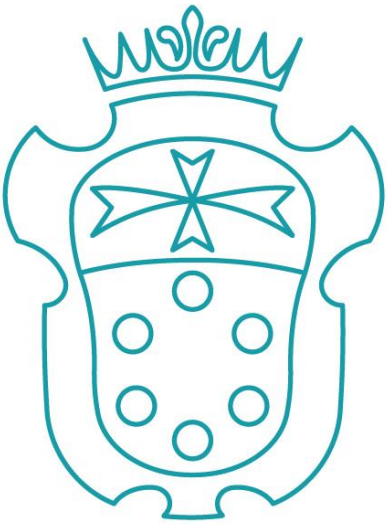
SCUOLA NORMALE SUPERIORE



PhD Thesis

in Biophysical Sciences

Development of high sensitivity and high specificity strategies for tissue microproteomics

Candidate

Davide Pellegrini

Supervisor

Prof. Liam McDonnell

Advisor

Prof. Gian Michele Ratto

Contents

Abstract

Mass-spectrometry based proteomics has become an indispensable tool for molecular biology and clinical research because of its ability to identify and quantify thousands of proteins. When combined with laser capture microdissection (LCM), MS-based proteomics may be used to investigate disease-associated changes in the proteome of specific tissue regions or cell populations. Such specificity is essential because different anatomical regions often have distinct and diverse functions and may behave differently under pathological conditions. However, the number of proteins that may be identified and quantified decreases with smaller sample amounts. Strict anatomical/cellular specification usually yields micrograms or submicrograms of protein, and thus ultrasensitive microproteomics protocols are required to analyze these small sample amounts while maintaining high proteome coverage.

Recent advances in liquid chromatography (LC) and MS equipment have improved the analysis of low sample amounts. The development of mass spectrometers with increased sequencing speed and ion transmission, have resulted in an increase in dynamic range and sensitivity. The advances in ultra high-pressure liquid chromatography (HPLC) has enabled the routine use of long columns (≥50 cm) with smaller internal diameter and smaller particle sized (< 5 µm) further increasing peptide separation resolution. However, the developments in LC-MS sensitivity have outpaced developments in sensitive sample preparation protocols.

In this PhD thesis, I will present my 4-years research on the development of ultransensitive microproteomics strategies for the molecular characterization of tissues. During my PhD I developed and optimized an ultrasensitive proteomic workflow for the analysis of small sample amounts, and I applied it to biomedical case studies.

First, I compared the digestion efficiency of the Filter-Aided Sample Preparation protocol (FASP) and the Single-Pot, Solid Phase-enhanced Sample Preparation protocol (SP3) with

the conventional urea based in-solution digestion (ISD) method for different amounts of HeLa cells. The SP3 protocol, based on the use of carboxylate coated paramagnetic beads, outperformed the FASP and ISD protocols for the analysis of small sample amounts, providing the identification of about 3000 protein groups from 1 µg of HeLa lysate. As a proof of principle, I applied the optimized SP3 protocol to the characterization of the brain of a mouse model of glioblastoma. Laser capture microdissection provided the specificity required to isolate different anatomical regions of the brain (healthy, border and tumor regions), while the SP3 digestion protocol provided the sensitivity required for the analysis of heterogeneous and complex samples.

To ensure accurate relative quantification and increase the proteome coverage I optimized in-solution and on-column Tandem Mass Tags (TMT) labeling and peptide fractionation of low sample amounts. Preliminary experiments revealed very low labeling efficiency when standard labeling conditions were applied to volume limited samples. Following an exhaustive optimized of in-solution and on-column TMT labeling the final conditions provided a TMT labeling efficiency (for 1 µg of HeLa digest) even greater than that obtained using standard methods on high sample amounts (25-50 µg of digest). Moreover, high-pH reversed phase fractionation increased proteome coverage by approximately 140% relative to single long gradient analyses.

One of the challenges of working with microdissected tissues or other samples characterized by low total protein content, is the need to estimate total protein content (essential knowledge for accurate quantitation). Previously adjacent sections were used just for the protein content estimation, which is non-ideal because tissue histologies may differ (especially for small pathological features with a specific histology). To address this, I developed a colorimetric assay for protein content estimation. I modified the microBCA protein assay to be able to measure proteins in just 1 µL and in the presence of the reagents commonly used in lysis buffers (as SDS, EDTA, EGTA, etc.). This modified microBCA assay allowed a reproducible quantification of the protein content of each individual sample down to a concentration of 15 ng/µL. The final optimized quantitative workflow for the proteomic analysis of tissue samples comprised laser capture microdissection, protein content estimation with the modified MicroBCA assay, SP3 digestion, TMT

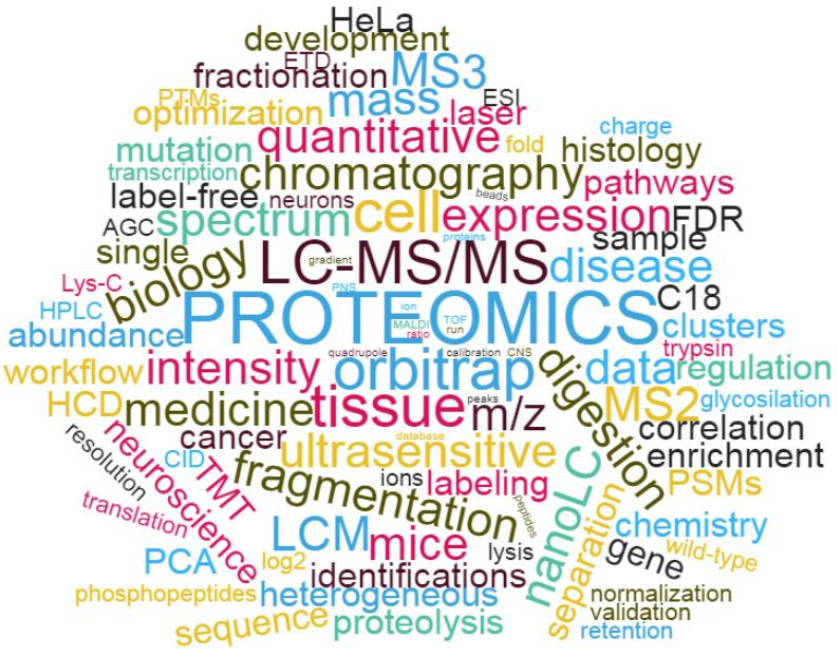
labeling, high-pH reversed phase fractionation and injection in a nanoLC system coupled with an Orbitrap Fusion mass spectrometer. As a proof of principle, I applied the optimized workflow to the proteomic characterization of mouse kidney substructures.

Finally, I applied the optimized workflow to the characterization of the central and peripheral nervous system of a mouse model of Krabbe disease (the Twitcher mouse). I compared the proteomes extracted from the corpus callosum, motor cortex and sciatic nerves of five Twitcher and five control wild type mice. The results on the proteome changes in the Twitcher mouse provided new insights into the molecular mechanisms of Krabbe disease showing neuroinflammation, activation of immune response, accumulation of lysosomal proteins, demyelination, membrane rafts disruption and reduced nervous system development.

Altogether, the microproteomic protocol developed during my PhD represents a powerful tool for the proteomic characterization of pathological tissues. Moreover, the research study on Krabbe disease represents the first in-depth proteomic characterization of the Twitcher mouse and a starting point for future functional experiments to study the pathogenesis of Krabbe disease and new possible therapies.

Chapter 1

General Introduction



1.1. Introduction to proteomics

Proteins are large biomolecules that are made of a chain of hundreds to thousands of amino acids. There are twenty common amino acids that can be combined to make a protein. The amino acid sequence of each protein is encoded in the DNA and helps determine the specific protein 3D structure and function.

All biological systems generate proteins via the transcription and translation processes, according to the central dogma of molecular biology [1]. During transcription the information stored in the gene's DNA is transferred to mRNA. In the translation step, the mRNA molecule interacts with ribosomes, which are specialized complexes able to read the mRNA nucleotide sequence. A family of tRNA's, which recognize the codon of three nucleotide bases that code for each amino acid, carry the amino acid to the ribosome where they are assembled into the protein one amino acid at a time.

Proteins represent the main functional machinery of cells, as they carry out the tasks specified by the information encoded in genes. The term "proteomics" was coined in 1995 and was defined as the large scale systematic study of all proteins, the proteome, of cells, tissues or organisms, in analogy to genomics [2][3]. This includes not only the identification and quantification of proteins, but also the determination of their cellular localization and the study of their interactions and functions [4]. The growth of proteomics was a result of the mapping and sequencing of the complete genome from different species, as they provided an essential resource for the identification of experimental protein sequences, namely the predicted amino acid sequences [5][6].

Proteins are the primary functional units of all cellular processes, and many aspects of the proteome cannot be predicted from the study of genes alone. For example, perturbations in protein expression levels, proteolytic processing that changes the length of the primary amino acid sequence, and post-translational modification states are all directly related to the molecular mechanisms of diseases [7]. Thus, proteomics has become a powerful tool for systems biology as it can provide a global picture of the molecular state of cells. This information is used by biologists to study how alterations of signaling pathways influence

cellular functions and by medical researchers to study the mechanisms of diseases and identify potential therapies.

The complexity of the proteome far exceeds that of the genome. There are approximately 20000 protein-coding genes in the human genome, which result in at least 20000 canonical proteins [8]. However many different protein variants can be produced from a single gene because of alternative splicing, single amino acid variants, post-translational modifications [9], and enzymatic cleavage (e.g. the protein proopiomelanocortin is a pre-protein that is enzymatically cleaved to produce more than 10 different neuropeptides). Protein expression levels also span a very wide dynamic range, $>10^6$. These characteristics, high complexity and high dynamic range, represent a significant analytical challenge for protein analysis, more so because of the absence of protein amplification techniques analogous to PCR. Mass spectrometry (MS) and affiliated technologies have provided the sensitivity, selectivity and throughput required for proteomics analysis. The rapid technological and methodological developments in the MS field have led to the initiation of the era of MS-based proteomics.

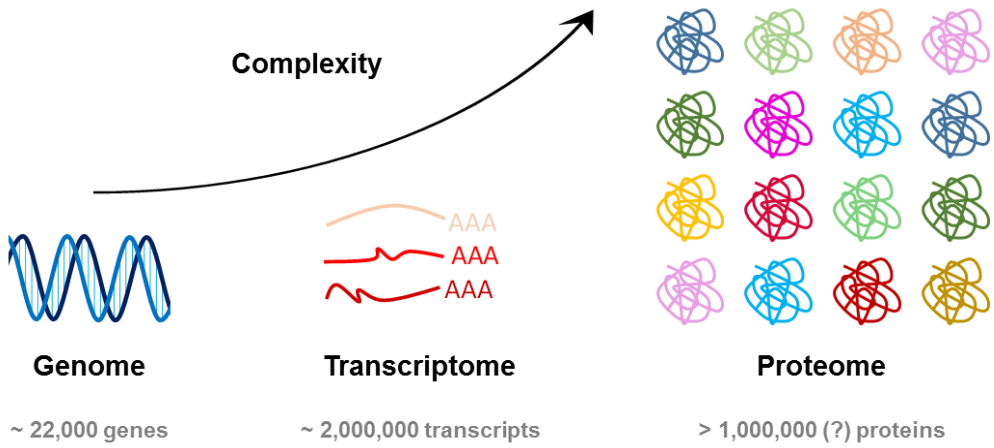


Figure 1.1. *Schematic comparing genome, transcriptome, and proteome diversity. The changes at the mRNA level and the different post-translational modifications lead to a vast increase in proteome complexity.*

A recent study by Wang *et al.* compared proteome and transcriptome abundances in 29 healthy human tissues [10]. The analysis revealed that hundreds of proteins could not be detected even for highly expressed mRNAs, that there are strong differences between mRNA and protein expression levels within and across tissues, and that protein expression can be more stable. The correlation between gene expression and protein expression is often poor, particularly in cancer. For example the Spearman correlation between gene expression and protein expression in pancreatic cancer is approximately 0.1, or barely above random [11]. It is thus essential to characterize which protein isoforms are expressed (protein identification) and at what levels (protein quantitation), particularly for biomedical applications.

1.2. Mass spectrometry-based proteomics

Mass spectrometry (MS) has established itself as the most powerful analytical tool for the analysis of proteins [12]. MS-based proteomics technologies have largely displaced earlier research tools for protein analysis, such as two-dimensional gel electrophoresis, because of their much higher sensitivity and their ability to deal with complex protein mixtures, as well as providing a much higher throughput [13].

MS-based proteomics relies on the availability of genome sequence databases and has benefited from rapid technical and conceptual advances in MS technology, separation sciences, ion chemistry, and bioinformatics. The widespread use of MS for protein analysis started in the late 1980s with the emergence of the macromolecular ionization techniques electrospray ionization (ESI) [14] and matrix-assisted laser desorption ionization (MALDI) [15] (for which Prof. John Fenn and Koichi Tanaka were jointly awarded the 2002 Nobel Prize in Chemistry). Since then, the MS instrumentation developments have led to a terrific increase in sensitivity, mass accuracy and throughput, enabling the in-depth characterization of proteomes and expanding the knowledge about protein structure, function, modification and protein dynamics [16].

Modern MS-based proteomics is overwhelmingly based on ESI instrumentation. These experiments can be divided into two different approaches for protein identification: bottom-up and top-down (Figure 1.2) [17]. In bottom-up proteomics, proteins are digested with a proteolytic enzyme, typically trypsin, and the resulting peptides mixture is analyzed by MS; in top-down proteomics MS analysis is performed directly on the intact proteins.

Bottom-up, also called shotgun, proteomics is the most widely used approach for protein identification and the characterization of complex samples, since peptides are more amenable to MS-based analysis than intact proteins [18]. However, only a fraction of the total peptide population is typically identified. A consequence of the limited protein coverage is reduced information about post-translational modifications (PTMs), and the loss of any combinatorial information about PTM's situated on different proteolytic peptides. In principle top-down mass spectrometry can provide almost complete sequence coverage of a protein, revealing their different modification states [19]. Indeed, the advantages of top-down proteomics are the potential access to the complete primary sequence of the protein and the ability to identify and locate PTMs. However, the long acquisition times required and the resulting very complex spectra make the coupling of MS with online chromatographic systems very challenging, limiting the applicability of top-down to isolated proteins or simple protein mixtures [20].

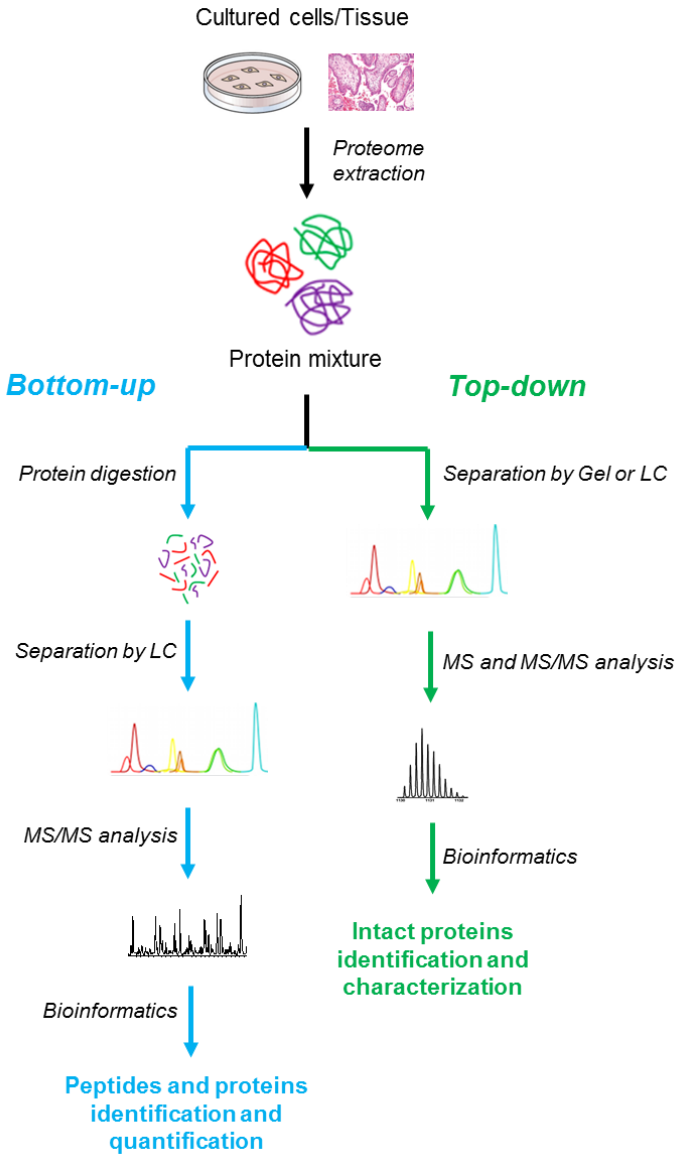


Figure 1.2. *Schematic of bottom-up and top-down approaches in MS-based proteomics.*

Middle-down proteomics strategies have been developed to combine the benefits of bottom-up and top-down approaches [21][22][23]. Middle-down methods are based on the proteolysis of proteins with enzymes that target less abundant amino acids than trypsin, to

generate longer peptides (20-100 amino acid residues), resulting in a less complex peptide mixture [24].

The choice of the proteomics approach ultimately depends on the aim of the experiment. Bottom-up proteomics is the method of choice for the qualitative and quantitative characterization of the proteomes from cells/tissues, while top-down is mainly used for the characterization of PTMs and proteoforms (*i.e.* all the different molecular forms in which the protein product of a single gene can be found).

A typical bottom-up proteomics workflow consists of four steps: (i) sample preparation, (ii) peptide separation, (iii) MS acquisition, and (iv) data analysis. The fine details of the analytical conditions are then dependent on the type of sample analyzed (*e.g.* in-vitro cell culture, whole organs, microdissected tissues) and the aim of the experiment (*e.g.* biomarker discovery, PTM analysis, quantitative analysis). The following paragraphs describe the different steps of a bottom-up proteomics workflow.

1.2.1. Sample preparation

Sample preparation is a critical step of the proteomics workflow as it influences the sensitivity, reproducibility and robustness of any experiment. In bottom-up experiments, all sample preparation protocols have five key elements: (i) protein extraction and denaturation, (ii) protein reduction, (iii) protein alkylation, (iv) protein digestion, (v) peptide clean-up/enrichment.

Protein extraction is usually performed using lysis buffers containing chaotropes (urea), surfactants (SDS, SDC), salts (TrisHCl, NaCl) or organic solvents (ACN, MeOH, TFE) that are able to disrupt the intra- and inter-protein interactions, promoting protein denaturation and solubilization. Cell lysis can also be performed by physical procedures such as ultrasonication [25], freeze-thaw cycles [26], and pressure-cycling based systems[27].

Proteins are then reduced and alkylated to open-up all disulfide bonds, thereby ensuring all protease cleavage sites are more accessible to the enzyme and the proteolytic digestion more complete. Protein reduction is usually performed using DTT or TCEP, which cleave the disulfide bonds [28], following which alkylation is performed with iodoacetamide or chloroacetamide to block the free–SH groups, preventing the reformation of disulfide bonds [29].

Protein digestion is performed using proteolytic enzymes that cleave proteins at specific amino acid residues. The cleavage involves hydrolysis of the amide bond before or after specific residues. The proteolytic enzyme trypsin cleaves the protein chain at the carboxyl side of arginine and lysine, and is the enzyme most often used for bottom-up proteomics on account of its robustness and its generation of peptides with a basic N-terminal and C-terminal (Arg/Lys). The basic termini are very useful during for identification of the proteolytic peptides using tandem mass spectrometry because it favors the detection of structural fragments that contain the N-terminus and, separately, the C-terminus.

Trypsin is often used in combination with Lys-C to increase the digestion efficiency of complex protein mixtures [30]. The resulting mixtures of proteolytic peptides are then purified or subjected to an enrichment protocol for the analysis of specific PTM's, e.g. protein phosphorylation. Enrichment is required for PTMs studies because many such peptides are frequently of lower abundance and would be masked by the non-modified peptides (without the enrichment step). Enrichment strategies are mainly focused on affinity chromatography or antibody-based immune-precipitation methods and the protocol of choice depends on the type of PTM of interest [31].

The last step of the proteomics workflow before LC-MS/MS analysis is the purification of the peptide mixture. Most of the buffers and chemicals used in the proteomics sample preparation workflow are not compatible with LC-MS/MS systems, as they may contain salts, detergents, and other factors that can clog the chromatographic system and/or induce ion suppression, adversely affecting robustness and sensitivity. Peptides can be desalted off-line using reverse-phase solid phase extraction cartridges[32], tips [33], magnetic beads [34], or online using trap columns before MS analysis [35]. Ù

1.2.2. Peptide separation

The proteolytic digestion of a cell lysate or a tissue extract generates hundreds of thousands of peptides that then need to be detected and characterized by MS analysis. The success of a proteomics study depends on several factors; the extent of proteome coverage determines the ability detect lower abundant proteins, and the sequence coverage of each identified protein is intrinsically related to the confidence of the identification and the ability to characterize proteoforms (e.g. due to PTM's) [36]. Liquid separations, such as liquid chromatography (LC) or capillary electrophoresis (CE), are used to distribute the high complexity of the peptide sample over the course of the LC/CE separation, and thereby reduce sample complexity.

Advances over the last decade in ultra high-pressure liquid chromatography (HPLC) have enabled the routine use of long columns (≥50 cm) with smaller internal diameter and smaller particle sizes (< 5 µm), which have increased peptide separation resolution [37]. Moreover, the development of nanoLC/low flow rate electrospray (nanoESI) platforms, which operates at flow rates in the range of 20-400 nL/min, have led to a significant increase in sensitivity, enabling the analysis of sub-microgram sample amounts [38].

The most common methods for peptide separation include reversed phase (RP) and strong cation exchange (SCX) chromatography, which separate peptides based on hydrophobicity and the number of positive charges, respectively [39]. Two dimensional LC may be used to further increase the separation capabilities of the LC step [40] and is most frequently performed using SCX for the first dimension and RP as the second dimension, because of the lower resolving power of SCX and its high orthogonality to RP. SCX is generally used to pre-fractionate the samples prior RP-LC-MS analysis, this offline pre-fractionation results in the conversion of a single highly complex peptide sample into multiple less complex peptide fractions, which are then individually analyzed by RP-LC and results in a significant increase in the number of peptide identifications. This setup takes advantage of the superior chromatographic resolving power of RP stationary phase and the compatibility

of the eluent used in RP-LC with ESI, allowing it to be directly interfaced to the MS for instant peptide mass measurements [41]. Alternatively, two RP separations at different pH has established itself as a powerful and convenient option for in-depth LC-MS/MS based proteomics: a simple change in pH has been shown to provide identical orthogonality to SCX-RP [42].

The choice of separation technology ultimately depends on the goal of the experiment, alternative forms of chromatography routinely used in proteomics include size exclusion chromatography (SEC), affinity chromatography (AC), hydrophilic interaction chromatography (HILIC) [43], and capillary electrophoresis (CE). CE has recently emerged as a technique complementary to RP-LC as the separation is based on differences in charge density, resulting in a high degree of orthogonality to RP-LC [44]. CE has been found to be particularly suited for the analysis of PTMs that can alter charge state, including phosphorylation and glycosylation [45][46]. Furthermore, the ultra-low flow capabilities of CE have been exploited for ultra-sensitive analysis, including single cell analysis; the ultra-low flow means minimal sample consumption, typically in the order of several nanoliters, and offers a high resolving power on short time scales [47][48].

1.2.3. MS detection

The mass spectrometer (MS) consists of an ionization source, a mass analyzer and a detector. The ion source generates gas-phase (pseuodo) molecular ions from the peptides in sample; ii) the mass analyzer separates these ions according to their *m/z* ratio; iii) the detector is used to record the abundance of each *m/z*-separated ion and convert that information into digitized signals. The following paragraphs describe in more detail each of these parts of the MS workflow.

1.2.3.1. Ionization

Proteomics is overwhelmingly based on the ionization techniques as electrospray ionization (ESI) [14] and, to a much lesser extent, matrix-assisted laser desorption

ionization (MALDI) [49]. Both ESI and MALDI can generate intact pseudomolecular ions (i.e. $[M+nH]^{n+}$) with minimal fragmentation [50].

In MALDI the peptides/proteins are mixed in solution with an excessive of a matrix, which is usually a small organic acid that absorbs UV light (*e.g.* sinapinic acid, α-cyano-4-hydroxycinnamic acid, 2,5-dihydroxybenzoic acid). Upon solvent evaporation, a solid deposit is obtained containing peptide-doped matrix crystals. The sample is then irradiated with a pulsed UV laser, pulse length <10 ns. Absorbance of the laser pulse energy, with pulse lengths shorter than the time required for energy redistribution throughout the matrix, leads to an almost-explosive phase change in the laser-super heated matrix. This phase change leads to a directed motion of matrix molecules, matrix clusters, and peptides/proteins into the gas phase (Figure 1.3.A). Singly protonated ions ($[M+H]^+$) are formed in the hot plume by matrix-to-analyte proton transfers [51]. MALDI suffers from some disadvantages such as the impossibility to couple it online with a chromatographic system and strong dependence on the sample preparation methods [50][12]. The main advantages of MALDI are the low complexity of the mass spectra (MALDI mass spectra are characterized by singly charged ions), the relatively high compatibility with salts and detergents and the possibility to perform imaging experiments using focused laser beams [52][53].

ESI was first introduced in 1989 by Fenn and workers as a soft ionization technique to analyze intact proteins [14]. ESI is characterized by little/ no fragmentation, and even weak non-covalent interactions may be preserved [54]. ESI is performed by applying a potential difference of 1-6 kV, under atmospheric pressure, between a liquid passing through a thin needle (*e.g.* the emitter) and a counter electrode (the MS inlet). The high electric field induces charge accumulation at the liquid emerging from the emitter. The forces applied to the ions in the liquid by the strong electric field leads to the formation of a Taylor cone, from which a jet of liquid emerges that releases charged droplets (Figure 1.3.B). These droplets are subject to solvent evaporation during their passage to and through the MS inlet; during solvent evaporation, the surface charge density increases until the droplet becomes unstable. Ostensibly one could consider the "Rayleigh" limit when coulombic repulsion exceeds surface tension, but in reality asymmetric fission events occur before the

Rayleigh limit because the coulombic repulsion exacerbates any deviation from spherical droplets, and favors the emergence of highly charged jets that generate smaller, highly charged, satellite droplets. This process continues until gas-phase protonated peptide/protein ions molecules are formed, which are then separated in the mass analyzer.

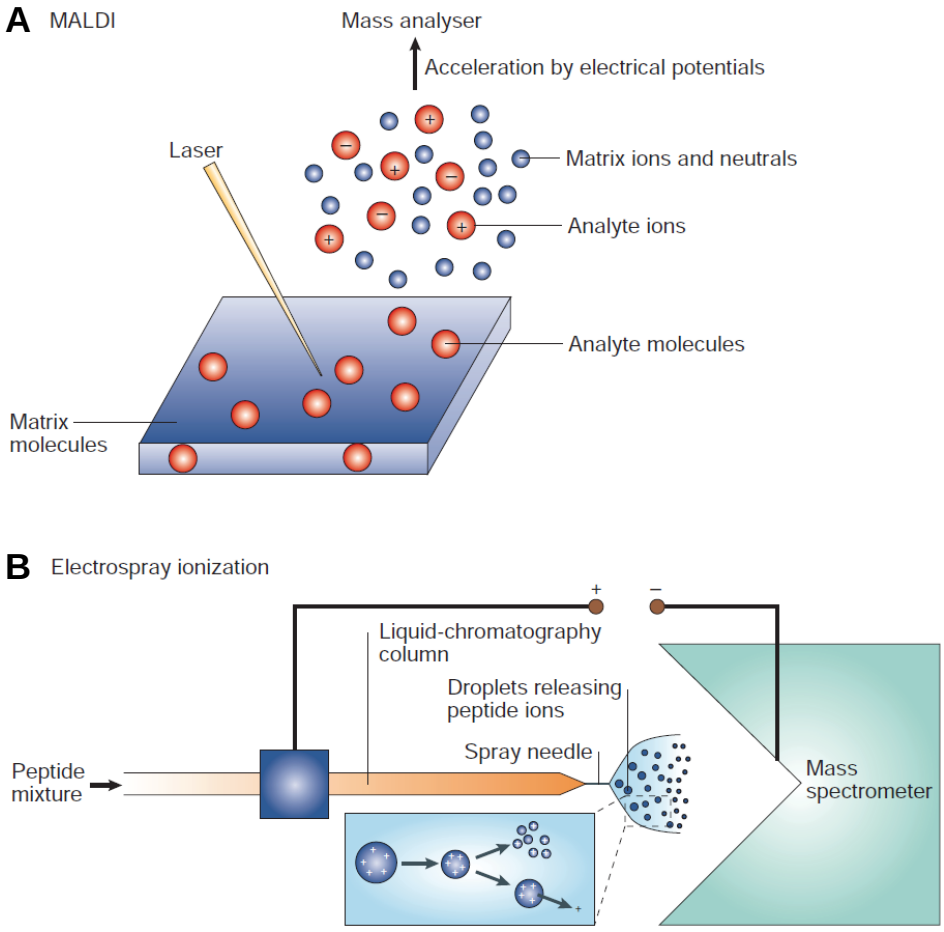


Figure 1.3. *The most used soft ionization techniques used in proteomics experiments. (A) Matrix Assisted Laser Desorption Ionization (MALDI); (B) Electrospray ionization (ESI). Figure adopted from* [55].

The pseudomolecular ions (i.e. $[M+nH]^{n+}$) generated by ESI are characterized by different charge states. The basic sites on the peptide/protein can bind the protons present in the final droplets. The charge state distribution reflects the number of basic sites available in the peptide/protein (larger proteins have higher charge states), the number of protons available in the final droplets (a stochastic phenomenon and so resulting in a distribution of charge states), as well as the thermodynamics of protonation. This latter aspect is important because it dictates that the degree of ionization for any specific peptide/protein depends on how well it competes with all other peptides/proteins and other cofactors present in the final droplets of the ESI process. The presence of many different peptides/proteins means only the more abundant and basic peptides/proteins are likely to be ionized to a sufficient degree to be analyzed with sufficient sensitivity. Similarly, cofactors that compete for charge lower sensitivity.

The multiple charging of ESI has numerous advantages, one in particular is that it enables the analysis of large molecules since the *m/z* values of the $[M+nH]^{n+}$ ions fall in the range of where many mass analyzers offer higher performance (e.g. mass resolution). Another significant advantage of ESI is that it can be directly coupled with capillary electrophoresis and UPLC systems, enabling the analysis of small and large molecules of a wide range of polarities in complex mixtures [56]. The development of micro-ESI and nano-ESI sources, in which the flow rate is in the range of μL/min and nL/min respectively, have greatly improved sensitivity due to further improved ionization efficiency [57][58][59].

1.2.3.2. Mass analyzers

Mass analyzers separate ions according to their mass-to-charge ratio (*m/z)*. Mass analyzers can be further divided into two categories, separate-ions-in-space (e.g. Quadrupole and Time of Flight mass analyzers) or separate-ion-in-time (e.g. Ion Trap, Orbitrap, and Fourier Transform Ion Cyclotron Resonance analyzers). Each mass analyzer has its advantages and limitations, and the choice of mass analyzer is ideally dependent on the goal of the experiment, but in practical purposes is dependent on the availability of these high cost instruments (to purchase and maintain). In proteomics higher speed, higher mass

resolution, and higher mass accuracy equate with increased depth of coverage, i.e. more protein identifications. For targeted proteomics, selectivity and sensitivity are the key performance metrics. Mass analyzers are now routinely combined in the so-called hybrid instruments characterized by increased versatility, throughput and parallel MS and MS/MS detection.

The most common mass analyzers used in proteomics are summarized below.

Quadrupole (Q) – The Quadrupole analyzer acts like an ion filter using the stability of the ion trajectories in oscillating electric fields to separate ions [50][60]. The quadrupole analyzer is made up of four parallel metal rods that have a bias DC applied to all four rods, a fixed DC potential difference between adjacent rods, and a radio frequency (RF) time-varying potential difference between adjacent rods (Figure 1.4A) [61]. Because of the oscillating RF voltage, each pair of rods will have successively positive and then negative potential. These oscillations result in complex ion trajectories dependent on the *m/z* ratio of the ions. The stability of the ions is best summarized with the Matheiu parameters and the stability diagram (Figure 1.4B):

$$a_u = \frac{8zeU}{mr_0^2\Omega^2} \qquad q_u = \frac{4zeV}{mr_0^2\Omega^2}$$

where Ω is the frequency of the RF potential, e = the charge of an electron; U = fixed DC voltage between adjacent rods, V = zero-to-peak RF voltage; m = mass of the ion, z= charge state of the ion, and r_0 = the field radius of the quadrupole (i.e. the radius of the circle that touches the inner most edge of all four rods).

Ion motion is only stable within defined ranges of the Mathieu parameters a_u and q_u. Figure 1.4B shows the stability diagram; only ions whose Mathieu parameters are within the stable region, highlighted in grey in the figure, will be transmitted through the quadrupole. For a given ratio of DC and RF voltages the quadrupole can be operated at close to the a_0

and q_0 point, in which only a single mass is transmitted through the quadrupole. As the Mathieu parameters a_u and q_u are both inversely proportional to *m/z* ratio, different ions can be transmitted just by changing the magnitude of the DC and RF voltages.

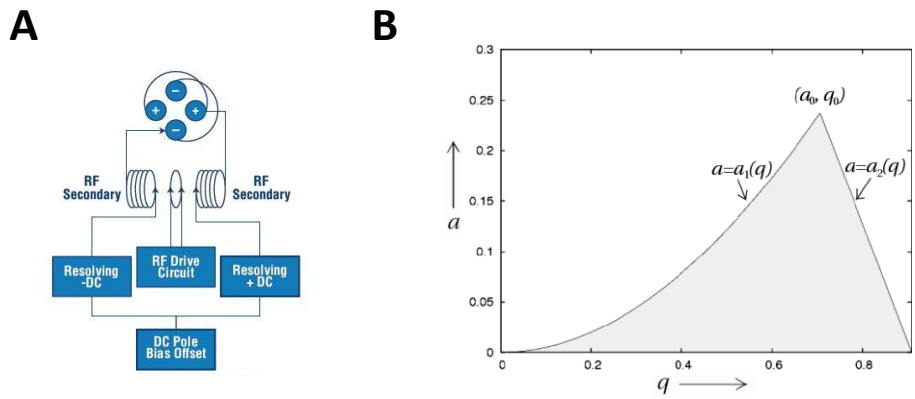


Figure 1.4. *Quadropole mass analyzer. A) Schematic of typical quadrupole power supply connections. B) The Mathieu stability diagram in two dimension . Figures adopted from* [61].

A quadrupole mass scan involves ramping the DC and RF potentials so that ions of increasing *m/z* ratio are transmitted through the analyzer (scanning mode). For targeted analysis the quadrupole is fixed at specific DC and RF potentials (selected ion monitoring mode). Selected ion monitoring enables signal averaging to increase sensitivity.

Quadrupoles are often used in series for tandem MS (*e.g.* triple quadrupole (QqQ) MS), or in combination with other mass analyzers for hybrid MS (*e.g.* quadrupole time-of-flight (Q-ToF) MS) [62]. In these types of mass spectrometers a precursor *m/z* is selected in the first quadrupole (Q0), which is then accelerated into a second quadrupole (Q1) containing an inert collision gas, the resulting fragment ions are then analyzed in the third mass analyzer (Q2 or ToF).

Time-of-Flight (ToF) – In ToF analyzers the *m/z* ratio of an ion is determined by its flight time through a drift tube of specified length that is under vacuum (Figure 1.5A) [63]. Ions are introduced into the ToF analyzer as a pulse, receiving the same initial kinetic energy.

As ions travel along the field free drift tube they are separated in time because differences in *m/z* ratio translate as a difference in velocity, and ultimately as a difference in flight time:

$$t = \frac{d}{\sqrt{2U}}\sqrt{\frac{m}{qe}} \qquad or\ more\ simply\ t \propto \sqrt{\frac{m}{q}}$$

where *d* is the length of the field free region, *U* is the accelerating voltage through which all ions were accelerated prior to entering the drift tube, *m* is the ion's mass, *q* the ion's charge state, and *e* is the charge of an electron.

For a given kinetic energy the time taken for an ion to reach the detector is proportional to the square root of the *m/z* ratio [16]. In an ideal scenario all ions would be accelerated through the same acceleration potential and have identical kinetic energy, plus there would be zero dispersion through the mass spectrometer so all ions of identical *m/z* would have an identical time-of-flight. In reality, owing to the discrete size of the ion cloud in the mass spectrometer all ions (of the same *m/z*) do not experience the same acceleration voltage; furthermore thermal distributions in velocity, imperfections in the electric fields of the ion-optics, and repulsion between ions, all combine to result in ions of identical *m/z* having a narrow range of flight times, leading to an experimental mass resolution that is substantially lower than the theoretical limit. Delayed extraction and the reflectron represent two distinct technological solutions to compensate for the range in flight times and thereby improve mass resolution, Figure 1.5B [64]. The reflectron consists of a shaped electric field gradient that reflects the ions back along the flight tube; fine tuning of the reflectron voltage gradient is used to time-focus the ions at the detector.

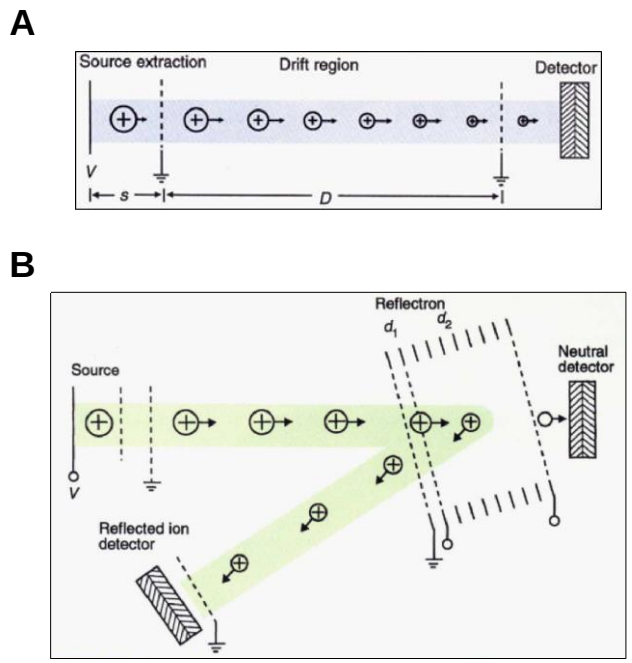


Figure 1.5. *Time-of-flight mass analyzer. Basic components of a linear ToF (A) and a reflectron ToF (B) mass spectrometer. Figure adopted from* [63].

Linear Ion Trap (LIT) – The LIT analyzer (also referred to as a 2D ion trap) consists of a quadrupole with two end-cap electrodes that can be used to axially constrain ions inside the quadrupole (Figure 1.6.). In the LIT the ions are confined in the radial dimension by the quadrupolar field of an RF-only quadrupole (a_u=0, so ions of wide *m/z* range have stable trajectories, see Quadrupole equation) and in the axial dimension by the axial potential well created using the cap electrodes. After injection of ions into the LIT, and collisional cooling using an inert background gas [65], the ions are focused onto the axis of the LIT and within the potential well of the axial field. The frequency of the secular motion of an ion in a quadrupole LIT is directly proportional to its *m/z* ratio and the applied RF voltage. The secular motion of selected ions or a range of ions can be excited by the use of an auxiliary waveform, which can be used to excite the ions out of the trap for detection, or to excite their motion collision induced dissociation (CID) [66].

The fundamentals of ion motion in a 2D LIT are very similar to the 3D ion trap (in which the ions are focused at a focal point), but has the distinct advantage of reduced space-charge effects as the ions are distributed along an axis, which results in a higher charge capacity (at least 400-fold higher) and increased sensitivity [67][68].

LIT analyzers are frequently used for the sequential trapping, isolation, fragmentation, and MS analysis of fragment ions (tandem in time). This process can be repeated so that fragments from the first MS/MS experiment can be isolated and subjected to an additional round of MS/MS (commonly referred to as MS^3). In principal the LIT enables MS^n but the vast majority experiments are performed with either MS^2 or MS^3.

A dual pressure LIT has been developed to increase trapping/fragmentation efficiency and sequencing speed [69]. The dual pressure LIT consist of two identical LITs separated by a center lens. The first LIT efficiently traps and fragments ions at high background gas pressures, whereas the second LIT realizes extremely fast scan speeds at reduced pressure.

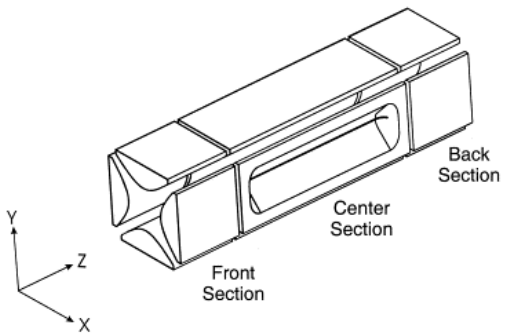


Figure 1.6. *Schematics of a Linear Ion Trap analyzer. Basic design of a 2D linear ion trap. Figure adopted from* [68].

Ion Cyclotron Resonance Mass Spectrometry (ICR-MS) was first applied in the mid 1950's [70] to measure very small mass differences at very high precision and remained mostly an academic tool until the application of FT methods in the early 1970's [71][72].

Today FTICR is synonymous with ultra high mass resolution, and finds widespread use in applications that demand super high mass resolution [16][73][74].

The mass analyzer of an FTICR consists of the ICR cell, which is situated inside a superconducting magnet (Figure 1.7.). The pressure inside the cell is kept very low, in the range of 10^{-10} to 10^{-11} mbar, in order to enable super high mass resolution measurements (collisions with residual gas molecules undermine such measurements).

Ions in a magnetic field move in a circular motion under the influence of the Lorentz force. This motion, the cyclotron motion, has a frequency proportional to the magnetic field and the *m/z* ratio. At room temperature the radius of the cyclotron motion is small, and too small to detect. The application of an RF field at an ion's cyclotron frequency, i.e. resonant excitation, will excite the ion's cyclotron motion and increase its radius. The higher orbits of the excited cyclotron motion can be detected by amplifying the image current imparted into a pair of detection plates; the Fourier transform of the free induction decay of the time-domain signal generates the frequency spectrum, from which the mass spectrum is then calculated using a calibration [50].

Hybrid FT-ICR instruments have been developed to provide parallelization of precursor mass selection and fragment ions detection. The major drawback of FTICR-MS is represented by the high purchase, operational and maintenance costs [75].

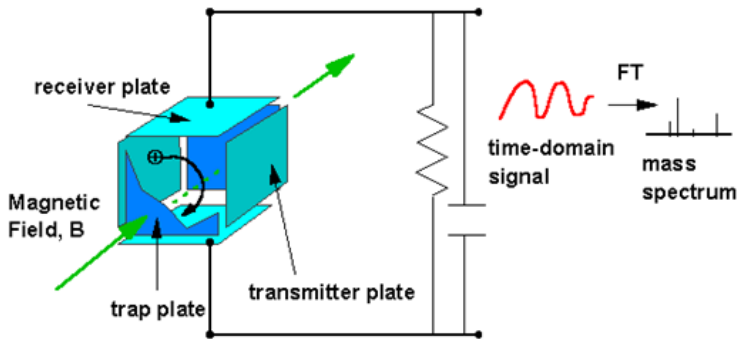


Figure 1.7. *Schematic of the FTICR-MS principle. Figure modified from* [76].

Orbitrap (OT) – The OT analyzer is an electrostatic ion trap that uses the Fourier transform to obtain simultaneously determine mass spectra spanning a wide mass range. The Orbitrap was designed by Makarov in 2000 [77] and is based on the Kingdon trap [78] that was first described in 1923. The first commercial instrument incorporating an OT analyzer (a LTQ Orbitrap mass spectrometer) was introduced by Thermo Electron (now Thermo Fisher Scientific) in 2005 [79]. The OT consists of an outer barrel-like electrode cut into two equals parts separated by a dielectric and a central spindle-like electrode (Figure 1.8.A) [80]. An electrostatic voltage is applied to the central electrode while the outer electrode is at ground potential. The ions are injected tangentially through a small gap in the external electrode and are trapped in the OT because of the balance between the electrical attraction to the inner electrode and the centrifugal forces. The electrostatic field has a quadro-logarithmic potential distribution that is obtained by the DC voltage applied between the outer and inner electrodes and the geometry of the trap. Stable ion trajectories combine rotation around the inner electrode and axial harmonic oscillation along its axis, resulting in intricate spirals (Figure 1.8.B). The frequency of the axial motion, which is independent of the kinetic energy and position of the ions, is proportional to the *m/z* ratio [77][81]. Axial oscillations are detected by the image current induced on the outer electrode, which is then converted to a voltage and amplified to produce a time domain transient. The transient signal is Fourier-transformed to the frequency domain from which the *m/z* spectrum is calculated (Figure 1.8.C) [82].

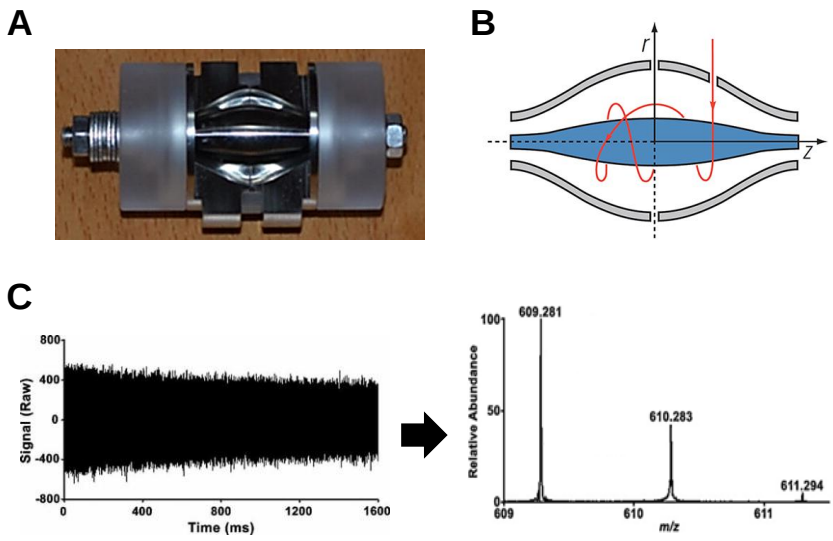


Figure 1.8. *(A) Cross-section of an Orbitrap mass analyzer (adopted from* [80]*). (B) Schematic of the Otbitrap geometry. Ions move in spiral trajectories around the inner electrode and their image current is recorded at the outer electrode resulting in a time domain transient. (C) The transient is Fourier-transformed resulting in the frequency domain spectrum, which is then converted to the m/z domain using a calibration routine to obtain the mass spectrum (adopted from* [82]*).*

Since the release of the first OT-based mass spectrometer, the expanding needs of the MS community have continuously inspired and contributed to the further development of OT-based technologies. OT has become the most popular mass analyzer in the proteomics field because its high resolution and mass accuracy enable accurate analysis of complex and low abundant samples. The Orbitrap Fusion (released in 2013) and the Orbitrap Fusion Lumos (released in 2015) represent the state of the art technology in the proteomics field. The Orbitrap Fusion (the mass spectrometer used for all the proteomics experiments in this thesis) is a hybrid mass spectrometer that combines quadrupole, dual pressure linear ion trap and Orbitrap technologies (Figure 1.9.), providing ultrahigh resolution (routinely 60-120K, going up to 500K) and high mass accuracy (<2 ppm) [83]. Ion isolation can be performed with either the quadrupole or the LIT, whereas fragment ions can be detected at high resolution in the orbitrap or at low resolution in the LIT. The hybrid configuration of this mass spectrometer enables fully parallelized operations as ion can be concurrently

isolated with one mass analyzer (quardupole) and detected with the other two mass analyzers (Orbitrap and LIT). The parallelization of operations maximize the use of the available ion current, resulting in increased amounts high quality data [79].

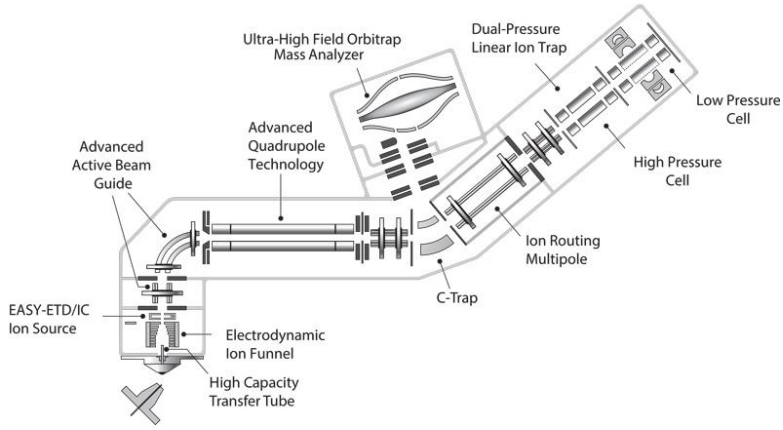


Figure 1.9. *Schematic of the Orbitrap Fusion mass spectrometer. Figure adopted from* [79].

1.2.3.3. Detectors

Ion detection consists of the conversion of the m/z-separated ions into a usable signal. As explained above, detection of ions in FTICR or OT mass spectrometers is achieved by means of a pair of plates within the mass analyzer region that detect the image current of the ions. For all other mass analyzers the most common detector is the electron multiplier (EM). A typical EM consists of a continuous dynode with a concave metal surface that catches the ions arriving from the mass analyzer. When an ion hits the surface it causes secondary electrons to be released from the atoms in the surface layer. These electrons are accelerated towards the exit of the dynode and they hit the wall of the dynode causing the emission of more and more electrons (Figure 1.10.), resulting in an exponential signal increase. Typical signal amplifications, or gains, are 10^4-10^7 for a single EM. The cascade of electrons is finally collected by a metal anode and the current is converted to a potential and measured [50].

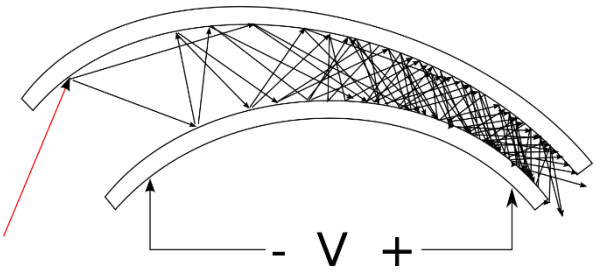


Figure 1.10. *Schematic of the continuous dynode electron multiplier.*

1.2.3.4. Peptide fragmentation

The determination of the *m/z* values of precursor ions is not sufficient to provide the unambiguous identification of large molecules. This is particularly true in proteomics as the *m/z* values of proteins and peptides are not unique, especially as mass accuracy decreases. Tandem MS provides information about the primary structure (amino acid sequence) of peptides, and thus their identity. In tandem MS, a peptide ion is isolated and fragmented. A mass spectrum of the resulting fragments (MS2 spectrum) is then acquired to get the *m/z* values of the fragment ions, which correspond to the peptide backbone cleavage fragments [55]. Figure 1.11. shows the Roepstorff–Fohlmann–Biemann nomenclature of the different peptide fragment ions [84].

