phosphopeptides (31).

Enrichment with antibodies involves phosphor-selective antibodies that target phosphopeptides. However, the binding efficiency to phosphoserine and –threonine is not as successful as the binding to phosphotyrosine. Since the antibodies are phospho-amino acid selective, it is necessary with several parallel immunoprecipitation reactions in order to improve the sensitivity of enrichment with antibodies (32).

Covalent condensation enrichment involves chemical modification of peptides. α-diazo resin and carbodiimide condensation modifies the phospho-amino acid. The phospho-amino acid is then covalently bound to a resin prior to elution with trifluoroacetic acid (TFA) to yield the original phosphopeptide. However, the β-elimination removes the phosphor group from serine/threonine and the peptide is coupled to a Michael addition, indicating that peptides with a Michael addition formerly was a phosphopeptide. The cons with covalent condensation is the failure to enrich all phosphor-amino acids equally; phosphoserine/threonine and phosphotyrosine are not successfully enriched at the same time (32).

1.4.2 Quantitative proteomics with tandem mass tags (TMT)

Labeling of proteins or peptides is a method where the relative amount of the proteome is determined. The tags are used to compare the peptide abundancy between the samples in order to understand the global protein expression and protein modifications that are fundamental for biological processes and eventually diseases. The labeling techniques available in quantitative proteomics are metabolic, isotopic and isobaric labeling (33).

Metabolic labeling involves incorporation of isotopes during cell growth, where the medium is filled with isotopic $^{13}C_6$-lysine and $^{13}C_6$-arginine due to the use of trypsin as the proteolytic enzyme. Also ^{15}N isotopes can be used in the medium. The most common approach is stable isotope labeling by amino acids in cell culture (SILAC) (34). The limitation with metabolic labeling is the lack of quantitative analysis for cells that cannot grow in cell culture e.g. human tissue.

However, in isotopic and isobaric labeling, the sample labeling is on peptide level. Isotopic tags are enzymatically incorporated as heavy isotopic atoms into peptides during proteolytic digestion. The commonly used methods are isotope-coded affinity tag (ICAT), ^{18}O labelling, dimethyl labeling and isotope-coded protein labeling (ISCL). The latter labeling method was developed due to the limitations of ICAT, which is dependent on cysteine residues in peptides (5, 35).

The commonly used method in isobaric labeling are isotope tag for relative and absolute quantification (iTRAQ) and tandem mass tags (TMT). The latter method is applied in this thesis. Isobaric, unlike isotopic, is a chemical reaction between the labeling reagent and the sample. The benefit of isobaric labeling is the simultaneous analysis of 4, 6 of 8 samples, referred to as multiplex tagging. The isobaric tags consists of a reporter ion that dissociates in the collision cell to indicate peptides that were successfully tagged, and a mass normalizer to ensure that all tags share an equal amount of mass as fig. 1.4-4 illustrates (5, 35-37).

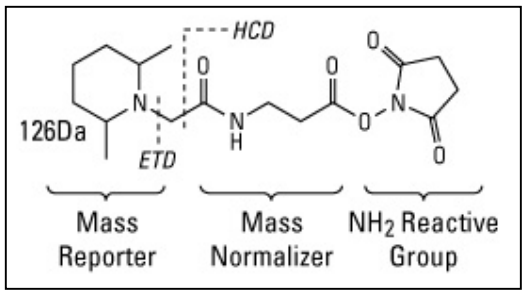


Figure 1.4-4: Scheme of TMTzero. Picture taken from the TMT reagent protocol.

1.4.3 EASY nanoLC

The liquid chromatography method applied in this thesis conducts the same principle as conventional liquid chromatographs. The EASY nanoLC is tailored for proteomics, being a downscaled version of high-pressure liquid chromatography (HPLC) in order to work with limited sample amounts. The principle of liquid chromatography is to separate peptides in a mobile phase and stationary phase as they have different properties, after desalting in the pre-column (38).

The mobile phase consist of water and acetonitrile in two separate containers, where the mobile phase is pumped through the system. This mobile phase is continuously regulated in order to control the retention of the peptides in the column. A scheme of the liquid chromatography in the EASY nanoLC is shown in fig 1.4-5 (38).

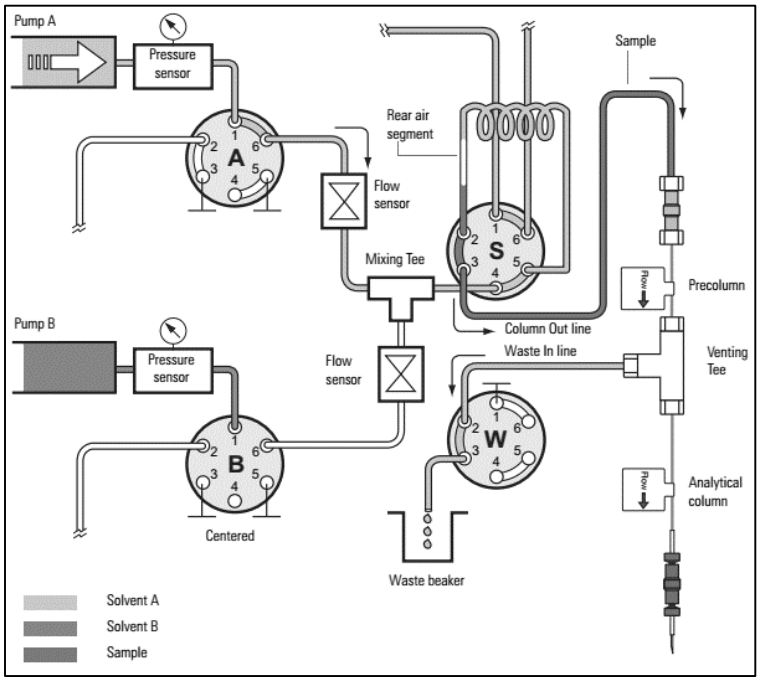


Figure 1.4-5: Scheme of the liquid chromatography process in the EASY nanoLC. Picture taken from EASY nLC operating manual.

This EASY nanoLC uses a two-column stationary phase, where the pre-column traps the analytes to desalt and rapid pre-concentration in order to clean up the volume injected prior to separation of peptides in latter column. The pre-column used is called Acclaim PepMap100 and the column used for ultra-high resolution separation is called Acclaim PepMap RSLC. Both columns are silica-based C_{18} and have an internal diameter of 75 µm. The length of the column may vary as the 5 cm column is suited for simple peptide mixtures and the longer columns are suited for more complex peptide mixtures (39)

1.4.4 Analyzing with LC-ESI/MS

The ion source utilized in this thesis is ESI. The principle behind ESI, as fig 1.4-6 illustrates, is injection of sample by the mobile phase into the MS, where the sample solution passes through a stainless steel needle or a narrow capillary. At the tip of the needle/capillary, a fine aerosol consisting highly charged droplets of mobile phase containing the analyte (approximately 2.5-6.0 kV) is formed. Due to the volatility of the mobile phase, the droplets will continuously reduce in size and simultaneously be flushed away by drying gas. The highly charged analytes are extracted into the MS though the skimmer, lens and poles in order to keep the flow of the charge analyte intact for further analysis (40-42).

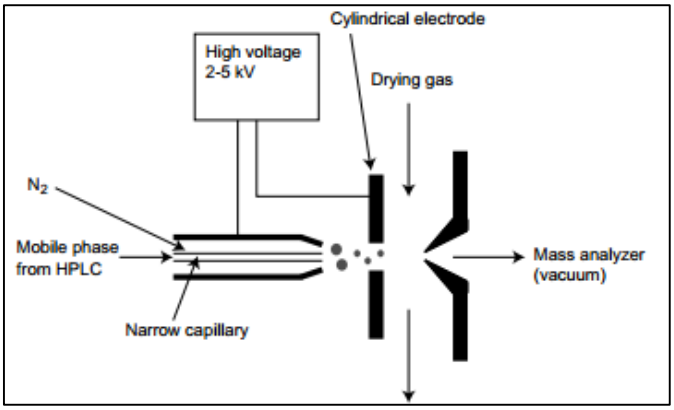


Figure 1.4-6: Scheme of Electrospray Ionization process (Hansen et. al (2011)).

In ESI, the analytes are ionized to molecular ion ($(M+1)^+$ or $(M-1)^-$) which are stable enough to avoid further decompensation into fragment ions. Fragmentation of the molecular ion occurs in the collision cell. It is thus essential that the mobile phase is adjusted in order to keep the analyte charged. It is also important to keep in mind that the mobile phase contains volatile component for efficient evaporation when the charged analytes enters the MS (40).

1.4.5 Quadrupole

The quadrupole is a mass filter based on four symmetrical rods parallel to each other with constant oscillating voltages. This creates an electrical field that lets ionized peptides pass through. The quadrupole can be set to analyze within a specific mass range, e.g. when conducting peptide fragmentation or when conducting targeted analyses. It can also be set to let all peptide ions pass through, known as full scan. In a data dependent analysis (DDA) the quadrupole is shifting between allowing all ions pass through (also called MS1 scans) and isolating selected peptides for fragmentation (called MS2 scans) (40)

1.4.6 Orbitrap mass analyzer

A curved linear trap (C-Trap) has proven to make the Orbitrap a practical mass analyzer. The peptide ions are accumulated in the C-Trap where the kinetic energy is dampened using nitrogen gas as collision gas. The radio frequency is decreased in a brief interval and the direct current potential is increased prior to acceleration into the Orbitrap (43).

The Orbitrap mass analyzer consist of a spindle-shaped central electrode with a bell-shaped electrode surrounding, as depicted in fig. 1.4-7, creating an electrical field to capture and confide peptide ions. The peptide ions cycles around the internal central electrode where the signals are converted to mass spectra by using Fast Fourier Transformation (FFT)(44).

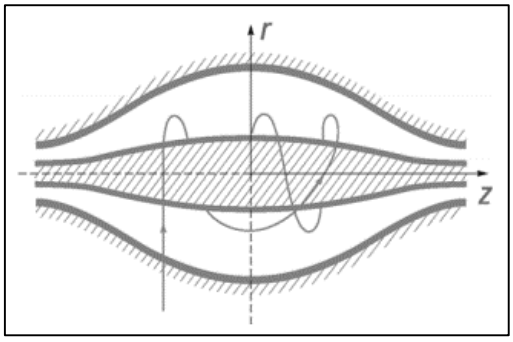


Figure 1.4-7: Orbitrap mass analyzer. Picture taken from Q Exactive operating manual.

1.4.7 Fragmentation of analyte molecules

In order to obtain the MS^2 of the peptides injected, the peptide ions are fragmented in a collision cell. The types of collision cell available for fragmentation are collision-induced dissociation/collisionally-activated dissociation (CID/CAD), higher energy collision dissociation (HCD), electron-capture dissociation (ECD) and electron transfer dissociation (ETD). The collision cell applied in this thesis is CID/CAD.

The structural information about the peptide ions are obtained from CID. The dissociation of peptide ions are induced by high-pressure collision with nitrogen gas. This increases the internal energy of the molecule, which drives the dissociation of the peptide ions into fragment ions (45).

The fragment ions flies back to the C-Trap and on to the Orbitrap, where the frequency of the fragment ions are converted to mass spectra, as explained in section 1.4-6.

The complete scheme of the mass spectrometer Q Exactive is illustrated belown in fig. 1.4-8.

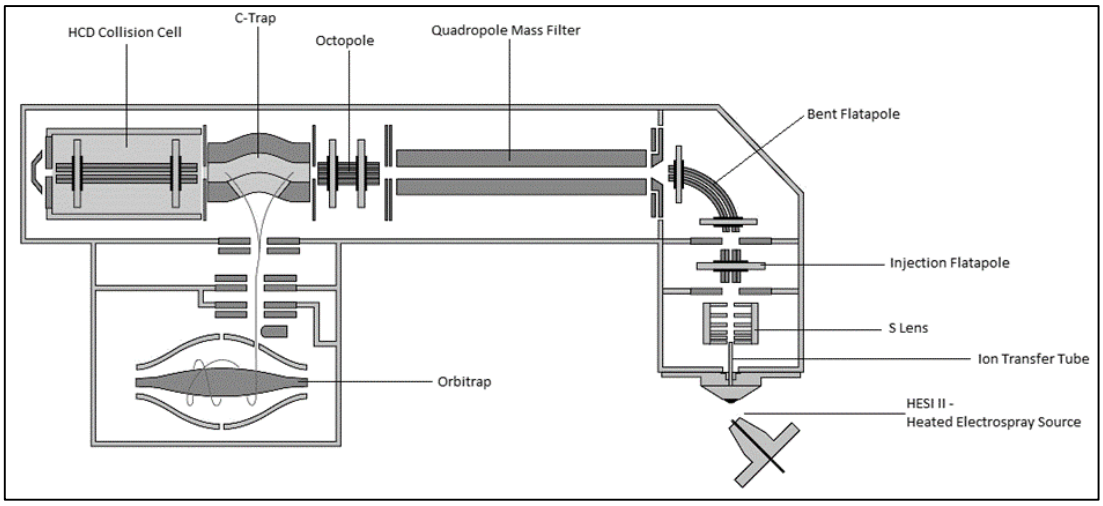


Figure 1.4-8: Detailed scheme of the components in the Q Exactive. Picture taken from Q Exactive operating manual.

1.5 Bio-Informatics

In order to identify peptides and thereby proteins from the raw files, different programs can be used to search protein databases. The following programs used for analyzing the data output from the Q Exactive in this thesis, were Proteome Discoverer, Peaks and PeptideShaker. Each program have different algorithm for analyzing file outputs, yielding slightly different results regarding phosphopeptide.

1.5.1 Proteome Discoverer

Proteome Discoverer (PD) is a program that holds search engines like Mascot and Sequest HT to search sequence databases with raw file outputs from the Q Exactive.

Mascot employs a probability-based algorithm. It scores identification by calculating the probability that an observed match between experimental data and a sequence database entry has occurred by chance. A good identification is an observed match with low probability. However due to inconvenience, Mascot does not base its scoring function on a significance value of 5% ($p<0.05$). The p-value is plotted in the logarithmic function, $-10Log_{10}(p)$ where the highest score is reckoned the best match. The score is typically of 70 or higher (46).

Sequest HT uses a different scoring algorithm. The fast cross correlation (fXCorr) was formerly based on FFT, but due to the expensiveness to compute, the algorithm was simplified. Eng et. al proved that there is no difference between the fXCorr and the FFT-based XCorr (47). From the fXCorr, a histogram of "dot-products" are obtained for each spectrum match. The histogram is then introduced to a linear least-square to determine the

expectation value (e-value). E-value operates equally as a p-value; explains the probability of match occurring at a random event (47).

1.5.2 SearchGUI/PeptideShaker

SearchGUI is a program with several built-in search engines in order to perform simultaneous search against a sequence database. Formerly known as OMSSAGUI until 2010, the name-change to SearchGUI was due to the inclusion of X! Tandem. To date, the interface holds search engines like MyriMatch, MS Amanda, MS-GF+, Comet, Tide and Andromeda. Tide and Andromeda will not be used this thesis to interpret the data. Unlike PD, where the user has to fill in the search parameters separately for each search engine, this is done once in SearchGUI and the parameters will automatically account for all engines (48). Each engine in SearchGUI employs different scoring algorithm to score peptide identified. A simplified workflow for how OMSSA matches experimental peptides to sequence database followed by scoring the peptide match hits is shown in fig. 1.5-1. OMSSA uses a probability based matching, which creates a mass ladder of m/z values in both experimental spectra and sequence database, in order to compare and find possible hits.

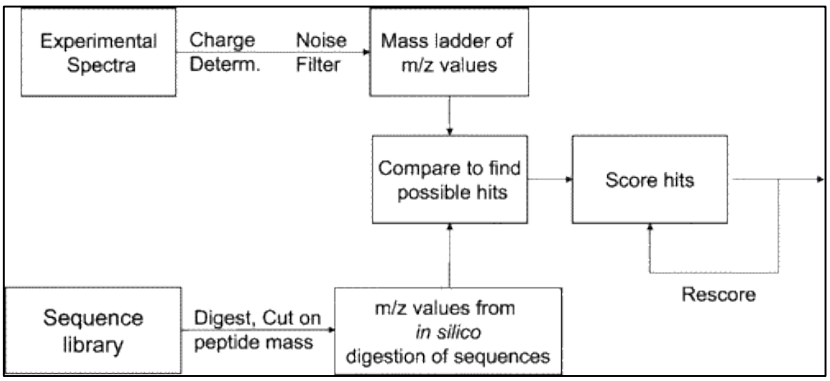


Figure 1.5-1: Peptide scoring worklflow in OMSSA (Geer et. al (2004)).

The possible hits are plotted in a Poisson distribution to calculate each spectrum matched, in order to determine the probability of a hit being a random event. From the Poisson distribution, OMSSA is then able to determine the e-value, which is the "expected number of random hits from a search library to a given spectrum" (49).

X! Tandem (50, 51), MyriMatch (52), Comet (47, 53) and MS Amanda (54) similarly employ the same workflow as OMSSA, but uses different scoring algorithms for generating "dot-products", or histogram of dots, by using different probabilistic distributions in order to calculate an e-value. An overview of the statistical models the search engines uses is

displayed in table 1.5-1. In addition, pre-processing in each engine is different before comparison with a sequence database, e.g. MyriMatch splits peaks in a spectrum into classes of A, B and C in a ratio of 1:2:4 (52)

Table 1.5-1: Statistical models the search engine uses to score identifications

Search algorithm	Statistical model
X! Tandem	Survival function
MyriMatch	Multivariable hypergeometric distribution
Comet	Fast Cross-correlation
MS Amanda	Cumulative binominal distribution

On the other hand, MS-GF+, interprets the experimental data differently. It bases it scoring on the following rules (55):

1. It should be adequate for the great majority of the spectra.
2. The algorithm for peptide spectrum matches (PSM) should be fast.
3. The algorithm for computing statistical significance (E-value) of PSMs should be fast and accurate.

The search engine operates by converting spectra into spectral vectors (S), database-matched peptides into peptide vectors (PV) and generating an MSGFScore of the matching (MSGFScore (PV,S)). The search engine has adopted the "generating function approach" to compute e-values for each peptide based on the PSM score distribution. According to the authors, adopting this approach improves the accuracy of the estimated e-values, thereby increasing the peptides identified. The general workflow applied in MS-GF+ is illustrated in fig. 1.5-2.

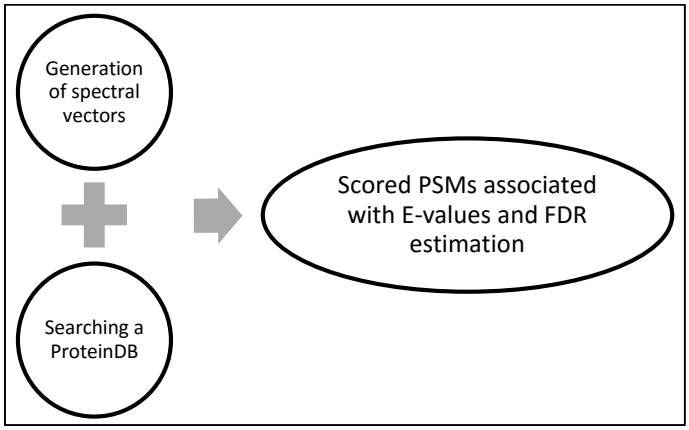


Figure 1.5-2: MS-GF+ workflow.Kim and Pevzner explains the approach MS-GF+ uses to search a protein database by computing suffix arrays in order to compare each peptide against all

spectra with the same precursor mass, unlike other search engines that compares one and one spectrum to all peptides (55).

After processing data with SearchGUI, the output data is post-processed with PeptideShaker. The post-processing is necessary to increase confidence and sensitivity. It does so by searching against a target-decoy for estimation of error occurred while searching through the algorithms, and uses these estimations to merge peptide spectrum matches (PSM) identified by the search engine from SearchGUI (56).

1.5.3 Peaks

Peaks uses an algorithm that differs completely from the ones in PD and SearchGUI. In fact, it does not compare experimental data to a sequence database. By performing de novo sequencing, Peaks computes the best possible sequence among all possible combinations of amino acids for each experimental MS^2. The computed sequences are used to generate a protein shortlist of 7000 by matching them to a sequence database. A shortlist for peptides is created from the protein shortlist, where the 512 top-scored peptides matches are kept for each tandem spectrum. The peptide shortlist is created by matching the peptides from the protein shortlist with tandem mass spectrum and a ”quick scorer” is used to rank PSM. The 512 peptide candidates are scored again by applying a more sophisticated scoring function to generate a p-value. The p-value is then plotted into,$-10Log_{10}(p)$, similarly to Mascot, giving the peptide an identification score (57, 58)

2 Aims of the study

The aim of this thesis is to develop a sample preparation method that is reliable enough to analyze the phosphoproteome of adipose tissue in humans. The enrichment method conducted in this thesis is Titanium dioxide (TiO2) enrichment. The sample preparation validity will be tested using HeLa cells. After validation, the labeling step with TMT reagent will be determined. Finally, search engines will be tested to determine which of these seems most compatible with the data.

This project is part of a comprehensive study regarding the effect of vitamin D on the proteome of adipose tissue**.**

3 Material and Methods

3.1 Material

3.1.1 Chemicals

Table 3.1-1: Chemicals and reagents

Chemical	Catalog #	Manufacturer/Provider
Sodium Deoxycholate (SDC)	D6750-25G	Merck Millipore
Ammonium Bicarbonate (AmBic)	A6141-500G	Sigma Aldrich
Triethylammonium Bicarbonate (TEAB)	17902-100ML	Sigma Aldrich
PhosSTOP cocktail	PHOSS-RO 04906845001	Roche
MilliQ Water	-	Millipore (USA)
Pierce™ BCA Protein Assay Kit	23227/23225	Thermo Fisher Scientific
Dithiothreitol (DTT)	3483-12-3	Sigma Aldrich
Iodoacetic acid (IAA)	I6125-10G	Sigma Aldrich
Calcium chloride (granules)	C1016-500G	Sigma Aldrich
Trypsin	T1426	Sigma Aldrich
Titanium Dioxide (TiO_2)	5020-75000	Titansphere, GL Sciences
Acetonitrile (ACN)	34851-2.5L	Sigma Aldrich
Acetonitrile (anhydrous)	271004-100ML	
Trifluoroacetic acid (TFA)	61020	Riedel-de-Haën
Glycolic acid (GA)	804104	Merck Millipore
Ammonium hydroxide	09861-1L	J.T Chemicals B.V.
Formic acid (FA)	33015-1L	Sigma Aldrich
TMTzero	90067	Thermo Fisher Scientific
Hydroxylamine	55459-50G	Sigma Aldrich

3.1.2 Material

Table 3.1-2: Materials

Material	Description	Manufacturer/Provider
Pipette tip	Tip	Thermo Scientific
GEL-tip		VWR
Micro Pipettor	0.5-2.5 µL 0.5-10 µL 10-100 µL 10-200 µL 100-300 µL (multiple) 100-1000 µL 1000-5000 µL	Eppendorf Research
Tissue Culture plate, 96-well	Micro-well plate	BD Falcon
Vortex	Mixer	Heidolph® Reax Top
Incubator	Oven	Termaks
Biofuge® Fresco	Centrifuge	Heraeus®
Eppendorf Concentrator 5301	Evaporator	Eppendorf
Protein LoBind Eppendorf Safe-Lock Tube Eppendorf	Eppendorf tube	Eppendorf
MagNa Lyser Green Beads	Sample vial	Roche
MagNa Lyser instruments	Cell lyser	Roche
LC-MS vial	Auto sampler	Sun-Sri
EASY nanoLC	Liquid chromatography	Thermo Scientific
Q Exactive™ Hybrid Quadrupole-Orbitrap	Mass Spectrometer	Thermo Scientific

3.2 Sample preparation

3.2.1 Determining the concentration of proteins

Cell lysis

Cell lysate buffer:

- 1% SDC dissolved in 50 mM AmBic
- 1% SDC and phosphatase cocktail inhibitor tablet dissolved in 50 mM AmBic

Lysing the adipose tissue

Adipose tissue was weighed to 50 mg in two separate MagNa Lyser vials. 900 µL of each cell lysate buffer was added to each vial. Samples were lysed by MagNa Lyser at 3425xG for 60 seconds (twice) and cooled down on ice for 30 minutes prior to centrifugation.

Separating lipids from liquid

Cell lysates were centrifuged by Heraeus Biofuge Fresco microliter centrifuge for separation of lipids from liquid at 16060xG (4° C) for 15 minutes. Subnatant was carefully transferred to a LoBind Protein eppendorf tube with a needle mounted on a syringe.

BCA Assay

Preparation of the BCA Assay

The determination of protein concentration is required for the latter stages in the experiment in order to use the right volumes of reagents when digesting proteins. The "microplate procedure" in bicinchoninic acid assay (BCA Assay) protocol published by Thermo scientific (Pierce BCA Protein Assay) was conducted for determining protein concentration (59). The suggested dilutions of samples for the BCA assay is shown in table 3.2-1.

Table 3.2-1: Dilutions for the BCA Assay for the micro-column experiment

Dilution step	Sample Amount	Source	Diluent (MilliQ water) amount	Dilution
1 PS	10 µL	PS Protein sample	140 µL	1:15
1 NPS	10 µL	NPS Protein sample	140 µL	1:15
2 PS	10 µL	PS Protein sample	490 µL	1:50
2 NPS	10 µL	NPS Protein sample	490 µL	1:50
3 PS	10 µL	PS Protein sample	990 µL	1:100
3 NPS	10 µL	NPS Protein sample	990 µL	1:100

The standards had dilutions ranging from 0-250 µg/mL. 25 µL of dilutions and BCA standards A-F were pipetted into a 96-well plate. 200 µL of working reagent (WR, Reagent A: Reagent B ratio of 50:1) was pipetted into each well containing diluted sample and standards, as shown in fig. 3.2-1.

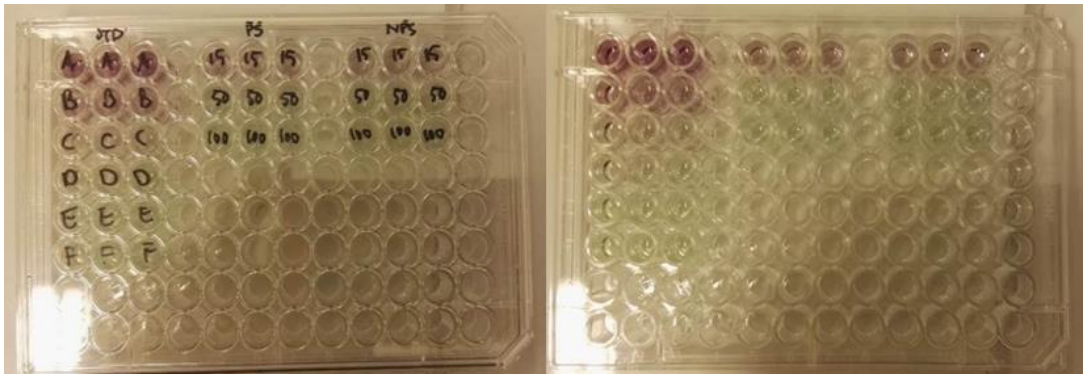


Figure 3.2-1: Micro well plates filled standards (A-F) and dilutions of PS and NPS (1:15, 1:50, 1:100)

Determining the protein concentration

The microplate was incubated at 37° C for 30 minutes, and absorbance (A) was measured by SpectraMax 190 Microplate Reader (A562nm). The results were exported to Microsoft Excel for determination of protein concentration, based on linear regression performed on standards A-F.

3.2.2 Protein digestion

Reduction and carbamidomethylation

Reduction buffer: 200 mM DTT stock

Alkylation buffer: 200 mM IAA stock

Reduction

The desired amount of protein conducted was calculated to approximately 500 µg, based on BCA results. The SDC concentration in samples were reduced to 0.5% by adding 50 mM AmBic. Dithiothreitol (DTT) is a reduction agent used for breaking cysteine bridges in proteins to form free cysteine residues. After reduction, it is necessary to carbamidomethylate free cysteine residues in samples to prevent reformation of cysteine bridges. The alkylation buffer used was iodoacetic acid (IAA).

DTT stock solution was added to samples to give a final concentration of 5 mM. The samples were incubated at 56° C for 30 minutes in an oven (Termaks incubator) with gentle shaking.

IAA addition

Fresh 200 mM IAA stock was prepared for the alkylation step and added to a final concentration of 15 mM followed by incubation at RT for 30 minutes protected from light. Additional DTT was added to a final concentration of 5 mM to react with free IAA in samples and yielded the following reaction mechanism:

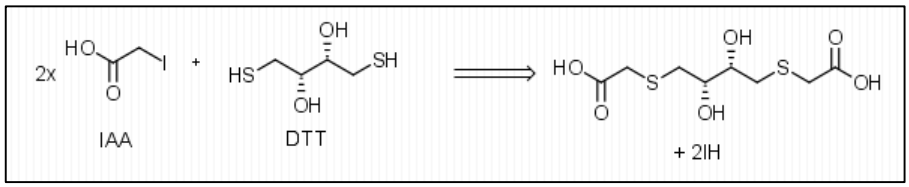


Figure 3.2-2: Reaction scheme - DTT reacting with free IAA

Proteolytic digestion

$CaCl_2$ addition

Digestion buffer: 10 mM $CaCl_2$ stock

$CaCl_2$ increases the stability of trypsin, thereby reducing autolysis which can falsify the results.(60).

10 mM $CaCl_2$ was added to a final concentration of 1 mM.

Trypsinization

Proteolytic enzyme: 2 µg/µl Trypsin

Stock trypsin was made with a concentration of 2 µg/µL. Trypsin was added to samples according to the protein concentration to give a ratio of 1:40 (w/w). Samples were carefully mixed with a pipette prior to incubation over night (16-18 hours) at 37°C in an oven (Termaks incubator).

3.3 Batch-mode experiment

Sample preparation from cell lysis to tryptic digest were conducted as described in section 3.2.1 and 3.2.2.

Digestion stop and SDC precipitation: 10 % TFA stock

SDC *removal*

10% TFA was added to a final concentration of 1% to quench digestion and precipitate SDC from samples. Samples were centrifuged at 4° C, with 16060xG (13000 rpm) for 15 minutes.

After centrifugation, the supernatant was transferred to new vials. The old vial was centrifuged again at same settings and the supernatant was added to the new vial.

At this stage of the experiment, it was decided that part of the sample would be run as it was on the LC-MS system, while the other part would be enriched.

The samples were thereafter dried using a speed-vac (Eppendorf Concentrator 5301)

Preparation of enrichment buffers

Table 3.3-1: Buffers for the TiO_2 enrichment

Buffer	Components	Volume of components
Loading buffer	ACN 99.8% Glycolic acid (GA) 1M MilliQ Water TFA 100%	800 µL 100 µL from 10M stock 50 µL 50 µL
Washing buffer 1	ACN 99.8% MilliQ Water TFA 100%	800 µL 190 µL 10 µL
Washing buffer 2	ACN 99.8% MilliQ Water TFA 100%	100 µL 899 µL 1 µL
Elution buffer	Ammonium Hydroxide 26.5%, pH 11.3 MilliQ Water	40µL 960µL
Elution buffer for filter (Micro column)	ACN 99.8% MilliQ Water	300 µL 700 µL
Acidification	TFA 10% Formic acid (FA) 100%	20 µL 16 µL

The phosphopeptide enrichment method used in this experiment was based on the publication by Dickhut et al.(61), but some alterations were made in order to optimize the protocol for this experiment:

- Washed with 200 µL instead of 100 µL with both washing buffer 1 and washing buffer 2.
- Eluted with 200 µL instead of 100 µL, and 60 µL instead of 30 µL the second time.
- Acidified with 20 µL 10% TFA and 16 µL 100% FA

In brief, TiO_2 was mixed with the loading buffer. The tip of the pipette was cut off in order to transfer the beads from the TiO_2-loading buffer mixture to dried samples. Additional loading buffer was added to fill the remaining volume, which was in total supposed to be 1000 µL. The bead-to-peptide ratio worked with was 6:1, 3;1 and 1.5:1 (w/w) (fig. 3.3-1). The bead-to-peptide ratio 6:1 was incubated for 10 minutes followed by centrifuging at 3000xG to pool the beads. The supernatant was enriched again

with bead-to-peptide ratio 3:1 and then 1.5:1. The beads were collected in one tube for washing and elution of phosphopeptides. After elution, samples were acidified prior to desalting (61)

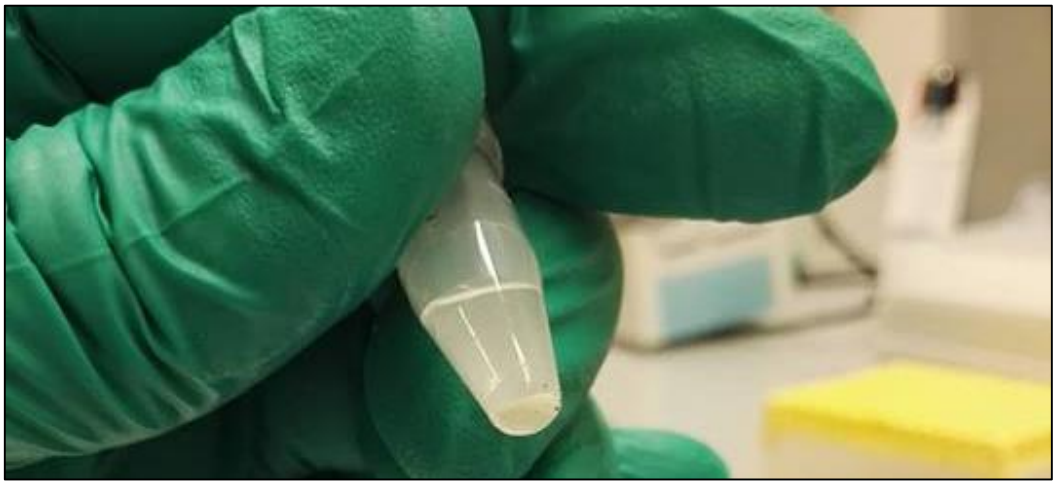


Figure 3.3-1: Incubation of sample in TiO beads

3.4 Micro-column experiment

The sample preparation from cell lysis to tryptic digest was conducted as described in section 3.2.1 and 3.2.2.

The phosphopeptide enrichment method used in this experiment was based on the publication by Dickhut et al.(61), but some alterations were made to optimize the protocol for this experiment:

- Washed with 200 µL instead of 100 µL with both washing buffer 1 and washing buffer 2.
- Eluted with 200 µL instead of 100 µL and 60 µL instead of 30 µL the second time.
- The filter was washed with 6 µL 30% ACN instead of 2 µL
- Acidified with 20 µL 10% TFA and 16 µL 100% FA

Samples were loaded on a micro-column. The micro-column was made by cutting the tip of a pipette that can hold 20-200 µL liquid, and using this cut-off pipette tip to carve out a piece of micro-glass fiber and load it in a new pipette tip (20-200 µL). A GEL-tip was used to push the glass fiber downwards the pipette making it work as a filter (fig. 3.4-1). Sample were loaded on the micro-column, centrifuged at 3000xG. The steps for incubating the bead –to-peptide ratios were conducted as described in section 3.3. After filtering the sample volume, the beads

where washed with loading buffer and washing buffer prior to elution of samples from beads and filter. The phosphopeptides were acidified prior to desalting. (61)

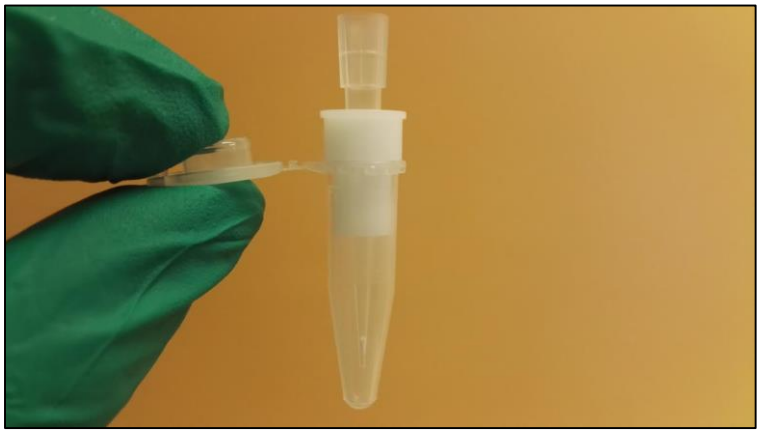


Figure 3.4-1: Sample loaded on micro column

3.5 TMTzero label (TMT^0)

Samples were prepared with 100 mM TEAB in this experiment, due to TMT reagents incompatibility with AmBic. The starting concentration of protein conducted was as described in section 3.2.1, but was doubled to 1000 µg. The samples were split into two after proteolytic digestion.

In brief, ACN (anhydrous) was pipetted into the TMT_0 vial. The vial was occasionally vortexed for five minutes, and then centrifuged at 2000xG. The ACN-TMT_0 mixture was added to samples and incubated for an hour. 5% hydroxylamine was added and incubated for 15 minutes to quench the reaction (62).

3.6 Mass spectrometry

Solid Phase Extraction desalting

The composition of the buffers for desalting is shown in table 3.5-1

Table 3.6-1: Desalting Buffers

Buffer	Components	Volume of components
Conditioning buffer	ACN ≥99.8% MilliQ Water	500µL 500µL
Washing buffer 1	TFA 10% MilliQ Water	100µL 900µL
Washing buffer 2	TFA 10% MilliQ Water	10µL 990µL
Elution buffer	ACN ≥99.8% Formic Acid 100% MilliQ Water	750µL 10µL 240µL

The OMIX C_{18} Protocol was conducted for desalting the samples. The substitution of Heptafluorobutyric acid (HFBA) to TFA is the only alteration made.

In brief, the C_{18}-tip was conditioned with conditioning buffer, and discarded with pipette plunger depressed. The tip was washed with washing buffer 1 and then discarded with the pipette plunger depressed. The washing step was done in two cycles. Sample were aspirated and dispensed in approximately 15 cycles and discarded with the plunger depressed. The tip was rinsed with washing buffer 2 in four cycles. Peptides were eluted from the tip with elution buffer. Care was taken to prevent air getting in the OMIX tip at any time during the procedure. Samples were dried in speed-vac (Eppendorf Concentrator 5301) and re-suspended in 0.1% TFA (63).