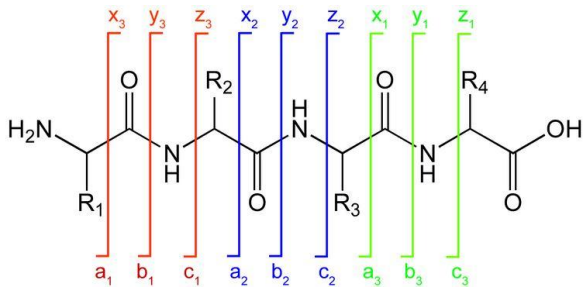


Figure 1.11. *Chemical structure of a peptide, together with the Roepstorff–Fohlmann–Biemann nomenclature of the peptide backbone fragments. Fragment ions generated by collision-induced dissociation (CID) are predominantly b- and y-ions. Fragment ions generated by electron-transfer dissociation (ETD) are typically c- and z-ions.*

Several fragmentation methods are now routinely used in proteomics research. Collision-induced dissociation (CID) or collision-activated dissociation (CAD) is a widely used fragmentation technique in proteomics [85]. In CID peptide ions are accelerated by applying an electric field to increase their kinetic energy and allowed to collide with background gas molecules (typically nitrogen, argon, helium or xenon). With each collision part of the ion's centre-of-mass kinetic energy is converted into internal energy; if when the internal energy exceeds a bond energy it results in peptide fragmentation. The site of bond breakage and the fragmentation mechanisms are described by the mobile proton model [86]. According to this model, protons are initially localized at the most basic residues of a peptide and upon activation they migrate across the peptide backbone. The proximity of a proton to the backbone amino groups weakens the C-N bonds and, in combination with the increase in internal energy, results in peptide fragmentation and the generation of mainly *b*- and *y*- fragment ions. CID can be subcategorized into resonant-excitation and beam-type CID. Resonant-excitation CID is performed in an ion trap (IT-CID) and only a small amount of energy is deposited onto the peptide with each collision; sustained excitation of the ions and collisions with the background gas over the millisecond time scale results in the slow heating (or slow increase of internal energy) of the ions until they fragment.

Beam type CID is applied in dedicated collision cells in tandem quadrupoles (QqQ, Q-TOF) or multipoles (HCD). This type of fragmentation is conceptually similar to ion trap CID; however, the ions are accelerated to higher collision energies and then collide with the background collision gas under single collision conditions. Accordingly, the internal energy needed to effect fragmentation is deposited in a single instance and fragmentation occurs on a much faster time scale (microseconds).

The slow heating nature and longer timescales of IT-CID means favor lower energy reactions, including slower rearrangement type reactions, are accessed. Conversely beam type CID is characterized by fewer chemical rearrangements. Furthermore the impartation of higher internal energy with a single collision frequently activates additional reaction resulting in more informative MS2 spectra. As a result, HCD fragmentation is particularly suited for the analysis of peptides modified with labile PTMs (such as phosphorylation) and with isobaric tags (as TMT) [87][88], which would otherwise be preferentially lost using IT-CID.

In 2004, Electron Transfer Dissociation (ETD) was proposed as a new fragmentation method for the sequencing of peptides modified with PTMs [89]. In ETD electrons are transferred from a gaseous anion (*e.g.* fluoranthene) to a peptide cation when they are in close proximity in an ion trap. The electron transfer results in charge reduction and fragmentation of peptide backbone bonds at the amine ($N\text{-}C_\alpha$) bond to produce c- and z-type ions (Figure 1.11.) [90]. Labile modification such as phosphorylation and glycosylation are left intact during the ETD fragmentation process, and so ETD has enjoyed wide success for the analysis of PTMs [91][92][93]. ETD fragmentation efficiency is proportional to the peptide's charge density and increases as the peptide charge state increases, resulting in the generation of only a few fragments for doubly charged peptides [94]. Since CID fragmentation is less efficient on peptides with more than 2 charges, ETD has developed as a complementary fragmentation technique to increase protein and peptide sequence coverage [95][96].

1.2.4. Data analysis

The final output of a single MS/MS experiment is a mass spectrum that contains the m/z ratios of the fragment ions. These spectra need to be interpreted to identify the peptides and consequently the proteins from which the peptides originate. In shotgun proteomics experiments tens of thousands of MS/MS spectra are produced and thus manual interpretation is not feasible. Instead interpretation is performed using automated search engines that compare the MS/MS spectra against a protein database. The protein database is first converted into a peptide database using the cleavage specificity of the enzyme used to proteolyze the proteins, from which theoretical MS/MS spectra are calculated using the peptide cleavage specificity of the tandem mass spectrometry technique employed. The scoring algorithms seek to determine which peptide (theoretical MS/MS spectrum) best matches the experimental data. Several algorithms have been developed for automated spectral matching, including SEQUEST [97], MASCOT [98] and MS AMANDA [99]. Different search engines use different algorithms for the peptide-spectrum match (PSM) but the search pipeline is similar [100]. The first step is the generation of a list of theoretical peptides from an *in silico* digestion of the protein database of interest. The masses of the theoretical peptides are then compared to the experimental precursor masses (measured in the MS1 acquisition) and only the peptide candidates matching within a given mass tolerance are kept for further analysis. The next step is the *in silico* fragmentation of the theoretical peptides that matched with the experimental precursor ions in order to generate theoretical MS/MS spectra. The theoretical MS/MS spectra are compared with the experimental spectra and a similarity score is calculated. The peptide with the highest score is the top ranking assignment for the MS/MS spectrum. Only when the top ranking assignment is significantly more likely than any other assignment is it considered a peptide spectrum match (PSM) [101]. From PSMs the original proteins present in the sample are then inferred [102]. However, when tens of thousands of spectra must be compared the risk of random matches is likely. Several statistical methods have been developed to estimate the rate of false discovery (FDR) [103], the most common of which is to perform a second search with a decoy database (derived from reversed or scrambled protein sequences) [104][105]. All the matches to the decoy database are

considered as false positives and the FDR can be estimated by dividing the number of PSMs that matched the decoy database by the total number of PSMs. The FDR can be lowered by applying stringent filters on the PSM parameters (*e.g.* matching score, peptide length, elution time and peptide charge). Generally, a 1% FDR is accepted for large-scale proteomics data [106].

1.2.5. Quantitative proteomics

A qualitative analysis of a proteome represents the identification of the proteins present in the sample. However, it is often also desired to know the amount of each protein, or the relative amount of proteins present in different samples. Protein quantitation is thus crucial as it allows the comparison of protein expression profiles between different conditions. The main challenge of MS-based protein quantitation is that the intensity of peptide ions is not solely dependent on the peptide concentration in the sample. The signals also depend on the sample preparation workflow (*e.g.* protein, extraction, digestion efficiency, peptide solubility) and on the ionization/transport/detection efficiency of each ion. The peptide ionization efficiency in turn depends on chemical-physical properties of the peptides (*e.g.* size, basicity, hydrophobicity), the composition of the solvent and the presence of co-eluting compounds [107]. Several strategies have been developed to allow accurate and reproducible quantitative proteomics experiments. The quantification methods can be divided into label-free and stable isotope labeling (Figure 1.12.) [107][108][109].

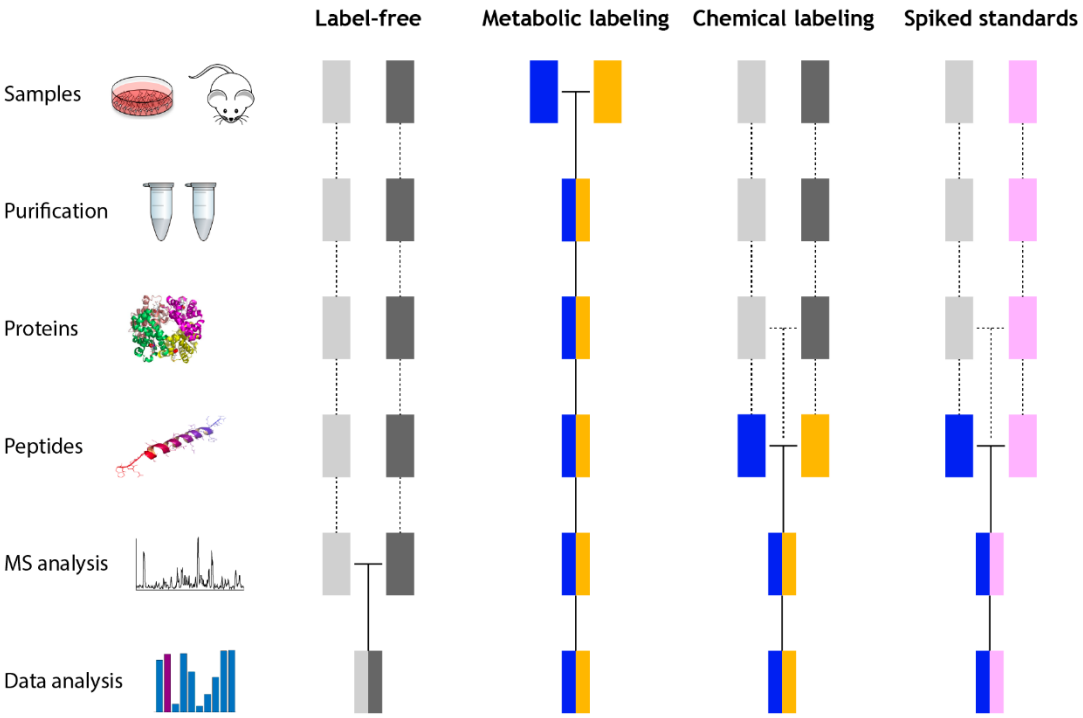


Figure 1.12. *Schematic of the different quantitative MS-based proteomics workflows. Grey boxes indicate un-labeled samples. Blue, orange and pink boxes indicate labeled samples. Horizontal lines indicate when samples are combined. Dashed lines indicate steps of the workflow that can be sources of quantification variance, as samples are processed separately. Figure modified from* [110].

1.2.5.1. Label free quantification

Label-free quantification methods aim to determine the relative amount of peptides between different LC-MS/MS runs. Quantification is based either on (i) the MS signal intensity of peptides (precursor signal intensity) [111] or (ii) the number of spectral matches (spectral counting) [112].

In the precursor signal intensity method, the ion chromatograms of every peptide are extracted from the LC-MS/MS runs and the MS peak areas are integrated over the chromatographic time scale. Since the quantification is performed at the MS1 level, high

mass accuracy mass spectrometers are preferred in order to reduce isobaric ions (same nominal m/z) adversely affecting quantitation precision.

The spectral counting strategy is based on the empirical observation that the number of acquired MS/MS spectra increases as the amount of the corresponding peptide/protein in the sample increases. Thus, the relative quantification of proteins across different samples is performed by comparing the number of MS/MS spectra for each protein. Even if this approach is less affected by interfering molecules, as it is based on MS2 spectra, it is still controversial because it does not measure any physical property of peptides and it assumes a linear response for every protein. However, the extraction and ionization efficiency of different peptides is influenced by the peptide chemical-physical properties and their chromatographic behavior, resulting in a large number of MS/MS spectra being required for reliable relative quantification [107].

Overall, the main advantage of label-free quantification methods is that no additional sample preparation steps are required. Moreover, there are no costs for labeling reagents and there is no limit to the number of samples that can be compared. However, label-free methods are the least accurate and reproducible because variations in sample preparation and LC conditions affect the peptide intensities determined by the mass spectrometer. This variability between different LC runs can be alleviated by applying normalization methods, whether matching median intensities, using spike-in standards as well as algorithms for peak alignment [113][114].

1.2.5.2. Stable isotope labeling quantification

Protein quantitation via labeling is achieved by introducing stable isotopes into/onto the proteins. Since different isotopes have the same chemical properties but different masses, different samples labeled with different isotopes can be discriminated by MS analysis. The advantage of these methods is that samples can be combined before the LC-MS/MS analysis, thereby excluding all sources of errors due to sample preparation and LC-MS/MS analysis [115]. Stable isotopes can be introduced at the protein or peptide level in three ways: (i) metabolically, (ii) chemically or (iii) by using spiked synthetic peptides (Figure 1.12.).

Metabolic labeling – Stable isotopes are introduced into proteins by metabolic labeling during cell growth. Cells, plants or animals are grown with a heavy isotope enriched food source that contains ^{15}N and/or ^{13}C instead of the naturally abundant ^{14}N and ^{12}C isotopes [116][117]. In this way, the synthetized proteins will contain (< 90%) heavy peptides. The most popular metabolic labeling method in proteomics is Stable Isotope Labeling by Amino acids in Cell culture (SILAC) [118]. SILAC allows a quantitative comparison of the protein expression profiles between two (or three) different populations of cells. One cell population is fed with normal growth medium. The other cell population is fed with a growth medium containing $^{13}C_6$-arginine and/or $^{13}C_6$-lysine, so that all tryptic peptides (the protease trypsin cleaves c-terminal to arginine and lysine) will carry at least one isotopically labeled amino acid. After metabolic labeling the two cell populations are combined, the proteins extracted and digested, and the resulting peptides analyzed by LC-MS/MS. Relative quantification is provided by the ratio between the intensities of the *heavy* and the *light* peptides. Despite the high quantification accuracy, the use of SILAC is limited by the costs and time for maintaining the model system and by the limited number of samples that can be compared [119].

Chemical labeling – Stable isotopes can be added onto post-biosynthetic samples that are not suited for metabolic labeling (*e.g.* tissues and body fluids) via chemical labeling. Chemical labels are targeted towards reactive sites on proteins or peptides, generally the side chains of lysine or cysteine. Several chemical labeling strategies have been developed for quantitative proteomics experiments. The first chemical labeling method to be reported for labeling samples at the protein level was Isotope-Coded Affinity Tags (ICAT) [120]. In ICAT two samples are labeled with a *light* or *heavy* tag that was designed to react with free cysteine residues (i.e. cysteines not involved in a disulfide bonds). After labeling, the two proteomes are combined and digested together in order to remove the variability related to protein digestion. The main limitation of ICAT is that only a small fraction of peptides (only those containing a cysteine) can be used for quantification.

Another group of chemical labels target primary amines (free N-termini of peptides and lysine residues). Tandem Mass Tags (TMT) [121] and Isobaric Tags for Relative and

Absolute Quantification (iTRAQ) [122] are examples of these and are arguably the most popular chemical labels in proteomics. Both iTRAQ and TMT tags consist of a mass reporter region, a cleavable linker, a mass normalization region and a protein reactive group. The mass reporter and mass normalizer regions incorporate stable isotopes in multiple configurations so that each mass reporter's mass can be resolved in an MS2 or MS3 spectrum, while the masses of the intact tags are identical. As a result, the peptides originating from different samples will be identical at the MS1 level, leading to less complex spectra with higher precursor intensities. The higher precursor intensities favor peptide identification. Upon precursor fragmentation using LIT-CID the isobaric tags are cleaved and the reporter ion intensities can be used for relative quantification [123]. The most important advantage of TMT and iTRAQ methods is the high multiplexing capability. Up to 16 samples can be simultaneously compared using TMT (TMT 16-plex) and up to 8 samples with iTRAQ (iTRAQ 8-plex). It was demonstrated that accuracy and precision are affected by contaminating isobaric ions that are co-isolated and co-fragmented together with the target precursor ions. This results in a compression of the reporter ion ratios, as the ratios tends towards unity due to interfering peptides originating from proteins with unchanged expression in the different samples [124]. The ratio compression is reduced by using MS3 acquisition methods, in which the reporter ion quantification is performed at the MS3 level, because the likelihood of co-eluting peptides having identical mass and fragment ions at the MS2 level also having identical mass (which is then selected for MS3) is much lower [125][126].

Stable isotope-labeled synthetic standards (SIS) – were originally described in 1983 for MS-based quantification of peptides [127] and can be used for relative and absolute quantification of proteins/peptides. Absolute quantification is achieved by the addition of a known amount of SIS to a protein digest and by comparing the MS signals to the endogenous peptides in the sample [110]. In contrast to metabolic and chemical labeling strategies, in which relative quantification is performed for a large number of proteins, SIS based methods require a synthetic peptide for every target peptide, and thus is normally used to quantify only a few proteins of interest, for instance for the validation of potential

biomarkers [128]. Moreover, SIS methods do not require any labeling of the sample, resulting in a simpler sample preparation workflow but with a much higher accuracy and precision with respect to label-free methods. SIS can be introduced into the samples before LC-MS/MS analysis (AQUA method) [129] or before protein digestion using biosynthesized heavy proteotypic peptide concatemers (QconCAT) [130] or synthetic heavy full length proteins (PSAQ) [131].

References

1. Crick, F. (1970) Central Dogma of Molecular Biology. *Nature*, **227** (5258), 561–563.

2. Abbott, A. (1999) A post-genomic challenge: learning to read patterns of protein synthesis . *Nature*, **402** (6763), 715–716.

3. Wilkins, M.R., Pasquali, C., Appel, R.D., Ou, K., Golaz, O., Sanchez, J.C., Yan, J.X., Gooley, A.A., Hughes, G., Humphery-Smith, I., Williams, K.L., and Hochstrasser, D.F. (1996) From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Arnino Acid Analysis. *Nat. Biotechnol.*, **14** (1), 61–65.

4. Fields, S. (2001) Proteomics in Genomeland. *Science*, **291** (5507), 1221–1224.

5. Patterson, S.D., and Aebersold, R.H. (2003) Proteomics: the first decade and beyond. *Nat. Genet.*, **33**, 311–323.

6. Tyers, M., and Mann, M. (2003) From genomics to proteomics. *Nature*, **422** (6928), 193–197.

7. Graves, P.R., and Haystead, T.A.J. (2002) Molecular Biologist's Guide to Proteomics. *Microbiol. Mol. Biol. Rev.*, **66** (1), 39–63.

8. Lander, E.S., et al., and Morgan, M.J. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409** (6822), 860–921.

9. Aebersold, R., et al., and Zhang, B. (2018) How many human proteoforms are there? *Nat. Chem. Biol.*, **14**, 206–214.

10. Wang, D., Eraslan, B., Wieland, T., Hallström, B., Hopf, T., Zolg, D.P., Zecha, J., Asplund, A., Li, L., Meng, C., Frejno, M., Schmidt, T., Schnatbaum, K., Wilhelm, M., Ponten, F., Uhlen, M., Gagneur, J., Hahne, H., and Kuster, B. (2019) A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.*, **15** (2), e8503.

11. Kosti, I., Jain, N., Aran, D., Butte, A.J., and Sirota, M. (2016) Cross-tissue Analysis of Gene and Protein Expression in Normal and Cancer Tissues. *Sci. Rep.*, **6** (1), 24799.

12. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature*, **422** (6928), 198–207.

13. Pandey, A., and Mann, M. (2000) Proteomics to study genes and genomes. *Nature*, **405** (6788), 837–846.

14. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science (80-.).*, **246** (4926), 64–71.

15. Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T., and Matsuo, T. (1988) Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.*, **2** (8), 151–153.

16. Domon, B., and Aebersold, R. (2006) Mass Spectrometry and Protein Analysis. *Science (80-.).*, **312** (5771), 212–217.

17. Chait, B.T. (2006) Mass Spectrometry: Bottom-Up or Top-Down? *Science (80-.).*, **314** (5796), 65–66.

18. Zhang, Y., Fonslow, B.R., Shan, B., Baek, M.-C., and Yates, J.R. (2013) Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.*, **113** (4), 2343–2394.

19. Kelleher, N.L. (2004) Top-Down Proteomics. *Anal. Chem.*, **76** (11), 196–203.

20. Angel, T.E., Aryal, U.K., Hengel, S.M., Baker, E.S., Kelly, R.T., Robinson, E.W., and Smith, R.D. (2012) Mass spectrometry-based proteomics: existing capabilities and future directions. *Chem. Soc. Rev.*, **41** (10), 3912–3928.

21. Yergey, J.A., Cotter, R.J., Heller, D., and Fenselau, C. (1984) Resolution requirements for middle molecule mass spectrometry. *Anal. Chem.*, **56** (12), 2262–2263.

22. Wu, C., Tran, J.C., Zamdborg, L., Durbin, K.R., Li, M., Ahlf, D.R., Early, B.P., Thomas, P.M., Sweedler, J. V, and Kelleher, N.L. (2012) A protease for "middle-down" proteomics. *Nat. Methods*, **9** (8), 822–824.

23. Garcia, B.A. (2010) What does the future hold for top down mass spectrometry? *J. Am. Soc. Mass Spectrom.*, **21** (2), 193–202.

24. Pandeswari, B., and Sabareesh, V. (2019) Middle-down approach: a choice to sequence and characterize proteins/proteomes by mass spectrometry. *RSC Adv.*, **9** (1), 313–344.

25. Araújo, J.E., Oliveira, E., Kouvonen, P., Corthals, G.L., Lodeiro, C., Santos, H.M., and Capelo, J.L. (2014) A journey through PROTEOSONICS. *Talanta*, **121**, 71–80.

26. Specht, H., Harmange, G., Perlman, D.H., Emmott, E., Niziolek, Z., Budnik, B., and Slavov, N. (2018) Automated sample preparation for high-throughput single-cell proteomics. *bioRxiv*, 399774.

27. Freeman, E., and Ivanov, A.R. (2011) Proteomics under Pressure: Development of Essential Sample Preparation Techniques in Proteomics Using Ultrahigh Hydrostatic Pressure. *J. Proteome Res.*, **10** (12), 5536–5546.

28. Getz, E.B., Xiao, M., Chakrabarty, T., Cooke, R., and Selvin, P.R. (1999) A Comparison between the Sulfhydryl Reductants Tris(2-carboxyethyl)phosphine and Dithiothreitol for Use in Protein Biochemistry. *Anal. Biochem.*, **273** (1), 73–80.

29. Hustoft, H.K. (2012) A Critical Review of Trypsin Digestion for LC-MS Based Proteomics, in *Integrative Proteomics* (eds.Malerod, H.), IntechOpen, Rijeka, pp. 73–92.

30. Zhang, Y., Fonslow, B.R., Shan, B., Baek, M., and Yates, J.R. (2013) Protein Analysis by Shotgun / Bottom-up Proteomics. *Chem. Re*, **113** (4), 2343–2394.

31. Zhao, Y., and Jensen, O.N. (2009) Modification-specific proteomics: Strategies for characterization of post-translational modifications using enrichment techniques. *Proteomics*, **9** (20), 4632–4641.

32. de Graaf, E.L., Pellegrini, D., and McDonnell, L.A. (2016) Set of Novel Automated Quantitative Microproteomics Protocols for Small Sample Amounts and Its Application to Kidney Tissue Substructures. *J. Proteome Res.*, **15** (12), 4722–4730.

33. Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.*, **2** (8), 1896–1906.

34. Hughes, C.S., Foehr, S., Garfield, D.A., Furlong, E.E., Steinmetz, L.M., and Krijgsveld, J. (2014) Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.*, **10** (757), 1–15.

35. Shen, Y., Tolić, N., Masselon, C., Paša-Tolić, L., Camp, D.G., Hixson, K.K., Zhao, R., Anderson, G.A., and Smith, R.D. (2004) Ultrasensitive Proteomics Using High-Efficiency On-Line Micro-SPE-NanoLC-NanoESI MS and MS/MS. *Anal. Chem.*, **76** (1), 144–154.

36. Manadas, B., Mendes, V.M., English, J., and Dunn, M.J. (2010) Peptide fractionation in proteomics approaches. *Expert Rev. Proteomics*, **7** (5), 655–663.

37. Köcher, T., Swart, R., and Mechtler, K. (2011) Ultra-High-Pressure RPLC Hyphenated to an LTQ-Orbitrap Velos Reveals a Linear Relation between Peak Capacity and Number of Identified Peptides. *Anal. Chem.*, **83** (7), 2699–2704.

38. Shen, Y., Zhao, R., Berger, S.J., Anderson, G.A., Rodriguez, N., and Smith, R.D. (2002) High-Efficiency Nanoscale Liquid Chromatography Coupled On-Line with Mass Spectrometry Using Nanoelectrospray Ionization for Proteomics. *Anal. Chem.*, **74** (16), 4235–4249.

39. Link, A.J. (2002) Multidimensional peptide separations in proteomics. *Trends Biotechnol.*, **20** (12), 8–13.

40. Horvatovich, P., Hoekman, B., Govorukhina, N., and Bischoff, R. (2010) Multidimensional chromatography coupled to mass spectrometry in analysing complex proteomics samples. *J. Sep. Sci.*, **33** (10), 1421–1437.

41. Di Palma, S., Hennrich, M.L., Heck, A.J.R., and Mohammed, S. (2012) Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *J. Proteomics*, **75** (13), 3791–3813.

42. Yang, F., Shen, Y., Camp, D.G., and Smith, R.D. (2012) High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. *Expert Rev. Proteomics*, **9** (2), 129–134.

43. Shi, Y., Xiang, R., Horváth, C., and Wilkins, J.A. (2004) The role of liquid chromatography in proteomics. *J. Chromatogr. A*, **1053** (1), 27–36.

44. Scriba, G.K.E. (2016) Separation of Peptides by Capillary Electrophoresis, in *Capillary Electrophoresis. Methods in Molecular Biology* (eds.Schmitt-Kopplin, P.), Springer New York, New York, NY, pp. 365–391.

45. Ludwig, K.R., Sun, L., Zhu, G., Dovichi, N.J., and Hummon, A.B. (2015) Over 2300 Phosphorylated Peptide Identifications with Single-Shot Capillary Zone Electrophoresis-Tandem Mass Spectrometry in a 100 min Separation. *Anal. Chem.*, **87** (19), 9532–9537.

46. Amon, S., Zamfir, A.D., and Rizzi, A. (2008) Glycosylation analysis of glycoproteins and proteoglycans using capillary electrophoresis-mass spectrometry strategies. *Electrophoresis*, **29** (12), 2485–2507.

47. Voeten, R.L.C., Ventouri, I.K., Haselberg, R., and Somsen, G.W. (2018) Capillary Electrophoresis: Trends and Recent Advances. *Anal. Chem.*, **90** (3), 1464–1481.

48. Lombard-Banek, C., Moody, S.A., Manzini, M.C., and Nemes, P. (2019) Microsampling Capillary Electrophoresis Mass Spectrometry Enables Single-Cell Proteomics in Complex Tissues: Developing Cell Clones in Live Xenopus laevis and Zebrafish Embryos. *Anal. Chem.*, **91** (7), 4797–4805.

49. Tanaka, K. (2003) The Origin of Macromolecule Ionization by Laser Irradiation (Nobel Lecture). *Angew. Chemie Int. Ed.*, **42** (33), 3860–3870.

50. de Hoffmann, E., and Stroobant, V. (2007) *Mass Spectrometry: Principles and*

Applications, John Wiley & Sons Ltd, Chichester, England.

51. Knochenmuss, R., and Zenobi, R. (2003) MALDI Ionization: The Role of In-Plume Processes. *Chem. Rev.*, **103** (2), 441–452.

52. Dreisewerd, K. (2014) Recent methodological advances in MALDI mass spectrometry. *Anal. Bioanal. Chem.*, **406** (9), 2261–2278.

53. Cornett, D.S., Reyzer, M.L., Chaurand, P., and Caprioli, R.M. (2007) MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. *Nat. Methods*, **4** (10), 828–833.

54. Przybylski, M., and Glocker, M.O. (1996) Electrospray Mass Spectrometry of Biomacromolecular Complexes with Noncovalent Interactions—New Analytical Perspectives for Supramolecular Chemistry and Molecular Recognition Processes. *Angew. Chemie Int. Ed. English*, **35** (8), 806–826.

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

56. Whitehouse, C.M., Dreyer, R.N., Yamashita, M., and Fenn, J.B. (1985) Electrospray interface for liquid chromatographs and mass spectrometers. *Anal. Chem.*, **57** (3), 675–679.

57. Wilm, M.S., and Mann, M. (1994) Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *Int. J. Mass Spectrom. Ion Process.*, **136** (2), 167–180.

58. Juraschek, R., Dülcks, T., and Karas, M. (1999) Nanoelectrospray—more than just a minimized-flow electrospray ionization source. *J. Am. Soc. Mass Spectrom.*, **10** (4), 300–308.

59. El-Faramawy, A., Siu, K.W.M., and Thomson, B.A. (2005) Efficiency of Nano-Electrospray Ionization. *J. Am. Soc. Mass Spectrom.*, **16** (10), 1702–1707.

60. Wolfgang, P., and Steinwedel, H. (1953) Notizen: Ein neues Massenspektrometer ohne Magnetfeld. *Zeitschrift für Naturforsch. A*, **8**, 448.

61. Extrel CMS, LLC. (2019, July 23). Using Graphical Tools to Understand Quadrupole Theory. AZoM. Retrieved on March 22, 2020 from https://www.azom.com/article.aspx?ArticleID=10996.

62. Chernushevich, I. V, Loboda, A. V, and Thomson, B.A. (2001) An introduction to quadrupole–time-of-flight mass spectrometry. *J. Mass Spectrom.*, **36** (8), 849–865.

63. Cotter, R.J. (1992) Time-of-flight mass spectrometry for the structural analysis of biological molecules. *Anal. Chem.*, **64** (21), 1027–1039.

64. Mamyrin, B.A. (1994) Laser assisted reflectron time-of-flight mass spectrometry. *Int. J. Mass Spectrom. Ion Process.*, **131**, 1–19.

65. Chernushevich, I. V, and Thomson, B.A. (2004) Collisional Cooling of Large Ions in Electrospray Mass Spectrometry. *Anal. Chem.*, **76** (6), 1754–1760.

66. Xu, F., Dang, Q., Dai, X., Fang, X., Wang, Y., Ding, L., and Ding, C.-F. (2016) Characteristics of Ion Activation and Collision Induced Dissociation Using Digital Ion Trap Technology. *J. Am. Soc. Mass Spectrom.*, **27** (8), 1351–1356.

67. Blackler, A.R., Klammer, A.A., MacCoss, M.J., and Wu, C.C. (2006) Quantitative Comparison of Proteomic Data Quality between a 2D and 3D Quadrupole Ion Trap. *Anal. Chem.*, **78** (4), 1337–1344.

68. Schwartz, J.C., Senko, M.W., and Syka, J.E.P. (2002) A two-dimensional quadrupole ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.*, **13** (6), 659–669.

69. Olsen, J. V, Schwartz, J.C., Griep-Raming, J., Nielsen, M.L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E.R., Senko, M., Makarov, A., Mann, M., and Horning, S. (2009) A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. *Mol. Cell. Proteomics*, **8** (12), 2759–2769.

70. Hipple, J.A., Sommer, H., and Thomas, H.A. (1949) A Precise Method of Determining the Faraday by Magnetic Resonance. *Phys. Rev.*, **76** (12), 1877–1878.

71. Cooley, J.W., and Tukey, J.W. (1965) An Algorithm for the Machine Calculation of Complex Fourier Series. *Math. Comput.*, **19** (90), 297–301.

72. Comisarow, M.B., and Marshall, A.G. (1974) Fourier transform ion cyclotron resonance spectroscopy. *Chem. Phys. Lett.*, **25** (2), 282–283.

73. Schaub, T.M., Hendrickson, C.L., Horning, S., Quinn, J.P., Senko, M.W., and Marshall, A.G. (2008) High-Performance Mass Spectrometry: Fourier Transform Ion Cyclotron Resonance at 14.5 Tesla. *Anal. Chem.*, **80** (11), 3985–3990.

74. Smith, D.F., Podgorski, D.C., Rodgers, R.P., Blakney, G.T., and Hendrickson, C.L. (2018) 21 Tesla FT-ICR Mass Spectrometer for Ultrahigh-Resolution Analysis of Complex Organic Mixtures. *Anal. Chem.*, **90** (3), 2041–2047.

75. Bogdanov, B., and Smith, R.D. (2005) Proteomics by FTICR mass spectrometry: Top down and bottom up. *Mass Spectrom. Rev.*, **24** (2), 168–200.

76. Sage, E. (2013) New concept of mass spectrometer based on arrays of resonating nanostructures.

77. Makarov, A. (2000) Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.*, **72** (6), 1156–1162.

78. Kingdon, K.H. (1923) A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Phys. Rev.*, **21** (4), 408–418.

79. Eliuk, S., and Makarov, A. (2015) Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu. Rev. Anal. Chem.*, **8** (1), 61–80.

80. Zubarev, R.A., and Makarov, A. (2013) Orbitrap Mass Spectrometry. *Anal. Chem.*, **85** (11), 5288–5296.

81. Hu, Q., Noll, R.J., Li, H., Makarov, A., Hardman, M., and Graham Cooks, R. (2005) The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.*, **40** (4), 430–443.

82. Perry, R.H., Cooks, R.G., and Noll, R.J. (2008) Orbitrap mass spectrometry: Instrumentation, ion motion and applications. *Mass Spectrom. Rev.*, **27** (6), 661–

699.

83. Senko, M.W., Remes, P.M., Canterbury, J.D., Mathur, R., Song, Q., Eliuk, S.M., Mullen, C., Earley, L., Hardman, M., Blethrow, J.D., Bui, H., Specht, A., Lange, O., Denisov, E., Makarov, A., Horning, S., and Zabrouskov, V. (2013) Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. *Anal. Chem.*, **85** (24), 11710–11714.

84. Roepstorff, P., and Fohlman, J. (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.*, **11** (11), 601.

85. McLuckey, S.A. (1992) Principles of collisional activation in analytical mass spectrometry. *J. Am. Soc. Mass Spectrom.*, **3** (6), 599–614.

86. Paizs, B., and Suhai, S. (2005) Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.*, **24** (4), 508–548.

87. Boersema, P.J., Mohammed, S., and Heck, A.J.R. (2009) Phosphopeptide fragmentation and analysis by mass spectrometry. *J. Mass Spectrom.*, **44** (6), 861–878.

88. Olsen, J. V, Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. (2007) Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods*, **4** (9), 709–712.

89. Syka, J.E.P., Coon, J.J., Schroeder, M.J., Shabanowitz, J., and Hunt, D.F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.*, **101** (26), 9528–9533.

90. Pitteri, S.J., Chrisman, P.A., and McLuckey, S.A. (2005) Electron-Transfer Ion/Ion Reactions of Doubly Protonated Peptides: Effect of Elevated Bath Gas Temperature. *Anal. Chem.*, **77** (17), 5662–5669.

91. Kim, M.-S., and Pandey, A. (2012) Electron transfer dissociation mass spectrometry in proteomics. *Proteomics*, **12** (4-5), 530–542.

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

93. Mikesh, L.M., Ueberheide, B., Chi, A., Coon, J.J., Syka, J.E.P., Shabanowitz, J., and Hunt, D.F. (2006) The utility of ETD mass spectrometry in proteomic analysis. *Biochim. Biophys. Acta - Proteins Proteomics*, **1764** (12), 1811–1822.

94. Good, D.M., Wirtala, M., McAlister, G.C., and Coon, J.J. (2007) Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry. *Mol. Cell. Proteomics*, **6** (11), 1942–1951.

95. Swaney, D.L., McAlister, G.C., and Coon, J.J. (2008) Decision tree–driven tandem mass spectrometry for shotgun proteomics. *Nat. Methods*, **5** (11), 959–964.

96. Brunner, A.M., Lössl, P., Liu, F., Huguet, R., Mullen, C., Yamashita, M., Zabrouskov, V., Makarov, A., Altelaar, A.F.M., and Heck, A.J.R. (2015) Benchmarking Multiple Fragmentation Methods on an Orbitrap Fusion for Top-down Phospho-Proteoform Characterization. *Anal. Chem.*, **87** (8), 4152–4158.

97. Eng, J.K., McCormack, A.L., and Yates, J.R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.*, **5** (11), 976–989.

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

99. Dorfer, V., Pichler, P., Stranzl, T., Stadlmann, J., Taus, T., Winkler, S., and Mechtler, K. (2014) MS Amanda, a Universal Identification Algorithm Optimized for High Accuracy Tandem Mass Spectra. *J. Proteome Res.*, **13** (8), 3679–3684.

100. Ivanov, M. V, Levitsky, L.I., Lobas, A.A., Tarasova, I.A., Pridatchenko, M.L., Zgoda, V.G., Moshkovskii, S.A., Mitulovic, G., and Gorshkov, M. V (2015) Peptide identification in “shotgun” proteomics using tandem mass spectrometry: Comparison of search engine algorithms. *J. Anal. Chem.*, **70** (14), 1614–1619.

101. Granholm, V., and Käll, L. (2011) Quality assessments of peptide–spectrum matches in shotgun proteomics. *Proteomics*, **11** (6), 1086–1093.

102. Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002) Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Anal. Chem.*, **74** (20), 5383–5392.

103. Aggarwal, S., and Yadav, A.K. (2016) False Discovery Rate Estimation in Proteomics BT - Statistical Analysis in Proteomics (eds.Jung, K.), Springer New York, New York, NY, pp. 119–128.

104. Higgs, R.E., Knierman, M.D., Bonner Freeman, A., Gelbert, L.M., Patil, S.T., and Hale, J.E. (2007) Estimating the Statistical Significance of Peptide Identifications from Shotgun Proteomics Experiments. *J. Proteome Res.*, **6** (5), 1758–1767.

105. Elias, J.E., and Gygi, S.P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods*, **4** (3), 207.

106. Weatherly, D.B., Atwood, J.A., Minning, T.A., Cavola, C., Tarleton, R.L., and Orlando, R. (2005) A Heuristic Method for Assigning a False-discovery Rate for Protein Identifications from Mascot Database Search Results. *Mol. Cell. Proteomics*, **4** (6), 762–772.

107. Smith, B.J., Martins-de-Souza, D., and Fioramonte, M. (2019) A Guide to Mass Spectrometry-Based Quantitative Proteomics BT - Pre-Clinical Models: Techniques and Protocols (eds.Guest, P.C.), Springer New York, New York, NY, pp. 3–39.

108. Schubert, O.T., Röst, H.L., Collins, B.C., Rosenberger, G., and Aebersold, R. (2017) Quantitative proteomics: challenges and opportunities in basic and applied research. *Nat. Protoc.*, **12** (7), 1289–1294.

109. Cox, J., and Mann, M. (2011) Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. *Annu. Rev. Biochem.*, **80** (1), 273–299.

110. Ong, S.-E., and Mann, M. (2005) Mass spectrometry–based proteomics turns quantitative. *Nat. Chem. Biol.*, **1** (5), 252–262.

111. Chelius, D., and Bondarenko, P. V (2002) Quantitative Profiling of Proteins in Complex Mixtures Using Liquid Chromatography and Mass Spectrometry. *J. Proteome Res.*, **1** (4), 317–323.

112. Liu, H., Sadygov, R.G., and Yates, J.R. (2004) A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Anal. Chem.*, **76** (14), 4193–4201.

113. Välikangas, T., Suomi, T., and Elo, L.L. (2016) A systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief. Bioinform.*, **19** (1), 1–11.

114. Christin, C., Hoefsloot, H.C.J., Smilde, A.K., Suits, F., Bischoff, R., and Horvatovich, P.L. (2010) Time Alignment Algorithms Based on Selected Mass Traces for Complex LC-MS Data. *J. Proteome Res.*, **9** (3), 1483–1495.

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

116. Oda, Y., Huang, K., Cross, F.R., Cowburn, D., and Chait, B.T. (1999) Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci.*, **96** (12), 6591–6596.

117. Jiang, H., and English, A.M. (2002) Quantitative Analysis of the Yeast Proteome by Incorporation of Isotopically Labeled Leucine. *J. Proteome Res.*, **1** (4), 345–350.

118. Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002) Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics*, **1** (5), 376–386.

119. Chen, X., Wei, S., Ji, Y., Guo, X., and Yang, F. (2015) Quantitative proteomics

using SILAC: Principles, applications, and developments. *Proteomics*, **15** (18), 3175–3192.

120. Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.*, **17** (10), 994–999.

121. Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., and Hamon, C. (2003) Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.*, **75** (8), 1895–1904.

122. Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D.J. (2004) Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. *Mol. Cell. Proteomics*, **3** (12), 1154–1169.

123. Zhang, L., and Elias, J.E. (2017) Relative Protein Quantification Using Tandem Mass Tag Mass Spectrometry BT - Proteomics: Methods and Protocols (eds.Comai, L., Katz, J.E., and Mallick, P.), Springer New York, New York, NY, pp. 185–198.

124. Ow, S.Y., Salim, M., Noirel, J., Evans, C., Rehman, I., and Wright, P.C. (2009) iTRAQ Underestimation in Simple and Complex Mixtures: "The Good, the Bad and the Ugly." *J. Proteome Res.*, **8** (11), 5347–5355.

125. Ting, L., Rad, R., Gygi, S.P., and Haas, W. (2011) MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Meth*, **8** (11), 937–940.

126. McAlister, G.C., Nusinow, D.P., Jedrychowski, M.P., Wühr, M., Huttlin, E.L., Erickson, B.K., Rad, R., Haas, W., and Gygi, S.P. (2014) MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.*, **86** (14), 7150–7158.

127. Desiderio, D.M., and Kai, M. (1983) Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of

peptides in biologic tissue. *Biomed. Mass Spectrom.*, **10** (8), 471–479.

128. Pan, S., Zhang, H., Rush, J., Eng, J., Zhang, N., Patterson, D., Comb, M.J., and Aebersold, R. (2005) High Throughput Proteome Screening for Biomarker Detection. *Mol. Cell. Proteomics*, **4** (2), 182–190.

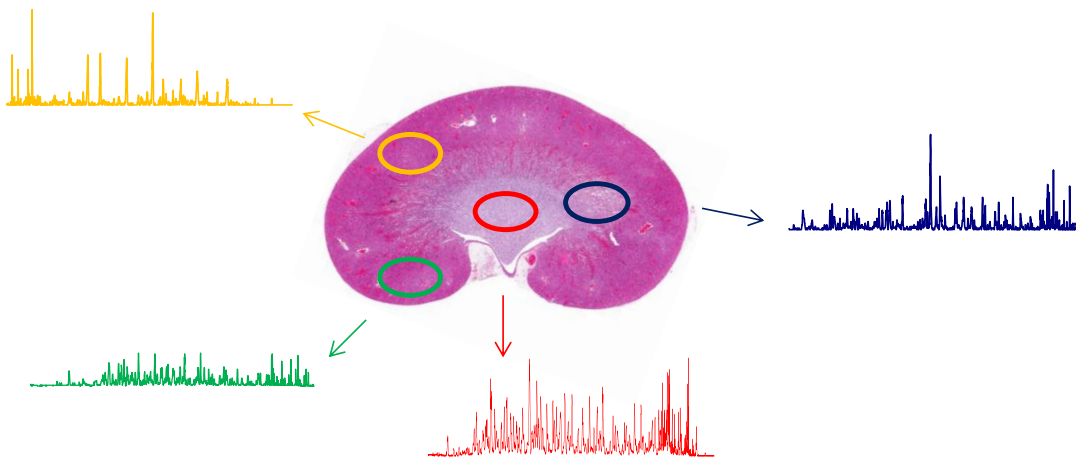
129. Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W., and Gygi, S.P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci.*, **100** (12), 6940–6945.

130. Pratt, J.M., Simpson, D.M., Doherty, M.K., Rivers, J., Gaskell, S.J., and Beynon, R.J. (2006) Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat. Protoc.*, **1** (2), 1029–1043.

131. Brun, V., Dupuis, A., Adrait, A., Marcellin, M., Thomas, D., Court, M., Vandenesch, F., and Garin, J. (2007) Isotope-labeled Protein Standards. *Mol. Cell. Proteomics*, **6** (12), 2139–2149.

Chapter 2

Clinical microproteomics: tissue characterization for molecular pathology



2.1. Clinical proteomics for molecular pathology

Proteins are intrinsic to all cellular processes, it therefore follows that the study of the proteome is crucial for the determination of disease pathophysiology [1]. The main goal of much present-day proteomics research is the development of new strategies for clinical diagnosis, prognosis, patient monitoring and the identification of drugable targets. [2]. In recent years, a lot of different technologies have been proposed for the proteomics characterization of both body fluids and tissues, and MS has rapidly become a powerful tool for molecular pathology.

Histopathology is the gold-standard technique used by pathologists for disease diagnosis and it is used in research labs to validate the results of translational studies [3]. Histopathological characterization and associated molecular tests of patient tissue samples represents the definitive method for the confirmation of the presence/absence of disease, and for disease grading and progression. However, the manual examination under a microscope not only limits the throughput of tissue analysis, but it is also subjective as it depends on the pathologist's experience.

The application of proteomics technologies to tissue characterization has aided the pathologist in the diagnostic process [4]. The key advantages of MS-based methods over histopathological examinations are the much higher sensitivity and throughput. Moreover, the conventional histopathological annotation provides only a morphological characterization of the tissue, while MS-based methods enable the identification of the molecular signatures of the different cell populations present in the tissue. Such unambiguous molecular information can be used to determine the presence and the progression of a disease and the response to therapies, thus providing prognostic and diagnostic value [5].

MS-based proteomics approaches for clinical tissue characterization can be divided into untargeted and targeted methods. Untargeted proteomics refers to the analysis of clinical samples aimed at the identification of putative biomarkers. MALDI mass spectrometry imaging (MSI) represents a unique approach to assess the spatial distribution of hundreds

of different molecules (proteins, peptides, lipids and metabolites) from tissue sections in a label-free fashion [6]. The MALDI-MSI molecular profiles can be aligned with histological images to translate the molecular information to tissue morphology and to complement histopathological evaluation for a better therapeutic management. MSI-based molecular pathology has been extensively used in clinical research to uncover molecular changes associated with disease progression, diagnosis and prognosis and to study the distribution of drugs and their metabolites [7][8].

Shotgun proteomics by LC-MS/MS analysis represents the most common untargeted proteomics approach for tissue biomarker discovery studies [9]. LC-MS/MS provides high resolution analysis of complex samples, enabling the identification and quantification of thousands of proteins. Unlike MALDI-MSI, proteins must be extracted from the tissue prior to LC-MS/MS analysis, so the spatial information is lost. However, the combination of laser microdissection strategies with shotgun proteomics enables the analysis of specific regions of the tissue, enabling the in-depth characterization of histology-defined regions-of-interest [10][11].

Targeted proteomics experiments are focused on the identification and quantification of known biomarkers [9]. Targeted experiments provide high sensitivity, reproducibility and throughput because only a predefined set of peptides/proteins is measured [12]. Selected reaction monitoring (SRM) and multiple reaction monitoring (MRM) are established methods performed on triple-quadrupole (QqQ) instruments, where ion selection in Q1 and fragment ion detection in Q3 with narrow *m/z* windows (± 0.02 *m/z*) provide higher selectivity and sensitivity compared to full scan proteomics analysis [13]. Parallel reaction monitoring (PRM) is a method similar to SRM that can be performed on quadrupole-time of flight (Q-ToF) and quadrupole-Orbitrap instruments, in which the quadrupole selects the precursor ions and the full MS/MS spectra are acquired in the high resolution mass analyzer [14][15].

Regardless of the acquisition method, protein extraction from tissue represents the first step in shotgun proteomics of tissues. In the last years, several strategies have been developed to enable the proteomics analysis of histology defined tissue regions for better

tissue characterization. Such focused analyses are crucial for the molecular analysis of tissues characterized by a high degree of cellular heterogeneity.

2.2. Tissue heterogeneity

The main challenge in tissue proteomics is the heterogeneity of the tissue. For instance, malignant solid tumor tissues consist of tumor cells, tumor-associated normal epithelial cells and stromal cells, immune cells and vascular cells (Figure 2.1.). This complex network of different types of cells (tumor microenvironment) is crucial for tumor growth, invasion and metastasis [16].

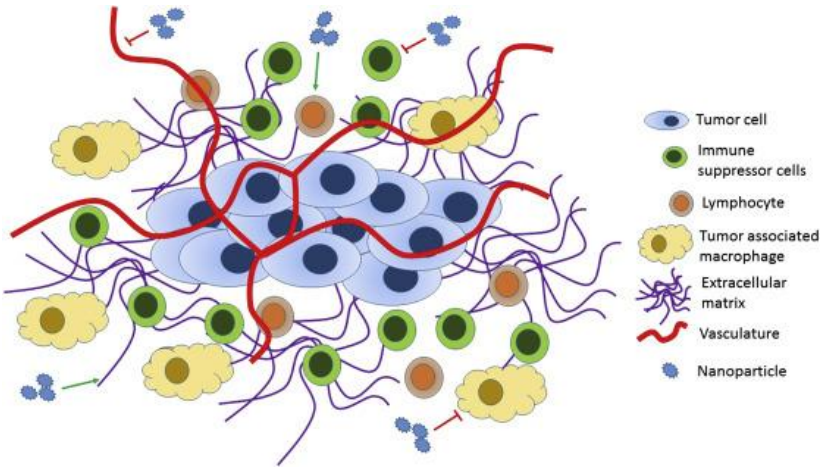


Figure 2.1. *Schematic presentation of components of the tumor microenvironment. Figure adopted from* [17].

MS-based proteomics has been successfully applied to analyze tumor and non-tumor tissues, early detection, tumor typing, tumor grading and tumor margins [18][19]. Similar considerations are also true for the neuroscience field since the central nervous system is made up of several hundred different cell types, which play roles in distinct developmental and behavioral disorders [20][21]. Moreover, several neurodegenerative disorders (*e.g.* Alzheimer's disease, Down's syndrome, Huntington's disease, Amyotrophic lateral sclerosis, ischemic stroke, frontotemporal lobar degeneration and Parkinson's disease) are associated only with specific cell types in the brain (Figure 2.2.) [22]. These disorders are characterized by a phenomenon called "selective vulnerability", which refers to the observation that degeneration in a specific disease is not diffuse but cell type and neuronal network specific [23].

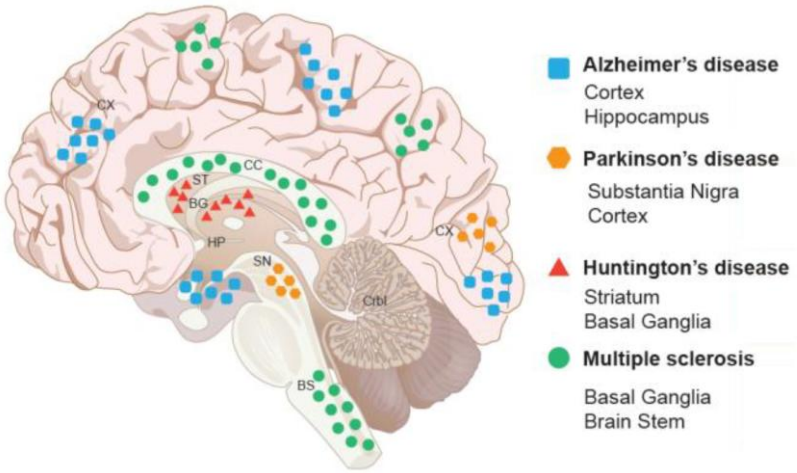


Figure 2.2. *Brain regions and neuronal cell types involved in neurodegenerative disease. Regions affected by different neurodegenerative diseases are indicated by different colors. Abbreviations: Basal ganglion (BG), Brain stem (BS), Cerebellum (Crbl), Corpus callosum (CC), Cortex (Cx), Hippocampus (Hp), Striatum (St), Substantia Nigra (SN). Figure modified from* [24].

Heterogeneous tissues may confound proteomics analysis because it is extremely difficult to separate the contribution of the different cell populations present in the sample. Different cell populations in the same tissue may exhibit different phenotypes because of the influence of the microenvironment [25]. The ability to investigate the proteome of single cell populations can provide information about interesting and important biological phenomena that may be obscured in bulk measurements. Population distributions can mask the presence of small subpopulations of cells, as the measured response is averaged across the bulk population. Since different cell populations generally do not contribute equally to the biology of the system, a bulk experiment can fail to illuminate the biological mechanisms [26]. In order to properly understand the molecular signals within these highly heterogeneous tissues, there is the need to focus the experiments on specific cells (specific cell type and specific environment).

2.3. Laser Microdissection

Laser microdissection strategies have emerged as the method of choice for the isolation of specific cells or specific tissue regions [27]. This technology is able to harvest the target cells directly from the tissue by means of a laser coupled with a microscope and can be applied to histological specimens, formalin-fixed paraffin-embedded (FFPE) or fresh-frozen tissues and stained or unstained tissue sections. After delineation of the tissue region of interest, the target cells are automatically isolated and can be used for downstream molecular characterization, including DNA/RNA sequencing and proteomics analysis.

Laser microdissection systems can use IR or UV laser sources and can operate in *capture* or *ablation* mode [28]. When operating in the *ablation* mode, a highly focused laser beam ablates the tissue surrounding the region of interest, which is then detached from the rest of the tissue. Laser-capture microdissection (LCM) is based on the direct isolation of the cells of interest [29]. The tissue slice is attached to a glass slide coated with a thermoplastic polyethylene naphthalate (PEN) membrane. A focused UV laser is used to cut around the

ROI and a defocused laser pulse is used to generate a photonic force to catapult the tissue ROI directly into the cap of a collection tube (Figure 2.3.).

LCM provides the isolation of the target cells (from several thousand cells to single cells) with high selectivity and it has become a powerful tool for tissue proteomics, since it represents a bridge between histology and LC-MS based proteomics methods. LCM has been widely used to address the tissue heterogeneity challenge in various medical research areas including neuroscience, oncology, forensic science and biomarker discovery [30]. Recently, LCM coupled with downstream shotgun proteomics analysis has been applied to study the neurons and protein aggregates in Alzheimer's disease patients [10][31], neurons and blood brain barrier structures to study ischemia [32] and in oncology for early cancer detection, tumor microenvironment characterization and drug resistance [33][34][35].

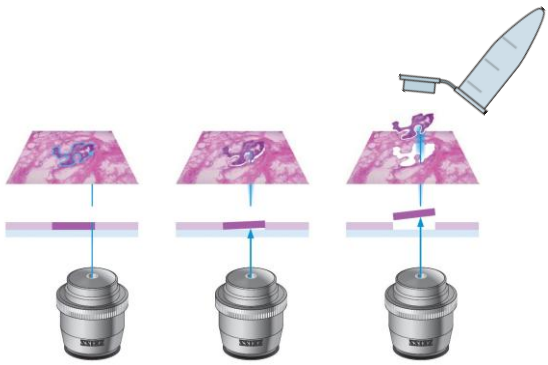


Figure 2.3. *Schematics of the laser capture microdissection process. A focused laser beam is used to cut the margins of an ROI, following which a defocused laser pulse catapults the piece of tissue in to the cap of a collection tube.*

The major limitation of LCM is the need to identify the target cells on the basis of morphological characteristics, which requires personnel trained in the histology of the organ/disease of interest. Any staining protocol used to aid histological characterization must be carefully selected to avoid interference with the downstream analyses. Moreover,

the amount of tissue material retrieved from LCM can be very small if strict cellular and environmental specifications are imposed, resulting in lower proteome coverage [36].

2.4. Microproteomics workflows for microdissected tissue analysis

In the last few years, the development of sample preparation protocols that provide sufficient sensitivity for the processing of small microdissected samples has led to the initiation of the "microproteomics" era [37]. The aim of microproteomics is the development of high sensitivity strategies that can be applied to the analysis of the low protein amounts (in the order of few micrograms) that can be extracted from microdissected tissues. These microproteomics strategies generally combine LCM with efficient protein extraction procedures and specific digestion protocols that minimize sample losses to achieve high proteome coverage, which is required for a comprehensive molecular characterization.

The conventional proteomics workflow is based on in-solution protein digestion (ISD) using urea as lysis buffer and solid phase extraction procedures for peptide purification prior to LC-MS/MS analysis. Urea is a chaotrope that stabilizes denatured proteins by hydrogen bonding promoting protein solubilization [38]. Although urea has been successfully applied to protein extraction from tissue samples ([39][40]) there are several drawbacks that can affect protein identification, especially when small microdissected tissues are analyzed. Urea leads to carbamylation and other chemical modification on lysine and arginine residues that not only can affect protein digestion, by interfering with proteolytic enzymes, but also can affect the protein identification process as peptides with an unknown chemical modification will not be identified [41][42]. Moreover, urea inhibits trypsin activity, thus samples have to be diluted prior to enzymatic digestion in order to reduce the urea concentration. In the case of small sample amounts, such dilution steps are best avoided because it increases peptide losses due to adsorption/absorption during sample handling; for microproteomics such losses are significant.

A more efficient strategy to extract proteins from tissues relies on the use of strong ionic surfactants such as sodium dodecyl sulfate (SDS) and sodium deoxycholate (SDC). These detergents solubilize proteins by disrupting hydrophobic interaction and result in a higher protein extraction yield than urea-based buffers [43][44]. However, the detergents are frequently incompatible with LC-MS/MS analysis and must be removed prior to the analysis, because they would otherwise dominate the mass spectra.

Several purification strategies have been developed to remove (or reduce) detergents including affinity-based methods, membrane filtration, electrophoresis and protein precipitation [45][46]. Although these purification methods provide an efficient removal of detergents, they introduce extra sample preparation steps that affect protein recovery because of extensive sample losses. Thus, microproteomics protocols have been developed that combine the high protein extraction efficiency of strong detergents with sample purification procedures that minimize sample losses, specifically protocols in which all steps are performed within a single vessel. These include Filter-Aided Sample Preparation (FASP) and Single-Pot Solid Phase-enhanced Sample Preparation (SP3).

2.4.1. Filter-Aided Sample Preparation (FASP)

The FASP protocol was reported by Wisńiewski *et al.* in 2009 and allows the use of highly concentrated SDS for “universal” protein solubilization [47]. The sample purification is performed by ultrafiltration to remove the SDS prior to LC-MS/MS analysis. Samples (either cells or tissue specimens) are solubilized in 4% SDS and then retained and concentrated into microliter volumes in an ultrafiltration device. The filter unit represents the key element of the FASP protocol as it provides detergent removal, buffer exchange, protein alkylation and digestion (Figure 2.4.). The filter unit (3 or 10 kDa cut off) retains high molecular weight compounds (as proteins), while small molecular weight compound (SDS, salts, impurities and peptides) can pass through the filter. After lysis in 4% SDS the samples are diluted with an urea buffer and loaded on the filter unit. Buffer exchange is

performed on the filter by centrifugation to remove SDS. Proteins are then alkylated and digested on the filter and peptides are eluted by centrifugation.

The FASP protocol has proven to work efficiently with high sample amounts, while incomplete peptide recovery has been observed when applied to small sample amounts. The sample losses observed during the processing of small sample amounts were reportedly minimized by using alternative reagents, resulting in improved sensitivity, recovery and proteome coverage [48].

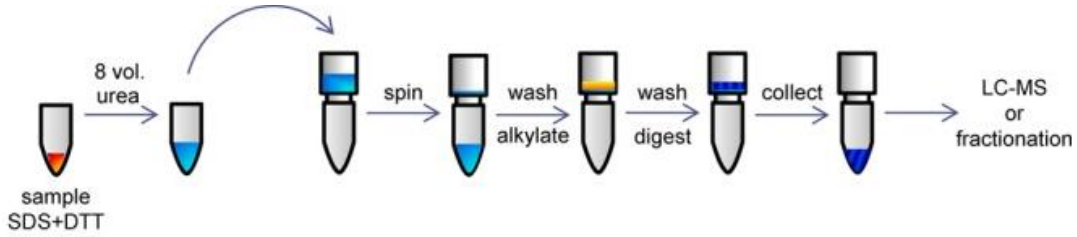


Figure 2.4. *Schematics of the FASP protocol. Figure modified from* [48].

The FASP protocol has been successfully applied to a wide variety of cells and tissue samples [49][50], including the analysis of 500 laser microdissected cells from an FFPE tissue specimens [51].

2.4.2. Single-Pot Solid Phase-enhanced Sample Preparation (SP3)

The SP3 protocol was developed by Hughes *et al.* in 2014 [52]. The advantage of the SP3 protocol is that all proteomics sample preparation steps are performed in a single tube, minimizing sample transfers. This results in a significant reduction of the sample losses associated with conventional sample preparation protocols. Moreover, cell lysis is achieved using an SDS-based buffer that maximizes protein extraction and solubilization.

The SP3 protocol is based on the use of carboxylate-coated paramagnetic beads that are able to interact with proteins and peptides via a mechanism similar to HILIC [53] and ERLIC [54]. The addition of an organic solvent to an aqueous solution containing paramagnetic beads promotes trapping of proteins/peptides into a solvation layer that forms on the surface of the carboxylate coated paramagnetic beads. The interaction between proteins/peptides and the beads can be modulated by varying either the percentage of organic solvent or the pH. At high organic solvent percentages and low pH proteins and peptides are retained in the hydrophilic layer on the surface of the beads (HILIC like interaction). By decreasing the proportion of organic solvent and increasing the pH proteins/peptides are repelled by the beads because of the negatively charged carboxylate groups on the surface of the beads (ERLIC like mechanisms). Thus, once immobilized on the beads, proteins/peptides can be rinsed with different buffers while using a magnet to remove contaminating agents such as salts and detergents. In this way all the proteomics sample preparation protocol can be performed in a single tube (Figure 2.5.). After cell lysis in an SDS-based buffer, a protein cleanup step by buffer exchange is performed to remove SDS before enzymatic digestion. The buffer exchange is achieved by trapping the proteins in the solvation layer of the beads. Following enzymatic digestion the peptides can be labeled (if quantification is needed) or purified with a similar procedure for the protein cleanup. Once all sample preparation steps have been completed the peptides can be eluted and then analyzed by LC-MS/MS.

Figure 2.5. *Schematics of the SP3 workflow. Figure modified from* [52].

The SP3 has been shown to be compatible with isotope labeling strategies for quantitative proteomics, providing a labeling efficiency higher that 97% [52].

The SP3 method has been compared to FASP by analyzing 10 μg of a yeast lysate. A similar depth of proteome coverage was observed and the physicochemical properties of the peptides identified did not show significant differences, indicating that the functionalized beads did not introduce a bias into the type of peptides that were identified. The feasibility of SP3 for the analysis of small sample amounts has been demonstrated by testing the procedure on different amount of HeLa cells (from 500k to 1k cells) and on single *Drosophila* embryos. The high sensitivity of the SP3 protocol enabled the possibility to study the proteome changes associated with the developmental stages of embryos [52].

2.5. Aim of the thesis

This thesis describes the development and optimization of an ultrasensitive microproteomics workflow to perform MS-based proteomics characterization of pathological tissues. All the steps of the conventional proteomics pipeline have been optimized to increase the depth of coverage that can be achieved from the analysis of microdissected tissue samples. The combination of laser capture microdissection with the ultrasensitive proteomic workflow enabled the proteomics characterization of specific tissue substructures. The workflow was tested on different types of murine tissues (brain and kidney) and was applied to different biomedical case studies in the oncology and neuroscience fields.

In Chapter 3 I compared three sample preparation protocols to find that most suited to the analysis of small sample amounts. The conventional urea based in-solution digestion (ISD) was compared with Filter-Aided Sample Preparation (FASP) and Single Pot Solid Phase-enhanced Sample Preparation (SP3) on different amounts of cultured HeLa cells (from 1 million down to 5000 cells). The different protocols were evaluated based on the depth of

proteome coverage (*i.e.* number of proteins and peptides identified) and on the basis of the digestion efficiency. The SP3 protocol was found to perform best when analyzing small sample amounts, as it provided the highest number of protein identifications and the highest digestion efficiency. A further optimization of the SP3 protocol resulted in the identification of approximately 3000 protein groups from 1 µg of protein. The optimized SP3 protocol was successfully applied, in combination with laser capture microdissection (LCM), to characterize different histological regions of the brain of a mouse model of glioblastoma.

In Chapter 4 I developed a quantitative ultrasensitive proteomics workflow for the analysis of tissue samples. The previously optimized LCM and SP3 digestion steps were integrated with a colorimetric assay for protein content estimation, a Tandem Mass Tags (TMT) labeling protocol and a peptide fractionation protocol. The integration of the different steps resulted in a complete quantitative ultrasensitive proteomics workflow that can be applied to characterize any type of microdissected tissue sample.

The colorimetric assay (a modified microBCA assay) enabled the measurement of the protein content in just 1 µL of sample and in the presence of the detergents and salts commonly used in lysis buffers. The low sample volume required for protein concentration estimation allowed protein quantitation of each individual sample (a crucial step for proteomics as the amount of enzyme that must be used is dependent on the total amount of protein present in the sample).

To ensure accurate relative quantification I optimized in-solution and on-column TMT labeling procedures for low sample amounts. Preliminary results revealed very low labeling efficiency when standard labeling conditions were applied to low sample amounts. Following an exhaustive optimization of in-solution and on on-column TMT labeling, the final conditions provided a TMT labeling efficiency (for 1 µg of HeLa digest) even greater than that obtained using standard methods on high sample amounts (25-50 µg of digest).

An automated peptide fractionation protocol was also developed to increase the depth of proteome coverage. The high-pH reversed phase fractionation increased the proteome coverage by approximately 140% relative to a single long gradient analyses.

The optimized workflow was then applied as a proof-of-principle to the characterization of mouse kidney substructures. The analysis provided the identification of more than 300 protein groups that exhibited region-specific expression profiles.

Chapter 5 describes the application of the quantitative proteomics workflow to the characterization of the central and peripheral nervous system of a mouse model of Krabbe disease (the Twitcher mouse). This work represents the first in-depth proteomics characterization of the Twitcher mouse. I compared the proteomes extracted from the corpus callosum, motor cortex and sciatic nerves of five Twitcher and five control wild type mice. The results on the proteome changes in the Twitcher mouse provided new insights into the molecular mechanisms of Krabbe disease showing neuroinflammation, activation of immune response, accumulation of lysosomal proteins, demyelination, membrane raft disruption and reduced nervous system development in the Twitcher mouse.

In Chapter 6 an outlook is described regarding how a further increase in sensitivity and specificity can be used to achieve cell population-specific and single-cell proteomics. Recently, several strategies have been proposed to study the proteome and the molecular profile of single cells, opening exciting frontiers and a new era in biomedical research. Many pressing medical needs, such as disease diagnosis, stem cell differentiation, rare cell profiling, protein dynamics and biological heterogeneity, have long been limited by the lack of technologies to detect proteins in single cells. These technological advances could help revolutionize medicine and will further deepen our understanding of biological systems, both fundamental and involved in disease progression.

References

1. Silberring, J., and Ciborowski, P. (2010) Biomarker discovery and clinical proteomics. *TrAC Trends Anal. Chem.*, **29** (2), 128–140.

2. Hanash, S. (2003) Disease proteomics. *Nature*, **422** (6928), 226–232.

3. Norris, J.L., and Caprioli, R.M. (2013) Imaging mass spectrometry: A new tool for pathology in a molecular age. *Proteomics - Clin. Appl.*, **7** (11–12), 733–738.

4. Kriegsmann, J., Casadonte, R., Kriegsmann, K., Longuespée, R., and Kriegsmann, M. (2018) Mass spectrometry in pathology – Vision for a future workflow. *Pathol. - Res. Pract.*, **214** (8), 1057–1063.

5. Longuespée, R., Casadonte, R., Schwamborn, K., Reuss, D., Kazdal, D., Kriegsmann, K., von Deimling, A., Weichert, W., Schirmacher, P., Kriegsmann, J., and Kriegsmann, M. (2018) Proteomics in Pathology. *Proteomics*, **18** (2), 1700361.

6. Caprioli, R.M., Farmer, T.B., and Gile, J. (1997) Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS. *Anal. Chem.*, **69** (23), 4751–4760.

7. Aichler, M., and Walch, A. (2015) MALDI Imaging mass spectrometry: current frontiers and perspectives in pathology research and practice. *Lab. Investig.*, **95**, 422.

8. McDonnell, L.A., Angel, P.M., Lou, S., and Drake, R.R. (2017) Chapter Eleven - Mass Spectrometry Imaging in Cancer Research: Future Perspectives, in *Applications of Mass Spectrometry Imaging to Cancer*, vol. 134, Academic Press, pp. 283–290.

9. Frantzi, M., Latosinska, A., Kontostathi, G., and Mischak, H. (2018) Clinical Proteomics: Closing the Gap from Discovery to Implementation. *Proteomics*, **18** (14), 1700463.

10. Drummond, E.S., Nayak, S., Ueberheide, B., and Wisniewski, T. (2015) Proteomic

analysis of neurons microdissected from formalin-fixed, paraffin-embedded Alzheimer's disease brain tissue. *Sci. Rep.*, **5**, 15456.

11. Clair, G., Piehowski, P.D., Nicola, T., Kitzmiller, J.A., Huang, E.L., Zink, E.M., Sontag, R.L., Orton, D.J., Moore, R.J., Carson, J.P., Smith, R.D., Whitsett, J.A., Corley, R.A., Ambalavanan, N., and Ansong, C. (2016) Spatially-Resolved Proteomics: Rapid Quantitative Analysis of Laser Capture Microdissected Alveolar Tissue Samples. *Sci. Rep.*, **6** (November), 1–13.

12. Shi, T., Song, E., Nie, S., Rodland, K.D., Liu, T., Qian, W.-J., and Smith, R.D. (2016) Advances in targeted proteomics and applications to biomedical research. *Proteomics*, **16** (15–16), 2160–2182.

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

14. Peterson, A.C., Russell, J.D., Bailey, D.J., Westphall, M.S., and Coon, J.J. (2012) Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Mol. Cell. Proteomics*, **11** (11), 1475–1488.

15. Gallien, S., Duriez, E., Crone, C., Kellmann, M., Moehring, T., and Domon, B. (2012) Targeted Proteomic Quantification on Quadrupole-Orbitrap Mass Spectrometer. *Mol. Cell. Proteomics*, **11** (12), 1709–1723.

16. Dagogo-Jack, I., and Shaw, A.T. (2017) Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.*, **15**, 81.

17. Siegler, E.L., Kim, Y.J., and Wang, P. (2016) Nanomedicine targeting the tumor microenvironment: Therapeutic strategies to inhibit angiogenesis, remodel matrix, and modulate immune responses. *J. Cell. Immunother.*, **2** (2), 69–78.

18. van der Merwe, D., Oikonomopoulou, K., Marshall, J., and Diamandis, E.P.B.T.-A. in C.R. (2006) Mass Spectrometry: Uncovering the Cancer Proteome for Diagnostics, in *Genomics in Cancer Drug Discovery and Development*, vol. 96, Academic Press, pp. 23–50.

19. Wang, P., Whiteaker, J.R., and Paulovich, A.G. (2009) The evolving role of mass spectrometry in cancer biomarker discovery. *Cancer Biol. Ther.*, **8** (12), 1083–1094.

20. Kitchen, R.R., Rozowsky, J.S., Gerstein, M.B., and Nairn, A.C. (2014) Decoding neuroproteomics: integrating the genome, translatome and functional anatomy. *Nat. Neurosci.*, **17**, 1491.

21. Hosp, F., and Mann, M. (2017) A Primer on Concepts and Applications of Proteomics in Neuroscience. *Neuron*, **96** (3), 558–571.

22. Wilson, S.R., and Nairn, C.A. (2018) Cell-Type-Specific Proteomics: A Neuroscience Perspective. *Proteomes* , **6** (4).

23. Fu, H., Hardy, J., and Duff, K.E. (2018) Selective vulnerability in neurodegenerative diseases. *Nat. Neurosci.*, **21** (10), 1350–1358.

24. Hussain, R., Zubair, H., Pursell, S., and Shahab, M. (2018) Neurodegenerative Diseases: Regenerative Mechanisms and Novel Therapeutic Approaches. *Brain Sci.* , **8** (9).

25. Altschuler, S.J., and Wu, L.F. (2010) Cellular Heterogeneity: Do Differences Make a Difference? *Cell*, **141** (4), 559–563.

26. Calvo, K.R., Liotta, L.A., and Petricoin, E.F. (2005) Clinical Proteomics: From Biomarker Discovery and Cell Signaling Profiles to Individualized Personal Therapy. *Biosci. Rep.*, **25** (1–2), 107–125.

27. Xu, B.J. (2010) Combining laser capture microdissection and proteomics: Methodologies and clinical applications. *Proteomics - Clin. Appl.*, **4** (2), 116–123.

28. Espina, V., Wulfkuhle, J.D., Calvert, V.S., VanMeter, A., Zhou, W., Coukos, G., Geho, D.H., Petricoin, E.F., and Liotta, L.A. (2006) Laser-capture microdissection. *Nat. Protoc.*, **1** (2), 586–603.

29. Schütze, K., Niyaz, Y., Stich, M., and Buchstaller, A.B.T.-M. in C.B. (2007) Noncontact Laser Microdissection and Catapulting for Pure Sample Capture, in

Laser Manipulation of Cells and Tissues, vol. 82, Academic Press, pp. 647–673.

30. Espina, V., Heiby, M., Pierobon, M., and Liotta, L.A. (2007) Laser capture microdissection technology. *Expert Rev. Mol. Diagn.*, **7** (5), 647–657.

31. Lutz, M.B., and Peng, J. (2018) Deep Profiling of the Aggregated Proteome in Alzheimer's Disease: From Pathology to Disease Mechanisms. *Proteomes* , **6** (4).

32. García-Berrocoso, T., Llombart, V., Colàs-Campàs, L., Hainard, A., Licker, V., Penalba, A., Ramiro, L., Simats, A., Bustamante, A., Martínez-Saez, E., Canals, F., Sanchez, J.-C., and Montaner, J. (2018) Single Cell Immuno-Laser Microdissection Coupled to Label-Free Proteomics to Reveal the Proteotypes of Human Brain Cells After Ischemia. *Mol. Cell. Proteomics*, **17** (1), 175–189.

33. Posadas, E.M., Simpkins, F., Liotta, L.A., MacDonald, C., and Kohn, E.C. (2005) Proteomic analysis for the early detection and rational treatment of cancer—realistic hope? *Ann. Oncol.*, **16** (1), 16–22.

34. Bateman, N.W., and Conrads, T.P. (2018) Recent advances and opportunities in proteomic analyses of tumour heterogeneity. *J. Pathol.*, **244** (5), 628–637.

35. An, Y., Zhou, L., Huang, Z., Nice, E.C., Zhang, H., and Huang, C. (2019) Molecular insights into cancer drug resistance from a proteomics perspective. *Expert Rev. Proteomics*, **16** (5), 413–429.

36. Altelaar, A.F.M., and Heck, A.J.R. (2012) Trends in ultrasensitive proteomics. *Curr. Opin. Chem. Biol.*, **16** (1), 206–213.

37. Gutstein, H.B., Morris, J.S., Annangudi, S.P., and Sweedler, J. V (2008) Microproteomics: Analysis of protein diversity in small samples. *Mass Spectrom. Rev.*, **27** (4), 316–330.

38. Lim, W.K., Rösgen, J., and Englander, S.W. (2009) Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. *Proc. Natl. Acad. Sci.*, **106** (8), 2595 LP – 2600.

39. Luebker, S.A., and Koepsell, S.A. (2016) Optimization of Urea Based Protein

Extraction from Formalin-Fixed Paraffin-Embedded Tissue for Shotgun Proteomics. *Int. J. Proteomics*, **2016**, 4324987.

40. Dapic, I., Uwugiaren, N., Jansen, P.J., and Corthals, G.L. (2017) Fast and Simple Protocols for Mass Spectrometry-Based Proteomics of Small Fresh Frozen Uterine Tissue Sections. *Anal. Chem.*, **89** (20), 10769–10775.

41. Garbis, S., Lubec, G., and Fountoulakis, M. (2005) Limitations of current proteomics technologies. *J. Chromatogr. A*, **1077** (1), 1–18.

42. Sun, S., Zhou, J.-Y., Yang, W., and Zhang, H. (2014) Inhibition of protein carbamylation in urea solution using ammonium-containing buffers. *Anal. Biochem.*, **446**, 76–81.

43. Ericsson, C., and Nister, M. (2011) Protein extraction from solid tissue. *Methods Mol. Biol.*, **675**, 307–312.

44. Shen, K., Sun, J., Cao, X., Zhou, D., and Li, J. (2015) Comparison of Different Buffers for Protein Extraction from Formalin-Fixed and Paraffin-Embedded Tissue Specimens. *PLoS One*, **10** (11), e0142650–e0142650.

45. Hengel, S.M., Floyd, E., Baker, E.S., Zhao, R., Wu, S., and Paša-Tolić, L. (2012) Evaluation of SDS depletion using an affinity spin column and IMS-MS detection. *Proteomics*, **12** (21), 3138–3142.

46. Feist, P., and Hummon, A.B. (2015) Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *Int. J. Mol. Sci.*, **16** (2), 3537–3563.

47. Wiśniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods*, **6**, 359.

48. Erde, J., Loo, R.R.O., and Loo, J.A. (2014) Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J. Proteome Res.*, **13** (4), 1885–1895.

49. Maurer, M., Muller, A.C., Parapatics, K., Pickl, W.F., Wagner, C., Rudashevskaya,

E.L., Breitwieser, F.P., Colinge, J., Garg, K., Griss, J., Bennett, K.L., and Wagner, S.N. (2014) Comprehensive comparative and semiquantitative proteome of a very low number of native and matched epstein-barr-virus-transformed B lymphocytes infiltrating human melanoma. *J. Proteome Res.*, **13** (6), 2830–2845.

50. Wisniewski, J.R., Dus, K., and Mann, M. (2013) Proteomic workflow for analysis of archival formalin-fixed and paraffin-embedded clinical samples to a depth of 10 000 proteins. *Proteomics. Clin. Appl.*, **7** (3–4), 225–233.

51. Wiśniewski, J.R., Ostasiewicz, P., and Mann, M. (2011) High Recovery FASP Applied to the Proteomic Analysis of Microdissected Formalin Fixed Paraffin Embedded Cancer Tissues Retrieves Known Colon Cancer Markers. *J. Proteome Res.*, **10** (7), 3040–3049.

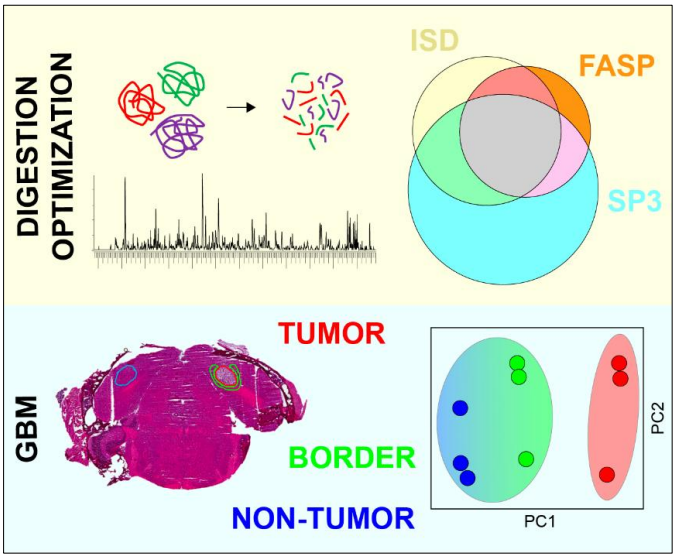
52. Hughes, C.S., Foehr, S., Garfield, D.A., Furlong, E.E., and Steinmetz, L.M. (2014) Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.*, **10** (757), 1–15.

53. Alpert, A.J. (1990) Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J. Chromatogr. A*, **499**, 177–196.

54. Alpert, A.J. (2008) Electrostatic Repulsion Hydrophilic Interaction Chromatography for Isocratic Separation of Charged Solutes and Selective Isolation of Phosphopeptides. *Anal. Chem.*, **80** (1), 62–76.

Chapter 3

Optimization of a Protein Digestion Protocol for Label-Free Microproteomics: Application to the Characterization of a Mouse Model of Glioblastoma



3.1. Summary

In this work I optimized a protein digestion protocol for the analysis of small tissue samples. I first compared *Filter-Aided Sample Preparation* (FASP) with *Single-Pot Solid Phase-enhanced Sample Preparation* (SP3) with a reference method based on in-solution proteolytic digestion (ISD) for the analysis of different amounts of material (from 10^6 HeLa cells down to 5×10^3 cells). The SP3 protocol was found to perform better than both FASP and ISD protocols for the analysis of small sample amounts, in terms of number of identification (2036 protein groups from 5e3 cells) and digestion efficiency (80% of 0 missed cleavages for the 5×10^3 cell samples). A further optimization of the SP3 protocol (*i.e.* type of sample tubes, magnets and concentration of carboxylate paramagnetic beads solution) increased the number of identified proteins by approximately 35% (3000 protein groups identified from 1 µg of HeLa cells) with a very high inter-batch reproducibility (<6% variation coefficient).

The performance of the optimized SP3 digestion protocol for microdissected tissue samples was then assessed by analyzing different histological brain regions of a mouse model of glioblastoma.

3.2. Introduction

Mass spectrometry (MS)-based proteomics has matured into the method of choice for the large-scale identification and characterization of proteins associated with biological processes and disease [1][2]. Nevertheless, challenges remain. For instance the analysis of proteins and polypeptides from sample-limited sources such as single cells [3], purified cell populations[4], subcellular organelles, exosomes, or small histologically defined regions of tissue isolated by laser capture microdissection [5][6][7].

Recent advances in liquid chromatography (LC) and MS equipment have greatly improved the analysis of low sample amounts. The development of mass spectrometers has increased sequencing speed [8][9][10][11] and ion transmission [12][13], resulting in an increase in dynamic range and sensitivity. Advances in ultra-high-pressure liquid chromatography (HPLC) has enabled the routine use of long columns (≥50 cm) with smaller internal diameter and smaller particle sized (< 5 µm) further increasing peptide separation resolution [14], resulting in the identification of more than 4800 protein groups in a single LC-MS run [15][16]. However, the developments in LC-MS sensitivity have outpaced developments in sensitive sample preparation protocols.

Several sample preparation steps are required for the proteomics analysis of samples; namely, cell lysis and protein denaturation, reduction and alkylation, proteolytic digestion and peptide purification before the LC-MS analysis [17]. To maximize the yield of protein extraction several detergents and chaotropes can be used. However, most of these chemicals inhibits protease activity and/or are not compatible with LC-MS. Thus, the main challenge in the proteomic analysis of small sample amounts is represented by sample losses occurring during the different steps of the sample preparation workflow.

In this work, we optimized a highly sensitive protein digestion protocol suited for the analysis of small sample amounts. We compared F*ilter-Aided Sample Preparation protocol* (FASP) [18] and S*ingle-Pot Solid Phase-enhanced Sample Preparation protocol* (SP3) [19] with the conventional urea based *in-solution digestion* (ISD). The FASP and SP3 are two recently published digestion protocols based on the use of strong detergents

(as SDS) to increase protein extraction. The SDS is then removed by a series of filtration steps on a molecular weight cutoff filter with the FASP protocol, or by retention of proteins/peptides on carboxylate modified paramagnetic beads with the SP3 protocol. These two protocols combine the benefit of using SDS to maximize protein extraction with a "single-pot" strategy to minimize losses during sample preparation. We compared the digestion efficiency and the total number of identified proteins and peptides obtained with the FASP, SP3 and ISD digestion protocols for different sample amounts in order to determine which was best suited for the characterization of low sample amounts. We then further optimized the best performing digestion protocol to increase the proteome coverage and reproducibility.

Finally, as a proof of principle, the optimized sample preparation protocol was applied to characterize the brain tumor of a mouse model of glioblastoma using just 0.8 mm^2 tissue areas of 12 µm thick tissue sections, and benchmarked the results against data from recently published tissue proteomics studies.

3.3. Experimental procedures

3.3.1. Materials

Trypsin/LysC mix, Mass Spec grade, was purchased from Promega (Madison, WI). Tandem Mass Tags (TMT 6-plex) were purchased from Thermo Fisher Scientific (Rockford, IL). All other reagents and solvents were purchased from Sigma Aldrich (St. Louis, MO). Magnets were purchased from Supermagnete (Gottmadingen, Deutschland).

3.3.2. Sample preparation

HeLa whole cell pellets (2.5e9) were purchased from IpraCell (Mons, Belgium), suspended in PBS and divided into aliquots of 5e3 to 1e6 cells. Cell pellets were stored at -80 °C until use. The optimization of the digestion protocol was performed on 1e6, 5e5, 1e5 and 5e3 cell pellets applying the ISD, FASP and SP3 protocols (described below). The microdissected tissue samples were digested with the optimized SP3 protocol.

3.3.2.1. In-Solution Digestion (ISD) protocol

HeLa cell pellets were suspended in a lysis buffer consisting of 8 M urea in 50 mM ammonium bicarbonate, 1 tablet of Complete mini EDTA-free Cocktail (Roche) and sonicated for 20 s (5 cycles of 2 s ON, 2 s OFF) with a Q-Sonic sonicator (QSONICA, Newton, Connecticut). After centrifugation at 20000g for 1.5 h at 4 °C the supernatant was assayed for protein content using the Qubit protein assay standard procedure (Life Technologies). Protein reduction and alkylation were performed by adding 2 µL of 2 mM dithiothreitol (DTT) at 56 °C for 25 min and 4 µL of 4 mM iodoacetamide (IAM) at room temperature for 30 min, respectively. Alkylation was quenched by adding 2 µL DTT to a final concentration of 4 mM. Enzymatic digestion was performed in two steps. The first step was performed directly with LysC at 37 °C for 4 h (1:75 enzyme/protein). The second

digestion was performed with trypsin (1:100 enzyme/protein) at 37 °C with overnight incubation after diluting the lysis buffer to 2 M urea with ammonium bicarbonate. Digestion was stopped by adding 1% formic acid. Desalting was performed by solid phase extraction using SepPak cartridges (Waters, 1 mg, C18) and a membrane vacuum pump. Cartridges were primed with 1 mL of 100% ACN and equilibrated twice with 1 mL of 0.1% formic acid. Samples were acidified to pH 2 with formic acid and loaded on the cartridges. The loaded cartridges were then washed twice with 1 mL of 0.1% formic acid and the peptides eluted twice with 250 µL of 80% ACN 0.1% formic acid. The purified samples were then dried down with a speedvac and stored at -20 °C until they were analyzed. For LC-MS/MS analysis the dried peptides were resuspended in 20 µL of 10% formic acid.

3.3.2.2. Filter-Aided Sample Preparation (FASP) protocol

HeLa cell pellets were resuspended in a lysis buffer containing 4% sodium dodecyl sulfate (SDS), 0.1 M DTT and 1 tablet of Complete mini EDTA-free Cocktail (Roche) in 100 mM tris(hydroxymethyl)aminomethane hydrochloride (Tris/HCl) pH 7.6 [20]. After incubation at 95 °C for 5 min, the samples were sonicated for 20 s (5 cycles of 2 s ON, 2 s OFF) with a Q-Sonic sonicator. The protein content was quantified with the Qubit protein assay standard procedure (Life Technologies). 30 µL of protein extract were mixed with 200 µL of UA buffer (8 M urea in 0.1M Tris/HCl pH 8.5) on a 10 kDa filter unit (Microcon YM-10, Millipore) and centrifuged at 14000g for 40 min. The filters were then rinsed twice with 200 µL of UA buffer and centrifuged at 14000g for 40 min. Alkylation was performed on the filters by adding 100 µL of 50 mM IAM in UA buffer. After vortex mixing and incubation at RT for 5 min, the filter units were centrifuged at 14000g for 30 min. The filters were then rinsed twice with 100 µL of UB buffer (8 M urea in 0.1M Tris/HCl pH 8) and centrifuged at 14000g for 40 min.

Enzymatic digestion was performed in two steps. The first step was performed by adding 40 µL of LysC (1:50 enzyme/protein) in UB buffer. Samples were then incubated overnight in a humidified chamber at 37 °C. After the first digestion step the filter units were transferred to new collection tubes. The second digestion was performed by adding 120 µL of trypsin (1:100 enzyme/protein) in 50 mM ammonium bicarbonate. Samples

were incubated at RT for 4 h and then centrifuged at 14000g for 40 min. The filter units were then rinsed with 50 µL of 0.5 M NaCl (aq.) and centrifuged at 14000g for 20 min. The filtrate was acidified with 1% formic acid and purified by solid phase extraction with a SepPak cartridge. The purified samples were then dried with a speedvac and stored at -20 °C until use, whereby they were resuspended in 20 µL of 10% formic acid prior to LC-MS/MS analysis.

3.3.2.3. Single-Pot Solid-Phase-enhanced Sample Preparation (SP3) protocol

HeLa cell pellets and microdissected tissue samples were suspended in a lysis buffer containing 0.5% SDS, 50% trifluoroethanol (TFE), 1 tablet of Complete mini EDTA-free Cocktail (Roche) in 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 8.5 and sonicated at 4 °C using a Bioruptor Pico sonicator (Diagenode, Seraing, Belgium – 10 cycles of 30 s ON and 30 s OFF) [19]. For the analysis of fresh-frozen GBM tissues we used a lysis buffer containing 0.5% SDS, 50% TFE, 2.5 mM ethylenediaminetetraacetic acid (EDTA), 2.5 mM ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid (EGTA), 1 tablet of Complete mini EDTA-free Cocktail (Roche) in 50 mM HEPES pH 8.5. The protein content was quantified with the Qubit protein assay standard procedure (Life Technologies). Samples were mixed with 2 µL of paramagnetic beads (20 mg/mL or 100 mg/mL solution of 50% Speedbeads A (GE45152105050250, Sigma) and 50% Speedbeads B (GE65152105050250, Sigma)). Proteins were denaturated at 95 °C for 5 min. Reduction and alkylation were performed by adding 2 µL of 200 mM DTT for 30 min at 45 °C, followed by 2 µL of 400 mM IAM for 30 min at RT in the dark. The alkylation reaction was quenched with 2 µL of 200 mM DTT.

Prior to the protein digestion step the proteins were purified on the beads by adding 100% ACN to get a final concentration of 50% ACN. After incubation for 8 min at RT the samples were placed on a magnet and incubated for 2 min. The supernatant was then discarded and the beads rinsed with 200 µL of 70% ethanol (EtOH). After 30 s of incubation on the magnet the supernatant was discarded and the EtOH wash was repeated one more time. Purified proteins were then eluted from the beads with 10 µL of 50 mM HEPES pH 8 and sonication at 18 °C for 5 min with the Bioruptor Pico. Protein digestion

was performed overnight at 37 °C by adding 2 µL of a Trypsin/LysC equimolar solution in 50 mM HEPES pH 8 (1:25 enzyme/protein). Peptides were then purified on the beads by adding 100% ACN to get a final concentration of 95% ACN. After incubation for 8 min at RT the samples were placed on a magnet and incubated for 2 min. The supernatant was then discarded and the beads rinsed with 180 µL of 100% ACN. After 30 s of incubation on the magnet the supernatant was discarded. The purified peptides were eluted from the beads with 10 µL of 2% DMSO and sonication at 18 °C for 5 min using the Bioruptor Pico. Samples were placed on the magnet, incubated for 2 min and the supernatant transferred to Lobind tubes (Eppendorf). Samples were then centrifuged at 20000g for 30 min to pellet any residual beads. The supernatant was then stored at -20 °C and diluted 1:1 with 10% formic acid prior to LC-MS/MS analysis.

3.3.2.4. Mice

The murine glioma GL261 cell line was grown in complete Dulbecco's modified Eagle's medium (DMEM) containing 10% newborn calf serum, 4.5 g/L glucose, 2 mM glutamine, 100 UI/mL penicillin and 100 mg/mL streptomycin at 37 °C in 5% CO_2 with media changes three times per week [21].

To induce glioma formation, C57BL/6 mice (12-14 weeks old) received a stereotaxically guided injection of 40000 GL261 cells into the visual cortex (2 mm lateral to the midline and in correspondence with lambda) using fine glass micropipettes (tip diameter 40 µm) [21][22].

GBM mouse brains were obtained using *in-situ* funnel-freezing [23] three weeks after GL261 cell inoculation. Briefly, animals were anesthetized by intraperitoneal injection using avertin (2,2,2-tri-bromoethanol 20 µL/g of body weight). An incision was made to expose the skull and a funnel was placed onto the skull; the skin was then raised around the funnel and secured with four sutures. Liquid nitrogen was slowly poured for 3 min and the entire animal was then frozen in liquid nitrogen. This procedure was applied to preserve the molecular integrity of the tissues as many biomolecules are known to undergo fast post-mortem degradation [23][24][25]. Extracted brains were stored at -80 °C until use.

All animal experiments conformed to the European Communities Council Directive n° 86/609/EEC and were approved by the Italian Ministry of Health.

All experiments were performed on two animals; laser capture microdissection, sample preparation and LC-MS/MS analysis were performed in technical triplicate for each animal.

3.3.2.5. Laser Capture Microdissection

Coronal tissue sections of 12 µm thickness were obtained using a Leica CM1950 cryostat and were then thaw mounted onto PEN membrane slides (previously conditioned in UV light for 30 minutes). After a light hematoxylin staining, laser capture microdissection (LCM) was performed using a PALM Technologies system (Carl Zeiss MicroImaging GmbH, Munchen, Germany) consisting of a PALM MicroBeam, a RoboStage and PALM RoboMover (PALM Robo software, version 4.6 Pro). LCM was performed using an x40 ocular lens. Small regions of tumor, healthy and border tissue, each of approximately 0.8 mm^2, were isolated and collected into LoBind tubes and then stored at -80 °C until use.

3.3.3. LC-MS/MS analysis

Peptides were suspended in 10% formic acid, injected and analyzed using an Easy-nLC1000 (Thermo Scientific) coupled to an Orbitrap Fusion (Thermo Scientific). For the comparison of sample preparation protocols (ISD, FASP, and SP3) peptides were first trapped online on a nanoviper trap column (2 cm x 100 µm, C18, 5 µm, 100 Å, Thermo Scientific) and separated using an Easyspray analytical column (ES803: 50 cm x 75 µm, C18, 2 µm, 100 Å, Thermo Scientific), a flow rate of 300 nL/min and a 100 min gradient. Briefly, peptides were loaded at 800 bar followed by a non-linear gradient: t=0-1 min, 5%B; t=76 min, 22%B; t=85 min, 32%B; t=90 min, 90%B; t=100 min, 90%B. Buffer A consisted of 0.1% formic acid and buffer B of 99.9% ACN and 0.1% formic acid. MS analysis was performed using a top speed 3 s cycle time, with an MS1 scan in the Orbitrap at 120K resolution, 375-1500 *m/z*, and an AGC target of 4e5. MS2 scans were performed

in the ion trap using a 1.6 *m/z* isolation window, HCD at 30% NCE, 2e3 AGC target and 300 ms maximum injection time.

Mouse GBM microdissected tissue samples were analyzed using a 75 min gradient: t=0-1 min, 5% B; t=51 min, 25% B; t=58 min, 35% B; t=64 min, 90% B; t=75 min, 90% B. Buffer A consisted of 0.1% formic acid and buffer B of 99.9% ACN and 0.1% formic acid. MS analysis was performed using a top speed 2 s cycle time, with an MS1 scan in the Orbitrap at 120K resolution, 375-1500 *m/z*, and an AGC target of 5e5. MS2 scans were performed in the ion trap using a 1.6 *m/z* isolation window, HCD at 32% NCE, 2e3 AGC target and 300 ms maximum injection time.

3.3.4. Data analysis

Raw data files were analyzed using Proteome Discoverer 2.1 (Thermo Scientific) with the SequestHT search engine. For the comparison of sample preparation protocols (ISD, FASP, SP3) using HeLa cells the LC-MS/MS spectra were compared with the Swissprot Homo sapiens database (ver2015-07-22, 42082 sequences) supplemented with a common contaminant database (246 sequences) allowing 2 missed cleavages, cysteine carbamidomethylation as static modification and methionine oxidation as dynamic modification, 20 ppm precursor mass tolerance and 0.6 Da fragment mass tolerance. For the microdissected mouse brain tumor samples the LC-MS/MS spectra were compared with the Swissprot Mus Musculus database (v2015-07-22, 24751 sequences) supplemented with the common contaminants database.

Results were filtered for 1% FDR using the Percolator algorithm [26] and additionally filtered for a minimum Xcorr score of 1.8. Protein intensities were calculated considering the three most abundant peptides with a co-isolation threshold of 50% and taking into account only unique peptides. Search results of the mouse brain tumor experiments were exported as txt files and processed with Perseus 1.5 [27]. Protein intensities were log2 transformed and subjected to a median normalization. Data were filtered such that each protein was quantified in all the three replicates for each brain region. Principal

Component Analysis was performed on the normalized and filtered datasets. Significantly different protein levels between the three different brain regions (tumor, healthy and border) were calculated using a two-sided Student's *t*-test using a permutation-based FDR cutoff (250 randomizations, FDR 0.001, S0 0.5). Gene ontology was performed with WebGestalt software [28] using the Overrepresentation Enrichment Analysis (ORA) method. The reference protein list included all proteins groups identified in all nine samples. A Benjamini-Hochberg (BH) method for multiple test adjustment was used and the FDR was set at 0.05.

3.4. Results and discussion

3.4.1. Protein digestion protocol comparison

We compared three protein digestion protocols using different sample amounts in to find which is the most suited for the analysis of small sample amounts. We compared In-Solution Digestion (ISD) with Filter-Aided Sample Preparation (FASP) and Single-Pot Solid Phase-enhanced Sample Preparation (SP3). We assessed the protocols using HeLa cell pellets containing 1e6, 1e5 and 5e3 cells. Figure 3.1. shows the number of protein groups and PSMs identified with the three protocols for the different amounts of HeLa cells.

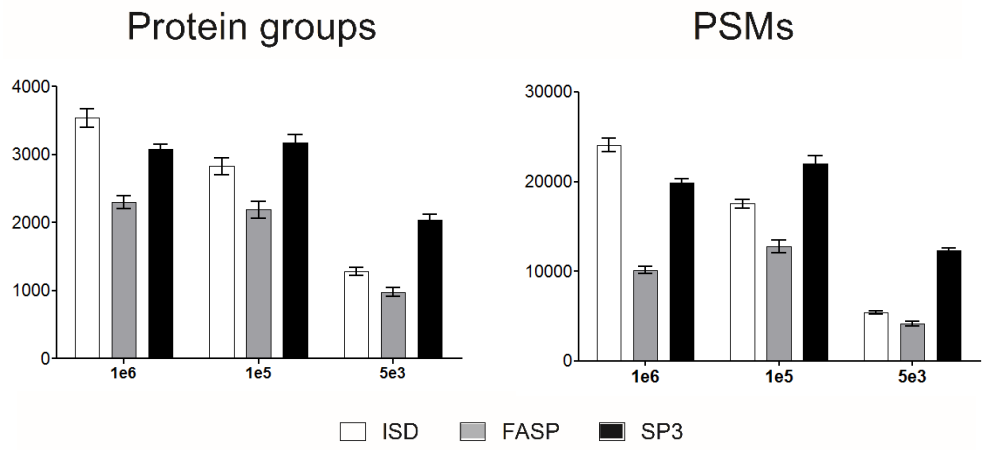


Figure 3.1. *Protein digestion protocol comparison. The bar graphs indicate the number of protein groups and PSMs identified using In-Solution Digestion (ISD, white), Filter-Aided-Sample Preparation (FASP, grey) and Single-Pot Solid Phase-enhanced Sample Preparation (SP3, black). Error bars represent standard deviation (N=3).*

The number of identifications for each protocol decreased with the amount of starting material. The ISD protocol led to the identification of 3544, 2830 and 1284 protein groups from the 1e6, 1e5 and 5e3 HeLa cell samples, respectively. The FASP protocol led to the identification of 2299, 2190 and 980 protein groups from the 1e6, 1e5 and 5e3 HeLa cell samples, respectively. The SP3 protocol provided the best results, with the identification of 3083, 3178 and 2036 protein groups from the 1e6, 1e5 and 5e3 HeLa cell samples, respectively. The decreasing number of protein groups with decreasing sample amounts was not observed for the 1e6 and 1e5 cells samples digested with the SP3 protocol, presumably because of the interference of DNA with the paramagnetic beads: we observed that when the number of cells is too high, the paramagnetic beads used in the SP3 protocol become coated in DNA (data not shown). This coating is thought to adversely affect the ability of the beads to bind the proteins, and thus hinders the efficiency of the SP3 method during the on-bead purification steps.

The results demonstrate that, in general, the number of identified proteins decreases with lower sample amounts. This is due to the fact that protein adsorption and losses during the different steps of the sample preparation protocols become more significant when the amount of starting material is very low. It is also known that digestion efficiency decreases with lower amounts of starting material, because of protein dilution (in the absence of a concomitant change in the working volume of the digestion). The FASP protocol provided the smallest number of protein identifications for all sample amounts investigated, suggesting that, despite the use of SDS to increase the protein extraction yield, a significant amount of material was lost during the different filtration steps. The ISD protocol provided the highest number of identifications for large amounts of starting material (1e6 HeLa cells), while the SP3 protocol outperformed both ISD and FASP for low sample amounts (1e5 and 5e3 HeLa cells).

We further investigated the digestion efficiency of the three protocols by examining the peptide charges distributions (Figure 3.2.A) and the number of missed cleavages (Figure 3.2.B).