Determination of injection volume

The concentration of peptide injected was 0.5 µg, but was increased to 1 µg in the latter experiments. The corresponding in µL was determined using Thermo Scientific NanoDrop 1000 Spectrophotometer. The menu on the application chosen to measure the absorbance (A) was "Protein A280" (A205nm)

3.6.1 Liquid chromatography parameters

Table 2.5-2 shows the mobile phase composition used in the EASY nanoLC to gradient elute the analytes in the sample. The type of column used are silica-based C_{18} with an internal diameter of 75 µm µm and a length of 2 cm (Acclaim PepMap 100) and 50 cm (PepMap RSLC)

Mobile phase consists of:

A: 0.1% formic acid in MilliQ water

B: 0.1% formic acid in ACN

Table 3.6-2: Gradient elution of analyte

Step	Time (min)	Flow (nL/min)	%A	%B
1	Initial	200	2	98
2	5	200	5	95
3	65	200	40	60
4	71	200	100	0
5	76	200	100	0

3.6.2 Mass spectrometer parameters

The mass spectrometer used to analyze samples was Q Exactive by Thermo Fisher scientific, with ESI as ion source and CID as collision method. The scan parameters used to analyze the analyte is shown in table 3.5-3

Table 3.6-3: Q Exactive scan analysis parameters

Full MS	**Labeled sample**	**Unlabeled sample**
Microscans	1	1
Resolution	70000	70000
AGC target	$3x10^6$	$3x10^6$
Maximum Inject Time	120 ms	120 ms
Number of scan ranges	1	1
Scan range	400 to 2000 m/z	400 to 2000 m/z
dd-MS/MS		
Microscans	1	1
Resolution	17500	17500
AGC target	$5x10^4$	$5x10^4$
Maximum Inject Time	250 ms	250 ms
Loop count	15	15
Isolation window	2.0 m/z	2.0 m/z
Fixed first mass	125 m/z	-
NCE/Stepped NCE	27	27
dd-settings		
Underfill ratio	1.0 %	1.0 %
Intensity threshold	$2.0x10^3$	$2.0x10^3$
Charge exclusion	Unassigned 1^+, 6^+-8^+, $>8^+$	Unassigned 1^+, 6^+-8^+, $>8^+$
Dynamic exclusion	12.0 s	12.0 s

3.6.3 Database Search

Search parameters

The raw files from the Q Exactive were processed using PD, PeptideShaker and Peaks. The search parameters used are listed in table 3.5-4. Regarding SearchGUI, the raw files were converted to mgf files with MSConvert before searching protein database. The workflow is illustrated in fig. 3.5-1.

Table 3.6-4: Search parameters for the search algorithms

Spectrum Matching	Fixed modification	Carbamidomethylation (C) TMT (This was only used when there was TMT in the sample)
	Variable modification	Phosphorylation (STY) Oxidation (M)
Spectrum Properties filter	Ion Types	Immonium ions Reporter ions b-ions y-ions Precursor ions Related ions
	Neutral Losses	H_2O NH_3 H_3PO_4 HPO_3 CH_4OS
	Minimum precursor mass Maximum precursor mass	350 Da 9000 Da
Tolerances	Precursor Mass Tolerance Fragment Mass Tolerance	10 ppm 0.02 Da
Input data	Maximum missed cleavages Peptide length	4 8-30
Decoy database search	Target FDR (strict) Target FDR (relaxed)	0.01 0.05

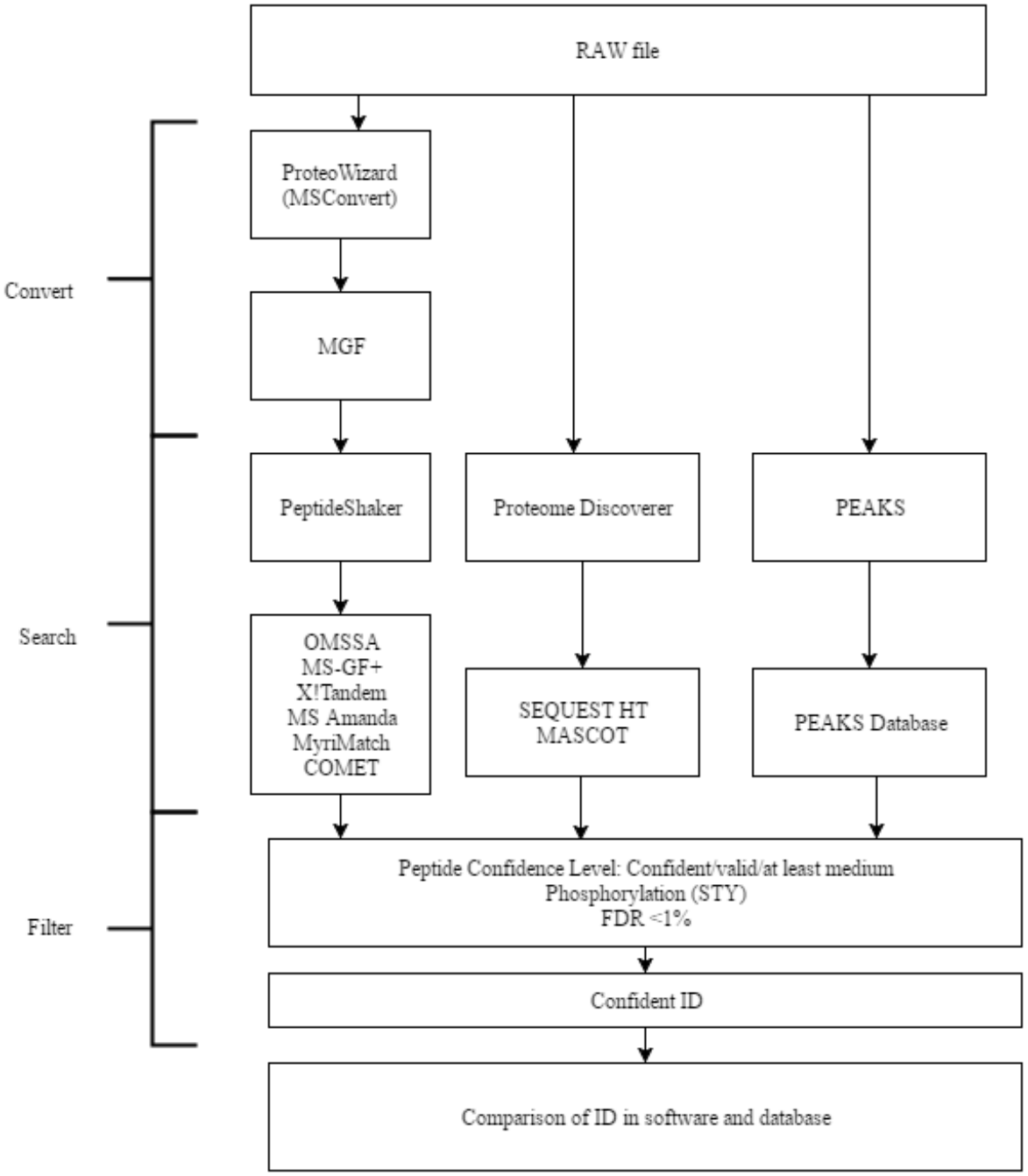


Figure 3.6-1 Bio-informatics workflow

4 Result and discussion

4.1 Choice of search engine for quantitative analysis

4.1.1 Number of identifications in each search engine

Proteome Discover 2.1 (PD), Peaks and PeptideShaker searched through the raw files from the Q Exactive. The search engines were compared to each other in order to determine which search engine that was suitable to conduct for interpreting data. The search engine choice was based on highest number of correct and definite identified phosphopeptides as well as designation of the fragment spectra. The chart shows the number of phosphopeptides identified in each search engine. Mascot (252) identified the highest number of phosphopeptides, followed by Peaks (239), PeptideShaker (197) and Sequest HT (146) (fig. 4.1-1).

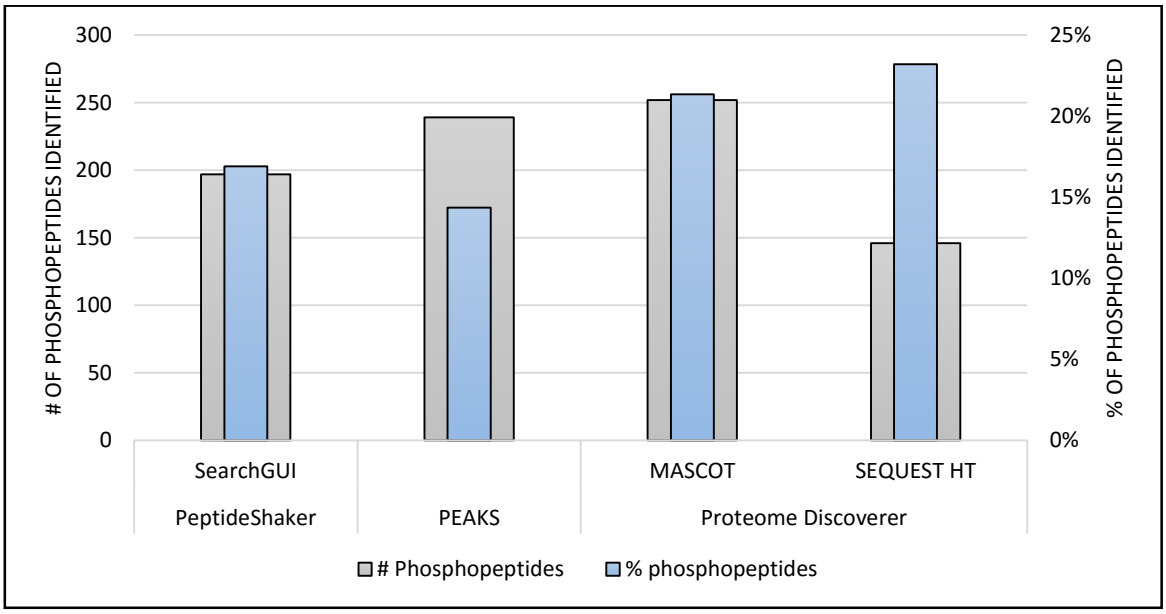


Figure 4.1-1: Comparison of identification algorithm using "Micro Column without PhosSTOP" sample as an exampleThe percentage (blue bars) explains the number of phosphopeptides among peptides confidently identified. Sequest HT identified the fewest number of phosphopeptides (grey bars) of the search engines, but simultaneously 23% of the peptides identified were phosphopeptides, which showed to be the highest unlike other search engines used. Peaks identified 239 phosphopeptides, which was calculated to 14% of peptides. Mascot showed to be remarkable with 252 phosphopeptides (21% of peptides)

Mascot and Sequest HT searched the data in PD, where Mascot proved to be significantly better in identification. As previously mentioned, both search engines conducted different algorithms in order to score peptides. Neutral losses may cause problems for the search algorithms to properly score peptides, which could possibly explain the fewer phosphopeptides identified by Sequest HT. The collision method that CID performed may encounter challenges while fragmenting phosphopeptides with multiple phosphor-group.

Details regarding the mechanism of CID is presented in section 4.7 together with the reason this type of fragmentation may cause problems for the scoring function of the search engines

The sequences identified in Mascot and Sequest HT were further examined, and results were filtered by altering the engines settings' confidence to 'at least medium' and modifications equal to 'phospho'. The confident sequences were obtained and plotted in a Venn diagram (fig. 4.1-2)

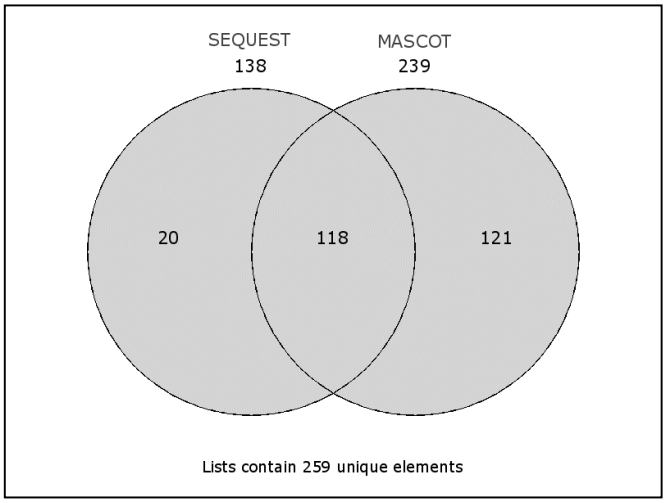


Figure 4.1-2: Overlap of phosphopeptides identified by Sequest HT and Mascot for human adipose tissue. Mascot had in total 239 unique identifications, whereas Sequest HT had 138 unique identifications.

118 phosphopeptides were common in both search engines, whereas Mascot was unique in 121 phosphopeptides compared to the 20 identified by Sequest HT. The supporting table (table 4.1-1) displays the total quantity of phosphopeptides each search engine identified as well as the number of unique phosphopeptides.

Table 4.1-1: Phosphopeptides identified in Mascot and Sequest HT

List names	Number of phosphopeptides	Number of unique phosphopeptides
Mascot (Micro column without PhosSTOP)	252	239
Sequest HT (Micro column without PhosSTOP)	146	138
Overall number of unique elements		259

Mascot had a 42% higher number of phosphopeptides in total and unique phosphopeptide identification than Sequest HT. Based on these statistics, Mascot was chosen for the data interpretation output in PD.

The identifications from Peaks and PeptideShaker were also examined by plotting the sequences in a Venn diagram in addition to the identifications from Mascot. All

identifications were assumed confident in Peaks as this engine did not have any classification options. Whereas in the PeptideShaker, the confident phosphopeptides were classified as either ‘valid’ or ‘invalid’.

It was previously shown (fig.4.1-1) that Mascot identified the most phosphopeptide, and that the new plot revealed that Mascot still had higher unique identifications than Peaks and PeptideShaker (fig. 4.1-3).

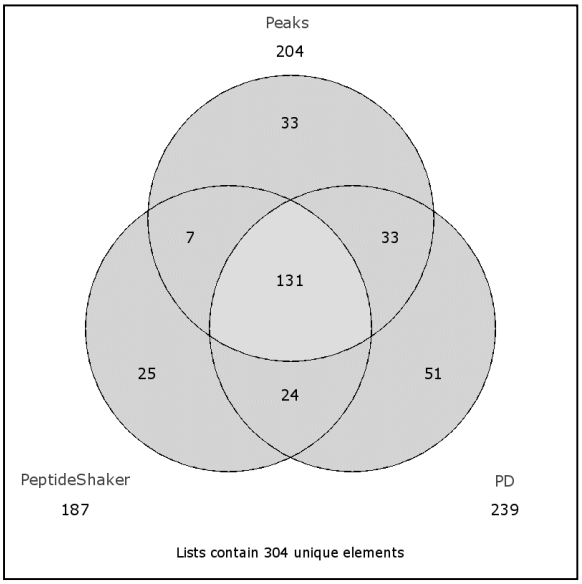


Figure 4.1-3: Overlap of phosphopeptides identified by Peaks, PeptideShaker and Mascot All search engines commonly identified 131 phosphopeptides. Mascot managed to identify 51 phosphopeptides that were not present in Peaks or PeptideShaker. However, Peaks and PeptideShaker found phosphopeptides that were not present in Mascot, but the numbers were lower.

Mascot and Peaks scores identification by generating a p-value that is plotted into the aforementioned logarithmic formula, whereas the search engines in SearchGUI generates ‘dot-products’ that is introduced to linear least square to determine the e-value. This thesis does not look in to the scoring function, despite the fact that the search engines generating p-values seemed to be more suited for identifying phosphopeptides. However, this is a preliminary association and needs to be investigated further by larger sample sets.

In the experiments conducted towards the end of the project, both Peaks and PeptideShaker managed to identify a greater number of phosphopeptide than Mascot, while Sequest HT still identified the lowest number of phosphopeptides. An issue encountered with Peaks is the filtering options when addressing missed cleavages as there is no possibility to attain data regarding missed cleavage from phosphopeptides or alone. The Peaks group were notified about the issue and replied; “*there currently is no way to automatically extract peptides containing a missed cleavage*”.

In SearchGUI, the samples were searched simultaneously, and post-processed by PeptideShaker. The built-in data filter in PeptideShaker exported data to excel, where the user could obtain the needed information. It was possible to create a set of filter the user wish to include. In numerous attempts when including 'spectrum file', which showed the exact sample the phosphopeptide originated from, it seemed impossible to find the information linking the sample name and peptide.

4.1.2 Search engine interpretation of fragment spectra.

Below, images of fragments spectra of the phosphopeptide sequences 'EDEIpSPPPPNPVVK', 'IGHHpSTDDSSAYR', 'LSPpSPTSQR', 'HpTFMGVVSLGSPpSGEVSHPR' and 'SPPpSPVER' are presented to show how the search engines interprets the data and the information given about the phosphopeptide sequence fragments.

The aforementioned search engines were able to designate the fragment spectra with a relatively high accuracy. PD and PeptideShaker have a complex set of annotated fragments with multiple charged ions and neutral losses. Despite the complexity, fragments in PD and their corresponding annotations appeared quite clearly; whereas PeptideShaker overlaps are present making manual interpretation challenging. Decreasing the annotation level gave a clearer view, while simultaneously omitting less intense peaks. This was not observed in all fragment spectra, but predominantly in sequences where all b- and y-ions were identified, e.g. 'IGHHpSTDDSSAYR' and 'HpTFMGVVSLGSPpSGEVSHPR'.

In each spectra, the sequence with its corresponding fragment peak has been marked with a corner bracket (brackets pointing upwards denote y-ions while brackets pointing downwards denote b-ions). Highlighted corner brackets in red indicates fragment(s) one search engine identified, while being missed by the other.

One particular peptide that differed in PD and Peaks from PeptideShaker, in terms of phosphorylation site, was 'HpTFMGVVSLGSPpSGEVSHPR'. PD and Peaks determined the site to be at S13, while PeptideShaker determined T2 as the phosphorylation site. Neither PD nor Peaks recognized T2 as phosphorylation site. However, "Spectrum ID" tab in PeptideShaker, the sequence with phosphorylation at S13 was found although omitted by PeptideShaker (fig. 4.1-4). This was due to loss of H_2O versus H_3PO_4 of same mass fragment. Yet all three search engine registered the same retention time at 40.38 minutes, thus indicating the same peptide.