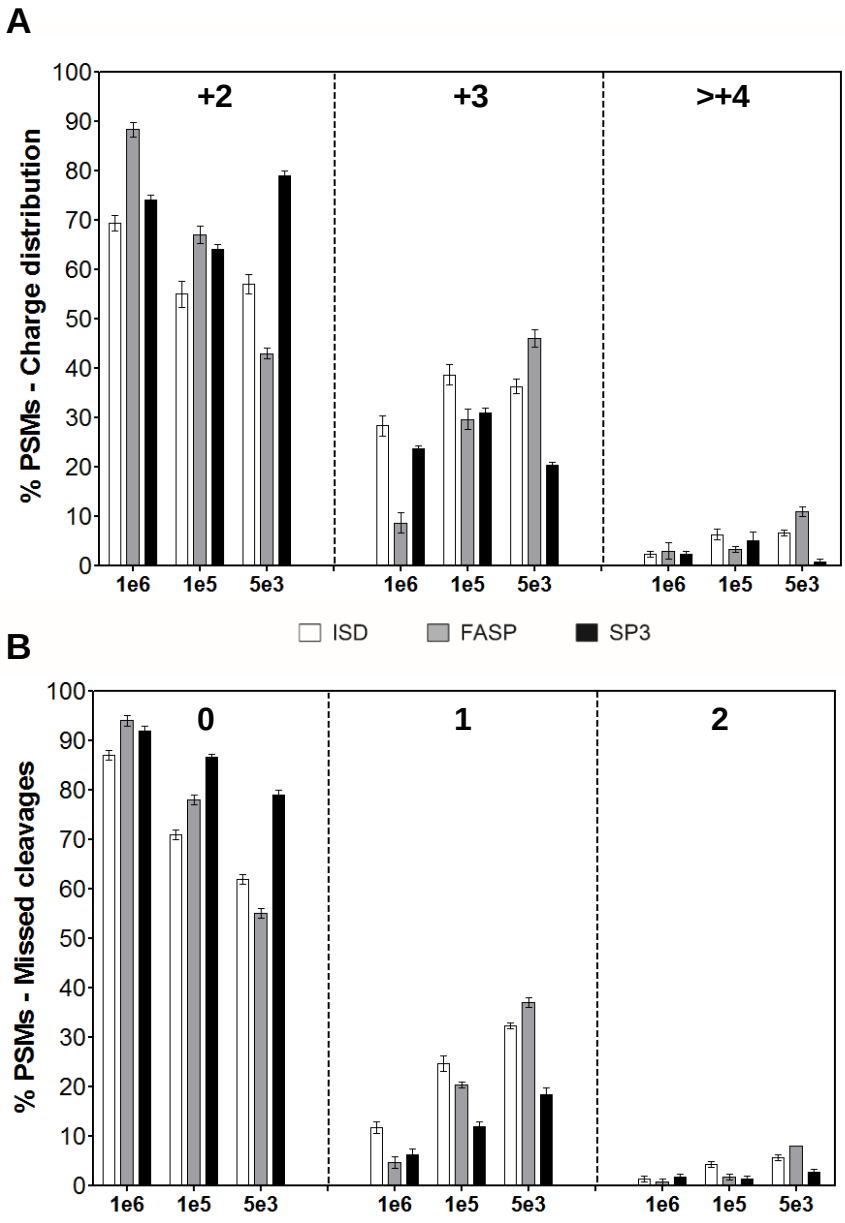


Figure 3.2. *Protein digestion efficiency comparisons for samples containing 1e6, 1e5 and 5e3 HeLa cells. A) Peptide charge state distribution: percentage of PSMs with +2, +3 and >+4 charge states. B) Missed cleavages: percentage of PSMs with 0, 1 and 2 missed cleavages. White, grey and black bars indicate ISD, FASP and SP3 protocols, respectively. Error bars represent standard deviation (N=3).*

The distributions of peptide charge states and missed cleavages provide insights into the efficiency of the digestion process. Across an entire proteome the peptide charge state distribution is related to the length of the proteolytic peptides, while the distribution of the missed cleavages indicate the efficiency of the proteolysis [29]. Tryptic peptides usually contain at least two positive charges: at the N-terminus and the C-terminal arginine/lysine residues (trypsin cleaves C-terminal to arginine and lysine residues). Higher charge states are due to the presence of basic residues that become protonated during the ESI process [30][31]. A complete tryptic digestion would lead to tryptic peptides in which the majority are doubly charged [32][33][34]. A higher percentage of higher charge state peptides indicates longer peptides and thus a less efficient digestion process. The percentage of doubly charged peptides and 0 missed cleavages was found to decrease with decreasing sample amounts (from 1e6 to 5e3 cells) except for the SP3 protocol, which gave the highest percentage of doubly charged peptides and 0 missed cleavages for the 5e3 cells sample. For high sample amounts (1e6 cells) the FASP protocol provided the highest percentage of doubly charged peptides and 0 missed cleavages (89.1% and 93.8%, respectively). Overall, all digestion protocols provided a high percentage of doubly charged peptides and 0 missed cleavages for the 1e6 and 1e5 cell samples, demonstrating a good digestion efficiency for high sample amounts. The observation that the FASP protocol generated a low number of identified proteins (despite a good digestion efficiency) indicates sample losses during the filtration steps and/or incomplete removal of SDS (which leads to ion suppression during the ionization process). For low sample amounts (5e3 cells), ISD and FASP provided a low percentage of doubly charged peptides (<60%) and 0 missed cleavages (<65%), indicating inefficient digestion. Conversely, the SP3 protocol showed a good digestion efficiency even for low sample amounts, with approximately 80% doubly charged peptides and 80% zero missed cleavages.

Figure 3.3 shows Venn diagrams indicating the overlapping protein groups identified with the different sample preparation protocols for the three sample amounts.

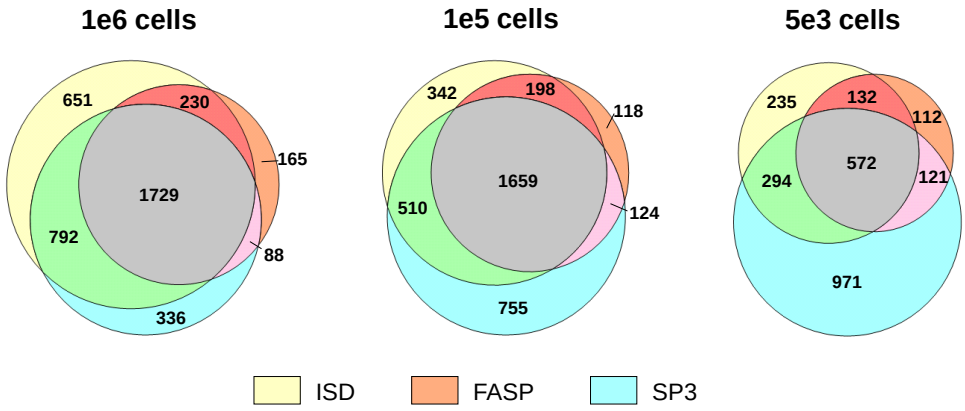


Figure 3.3. *Venn diagrams showing the overlap in protein groups between the ISD (yellow circle), FASP (orange circle) and SP3 (cyan circle) protocols for the 1e6, 1e5 and 5e3 HeLa cell samples.*

The overlap between the ISD, FASP and SP3 protocols was 43% and 45% for the 1e6 and 1e5 cell samples, respectively. The overlap decreased to 23% for the 5e3 cell sample because the SP3 method was much more effective for this low sample amount, and led to the identification of many more proteins (the overlap between ISD and FASP was still 48%). The differences in proteome coverage demonstrate a higher sensitivity of the SP3 protocol for experiments characterized by low sample amounts, with respect to the ISD and FASP protocols. In order to confirm that differences in proteome coverage were due to a higher sensitivity of the SP3 protocol rather than a bias towards specific cellular compartments, we performed a gene ontology analysis to classify the cellular localization of the protein groups identified with the three protocols. We checked the percentage of protein groups that are linked to the cellular components: nucleus, membrane, mitochondrion, cytoskeleton, endoplasmatic reticulum, chromosome, Golgi apparatus and ribosome (Figure 3.4.A-C).

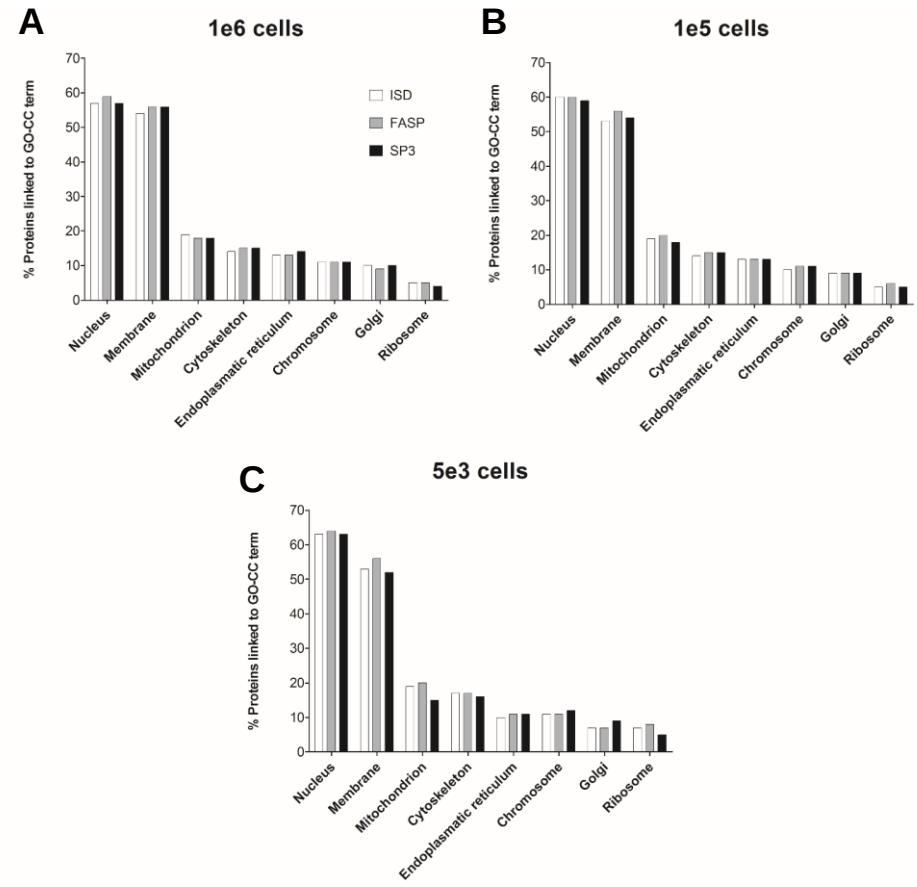


Figure 3.4. *Gene ontology cellular component analysis. The bar graphs show the percentage of protein groups that are linked to a GO term for the 1e6 (A), 1e5 (B) and 5e3 (C) cell samples. White, grey and black bars indicate ISD, FASP and SP3 protocols, respectively.*

The coverage of cellular compartments was similar for all three sample preparation protocols and for all sample amounts investigated. This indicates that the different protocols did not bias protein extraction, and thus the higher proteome coverage obtained with the SP3 protocol for low sample amounts was due to higher detection sensitivity, which in turn may be the result of: (i) a lower number of sample transfers and sample losses, (ii) higher sensitivity due to the unbiased interaction between proteins/peptides and the carboxylated beads, (iii) a higher digestion efficiency.

On the basis of these results we decided to adopt the SP3 protocol as the digestion protocol of choice for the analysis of small sample amounts.

3.4.2. SP3 digestion protocol optimization

We optimized the SP3 digestion protocol to further increase proteome coverage from low sample amounts. The initial version of the SP3 protocol was performed in 96-well plates using neodymium ring magnets (outer diameter 10 mm, inner diameter 4 mm, height 5 mm, magnetization N42, strength 24.5 N). We decided to test the SP3 protocol in PCR tubes because with the 96 well plates it was not possible to sonicate the samples with the Bioruptor Pico. When using the 96 well plate format samples had to be lysed and sonicated in an Eppendorf tube and then transferred to the 96 well plate. Furthermore the ring magnets were not strong enough to retain all the beads during the purification steps. Thus, we performed the SP3 digestion protocol in PCR tubes (Sarstedt) and using neodymium cube magnets (length 12 mm, magnetization N48, strength 6.3 Kg). With this configuration it was possible to completely eliminate sample transfers during the procedure (the samples were kept in the same PCR tube from cell lysis to the final peptide cleanup) and to reduce the amount of beads lost during the purification steps (the amount of beads in the purified samples was negligible).

We also increased the number of beads used for the SP3 protocol to increase protein recovery. The original SP3 protocol used a stock solution of 20 mg/mL carboxylate modified paramagnetic beads, and we tested a 5x (100 mg/mL) and 10x (200 mg/mL) more concentrated bead solutions. The 10x bead solution was found to be too concentrated for the sample volumes used here (10-20 µL), as the sample solution was too viscous and many beads were lost during the purification steps. The 5x bead solution led to higher protein recovery and a higher number of identified protein groups, compared with the 1x bead solution. Figure 3.5. shows the number of protein groups identified with the final optimized SP3 protocol performed on 1 µg of HeLa cells (~3300 cells) and those obtained with the first version of the SP3 protocol applied on 300 µg of HeLa cells (~1e6 cells).

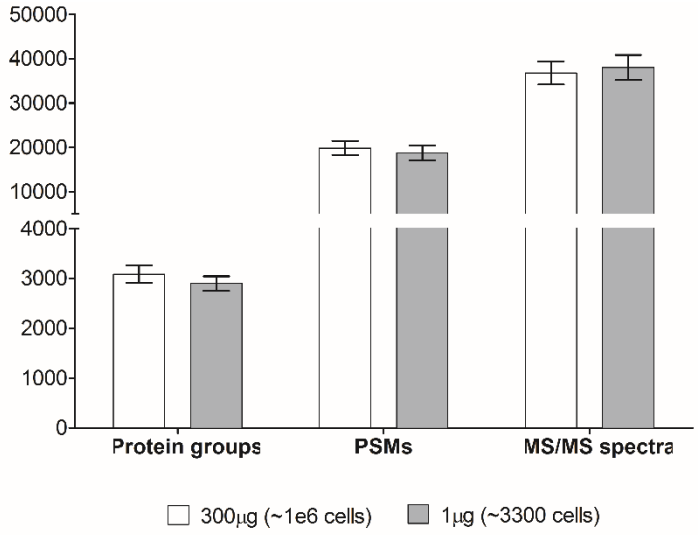


Figure 3.5. *Comparison between the optimized SP3 protocol (grey bars) performed on 1 µg of HeLa cells and the initial protocol (white bars) performed on 300 µg of HeLa cells. Number of identified protein groups, PSMs and acquired MS/MS spectra are reported. Error bars represent standard deviation (N=3).*

The results obtained with the optimized SP3 protocol on 1 µg of HeLa cells were with those obtained with the initial (non-optimized) SP3 protocol performed on 300 µg of HeLa cells. Almost 3000 protein groups were identified from 1 µg of HeLa cells, corresponding to approximately 35% more identifications when compared with the initial SP3 version (re. figure 3.1).

We then assessed the intra-batch and inter-batch reproducibility of the optimized SP3 digestion protocol by analyzing 1 µg of HeLa cells in triplicate on different days. Figure 3.6. shows the number of identified protein groups, PSMs and acquired MS/MS spectra of two batches of triplicates performed on different days and the average between the two batches. The variability in the number of protein groups was found to be less than 6% (coefficient of variance) for both the intra- and inter-batch comparisons, demonstrating that the SP3 protocol is highly reproducible.

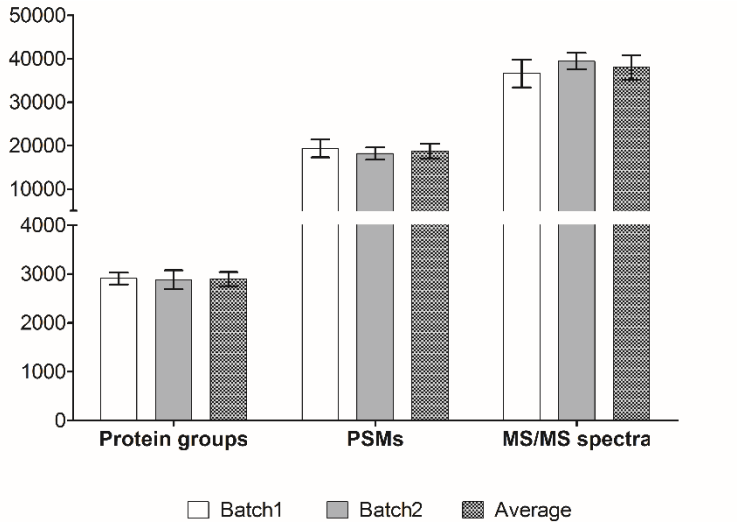


Figure 3.6. *Reproducibility of the SP3 protocol on 1 µg of HeLa cells (~3300 cells). Number of identified protein groups, PSMs and acquired MS/MS spectra for two sets of triplicates performed on different days (batch 1 and batch 2) and inter-day average. Error bars represent standard deviation (N=3 for batch1 and batch 2, N=6 for average).*

3.4.3. Characterization of a mouse model of glioblastoma

To investigate the potential of the ultrasensitive proteomics technique for molecular pathology we applied the optimized SP3 protocol to characterize the proteome of a mouse model of glioblastoma. The proteomics analysis of tissues is more challenging than the analysis of cell lines because:

i) the extraction and solubilization of proteins from tissue requires stronger conditions [19];

ii) laser microdissection has to be used to isolate specific histological regions of interest [35][36];

iii) the amount of protein extracted from small samples of microdissected tissue can be very small [37][38] and variable.

In this study we applied the optimized SP3 protocol to study the proteome of different histological regions of a mouse model of GBM. We used laser capture microdissection to excise 0.8 mm^2 regions from the tumor mass, tumor border and contralateral non-tumor region (Figure 3.7.). We isolated three technical replicates for each tissue region, isolated from three serial sections of the same mouse brain. The protein content estimation was performed on 0.8 mm^2 regions excised from an additional serial brain section using the Qubit protein quantitation assay. After LCM, samples were lysed and digested using the optimized SP3 protocol.

Figure 3.7. *Tissue section of the GBM containing mouse brain, stained with hematoxylin and eosin. Red, green and blue areas indicate ROIs isolated from the tumor, tumor border and contralateral non-tumor region, respectively, from an adjacent tissue section.*

The LC-MS/MS analysis of the small microdissected tissue samples resulted in the identification of 2903, 1814 and 2109 protein groups from the tumor mass, tumor border and non-tumor regions, respectively. Figure 3.8 summarizes the number of identified protein groups, PSMs and MS/MS spectra. The results demonstrate the high sensitivity of

the optimized SP3 protocol even when applied to small microdissected tissue samples. Furthermore, the optimized protocol provided a higher proteome coverage compared to other recently published tissue proteomics investigations [39][40][41][42][43].

While the average number of identifications demonstrated the high sensitivity of the optimized SP3 protocol, the variability in the number of identifications indicated that the tissue analysis workflow required further optimization. While the variability of the number of protein groups identified from the tumor region was very low (CV<4%), for the tumor border and non-tumor regions the variability was higher (CV = 12% and 16%, respectively).

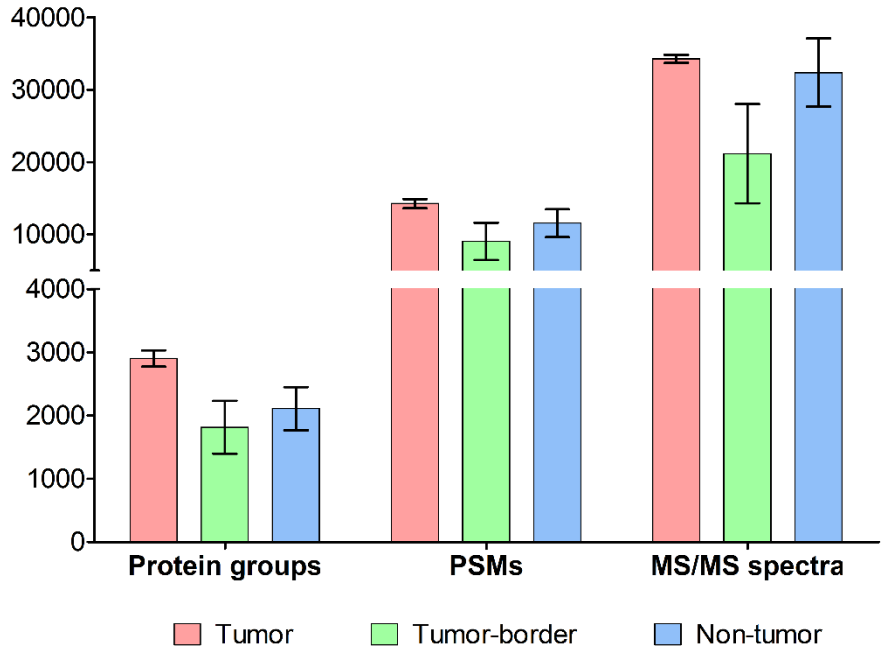


Figure 3.8. *Protein identification metrics: number of protein groups, PSMs and MS/MS spectra acquired from 0.8 mm^2 microdissected tissue samples from the tumor (red), tumor border (green) and contralateral non-tumor (blue) regions. Error bars indicate standard deviation (N=3).*

The variability can be due to the heterogeneity in the brain tissue, in the number and type of cells, and thus the amount and type of protein extracted per unit of area. The tumor mass is characterized by very high cellularity and, consequently, is histologically quite

homogenous. The cellular make-up of the border regions and the healthy regions is highly dependent on the anatomical region, and can differ between serial sections. Such cellular variation will contribute to differences in the levels and identities of the proteins that can be identified from localized regions of tissue. In this experiment we did not quantify the protein content of the microdissected tissue sample used for the LC-MS/MS analysis, but instead used sequential tissue sections for the protein estimation because the volume required for the Qubit protein quantitation assay was 20 µL (which represents the whole sample volume). For this reason, a protein quantification assay that allows the quantification of the protein content of each individual sample (subject to LC-MS/MS) is needed to reduce the variability due to tissue heterogeneity. This issue will be addressed in the next Chapter.

We performed a statistical analysis on the protein expression levels to study the changes in the proteome associated with the tumor. Figure 3.9. shows the score plots following a Principal Component Analysis (PCA) of the normalized protein expression levels. It can be seen that the tumor samples formed a well-defined cluster, separated by PC1 from the tumor border and non-tumor samples. The high percentage of variance explained by PC1 (67.9%) clearly demonstrated that the tumor proteome is highly different, in terms of protein expression levels, from the tumor border and non-tumor regions. The tumor border and non-tumor samples were also separated by PC1, in which the tumor border lies between the healthy control regions and the tumor core (in agreement with the results of spatially correlated proteomics [44], which have demonstrated that tumor associated changes can extend beyond the histological borders into adjacent healthy tissues).

In this study we considered as tumor border the tissue region within 150 µm of the tumor mass. Furthermore, the border region was isolated from all around the tumor. Recent work from our group using imaging mass spectrometry has demonstrated that the border region, as defined by having a perturbed proteome signature, may not be uniformly distributed around the tumor and maybe highly localized to the tumor interface zone [45]. Accordingly a combination of imaging mass spectrometry to define the affected tumor border, and LCM and LC-MS/MS to quantify the region-specific changes in protein

expression, would better uncover the molecular changes that occur in the tumor border region.

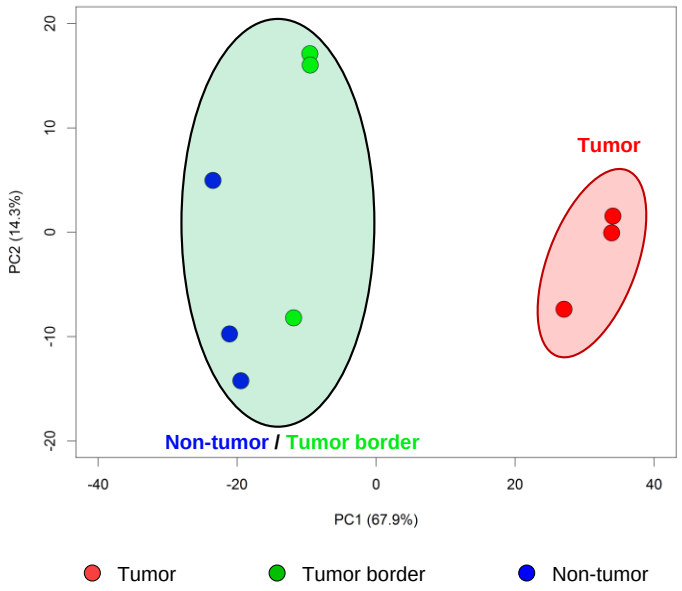


Figure 3.9. *Principal Component Analysis of the quantitative proteomics data from the tumor, tumor border and non-tumor samples. Red, green and blue circles indicate tumor, tumor border and non-tumor samples, respectively.*

Significantly different protein levels between non-tumor, tumor and tumor-border regions were calculated using a two-sided Student's *t*-test with FDR correction for multiple testing. Figure 3.10. show the corresponding volcano plots, in which the significance (expressed as –log of the calculated p-value) is plotted as a function of fold change.

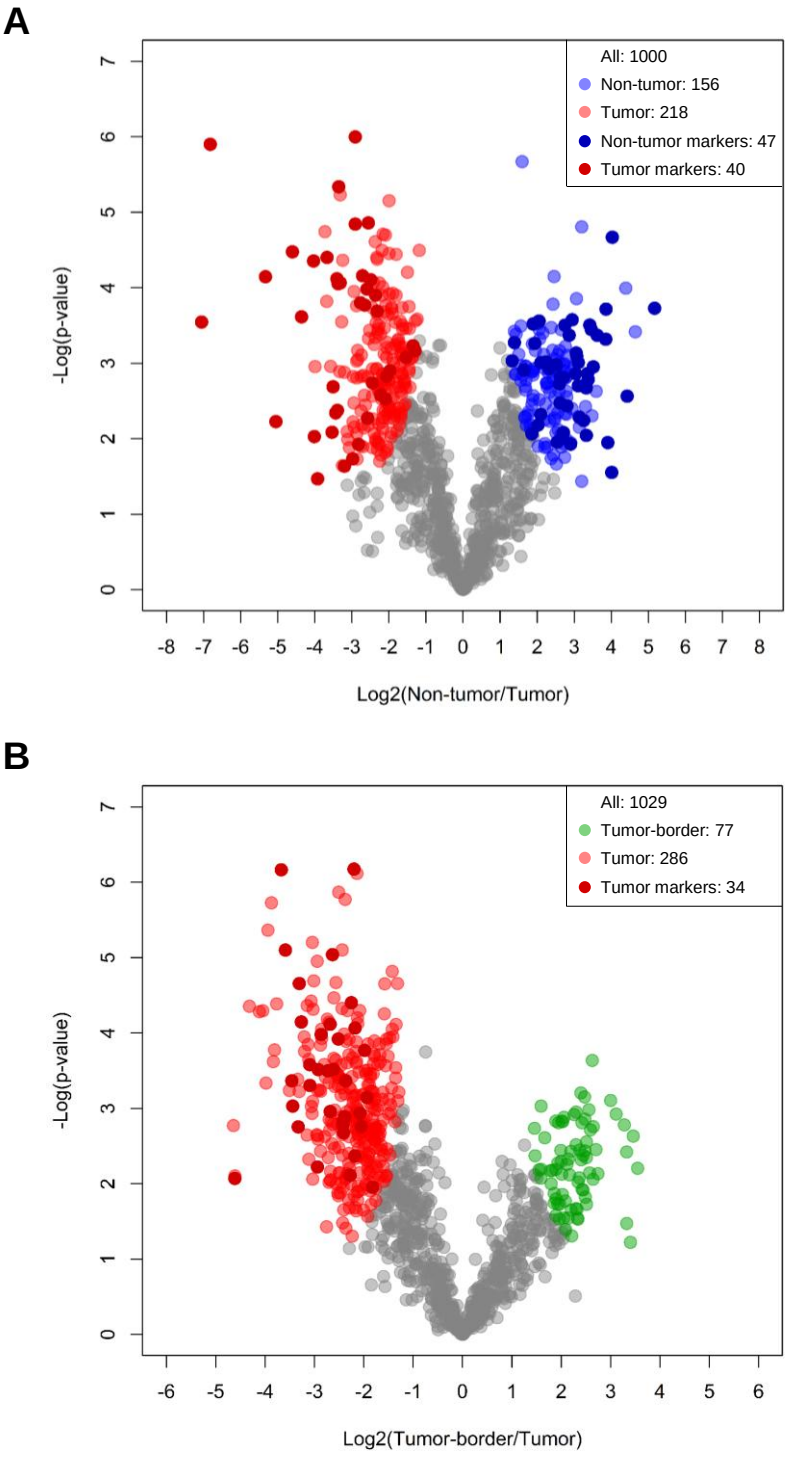
When comparing the protein expression levels between non-tumor and tumor regions almost 40% of all quantified proteins were deregulated. Specifically, in the non-tumor region 156 protein groups were significantly up-regulated and 218 up-regulated in the tumor region. Moreover, we detected 87 marker proteins known to be up-regulated in the non-tumor region or in the tumor region [46][47][48][49][50][51][52][53][54][55][56]

[57][58][59][60][61][62][63][64][65][66][67][68][69], demonstrating the reliability of the data.

The comparison between the tumor-border and tumor region showed 286 protein groups significantly up-regulated in the tumor region and 77 up-regulated in the tumor-border region. Also in this case we detected 34 known tumor marker proteins among the protein groups up-regulated in the tumor region.

We also compared the protein expression levels between the non-tumor and tumor-border regions. As discussed above the quantitative proteomics data from the healthy regions and the tumor border were more variable, reflecting the underlying cellular heterogeneity and the difficulty of defining the border zone affected by the tumor. This variability resulted in an increase in the standard deviation between the replicates. While it would have been possible to remove samples as outliers, such removal would be too arbitrary. Thus, we decided to perform the Student's *t*-test considering all samples per region, in order to avoid false positives. The high variability in the dataset meant that only 5 proteins were found to be significantly up-regulated in the tumor-border region (with respect to the healthy region). The proteins up-regulated in the tumor-border region included GFAP, VIM, MSN, and YBX1, which are known to be involved in glioma progression, malignancy and poor survival [63][68][69].

Gene Ontology (GO) analysis was performed to determine which pathways are represented in the differentially expressed proteins. We performed an Overrepresentation Enrichment Analysis (ORA) using WebGestalt on the protein groups that were found up-regulated in the non-tumor, tumor-border and tumor regions in the non-tumor vs tumor and tumor-border vs tumor pairwise comparisons. The low number of differentially expressed proteins between the non-tumor and tumor-border regions was insufficient for the GO analysis. Figure 3.11. shows the enriched KEGG pathways, in which the bubble graphs show the number of differentially expressed proteins contributing to the terms and the FDR value.


A
All: 1000
Non-tumor: 156
Tumor: 218
Non-tumor markers: 47
Tumor markers: 40
-Log(p-value)
Log2(Non-tumor/Tumor)
B
All: 1029
Tumor-border: 77
Tumor: 286
Tumor markers: 34
-Log(p-value)
Log2(Tumor-border/Tumor)

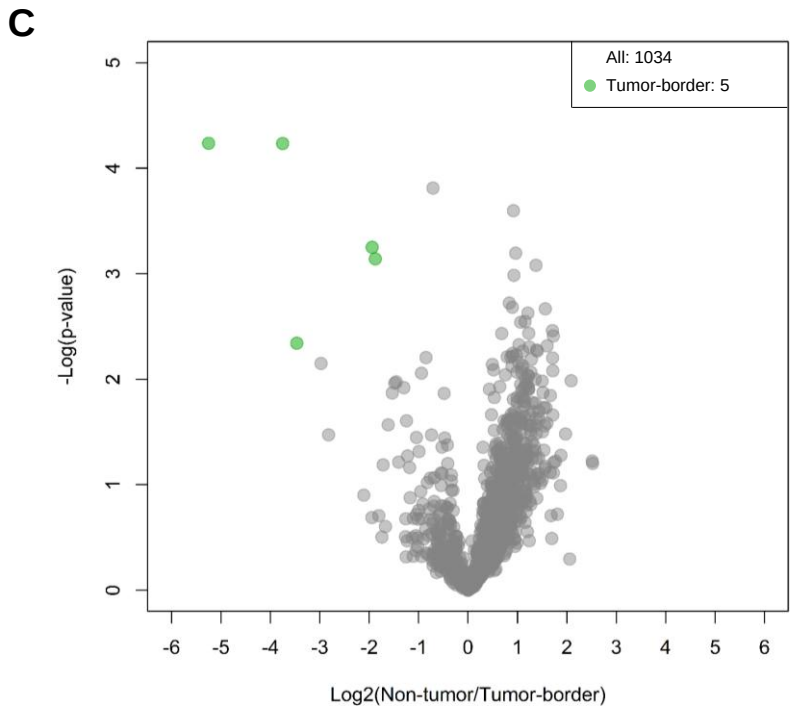


Figure 3.10. *Comparison of protein expression in tumor, non-tumor and tumor-border regions. Volcano plots for non-tumor vs tumor (A), tumor-border vs tumor (B) and non-tumor vs tumor-border (C). Grey circles indicate protein-groups with non-significant changes in protein expression (log2(fold change) <1 and >-1, FDR >0.001). Blue circles indicate protein groups significantly up-regulated in the non-tumor region (log2(fold change) >|-1|, FDR <0.001). Red circles indicate protein groups significantly up-regulated in the tumor region (log2(fold change) >|-1|, FDR <0.001). Green circles indicate protein groups significantly up-regulated in the tumor-border region (log2(fold change) >|-1|, FDR <0.001). Dark blue and dark red circles indicate markers of non-tumor and tumor regions, respectively.*

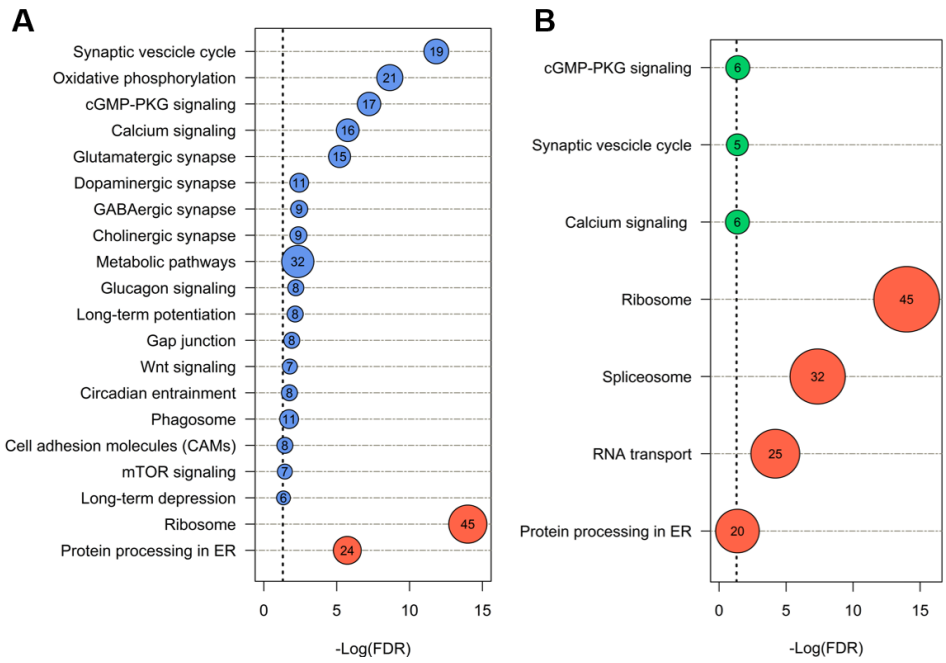


Figure 3.11. *Gene ontology Overrepresentation Enrichment Analysis (ORA) performed on the protein groups that were significantly up-regulated in the non-tumor, tumor-border and tumor regions for the non-tumor vs tumor (A) and tumor-border vs tumor (B) pairwise comparisons. The number in the bubble indicates the number of protein groups that contributed to the KEGG pathway. Blue, red and green bubbles indicate KEGG pathways enriched in the protein groups that were significantly up-regulated in the non-tumor, tumor and tumor-border region, respectively.*

The GO analysis revealed that the main pathways enriched in the up-regulated protein groups in the tumor region are related to gene expression regulation, translation and protein processing in the endoplasmatic reticulum. We found 45 ribosomal proteins up-regulated in the tumor region. Ribosomes are the translational machinery for protein synthesis from messenger RNA. Ribosome biogenesis and protein translation are finely tuned to regulate cell growth, proliferation and differentiation during development [70]. However, this type of regulation is abnormal in cancer cells. Ribosomal proteins of both small and large subunits (40S and 60S) have been found to be expressed at higher levels in a number of different cancers, including glioblastoma and other brain tumors [71][72][73].

Among these, we found an up-regulation in the tumor region of ribosomal proteins S11, S15a and S20, which have been reported to be related to glioblastoma cell proliferation and poor patient prognosis [74][75][76].

The main pathways enriched in the protein groups up-regulated in the non-tumor region were related to synaptic transmission (*e.g.* synaptic vescicle cycle, cGMP-PKG signaling, calcium signaling, long-term potentiation, Wnt signaling) and metabolic processes (*e.g.* metabolic processes, oxidative phosphorylation, glucagon signaling). Several marker proteins of different types of synapses (glutamatergic, dopaminergic, GABAergic, and cholinergic synapses) were found to be up-regulated in the non-tumor and tumor-border region, which is to be expected as GBM is a glial cell tumor [46][77]. We also found a down-regulation in the tumor region of several solute carrier transporter proteins, including SLC1A2, SLC17A7 and SLC8A2, which have already been shown to be down-regulated in glioblastoma and to be involved in metastasis and tumor growth [50][51].

Among the pathways up-regulated in the non-tumor region we also found oxidative phosphorylation pathway. The down-regulation of oxidative phosphorylation in the tumor region is in agreement with the Warburg effect, in which tumor cells metabolism switches toward aerobic glycolysis [78]. Moreover, the down-regulation in the tumor region of several subunit of the mitochondrial complex 1 (*e.g.* NDUFA10, NDUFB9, NDUFB10, NDUFS3, NDUFS4, NDUFS5) is consistent with the mitochondrial dysfunction in glioblastoma cells reported in previous studies [79][80][81].

3.5. Conclusions

We optimized a protein sample preparation protocol for the analysis of small sample amounts and microdissected samples. We compared three different protocols on different amounts of starting material to determine the best suited for the processing of small sample amounts. We compared the SP3 and FASP digestion protocols with the ISD protocol. We observed a dramatic decrease in the number of proteins and peptides identified with ISD when the amount of starting material was reduced from 1e6 to 5e3 HeLa cells. We observed the same trend when using the FASP protocol. The SP3 protocol provided the highest number of protein identifications for the 5e3 cell sample, with a 2-fold increase in the number of protein groups and PSMs identified (2036 and 12327, respectively) compared to the FASP and ISD protocol. The SP3 protocol was also found to have the highest digestion efficiency with 80% of peptides having zero missed cleavages for the 5e3 cells sample. These results clearly indicated that the SP3 protocol is that most suited for the processing of low sample amounts. The high sensitivity of the SP3 protocol is due to the fact that it is a single-pot protocol, in which sample losses are minimized because of the use of carboxylate-coated paramagnetic beads to retain proteins and peptides and thus reduces sample-transfer-associated losses.

We then tested different bead concentrations, different magnets and sample holders to reduce losses during the SP3 procedure, and thus increase further the proteome coverage from low sample amounts. In the final optimized conditions, the SP3 protocol led to the identification of 3000 protein groups from 1 µg of HeLa cells (3300 cells), the same number obtained from the analysis of 1e6 cells with the initial, non-optimized SP3 protocol.

As a proof of principle, the SP3 digestion protocol was tested on 0.8 mm^2 microdissected tissues samples from a mouse model of glioblastoma. We characterized the proteome extracted from the tumor, non-tumor and tumor-border brain regions. The optimized SP3 protocol provided the identification of more than 2000 protein groups per sample, providing a depth of coverage much greater than that in recent tissue proteomics publications. More than 300 protein groups were found to be significantly different

between the tumor and non-tumor regions and between the tumor and tumor-border regions. We also detected 87 marker proteins known to be up-regulated in the non-tumor region or in the tumor region, demonstrating the high reliability of the data.

The comparison between the non-tumor and tumor-border regions highlighted only a few proteins, all of them related to tumor progression and metastasis. This is thought to be due to the difficulty of defining the border regions affected by the tissue, which is affecting the variability within the replicates. Future work will focus on using imaging mass spectrometry to define the affected tumor borders, which can then be characterized further by the SP3 method (the optimized protocol reported in this thesis, or further refined variations thereof). Altogether, the results demonstrated that the SP3 protocol provided sufficient sensitivity to study the proteomes of specific histological regions of complex and heterogeneous tissues.

References

1. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature*, **422** (6928), 198–207.

2. Altelaar, A.F.M., Munoz, J., and Heck, A.J.R. (2012) Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.*, **14** (1), 35–48.

3. Ong, T.H., Kissick, D.J., Jansson, E.T., Comi, T.J., Romanova, E. V, Rubakhin, S.S., and Sweedler, J. V (2015) Classification of Large Cellular Populations and Discovery of Rare Cells Using Single Cell Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. *Anal. Chem.*, **87** (14), 7036–7042.

4. Di Palma, S., Stange, D., van de Wetering, M., Clevers, H., Heck, A.J.R., and Mohammed, S. (2011) Highly Sensitive Proteome Analysis of FACS-Sorted Adult Colon Stem Cells. *J. Proteome Res.*, **10** (8), 3814–3819.

5. Han, M.H., Hwang, S.I., Roy, D.B., Lundgren, D.H., Price, J. V, Ousman, S.S., Fernald, G.H., Gerlitz, B., Robinson, W.H., Baranzini, S.E., Grinnell, B.W., Raine, C.S., Sobel, R.A., Han, D.K., and Steinman, L. (2008) Proteomic analysis of active multiple sclerosis lesions reveals therapeutic targets. *Nature*, **451** (7182), 1076–1081.

6. Umar, A., Kang, H., Timmermans, A.M., Look, M.P., Meijer-van Gelder, M.E., den Bakker, M.A., Jaitly, N., Martens, J.W.M., Luider, T.M., Foekens, J.A., and Paša-Tolić, L. (2009) Identification of a Putative Protein Profile Associated with Tamoxifen Therapy Resistance in Breast Cancer. *Mol. Cell. Proteomics*, **8** (6), 1278–1294.

7. Wiśniewski, J.R., Ostasiewicz, P., and Mann, M. (2011) High Recovery FASP Applied to the Proteomic Analysis of Microdissected Formalin Fixed Paraffin Embedded Cancer Tissues Retrieves Known Colon Cancer Markers. *J. Proteome Res.*, **10** (7), 3040–3049.

8. Andrews, G.L., Simons, B.L., Young, J.B., Hawkridge, A.M., and Muddiman, D.C. (2011) Performance Characteristics of a New Hybrid Quadrupole Time-of-Flight Tandem Mass Spectrometer (TripleTOF 5600). *Anal. Chem.*, **83** (13), 5442–5446.

9. Hebert, A.S., Richards, A.L., Bailey, D.J., Ulbrich, A., Coughlin, E.E., Westphall, M.S., and Coon, J.J. (2014) The One Hour Yeast Proteome. *Mol. Cell. Proteomics*, **13** (1), 339–347.

10. Kelstrup, C.D., Jersie-Christensen, R.R., Batth, T.S., Arrey, T.N., Kuehn, A., Kellmann, M., and Olsen, J. V. (2014) Rapid and Deep Proteomes by Faster Sequencing on a Benchtop Quadrupole Ultra-High-Field Orbitrap Mass Spectrometer. *J. Proteome Res.*, **13** (12), 6187–6195.

11. Beck, S., Michalski, A., Raether, O., Lubeck, M., Kaspar, S., Goedecke, N., Baessmann, C., Hornburg, D., Meier, F., Paron, I., Kulak, N.A., Cox, J., and Mann, M. (2015) The Impact II, a Very High-Resolution Quadrupole Time-of-Flight Instrument (QTOF) for Deep Shotgun Proteomics. *Mol. Cell. Proteomics*, **14** (7), 2014–2029.

12. Olsen, J. V, Schwartz, J.C., Griep-Raming, J., Nielsen, M.L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E.R., Senko, M., Makarov, A., Mann, M., and Horning, S. (2009) A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. *Mol. Cell. Proteomics*, **8** (12), 2759–2769.

13. Tang, K., Shvartsburg, A.A., Lee, H.-N., Prior, D.C., Buschbach, M.A., Li, F., Tolmachev, A. V, Anderson, G.A., and Smith, R.D. (2005) High-Sensitivity Ion Mobility Spectrometry/Mass Spectrometry Using Electrodynamic Ion Funnel Interfaces. *Anal. Chem.*, **77** (10), 3330–3339.

14. Köcher, T., Swart, R., and Mechtler, K. (2011) Ultra-High-Pressure RPLC Hyphenated to an LTQ-Orbitrap Velos Reveals a Linear Relation between Peak Capacity and Number of Identified Peptides. *Anal. Chem.*, **83** (7), 2699–2704.

15. Pirmoradian, M., Budamgunta, H., Chingin, K., Zhang, B., Astorga-Wells, J., and

Zubarev, R.A. (2013) Rapid and Deep Human Proteome Analysis by Single-dimension Shotgun Proteomics. *Mol. Cell. Proteomics*, **12** (11), 3330–3338.

16. Yamana, R., Iwasaki, M., Wakabayashi, M., Nakagawa, M., Yamanaka, S., and Ishihama, Y. (2013) Rapid and Deep Profiling of Human Induced Pluripotent Stem Cell Proteome by One-shot NanoLC–MS/MS Analysis with Meter-scale Monolithic Silica Columns. *J. Proteome Res.*, **12** (1), 214–221.

17. López-Ferrer, D., Cañas, B., Vázquez, J., Lodeiro, C., Rial-Otero, R., Moura, I., and Capelo, J.L. (2006) Sample treatment for protein identification by mass spectrometry-based techniques. *TrAC - Trends Anal. Chem.*, **25** (10), 996–1005.

18. Wiśniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods*, **6** (5), 359–362.

19. Hughes, C.S., Foehr, S., Garfield, D.A., Furlong, E.E., and Steinmetz, L.M. (2014) Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.*, **10** (757), 1–15.

20. Wisniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat Meth*, **6** (5), 359–362.

21. Vannini, E., Panighini, A., Cerri, C., Fabbri, A., Lisi, S., Pracucci, E., Benedetto, N., Vannozzi, R., Fiorentini, C., Caleo, M., and Costa, M. (2014) The bacterial protein toxin, cytotoxic necrotizing factor 1 (CNF1) provides long-term survival in a murine glioma model. *BMC Cancer*, **14** (449), 1–10.

22. Vannini, E., Olimpico, F., Middei, S., Ammassari-Teule, M., de Graaf, E.L., McDonnell, L., Schmidt, G., Fabbri, A., Fiorentini, C., Baroncelli, L., Costa, M., and Caleo, M. (2016) Electrophysiology of glioma: a Rho GTPase-activating protein reduces tumor growth and spares neuron structure and function. *Neuro. Oncol.*, **18** (12), 1634–1643.

23. Mulder, I.A., Esteve, C., Wermer, M.J.H., Hoehn, M., Tolner, E.A., den Maagdenberg, A.M.J.M., and McDonnell, L.A. (2016) Funnel-freezing versus heat-stabilization for the visualization of metabolites by mass spectrometry imaging in a

mouse stroke model. *Proteomics*, **16** (11–12), 1652–1659.

24. Hattori, K., Kajimura, M., Hishiki, T., Nakanishi, T., Kubo, A., Nagahata, Y., Ohmura, M., Yachie-Kinoshita, A., Matsuura, T., Morikawa, T., Nakamura, T., Setou, M., and Suematsu, M. (2010) Paradoxical ATP Elevation in Ischemic Penumbra Revealed by Quantitative Imaging Mass Spectrometry. *Antioxid. Redox Signal.*, **13** (8), 1157–1167.

25. Sugiura, Y., Honda, K., Kajimura, M., and Suematsu, M. (2013) Visualization and quantification of cerebral metabolic fluxes of glucose in awake mice. *Proteomics*, **14** (7–8), 829–838.

26. Käll, L., Canterbury, J.D., Weston, J., Noble, W.S., and MacCoss, M.J. (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods*, **4** (11), 923–925.

27. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods*, **13** (9), 731–740.

28. Wang, J., Vasaikar, S., Shi, Z., Greer, M., and Zhang, B. (2017) WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.*, **45** (W1), 130–137.

29. Wierenga, S.K., Zocher, M.J., Mirus, M.M., Conrads, T.P., Goshe, M.B., and Veenstra, T.D. (2002) A method to evaluate tryptic digestion efficiency for high-throughput proteome analyses. *Rapid Commun. Mass Spectrom.*, **16** (14), 1404–1408.

30. Tang, X.J., Thibault, P., and Boyd, R.K. (1993) Fragmentation reactions of multiply-protonated peptides and implications for sequencing by tandem mass spectrometry with low-energy collision-induced dissociation. *Anal. Chem.*, **65** (20), 2824–2834.

31. Sonsmann, G., Römer, A., and Schomburg, D. (2002) Investigation of the influence of charge derivatization on the fragmentation of multiply protonated peptides. *J.*

Am. Soc. Mass Spectrom., **13** (1), 47–58.

32. Tang, X.-J., Boyd, R.K., and Bertrand, M.J. (1992) An investigation of fragmentation mechanisms of doubly protonated tryptic peptides. *Rapid Commun. Mass Spectrom.*, **6** (11), 651–657.

33. Tabb, D.L., Smith, L.L., Breci, L.A., Wysocki, V.H., Lin, D., and Yates, J.R. (2003) Statistical Characterization of Ion Trap Tandem Mass Spectra from Doubly Charged Tryptic Peptides. *Anal. Chem.*, **75** (5), 1155–1163.

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

35. Clair, G., Piehowski, P.D., Nicola, T., Kitzmiller, J.A., Huang, E.L., Zink, E.M., Sontag, R.L., Orton, D.J., Moore, R.J., Carson, J.P., Smith, R.D., Whitsett, J.A., Corley, R.A., Ambalavanan, N., and Ansong, C. (2016) Spatially-Resolved Proteomics: Rapid Quantitative Analysis of Laser Capture Microdissected Alveolar Tissue Samples. *Sci. Rep.*, **6**, 39223.

36. Mukherjee, S., Rodriguez-Canales, J., Hanson, J., Emmert-Buck, M.R., Tangrea, M.A., Prieto, D.A., Blonder, J., and Johann, D.J. (2013) Proteomic Analysis of Frozen Tissue Samples Using Laser Capture Microdissection BT - Proteomics for Biomarker Discovery (eds.Zhou, M., and Veenstra, T.), Humana Press, Totowa, NJ, pp. 71–83.