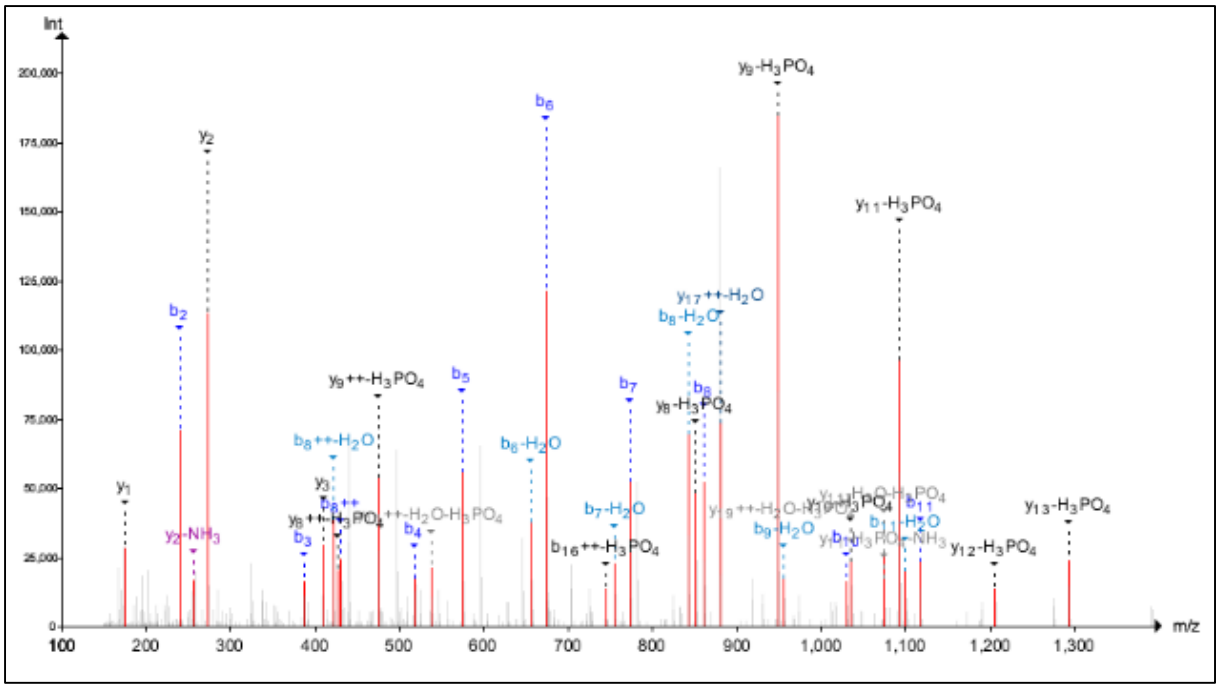


Figure 4.1-4: Spectrum of HTFMGVVSLGSPpSGEVSHPRThe peptide was identified by X! Tandem, MS-GF+ and MyriMatch. The spectrum match was, despite 100% confident, ranked as 1, while the same peptide with phosphorylation site at T2 with 74% (Comet) and 97% (MS-GF+) confidence was ranked as 5 and retained by PeptideShaker.

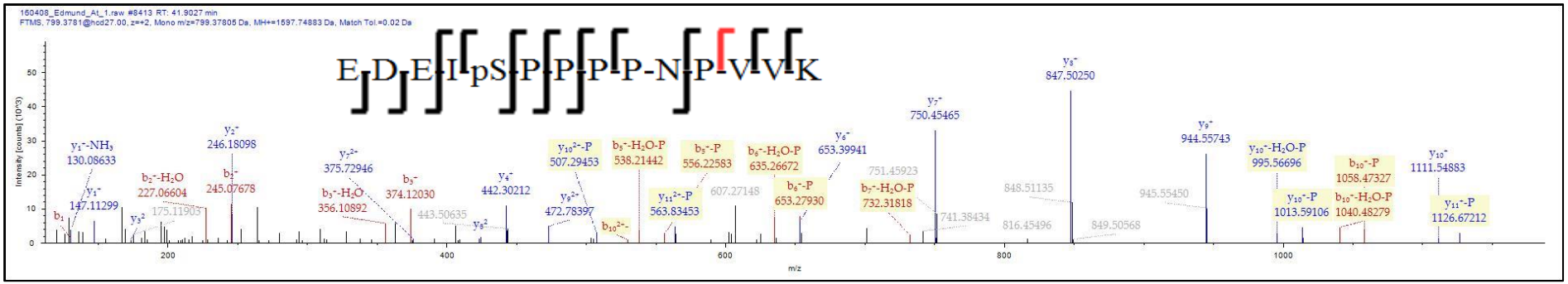
Peaks was able to identify all fragments in 'HpTFMGVVSLGSPpSGEVSHPR', but failed to identify y_{15}, y_{17}, y_{19} (found in PD) and y_{15}, y_{16}, y_{19} (found in PeptideShaker).

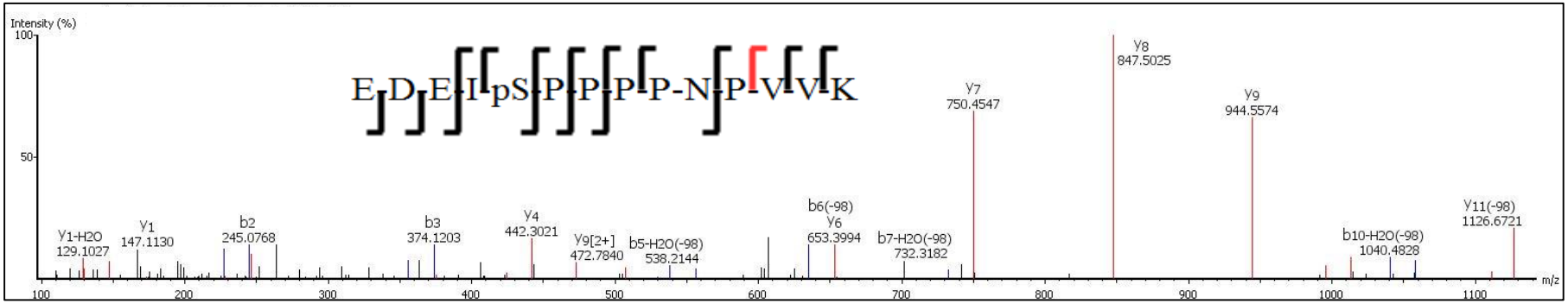
In the 'LSPpSPTSQR' fragment spectra, the search engines successfully identified all y-ion, though only Peaks managed to measure the intensity of the b_5-ion in addition b_2. This fragment is not visible in the spectra, due to a low intensity at 0.5%.

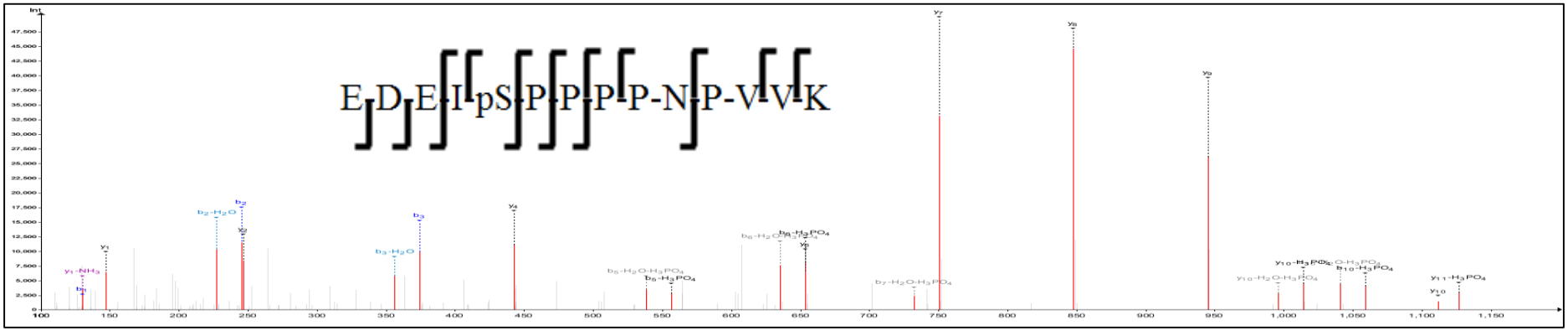
What was noticeable in PeptideShaker was the uncertainty identifying phosphorylation sites. PeptideShaker highlighted potential phosphorylation site in peptides denoted in red. Of all the illustrated peptides, 'EDEIpSPPPPNPVVK' was the only peptide characterized with confident phosphorylation site. The peptide sequence 'SPPpSPVER' was not a valid identification in PeptideShaker.

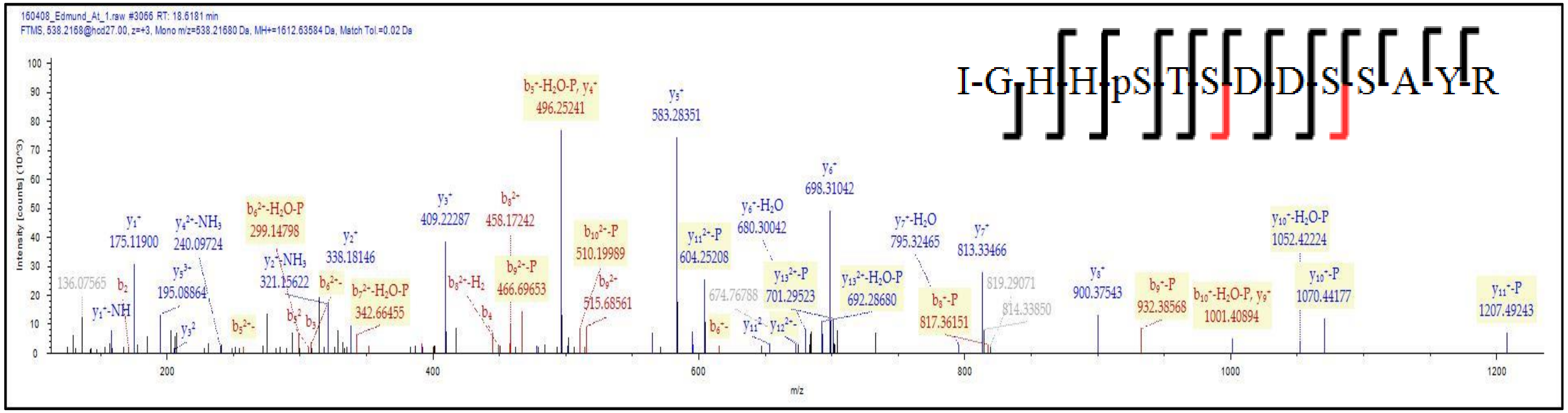
These findings suggest that each individual search engine interprets phosphopeptide fragment spectra differently, which could affect the scoring function.

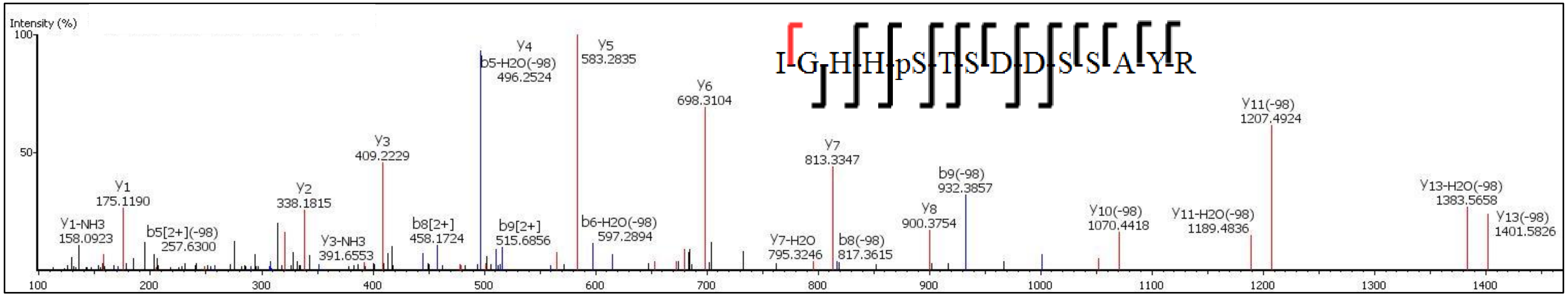
Based on number of total and unique peptides identified in each search engine, user-friendly settings, and the ability to obtain good fragment spectra from phosphopeptides Mascot was chosen as the default search engine for data interpretation.

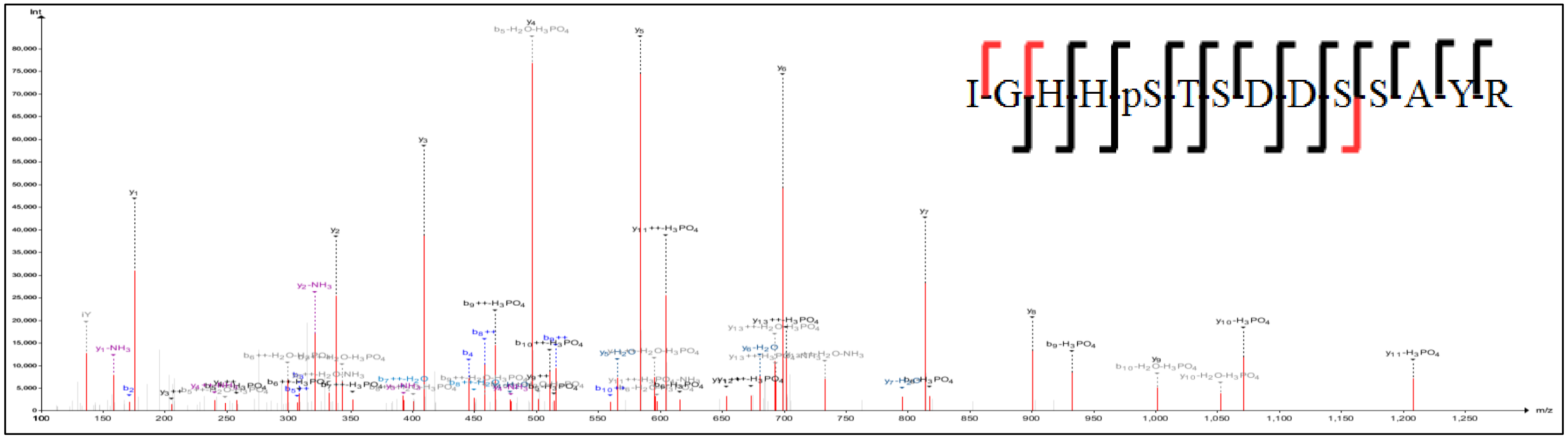

160408_Edmund_At_1.raw #8413 RT: 41.9027 min
FTMS, 799.3781@hcd27.00, z=+2, Mono m/z=799.37805 Da, MH+=1597.74883 Da, Match Tol.=0.02 Da
E D E I pS P P P P-N P-V V K
Intensity [counts] (10^3)
m/z
y1-NH3 130.08633
y2+ 246.18098
b2-H2O 227.06604
b2+ 245.07678
y1+ 147.11299
y3+2 375.72946
b3-H2O 356.10892
b3+ 374.12030
y4+ 442.30212
y9+2 472.78397
y10+2-P 507.29453
b5+-H2O-P 538.21442
y11+2-P 563.83453
b5+-P 556.22583
b6+-H2O-P 635.26672
b6+-P 653.27930
y6+ 653.39941
b7+-H2O-P 732.31818
y7+ 750.45465
y8+ 847.50250
y9+ 944.55743
y10+-H2O-P 995.56696
y10+-P 1013.59106
b10+-P 1058.47327
b10+-H2O-P 1040.48279
y10+ 1111.54883
y11+-P 1126.67212

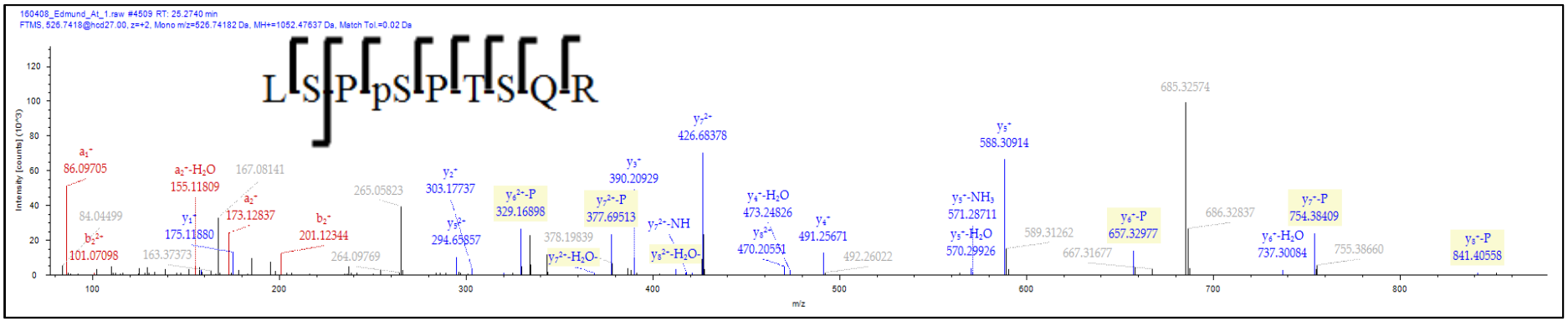

Intensity (%)
E D E I pS P P P P-N P-V V K
y8 847.5025
y7 750.4547
y9 944.5574
y1-H2O 129.1027
y1 147.1130
b2 245.0768
b3 374.1203
y4 442.3021
y9[2+] 472.7840
b5-H2O(-98) 538.2144
b6(-98) y6 653.3994
b7-H2O(-98) 732.3182
b10-H2O(-98) 1040.4828
y11(-98) 1126.6721
m/z

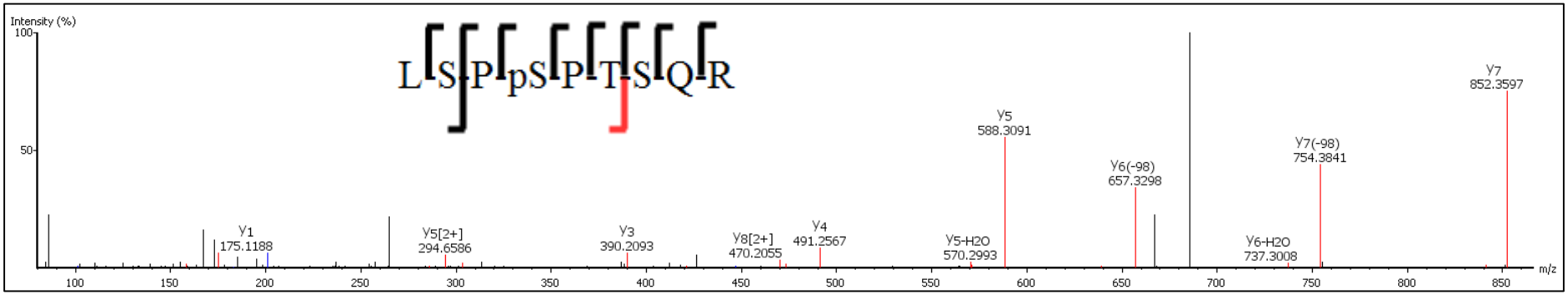

E D E I pS P P P P-N P-V V K

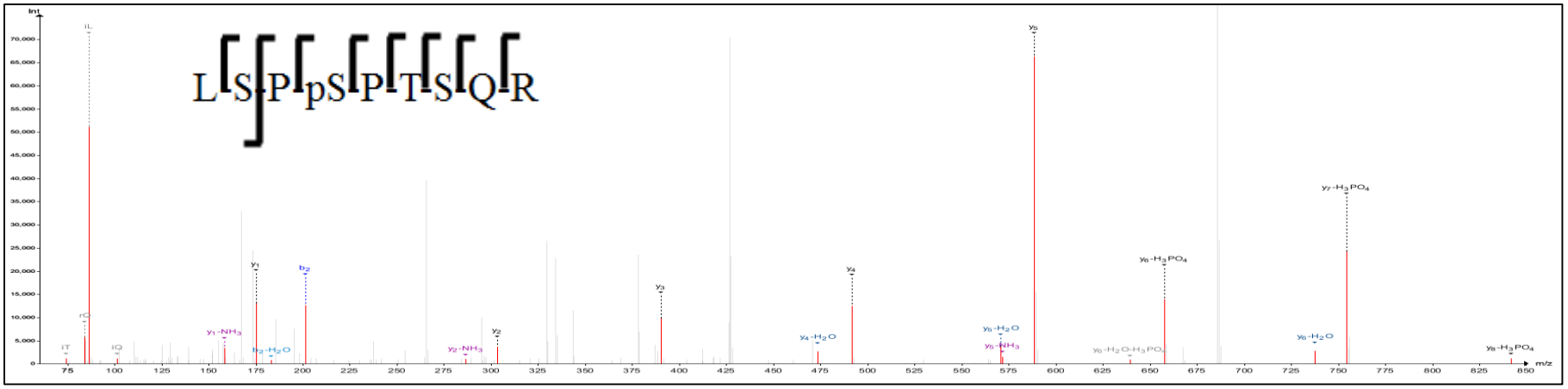

160408_Edmund_At_1.raw #3066 RT: 18.6181 min
FTMS, 538.2168@hcd27.00, z=+3, Mono m/z=538.21680 Da, MH+=1612.63584 Da, Match Tol.=0.02 Da
I-G-H-H-pS-T-S-D-D-S-S-A-Y-R
Intensity [counts] (10^3)
m/z
y1+ 175.11900
y4 2+-NH3 240.09724
y3 3+ 195.08864
b6 2+-H2O-P 299.14798
y2+-NH3 321.15622
y2+ 338.18146
b7 2+-H2O-P 342.66455
y3+ 409.22287
b8 2+ 458.17242
b8 2+-H2
b9 2+-P 466.69653
b5+-H2O-P, y4+ 496.25241
b10 2+-P 510.19989
b9 2+ 515.68561
y5+ 583.28351
y11 2+-P 604.25208
y6+-H2O 680.30042
y13 2+-P 701.29523
y6+ 698.31042
y13 2+-H2O-P 692.28680
y7+-H2O 795.32465
y7+ 813.33466
b8+-P 817.36151
819.29071
814.33850
y8+ 900.37543
b9+-P 932.38568
b10+-H2O-P, y9+ 1001.40894
y10+-H2O-P 1052.42224
y10+-P 1070.44177
y11+-P 1207.49243
136.07565
674.76788