37. Gutstein, H.B., and Morris, J.S. (2007) Laser capture sampling and analytical issues in proteomics. *Expert Rev. Proteomics*, **4** (5), 627–637.

38. Altelaar, A.F.M., and Heck, A.J.R. (2012) Trends in ultrasensitive proteomics. *Curr. Opin. Chem. Biol.*, **16** (1), 206–213.

39. Johann, D.J., Rodriguez-Canales, J., Mukherjee, S., Prieto, D.A., Hanson, J.C., Emmert-Buck, M., and Blonder, J. (2009) Approaching Solid Tumor Heterogeneity on a Cellular Basis by Tissue Proteomics Using Laser Capture Microdissection and Biological Mass Spectrometry. *J. Proteome Res.*, **8** (5), 2310–2318.

40. Cha, S., Imielinski, M.B., Rejtar, T., Richardson, E.A., Thakur, D., Sgroi, D.C., and Karger, B.L. (2010) In Situ proteomic analysis of human breast cancer epithelial cells using laser capture microdissection (LCM)-LC/MS: Annotation by protein set enrichment analysis (PSEA) and gene ontology (GO). *Mol. Cell. Proteomics*, **9** (11), 2529–2544.

41. Liu, N.Q., Braakman, R.B.H., Stingl, C., Luider, T.M., Martens, J.W.M., Foekens, J.A., and Umar, A. (2012) Proteomics Pipeline for Biomarker Discovery of Laser Capture Microdissected Breast Cancer Tissue. *J. Mammary Gland Biol. Neoplasia*, **17** (2), 155–164.

42. Fang, X., Wang, C., Balgley, B.M., Zhao, K., Wang, W., He, F., Weil, R.J., and Lee, C.S. (2012) Targeted Tissue Proteomic Analysis of Human Astrocytomas. *J. Proteome Res.*, **11** (8), 3937–3946.

43. Zhang, L., Lanzoni, G., Battarra, M., Inverardi, L., and Zhang, Q. (2017) Proteomic profiling of human islets collected from frozen pancreata using laser capture microdissection. *J. Proteomics*, **150**, 149–159.

44. Oppenheimer, S.R., Mi, D., Sanders, M.E., and Caprioli, R.M. (2010) Molecular Analysis of Tumor Margins by MALDI Mass Spectrometry in Renal Carcinoma. *J. Proteome Res.*, **9** (5), 2182–2190.

45. Dilillo, M., Ait-Belkacem, R., Esteve, C., Pellegrini, D., Nicolardi, S., Costa, M., Vannini, E., Graaf, E.L. de, Caleo, M., and McDonnell, L.A. (2017) Ultra-High Mass Resolution MALDI Imaging Mass Spectrometry of Proteins and Metabolites in a Mouse Model of Glioblastoma. *Sci. Rep.*, **7** (1), 603–614.

46. Polisetty, R.V., Gautam, P., Sharma, R., Harsha, H.C., Nair, S.C., Gupta, M.K., Uppin, M.S., Challa, S., Puligopu, A.K., Ankathi, P., Purohit, A.K., Chandak, G.R., Pandey, A., and Sirdeshmukh, R. (2012) LC-MS/MS Analysis of Differentially Expressed Glioblastoma Membrane Proteome Reveals Altered Calcium Signaling and Other Protein Groups of Regulatory Functions. *Mol. Cell. Proteomics*, **11** (6), 1–15.

47. Patel, V.N., Gokulrangan, G., Chowdhury, S.A., Chen, Y., Sloan, A.E., Koyutürk, M., Barnholtz-Sloan, J., and Chance, M.R. (2013) Network Signatures of Survival in Glioblastoma Multiforme. *PLOS Comput. Biol.*, **9** (9), e1003237.

48. Takita, J., Chen, Y., Okubo, J., Sanada, M., Adachi, M., Ohki, K., Nishimura, R., Hanada, R., Igarashi, T., Hayashi, Y., and Ogawa, S. (2011) Aberrations of NEGR1 on 1p31 and MYEOV on 11q13 in neuroblastoma. *Cancer Sci.*, **102** (9), 1645–1650.

49. Bi, G., Yan, J., Sun, S., and Qu, X. (2017) PRRT2 inhibits the proliferation of glioma cells by modulating unfolded protein response pathway. *Biochem. Biophys. Res. Commun.*, **485** (2), 454–460.

50. Tong, H., Yu, X., Lu, X., and Wang, P. (2015) Downregulation of solute carriers of glutamate in gliosomes and synaptosomes may explain local brain metastasis in anaplastic glioblastoma. *IUBMB Life*, **67** (4), 306–311.

51. Qu, M., Yu, J., Liu, H., Ren, Y., Ma, C., Bu, X., and Lan, and Q. (2017) The Candidate Tumor Suppressor Gene SLC8A2 Inhibits Invasion, Angiogenesis and Growth of Glioblastoma. *Mol. Cells*, **40** (10), 761–772.

52. Gao, Y.-F., Mao, X.-Y., Zhu, T., Mao, C.-X., Liu, Z.-X., Wang, Z.-B., Li, L., Li, X., Yin, J.-Y., Zhang, W., Zhou, H.-H., and Liu, Z.-Q. (2016) COL3A1 and SNAP91: novel glioblastoma markers with diagnostic and prognostic value. *Oncotarget*, **7** (43), 70494–70503.

53. Balbous, A., Cortes, U., Guilloteau, K., Villalva, C., Flamant, S., Gaillard, A., Milin, S., Wager, M., Sorel, N., Guilhot, J., Bennaceur-Griscelli, A., Turhan, A., Chomel, J.-C., and Karayan-Tapon, L. (2014) A mesenchymal glioma stem cell profile is related to clinical outcome. *Oncogenesis*, **3** (3), e91.

54. Yanamandra, N., Gumidyala, K. V, Waldron, K.G., Gujrati, M., Olivero, W.C., Dinh, D.H., Rao, J.S., and Mohanam, S. (2004) Blockade of cathepsin B expression in human glioblastoma cells is associated with suppression of angiogenesis. *Oncogene*, **23** (12), 2224–2230.

55. Hassan, M.K., Kumar, D., Naik, M., and Dixit, M. (2018) The expression profile and prognostic significance of eukaryotic translation elongation factors in different cancers. *PLoS One*, **13** (1), e0191377.

56. Liang, H., Ding, X., Zhou, C., Zhang, Y., Xu, M., Zhang, C., and Xu, L. (2012) Knockdown of eukaryotic translation initiation factors 3B (EIF3B) inhibits proliferation and promotes apoptosis in glioblastoma cells. *Neurol. Sci.*, **33** (5), 1057–1062.

57. Ren, M., Zhou, C., Liang, H., Wang, X., and Xu, L. (2015) RNAi-Mediated Silencing of EIF3D Alleviates Proliferation and Migration of Glioma U251 and U87MG Cells. *Chem. Biol. Drug Des.*, **86** (4), 715–722.

58. Friedlein, K., Bozhkov, Y., Hore, N., Merkel, A., Sommer, B., Brandner, S., Buchfelder, M., Savaskan, N.E., and Eyüpoglu, I.Y. (2015) A new functional classification system (FGA/B) with prognostic value for glioma patients. *Sci. Rep.*, **5**, 12373.

59. Angelopoulou, E., Piperi, C., Adamopoulos, C., and Papavassiliou, A.G. (2016) Pivotal role of high-mobility group box 1 (HMGB1) signaling pathways in glioma development and progression. *J. Mol. Med.*, **94** (8), 867–874.

60. Deng, J., Chen, S., Wang, F., Zhao, H., Xie, Z., Xu, Z., Zhang, Q., Liang, P., Zhai, X., and Cheng, Y. (2016) Effects of hnRNP A2/B1 Knockdown on Inhibition of Glioblastoma Cell Invasion, Growth and Survival. *Mol. Neurobiol.*, **53** (2), 1132–1144.

61. Shen, F., Zhang, Y., Yao, Y., Hua, W., Zhang, H., Wu, J., Zhong, P., and Zhou, L. (2014) Proteomic analysis of cerebrospinal fluid: toward the identification of biomarkers for gliomas. *Neurosurg. Rev.*, **37** (3), 367–380.

62. Carbonell, W.S., DeLay, M., Jahangiri, A., Park, C.C., and Aghi, M.K. (2013) β1 Integrin Targeting Potentiates Antiangiogenic Therapy and Inhibits the Growth of Bevacizumab-Resistant Glioblastoma. *Cancer Res.*, **73** (10), 3145–3154.

63. Zhu, X., Morales, F.C., Agarwal, N.K., Dogruluk, T., Gagea, M., and Georgescu,

M.-M. (2013) Moesin Is a Glioma Progression Marker That Induces Proliferation and Wnt/β-Catenin Pathway Activation via Interaction with CD44. *Cancer Res.*, **73** (3), 1142–1155.

64. Beadle, C., Assanah, M.C., Monzo, P., Vallee, R., Rosenfeld, S.S., and Canoll, P. (2008) The role of myosin II in glioma invasion of the brain. *Mol. Biol. Cell*, **19** (8), 3357–3368.

65. Li, J., Yin, C., Okamoto, H., Mushlin, H., Balgley, B.M., Lee, C.S., Yuan, K., Ikejiri, B., Glasker, S., Vortmeyer, A.O., Oldfield, E.H., Weil, R.J., and Zhuang, Z. (2008) Identification of a novel proliferation-related protein, WHSC1 4a, in human gliomas. *Neuro. Oncol.*, **10** (1), 45–51.

66. Kwiatkowska, A., Didier, S., Fortin, S., Chuang, Y., White, T., Berens, M.E., Rushing, E., Eschbacher, J., Tran, N.L., Chan, A., and Symons, M. (2012) The small GTPase RhoG mediates glioblastoma cell invasion. *Mol. Cancer*, **11** (1), 65.

67. Xu, D., Liu, A., Wang, X., Chen, Y., Shen, Y., Tan, Z., and Qiu, M. (2018) Repression of Septin9 and Septin2 suppresses tumor growth of human glioblastoma cells. *Cell Death Dis.*, **9** (5), 514.

68. Lin, L., Wang, G., Ming, J., Meng, X., Han, B., Sun, B., Cai, J., and Jiang, C. (2016) Analysis of expression and prognostic significance of vimentin and the response to temozolomide in glioma patients. *Tumor Biol.*, **37** (11), 15333–15339.

69. Zheng, J., Dong, W., Zhang, J., Li, G., and Gong, H. (2016) YB-1, a new biomarker of glioma progression, is associated with the prognosis of glioma patients. *Acta Biochim. Biophys. Sin. (Shanghai).*, **48** (4), 318–325.

70. Ruggero, D., and Pandolfi, P.P. (2003) Does the ribosome translate cancer? *Nat. Rev. Cancer*, **3**, 179.

71. Barnard, G.F., Mori, M., Staniunas, R.J., Begum, N.A., Bao, S., Puder, M., Cobb, J., Redman, K.L., Steele, G.D., and Chen, L.B. (1995) Ubiquitin fusion proteins are overexpressed in colon cancer but not in gastric cancer. *Biochim. Biophys. Acta - Mol. Basis Dis.*, **1272** (3), 147–153.

72. Jing, L.M.-D. and X. (2007) Ribosomal Proteins and Colorectal Cancer. *Curr. Genomics*, **8** (1), 43–49.

73. Bastide, A., and David, A. (2018) The ribosome, (slow) beating heart of cancer (stem) cell. *Oncogenesis*, **7** (4), 34.

74. Zhang, C., Fu, J., Xue, F., Ryu, B., Zhang, T., Zhang, S., Sun, J., Xu, X., Shen, Z., Zheng, L., and Chen, X. (2016) Knockdown of ribosomal protein S15A induces human glioblastoma cell apoptosis. *World J. Surg. Oncol.*, **14**, 129.

75. Yao, Y., Liu, Y., Lv, X., Dong, B., Wang, F., Li, J., Zhang, Q., Xu, R., and Xu, Y. (2016) Down-regulation of ribosomal protein S15A inhibits proliferation of human glioblastoma cells in vivo and in vitro via AKT pathway. *Tumor Biol.*, **37** (4), 4979–4990.

76. Yong, W.H., Shabihkhani, M., Telesca, D., Yang, S., Tso, J.L., Menjivar, J.C., Wei, B., Lucey, G.M., Mareninov, S., Chen, Z., Liau, L.M., Lai, A., Nelson, S.F., Cloughesy, T.F., and Tso, C.-L. (2015) Ribosomal Proteins RPS11 and RPS20, Two Stress-Response Markers of Glioblastoma Stem Cells, Are Novel Predictors of Poor Prognosis in Glioblastoma Patients. *PLoS One*, **10** (10), e0141334–e0141334.

77. Liu, M., Xu, Z., Du, Z., Wu, B., Jin, T., Xu, K., Xu, L., Li, E., and Xu, H. (2017) The Identification of Key Genes and Pathways in Glioma by Bioinformatics Analysis. *J. Immunol. Res.*, **2017**, 1278081.

78. Weinhouse, S., Warburg, O., Burk, D., and Schade, A.L. (1956) On Respiratory Impairment in Cancer Cells. *Science (80-.).*, **124** (3215), 267 LP-272.

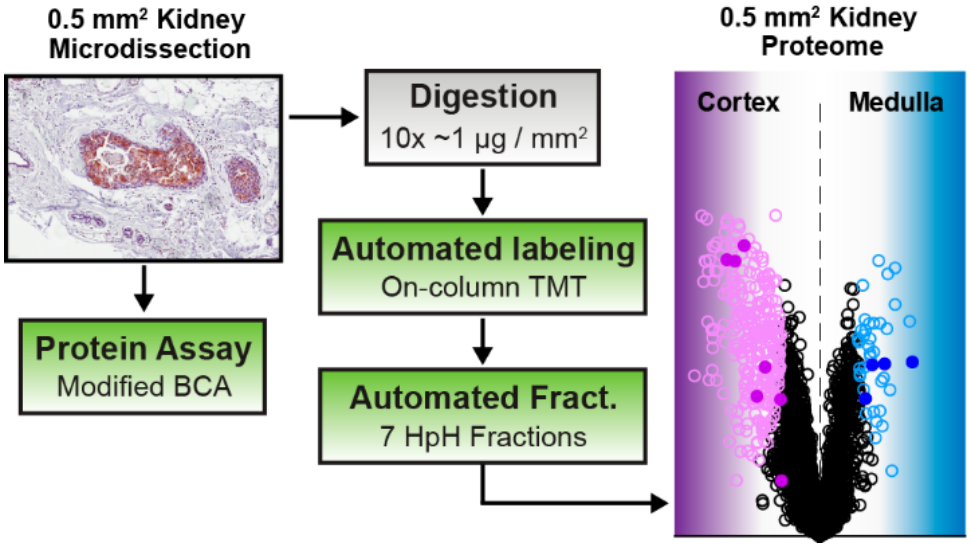
79. Agnihotri, S., and Zadeh, G. (2016) Metabolic reprogramming in glioblastoma: the influence of cancer metabolism on epigenetics and unanswered questions. *Neuro. Oncol.*, **18** (2), 160–172.

80. Deighton, R.F., Le Bihan, T., Martin, S.F., Gerth, A.M.J., McCulloch, M., Edgar, J.M., Kerr, L.E., Whittle, I.R., and McCulloch, J. (2014) Interactions among mitochondrial proteins altered in glioblastoma. *J. Neurooncol.*, **118** (2), 247–256.

81. Tang, H., Li, J., Liu, X., Wang, G., Luo, M., and Deng, H. (2016) Down-regulation of HSP60 Suppresses the Proliferation of Glioblastoma Cells via the ROS/AMPK/mTOR Pathway. *Sci. Rep.*, **6**, 28388.

Chapter 4

Development of a Quantitative Microproteomics Workflow for the Analysis of Small Microdissected Tissue Samples: Application to Mouse Kidney Substructures



4.1. Summary

A quantitative microproteomics workflow for the analysis of small tissue samples is described. The conditions for in-solution and on-column tandem mass tags (TMT) labeling, for accurate relative quantification, were first optimized. It was found that the optimal conditions for in-solution labeling of 1 µg of HeLa cell digest (HEPES buffer pH 8.5, 1:20 protein:TMT ratio) provided a labeling efficiency higher than 99% at both N-termini and Lysine residues, with an overall overlabeling ratio lower than 2%. A good labeling efficiency (>96%) was also achieved using the optimized conditions for automated on-column labeling (NaH_2PO_4 pH 4.5, 1:40 protein:TMT ratio). We then optimized an automated high-pH fractionation protocol for low sample amounts and compared the LC-MS/MS results with those obtained using a single 4 h gradient run. The fractionation of 6 µg of HeLa cell digest doubled the number of unique peptides and increased proteome coverage 1.43 fold compared to the single long gradient experiment.

To facilitate the analysis of microdissected tissue samples, in which different regions of tissue may be characterized by different protein content, a modified MicroBCA assay was developed that consumes and detects down to 15 ng of protein, enabling the protein content of each tissue isolate to be measured. The performance of the microproteomics workflow for microdissected tissue samples was assessed by analyzing kidney substructures, using just 0.5 mm^2 and 1.0 mm^2 regions of tissue.

4.2. Introduction

Mass spectrometry is a powerful tool to assess the relative abundance of proteins in biological samples [1][2]. The characterization of changes in protein expression associated with pathological conditions can be used to investigate disease mechanisms and identify proteins that may act as biomarkers and/or therapeutic targets [3][4][5]. Several strategies have been developed for relative quantification measurements and to analyze relative protein expression patterns as a function of biological perturbation [6]. One of the most popular methods for protein relative quantification is stable isotope labeling [2]. Stable isotope tags have been introduced to proteins via metabolic labeling using heavy salts or amino acids (SILAC) [7][8], enzymatically via transfer of ^{18}O from water to peptides [9][10], or via chemical reactions using isotope-coded affinity tags (ICATs) [11]. Commercial ICATs include tandem mass tags (TMTs) [12] and isobaric tags for relative and absolute quantification (iTRAQ) [13].

TMT based relative quantification has distinguished itself as one of the most powerful techniques for relative quantitation because it enables multiplex comparisons of small sample cohorts [14]; recent advances now enable up to 11-plex comparisons [15][16][17].

In a typical TMT labeling workflow the peptide sample obtained after proteolytic digestion of the protein extract undergoes multiple sample handling steps that invariably involve some degree of sample loss. For example the peptide sample is dried twice, after desalting the peptide mixture prior to TMT labeling, and after desalting the TMT-labeled peptide mixture for fractionation-compatible buffer exchange or for concentration prior to LC-MS analysis. The typical workflow is labor- and time-intensive, and sample losses inevitably occur when transferring samples between labware, which adversely affects sensitivity and reproducibility. In this work we optimized the conditions for in-solution and on-column isobaric labeling with tandem mass tags (TMT). We optimized the reaction buffer, pH, protein/TMT ratios, and incubation times for the labeling of low sample amounts (1 µg of HeLa digest), and compared the results with those obtained using the typical method (*i.e.* following manufacturer's instructions) and a larger sample amount (50 µg of HeLa digest).

We also optimized a peptide fractionation protocol for low sample amounts to further increase the number of proteins identified and quantified by the analysis, *i.e.* increased proteome coverage. Offline pre-fractionation is usually performed using several hundreds of micrograms of protein in order to dig deep into the proteome [18][19][20][21][22]. Volume-limited samples typically contain less than one to maximally several micrograms of protein, posing a significant challenge for in-depth proteome analysis. Therefore, optimized fractionation methods for several micrograms of sample are needed. First, we examined whether peptide fractionation could be performed on the carboxylate coated paramagnetic beads using the SP3 protocol. We then compared strong cation exchange (SCX) and high-pH reversed phase (RP) offline fractionation using the AssayMap BRAVO platform in a completely automatized fashion. We then compared the results obtained from best performing method, high-pH RP fractionation, with those obtained using a single 4 h gradient run using the same amount of starting material. This comparison was performed using 6 µg proteins from a HeLa extract, which corresponds to the total protein amount from 6-plex TMT in which each channel corresponds to a just 1 µg of protein extract.

Protein quantification assays represent a crucial step in a bottom-up proteomics experiment; an accurate estimate of the protein content of each sample is important to ensure equal loading of protein into the different TMT channels, and to ensure a constant ratio of protein-to-proteolytic enzyme. Differences in the ratio of protein-to-proteolytic enzyme would affect proteolysis rates and thereby the relative protein quantitation measurements. Finally when comparing between LC-MS/MS experiments it is equally important to inject approximately the same amount of peptides into the LC-MS system. Analyzing the protein content of small tissue sections or microdissected tissue samples remains challenging because of the low amount of protein present and the high degree of cellular heterogeneity, resulting in varying protein concentrations within sample subsections [23][24]. To determine the protein content of small sample volumes and amounts I modified the MicroBCA assay to measure samples down to 1 µL with a concentration down to 15 ng/µL.

Finally, as a proof of principle, the newly developed microproteomic workflow was applied to quantitatively compare the proteomes of two mouse kidney substructures using tissue areas as small as 1 or 0.5 mm^2 per sample, and benchmarked against data from the human proteome atlas.

4.3. Experimental procedures

4.3.1. Materials

Trypsin/LysC mix Mass Spec grade was purchased from Promega (Madison, WI). Tandem Mass Tags (TMT 6-plex) were purchased from Thermo Fisher Scientific (Rockford, IL). All other reagents and solvents were purchased from Sigma Aldrich (St. Louis, MO).

4.3.2. Sample preparation

HeLa whole cell pellets (2.5e9) were ordered from IpraCell (Mons, Belgium), resuspended in PBS and divided into aliquots. Cell pellets were stored at -80 °C until analysis. For the TMT labeling and high-pH fractionation experiments HeLa cell pellets were lysed in an urea-based buffer and digested with the in-solution digestion (ISD) protocol (see paragraph 2.3.2.1. of Chapter 2). For the on-beads SP3 fractionation experiments HeLa cell pellets were lysed and digested according to the SP3 protocol (see paragraph 2.3.2.3. of Chapter 2).

Mouse kidney tissue sections were cut at 15 µm thickness, and after hematoxylin staining an area of 0.5 or 1 mm^2 per sample was excised from the cortex and medulla with a PALM microbeam laser capture microdissection system (Carl Zeiss MicroImaging, Munich, Germany). Protein extraction and proteolytic digestion was performed with the optimized SP3 protocol. Briefly, lysis and denaturation were performed in 20 µL of lysis buffer, sonicated for 10 min with a Bioruptor Pico (Diagenode, Belgium) and incubated for 5 min at 95 °C. After lysis 1 µL of sample was used for protein content estimation using the modified microBCA assay. Then 20 µL of 50% trifluoroethanol (TFE) in lysis buffer together with 2 µL of paramagnetic beads were added to the remaining 19 µL. The protein mixture was reduced, alkylated, washed and subsequently digested overnight at 37 °C using a Trypsin/LysC mixture (Promega) in a 1:25 ratio (protease/sample). After digestion

the resulting peptides were recovered by collecting the supernatant and two additional peptide elution steps using 20 µL of 2% DMSO each with intermittent sonication. Peptides were acidified to 5% formic acid in a final volume of 110 µL prior to automated desalting, on-column TMT labeling and high-pH fractionation.

4.3.3. Modified protein concentration assay

The protein content of each volume-limited sample was determined with a modified microBCA assay (Pierce). Briefly, 2 µL of modified working reagent (MA:MB:MC ratio 25:24:4) were mixed with 1 µL of sample and incubated at 60 °C for 1 h. Absorbance was measured at 570 nm on a spectrophotometer (Infinite 200 PRO with Nanoquant plate, Tecan). Calibration was performed using a BSA standard solutions (Pierce) prepared in SP3 lysis buffer and heated at 95 °C for 5 min.

4.3.4. In-solution TMT labeling of low sample amounts

In-solution TMT labeling experiments were performed on digests (proteolytic peptides) desalted with SepPack cartridges that, after elution, had been dried down with a speedvac and resuspended in 100 mM triethylammonium bicarbonate (TEAB, pH 8.5), 50 mM HEPES (pH 8.5), or 50 mM NaH_2PO_4 (pH 4.5) in LoBind tubes (Eppendorf). TMT 6-plex reagent (Thermo Scientific) solutions were prepared fresh by resuspending 800 mg of label in 100% ACN and kept on ice until use. A 50 µg HeLa digest was labeled with TMT for 1 h according to the manufacturer's protocol in 100 µL of 50 mM TEAB 25% ACN, followed by hydroxylamine quenching. Low sample amount aliquots of 1 µg each were labeled in a final volume of 10 µL, consisting of 8 µL of sample in buffer and 2 µL of TMT label in ACN. The TMT label was added in two equal steps of 1 µL each, each followed by an incubation of 30 min. The labeling reaction was quenched by adding 2 µL of 1.6% hydroxylamine and incubating for 15 min. After labeling the samples were

desalted using the peptide cleanup V2 protocol on the AssayMap BRAVO platform (Agilent). Briefly, C18 cartridges (Agilent, 5 µL bed volume) were primed with 100 µL ACN, equilibrated with 50 µL of buffer A (0.1% formic acid) and loaded with 100 µL of acidified digest at 5 µL/min. Subsequently, a cupwash and internal cartridge wash were performed with 50 µL of buffer A at 10 µL/min, followed by peptide elution with 30 µL of 80% ACN 0.1% formic acid at 5 µL/min. Purified TMT-labeled peptides were dried down with a speedvac, stored at -20 °C and resuspended in 10% formic acid prior to LC-MS/MS analysis.

4.3.5. Automated on-column TMT labeling of low sample amounts

For on-column TMT labeling the cleanup steps before and after labeling were combined into a single automated protocol adapted from the peptide cleanup V2 protocol using the AssayMap BRAVO platform. In brief, after automated cartridge priming, conditioning and sample loading the buffer was exchanged with labeling buffer L (50 mM NaH_2PO_4 pH 4.5). A random channel from a TMT 6-plex set was prepared fresh by solubilizing labels in 100% ACN and after dilution in buffer L they were kept on ice and introduced on-deck immediately before use. Labeling was achieved by loading 10 µL of TMT in buffer L at 1 µL/min, followed by a pause of 20 min and then repeated. The labeled peptides were then washed with Buffer A and eluted in 80% ACN 0.1% formic acid. When quenched, eluates were collected in wells containing 2 µL of 1% hydroxylamine and incubated for 10 min at RT. Desalted and labeled peptides were dried down in a speedvac, stored at -20 °C and resuspended in 10% formic acid prior to LC-MS/MS analysis.

4.3.6. SP3 peptide fractionation

SP3 fractionation was performed following the SP3 peptide cleanup. Peptides were eluted with five 10 µL elution buffers containing 90% ACN, 80% ACN, 60% ACN, 25% ACN

and 2% DMSO at pH 7 (neutral fractionation) and pH 10 (basic fractionation). After the addition of each elution buffer the bead-containing solution was sonicated with a Bioruptor for 5 min (30 s ON, 30 s OFF cycle), placed on a magnet for 2 min and the supernatant collected in a LoBind tube (Eppendorf). Fractions were then centrifuged at 20000 x g for 30 min to pellet any residual beads. The supernatant was stored at -20 °C and diluted 1:1 with 10% formic acid prior to LC-MS/MS analysis.

4.3.7. Automated Strong Cation Exchange (SCX) and high-pH reversed phase peptide fractionation

Strong cation exchange (SCX) and high-pH fractionation were performed using the AssayMap BRAVO platform and SCX and reversed phase S (RPS) cartridges, respectively, and fractionation protocol V1.0.

For SCX fractionation the SCX cartridges were primed and equilibrated with 100 µL of 400 mM ammonium formate, 1% formic acid and 25% ACN in water, and 50 µL of buffer D (1% formic acid 25% can in water), followed by sample loading in 50 µL of buffer D at 5 µL/min. Subsequently, cupwash and cartridge washes were performed with 50 and 10 µL of buffer D, respectively. Peptides were eluted with six 35 µL plugs at 5 µL/min using 40 mM ammonium formate, 25% ACN at pH 3.5, 4.0, 4.5, 5.0, 5.5 and 9.5.

For high-pH fractionation C18 cartridges were primed and equilibrated with 100 µL of ACN and 50 µL of buffer C (10 mM NH_4OH pH 10), followed by sample loading in 50 µL of buffer C at 5 µL/min. Subsequently the cup and cartridge were washed using 50 and 10 µL of buffer C, respectively. Peptides were eluted with six 35 µL plugs at 5 µL/min using 12, 18, 24, 30, 36 and 80% ACN in buffer C. The elution buffer plates were stored at 4 °C and introduced on-deck just after cartridge washes. Fractions including sample flow-through were transferred to a single 96-well plate and dried down with a speedvac and stored at -20 °C.

4.3.8. LC-MS/MS analysis

Peptides were resuspended in 10% formic acid, injected and analyzed using an Easy-nLC1000 (Thermo Scientific) coupled to an Orbitrap Fusion (Thermo Scientific). Peptides were first trapped online using a nanoviper trap column (2 cm x 100 μm, C18, 5 μm, 100 Å, Thermo Scientific) and separated using an Easyspray analytical column (ES803: 50 cm x 75 μm, C18, 2 μm, 100 Å, Thermo Scientific) and a flow rate of 300 nL/min.

TMT labeling experiments were analyzed using a 75 min gradient: t=0-1 min, 5% B; t=51 min, 25% B; t=58 min, 35% B; t=64 min, 90% B; t=75 min, 90% B. Buffer A consisted of 0.1% formic acid and buffer B of 99.9% ACN and 0.1% formic acid. MS analysis was performed using an Orbitrap Fusion configured for a top speed 3 s cycle time with MS1 scans in the Orbitrap (60K resolution, 375-1500 *m/z*, AGC target 4e5) and MS2 scans in the ion trap (1.6 *m/z* isolation window, HCD with 35% NCE, 2e3 AGC target, and 300 ms maximum injection time).

SP3, SCX and high-pH fractions were analyzed using a 145 min gradient, whereas the single-shot analysis was performed with a 4 h gradient. The 145 min gradient was as follows: t=0-1 min, 5%B; t=105 min, 22%B; t=120 min, 32%B; t=130 min, 90%B; t=145 min, 90%B. The 4 h gradient was t=0-1 min, 6%B; t=166 min, 23%; t=211 min, 33%B; t=226 min, 90%B; t=240 min, 90%B. MS analysis was performed with an Orbitrap Fusion configured for a top speed 3 s cycle time with MS1 scans in the Orbitrap (120K resolution, 375-1500 *m/z*, AGC target of 4e5) followed by MS2 scans in the ion trap (1.6 *m/z* isolation window, HCD at 35% NCE, 2e3 AGC target and 300 ms maximum injection time).

For the mouse kidney tissue experiments peptides were analyzed using the same 145 min gradient used for the SP3, SCX and high-pH RP fractions. MS analysis was performed with an Orbitrap Fusion configured for a top speed 3 s cycle time with MS1 scans in the Orbitrap (120K resolution, 375-1500 *m/z*, AGC target of 4e5) but with MS2 scans also in the Orbitrap (1.6 *m/z* isolation window, HCD at 37% NCE, 60K resolution, 5e5 AGC and 80 ms maximum injection time).

4.3.9. Data analysis

Raw data files were analyzed using Proteome Discoverer 2.1 (Thermo Scientific) with the SequestHT search engine. For the HeLa experiments spectra were matched with a forward-decoy Swissprot Homo sapiens database (ver2015-07-22, 42082 sequences) supplemented with a common contaminant database (246 sequences) and allowing up to two missed cleavages, cysteine carbamidomethylation as static modification and methionine oxidation as dynamic modification, 20 ppm precursor mass tolerance and 0.6 Da fragment mass tolerance. For the TMT labeled samples peptide N-terminal and lysine TMT 6-plex modifications were used as static modifications. For the assessment of TMT overlabeling the TMT tag was used as a static modification at the peptide N-terminal and at lysine residues, and as a dynamic modification at serine, threonine, histidine and tyrosine residues.

For the mouse kidney tissue experiments the LC-MS/MS spectra were matched with a forward-decoy Swissprot Mus Musculus (v2015-07-22, 24751 sequences) database supplemented with the common contaminants database and allowing up to two missed cleavages, cysteine carbamidomethylation and TMT tag at peptide N-terminal and lysine residues as static modifications, and methionine oxidation as a dynamic modification, 20 ppm precursor mass tolerance and 0.05 Da fragment mass tolerance.

Results were filtered for 1% FDR using the Percolator algorithm and additionally filtered for a minimum Xcorr score of 1.8. Protein intensities were calculated based on label intensities of the three most abundant peptides with a co-isolation threshold of 50% taking into account unique and razors peptides and normalized to total peptide amount. Search results were exported as txt files and processed with Perseus 1.5 [25]. For each TMT experiment the protein intensities were log2 transformed and subjected to median normalization. Data were filtered such that each protein was quantified in at least three cortex and three medulla samples. Significantly different protein levels between kidney cortex and medulla tissues were calculated using a two-sided Student's *t*-test using a permutation-based FDR cutoff (250 randomizations, FDR 0.001, S0 0.55). Gene ontology enrichment was performed with Panther [26] displaying only terms with a Bonferroni

corrected p-value below 0.05 and an over/under representation of at least 2-fold compared to the set of all identified proteins.

4.4. Results and discussion

4.4.1. In-solution and on-column TMT labeling

We optimized the conditions for the in-solution and on-column TMT labeling of low sample amounts in order to perform accurate relative quantification. Several TMT labeling protocols have been reported for relatively high sample amounts (from 25 μg and upwards) [27][28][29], but the optimal reaction conditions for the labeling of low sample amounts had not been investigated. All the optimization experiments were carried out on 1 μg aliquots of a digested HeLa lysate, in order to avoid any possible variability related to protein digestion and/or cell batch.

4.4.1.1. In-solution TMT labeling optimization

We assessed the efficiency of in-solution TMT labeling on low sample amounts (1 μg of HeLa digest) and compared it with the labeling of 50 μg of HeLa digest. First, we scaled down the labeling protocol recommended by the manufacturer (Thermo Fisher Scientific) in order to adapt it to the lower volume of the 1 μg samples. Then we investigated the effect of different reaction buffers, pH and protein/TMT ratios on the labeling efficiency. We tested three reaction buffers at different pH: 100 mM TEAB pH 8.5, 50 mM HEPES pH 8, 50 mM NaH_2PO_4 pH 4.5. TEAB and phosphate buffers are the recommended buffers for the in-solution and on-column TMT labeling, respectively. We tested also the HEPES buffer in order to assess SP3-TMT labeling. We also investigated the effect of the TMT excess by varying the protein/TMT ratio from 1/10 up to 1/40. Each experimental condition was tested in triplicate. We assessed the TMT labeling efficiency in terms of the percentage of PSMs with the TMT modification at lysine and N-termini residues (Figure 4.1.A), and also assessed the undesired overlabeling rate in terms of the percentage of

PSMs with TMT modification at serine, threonine, histidine and tyrosine residues (Figure 4.1.B).

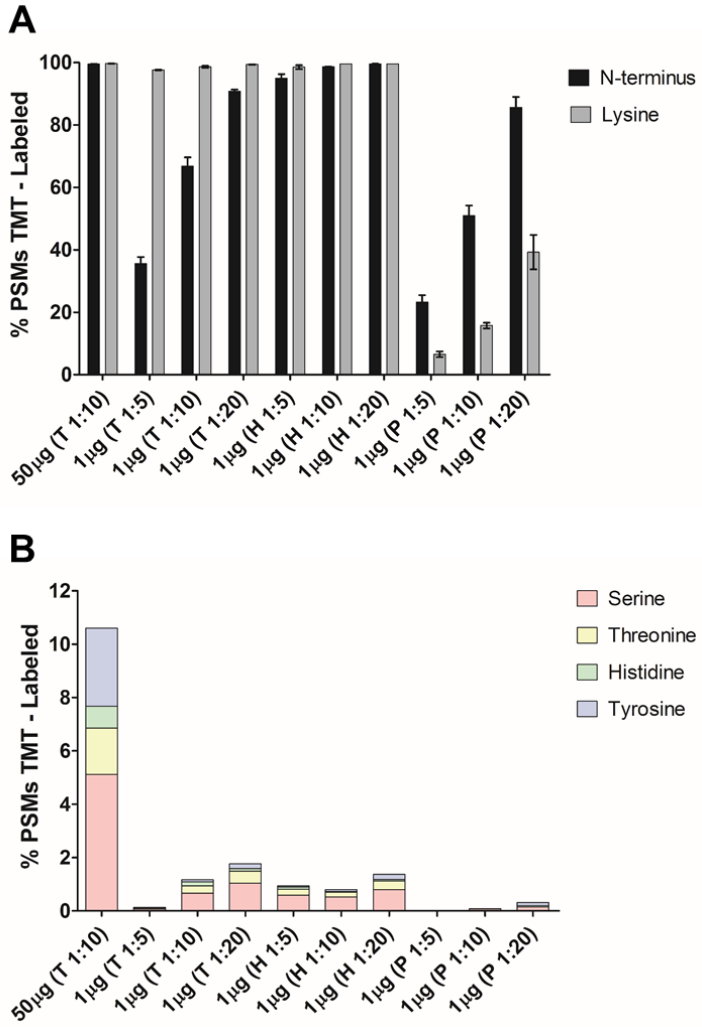


Figure 4.1. *A) TMT labeling efficiency expressed as the percentage of PSMs with the TMT modification at N-termini and lysine residues. B) Over-labeling rate expressed as the percentage of PSMS with the TMT modification at serine, histidine, threonine and tyrosine residues. Error bars represent standard deviation (N=3). Buffers used: 100 mM TEAB pH8.5 (T), 50 mM HEPES pH8.5 (H), and NaH2PO4 pH4.5 (P).*

When performing in-solution TMT labeling of 50 µg of digest using the manufacturer's recommended protocol (TEAB buffer, 1:10 protein/TMT ratio) a good labeling efficiency is achieved (>98%). However, for 1 µg of digest a simple downscaling of the TEAB-based protocol resulted in poor labeling efficiency of N-termini (65%), even at high levels of excess TMT (1:20). Although HEPES and TEAB have the exact same pH (8.5), the labeling efficiency using HEPES was significantly greater. Even a 1:5 TMT excess in HEPES outperformed a 1:10 TMT excess in TEAB, with 95% versus 65% N-terminal labeling, respectively. This result indicates that TMT labeling in HEPES is more efficient but also more economical. The labeling of 1 µg digest in HEPES using a 1:10 TMT excess achieved similar labeling efficiencies (>98%) as obtained using the manufacturer's recommended protocol with 50 µg digest and with a significant reduction in the undesired overlabeling rate. A 1:10 or even 1:20 TMT excess in HEPES resulted in <1.5% overlabeling compared with >10% observed for the 50 µg sample using the manufacturer's recommended protocol. Even though labeling a 1 µg sample in TEAB (1:20 TMT excess) did not result in full labeling of primary amines, the rate of overlabeling was greater than that observed with complete labeling in HEPES. The phosphate-based buffer (indicated with *P* in the above figure) was found to be far from optimal, with poor lysine labeling efficiencies of 5 to 40% even for 1:10 to 1:20 TMT excess.

The absolute number of PSMs was the highest for the labeling of 50 µg (Figure 4.2.) because of the much larger amount of protein available for the LC-MS/MS analysis. For the 1 µg samples the highest number of PSMs was obtained when using HEPES as reaction buffer. In summary for in-solution labeling of 1 µg samples 50 mM HEPES (pH 8.5) and a 1:10 or 1:20 TMT excess ratio gave the best results in terms of labeling efficiency and overlabeling rate.

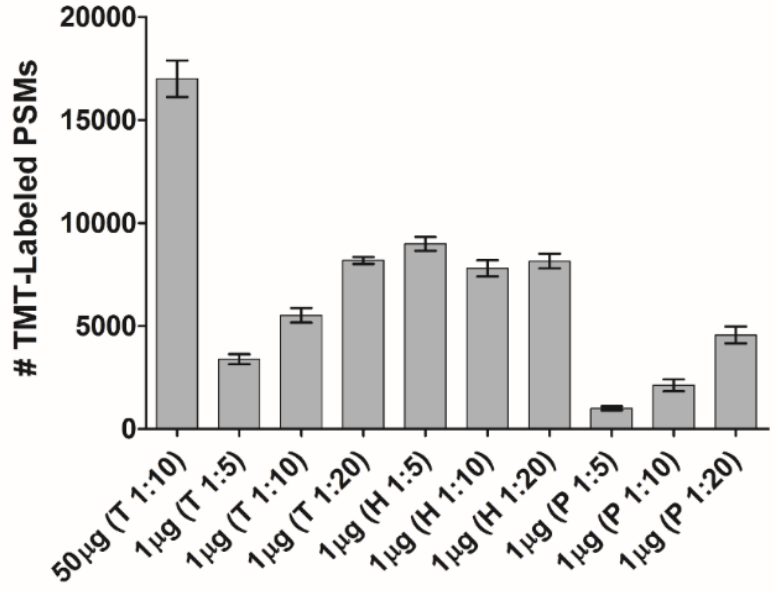


Figure 4.2. *Number of TMT labeled PSMs for all the tested conditions. Error bars represent standard deviation (N=3).*

Since HEPES is the digestion buffer used for the SP3 protocol, the in-solution TMT labeling can be easily performed after the SP3 digestion without the need to dry the samples or to perform a buffer exchange, representing a great advantage in terms of analysis time and avoiding peptide losses.

4.4.1.2. On-column TMT labeling

On-column TMT labeling was performed using the AssayMap Bravo platform. In all labeling experiments cartridges with RPS chemistry were used because pilot studies showed that although C18 cartridges show slightly less peptide breakthrough during the labeling process their labeling efficiency was significantly lower. On-column labeling was performed using 1 µg of digest in HEPES, the optimal in-solution buffer, or phosphate buffer, previously reported to be suitable for on-column TMT labeling [29]. Figure 4.3.A and 4.3.B show the TMT labeling efficiency and overlabeling rate, respectively, for all tested conditions.

The optimal conditions for in-solution labeling, 1:20 TMT excess in HEPES buffer, exhibited poor on-column labeling efficiency and also led to a greater degree of overlabeling when compared with the in-solution protocol. In contrast with the very poor in-solution labeling obtained using the phosphate buffer, it provided the highest efficiency for on-column labeling. When comparing the complete on-column to complete in-solution labeling experiments the overlabeling rate was higher for on-column labeling. Even for HEPES, which displayed very low overlabeling in solution, a 9-fold higher overlabeling rate was observed on-column. It was found that the on-column overlabeling could be alleviated by eluting the peptide/TMT mixture into hydroxylamine to quench the reaction. This quench enabled the degree of overlabeling to remain low, <0.5 %, while maintaining a high label efficiency (>95%) and provided the largest number of identified fully labeled peptides.

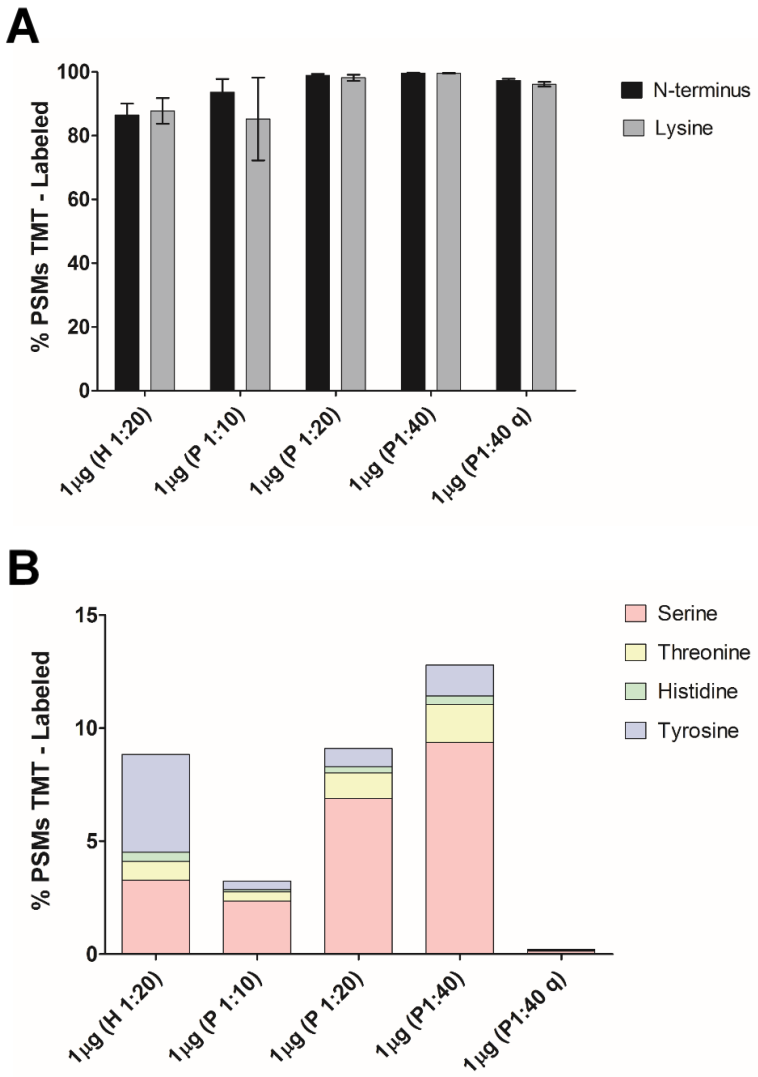


Figure 4.3. *A) TMT labeling efficiency expressed as the percentage of PSMs with the TMT modification at N-termini and lysine residues. B) Over-labeling rate expressed as the percentage of PSMS with the TMT modification at serine, histidine, threonine and tyrosine residues. Error bars represent standard deviation (N=3). Buffers used: 100 mM TEAB pH8.5 (T), 50 mM HEPES pH8.5 (H), and NaH2PO4 pH4.5 (P).*

Interestingly, on-column labeling increased the number of identified spectra (Figure 4.4.). A 146% increase in TMT labeled PSMs was observed when comparing the maximum number of PSMs for in-solution TMT labeling (1:5 in HEPES) with on-column TMT labeling (1:20 in phosphate buffer).

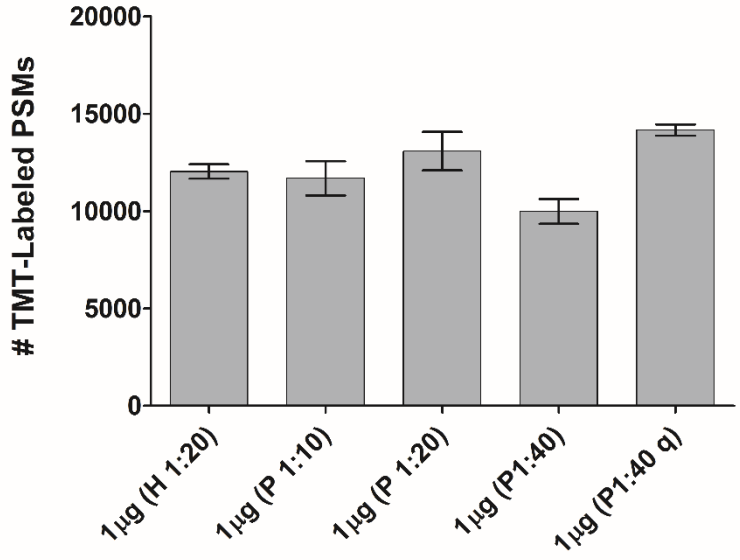


Figure 4.4. *Number of TMT labeled PSMs for all the tested conditions. Error bars represent standard deviation (N=3). Buffers used: 100 mM TEAB pH8.5 (T), 50 mM HEPES pH8.5 (H), and NaH2PO4 pH4.5 (P).*

The best on-column labeling protocol (phosphate buffer, 1:40 TMT excess, quenched with hydroxylamine) provided the best performance for 1 µg of sample for all conditions investigated (on-column or in-solution), showing similar results to labeling 50 µg of protein using the manufacturer's recommended protocol in terms of labeling efficiency (>96%).

4.4.2. Peptide fractionation

We optimized the conditions to perform peptide fractionation on the beads after the SP3 digestion protocol, and on-column using the AssayMap BRAVO platform. Peptide fractionation is performed to reduce the very high complexity of the peptide sample obtained after trypsin digestion, and thereby to identify and quantify more proteins. Fractionation divides the starting material over multiple LC-MS/MS runs and is normally applied to larger sample amounts [30][31] to ensure that there is sufficient protein available that the lower abundant proteins are present above the lower limit of quantitation/identification. When the amount of starting material is low, the lower abundant proteins may not be present in amounts above the lower limit of quantitation/identification and peptide fractionation has limited benefit. However, when a TMT experiment is performed the samples from each TMT channel are pooled prior to LC-MS/MS analysis. In this scenario even if the amount of starting material is 1 μg, the pooled sample is 6 or 10 μg (for 6-plex or 10-plex TMT, respectively) and fractionation can be performed to increase identify and quantify lower abundant proteins. Thus, we sought to optimize the conditions for peptide fractionation on 6 μg aliquots of HeLa digest.

4.4.2.1. SP3 fractionation

Fractionation can be carried out within the SP3 protocol by modulating the interactions between the peptides and the paramagnetic beads [32]. After the protein digestion step the peptides can be selectively eluted from the beads by changing the percentage of organic solvent (ACN) in the buffer and/or the pH. The addition of an organic solvent to an aqueous solution containing paramagnetic beads promotes trapping of peptides in a solvation layer on the hydrophilic surface of the beads. Thus, the elution of the peptides from the beads is achieved by reducing the percentage of organic solvent. The interaction between peptides and the carboxylic groups on the surface of the beads is also pH dependent. Acidic buffers promote a hydrophilic-interaction-based retention of peptides, whereas basic buffers induce elution of the peptides because of the repulsion between the negatively charged carboxylate groups.

The original report of the SP3 protocol [32] included a 2-step fractionation using two different elution buffers: 87% ACN in ammonium formate at pH 10 and 2% DMSO at pH 7. We tried to further optimize the SP3 fractionation protocol in order to increase the number of fractions, and thereby increase proteome coverage.

First we performed a pilot study using 1e5 HeLa cells (approximately 15 μg of proteins) to test the efficacy of SP3 fractionation. We eluted the peptides with five buffers at different organic solvent concentration: 90% ACN, 80% ACN, 60% ACN, 25% ACN and 2% DMSO. The effect of the pH was also investigated by performing the fractionation with buffers at pH 7 and 10. SP3 fractionation was assessed by determining the percentage of peptides shared between fractions: a lower percentage of overlap indicates a higher selectivity of the elution mechanism. Figure 4.5. shows the number of unique peptides identified in the six neutral and six basic SP3-fractions and the percentage of non overlapping peptides between adjacent fractions.

The neutral SP3-fractionation provided the largest number of unique peptides for all of the fractions. However, we found a high degree of overlap between adjacent fractions for both the neutral and basic fractionations. These results suggest that below a certain ACN percentage threshold the peptides were eluted from the beads with low selectivity. We performed another fractionation experiment reducing the number of fractions (90% ACN, 70% ACN, 40% ACN and 2% DMSO) but again the overlap between adjacent fractions was larger than 50% (data not shown). These results indicate that the interaction between the peptides and the carboxylate modified beads cannot be sufficiently modulated for effective on-bead fractionation.

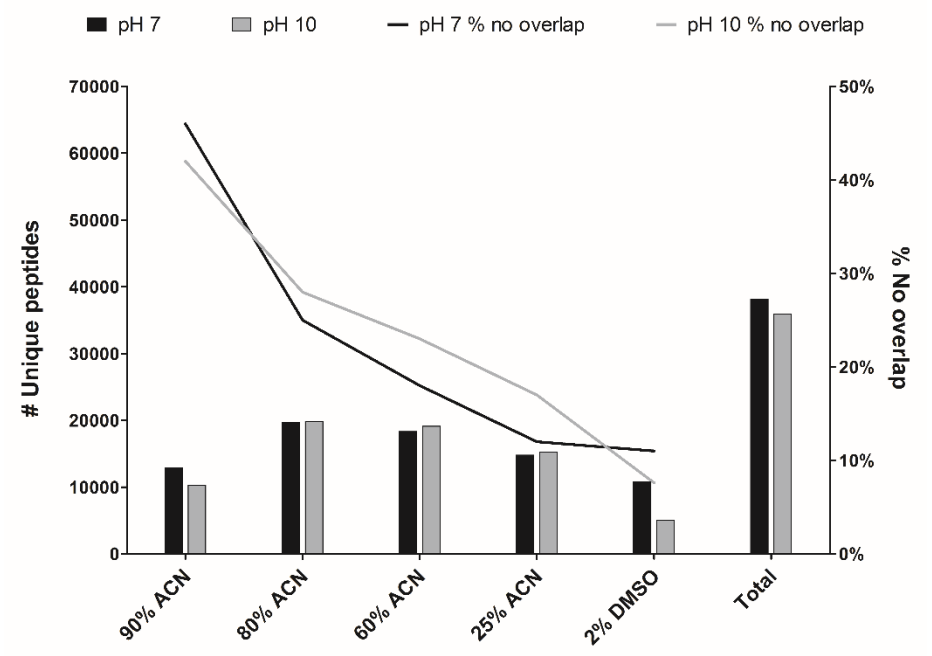


Figure 4.5. *Neutral vs basic SP3-fractionation of 15 µg of HeLa digest. Bar graph display the numbers of unique peptides identified in each fraction (left y-axis). The lines refer to the percentage of peptides that were only found in the specified fraction number and not in adjacent fractions (right y-axis).*

4.4.2.2. Peptide fractionation with AssayMap BRAVO platform

We optimized an automated peptide fractionation protocol using the AssayMap BRAVO platform. It is important that the fractionation mechanism is orthogonal to the reverse phase separation of the LC-MS/MS analysis in order to maximize peak capacity, resolution and proteome coverage. Strong cation exchange (SCX) and high-pH reversed phase (RP) fractionations have been extensively used for peptide fractionation of complex samples, as they provide a good orthogonality with the RP chromatographic separation [33][34][35]. SCX and RP separations are fairly orthogonal as they based on different separation mechanisms. In SCX negative functional groups interact with positively charged peptides at low pH, so peptides are separated according to differences in their electric charge by

using elution buffers with different pH. The stationary phase used for high-pH fractionation and the chromatographic separation is the same, but a high degree of orthogonality is achieved by changing the pH of the mobile phase. Since peptides are charged molecules comprised of acidic and basic functional groups, the change in pH influences their protonation state and thus their retention behavior. Neutralization of a charged residue decreases its hydrophobicity and consequently leads to a better retention. Generally acidic peptides are better retained at lower pH in which the carboxylic groups are protonated (increased hydrophobicity), while basic peptides are better retained in high pH conditions because of the deprotonation of the basic residues. Thus, a high-pH separation (fractionation) and a low-pH separation (chromatography) provides a high degree of orthogonality, similar to that obtained with a SCX-RP strategy [36].

We first compared the SCX and high-pH fractionation with the AssayMap BRAVO platform using 100 μg aliquots of HeLa digest. Figure 4.6. shows the number of unique peptides identified in the six SCX and six high-pH fractions and the percentage of non overlapping peptides between adjacent fractions. Even though SCX fractionation led to the identification of more unique peptides in most fractions, the total number of unique peptide sequences identified was greater for the high-pH fractionation because of a lower degree of overlap between adjacent fractions, indicating higher selectivity of the fractionation.

Since the high-pH fractionation provided the highest selectivity and number of identified peptides, we tested it on the fractionation of low sample amounts. We fractionated 6 μg of a HeLa digest to mimic the analysis of 1 μg of samples in a TMT 6-plex experiment and compared the results with a non-fractionated sample. A single 4 h run was compared with seven high-pH fractions, each of which was analyzed with a 2 h gradient (Figure 4.7.).

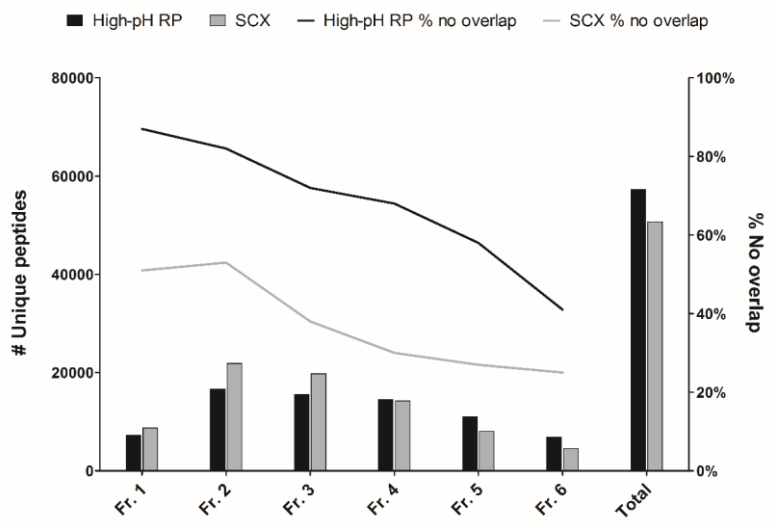


Figure 4.6. *SCX vs High-pH fractionation of 100 µg of HeLa digest. Bar graph display the numbers of unique peptides identified in each fraction (left y-axis). The lines refer to the percentage of peptides that were only found in the specified fraction number and not in adjacent fractions (right y-axis).*

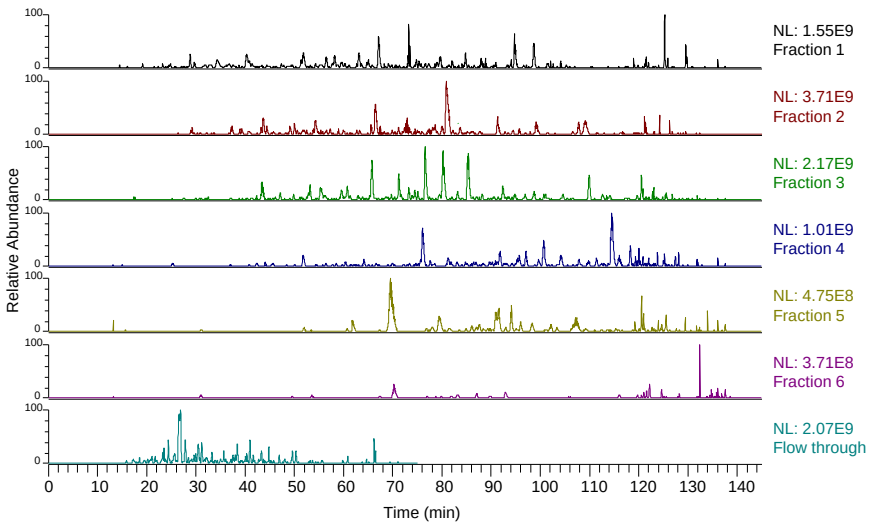


Figure 4.7. *Base peak chromatograms of the six fraction and flow through obtained with the automated high-pH fractionation on 6µg of HeLa digest.*

To prevent overloading of the LC column in the single-run analysis, only 5 µg were injected. Figure 4.8. shows the number of MS/MS acquired, unique peptides and protein groups identified with the single-run analysis and with high-pH fractionation.

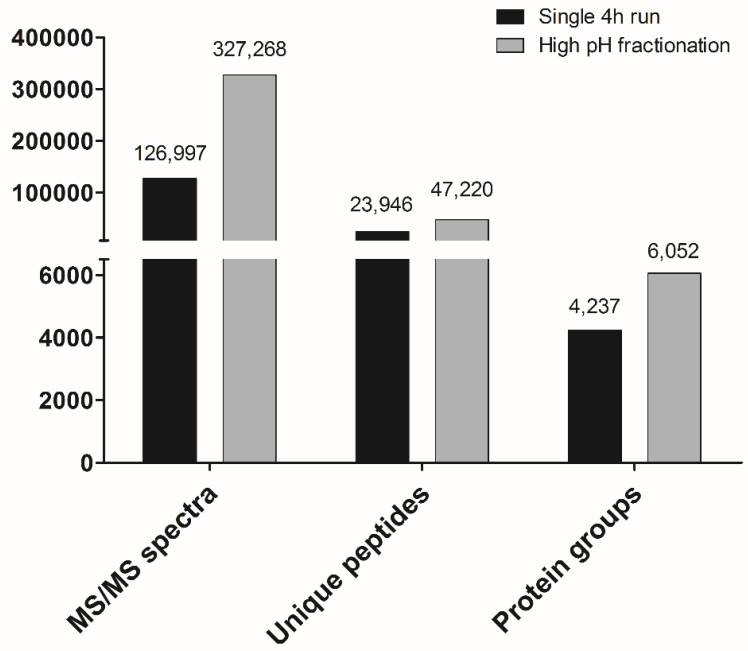


Figure 4.8. *Low sample amount proteome coverage for 6 µg of sample: single 4 h LC-MS/MS runs versus LC-MS/MS analysis of offline high-pH fractionated sample (seven fractions x 2 h gradients).*

The high-pH fractionation provided a higher proteome coverage compared to the single-run analysis even for small sample amounts. The fractionation increased the number of acquired spectra 2.6 fold and doubled the number of unique peptide identifications (23946 vs 47220). As a result proteome coverage was increased by 143%, from 4237 to 6052 protein groups (grouped from 10277 and 14645 identified protein accessions, respectively).

The number of unique peptides and protein groups identified with high-pH fractionation exceed that reported using custom designed Stagetips [37][38] or spintip-based [39][40] fractionation approaches to fractionate 200 µg of protein. However, in this study the amount of starting protein was 6 µg and we used commercial reversed-phase cartridges handled by a robot using controlled flowrate and volumes to ensure high-throughput and reproducible sample handling. Moreover, this fractionation strategy is easy to implement in a modular, offline and robotized proteomic workflow.

4.4.3. Protein determination of minute sample amounts

One of the challenges of working with microdissected tissues or other samples characterized by low sample amounts is the need to know the total protein content. Protein content quantification is a crucial step in shotgun proteomics not only for the estimation of the proper amount of proteolytic enzyme required for the digestion, but also for accurate quantitation of protein expression levels between different samples. Previously adjacent sections were used for the protein content estimation, which is non-ideal because tissue histologies may differ, especially for small pathological features with a specific histolog.

Many direct (tyrosine and tryptophane absorption method [41]), colorimetric (Lowry [42], Bradford [43] and BCA [44]) and fluorescence-based assays (Qubit [45], FluoroProfile [46] and Nano-Orange [47]) have been developed to quickly and reliably quantify protein amounts. However, many are incompatible with the lysis buffers commonly used for proteomics sample preparation. The microBCA assay is tolerant to detergents and a range of buffers. Although the microBCA protein determination assay is highly sensitive for medium sample volumes (2-40 ng/µL for 150 µL sample), it does not suffice if only a few microliter of sample can be spared for the protein determination assay. Therefore, we modified the microBCA assay to use less sample while increasing the range of detection.

Technological advances in spectrophotometers have enabled the measurements of only 1-2 µL of sample [48]. However, due to the great reduction of the light path length, detection limits are reduced. In order to compensate for the decreased absorbance due to the

shortened light path length and the presence of chelating agents in our lysis buffer (EDTA and EGTA) we increased the copper (MC buffer) concentration 4 times and increased the working reagent (WR) to sample ratio 2-fold. In this way by mixing 1 µL of sample with 2 µL of modified working reagent (MA:MB:MC ratio 25:24:4) we were able to use 2 µL for the absorbance measurement with the Nanoquant plate (Tecan). We obtained a reproducible BSA standard calibration curve between 5 and 120 ng/µL using only 1 µL of standard (Figure 4.9.).

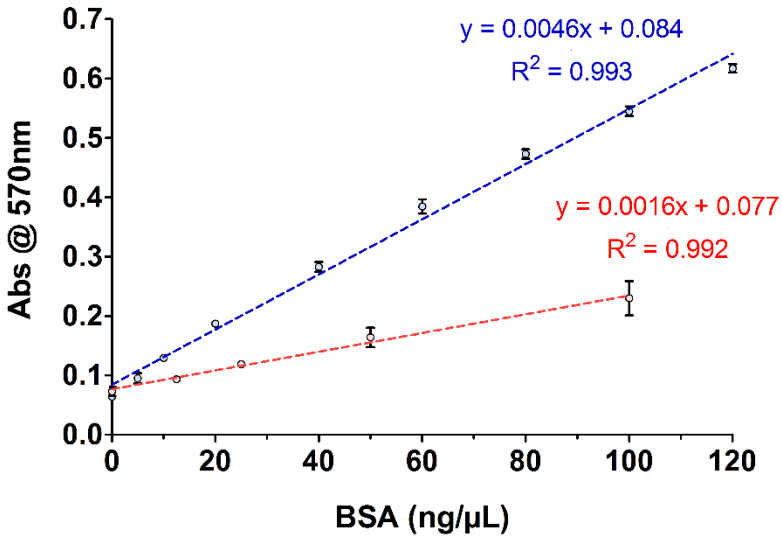


Figure 4.9. *Modified microBCA assay optimized to improve sensitivity and consume less sample (only 1 µL). Calibration curves obtained using BSA standard solutions prepared in SP3 lysis buffer (5-120 ng/µL linear dynamic range, corresponding to sample concentration of 15-360 ng/µL) using the original unmodified conditions (red) and the final optimized conditions (blue). Error bars represent standard deviation (N=3).*

The unmodified conditions (MA:MB:MC ratio 25:24:1, sample/WR ratio 1:1) showed low absorbance values and a low sensitivity, while the modified conditions provided almost a 3-fold increase in sensitivity. With the optimized method we could reproducibly measure down to 15 ng of protein from only 1 µL of sample with an average CV of 3% (average from nine BSA standards). For our tissue lysis in 20 µL of lysis buffer, this translates into measuring tissue protein contents ranging from 300 ng to 7.2 µg while consuming only 5% of sample.

We compared our modified MicroBCA assay with the most common protein quantitation assays in terms of analytical figures of merit, compatibility to SDS and costs (Table 4.1.). Altogether, considering the dynamic working range, sample volume required, lowest protein concentration detectable and percentage of SDS compatible with the assay, our modified MicroBCA assay represents the best method for protein quantitation, including in terms of costs.

Table 4.1. *Protein quantitation assays comparison*

Assay	Dynamic working range (ng/µL)*	Sample volume (µL)	Minimum protein content (µg)	SDS[#]	Cost / assay[$]	Ref.
MicroBCA	4-80 or **1-40**	150 or **1000**	0.6 or **1**	10%	6 or **43**	[49]
NanoOrange	0.25-250	100	25	0.25%	137	[50]
Lowry	6.5-9750	40 or **200**	0.26 or **1.3**	6.6%	6 or **32**	[51]
Bradford	5100-76500 or **2-50**	5 or **1000**	25.5 or **2**	6.37%	3 or **11**	[52]
Qubit	25-500 or **12.5-250**	1 or **20**	0.025 or **0.25**	2-0.1%	15 or **14**	[53]
FluoroProfile	6-200	5	0.03	0.2%	2	[54]
Modified microBCA	**15-360**	**1**	**0.015**	**10%**	**0.3**	

*Initial sample concentration (based on BSA calibration)

[#]Highest compatible concentration in the initial sample

[$]Reagent costs per sample in USD cents.

4.4.5. Mouse kidney tissue substructure proteome analysis

As a proof of principle we applied the workflow to investigate the proteome of the anatomical regions of the mouse kidney. Laser capture microdissection (LCM) was used to excise and isolate cortex and medulla regions of fresh frozen mouse kidney (Figure 4.10.), followed by the protein quantitation assay and microproteomics workflow.

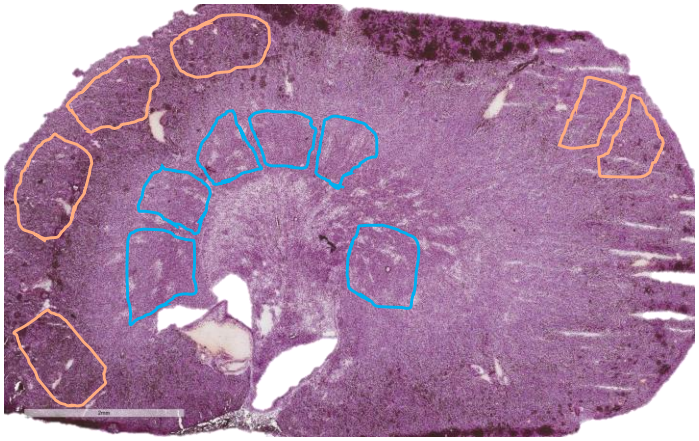


Figure 4.10. *Representative mouse kidney section stained with hematoxylin. Orange lines indicate ROIs isolated from the cortex, blue lines indicate ROIs isolated from the medulla.*

Five ROIs of 0.5 mm^2 and five ROIs of 1 mm^2 were microdissected from both the cortex and medulla. The resulting 20 samples were lysed using the SP3 protocol resulting in an average of 1.5 μg of protein per sample. After protein cleanup and digestion the proteolytic peptide samples were transferred to the AssayMap BRAVO platform for automated peptide desalting, 10-plex TMT labeling and high-pH fractionation. The peptides from each of the ten 0.5 mm^2 and ten 1 mm^2 samples were labeled with a different 10-plex TMT label. This resulted in two data sets, one from 0.5 mm^2 samples and the other from 1 mm^2 samples, each containing five replicates from the cortex and medulla. As seen in Figure

4.11. deep proteome coverage was achieved for a sample area of only 1 mm^2 (5002 protein groups), even when using TMT labeling (which reduces sequencing speed). A lower proteome depth (3440 protein groups) was obtained from the 0.5 mm^2 samples, reflecting the lower amount of protein available for the analysis. Nevertheless, when comparing our approach using only 1.5 μg per sample and analyzing only seven high-pH fractions in seven LC-MS runs, an equal or greater proteome coverage was achieved than that recently reported as part of a draft of the human proteome, and which used 450 μg of human kidney homogenate and which was separated into 24 high-pH fractions and analyzed by 24 LC-MS runs [18]. This result reflects the high sensitivity of the workflow, which provided a similar proteome coverage from >200-fold less starting material (1.5 vs 450 μg).

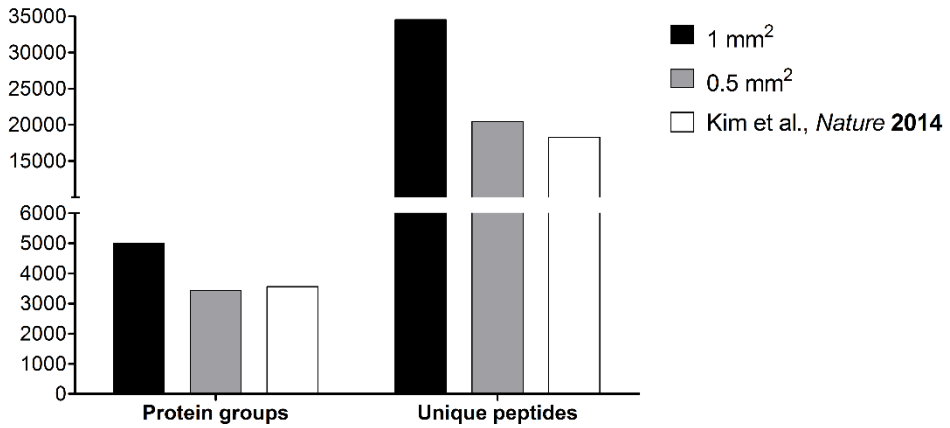


Figure 4.11. *Protein identification metrics: number of protein groups identified from the 1 mm^2 (black bars) and 0.5 mm^2 (grey bars) datasets compared with those obtained by Kim et al. from the proteomics analysis of a human kidney homogenate* [18].

A a number of differentially expressed proteins could be identified in the cortex and medulla, Figure 4.12.

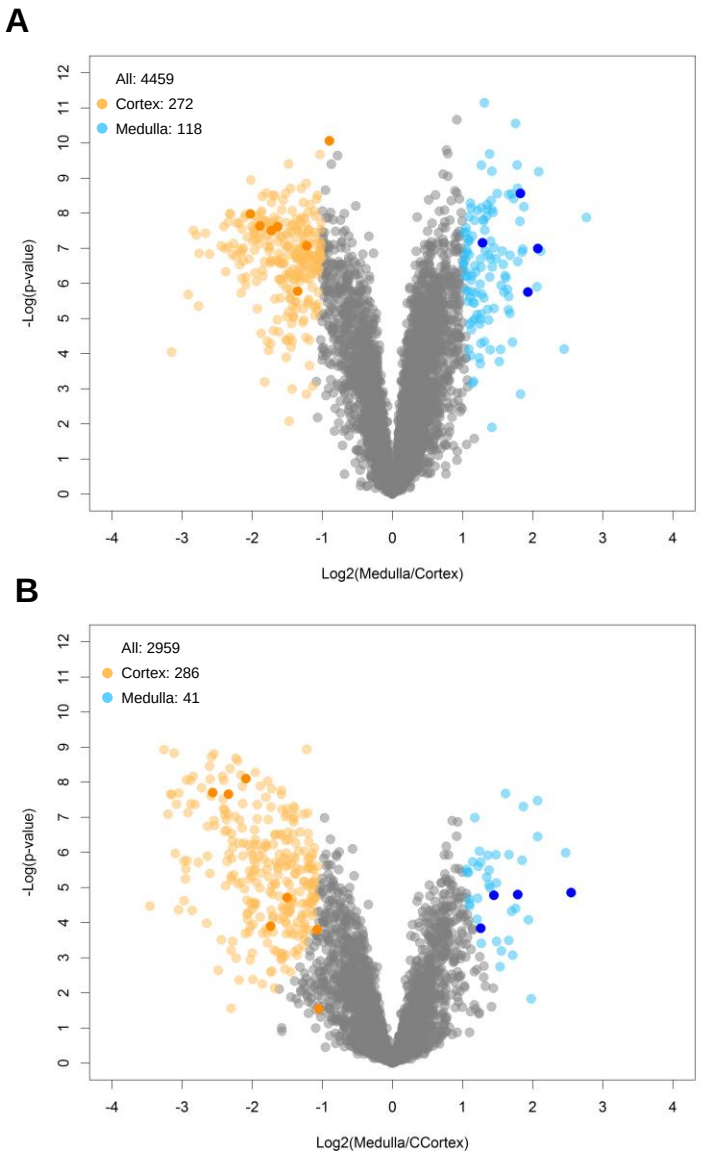


Figure 4.12. *Protein expression comparison between kidney cortex and medulla. Volcano plots for the 1 mm^2 (A) and 0.5 mm^2 (B) data sets. Grey circles indicate non significant protein groups (log2(fold change) <1 and >-1, FDR >0.001). Orange circles indicate protein groups significantly up-regulated in the medulla (log2(fold change) <-1, FDR <0.001). Blue circles indicate protein groups significantly up-regulated in the cortex (log2(fold change) >1, FDR <0.001). Dark orange circles indicate cortex markers (LRP2, GGT1, HPD, HRSP12, PKLR, SLC22A8, DPYS, SLC22A13). Dark blue circles indicate medulla markers (AqP2, CRYAB, UT1, UT2).*

Several marker proteins known to be localized to the different kidney layers (Human Protein Atlas) were detected in the layer-specific proteomes. Interestingly, the 0.5 mm^2 proteome shows more variation (lower log p-values) and larger ratios. This may reflect the greater sensitivity of smaller tissue samples to cellular heterogeneity (less averaging across more cells). Despite a slightly higher variation the proteomes obtained from 0.5 and 1 mm^2 are very similar in terms of relative expression levels and the set of significantly regulated proteins, as can be seen from the consistently regulated kidney markers and the high correlation between the two proteomes (Figure 4.13.).

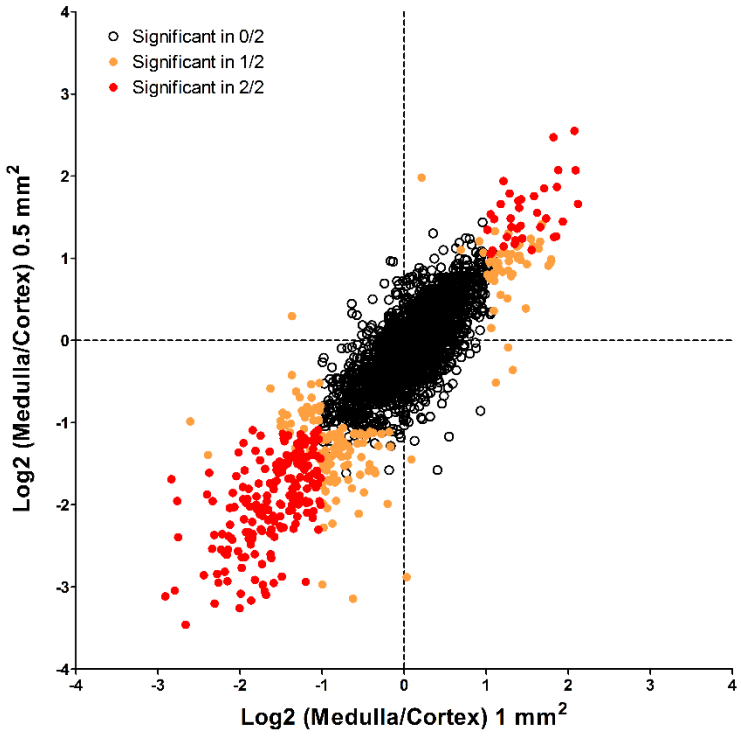


Figure 4.13. *Protein expression correlation between the 0.5 and 1 mm^2 datasets. Significantly different protein groups between cortex and medulla in one or both datasets are indicated with orange and red, respectively.*

The in-depth proteomics data of the cortex and medulla were then subject to a gene ontology analysis to assess the functional roles of the differentially expressed proteins. Gene ontology over- and under-representation analysis supported the well established function of each layer (Figure 4.14). For example, the main function of the cortex is to resorb ions, glucose, amino acids and organic acids, whereas in the medulla salts and water are secreted and resorbed. The functions and biological processes enriched in the cortex proteome recapitulate the transporting activity. Interestingly, in addition to transport proteins many metabolic enzymes are also enriched in the cortex, which could indicate other important processes in the cortex, including the modification, transformation or clearance of the resorbed amino acids, carbohydrates and other organic compounds. In the medulla water transport/homeostasis and salt transport are strongly enriched due to the relatively high levels of aquaporins and specific ion transporters.

The data also allowed the large-scale monitoring of kidney substructure-specific protein paralogue expression (Figure 4.15.). Whereas most of the detoxifying glutathione S-transferase proteins are expressed at similar levels in each substructure some are specifically expressed in the cortex (GSTA2, GSTM5, MGST1) or medulla (GSTM2, GSTM6). A Similar analysis was also performed for membrane transporter proteins from the solute carrier and ATPase families. Specific bicarbonate transporters (SLC family 4) are expressed in the medulla (SLC4A2 and SLC4A7) or cortex (SLC4A4), whereas only organic ion transporters (SLC family 22) were found to be significantly elevated in the cortex. Regarding ATPases, alpha, beta and gamma subunits (ATP1A1, ATP1A4, ATP1B1 and FXYD2) of the Na^+/K^+ transporting ATPase assembly were significantly up-regulated in the medulla, whereas many different H^+ transporting subunits showed elevated but insignificant up-regulation in the cortex.

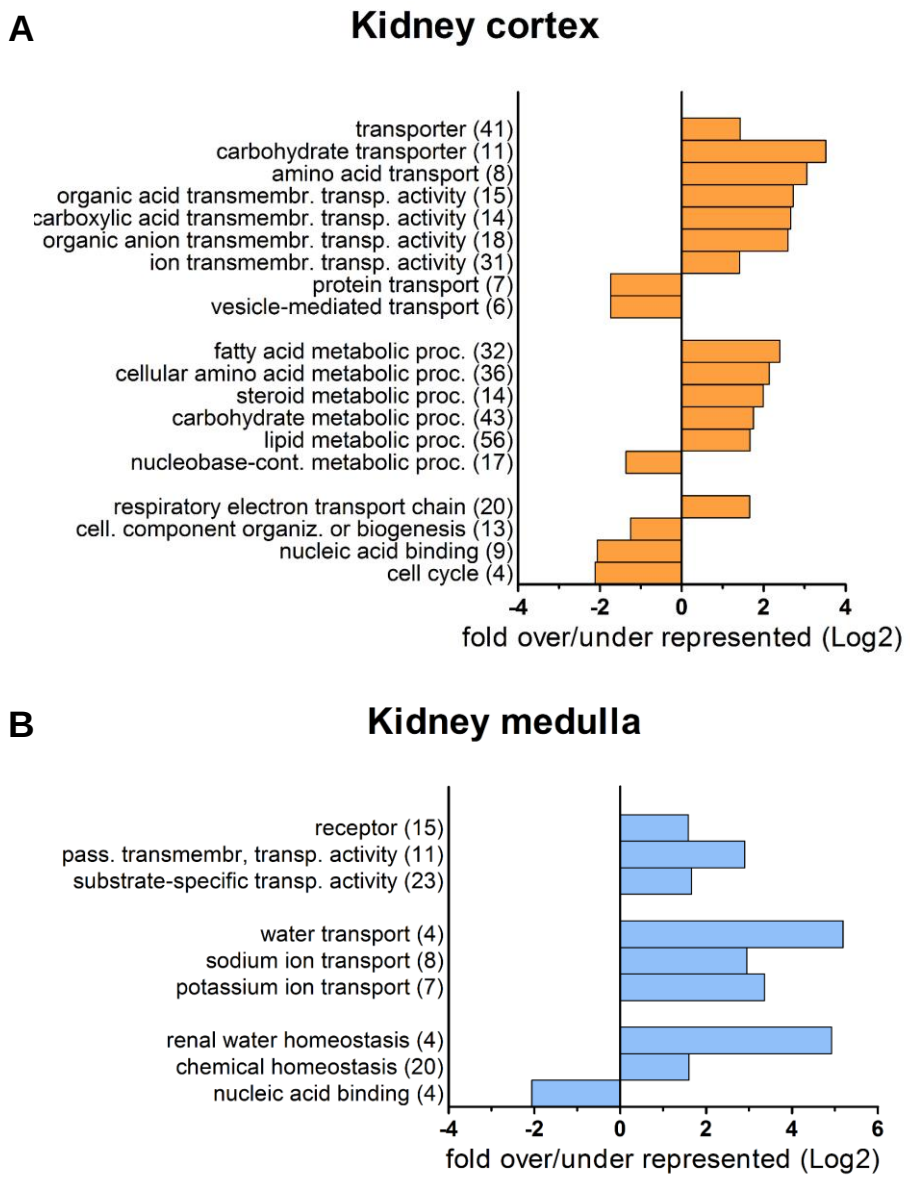


Figure 4.14. *Gene ontology over/under representation of proteins significantly up-regulated in the kidney cortex or medulla compared to all identified proteins using Panther analysis (1 mm^2 dataset, Bonferroni correction $p<0.05$). Number of proteins are indicated between brackets.*

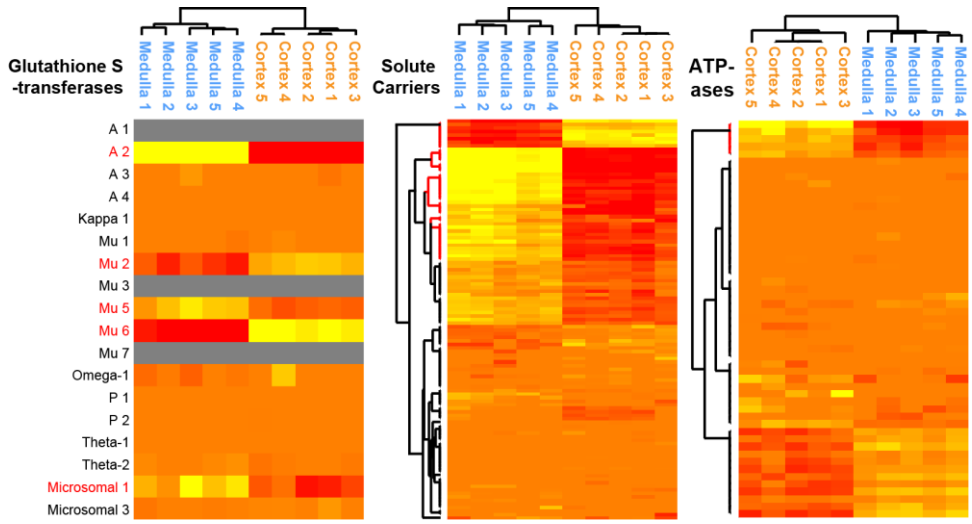


Figure 4.15. *Kidney substructure specific expression pattern of protein class members. Three classes of proteins were clustered to visualize high (>2-fold above median; red) or low (<2-fold below median; yellow) protein levels in all samples. Proteins/clusters marked in red showed significantly different protein levels between two tissue types, grey indicate not quantified protein levels due to lack of unique peptide sequences. Cluster analysis was performed on solute carriers family, ATPases family, glutathione S transferase family.*

4.5. Conclusions

We developed a quantitative microproteomic workflow for the analysis of small sample amounts and microdissected tissue samples. It was demonstrated that scaling down TMT labeling to 1 µg of sample using current protocols for in-solution or on-column labeling resulted is a poor labeling efficiency (< 50%). Therefore, new in-solution and on-column TMT labeling protocols were developed for low sample amounts. The optimal conditions for in-solution TMT labeling were obtained using HEPES buffer, which is the same buffer used for the SP3 digestion protocol, indicating that TMT labeling can be integrated within the SP3 digestion protocol. Automated on-column TMT labeling with a phosphate buffer was found to provide good labeling efficiency, representing an amenable alternative to in-solution labeling when high sample throughput is needed. Because of the 96-well parallel format of the AssayMap BRAVO platform, the same amount of time will be spent to label 1 sample as for 96 samples.

In addition to a reproducible quantitative workflow for low sample amounts, we sought to increase proteome coverage from limited amounts of sample. On-bead fractionation within the SP3 protocol exhibited poor selectivity as the percentage of overlapping protein groups between adjacent fractions was found to be greater than 50%. Conversely, automated on-column high-pH fractionation provided not only a good selectivity but also an increase in proteome coverage of about 143% when compared to a single long gradient analysis.

As a proof of principle the quantitative microproteomic workflow was tested on mouse kidney medulla and cortex substructures, using areas of 0.5 or 1 mm^2 of 15 µm thick tissue sections. We first developed a protein assay to quantify the total protein content from the microdissected tissue samples because established methods lacked the required sensitivity, and were incompatible with the reagents used for sample lysis. I developed a modified microBCA assay that enabled the quantification down to 15 ng of proteins using only 1 µL of sample.

The depth of proteome coverage for the 0.5 mm^2 kidney samples was comparable to the data from the recently described human proteome atlas; however, in our study >200-fold

less starting material (450 vs 1.5 μg per sample) and 20-fold less LC-MS run time was used (36 h per sample vs 17 h for 10 samples). The differences between the two studies is mainly due to the use of a more sensitive protocol and multiplexing. The selection and microdissection of cortex and medulla kidney substructures was used to create a detailed list of region-specific protein expression. Many well-documented functions of the different regions were molecularly confirmed and in addition further specified in detail because of the protein isoform-specific expression patterns.

The optimized set of microproteomics protocols has the potential to be used to study the proteomes of other heterogeneous low sample amount sources such as ex vivo tumors or specific histological tissue regions.

References

1. Rauniyar, N., and Yates, J.R. (2014) Isobaric Labeling-Based Relative Quantification in Shotgun Proteomics. *J. Proteome Res.*, **13** (12), 5293–5309.

2. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature*, **422** (6928), 198–207.

3. Chahrour, O., Cobice, D., and Malone, J. (2015) Stable isotope labelling methods in mass spectrometry-based quantitative proteomics. *J. Pharm. Biomed. Anal.*, **113**, 2–20.

4. Diamandis, E.P. (2004) Mass Spectrometry as a Diagnostic and a Cancer Biomarker Discovery Tool. *Mol. Cell. Proteomics*, **3** (4), 367–378.

5. Srinivas, P.R., Srivastava, S., Hanash, S., and Wright, G.L. (2001) Proteomics in Early Detection of Cancer. *Clin. Chem.*, **47** (10), 1901–1911.

6. Heck, A.J.R., and Krijgsveld, J. (2004) Mass spectrometry-based quantitative proteomics. *Expert Rev. Proteomics*, **1** (3), 317–326.

7. Conrads, T.P., Issaq, H.J., and Veenstra, T.D. (2002) New Tools for Quantitative Phosphoproteome Analysis. *Biochem. Biophys. Res. Commun.*, **290** (3), 885–890.

8. Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002) Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics*, **1** (5), 376–386.

9. Mirgorodskaya, O.A., Kozmin, Y.P., Titov, M.I., Körner, R., Sönksen, C.P., and Roepstorff, P. (2000) Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using 18O-labeled internal standards. *Rapid Commun. Mass Spectrom.*, **14** (14), 1226–1232.

10. Yao, X., Freas, A., Ramirez, J., Demirev, P.A., and Fenselau, C. (2001) Proteolytic 18O Labeling for Comparative Proteomics: Model Studies with Two Serotypes of

Adenovirus. *Anal. Chem.*, **73** (13), 2836–2842.

11. Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.*, **17** (10), 994–999.

12. Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., and Hamon, C. (2003) Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.*, **75** (8), 1895–1904.

13. Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D.J. (2004) Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. *Mol. Cell. Proteomics*, **3** (12), 1154–1169.

14. Altelaar, A.F.M., Frese, C.K., Preisinger, C., Hennrich, M.L., Schram, A.W., Timmers, H.T.M., Heck, A.J.R., and Mohammed, S. (2013) Benchmarking stable isotope labeling based quantitative proteomics. *J. Proteomics*, **88**, 14–26.

15. Smith, B.J., Martins-de-Souza, D., and Fioramonte, M. (2019) A Guide to Mass Spectrometry-Based Quantitative Proteomics, in *Pre-Clinical Models: Techniques and Protocols* (eds.Guest, P.C.), Springer New York, New York, NY, pp. 3–39.

16. Mertins, P., Tang, L.C., Krug, K., Clark, D.J., Gritsenko, M.A., Chen, L., Clauser, K.R., Clauss, T.R., Shah, P., Gillette, M.A., Petyuk, V.A., Thomas, S.N., Mani, D.R., Mundt, F., Moore, R.J., Hu, Y., Zhao, R., Schnaubelt, M., Keshishian, H., Monroe, M.E., Zhang, Z., Udeshi, N.D., Mani, D., Davies, S.R., Townsend, R.R., Chan, D.W., Smith, R.D., Zhang, H., Liu, T., and Carr, S.A. (2018) Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography–mass spectrometry. *Nat. Protoc.*, **13** (7), 1632–1661.

17. O'Neill, J.R. (2019) An Overview of Mass Spectrometry-Based Methods for

Functional Proteomics, in *Functional Proteomics: Methods and Protocols* (eds.Wang, X., and Kuruc, M.), Springer New York, New York, NY, pp. 179–196.

18. Kim, M.-S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S., Thomas, J.K., Muthusamy, B., Leal-Rojas, P., Kumar, P., Sahasrabuddhe, N.A., Balakrishnan, L., Advani, J., George, B., Renuse, S., Selvan, L.D.N., Patil, A.H., Nanjappa, V., Radhakrishnan, A., Prasad, S., Subbannayya, T., Raju, R., Kumar, M., Sreenivasamurthy, S.K., Marimuthu, A., Sathe, G.J., Chavan, S., Datta, K.K., Subbannayya, Y., Sahu, A., Yelamanchi, S.D., Jayaram, S., Rajagopalan, P., Sharma, J., Murthy, K.R., Syed, N., Goel, R., Khan, A.A., Ahmad, S., Dey, G., Mudgal, K., Chatterjee, A., Huang, T.-C., Zhong, J., Wu, X., Shaw, P.G., Freed, D., Zahari, M.S., Mukherjee, K.K., Shankar, S., Mahadevan, A., Lam, H., Mitchell, C.J., Shankar, S.K., Satishchandra, P., Schroeder, J.T., Sirdeshmukh, R., Maitra, A., Leach, S.D., Drake, C.G., Halushka, M.K., Prasad, T.S.K., Hruban, R.H., Kerr, C.L., Bader, G.D., Iacobuzio-Donahue, C.A., Gowda, H., and Pandey, A. (2014) A draft map of the human proteome. *Nature*, **509**, 575.

19. Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A.M., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J.-H., Bantscheff, M., Gerstmair, A., Faerber, F., and Kuster, B. (2014) Mass-spectrometry-based draft of the human proteome. *Nature*, **509**, 582.

20. Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J., and Gygi, S.P. (2003) Evaluation of Multidimensional Chromatography Coupled with Tandem Mass Spectrometry (LC/LC−MS/MS) for Large-Scale Protein Analysis: The Yeast Proteome. *J. Proteome Res.*, **2** (1), 43–50.

21. Wang, Y., Yang, F., Gritsenko, M.A., Wang, Y., Clauss, T., Liu, T., Shen, Y., Monroe, M.E., Lopez-Ferrer, D., Reno, T., Moore, R.J., Klemke, R.L., Camp, D.G., and Smith, R.D. (2011) Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells.

Proteomics, **11** (10), 2019–2026.

22. Washburn, M.P., Wolters, D., and Yates III, J.R. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.*, **19**, 242.

23. Craven, R.A., and Banks, R.E. (2001) Laser capture microdissection and proteomics: Possibilities and limitation. *Proteomics*, **1** (10), 1200–1204.

24. Altelaar, a. F.M., and Heck, A.J.R. (2012) Trends in ultrasensitive proteomics. *Curr. Opin. Chem. Biol.*, **16**, 206–213.

25. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods*, **13** (9), 731–740.

26. Mi, H., Poudel, S., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2016) PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res.*, **44** (D1), D336–D342.

27. Raijmakers, R., Berkers, C.R., de Jong, A., Ovaa, H., Heck, A.J.R., and Mohammed, S. (2008) Automated Online Sequential Isotope Labeling for Protein Quantitation Applied to Proteasome Tissue-specific Diversity. *Mol. Cell. Proteomics*, **7** (9), 1755–1762.

28. Dayon, L., Núñez Galindo, A., Corthésy, J., Cominetti, O., and Kussmann, M. (2014) Comprehensive and Scalable Highly Automated MS-Based Proteomic Workflow for Clinical Biomarker Discovery in Human Plasma. *J. Proteome Res.*, **13** (8), 3837–3845.

29. Böhm, G., Prefot, P., Jung, S., Selzer, S., Mitra, V., Britton, D., Kuhn, K., Pike, I., and Thompson, A.H. (2015) Low-pH Solid-Phase Amino Labeling of Complex Peptide Digests with TMTs Improves Peptide Identification Rates for Multiplexed Global Phosphopeptide Analysis. *J. Proteome Res.*, **14** (6), 2500–2510.

30. Pirmoradian, M., Budamgunta, H., Chingin, K., Zhang, B., Astorga-Wells, J., and

Zubarev, R.A. (2013) Rapid and Deep Human Proteome Analysis by Single-dimension Shotgun Proteomics. *Mol. Cell. Proteomics*, **12** (11), 3330–3338.

31. Stasyk, T., and Huber, L.A. (2004) Zooming in: Fractionation strategies in proteomics. *Proteomics*, **4** (12), 3704–3716.

32. Hughes, C.S., Foehr, S., Garfield, D.A., Furlong, E.E., and Steinmetz, L.M. (2014) Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.*, **10** (757), 1–15.

33. Manadas, B., Mendes, V.M., English, J., and Dunn, M.J. (2010) Peptide fractionation in proteomics approaches. *Expert Rev. Proteomics*, **7** (5), 655–663.

34. Nägele, E., Vollmer, M., Hörth, P., and Vad, C. (2004) 2D-LC/MS techniques for the identification of proteins in highly complex mixtures. *Expert Rev. Proteomics*, **1** (1), 37–46.

35. Vollmer, M., Hörth, P., and Nägele, E. (2004) Optimization of Two-Dimensional Off-Line LC/MS Separations To Improve Resolution of Complex Proteomic Samples. *Anal. Chem.*, **76** (17), 5180–5185.

36. Donato, P., Cacciola, F., Mondello, L., and Dugo, P. (2011) Comprehensive chromatographic separations in proteomics. *J. Chromatogr. A*, **1218** (49), 8777–8790.

37. Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.*, **2** (8), 1896–1906.

38. Han, D., Moon, S., Kim, Y., Kim, J., Jin, J., and Kim, Y. (2013) In-depth proteomic analysis of mouse microglia using a combination of FASP and StageTip-based, high pH, reversed-phase fractionation. *Proteomics*, **13** (20), 2984–2988.

39. Lawrence, R.T., Perez, E.M., Hernández, D., Miller, C.P., Haas, K.M., Irie, H.Y., Lee, S.-I., Blau, C.A., and Villén, J. (2015) The Proteomic Landscape of Triple-Negative Breast Cancer. *Cell Rep.*, **11** (4), 630–644.

40. Chen, W., Wang, S., Adhikari, S., Deng, Z., Wang, L., Chen, L., Ke, M., Yang, P., and Tian, R. (2016) Simple and Integrated Spintip-Based Technology Applied for Deep Proteome Profiling. *Anal. Chem.*, **88** (9), 4864–4871.

41. Edelhoch, H. (1967) *Spectroscopic Determination of Tryptophan and Tyrosine in Proteins*.

42. Lowry, O.H., Rosebrough, N.J., Farr, L.A., and Randall, R.J. (1951) Protein Measurement with the Folin phenol reagent. *J. Biol. Chem.*, **193** (1), 265–275.

43. Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, **72** (1), 248–254.

44. Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., and Klenk, D.C. (1985) Measurement of protein using bicinchoninic acid. *Anal. Biochem.*, **150** (1), 76–85.

45. Haughland, R.P. (2005) Protein Detection and Proteomics Technology, in *Molecular Probes Handbook, A Guide to fluorescent probes and labeling technologies*, 11thed., Life Technologies Corporation, pp. 413–472.

46. Bell, P.J.L., and Karuso, P. (2003) Epicocconone, A Novel Fluorescent Compound from the Fungus Epicoccum nigrum. *J. Am. Chem. Soc.*, **125** (31), 9304–9305.

47. Jones, L.J., Haugland, R.P., and Singer, V.L. (2003) Development and characterization of the NanoOrange protein quantitation assay: a fluorescence-based assay of proteins in solution. *Biotechniques*, **34** (4), 856–858.

48. Desjardins, P., Hansen, J.B., and Allen, M. (2009) Microvolume Protein Concentration Determination using the NanoDrop 2000c Spectrophotometer. *JoVE*, (33), e1610.

49. Thermo Scientific (2007) BCA ™ Protein Assay Kit. **23235**.

50. Thermo Scientific (2008) NanoOrange ® Protein Quantitation Kit. **MP 06666**.

51. Pierce Biotechnology (2011) Modified Lowry Protein Assay Reagent Kit. **23240**.

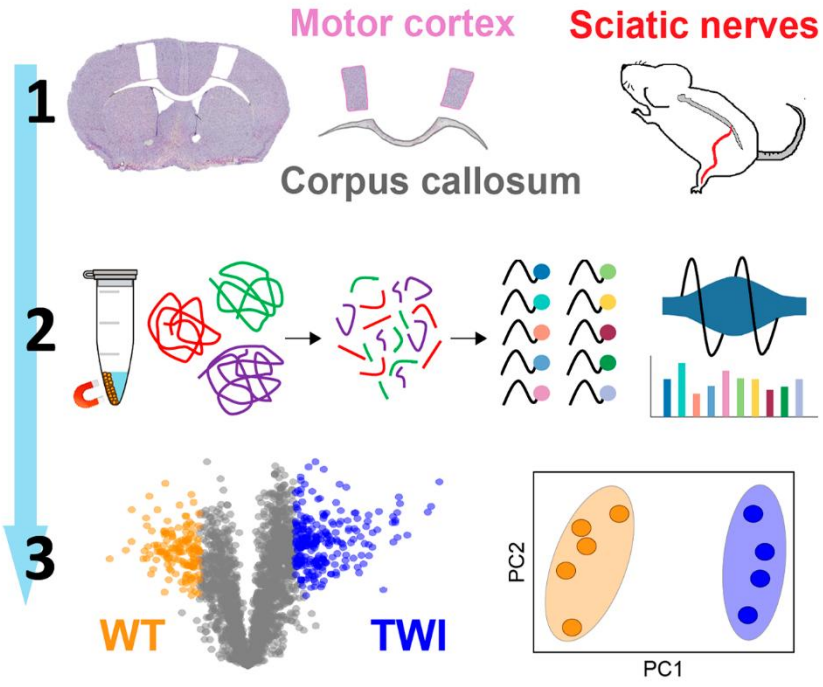
52. Thermo Scientific (2013) Coomassie (Bradford) Protein Assay Kit. **23200**.

53. Thermo Scientific (2015) Qubit ™ Protein Assay Kits. **MAN0006883**.

54. Thermo Scientific (2017) FluoroProfile® Protein Assay. **269**–**309102**.

Chapter 5

Quantitative Microproteomics Based Characterization of the Central and Peripheral Nervous System of a Mouse Model of Krabbe Disease



5.1. **Summary**

Krabbe disease is a rare, childhood lysosomal storage disorder caused by a deficiency of galactosylceramide beta-galactosidase (GALC). The major effect of GALC deficiency is the accumulation of psychosine in the nervous system and widespread degeneration of oligodendrocytes and Schwann cells, causing rapid demyelination. The molecular mechanisms of Krabbe disease are not yet fully elucidated and a definite cure is still missing.