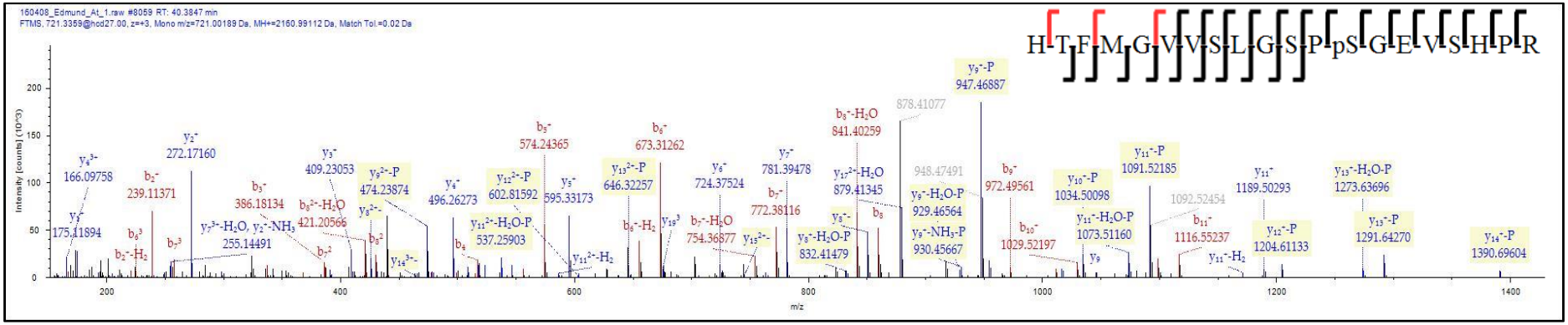

Intensity (%)
I-G-H-H-pS-T-S-D-D-S-S-A-Y-R
y1-NH3 158.0923
y1 175.1190
b5[2+](-98) 257.6300
y2 338.1815
y3-NH3 391.6553
y3 409.2229
b8[2+] 458.1724
y4 b5-H2O(-98) 496.2524
b9[2+] 515.6856
y5 583.2835
b6-H2O(-98) 597.2894
y6 698.3104
y7-H2O 795.3246
y7 813.3347
b8(-98) 817.3615
y8 900.3754
b9(-98) 932.3857
y10(-98) 1070.4418
y11-H2O(-98) 1189.4836
y11(-98) 1207.4924
y13-H2O(-98) 1383.5658
y13(-98) 1401.5826

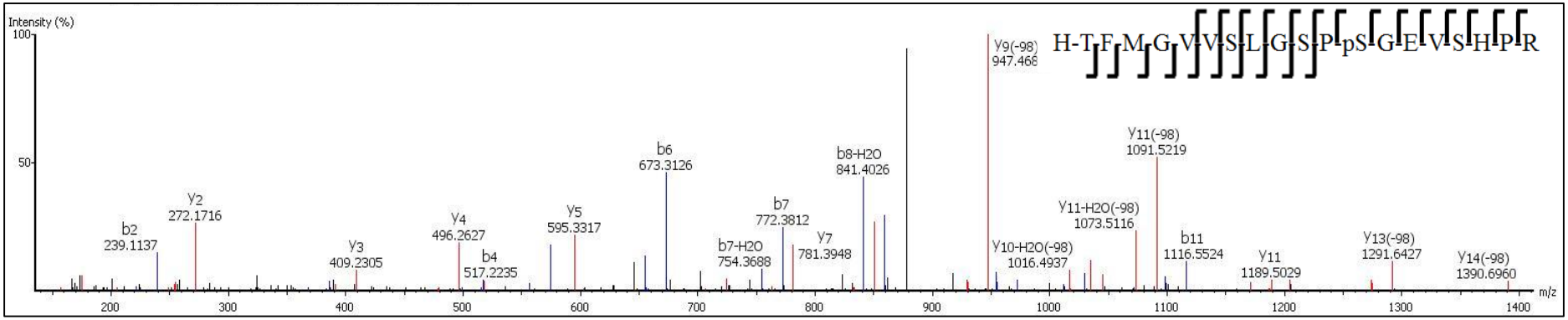

I-G-H-H-pS-T-S-D-D-S-S-A-Y-R
Int
m/z
y1
y1-NH3
y2
y2-NH3
y3
y4
b5-H2O-H3PO4
y5
y6
y7
y8
b9-H3PO4
y10-H3PO4
y11-H3PO4
y11++-H3PO4

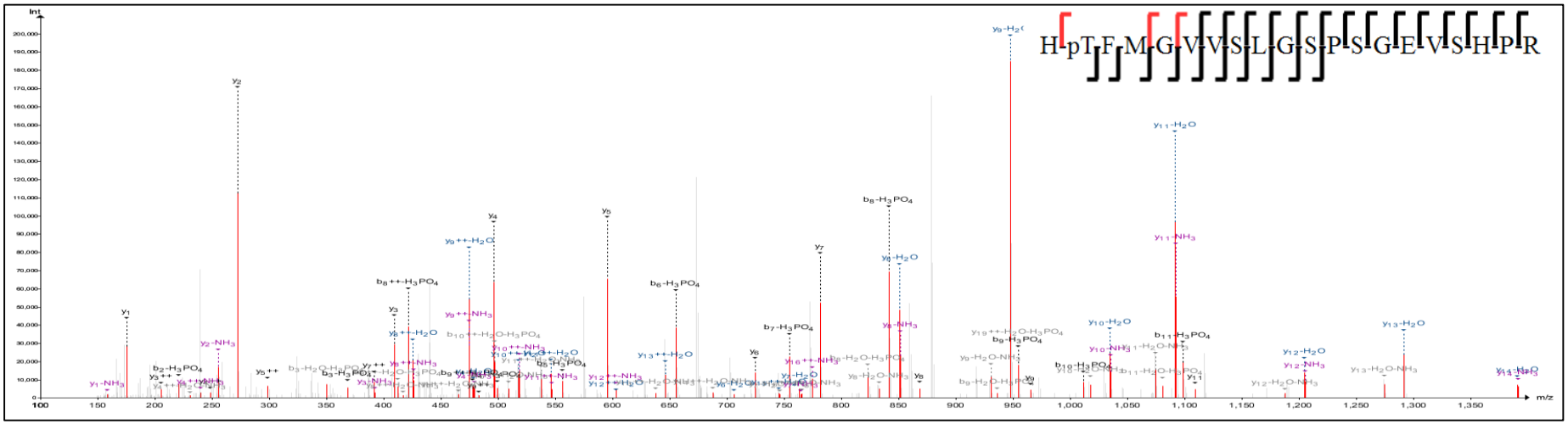

160408_Edmund_At_1.raw #4509 RT: 25.2740 min
FTMS, 526.7418@hcd27.00, z=+2, Mono m/z=526.74182 Da, MH+=1052.47637 Da, Match Tol.=0.02 Da
L[S[P[pS[P[T[S[Q[R
Intensity (counts) (10^3)
m/z
a1+ 86.09705
a2+-H2O 155.11809
167.08141
a2+ 173.12837
265.05823
y2+ 303.17737
y8 2+-P 329.16898
y3+ 390.20929
y7 2+ 426.68378
y5+ 588.30914
685.32574
84.04499
y1+ 175.11880
b2+ 201.12344
y8 2+ 294.65857
y2 2+-P 377.69513
y7 2+-NH 378.19839
y7 2+-H2O
y8 2+-H2O
y4+-H2O 473.24826
y8 2+ 470.20551
y4+ 491.25671
492.26022
y5+-NH3 571.28711
y5+-H2O 570.29926
589.31262
y6+-P 657.32977
667.31677
686.32837
y7+-P 754.38409
y4+-H2O 737.30084
755.38660
y8+-P 841.40558
b2 2+ 101.07098
163.37373
264.09769

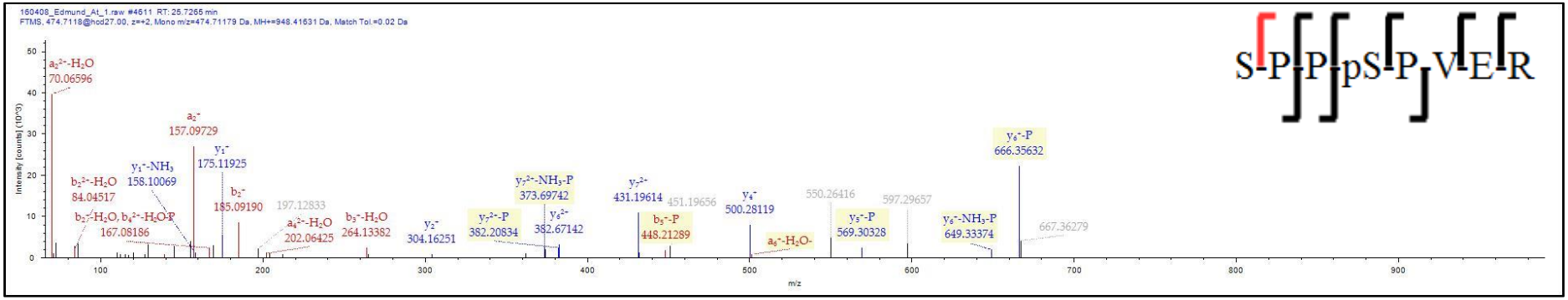

Intensity (%)
L[S[P[pS[P[T[S[Q[R
m/z
Y1 175.1188
Y5[2+] 294.6586
Y3 390.2093
Y8[2+] 470.2055
Y4 491.2567
Y5-H2O 570.2993
Y5 588.3091
Y6(-98) 657.3298
Y6-H2O 737.3008
Y7(-98) 754.3841
Y7 852.3597

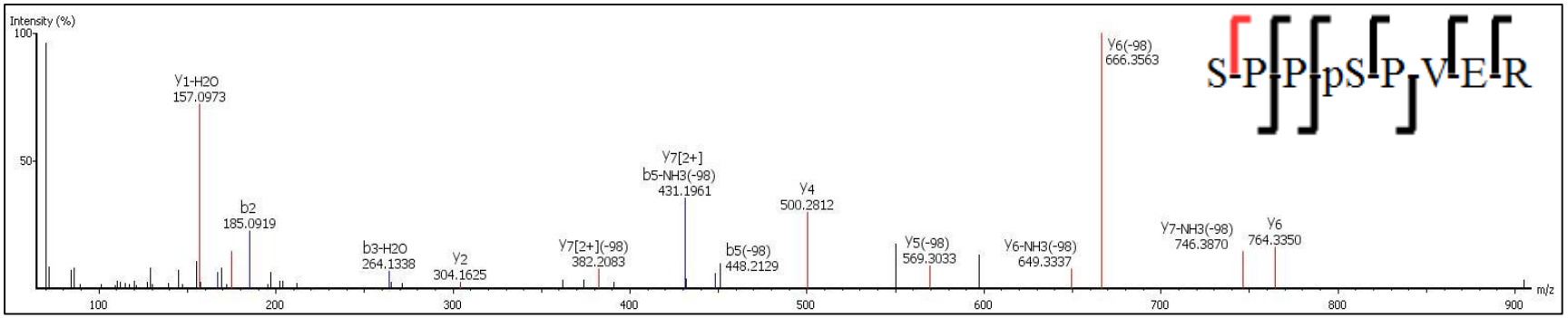

Int.
L[S[P[pS[P[T[S[Q[R
m/z

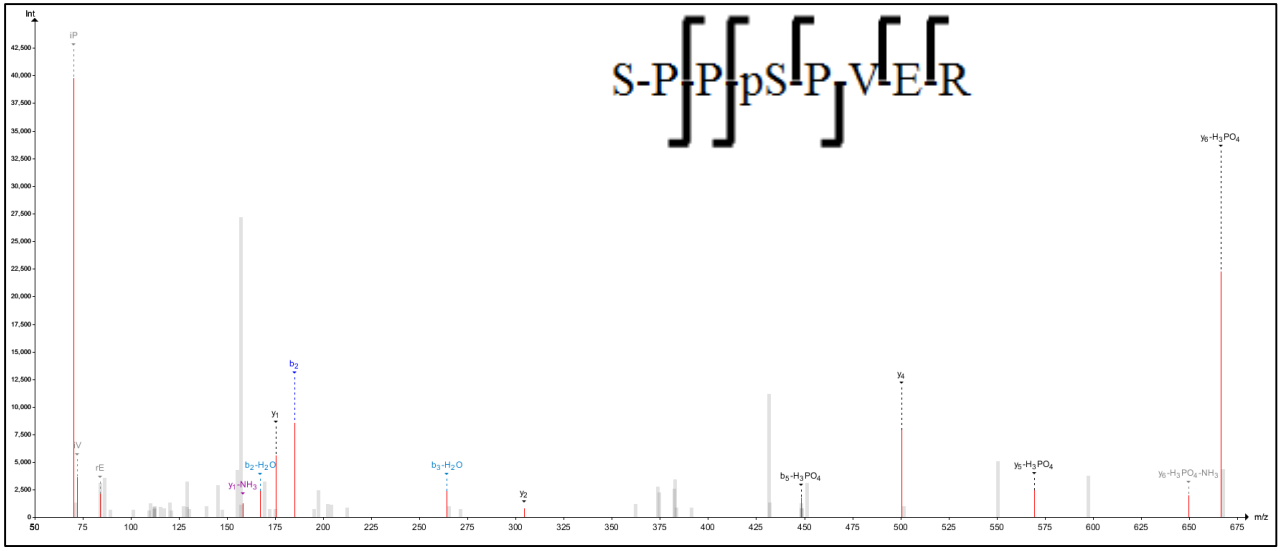

160408_Edmund_At_1.raw #8059 RT: 40.3847 min
FTMS, 721.3359@hcd27.00, z=+3, Mono m/z=721.00189 Da, MH+=2160.99112 Da, Match Tol.=0.02 Da
H-T-F-M-G-V-V-S-L-G-S-P-pS-G-E-V-S-H-P-R
Intensity [counts] (10^3)
m/z


Intensity (%)
H-T-F-M-G-V-V-S-L-G-S-P-pS-G-E-V-S-H-P-R
y9(-98) 947.468
b6 673.3126
b8-H2O 841.4026
y11(-98) 1091.5219
y2 272.1716
b2 239.1137
y3 409.2305
y4 496.2627
b4 517.2235
y5 595.3317
b7 772.3812
b7-H2O 754.3688
y7 781.3948
y11-H2O(-98) 1073.5116
y10-H2O(-98) 1016.4937
b11 1116.5524
y11 1189.5029
y13(-98) 1291.6427
y14(-98) 1390.6960
m/z


Int.
H-pT-F-M-G-V-V-S-L-G-S-P-S-G-E-V-S-H-P-R
m/z


160408_Edmund_At_1.raw #4611 RT: 25.7265 min
FTMS, 474.7118@hcd27.00, z=+2, Mono m/z=474.71179 Da, MH+=948.41631 Da, Match Tol.=0.02 Da
Intensity [counts] (10^3)
m/z
a2 2+-H2O 70.06596
a2+ 157.09729
y1+ 175.11925
y1+-NH3 158.10069
b2 2+-H2O 84.04517
b2+ 185.09190
197.12833
b2+-H2O, b4 2+-H2O-P 167.08186
a4 2+-H2O 202.06425
b3+-H2O 264.13382
y2+ 304.16251
y7 2+-P 382.20834
y3 2+-NH3-P 373.69742
y6 2+ 382.67142
y7 2+ 431.19614
451.19656
b5+-P 448.21289
y4+ 500.28119
a6+-H2O- 550.26416
y5+-P 569.30328
597.29657
y6+-NH3-P 649.33374
y6+-P 666.35632
667.36279
S-P-P-pS-P-V-E-R


Intensity (%)
y1-H2O 157.0973
b2 185.0919
b3-H2O 264.1338
y2 304.1625
y7[2+](-98) 382.2083
y7[2+] b5-NH3(-98) 431.1961
b5(-98) 448.2129
y4 500.2812
y5(-98) 569.3033
y6-NH3(-98) 649.3337
y6(-98) 666.3563
y7-NH3(-98) 746.3870
y6 764.3350
m/z
S-P-P-pS-P-V-E-R


Int.
m/z
S-P-P-pS-P-V-E-R

4.2 Phosphopeptide identification

The first set of results were enriched using the batch mode protocol published by Dickhut et. al. (61) There were both enriched and non-enriched samples with two of the samples containing a phosphate inhibitor cocktail (PhosSTOP). It could seem that PhosSTOP had negative effect on samples (fig. 4.2-1). This reason for this was not examined any further. In theory phosphate inhibitors are supposed to inhibit the action of phosphatase enzyme in order for the phosphor-group on serine/threonine/tyrosine to avoid dephosphorylation.

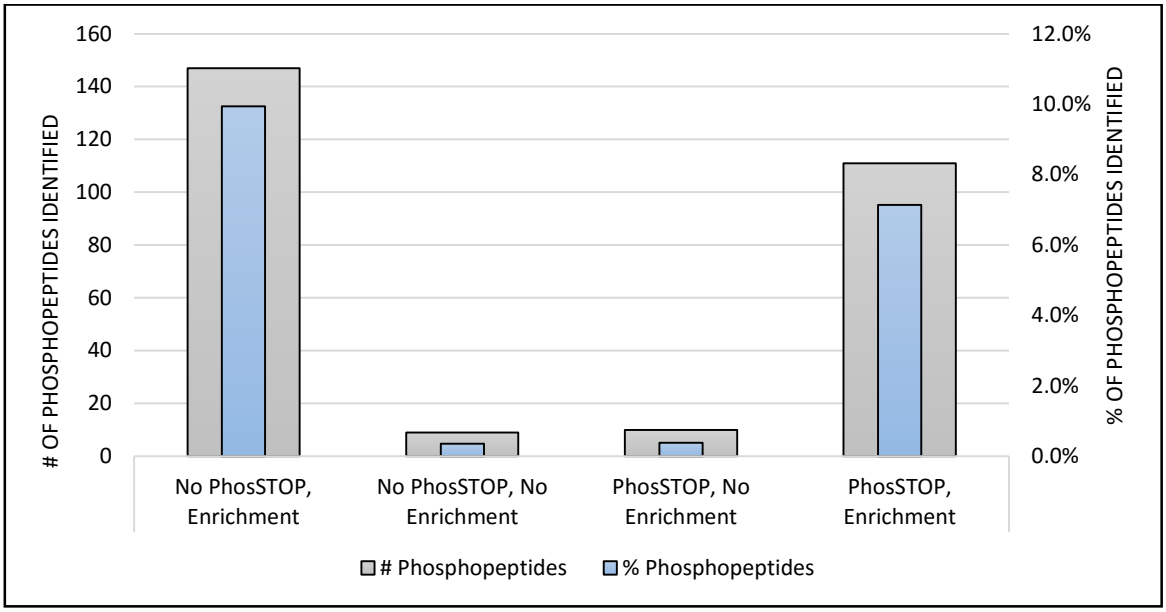


Figure 4.2-1: Phosphopeptide identifications in batch mode protocol. Mascot identified a higher number of phosphopeptides in the samples enriched compared to the enriched-free samples. The number of phosphopeptide identified in the enriched-free samples showed to be not significant even though there were PhosSTOP in one of them (No PhosSTOP, No Enrichment = 9 and PhosSTOP, No Enrichment = 10). In the enriched samples, however, the PhosSTOP-free sample identified a significant higher number of identification (147) compared to sample containing PhosSTOP (111).

Results from the batch mode protocol were less than expected, so the samples were enriched by conducting the micro column protocol in the following experiment. The non-enriched samples yielded extremely low phosphopeptides leading to the discontinuation of future experiments. The micro column protocol showed to identify more phosphopeptides than batch mode protocol. Samples without PhosSTOP yielded a higher number of phosphopeptides in comparison to the samples containing PhosSTOP (fig. 4.2-2).

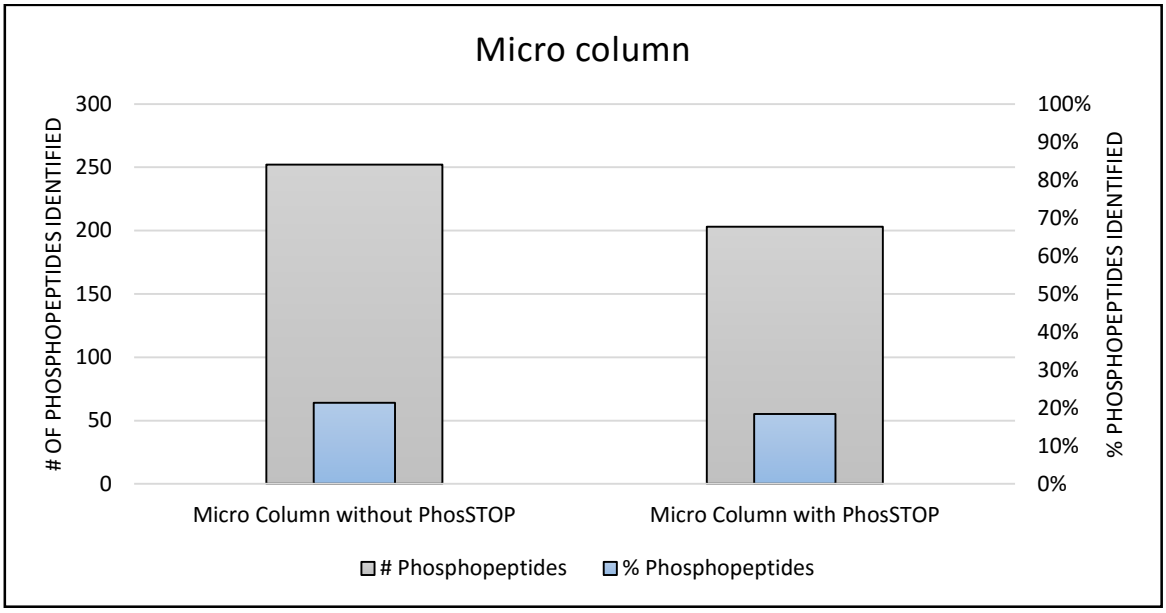


Figure 4.2-2: Phosphopeptide identifications in micro column protocol. Samples without PhosSTOP yielded 252 phosphopeptides, whereas PhosSTOP-containing samples yielded 203. Samples had also higher percentage of phosphopeptide (21%) than samples with PhosSTOP (18%)

As mentioned earlier, theoretically by adding PhosSTOP to samples, it was expected was to reduce the phosphatase activity thereby increasing the number of phosphopeptides. PhosSTOP is a cocktail of phosphatase inhibitors of different classes; Acid phosphatases, alkaline phosphatases, serine/threonine phosphatases, tyrosine phosphatases and dual-specificity phosphatases (64), with the intention to prevent cleavage of phosphor group from peptides and protein during sample preparation.

However the results show preparing the samples without PhosSTOP and using Metal-Oxide Affinity Chromatography (MOAC) enrichment method, in this case TiO_2, was able to extract a higher percentage of phosphopeptides than the other samples. The reason for this is unknown, but looking at the non-enriched samples (fig. 4.1-2) the sample with PhosSTOP yielded a slight increase in phosphopeptides identifications. Given that the sample was enriched and contained PhosSTOP, the identifications were less than in the sample not containing PhosSTOP while being enriched. Thus, there seems to be an association between PhosSTOP and the type of enrichment. The association was however was not examined. Aryal et. al (65) examined the compatibility of phosphatase inhibitors and TiO_2. Their findings showed that phosphatase inhibitor cocktails reduces the relative number and intensity of multiple phosphorylated peptide peaks (fig 4.2-4 and fig 4.2-5). This was also found when validating the method on HeLa cells(fig. 4.3-1).

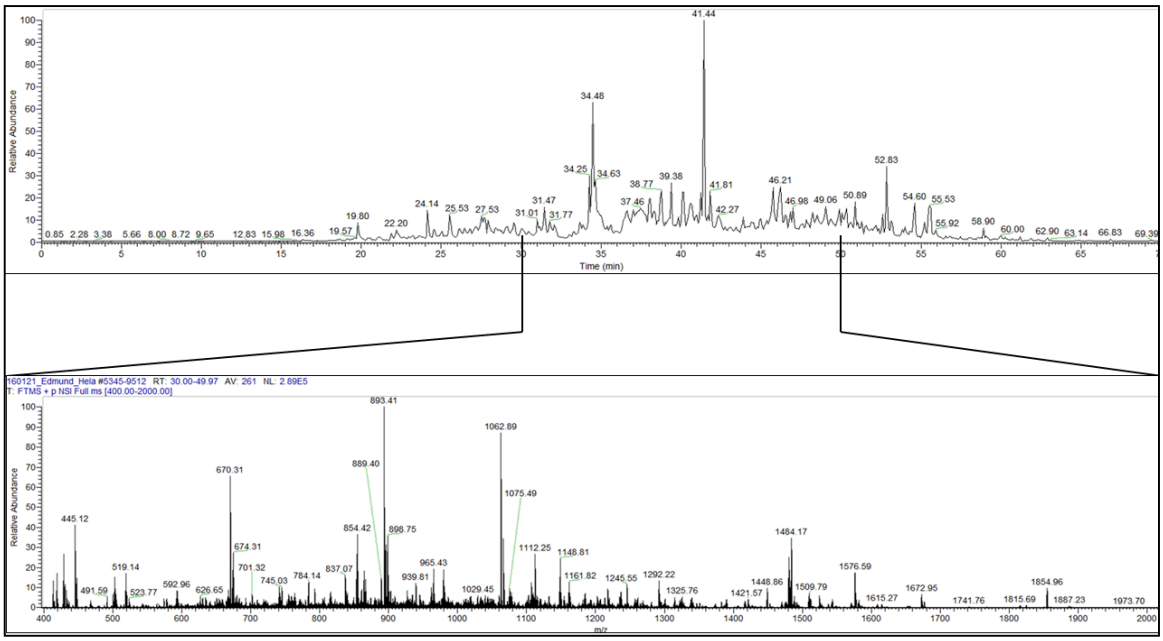


Figure 4.2-3: HeLa chromatogram (upper) and MS spectrum (lower) without phosphate inhibitorThe chromatogram zoomed in at retention time interval 30-50 min.

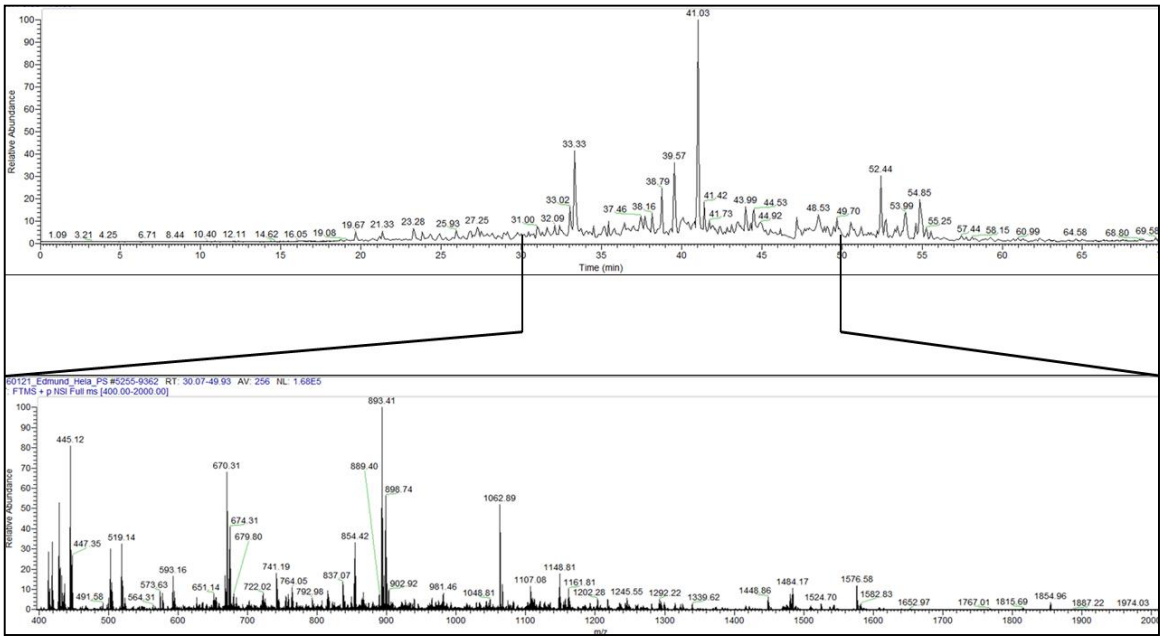


Figure 4.2-4: HeLa chromatogram (upper) and MS spectrum (lower) with phosphate inhibitorThe chromatogram zoomed in at retention time interval 30-50 min.

Regarding the chromatogram and MS spectra of HeLa Cells (fig. 4.2-3 and 4.2-4) with and without PhosSTOP and contradicting observations in the number of phosphopeptides that were identified. Consequently, no further test were conducted with PhosSTOP.

Determining the enrichment protocol that were to be conducted for the experiment was done by plotting confident sequences from batch mode and micro column in a Venn diagram (4.2-6)

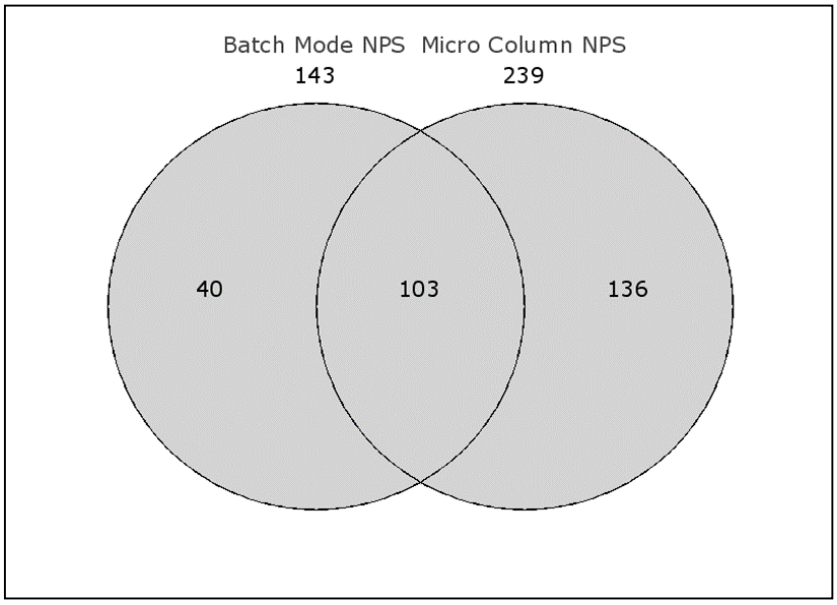


Figure 4.2-5: Overlap of phosphopeptide identified in batch mode protocol and micro column. The micro column protocol was more successful than the batch mode protocol with 40% higher phosphopeptide identifications. Both enrichment protocol commonly identified 103 phosphopeptides. The micro column procedure identified a significant higher number of unique phosphopeptides (239) compared to batch mode procedure (143).

Based on the number of phosphopeptides identified from the comparison of the enrichment protocols, the micro column protocol seemed to be more reliable and was chosen to be conducted in future experiments.

4.3 Validation of sample preparation method.

The validity of the procedure was tested by using HeLa cells (fig. 4.3-1).

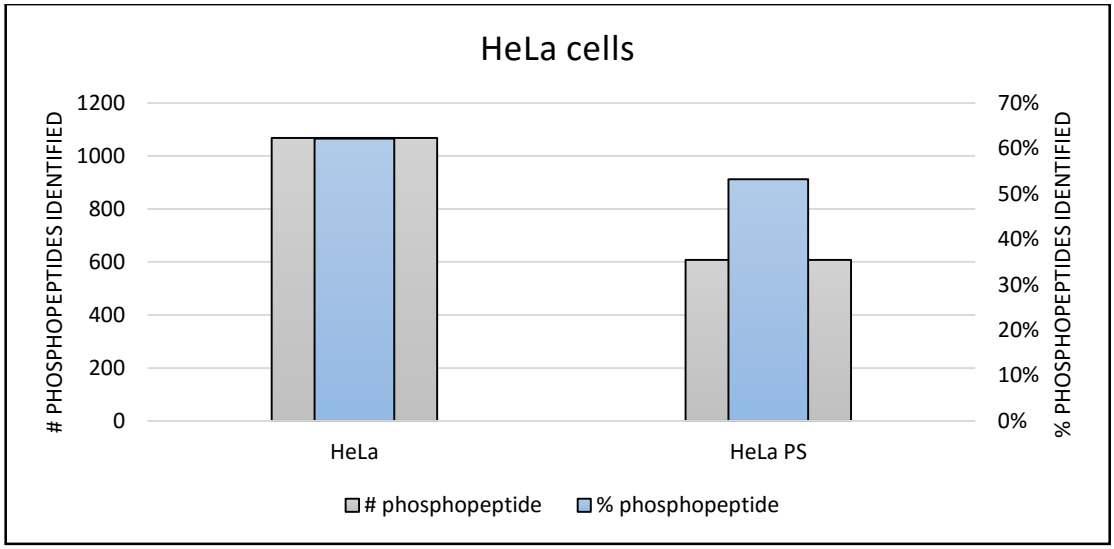


Figure 4.3-1: Validation of the approach on HeLa cells – Phosphopeptides identified.

Due to the number of phosphopeptide was lower than anticipated, the adipose tissue was tested using the sample preparation and micro column enrichment protocol published by Dickhut et. al (61) in parallel with HeLa cells. In addition, PhosSTOP was attempted to be removed by protein precipitation due to ethanol. This was in order to stop the activity of protein kinase in sample followed by removal prior to tryptic digest. As well as to evaluate if the PhosSTOP was removed by protein precipitation. With HeLa cells the number of phosphopeptides increased, while in adipose tissue the number of phosphopeptide decreased significantly (fig. 4.3-1)

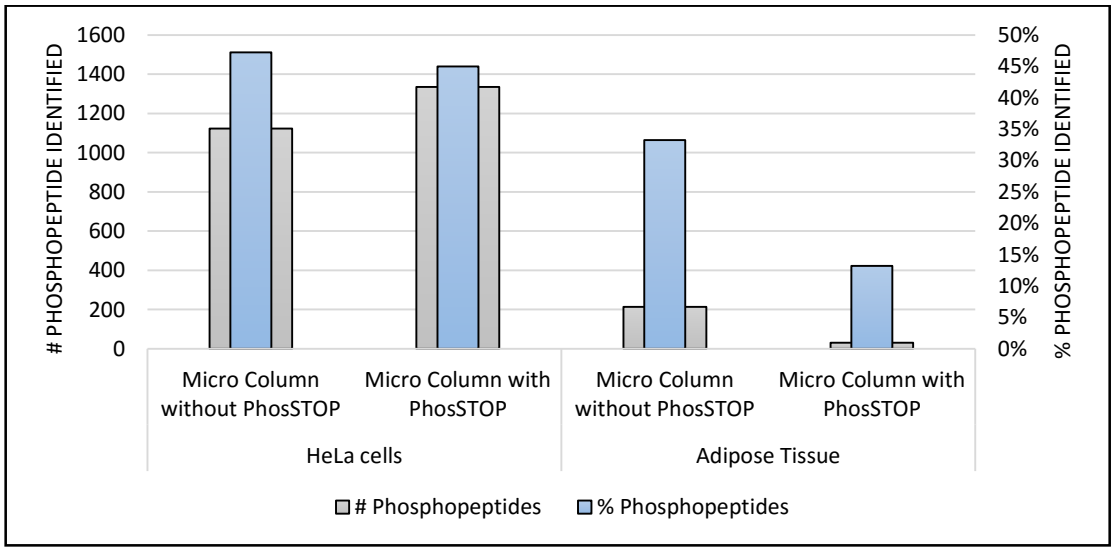


Figure 4.3-2: Phosphopeptide identified in HeLa cells from the protocol published by Dickhut et. al

The results show that the sample preparation protocol used during this project works, but adipose tissue as the sample for analysis is somehow problematic due to low identifications of phosphopeptides.

4.4 Correlation between peptide charge state and missed cleavage

Phosphor-group on Serine (pSer), Threonine (pThr) primarily and Tyrosine (pTyr) nearby enzymatic cleavage sites has shown to increase the likelihood that trypsin fails to cleave arginine and lysine.(66).

Serine proteases have a catalytic triad consisting of Histidine 57 (His57), Aspartate 102 (Asp102) and Serine 195 (Ser195). In the catalytic pocket, the arginine/lysine residue of a protein makes an ionic interaction with Asp189 so that cleavage can occur. The mechanism is as illustrated in fig. 4.4-1 (11).