Here we report the first in-depth characterization of the proteome of the Twitcher mouse, a spontaneous mouse model of Krabbe disease, to investigate the proteome changes in the Central and Peripheral Nervous System. We applied a TMT-based workflow to compare the proteomes of the corpus callosum, motor cortex and sciatic nerves of littermate Twitcher and wild-type mice. More than 700 protein groups exhibited differences in expression and included proteins involved in pathways that can be linked to Krabbe disease, such as inflammatory and defense response, lysosomal proteins accumulation, demyelination, reduced nervous system development and cell adhesion. These findings provide new insights on the molecular mechanisms of Krabbe disease, representing a starting point for future functional experiments to study the molecular pathogenesis of Krabbe disease. Data are available via ProteomeXchange with identifier PXD010594.

5.2. Introduction

Krabbe disease (KD), also known as globoid cell leukodystrophy, is a rare autosomal recessive sphingolipidosis and is one of a larger group of lysosomal storage disorders (LSDs). KD is a neurodegenerative disorder and occurs due to mutations in the β-galactocerebrosidase gene (*galc*). These mutations lead to a reduced, or loss of, activity of the encoded enzyme (β-galactocerebrosidase, GALC), leading to a disruption of myelin turnover in the central nervous system (CNS) and peripheral nervous system (PNS) [1][2].

The incidence of KD in the US is 1:100,000 and in 95% of cases onset occurs within the first 6 months of life (infantile KD) [3][4]. Patients develop progressive blindness, ataxia and psychomotor regression, with death typically occurring within 2 years [5][6]. Late onset KD (late infantile, late juvenile and adult form) is characterized by milder progression and severity [7][3]. Currently there are neither long-term therapies nor a definitive cure for KD. Hematopoietic stem cell transplantation (HSCT) is the only treatment available and has been shown to slow the course of the disease in pre-symptomatic infantile patients [8]. Additional treatment strategies, including gene therapy, substrate reduction therapy, chemical chaperones and enzyme replacement therapy, are currently under investigation in animal models of KD [9][10][11][12][13][14]. The Twitcher mouse represents the most common animal model of human KD, as it shows similar clinical and histopathological features [15].

The "psychosine hypothesis" is the established explanation for KD pathogenesis [16]. Defective GALC leads to impaired degradation of the two glycolipids: galactosylceramide (Gal-cer), the primary substrate of GALC, and psychosine (PSY) [17]. Whereas Gal-cer is also degraded by GM1 ganglioside β-galactosidase [18], there is not another degradation pathway for PSY. PSY is a cytotoxic sphingolipid that is produced by galactosylation of sphingosine by ceramide galactosyltransferase (CGT or UGT8). CGT is mainly expressed in myelinating cells and so in KD psychosine accumulates in oligodendrocytes and Schwann cells and is believed to be the main cause of demyelination [19][20]. The pathophysiological effect is not limited to oligodendrocytes and Schwann cells death, but includes also other cell types of the nervous system because the consequent accumulation

of myelin debris triggers an inflammatory response with astrocytosis and microgliosis [8][17][21]. Furthermore it has been shown that PSY accumulates in membrane microdomains affecting lipid rafts and signaling pathways [22][23][24].

Metabolic profiling of the Twitcher mouse has identified metabolic pathways influenced by KD, revealing decreased levels of long chain fatty acids, increased levels of short chain fatty acids and alteration of several metabolites involved in mitochondrial fuel selection, energy production, inflammation, neurotransmitter metabolism and osmotic regulation [25][26][27][28]. However, the proteome changes associated with KD and the pathogenic mechanisms are still not well understood.

MS-based proteomics has been used to study several LSDs, including Niemann-Pick type C disease [29][30][31], Gaucher disease [32][33] and Fabry disease [34][35]. Mass spectrometry has been extensively used to quantify psychosine in several tissues and cells used to study KD, including the Twitcher mouse brain [22], spinal cord [36], serum [37], newborn dried blood spots from infants [38] and a cell model of KD derived from the Twitcher mouse [39]. However, to date an in-depth proteome characterization of the central and peripheral nervous system during disease progression is lacking. Here we report the first in-depth characterization of the central and peripheral nervous system of the Twitcher mouse. 10-plex Tandem Mass Tags (TMT) based experiments were used to compare the proteomes of the corpus callosum, motor cortex and sciatic nerves of littermate Twitcher and wild-type mice.

5.3. Experimental procedures

5.3.1. Experimental design and statistical rationale

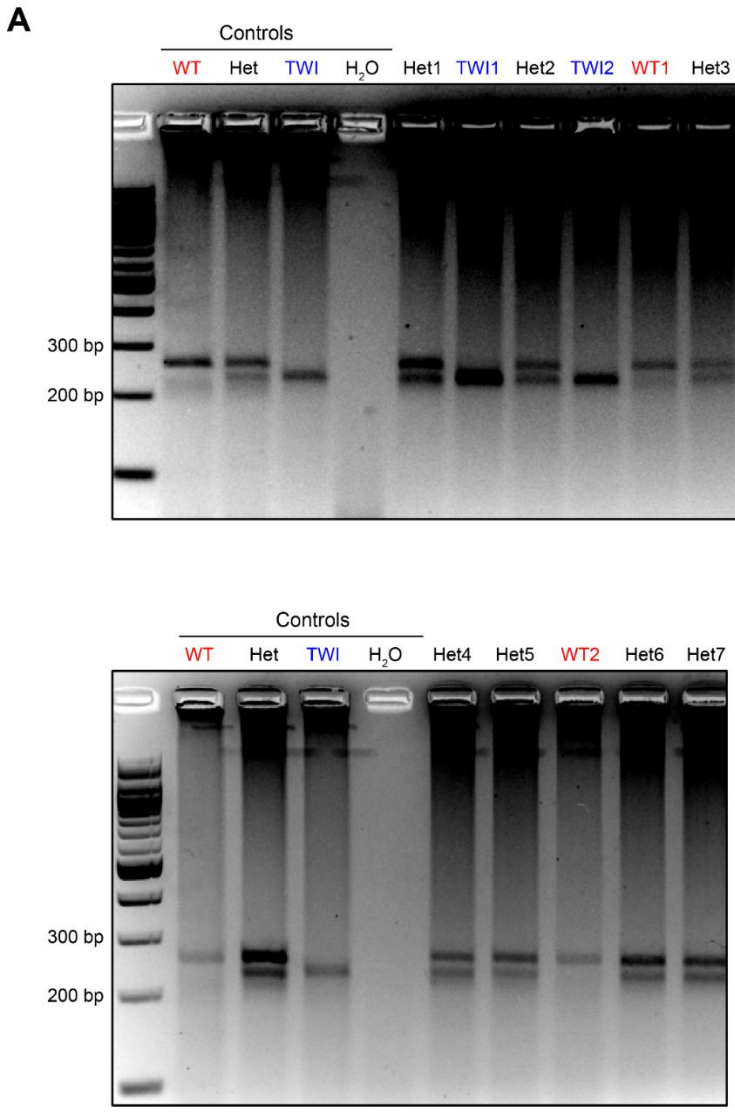
We investigated the proteome changes associated with Krabbe disease by comparing the proteomes extracted from the corpus callosum, motor cortex and sciatic nerves of littermate homozygous Twitcher and homozygous wild-type mice. We performed a TMT 10-plex experiment for each tissue region using independent biological replicates (n = 5) for each mouse type. The five TWI and five WT samples within each TMT 10-plex set were randomized using the Random.org list randomizer (www.random.org), and the expression levels of the confidently identified proteins compared using a two-sided Student's t-test with a permutation-based FDR cutoff (250 randomizations, FDR 0.01, S0 1).

5.3.2. Materials

Trypsin/LysC mix Mass Spec grade was purchased from Promega (Madison, WI). Tandem Mass Tags (TMT 10-plex) kits and microBCA protein assay kit were purchased from Thermo Fisher Scientific (Rockford, IL). Polyethylene naphthalate (PEN) membrane slides were purchased from Carl Zeiss (Carl Zeiss Microsystems GmbH, Göttingen, Germany). All other reagents and solvents were purchased from Sigma-Aldrich (St. Louis, MO).

5.3.3. Mice

Twitcher (strain B6.CE-Galctwi/J) animals were bred at the Center for Experimental Biomedicine of CNR in Pisa, authorized for the use of animals for scientific purposes by the Ministry of Health (Authorization No. 114/2003-A of 16/9/2003). Animals were maintained under standard housing conditions and used according to the protocols and ethical guidelines of the Italian (DLGS 26/2014; Permit number: PT5.15, July 2015) and European Union (2010/63/EU) laws. Experiments were conducted in parallel on Twitcher-wild type mice (WT) and littermate Twitcher-mutant homozygous mice (TWI), while Twitcher-mutant heterozygous mice (Het) were used for the TWI colony maintenance. Genomic DNA was extracted from the clipped tails of mice by Proteinase K digestion and subsequent genomic DNA extraction (EUROGOLD Tissue-DNA Mini kit, Euroclone) as previously described [40][40]. The genetic status of each mouse was determined from the genome analysis of the Twitcher mutation (Figure 5.1.), following the method reported by Sakai *et al.* [41].


A
Controls
WT
Het
TWI
H2O
Het1
TWI1
Het2
TWI2
WT1
Het3
TWI3
300 bp
200 bp
Controls
WT
Het
TWI
H2O
Het4
Het5
WT2
Het6
Het7
300 bp
200 bp

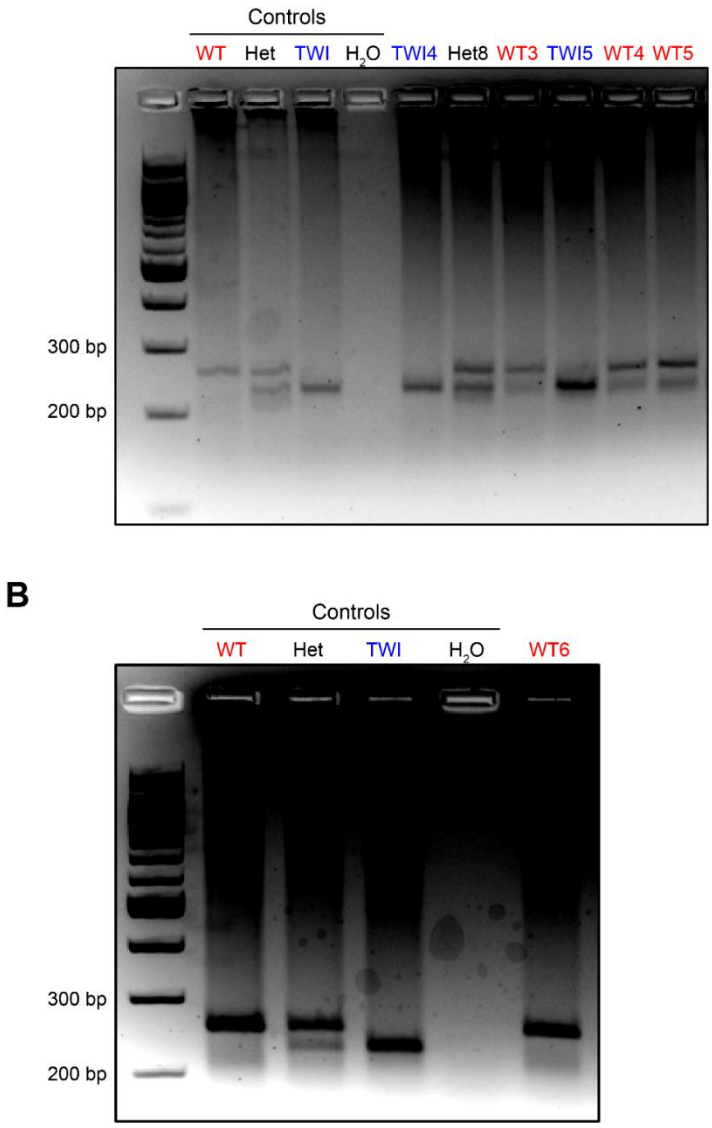


Figure 5.1. *Genotyping of the Galc mutation. (A) Gel electrophoresis of the littermate mice, indicating the five homozygous WT and five homozygous TWI mice selected for proteomics analysis. Heterozygous animals were not included in the experimental design. (B) Gel electrophoresis of the WT mouse mislabeled as TWI, clearly confirming its WT genotype. The genotype was determined by PCR amplification of a genomic DNA fragment in intron 16 of the GALC gene using forward and reverse primers, 5'-CCACTCCCATCCTTTCTCC-3' and 5'-GGCCATCACATTCGTCAGA-3' respectively.*

Five WT (3 male and 2 female) and five TWI (4 male and 1 female) mice were sacrificed by cervical dislocation, dissected and the brain extracted [42] and snap frozen in liquid nitrogen for 15 s. Sciatic nerves were dissected by making a 5 mm vertical incision along the thigh [43]. The muscles were split until the entire length of the sciatic nerve in the thigh region was exposed. The nerves were then gently lifted using forceps and removed by cutting at the proximal and distal ends. Brains and sciatic nerves were stored at -80°C until analysis.

5.3.4. Laser Capture Microdissection

Consecutive coronal brain sections, 10µm thick, were cut at 0.74-0.98 mm from Bregma using a cryostat (CM1950, Leica Microsystems Srl, Milan, Italy) and thaw mounted onto PEN membrane slides (previously conditioned in UV light for 30 minutes). After a light hematoxylin staining, small regions of 2.5 - 3.3mm^2 were isolated from the corpus callosum and the motor cortex using a PALM microbeam laser capture microdissection system (Carl Zeiss MicroImaging, Munich, Germany). Microdissection was performed using a x40 ocular lens and a 355nm laser for cutting the tissue and catapulting the isolated regions of tissue into adhesive cap tubes (Carl Zeiss). Samples were stored at -80°C until analysis.

5.3.5. Protein extraction and digestion

The isolated brain samples were dissolved in 20µL of lysis buffer consisting of 50% trifluoroethanol (TFE), 0.5% sodium dodecyl sulfate (SDS), 2.5mM ethylenediaminetetraacetic acid (EDTA), 2.5mM ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid (EGTA), 10mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 8.5 and protease inhibitor (cOmplete^{TM}, Mini, EDTA-free Inhibitor Cocktail). The sciatic nerves were homogenized in 100µL of the same lysis buffer.

Proteins were extracted by sonication at 4°C using a Bioruptor Pico (Diagenode, Seraing, Belgium - 10 cycles of 30s ON and 30s OFF). Protein quantification was performed on a 1µL aliquot of each sample using a modified microBCA assay [44].

Each anatomical region was compared using an identical amount of extracted protein, and which corresponded to the maximum amount that could be obtained from all ten animals, namely 1.5, 2.5 and 3µg for the corpus callosum, motor cortex and sciatic nerves, respectively. Protein digestion was performed using a modified SP3 protocol [44][45]. The protein extracts were mixed with 20µL of lysis buffer and 2µL of paramagnetic beads (100mg/mL solution of 50% Speedbeads A (GE45152105050250, Sigma) and 50% Speedbeads B (GE65152105050250, Sigma)) in 0.2mL PCR tubes (Sarstedt). Briefly, proteins were denaturated at 95°C for 5 min. Reduction (DTT 200mM), alkylation (IAM 400mM), protein purification and overnight trypsin/Lys-C digestion (1:25 enzyme/protein) steps were performed in the same tube.

5.3.6. TMT labeling

The protein expression levels in the tissue samples from all ten animals were compared using 10-plex TMT isobaric labeling. After digestion the samples were vortexed and sonicated for 5min with a Bioruptor Pico (5 cycles of 30s ON and 30s OFF). In-solution TMT labeling was performed using a 1:20 peptide/TMT proportion. Three TMT 10-plex reagents (0.8mg) were dissolved in 26, 16, and 13µL of ACN and used for the labeling of the peptides obtained from the corpus callosum, motor cortex and sciatic nerves, respectively (note the TMT solutions were made up using different volumes in order to ensure 2µL of the TMT solutions provided the correct excess for each tissue extract). The TMT label was added in two equal steps of 1µL, each followed by an incubation period of 30min. TMT labeling was quenched by adding 2µL of 4% hydroxylamine and incubating for 15min. After TMT labeling, the samples were transferred to 0.5mL LoBind tubes (Eppendorf) for peptide purification. Samples were rinsed with 100% ACN to promote peptide binding on the carboxylate coated beads. Peptides were washed twice with 70%

ethanol and once with 100% ACN. The purified peptides were eluted from the beads with a 2% DMSO aqueous solution. The samples from the five WT and five TWI mice were each labeled with a different, randomized TMT label; the labeled peptides from the ten mice were then combined in a 1:1:1:1:1:1:1:1:1:1 ratio. An aliquot corresponding to 1.5µg of total protein content was collected from each TMT set, diluted 1:1 with 10% formic acid and injected into a nanoLC system for the evaluation of the TMT labeling reaction efficiency. The remaining samples were dried down with a speedvac (Eppendorf) and stored at -20°C.

5.3.7. Automated high-pH fractionation

High-pH fractionation was performed using an AssayMap Bravo robot (Agilent Technologies) using the fractionation protocol V1.0 [44]. Briefly, the dried samples were resuspended in 10µL of 2% DMSO, mixed with 100µL of 10mM NH_4OH (pH 10) and loaded on to reversed-phase (RP-S) cartridges. Peptides were isocratically eluted with 35µL plugs of 12, 18, 24, 30, 36 and 80% ACN in 10mM NH_4OH (pH 10). The six fractions and the sample flow-through were transferred to 0.5mL LoBind tubes (Eppendorf), dried down with a speedvac and stored at -20°C.

5.3.8. LC-MS3 analysis

Peptides were resuspended in 10% formic acid and injected into an Easy-nLC1000 (Thermo Scientific) coupled to an Orbitrap Fusion (Thermo Scientific). Peptides were first trapped using a nanoviper trap column (2cm x 75µm, C18, 3µm, 100A; Thermo Scientific) and then separated using an Easyspray analytical column (ES803: 50cm x 75µm, C18, 2µm, 100A; Thermo Scientific) using a flow rate of 300nl/min and a 140min gradient. Peptides were loaded at 800 bar followed by a non-linear gradient: 0-1min, 8%B;

t=105min, 25%B; t=120min, 35%B; t=130min, 90%B; t=140min, 90%B. Buffer A consisted of 0.1% FA and Buffer B of 99.9% ACN and 0.1% FA.

The Orbitrap Fusion was operated in a data dependent top-speed method using a 2 second maximum cycle time and multi-notch synchronous precursor selection (SPS) for MS3 based TMT quantification. The survey scan was performed in the Orbitrap (*m/z* 375-1500, 120k resolution, AGC target 2e5, 50 ms maximum injection time). Monoisotopic precursor selection and a dynamic exclusion of 70s were adopted. Ions with charge states from 2+ to 7+ and intensity greater than 5e3 were selected for CID fragmentation (35% NCE) using an isolation window of 1.6 *m/z*.

MS2 spectra were recorded in the linear ion trap with a rapid scan rate, 5e3 AGC target and 125ms maximum injection time. Following fragmentation, multinotch (synchronous) precursor selection was performed to select the 8 most abundant fragment ions for HCD-MS3 (50% NCE). MS3 scanning was performed in the Orbitrap at a resolution of 60K, with an AGC target of 1e5 and 150ms maximum injection time.

5.3.9. Data analysis

Raw data files were analyzed using Proteome Discoverer 2.1 (Thermo Scientific). For each tissue type the raw files from the 7 fractions were merged and searched with the SEQUEST HT [46] search engine against the Mus Musculus Swiss-Prot protein database (July 2016, 16,808 entries) supplemented with a common contaminant database (246 entries). Searches were performed using the TMT reagents (+229.163 Da, lysine and N-termini) and carbamidomethyl (+57.021 Da, cysteine) as static modifications, methionine oxidation (+15.995 Da) as a dynamic modification, 20ppm precursor mass tolerance, 0.6 Da fragment mass tolerance, and 20ppm reporter ions tolerance. The search was performed using fully tryptic peptides with a minimum length of 6 amino acids and up to 2 missed cleavages. Results were filtered for a 1% false discovery rate (FDR) using the Percolator algorithm and additionally filtered for a minimum Xcorr score of 1.8. At least one unique peptide was required for definitive protein identification. Protein intensities were

calculated based on label intensities of the 3 most abundant peptides with a co-isolation threshold of 50% taking into account unique peptides.

TMT labeling efficiency and over labeling rate were evaluated using 1.5 µg aliquots of each TMT set. TMT labeling efficiency was evaluated by setting the TMT modification (+229.163 Da) at N-termini and lysine residues as dynamic and determining the percentage of labeled PSMs. TMT over labeling was evaluated by setting the TMT modification at N-termini and lysine residues as static, and TMT modification at serine, threonine, histidine and tyrosine as dynamic, and then determining the percentage of TMT-labeled serine, threonine, histidine and tyrosine.

Search results were exported as txt files and processed with Perseus 1.5 [47]. For each TMT experiment the protein intensities were log2 transformed and subject to a median normalization. Figure 5.2. shows the boxplots of the log2 transformed protein intensities before and after normalization for all the datasets.

Data were filtered such that each protein was quantified in at least four TWI and four WT mice. Principal Component Analysis was performed on the normalized and filtered datasets. Significantly different protein levels between TWI and WT mice for the three TMT experiments were calculated using a two-sided Student's *t*-test using a permutation-based FDR cutoff (250 randomizations, FDR 0.01, S0 1). Proteins were considered as differentially regulated if their adjusted p-value corresponded to an FDR lower or equal to 0.01 and their fold change (expressed as log2 ratio) was <-1 or >+1.

Gene ontology was performed with WebGestalt [48] using the Overrepresentation Enrichment Analysis (ORA) method. For each TMT experiment the input protein list included the proteins significantly up or down regulated in the TWI mice as well as the proteins that were quantified only in the TWI or WT mice (proteins quantified in at least 3 TWI or WT mice and not quantified in at least three WT or TWI mice). The reference protein list included all proteins identified in the TMT experiment. A Benjamini-Hochburg (BH) method for multiple test adjustment was used and the FDR was set at 0.05.

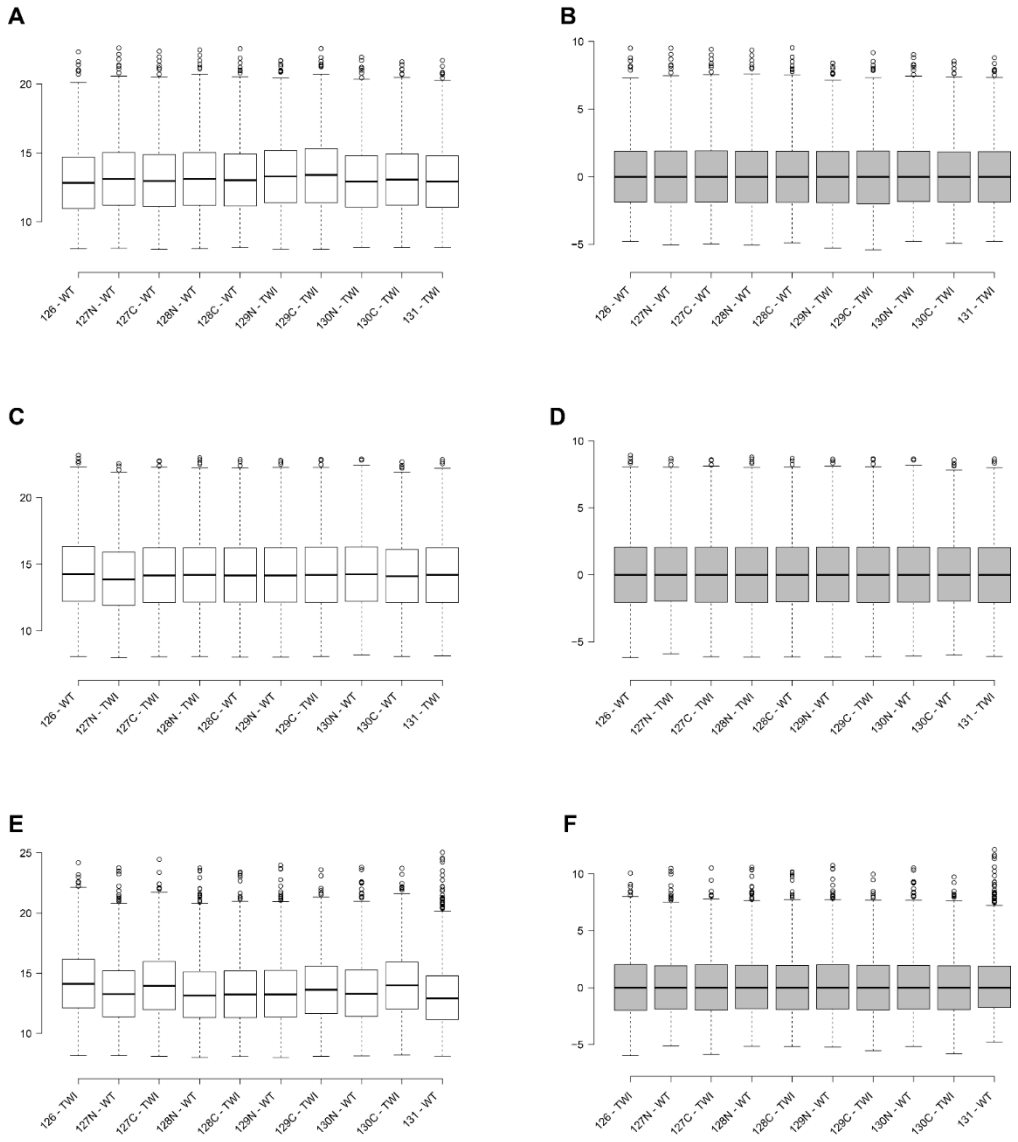


Figure 5.2. *Tukey box plots with whiskers to 1.5 interquartile range of the log2 transformed protein groups intensities for the corpus callosum dataset before (A) and after median normalization (B), motor cortex before (C) and after (D) median normalization and sciatic nerves before (E) and after (F) median normalization. For each box plot the horizontal line represents the median value.*

5.3.10. Western blot analysis

Western blot analysis of seven selected proteins were performed on the sciatic nerves extracts to validate the LC-MS/MS results. Proteins were extracted from the sciatic nerves and quantified identically as for the LC-MS/MS analysis (see *Protein extraction and digestion* section above). A WT mice pooled sample and a TWI mice pooled sample were prepared by combining the sciatic nerves extracts of three WT and three TWI mice, respectively. The western blot analysis was performed in technical triplicate on the two pooled samples. The samples were boiled in Laemli buffer containing b-mercaptoethanol (5% final concentration) for 5min and centrifuged at room temperature. The supernatants were used for gel electrophoresis (SDS-PAGE). Samples (25µg) were resolved by SDS-PAGE using Gel Criterion XT-Precasted polyacrylamide gel 4–12% Bis-Tris (Bio-Rad, Hercules, CA) and subsequently transferred to nitrocellulose membranes [49]. Immunodetection was performed for ATG16L1 (Abcam, Cambridge, UK; catalog No. ab188642), VAMP8 (Abcam; catalog No. ab76021), UGT8 (Abcam; catalog No. ab170351), CTSB (Abcam; catalog No. ab58802), SQSTM1 (Abcam; catalog No. ab56416), HEXB (Santa Cruz Biotechnology, Dallas, Texas (USA); catalog No. sc-376781), LAMP1 (Santa Cruz Biotechnology, Dallas, Texas (USA); catalog No. sc-20011), and α-tubulin (Sigma Aldrich; catalog No. T6074). On the following day, the blots were incubated with the corresponding peroxidase-linked secondary antibodies and after incubation the membranes were developed with Clarity enhanced chemiluminescent substrates (Bio-Rad). The chemiluminescent signal was acquired with an ImageQUANT LAS400 scanner (GE Healthcare Life Sciences, Uppsala, Sweden), and the density of immunoreactive bands quantified in ImageJ. The results were normalized to α-tubulin.

5.3.11. Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [50] partner repository with the data set identifier PXD010594.

Reviewer access via Pride -- https://www.ebi.ac.uk/pride/archive/login

Login: reviewer86461@ebi.ac.uk

Password: 40IWTMpZ

5.4. Results

We applied a quantitative microproteomics workflow for the characterization of the changes in the proteome of the central and peripheral nervous system of the Twitcher mouse, the most widely used animal model of Krabbe disease. For the characterization of the central nervous system we focused the analyses on the corpus callosum and the motor cortex. The corpus callosum represents one of the main regions of the brain showing extensive demyelination in Krabbe patients [51][52][53][54] and in the Twitcher mouse [55]. The motor cortex was selected because in humans and rodents KD leads to muscle weakness, spasticity and paralysis [2]. In order to investigate the effects of KD on the peripheral nervous system we chose the sciatic nerves as they show marked demyelination, decreased number of axons and axonopathy in the Twitcher mouse [40][56][57]. Figure 5.3. shows an overview of the experimental approach. 10-plex TMT experiments were used to compare the proteomes of the corpus callosum, motor cortex, and sciatic nerve of five TWI and five WT littermate mice. The proteomic data were validated by western blot analysis on the sciatic nerve extracts.

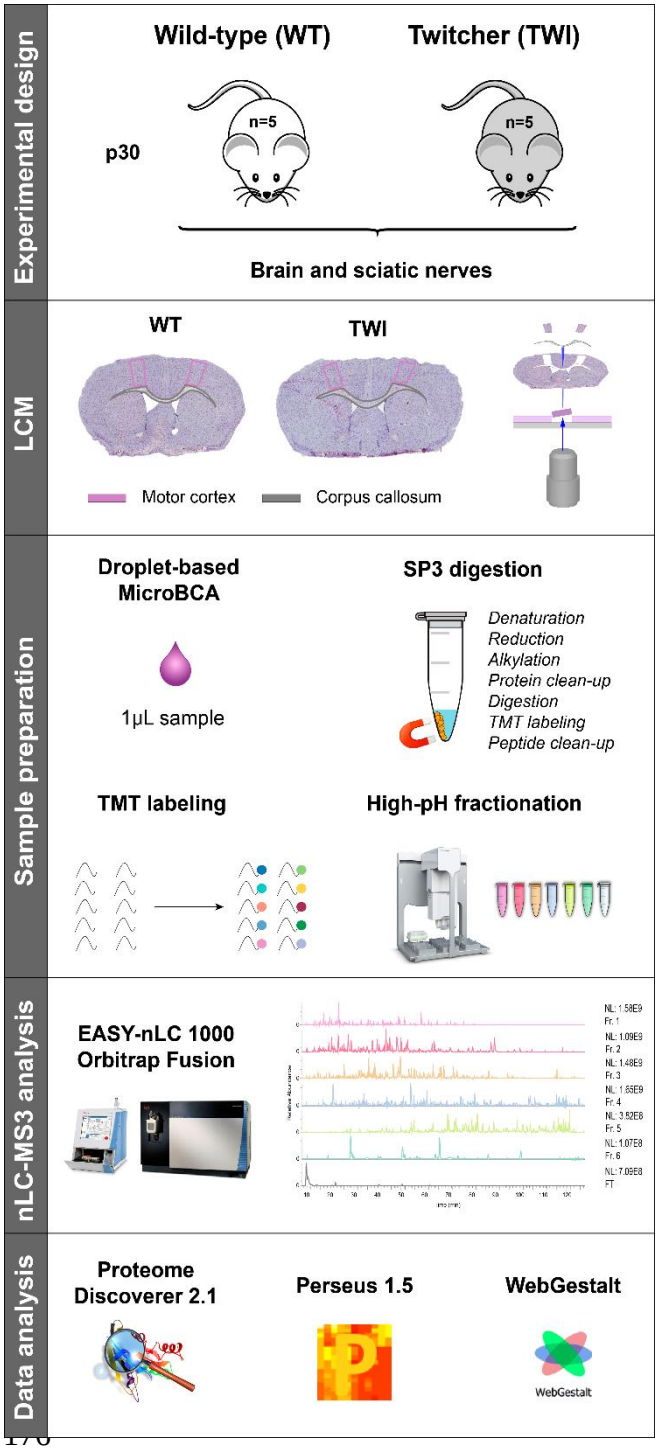


Figure 5.3. *Ultrasensitive microproteomic workflow for the characterization of CNS and PNS of the Twitcher mouse. The brains and sciatic nerves were collected from five TWI and five WT mice at 30 days of age. Laser capture microdissection was used to isolate ROIs from the corpus callosum and motor cortex. Proteins were extracted from the corpus callosum, motor cortex and sciatic nerves, quantified with a modified microBCA assay and digested with the SP3 protocol. Peptides from each dataset were pooled and labeled with TMT 10-plex reagents. Labeled peptides were purified and fractionated on a RPS cartridge. MS3 spectra were acquired on an Orbitrap Fusion mass spectrometer and proteins were identified with Proteome Discoverer 2.1. Data and statistical analysis were performed with Perseus and WebGestalt softwares. The proteomic data were validated by Western blot analysis.*

5.4.1. Tissue sampling and protein extraction

Laser capture microdissection (LCM) was used to isolate and excise small regions of interest (ROIs) from the corpus callosum and the motor cortex of the mouse brains. The brain morphology was found to be different between WT and TWI mice (Figure 5.4.). The corpus callosum of TWI mice exhibited a greater cell density and reduced white matter compared to WT mice. This morphological feature, due to demyelination of the corpus callosum, is consistent with the known histopathology of the Twitcher mouse brain [55]. The motor cortex did not show any gross morphological difference between TWI and WT mice.

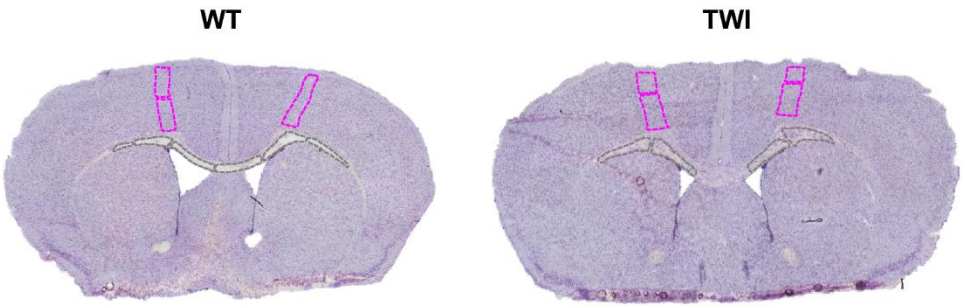


Figure 5.4. *Representative coronal brain sections stained with hematoxylin of a WT and TWI mouse. Violet dashed lines indicate ROIs isolated from the motor cortex, grey dashed lines indicate ROIs isolated from the corpus callosum.*

For each mouse, the small, localized regions of the corpus callosum and motor cortex were isolated from sequential tissue sections to ensure approximately 2μg of protein was available from each region of each mouse. The corpus callosum ROI was isolated from 4-8 brain sections per animal (total area of 2.6-3.6 mm^2) and the motor cortex ROI was isolated from 5-10 sections (total area of 2.5-4.2 mm^2). The protein amount extracted from each sample was estimated using a modified microBCA assay performed on a 1 μL aliquot

[44]. This assay allowed us to quantify the protein content of each sample using only 5% of its total volume. We extracted 1.5-2.4 µg of proteins from the corpus callosum samples and 2.5-4.2 µg from the motor cortex samples. Table 5.1. summarizes the number of tissue sections, ROI areas and protein amounts extracted from the corpus callosum and motor cortex.

Table 5.1. *Summary of the number of tissue sections, ROIs areas and protein amounts extracted from the corpus callosum and motor cortex microdissected samples.*

Brain region	# sections	ROIs area / mm^2	µg proteins
Corpus callosum	4 - 8	2.6 - 3.6	1.5 - 2.4
Motor cortex	5 - 10	3.3 - 5.1	2.5 - 4.2

5.4.2. Proteome profiling

The quantitative microproteomics workflow combined SP3 protein digestion [44][45], in-solution TMT labeling and high-pH fractionation followed by LC-MS3. The corpus callosum and motor cortex analyses used the maximum protein amount available from all mice for each microdissected region (1.5µg for the corpus callosum and 2.5µg for the motor cortex). The analysis of the sciatic nerve was performed on 3µg aliquots of the tissue extracts.

We assessed TMT labeling efficiency on the three datasets in terms of percentage of PSMs with the TMT modification at lysine and N-termini residues (Figure 5.5.A). The labeling efficiency was greater than 96% for all the datasets for both lysine and N-terminal residues. We also assessed the over labeling rate (TMT labeling of serine, threonine, histidine and tyrosine), Figure 5.5.B shows a bar graph indicating the percentage of PSMs with a TMT modification at serine, threonine, histidine and tyrosine residues. The total over labeling rate was less than 5% for all datasets.

The MS3 analyses of the high-pH fractionated TMT-labeled samples resulted in the identification of 3699, 4394 and 3388 protein groups from the corpus callosum, motor cortex and sciatic nerve isolates, respectively. Figure 5.5.C summarizes the number of identified protein groups, peptides PSMs and MS/MS spectra. For the determination of group-wise relative protein quantification the datasets were filtered such that each protein needed to be quantified in at least four TWI and four WT mice. This filtering reduced the number of protein groups to 2607, 3579 and 2350 for the corpus callosum, motor cortex and sciatic nerve, respectively.

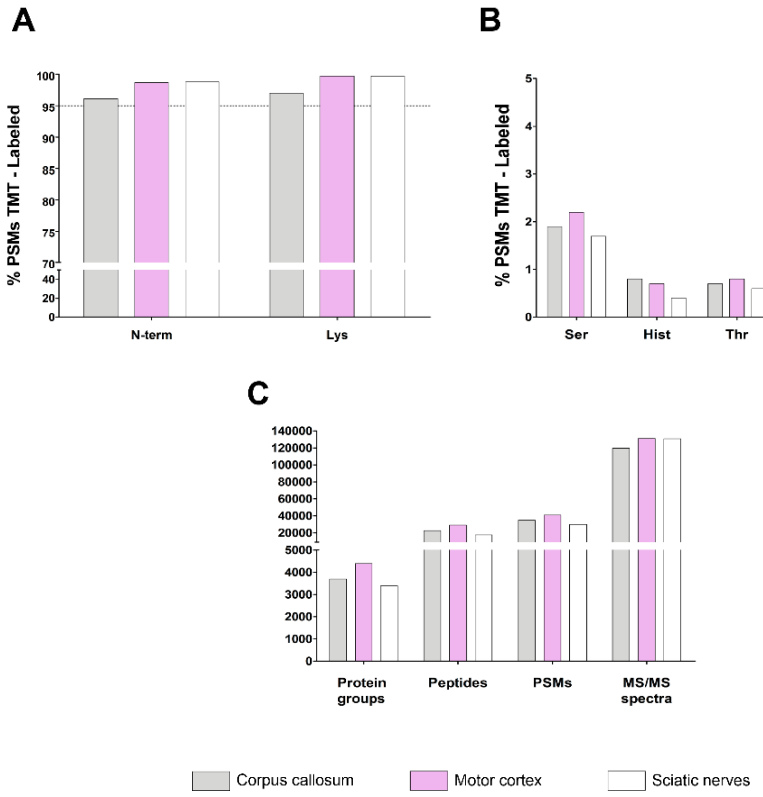


Figure 5.5. *(A) TMT labeling efficiency expressed as the percentage of PSMs with the TMT modification at N-termini and lysine residues. (B) Over-labeling rate expressed as the percentage of PSMS with the TMT modification at serine, histidine, threonine and tyrosine residues. (C) Protein identification metrics: number of identified protein groups, peptides, PSMs and acquired MS/MS*

spectra. Grey, violet and white bars refer to the corpus callosum, motor cortex and sciatic nerves datasets.

Figure 5.6. shows a histogram of the mass errors of the identified peptides, which demonstrates that >95% of all peptide identifications were within 2 ppm, and that >99.9% were within 5 ppm.

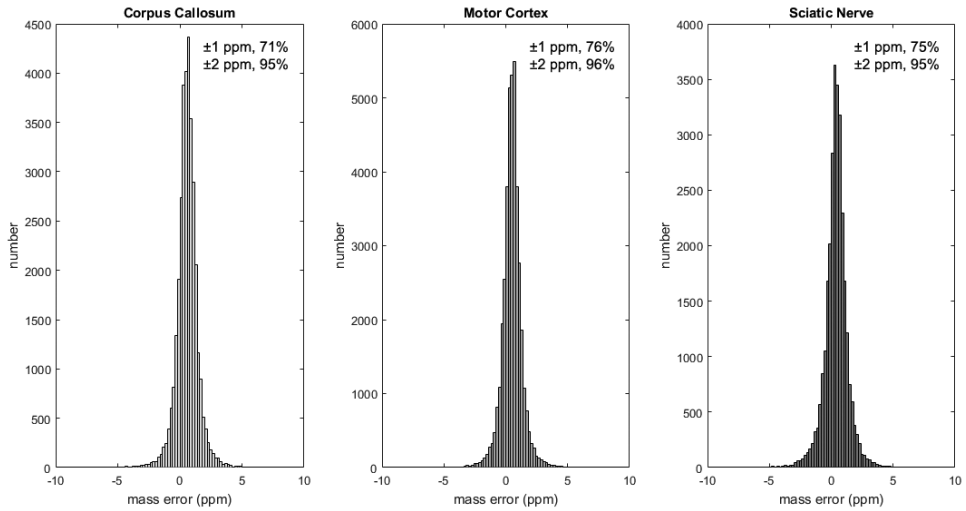


Figure 5.6. *Histograms of mass errors (ppm) for the peptides identified from the extracts of the corpus callosum, motor cortex, and sciatic nerve datasets, demonstrating that >95% of the identifications are within 2 ppm.*

5.4.3. Proteome changes in the CNS and PNS of the Twitcher mouse

Principal Component Analysis (PCA) was performed on the filtered, quantitative data matrices to summarize differences in protein expression between TWI and WT mice. The

PCA score plot revealed that one mouse, originally considered a TWI mouse, fell into the cluster of WT mice for each tissue region (Figure 5.7.). The incorrect assignment of this mouse was further indicated by a GALC enzymatic activity assay performed on the sciatic nerve extract, which showed a GALC activity comparable to that of WT mice. The initial genotyping of the littermate mice is shown in Figure 5.1., which clearly indicates the homozygous TWI and WT mice selected for the experiment. These results indicated a labeling error with one of the TWI mice selected for the experiment; a subsequent genotyping of the suspect TWI-annotated mouse confirmed its status as a WT mouse (Figure 5.1.). Thus, this mouse was considered as a WT in the subsequent statistical analysis.

In the PCA score plots of the corpus callosum and the sciatic nerve datasets (Figure 5.7.A, C) the same WT mouse was well separated by PC1 (thus maximum variance in the data) from both the TWI and WT clusters.

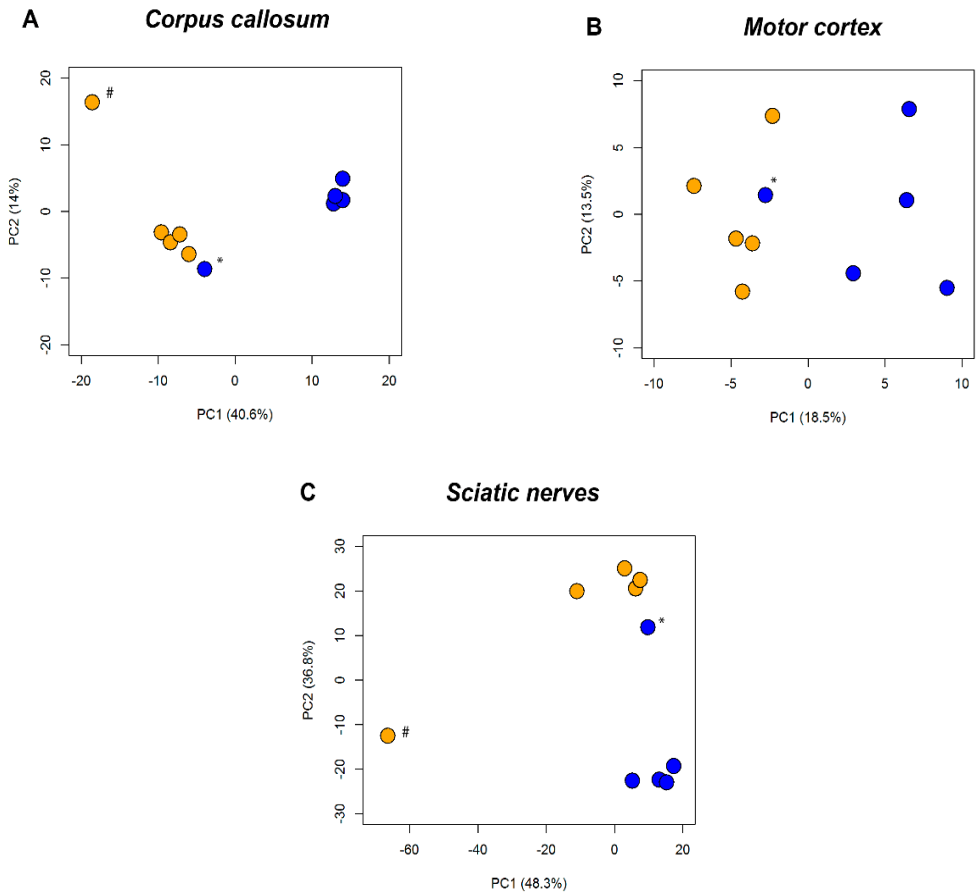


Figure 5.7. *Principal Component Analysis performed on the filtered protein groups expression values and considering all the analyzed samples for the corpus callosum (A), motor cortex (B) and sciatic nerves (C) datasets. Blue circles indicate TWI mice, orange circles indicate WT mice. * indicates the mouse originally considered erroneously as TWI, # indicates the WT mouse considered as an outlier in the corpus callosum and motor cortex dataset.*

We repeated the TMT experiment on another aliquot of the sciatic nerve extracts, from the same cohort of mice, and the same mouse remained an outlier and was thus excluded from subsequent statistical analysis. Once the outlier was removed the TWI and WT mice were well separated as two distinct clusters, separated by PC1, for all tissue regions (Figure 5.8.A-C). The percentage of variation explained by PC1 was 42.3%, 18.5%, and 68% for the corpus callosum, motor cortex, and sciatic nerve, respectively.

Significantly different protein levels between TWI and WT mice for the three datasets were calculated using a two-sided Student's *t*-test with FDR correction for multiple testing. Figure 5.8.D-F show the volcano plots for the datasets from the three tissue regions.

We found 75, 14 and 387 protein groups differentially expressed in the corpus callosum, motor cortex and sciatic nerves of TWI mice, respectively. Specifically, in the corpus callosum 63 protein groups were significantly up-regulated and 12 down-regulated in TWI mice. In the motor cortex 9 protein groups were significantly up-regulated and 5 down-regulated in TWI mice. The sciatic nerve dataset exhibited the largest difference in protein expression, with 16% of the total number of quantified protein groups being differentially expressed (244 protein groups significantly up-regulated in TWI mice and 143 down-regulated).

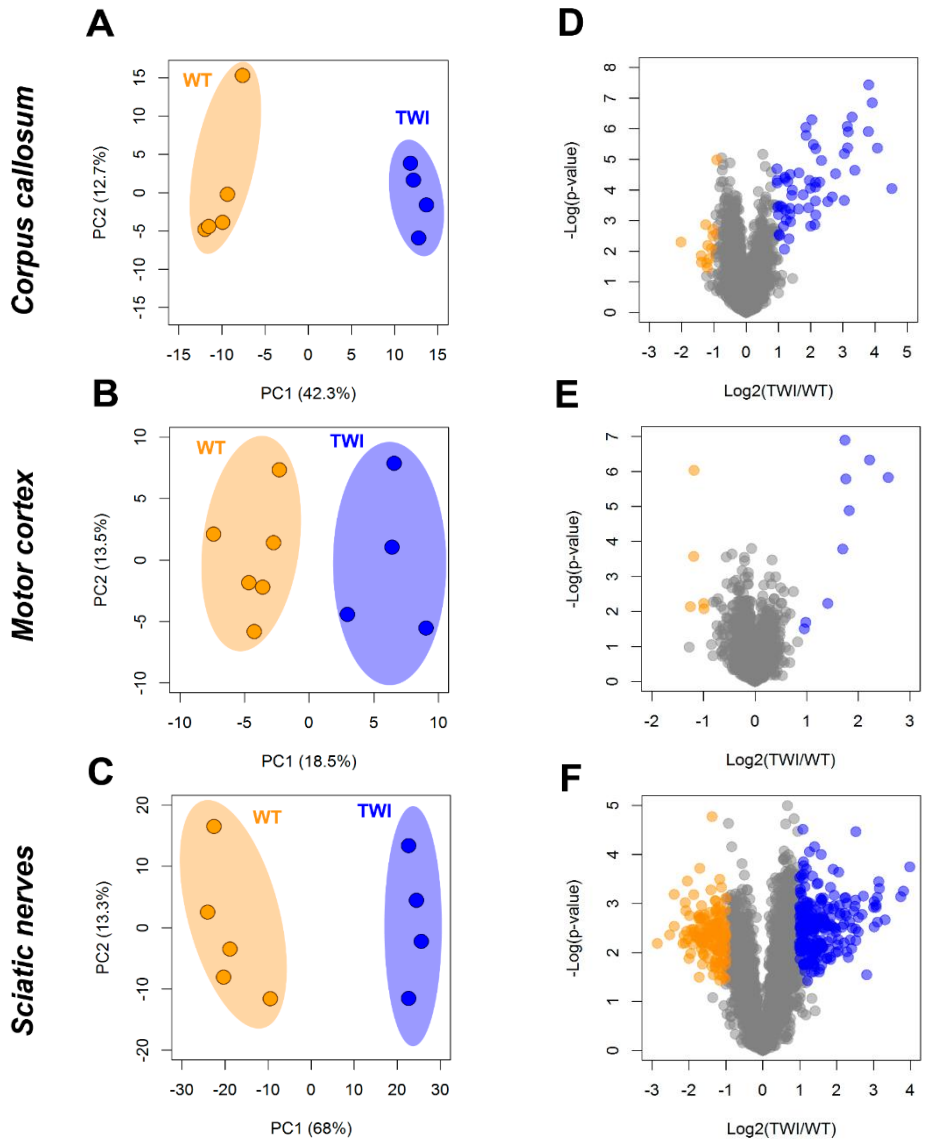


Figure 5.8. *Principal Component Analysis performed on the filtered protein groups expression values for the corpus callosum (A), motor cortex (B) and sciatic nerves (C) datasets. Blue circles indicate TWI mice, orange circles indicate WT mice. Volcano plots for the corpus callosum (D), motor cortex (E) and sciatic nerves (F) datasets. Grey circles indicate non significant protein groups (log2(fold change) < 1 and >-1, FDR > 0.01). Blue circles indicate protein groups significantly up-regulated in the TWI mice (log2(fold change) >1, FDR < 0.01). Orange circles indicate protein groups significantly down-regulated in the TWI mice (log2(fold change) < -1, FDR < 0.01).*

The proteomics data was also searched for highly differentially expressed protein groups, namely those that were quantifiable in only one group of mice (TWI or WT). The data was filtered for protein groups quantified in at least three TWI (or WT) mice but not quantified in at least three WT or TWI mice. We found 73 (corpus callosum), 14 (motor cortex) and 158 (sciatic nerve) protein groups that were quantified only in TWI mice, and 33 (corpus callosum), 18 (motor cortex) and 10 (sciatic nerve) that were quantified only in WT mice. Table 5.2. summarizes the number of protein groups identified, quantified and significantly different between TWI and WT mice for the three datasets. The total number of deregulated protein groups was 181, 46 and 554 for the corpus callosum, motor cortex and sciatic nerves datasets, respectively.