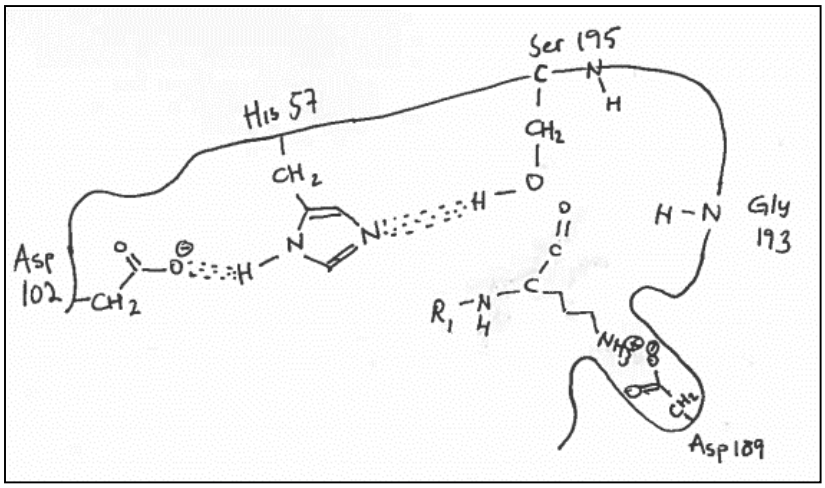


Figure 4.4-1: Scheme of proteolytic digestion of peptides

Dickhut et. al (66) found that pSer or pThr positioned p_2 or p_3 from cleavage site, which is arginine or lysine at position p_1, could impair the digestion of peptides. The possible mechanism for impairing proteolytic digestion of peptides is the formation of salt bridges between pSer/pThr and arginine/lysine, which means that the salt bridge formation between arginine/lysine and Asp189 does not occur. Other factors that could impair the enzymatic cleavage is proline positioned after arginine or lysine and the effects that aspartate and glutamate has in order to form a salt bridge (67).

Missed cleavage of tryptic peptides can also lead to higher peptide charge state as the basic residues will increase, and it was therefor of interest to have a look at the correlation between

peptide charge and missed cleavage of tryptic peptides ranged from zero to three (fig. 4.4-2). The chart shows that the average peptide charge decreases when the peptide contains a phosphor-group. The buffer used to re-dissolve the sample has a pH=2 meaning that pSer/pThr/pTyr are partially negatively charged (pKa of phosphor = 2.15 (68)) which could be the possible reason for the net decrease in average peptide charge state in phosphopeptides compared to non-phosphorylated peptides. There is, however, limited or no literature regarding this topic. Steen et al. has suggested that phosphor group on serine, threonine and tyrosine can affect the net charge of the peptide, which could affect the retention time of phosphopeptides eluted in the LC-MS (69).

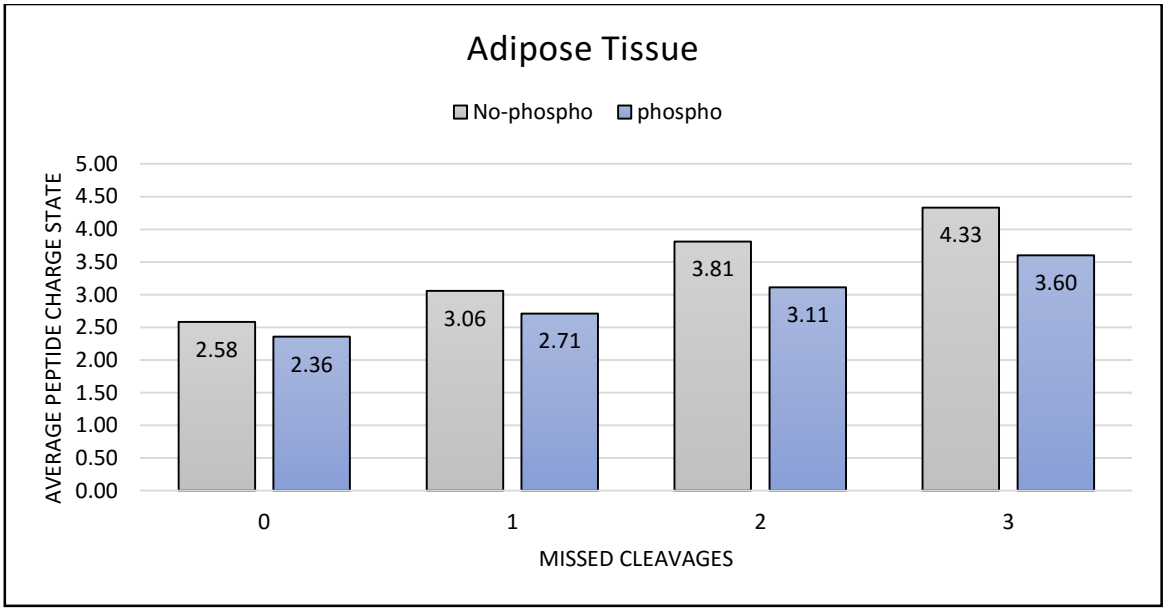


Figure 4.4-2: Average charge state of non-phosphorylated peptides and phosphorylated peptides in adipose tissue

This pattern is similar in HeLa cells (fig. 4.4-3) where the average charge of phosphopeptides are lower than non-phosphorylated peptides.

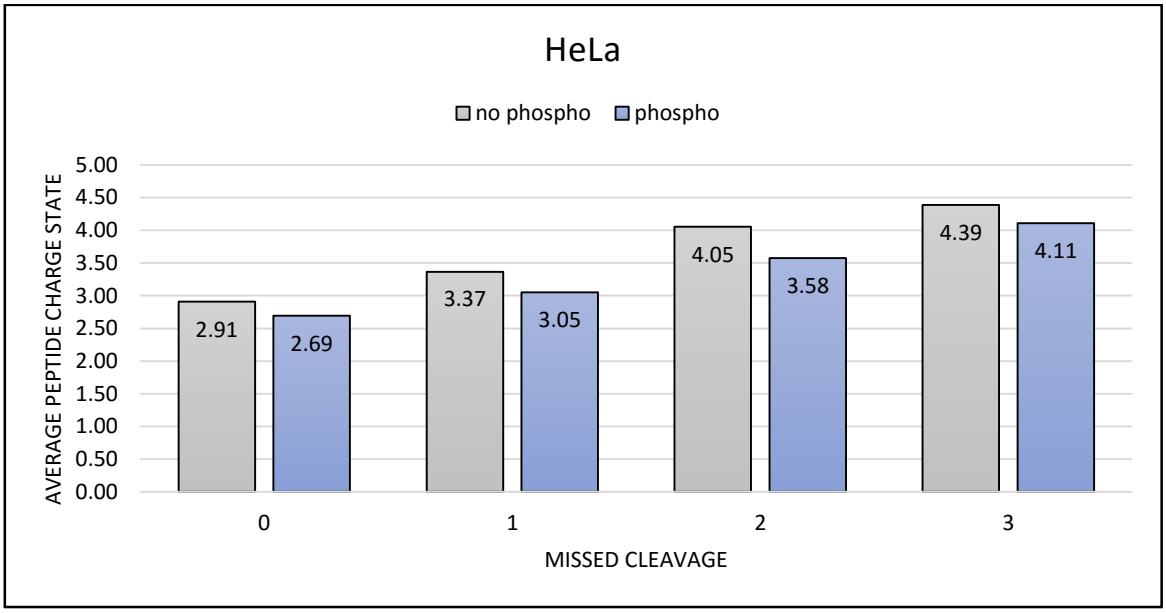


Figure 4.4-3: Average charge state of non-phosphorylated peptides and phosphorylated peptides in HeLa cells

4.5 Optimization of starting material

The starting material of the experiment was based on BCA assay results. The experiment was started with 100 µg proteins. Interpretation of the results on Mascot showed very low results of phosphopeptides (less than 100 confident identifications), and the relative abundance of peptides in the chromatogram and MS spectrum were very low. The samples, however, in the early stages were not enriched due to determination of the SDC concentration in samples. To compensate for loss of peptides during sample preparation, the amount of starting material was increased to 200 µg protein in the next round. The enrichment step was included in this round, yet still the number of identifications were slightly better than 100 µg as starting material, so it was decided to increase to 500 µg protein as starting material. These results showed to be satisfying. It was then decided to increase the injection concentration from 0.5 µg to 1 µg, which showed slightly higher relative abundancy in the chromatogram and MS spectrum, and in terms of phosphopeptide identification, the numbers of identification increased gaining approximately 100 more phosphopeptides.

4.6 Fragmentation of precursor ions

It was noticeable that Mascot had issues when correctly identifying the correct phosphorylation site in some peptide sequences, especially possible phosphorylation sites that were adjacent. An example of this shown in fig. 4.6-1 where Mascot has found two phosphorylation sites. The phosphorylation site at S7 is seems to be accurate whereas Mascot seems to be uncertain whether the other phosphorylation site is a S1 or S2.

p(50)S-p(50)S-P-F-K-V-p(100)S-P-L-T-F-G-R

Figure 4.6-1: Phosphopeptide sequence with uncertain phosphorylation site

The difficulty in providing the correct localization of phosphorylation site has been proved by Palumbo et. al that phosphor-group can be rearranged to another amino acid with a hydroxyl group, given that amino acids are in the same precursor peptide. It was suggested that this could be facilitated by protonated arginine (70). This normally occurred in the time space where the neutral loss was predominant in CID, which is at the time where the proton is not mobile or partially mobile (71). This could therefore give a low localization score Mascot. The mechanism CID used to fragment peptides is described as the 'mobile proton model'. In the collision cell, the inert nitrogen gas collides with protonated peptides and creates an internal energy distributed over the whole peptide. The collision makes the proton mobile, making it likely to change its position to a place in the peptide where it is less favorable energetically, weakening the amide bonds followed by dissociation of the protonated site. The fragment-ions yielded are b- and y-ions (71). The neutral loss, however, seems to be dependent on the ratio of peptide charge state and number of basic residues in the peptide. A charge state higher than number of basic residues yields lesser extent of neutral loss and the fragmentation occurs easily. Whereas a charge state lower than number of basic residue in the peptide requires higher energy in order to fragment the peptide backbone, and a low-energy pathway, e.g. neutral loss of phosphor, assisted by arginine or lysine is more favorable (72). An example of confident phosphopeptide is shown in fig. 4.6-2

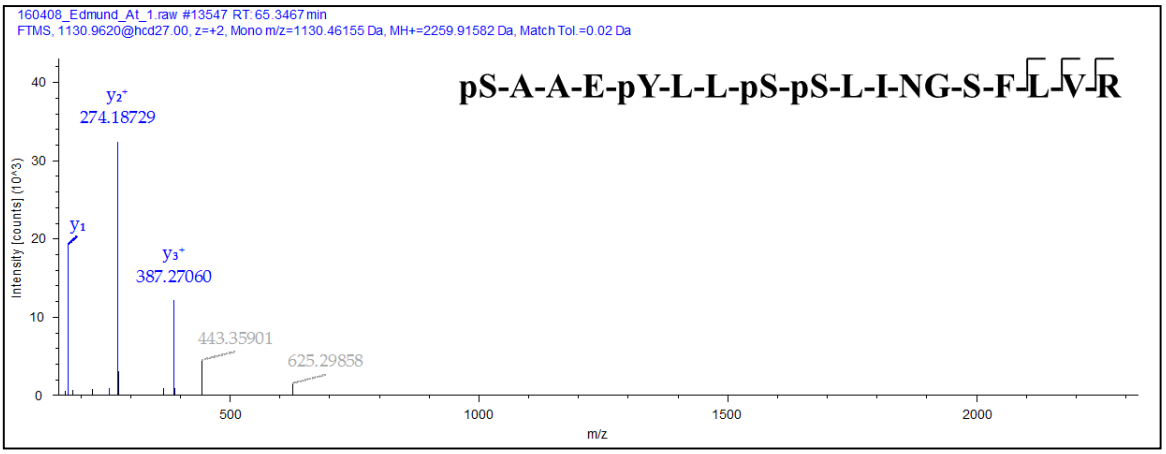


Figure 4.6-2: CID spectrum of a multiple phosphorylated peptideAll phosphoamino acids had localization scores of 80 out of 100 and was overall a confident identification. However, it is with care when claiming that the spectrum shows fragmentation of peptides.

The favorable fragmentation for phosphorylated peptides is ECD or ETD. These types of fragmentation techniques are similarly to each other and yet differently from CID. One of the major advantages of employing ETD for fragmentation of complex peptide mixtures compared to CID is the neutral loss of labile PTMs, which can hamper the fragmentation spectra. In ETD, neutral losses do not occur, giving good fragmentation. The mechanism of ETD is the transfer of electrons with a charge less than 0.2 eV to precursor ions. The electrons are transferred from a radical anion. The electron is captured by the peptide and charge-reduces the peptide into a radical cation. The basicity of the peptide (the amide carbonyl oxygen) increases to an extent that it is capable of taking a proton from the radical cationized residue yielding cleavage at the N-Cα bond to generate c- and z-dot ion products (71). Another advantage with ETD, although it is no restriction, is efficiency when it comes to peptide charge, where ETD seem to be more efficient than CID when the peptide charge is ≥3+ (73). The only disadvantage with ETD is the slower scan rate compared to CID (71, 73, 74)

ETD would therefore be a better option for the fragmentation of phosphopeptides and more suited for the search engine in terms of correct localization score of phosphorylation site.

4.7 Quantification with TMTzero (TMT_0) labelling

The peptides were labeled with TMT0 reagent to determine the labeling step in the procedure and observe the reagents compatibility with the sample. TMT in multiplex is used for the relative quantification of the proteome.

Samples were labeled before and after enrichment with TiO2. Labeling before enrichment, 50 mM AmBic as buffer was used for the digestion of proteins and re-suspended in 100 mM TEAB buffer prior to labeling. Labeling before enrichment, 100 mM TEAB as digestion buffer was used. Because of TMT reagents activity with amines, we changed the buffer to TEAB to avoid reaction with the buffer. TMT reagent reacts with Lys and N-terminal in peptides so it is crucial to remove any primary amine-based buffers (62). The chart (fig. 4.7-1) shows the number of phosphopeptides identified when labeling before enrichment and after enrichment. Labelling after enrichment seems to cause a greater loss of phosphopeptides than labelling before enrichment with only three phosphopeptides labeled after enrichment.

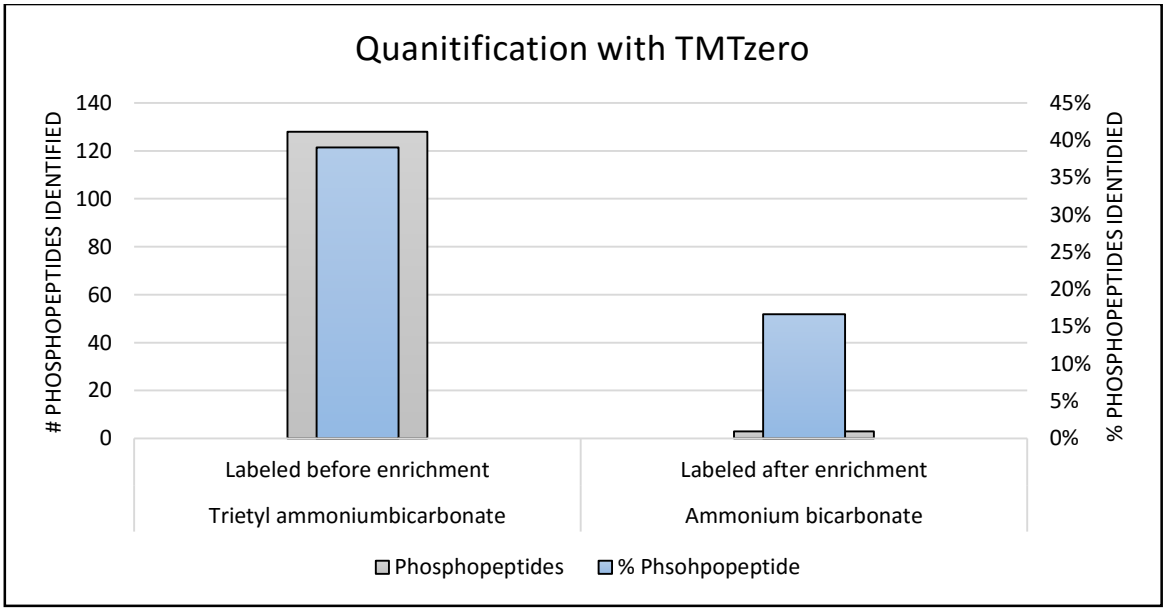


Figure 4.7-1: Labeled phosphopeptides identifications. Labeling before enrichment yielded 128 TMT-tagged phosphopeptides, labeling after enrichment yielded 3 TMT tagged phosphopeptides.

Three replicates of labeled and unlabeled were run. The pre-processed samples were split after proteolytic digest in to two, so that for each sample that was split, had a replicate that was TMT labeled. It is possible to see from fig. 4.7-2 that labeling in general causes loss of phosphopeptides. The next matter examined was commonly phosphopeptides identified in unlabeled and labeled samples. The confident sequence were extracted and plotted in a Venn diagram. From the unlabeled samples, 186 phosphopeptides were commonly identified and 71 commonly identified phosphopeptides in the labeled samples (fig 4.7-3). The overlapping sequences from unlabeled and labeled samples were plotted in a Venn diagram to get an overview of the commonly identified phosphopeptides in all six samples, which was found to be 50 phosphopeptides. In terms of uniqueness, 479 phosphopeptides were unique

identification in unlabeled samples, whereas 189 phosphopeptides were unique in the labeled samples (fig. 4.7-4).

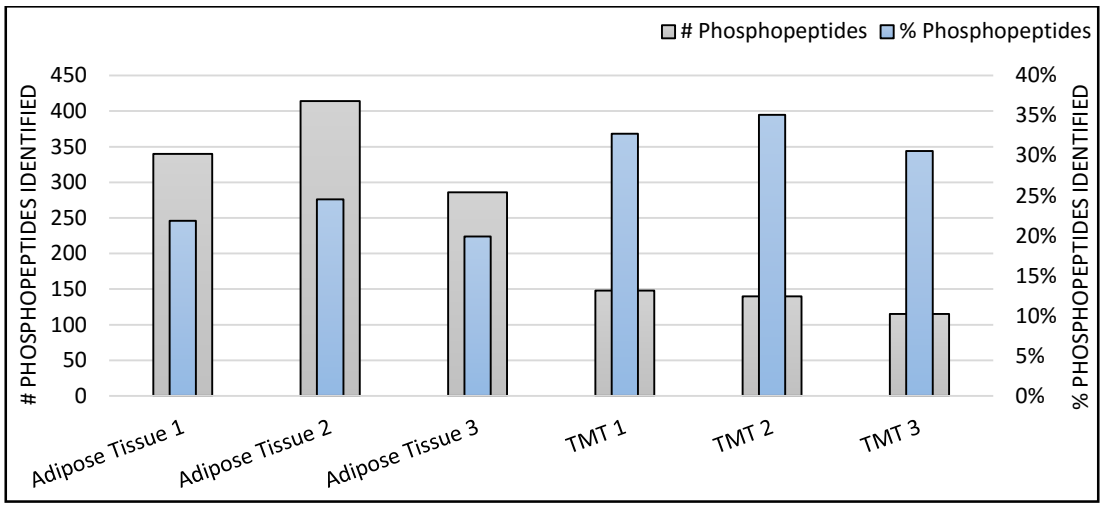


Figure 4.7-2: Phosphopeptide identified in adipose tissue 1-3 and TMT labeled adipose tissue 1-3

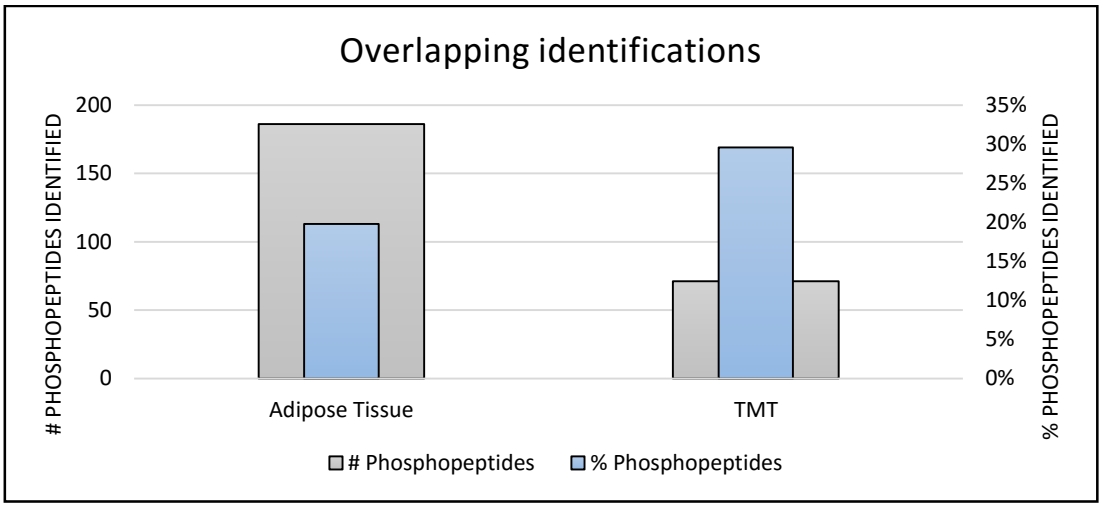


Figure 4.7-3: Phosphopeptides identified in labeled and unlabeled adipose tissue samples

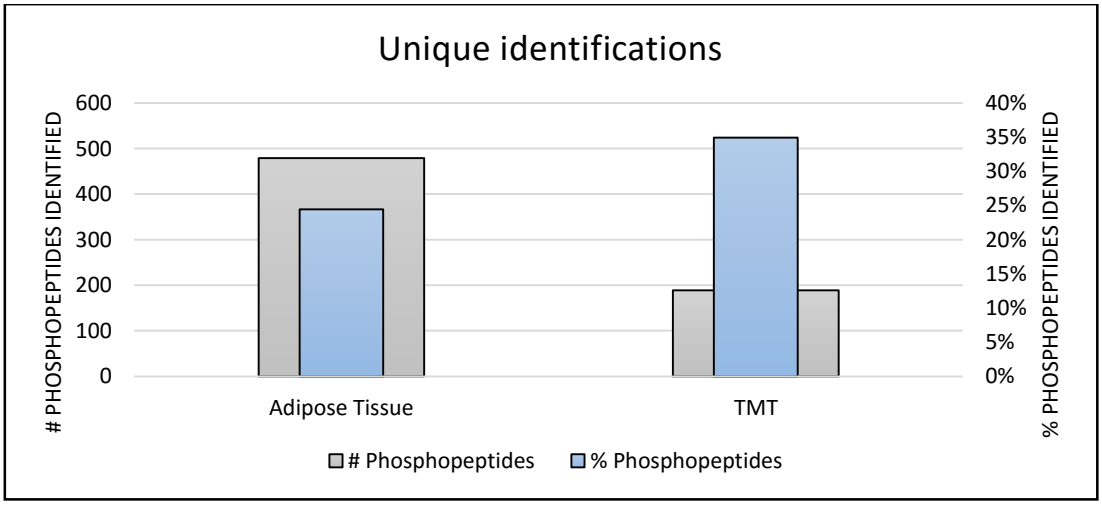


Figure 4.7-4: Number of phosphopeptide identifications from overlapping labeled and unlabeled adipose tissue samples

5 Conclusion

In this master thesis, a novel bottom-up based procedure has been developed and presented. Two enrichment protocols published by Dickhut et. al were tested in order to yield the highest number of phosphopeptides in which micro column protocol was determined to be conducted based on higher yield in confident phosphopeptides than batch mode protocol. In terms of search engines, Mascot was conducted as it proved to be most compatible with the data compared to other search engines utilized in this thesis.

The validity of the procedure was tested by employing it on HeLa cells, which proved to work successfully.

TMT0 reagent was tested on samples to determine which step it was appropriate to label the samples without losing high number of phosphopeptides. Labeling before enrichment seemed to be convincing compared to labeling after enrichment, and was based on the outcome conducted before enrichment.

6 Future aspects

In attempt to improve the validity of the sample preparation method, it would be interesting to test the procedure on freshly extracted adipose tissue. This due to the comparison of fresh adipose tissue against frozen adipose tissue in order to exclude the difference in number of identifications.

It would be a better for the project if the collision cell used ETD fragmentation technique as this type of technique gives no neutral loss and is better suited for search engines in terms of identification of phosphopeptides.

The results from the project provides information about the phospho proteome in adipose tissue, which can be further examined and compared to adipose tissue associated changes in the clinical aspect, for example patients who frequently administrate vitamin D supplements versus patients not taking vitamin D supplements.

References

1. Mishra NC. Introduction to proteomics : principles and applications. Hoboken, N.J.: Wiley; 2010. 7-8, 25, 103-13 p.

2. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003 Mar;422(6928):198-207.

3. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Molecular & Cellular Proteomics. 2005 Oct;4(10):1487-502.

4. Zhu W, Smith JW, Huang CM. Mass spectrometry-based label-free quantitative proteomics. Journal of Biomedicine and Biotechnology. 2010;2010:840518.

5. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: a critical review. Analytical and Bioanalytical Chemistry. 2007 Oct;389(4):1017-31.

6. Aebersold R. Quantitative Proteome Analysis: Methods and Applications. Journal of Infectious Diseases. 2003 Jun;187(Supplement 2):S315-S20.

7. Fleet JC. Genomic and proteomic approaches for probing the role of vitamin D in health. American Journal of Clinical Nutrition. 2004 Dec;80(6 Suppl):1730S-4S.

8. Mutt SJ, Hypponen E, Saarnio J, Jarvelin MR, Herzig KH. Vitamin D and adipose tissue-more than storage. Frontiers in Physiology. 2014 Jun;5:228.

9. Crick F. Central Dogma of Molecular Biology. Nature. 1970 Aug;227(5258):561-3.

10. Nelson DL, Cox MM, Lehninger AL. Lehninger principles of biochemistry. 5th ed. New York: Freeman; 2008. 71-140 p.

11. Branden C, Tooze J. Introduction to Protein Structure. New York: Garland Publishing; 1991. 3-5 p.

12. Green K, Garneau-Tsodikova S. Posttranslational Modification of Proteins. In: Liu H-W, Mander L, editors. Comprehensive Natural Products II. Oxford: Elsevier; 2010. p. 436-42.

13. Walsh C, Garneau-Tsodikova S, Gatto Jr. G. Protein posttranslational modifications: the chemistry of proteome diversifications. Angewandte Chemie International Edition. 2005 Dec;44(45):7345.

14. Kehoe JW, Bertozzi CR. Tyrosine sulfation: a modulator of extracellular protein–protein interactions. Chemistry and Biology. 2000 Mar;7(3):R57-R61.

15. Monigatti F, Hekking B, Steen H. Protein sulfation analysis—A primer. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics. 2006 Dec;1764(12):1904-13.

16. Colaianni G, Colucci S, Grano M. Anatomy and Physiology of Adipose Tissue. In: Lenzi A, Migliaccio S, Donini ML, editors. Multidisciplinary Approach to Obesity: From Assessment to Treatment. Cham: Springer International Publishing; 2015. p. 3-12.

17. Coelho M, Oliveira T, Fernandes R. Biochemistry of adipose tissue: an endocrine organ. Archives of Medical Science. 2013 Apr;9(2):191-200.

18. Bernlohr DA, Jenkins AE, Bennaars AA. Adipose tissue and lipid metabolism. In: Vance JE, Vance DE, editors. Biochemistry of lipids, lipoproteins, and membranes 4th edition. Amsterdam: Elsevier; 2002. p. 263-89.

19. Rosen ED, MacDougald OA. Adipocyte differentiation from the inside out. Nature Reviews: Molecular Cell Biology. 2006 Dec;7(12):885-96.

20. Nakagami H. The mechanism of white and brown adipocyte differentiation. Diabetes & Metabolism Journal. 2013 Apr;37(2):85-90.

21. Dani C, Billon N. Adipocyte Precursors: Developmental Origins, Self-Renewal, and Plasticity. In: Symonds EM, editor. Adipose Tissue Biology. New York, NY: Springer New York; 2012. p. 1-16.

22. Klingenspor M, Fromme T. Brown Adipose Tissue. In: Symonds EM, editor. Adipose Tissue Biology. New York, NY: Springer New York; 2012. p. 39-69.

23. Fruhbeck G, Gomez-Ambrosi J, Muruzabal FJ, Burrell MA. The adipocyte: a model for integration of endocrine and metabolic signaling in energy metabolism regulation. American Journal of Physiology: Endocrinology and Metabolism. 2001 Jun;280(6):E827-47.

24. Conde J, Scotece M, Gomez R, Lopez V, Gomez-Reino JJ, Lago F, et al. Adipokines: biofactors from white adipose tissue. A complex hub among inflammation, metabolism, and immunity. Biofactors. 2011 Nov-Dec;37(6):413-20.

25. Kershaw EE, Flier JS. Adipose tissue as an endocrine organ. Journal of Clinical Endocrinology and Metabolism. 2004 Jun;89(6):2548-56.

26. Ronti T, Lupattelli G, Mannarino E. The endocrine function of adipose tissue: an update. Clinical Endocrinology. 2006 Apr;64(4):355-65.

27. Han X, Aslanian A, Yates JR, 3rd. Mass spectrometry for proteomics. Current Opinion in Chemical Biology. 2008 Oct;12(5):483-90.

28. Chait BT. Chemistry. Mass spectrometry: bottom-up or top-down? Science. 2006 Oct;314(5796):65-6.

29. Resing KA, Ahn NG. Proteomics strategies for protein identification. FEBS Letters. 2005 Feb;579(4):885-9.

30. Dalmasso E, Caseñas D, Miller S. Top-Down, Bottom-Up: The Merging of Two High-Performance Technologies [Internet] Stanford Univeristy, California: Bio-Rad Laboratories, Inc; 2009 [cited 2016.02.08]. Available from: http://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_5908A.pdf.

31. Dunn JD, Reid GE, Bruening ML. Techniques for phosphopeptide enrichment prior to analysis by mass spectrometry. Mass Spectrometry Reviews. 2010 Jan-Feb;29(1):29-54.

32. Fila J, Honys D. Enrichment techniques employed in phosphoproteomics. Amino Acids. 2012 Sep;43(3):1025-47.

33. Quantitative Proteomics [Internet]. Thermo Fisher Scientific Inc; [cited 2016.04.23]. Available from: https://www.thermofisher.com/no/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/quantitative-proteomics.html.

34. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Molecular & Cellular Proteomics. 2002 May;1(5):376-86.

35. Sap K, Demmers J. Labeling Methods in Mass Spectrometry Based Quantitative Proteomics. In: Leung H-C, editor. Integrative Proteomics. Netherlands: InTech; 2012.

36. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Analytical Chemistry. 2003 Apr;75(8):1895-904.

37. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Molecular & Cellular Proteomics. 2004 Dec;3(12):1154-69.

38. EASY-nLC Series: Getting Started Guide [Internet]. USA: Thermo Fisher Scientific Inc; 2013 [cited 2016.04.17]. Available from: http://www.thermoscientific.com/content/dam/tfs/ATG/CMD/CMD%20Documents/lc-associations/EASYnLC-Getting-Started-60053-97229-rev-C-01-13.pdf.

39. Acclaim PepMap and Acclaim RSLC Columns for High-Resolution Peptide Mapping [Internet]. USA: Thermo Fisher Scientific Inc; 2013 [cited 2016.04.15]. Available from: https://www.thermoscientific.com/content/dam/tfs/ATG/CMD/CMD%20Documents/Product%20Manuals%20%26%20Specifications/Chromatography%20Columns%20and%20Supplies/HPLC%20Columns/HPLC%20Columns%20(3um)/70762-DS-Acclaim-PepMap-23Feb10-LPN2234-01.pdf.

40. Hansen S, Pedersen-Bjergaard S, Rasmussen K. Introduction to Pharmaceutical Chemical Analysis: Mass Spectrometry. United Kingdom: John Wiley & Sons, Ltd; 2011. 231-59 p.

41. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. Science. 1989 Oct;246(4926):64-71.

42. Ho CS, Lam CW, Chan MH, Cheung RC, Law LK, Lit LC, et al. Electrospray ionisation mass spectrometry: principles and clinical applications. Clinical Biochemist. 2003 Feb;24(1):3-12.

43. Olsen JV, de Godoy LM, Li G, Macek B, Mortensen P, Pesch R, et al. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Molecular & Cellular Proteomics. 2005 Dec;4(12):2010-21.

44. Q Exactive™ [Internet]. USA: Thermo Fisher Scientific Inc; 2010 [cited 2016.04.17]. Available from: http://www.thermoscientific.com/content/dam/tfs/ATG/CMD/cmd-support/q-exactive/operations-and-maintenance/operators-manuals/1249360-Exactive-Operating-Rev-C.pdf.

45. Watson JT, Sparkman OD. Introduction to Mass Spectrometry. United Kingdom: John Wiley & Sons, Ltd; 2008. 173-228 p.

46. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999 Dec;20(18):3551-67.

47. Eng JK, Fischer B, Grossmann J, Maccoss MJ. A fast SEQUEST cross correlation algorithm. Journal of Proteome Research. 2008 Oct;7(10):4598-602.

48. Vaudel M, Barsnes H, Berven FS, Sickmann A, Martens L. SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. Proteomics. 2011 Mar;11(5):996-9.

49. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, et al. Open mass spectrometry search algorithm. Journal of Proteome Research. 2004 Sep-Oct;3(5):958-64.

50. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. Bioinformatics. 2004 Jun;20(9):1466-7.

51. Fenyo D, Beavis RC. A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. Analytical Chemistry. 2003 Feb;75(4):768-74.

52. Tabb DL, Fernando CG, Chambers MC. MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. Journal of Proteome Research. 2007 Feb;6(2):654-61.

53. Eng JK, Jahan TA, Hoopmann MR. Comet: an open-source MS/MS sequence database search tool. Proteomics. 2013 Jan;13(1):22-4.

54. Dorfer V, Pichler P, Stranzl T, Stadlmann J, Taus T, Winkler S, et al. MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. Journal of Proteome Research. 2014 Aug;13(8):3679-84.

55. Kim S, Pevzner PA. MS-GF+ makes progress towards a universal database search tool for proteomics. Nature Communications. 2014 Oct;5:5277.

56. Vaudel M, Burkhart JM, Zahedi RP, Oveland E, Berven FS, Sickmann A, et al. PeptideShaker enables reanalysis of MS-derived proteomics data sets. Nature Biotechnology. 2015 Jan;33(1):22-4.

57. Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, et al. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Communications in Mass Spectrometry. 2003 Sep;17(20):2337-42.

58. Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, et al. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Molecular & Cellular Proteomics. 2012 Apr;11(4):M111 010587.

59. Pierce BCA Protein Assay Kit [Internet]. USA: Thermo Fisher Scientific Inc; 2015 [cited 2015.09.03]. Available from: https://tools.thermofisher.com/content/sfs/manuals/MAN0011430_Pierce_BCA_Protein_Asy_UG.pdf.

60. Papaleo E, Fantucci P, De Gioia L. Effects of Calcium Binding on Structure and Autolysis Regulation in Trypsins. A Molecular Dynamics Investigation. Journal of Chemical Theory and Computation. 2005 Nov;1(6):1286-97.

61. Dickhut C, Radau S, Zahedi RP. Fast, efficient, and quality-controlled phosphopeptide enrichment from minute sample amounts using titanium dioxide. Methods in Molecular Biology. 2014 Apr;1156.

62. Pierce Biotechnology. TMT Mass Tagging Kits and Reagents [Internet] Rockford, USA: Thermo Scientific; 2012 [cited 2016.04.13]. Available from: http://www.biotech.cornell.edu/sites/default/files/uploads/Documents/Proteomics_protocols/Protocol%2017_TMT%20labeling.pdf.

63. Agilent Bond Elut OMIX [Internet]. USA: Agilent Technologies; 2011 [cited 2015.07.15]. Available from: https://www.agilent.com/cs/library/primers/public/5990-9049EN-Omix-Sept11-lo.pdf.

64. Rode H-J, Fernholz E. PhosSTOP - Phosphatase inhibitor cocktail tablets: A new reagent to freeze phosphorylation states Penzberg, Germany: Roche; 2007 [cited 2016.04.05.]. Available from: http://wolfson.huji.ac.il/purification/PDF/Protease_Inhibitors/ROCHE_PhosphInhibitCocktail.pdf.

65. Aryal UK, Ross AR. Enrichment and analysis of phosphopeptides under different experimental conditions using titanium dioxide affinity chromatography and mass spectrometry. Rapid Communications in Mass Spectrometry. 2010 Jan;24(2):219-31.

66. Dickhut C, Feldmann I, Lambert J, Zahedi RP. Impact of digestion conditions on phosphoproteomics. Journal of Proteome Research. 2014 Jun;13(6):2761-70.

67. Siepen JA, Keevil EJ, Knight D, Hubbard SJ. Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics. Journal of Proteome Research. 2007 Jan;6(1):399-408.

68. Phosphoric Acid - National Center for Biotechnology Information [Internet]. PubChem Compound Database 2004 [updated 2016.04.24; cited 2016.05.05]. Available from: http://pubchem.ncbi.nlm.nih.gov/compound/1004#section=Top.

69. Steen H, Jebanathirajah JA, Rush J, Morrice N, Kirschner MW. Phosphorylation Analysis by Mass Spectrometry: Myths, Facts, and the Consequences for Qualitative and Quantitative Measurements. Molecular & Cellular Proteomics. 2006 Jan;5(1):172-81.

70. Palumbo AM, Reid GE. Evaluation of gas-phase rearrangement and competing fragmentation reactions on protein phosphorylation site assignment using collision induced dissociation-MS/MS and MS3. Analytical Chemistry. 2008 Dec;80(24):9735-47.

71. Boersema PJ, Mohammed S, Heck AJ. Phosphopeptide fragmentation and analysis by mass spectrometry. Journal of Mass Spectrometry. 2009 Jun;44(6):861-78.

72. Palumbo AM, Tepe JJ, Reid GE. Mechanistic insights into the multistage gas-phase fragmentation behavior of phosphoserine- and phosphothreonine-containing peptides. Journal of Proteome Research. 2008 Feb;7(2):771-9.

73. Molina H, Horn DM, Tang N, Mathivanan S, Pandey A. Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America. 2007 Feb;104(7):2199-204.

74. Quan L, Liu M. CID,ETD and HCD Fragmentation to Study Protein Post-Translational Modification. Modern Chemistry & Applications. 2013 Nov;1(1).