Table 5.2. *Summary of the number of protein groups identified, quantified, up-regulated and down-regulated in TWI mice, quantified only in TWI and quantified only in WT mice for the corpus callous, motor cortex and sciatic nerves datasets.*

# Protein groups	Corpus callosum	Motor cortex	Sciatic nerves
Identified	3699	4394	3388
Quantified	2607	3579	2350
Up-regulated in TWI	63	9	243
Down-regulated in TWI	12	5	143
Quantified only in TWI	73	14	158
Quantified only in WT	33	18	10

5.4.4. Gene Ontology Enrichment Analysis

Gene Ontology (GO) analysis was performed to understand which GO terms are represented in the differentially expressed proteins. We performed an Overrepresentation Enrichment Analysis (ORA) using WebGestalt on the corpus callosum and sciatic nerve

datasets (the small number of differentially expressed proteins in the motor cortex dataset was insufficient for the analysis). Figure 5.9. shows the enriched GO terms for both datasets, in which the bubble graphs show the number of differentially expressed proteins contributing to the terms and the FDR value. Most of the enriched biological processes in the TWI mice are related to inflammatory and defense response. In the sciatic nerves dataset we also found down regulation of processes consistent with axon demyelination (reduced axon development, reduced neuron ensheathment, and reduced nervous system development) as well as reduced microtubule cytoskeleton organization. KEGG pathways upregulated in the TWI mice included those related to inflammatory response (e.g. antigen processing and presentation, leukocyte transendothelial migration, and complement pathway) and to phagocytosis and lysosomes. The glycosaminoglycan degradation pathway, a subclass of the lysosome pathway, was also found to be up-regulated in the TWI mice.

A cellular component ontology analysis was performed to investigate the subcellular localization of the deregulated proteins because Golgi apparatus, endosomes and lysosomes have previously been demonstrated to play critical roles in LSD's [31][58][59]. Figure 5.10. demonstrates that a significant number of the differentially regulated proteins in TWI mice were linked to these organelles.

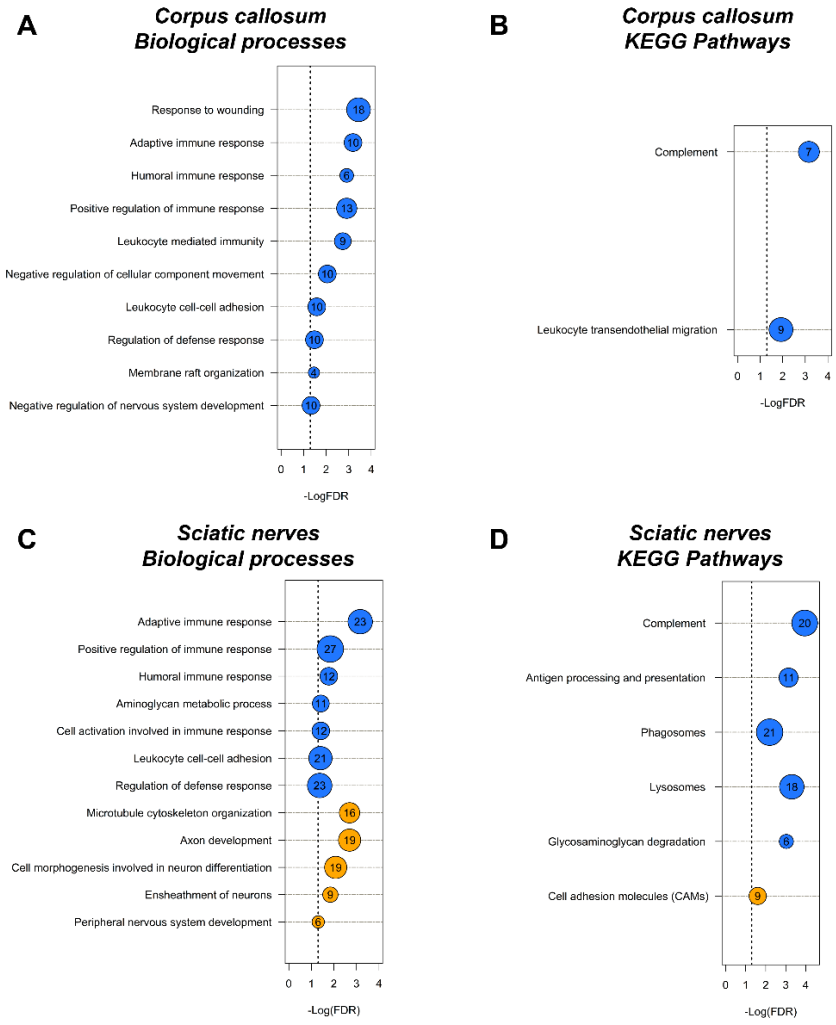


Figure 5.9. *Gene ontology Overrepresentation Enrichment Analysis (ORA) performed on the protein groups that are significantly up- or down-regulated between TWI and WT mice for the corpus callosum and sciatic nerves datasets. A) Biological processes enriched in the corpus callosum dataset, B) KEGG pathways enriched in the corpus callosum dataset, C) biological processes enriched in the sciatic nerves dataset, D) KEGG pathways enriched in the sciatic nerves dataset. The number in the bubble indicates the number of protein groups in the experimental dataset that matched with the corresponding GO term. Blue and orange bubbles indicate biological processes/pathways enriched in the protein groups that are significantly up- and down-regulated in TWI mice, respectively.*

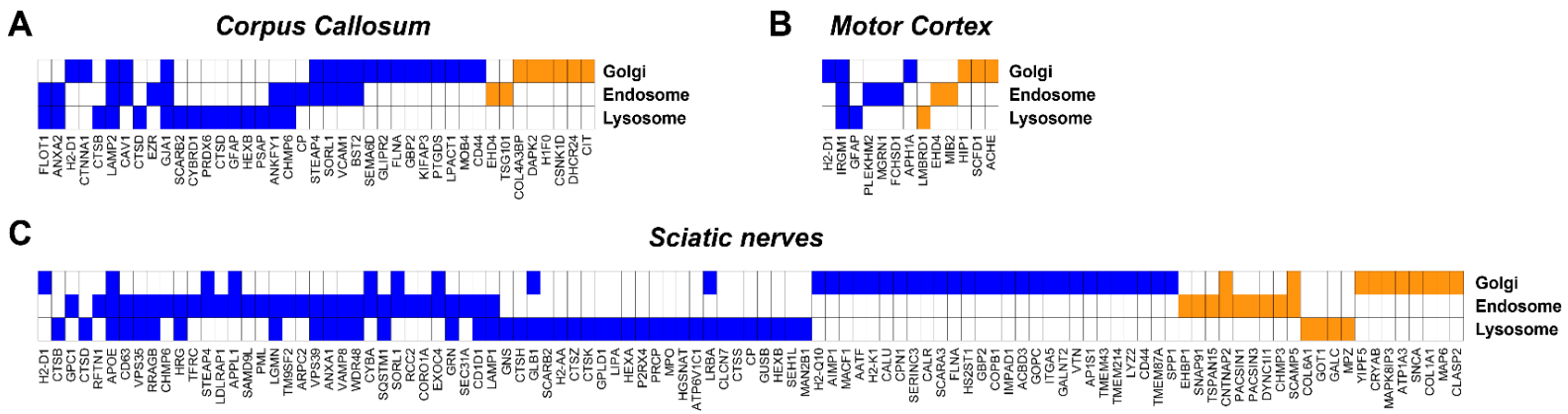


Figure 5.10. *GO - Cellular component analysis. List of deregulated protein groups that are localized on Golgi apparatus, endosomes and lysosomes for the corpus callosum (A), motor cortex (B), and sciatic nerves (C) datasets. Blue and orange indicate protein groups that are significantly up- and down-regulated in TWI mice, respectively.*

5.4.5. Validation by Western blot analysis

A set of the differentially expressed proteins, including proteins related to lysosomes, autophagy and psychosine synthesis, were selected for validation by western blot analysis. These validation experiments were performed using sciatic nerves extracts. Figure 5.11. shows the western blots of the seven selected proteins. The relative expression of all the proteins selected for validation by western blot analysis was in agreement with the LC-MS/MS data. ATG16L1 and UGT8 were down-regulated in the TWI mice. ATG16L1 belongs to the autophagy-related proteins and plays a crucial role in the autophagy pathway as part of a complex with autophagy proteins ATG5 and ATG12 [60]. Down-regulation of ATG16L1 may indicate a reduced efficiency of autophagosome assembly. UGT8 (or CTG) is the enzyme that catalyzes the synthesis of psychosine[61]. The down-regulation of UGT8 in the TWI mice may indicate a homeostatic response of the cell to the accumulation of psychosine. LAMP1, HEXB, CTSB and VAMP8 are lysosomal proteins that were up-regulated in the TWI mice. Deregulation of the lysosomal pathway is a general hallmark of all LSD's, including KD [58][62]. SQSTM1 (also known as ubiquitin binding protein p62) is an autophagy substrate and a marker used to study autophagic flux [63]. SQSTM1 was detected at higher levels in the TWI mice.

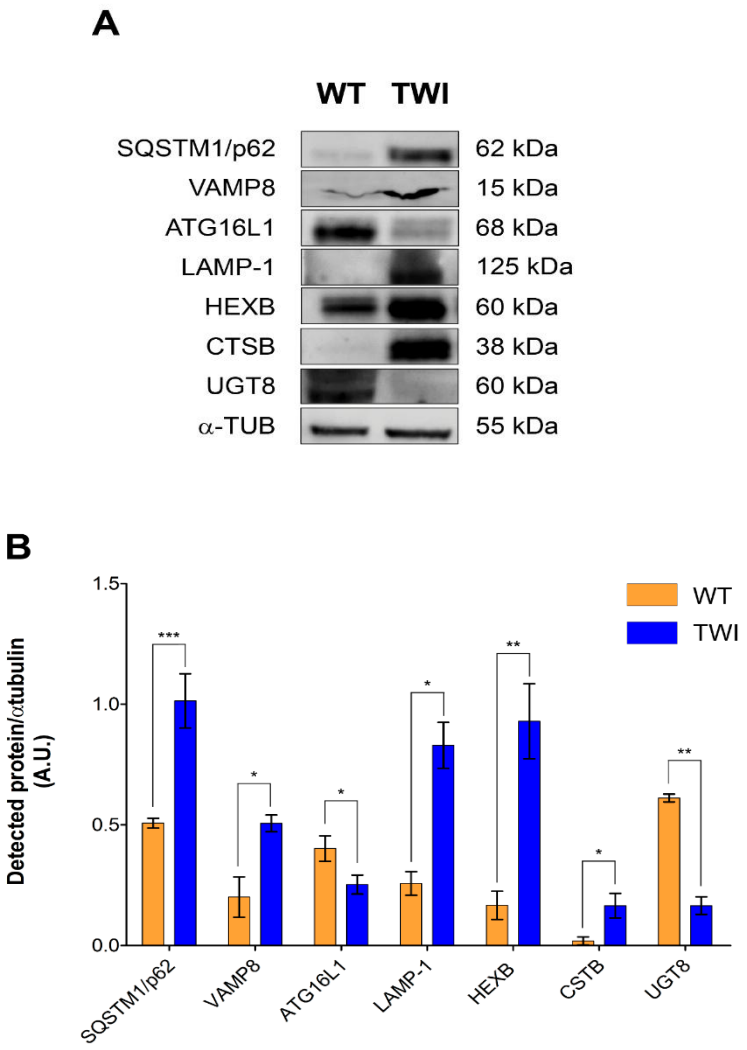


Figure 5.11. *Western blots showing the expression levels of SQSTM1, VAMP8, ATG16L1, LAMP-1, HEXB, CTSB and UGT8 in the sciatic nerve extracts of WT mice compared to TWI mice. Results were normalized to α-tubulin. A WT mice pooled sample and a TWI mice pooled sample were prepared by combining the sciatic nerves extracts of three WT and three TWI mice, respectively. The western blot analysis was performed in technical triplicate on the two pooled samples. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$; Student's t-test. Blue and orange bars indicate TWI and WT mice, respectively. Error bars indicate standard deviation.*

5.5. Discussion

This work represents the first mass spectrometry-based in-depth characterization of the Twitcher mouse proteome to study changes associated with KD. 10-plex TMT experiments were used to compare the proteomes of the corpus callosum, motor cortex and sciatic nerves of five TWI and five WT control mice. LCM was combined with a microproteomics approach [44] to focus the analysis on localized anatomical regions from individual animals; high pH fractionation [64] and multinotch MS3 was used to ensure high proteome coverage and higher relative quantitation precision [65][66], and together enabled the relative quantitation of 3000 to 4000 protein groups from the microdissected samples. Statistical analysis highlighted 181, 46 and 555 protein groups differentially expressed between TWI and WT mice in the corpus callosum, motor cortex and sciatic nerves, respectively. The expression levels of several proteins were further validated by western blot analysis of the sciatic nerve extracts. The roles of the differentially expressed proteins in the central and peripheral nervous system of the Twitcher mouse are discussed below.

5.5.1. Proteome changes in the Peripheral Nervous System of the Twitcher mouse

The analysis of the sciatic nerve extracts showed an activation of inflammatory response in the TWI mouse. Gene ontology analysis, Figure 5.9., revealed an up-regulation of several biological processes and pathways related to immune and defense response, antigen processing and presentation pathway, complement and phagosome pathways, indicating a marked inflammation of the peripheral nervous system.

Thirteen proteins belonging to the complement pathway were up-regulated in the TWI mouse (C1QA, C1QB, C1SA, C3, C5, C5AR1, C8A, C8B, CFB, CD93, ITGAM, ITGAX, VTN). CD55, an inhibitor of complement activation, was also found to be significantly down-regulated in the TWI mouse, indicating an activation of the complement pathway.

Moreover, the up-regulation of several macrophage markers (*e.g.* CSF1R, MPEG1, MSR1) suggests macrophage infiltration of the sciatic nerves. Complement activation has been shown to be involved in the initiation and/or progression of inflammation by attracting macrophages, stimulation of phagocytosis and tissue injury in diseases of both CNS and PNS, including Alzheimer's [67] and Gaucher disease [68]. The observed up-regulation of 21 proteins involved in phagocytosis (*e.g.* MCHI, MCHII, CORO1A, CALR and FCGR1) may be an effect of complement activation.

Neuroinflammation is a well-known feature of KD that was believed to be a consequence of demyelination [69]. Recent studies showed a neuroimmune activation in murine models of KD several weeks before symptoms onset, suggesting that neuroinflammation precedes demyelination [70]. The finding that complement proteins are up-regulated in TWI mice may be of interest since complement activation has been shown to be involved in progressive demyelination of the PNS in Guillain-Barré syndrome [71] and Miller Fisher syndrome [72], and to induce tissue inflammation in Gaucher disease [68]. In addition, 53 proteins associated with the PNS, axon development and neuron ensheathment (*e.g.* MAG, MPZ, MPZL1, JAM3, NEFL, NCAM2, STXBP1) were found to be down-regulated in the TWI mouse, indicating neuron damage and extensive demyelination.

Our data show a marked neuroinflammation, demyelination and complement up-regulation at the terminal stage of the disease (the mice were sacrificed at postnatal day 30 and the mean lifespan of the TWI mouse is 35-40 days). Future work will focus on younger TWI mice to address the role of complement activation in the pathogenesis of KD.

The lysosome pathway was found to be up-regulated in the TWI mouse (Figure 5.9.). Defective lysosomal function is a general hallmark of LSD's and leads to an accumulation of nondegraded macromolecules and metabolites [58]. We found 40 lysosomal proteins up-regulated in the TWI mouse (Figure 5.10.), including lysosome membrane proteins (*e.g.* LAMP1 and VAMP8), hydrolases (*e.g.* CTSB, CTSK, CTSH) and vacuolar H+-ATPases (ATP6V1C1). A subset of lysosomal proteins associated with glycosaminoglycan (GAG) degradation were also up-regulated in the TWI mice, specifically five hydrolases (GUSB, GLB1, HEXA, HEXB and GNS) and a transferase (HGSNAT), which are involved in the degradation of hyaluronan, heparan sulfate, chondroitin sulfate and keratan

sulfate [73]. GAGs are important constituents of the cell membrane and the extracellular matrix and play important roles not only in cell adhesion but also in inflammation, neurodevelopment and neuropathology [74]. GAGs have been shown to be involved in leukocyte transmigration at inflammatory sites and in the modulation of chemokines [75]. Accumulation and impaired degradation of specific GAGs in mucopolysaccharidosis disorder leads to cell death and a chronic inflammatory response resulting in neurodegeneration [76]. Our data indicate altered GAG degradation in TWI mice, which has not been reported previously. Further studies are needed to elucidate the potential role of GAG dysregulation in KD pathogenesis and/or progression.

The up-regulation of hydrolases suggests inefficient lysosomal degradation, while the up-regulation of lysosomal membrane proteins (e.g. LAMP1) may indicate a larger number of lysosomes in the TWI mouse [77]. An accumulation of lysosomes was already observed in induced neurons derived from an adult onset KD patient [78]. The accumulation of lysosomes and lysosomal proteins may indicate a disruption of the autophagy pathway, which has previously been shown in several LSD's [79][59]. Autophagy was shown to be active in a cell line model of KD after psychosine administration, and the treatment with Lithium (an autophagy stimulator) improved cell viability [80]. Here ATG16L1 was found to be down-regulated in the TWI mouse. ATG16L1 belongs to the autophagy related protein family and forms part of a large protein complex that is necessary for autophagy [81]. Specifically, ATG16L1 forms a complex with ATG12 and ATG5 that localizes to phagophores and pre-phagophore structures [82] and is essential for the proper elongation of the nascent autophagosome and for the lipidation of LC3 [83][84]. Down-regulation of ATG16L1 may lead to a reduced rate of autophagosome assembly, and thus a reduced rate of autophagy [85]. A deregulation of the autophagic flux in TWI mice also explains the observed up-regulation of p62. The ubiquitin-binding protein p62 can bind to LC3 on the autophagosome membrane and be degraded by autophagy or target other proteins for degradation[86]. Accumulation of p62 and formation of p62 aggregates have been observed to occur upon autophagy inhibition and in several LSDs characterized by a disruption of the autophagy pathway, including mucopolysaccharidoses, mucolipidoses, Niemann-Pick C1, Pompe, Gaucher and Fabry diseases [79]. Thus, the accumulation of p62 is consistent with a reduced autophagic activity due to down-regulation of ATG16L1.

Lysosomal protein accumulation could also be due to decreased autophagosome-lysosome fusion. A significant down-regulation of 16 proteins associated with microtubule cytoskeleton organization (*e.g.* MAP6 MAP7D2, CHMP3, CLASP2) were observed in TWI mice (Figure 5.9.C), MAP6 and MAPK8IP3 have already been shown to be involved in the active transport of lysosomes [87][88]. The two kinesin motor proteins (KIF5C and KIF21A), which play an active role in the anterograde transport of lysosomes [89], were also found to be down regulated in TWI mice. Moreover we observed a down-regulation of CHMP3, which is a protein that forms part of the ESCRT-III complex whose inactivation has been shown to induce autophagosome accumulation and autophagy disruption [90].

An accumulation of lysosomal proteins may also cause a destabilization of the lysosomal membrane that leads to lysosome membrane permeabilization (LMP) [91]. LMP have been shown to play a crucial role in lysosomal cell death (LCD) because of the release of lysosomal proteins in the cytosol. Cathepsins are believed to be the main mediators of LCD, even if they can also trigger LCD [92]. Among the lysosomal proteins that were found to be up-regulated in TWI mice there were 6 cathepsins: CTSB, CTSD, CTSH, CTSK, CTSS, CTSZ. In particular, CTSB and CTSD are involved not only in the triggering of LMP but also in cell death since they remain active at neutral pH, and so they remain active in the cytosol [93]. LMP may also be induced by reactive oxygen species (ROS) and apolipoproteins [91][94]. ROS production in Krabbe disease has been linked to psychosine accumulation, which activates secretory phospholipase A2 in oligodendrocytes MO3.13 [95]. Here we found an up-regulation in TWI mice of PLAA (an activator of phospholipase A2) and 5 apolipoproteins (APOA1, APOB, APOBR, APOD, APOE), supporting the hypothesis of LMP in the TWI mouse. APOE is a high density lipoprotein produced mostly by astrocytes in the CNS and secreted by macrophages at the site of injury in the PNS [96][97]. Accumulation of APOE has been shown to delay the regeneration of sciatic nerves, contributing to degeneration of the nervous system [97]. Moreover, APOE seems to impair autophagy in astrocytes leading to a reduced capacity of Aβ plaque clearance in the CNS of Alzheimer's disease patients [98]. Apolipoproteins have not previously been reported to accumulate in the TWI mouse and future studies will be needed to establish their role in KD.

Among the proteins down-regulated in the TWI mouse we also found UGT8, which is the enzyme that catalyzes the synthesis of psychosine. This result is in agreement with previous studies that reported a down-regulation of UGT8 in the brain and spinal cord of the TWI mouse at a late stage of the disease [99][100][61]. The down-regulation of UGT8 in TWI mice is thought to be a homeostatic response of the cell to psychosine accumulation.

5.5.2. Proteome changes in the Central Nervous System of the Twitcher mouse

We characterized the Central Nervous System of the TWI mouse by analyzing the corpus callosum and motor cortex regions.

The corpus callosum dataset contained a larger number of deregulated proteins than the motor cortex, consistent with the known mechanisms of KD in which the accumulation of psychosine occurs mainly in myelinating cells (such as oligodendrocytes and Schwann cells), impairing remyelination and damaging the brain's white matter [2].

The main biological processes up-regulated in the CNS of the TWI mouse are related to inflammatory and defense response and leukocyte infiltration (Figure 5.9.A-B), as we found in the PNS. We found an up-regulation of the Complement and JAK/STAT pathways in both the corpus callosum and motor cortex, indicating neuroinflammation in the TWI mouse brain. The JAK/STAT pathway is essential for both innate and adaptive immunity and its aberrant activation has previously been reported in the neuroinflammatory diseases Multiple Sclerosis and Parkinson's Disease [101]. The use of JAK1 and JAK2 inhibitors in several murine models of Multiple Sclerosis has been shown to suppress clinical symptoms, reduce demyelination and suppress the production of pro inflammatory cytokines and chemokines [102]. The up-regulation of both STAT1 and STAT3 in the CNS of the TWI mouse suggests that the JAK/STAT pathway may be a therapeutic target for KD treatment.

We also found an up-regulation in the TWI mouse of GFAP and VIM, which are markers of reactive astrocytes and microglia activation [103][104]. GFAP and VIM up-regulation in the TWI mouse have been previously shown to occur before demyelination and in concomitance with macrophage infiltration, suggesting that CNS astrogliosis and inflammation may precede demyelination [105][106].

Several lysosomal proteins were found up-regulated in the corpus callosum of the TWI mouse (Figure 5.10.A-B), including HEXB, CTSB, CTSD and LAMP2. These results are in agreement with that found for the PNS of the TWI mouse, indicating that the accumulation of nondegraded material and the related biological alterations (see section "*Proteome changes in the Peripheral Nervous System of the Twitcher mouse")* also occurs in the CNS of the TWI mouse.

Five proteins involved in membrane raft organization (ANXA2, CAV1, FLOT1 and DOCK2) were found to be up-regulated in the corpus callosum of the TWI mouse. Disruption of lipid raft domains have been shown to occur in sphingolipid storage disorders and KD as a consequence of undigested lipids and psychosine accumulation [22][107]. Lipid rafts are involved in several signaling pathways and an alteration in raft composition can lead to altered membrane fluidity, and deregulation of cell signaling, influencing survival signals such as PKC, Akt and ERK [23][24].

Among the proteins down-regulated in the motor cortex of the TWI mouse we found UGT8, and which was also found to be down regulated in the sciatic nerve (see section "*Proteome changes in the Peripheral Nervous System of the Twitcher mouse"*). This result is in agreement with previous studies, which reported a down-regulation of UGT8 at late stages of KD in both CNS and PNS [99].

5.6. Conclusions

This work represents the first in-depth characterization of the Twitcher mouse proteome to study Krabbe disease. We applied a quantitative microproteomic workflow to detect changes associated with Krabbe disease in the Central and Peripheral Nervous System of the Twitcher mouse. The proteomes extracted from the corpus callosum, motor cortex and sciatic nerves of five Twitcher and five wild-type mice were compared at postnatal day 30. More than 3300 protein groups were identified for each dataset. Statistical analysis revealed a total number of deregulated protein groups of 181, 46 and 554 for the corpus callosum, motor cortex and sciatic nerves datasets, respectively.

Most of the enriched biological processes and pathways in the TWI mice were related to neuroinflammation, immune response, accumulation of lysosomal proteins, demyelination, membrane raft organization and reduced nervous system development. These results on the proteome changes in the Twitcher mouse help provide new insights into the molecular mechanisms of Krabbe disease. Future experiments will focus on the proteomic characterization of younger mice (pre-symptomatic and at symptoms onset) in order to understand how these deregulated pathways are linked to the pathogenesis of Krabbe disease and specifically, which of these trigger damage to the CNS and PNS and can act targets for new therapies.

References

1. Wenger, D.A., Escolar, M.L., Luzi, P., and Rafi, M.A. (2001) Krabbe disease (Globoid Cell Leukodystrophy), in *The Online Metabolic and Molecular Bases of Inherited Disease*, 8ed., McGraw-Hill, New York.

2. Graziano, A.C.E., and Cardile, V. (2015) History, genetic, and recent advances on Krabbe disease. *Gene*, **555** (1), 2–13.

3. Duffner, P.K., Barczykowski, A., Kay, D.M., Jalal, K., Yan, L., Abdelhalim, A., Gill, S., Gill, A.L., and Carter, R. (2012) Later onset phenotypes of Krabbe disease: Results of the world-wide registry. *Pediatr. Neurol.*, **46** (5), 298–306.

4. Barczykowski, A.L., Foss, A.H., Duffner, P.K., Yan, L., and Carter, R.L. (2012) Death rates in the U.S. due to Krabbe disease and related leukodystrophy and lysosomal storage diseases. *Am. J. Med. Genet. Part A*, **158 A** (11), 2835–2842.

5. Hagberg, B., Sourander, P., and Svennerholm, L. (1963) Diagnosis of Krabbe's infantile leucodystrophy. *J. Neurol. Neurosurg. Psychiatry*, **26** (3), 195–198.

6. Duffner, P.K., Jalal, K., and Carter, R.L. (2009) The Hunter's Hope Krabbe Family Database. *Pediatr. Neurol.*, **40** (1), 13–18.

7. Kolodny, E.H., Raghavan, S., and Krivit, W. (1991) Late-Onset Krabbe Disease (Globoid Cell Leukodystrophy): Clinical and Biochemical Features of 15 Cases. *Dev. Neurosci.*, **13**, 232–239.

8. Nicaise, A.M., Bongarzone, E.R., Crocker, S.J., and Biology, C. (2017) HHS Public Access. **94** (11), 1049–1061.

9. Rafi, M.A., Rao, H.Z., Luzi, P., Curtis, M.T., and Wenger, D.A. (2012) Extended normal life after AAVrh10-mediated gene therapy in the mouse model of krabbe disease. *Mol. Ther.*, **20** (11), 2031–2042.

10. Lattanzi, A., Salvagno, C., Maderna, C., Benedicenti, F., Morena, F., Kulik, W., Naldini, L., Montini, E., Martino, S., and Gritti, A. (2014) Therapeutic benefit of

lentiviral-mediated neonatal intracerebral gene therapy in a mouse model of globoid cell leukodystrophy. *Hum. Mol. Genet.*, **23** (12), 3250–3268.

11. Biswas, S., Biesiada, H., Williams, T.D., and LeVine, S.M. (2003) Substrate reduction intervention by L-cycloserine in twitcher mice (globoid cell leukodystrophy) on a B6;CAST/Ei background. *Neurosci. Lett.*, **347** (1), 33–36.

12. Berardi, A.S., Pannuzzo, G., Graziano, A., Costantino-Ceccarini, E., Piomboni, P., and Luddi, A. (2014) Pharmacological chaperones increase residual β-galactocerebrosidase activity in fibroblasts from krabbe patients. *Mol. Genet. Metab.*, **112** (4), 294–301.

13. Lee, W.C., Tsoi, Y.K., Troendle, F.J., DeLucia, M.W., Ahmed, Z., Dicky, C.A., Dickson, D.W., and Eckman, C.B. (2007) Single-dose intracerebroventricular administration of galactocerebrosidase improves survival in a mouse model of globoid cell leukodystrophy. *FASEB J.*, **21** (10), 2520–2527.

14. Qin, E.Y., Hawkins-Salsbury, J.A., Jiang, X., Reddy, A.S., Farber, N.B., Ory, D.S., and Sands, M.S. (2012) Bone marrow transplantation increases efficacy of central nervous system-directed enzyme replacement therapy in the murine model of globoid cell leukodystrophy. *Mol. Genet. Metab.*, **107** (1–2), 186–196.

15. Kobayashi, T., Yamanaka, T., Jacobs, J.M., Teixeira, F., and Suzuki, K. (1980) The twitcher mouse: an enzymatically authentic model of human globoid cell leukodystrophy (Krabbe disease). *Brain Res.*, **202** (2), 479–483.

16. Miyatake, T., and Suzuki, K. (1972) Globoid cell leukodystrophy: Additional deficiency of psychosine galactosidase. *Biochem. Biophys. Res. Commun.*, **48** (3), 538–543.

17. Suzuki, K. (2003) Globoid Cell Leukodystrophy (Krabbe's Disease): Update. *J. Child Neurol.*, **18** (9), 595–603.

18. Kobayashi, T., Shinnoh, N., Goto, I., and Kuroiowa, Y. (1985) Hydrolysis of Galactosylceramide Is Catalyzed by Two Genetically Distinct Acid β-Galactosidases. *J. Biol. Chem.*, **260** (28), 14982–14987.

19. Suzuki, K. (1998) Twenty five years of the "psychosine hypothesis": a personal perspective of its history and present status. *Neurochem. Res.*, **23** (3), 251–259.

20. Voccoli, V., Tonazzini, I., Signore, G., Caleo, M., and Cecchini, M. (2014) Role of extracellular calcium and mitochondrial oxygen species in psychosine-induced oligodendrocyte cell death. *Cell Death Dis.*, **5**, e1529.

21. Suzuki, K. (2003) Evolving perspective of the pathogenesis of globoid cell leukodystrophy (Krabbe disease). *Proc. Japan Acad. Ser. B*, **79B** (1), 1–8.

22. White, A.B., Givogri, M.I., Lopez-Rosas, A., Cao, H., van Breemen, R., Thinakaran, G., and Bongarzone, E.R. (2009) Psychosine Accumulates in Membrane Microdomains in the Brain of Krabbe Patients, Disrupting the Raft Architecture. *J. Neurosci.*, **29** (19), 6068–6077.

23. Auria, L.D., Reiter, C., Ward, E., Moyano, A.L., Marshall, M.S., Nguyen, D., Scesa, G., Hauck, Z., Breemen, R. Van, Givogri, M.I., and Bongarzone, E.R. (2017) Psychosine enhances the shedding of membrane microvesicles : Implications in demyelination in Krabbe ' s disease. 1–19.

24. Sural-fehr, T., and Bongarzone, E.R. (2016) How membrane dysfunction influences neuronal survival pathways in sphingolipid storage disorders. *J. Neurosci. Res.*, **94** (11), 1042–1048.

25. W Esch, S., Williams, T., Biswas, S., Chakrabarty, A., and M LeVine, S. (2003) *Sphingolipid profile in the CNS of the twitcher (globoid cell leukodystrophy) mouse: A lipidomics approach*.

26. Assunta, Z., Elena, D., Anna, B., Paola, P., Elvira, C., and Alice, L. (2014) GC- EI- MS Analysis of Fatty Acid Composition in Brain and Serum of Twitcher Mouse. *Lipids*, **49** (11), 1115–1125.

27. Meisingset, T.W., Ricca, A., Neri, M., Sonnewald, U., and Gritti, A. (2013) Region- and Age-Dependent Alterations of Glial-Neuronal Metabolic Interactions Correlate with CNS Pathology in a Mouse Model of Globoid Cell Leukodystrophy. *J. Cereb. Blood Flow Metab.*, **33** (7), 1127–1137.

28. I., W.N., Lawrence, W., Laura, F.M., and Daesung, S. (2016) Metabolic profiling reveals biochemical pathways and potential biomarkers associated with the pathogenesis of Krabbe disease. *J. Neurosci. Res.*, **94** (11), 1094–1107.

29. Cologna, S.M., Jiang, X.-S., Backlund, P.S., Cluzeau, C.V.M., Dail, M.K., Yanjanin, N.M., Siebel, S., Toth, C.L., Jun, H., Wassif, C.A., Yergey, A.L., and Porter, F.D. (2012) Quantitative Proteomic Analysis of Niemann-Pick Disease, Type C1 Cerebellum Identifies Protein Biomarkers and Provides Pathological Insight. *PLoS One*, **7** (10), e47845.

30. Sleat, D.E., Wiseman, J.A., Sohar, I., El-Banna, M., Zheng, H., Moore, D.F., and Lobel, P. (2012) Proteomic analysis of mouse models of Niemann-Pick C disease reveals alterations in the steady-state levels of lysosomal proteins within the brain. *Proteomics*, **12** (23–24), 3499–3509.

31. Rauniyar, N., Subramanian, K., Lavallée-Adam, M., Martínez-Bartolomé, S., Balch, W.E., and Yates, J.R. (2015) Quantitative Proteomics of Human Fibroblasts with I1061T Mutation in Niemann–Pick C1 (NPC1) Protein Provides Insights into the Disease Pathogenesis. *Mol. Cell. Proteomics* , **14** (7), 1734–1749.

32. Jones, E.E., Zhang, W., Zhao, X., Quiason, C., Dale, S., Shahidi-Latham, S., Grabowski, G.A., Setchell, K.D.R., Drake, R.R., and Sun, Y. (2017) Tissue Localization of Glycosphingolipid Accumulation in a Gaucher Disease Mouse Brain by LC-ESI-MS/MS and High-Resolution MALDI Imaging Mass Spectrometry. *SLAS Discov. Adv. Life Sci. R&D*, **22** (10), 1218–1228.

33. Vissers, J.P.C., Langridge, J.I., and Aerts, J.M.F.G. (2007) Analysis and Quantification of Diagnostic Serum Markers and Protein Signatures for Gaucher Disease. *Mol. Cell. Proteomics* , **6** (5), 755–766.

34. Moore, D.F., Krokhin, O. V, Beavis, R.C., Ries, M., Robinson, C., Goldin, E., Brady, R.O., Wilkins, J.A., and Schiffmann, R. (2007) Proteomics of specific treatment-related alterations in Fabry disease: A strategy to identify biological abnormalities. *Proc. Natl. Acad. Sci.*, **104** (8), 2873–2878.

35. Manwaring, V., Heywood, W.E., Clayton, R., Lachmann, R.H., Keutzer, J., Hindmarsh, P., Winchester, B., Heales, S., and Mills, K. (2013) The Identification of New Biomarkers for Identifying and Monitoring Kidney Disease and Their Translation into a Rapid Mass Spectrometry-Based Test: Evidence of Presymptomatic Kidney Disease in Pediatric Fabry and Type-I Diabetic Patients. *J. Proteome Res.*, **12** (5), 2013–2021.

36. Whitfield, P.D., Sharp, P.C., Taylor, R., and Meikle, P. (2001) Quantification of galactosylsphingosine in the twitcher mouse using electrospray ionization-tandem mass spectrometry. *J. Lipid Res.* , **42** (12), 2092–2095.

37. Zanfini, A., Dreassi, E., Berardi, A., Governini, L., Corbini, G., Costantino-Ceccarini, E., Balestri, P., and Luddi, A. (2013) Quantification of psychosine in the serum of twitcher mouse by LC–ESI-tandem-MS analysis. *J. Pharm. Biomed. Anal.*, **80**, 44–49.

38. Escolar, M.L., Kiely, B.T., Shawgo, E., Hong, X., Gelb, M.H., Orsini, J.J., Matern, D., and Poe, M.D. (2017) Psychosine, a marker of Krabbe phenotype and treatment effect. *Mol. Genet. Metab.*, **121** (3), 271–278.

39. Ribbens, J.J., Moser, A.B., Hubbard, W.C., Bongarzone, E.R., and Maegawa, G.H.B. (2018) Characterization and application of a disease-cell model for a neurodegenerative lysosomal disease. *Mol. Genet. Metab.* , **111** (2), 172–183.

40. Cappello, V., Marchetti, L., Parlanti, P., Landi, S., Tonazzini, I., Cecchini, M., Piazza, V., and Gemmi, M. (2016) Ultrastructural Characterization of the Lower Motor System in a Mouse Model of Krabbe Disease. *Sci. Rep.*, **6** (1), 1.

41. Sakai, N., Inui, K., Tatsumi, N., Fukushima, H., Nishigaki, T., Taniike, M., Nishimoto, J., Tsukamoto, H., Yanagihara, I., Ozono, K., and Okada, S. (1996) Molecular Cloning and Expression of cDNA for Murine Galactocerebrosidase and Mutation Analysis of the Twitcher Mouse, a Model of Krabbe's Disease. *J. Neurochem.*, **66** (3), 1118–1124.

42. Soueid, J., Nokkari, A., and Makoukji, J. (2015) Techniques and Methods of

Animal Brain Surgery, in *Brain Neurotrauma: Molecular, Neuropsychological, and Rehabilitation Aspects.*, 1sted., CRC Press/Taylor & Francis, Boca Raton (FL).

43. Bala, U., Tan, K.-L., Ling, K.-H., and Cheah, P.-S. (2014) Harvesting the maximum length of sciatic nerve from adult mice: a step-by-step approach. *BMC Res. Notes*, **7** (1), 714.

44. de Graaf, E.L., Pellegrini, D., and McDonnell, L.A. (2016) Set of Novel Automated Quantitative Microproteomics Protocols for Small Sample Amounts and Its Application to Kidney Tissue Substructures. *J. Proteome Res.*, **15** (12), 4722–4730.

45. Hughes, C.S., Foehr, S., Garfield, D.A., Furlong, E.E., and Steinmetz, L.M. (2014) Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.*, **10** (757), 1–15.

46. Eng, J.K., McCormack, A.L., and Yates, J.R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.*, **5** (11), 976–989.

47. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods*, **13** (9), 731–740.

48. Wang, J., Vasaikar, S., Shi, Z., Greer, M., and Zhang, B. (2017) WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.*, **45** (W1), 130–137.

49. Tonazzini, I., Meucci, S., Van Woerden G, M., Elgersma, Y., and Cecchini, M. (2016) Impaired Neurite Contact Guidance in Ubiquitin Ligase E3a (Ube3a) Deficient Hippocampal Neurons on Nanostructured Substrates. *Adv. Healthc. Mater.*, **5** (7), 850–862.

50. Vizcaíno, J.A., Csordas, A., del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q.-W., Wang, R., and Hermjakob, H. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.*, **44** (D1), D447–D456.

51. Nathan, F., Allen, A., Neville, I., Charbel, S., and Lena, N. (2016) Magnetic resonance imaging findings of central nervous system in lysosomal storage diseases: A pictorial review. *J. Med. Imaging Radiat. Oncol.*, **61** (3), 344–352.

52. Poretti, A., Meoded, A., Bunge, M., Fatemi, A., Barrette, P., Huisman, T.A.G.M., and Salman, M.S. (2018) Novel diffusion tensor imaging findings in Krabbe disease. *Eur. J. Paediatr. Neurol.*, **18** (2), 150–156.

53. Abdelhalim, A.N., Alberico, R.A., Barczykowski, A.L., and Duffner, P.K. (2018) Patterns of Magnetic Resonance Imaging Abnormalities in Symptomatic Patients With Krabbe Disease Correspond to Phenotype. *Pediatr. Neurol.*, **50** (2), 127–134.

54. Zarifi, M.K., Tzika, A.A., Astrakas, L.G., Young Poussaint, T., Anthony, D.C., and Darras, B.T. (2001) Magnetic Resonance Spectroscopy and Magnetic Resonance Imaging Findings in Krabbe's Disease. *J. Child Neurol.*, **16** (7), 522–526.

55. Dolcetta, D., Amadio, S., Guerrini, U., Givogri, M.I., Perani, L., Galbiati, F., Sironi, L., Del Carro, U., Roncarolo, M.C., and Bongarzone, E. (2005) Myelin deterioration in Twitcher mice: Motor evoked potentials and magnetic resonance imaging as in vivo monitoring tools. *J. Neurosci. Res.*, **81** (4), 597–604.

56. Vito, G., Cappello, V., Tonazzini, I., Cecchini, M., and Piazza, V. (2016) RP CARS reveals molecular spatial order anomalies in myelin of an animal model of Krabbe disease. *J. Biophotonics*, **10** (3), 385–393.

57. Teixeira, C.A., Miranda, C.O., Sousa, V.F., Santos, T.E., Malheiro, A.R., Solomon, M., Maegawa, G.H., Brites, P., and Sousa, M.M. (2014) Early axonal loss accompanied by impaired endocytosis, abnormal axonal transport, and decreased microtubule stability occur in the model of Krabbe's disease. *Neurobiol. Dis.*, **66**, 92–103.

58. Platt, F.M., Boland, B., and van der Spoel, A.C. (2012) Lysosomal storage disorders: The cellular impact of lysosomal dysfunction. *J. Cell Biol.*, **199** (5), 723 LP – 734.

59. Sarkar, S., Carroll, B., Buganim, Y., Maetzel, D., Ng, A.H.M., Cassady, J.P.,

Cohen, M.A., Chakraborty, S., Wang, H., Spooner, E., Ploegh, H., Gsponer, J., Korolchuk, V.I., and Jaenisch, R. (2013) Impaired Autophagy in the Lipid-Storage Disorder Niemann-Pick Type C1 Disease. *Cell Rep.*, **5** (5), 1302–1315.

60. Rogov, V., Dötsch, V., Johansen, T., and Kirkin, V. (2014) Interactions between Autophagy Receptors and Ubiquitin-like Proteins Form the Molecular Basis for Selective Autophagy. *Mol. Cell*, **53** (2), 167–178.

61. Kunihiko, S., and Kinuko, S. (2018) The Twitcher Mouse: A Model for Krabbe Disease and for Experimental Therapies. *Brain Pathol.*, **5** (3), 249–258.

62. Futerman, A.H., and van Meer, G. (2004) The cell biology of lysosomal storage disorders. *Nat. Rev. Mol. Cell Biol.*, **5**, 554.

63. Liu, W.J., Ye, L., Huang, W.F., Guo, L.J., Xu, Z.G., Wu, H.L., Yang, C., and Liu, H.F. (2016) p62 links the autophagy pathway and the ubiqutin–proteasome system upon ubiquitinated protein degradation. *Cell. Mol. Biol. Lett.*, **21** (1), 29.

64. Manadas, B., Mendes, V.M., English, J., and Dunn, M.J. (2010) Peptide fractionation in proteomics approaches. *Expert Rev. Proteomics*, **7** (5), 655–663.

65. Altelaar, a. F.M., Frese, C.K., Preisinger, C., Hennrich, M.L., Schram, A.W., Timmers, H.T.M., Heck, A.J.R., and Mohammed, S. (2012) Benchmarking stable isotope labeling based quantitative proteomics. *J. Proteomics*, **88**, 14–26.

66. McAlister, G.C., Nusinow, D.P., Jedrychowski, M.P., Wühr, M., Huttlin, E.L., Erickson, B.K., Rad, R., Haas, W., and Gygi, S.P. (2014) MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.*, **86** (14), 7150–7158.

67. Crehan, H., Hardy, J., and Pocock, J. (2012) Microglia, Alzheimer's Disease, and Complement. *Int. J. Alzheimers. Dis.*, **2012**, 1–10.

68. Pandey, M.K., Burrow, T.A., Rani, R., Martin, L.J., Witte, D., Setchell, K.D., Mckay, M.A., Magnusen, A.F., Zhang, W., Liou, B., Köhl, J., and Grabowski, G.A. (2017) Complement drives glucosylceramide accumulation and tissue inflammation

in Gaucher disease. *Nature*, **543**, 108.

69. B., P.G., and A., P.M. (2016) Neuroimmune mechanisms in Krabbe's disease. *J. Neurosci. Res.*, **94** (11), 1341–1348.

70. Snook, E.R., Fisher-Perkins, J.M., Sansing, H.A., Lee, K.M., Alvarez, X., MacLean, A.G., Peterson, K.E., Lackner, A.A., and Bunnell, B.A. (2014) Innate Immune Activation in the Pathogenesis of a Murine Model of Globoid Cell Leukodystrophy. *Am. J. Pathol.*, **184** (2), 382–396.

71. Kaida, K., and Kusunoki, S. (2009) Guillain–Barré syndrome: update on immunobiology and treatment. *Expert Rev. Neurother.*, **9** (9), 1307–1319.

72. K., H.S., D., H.P., A., G.J., R., W.E., G., S.R.A., and J., W.H. (2005) Complement inhibition abrogates nerve terminal injury in Miller Fisher syndrome. *Ann. Neurol.*, **58** (2), 203–210.

73. Freeze, H.H. (2009) Genetic Disorders of Glycan Degradation, in *Essentials of Glycobiology*, 3rded., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), pp. 1–23.

74. Smith, P.D., Coulson-Thomas, V.J., Foscarin, S., Kwok, J.C.F., and Fawcett, J.W. (2015) "GAG-ing with the neuron": The role of glycosaminoglycan patterning in the central nervous system. *Exp. Neurol.*, **274**, 100–114.

75. Parish, C.R. (2006) The role of heparan sulphate in inflammation. *Nat. Rev. Immunol.*, **6**, 633.

76. Archer, L.D., Langford-Smith, K.J., Bigger, B.W., and Fildes, J.E. (2014) Mucopolysaccharide diseases: A complex interplay between neuroinflammation, microglial activation and adaptive immunity. *J. Inherit. Metab. Dis.*, **37** (1), 1–12.

77. Gowrishankar, S., Yuan, P., Wu, Y., Schrag, M., Paradise, S., Grutzendler, J., De Camilli, P., and Ferguson, S.M. (2015) Massive accumulation of luminal protease-deficient axonal lysosomes at Alzheimer's disease amyloid plaques. *Proc. Natl. Acad. Sci.*, **112** (28), E3699–E3708.

78. Lim, S.M., Choi, B.-O., Oh, S., Choi, W.J., Oh, K.-W., Nahm, M., Xue, Y., Choi, J.H., Choi, J.Y., Kim, Y.-E., Chung, K.W., Fu, X.-D., Ki, C.-S., and Kim, S.H. (2016) Patient fibroblasts-derived induced neurons demonstrate autonomous neuronal defects in adult-onset Krabbe disease. *Oncotarget*, **7** (46).

79. Lieberman, A.P., Puertollano, R., Raben, N., Slaugenhaupt, S., Walkley, S.U., and Ballabio, A. (2012) Autophagy in lysosomal storage disorders. *Autophagy*, **8** (5), 719–730.

80. Del Grosso, A., Antonini, S., Angella, L., Tonazzini, I., Signore, G., and Cecchini, M. (2016) Lithium Improves Cell Viability in Autophagy Activation. *J. Neurosci. Res.*, **1260** (8), 1246–1260.

81. Behrends, C., Sowa, M.E., Gygi, S.P., and Harper, J.W. (2010) Network organization of the human autophagy system. *Nature*, **466**, 68.

82. Rubinsztein, D.C., Shpilka, T., and Elazar, Z. (2012) Mechanisms of Autophagosome Biogenesis. *Curr. Biol.*, **22** (1), R29–R34.

83. Mizushima, N., Yoshimori, T., and Ohsumi, Y. (2011) The Role of Atg Proteins in Autophagosome Formation. *Annu. Rev. Cell Dev. Biol.*, **27** (1), 107–132.

84. Cadwell, K., Liu, J.Y., Brown, S.L., Miyoshi, H., Loh, J., Lennerz, J.K., Kishi, C., Kc, W., Carrero, J.A., Hunt, S., Stone, C.D., Brunt, E.M., Xavier, R.J., Sleckman, B.P., Li, E., Mizushima, N., Stappenbeck, T.S., and Virgin IV, H.W. (2008) A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells. *Nature*, **456**, 259.

85. Nakamura, S., and Yoshimori, T. (2017) New insights into autophagosome–lysosome fusion. *J. Cell Sci.*, **130** (7), 1209–1216.

86. Watanabe, Y., and Tanaka, M. (2011) p62/SQSTM1 in autophagic clearance of a non-ubiquitylated substrate. *J. Cell Sci.*, **124** (16), 2692–2701.

87. Schwenk, B.M., Lang, C.M., Hogl, S., Tahirovic, S., Orozco, D., Rentzsch, K., Lichtenthaler, S.F., Hoogenraad, C.C., Capell, A., Haass, C., and Edbauer, D.

(2014) The FTLD risk factor TMEM106B and MAP6 control dendritic trafficking of lysosomes. *EMBO J.*, **33** (5), 450–467.

88. Drerup, C.M., and Nechiporuk, A. V (2013) JNK-Interacting Protein 3 Mediates the Retrograde Transport of Activated c-Jun N-Terminal Kinase and Lysosomes. *PLOS Genet.*, **9** (2), e1003303.

89. Pu, J., Guardia, C.M., Keren-Kaplan, T., and Bonifacino, J.S. (2016) Mechanisms and functions of lysosome positioning. *J. Cell Sci.*, **129** (23), 4329–4339.

90. Lee, J.-A., Beigneux, A., Ahmad, S.T., Young, S.G., and Gao, F.-B. (2007) ESCRT-III Dysfunction Causes Autophagosome Accumulation and Neurodegeneration. *Curr. Biol.*, **17** (18), 1561–1567.

91. Michaela, M., and M., R.K. (2015) Lysosomal proteins in cell death and autophagy. *FEBS J.*, **282** (10), 1858–1870.

92. Aits, S., and Jäättelä, M. (2013) Lysosomal cell death at a glance. *J. Cell Sci.*, **126** (9), 1905–1912.

93. Boya, P., and Kroemer, G. (2008) Lysosomal membrane permeabilization in cell death. *Oncogene*, **27**, 6434.

94. Ji, Z.S., Müllendorff, K., Cheng, I.H., Miranda, R.D., Huang, Y., and Mahley, R.W. (2006) Reactivity of apolipoprotein E4 and amyloid β peptide: Lysosomal stability and neurodegeneration. *J. Biol. Chem.*, **281** (5), 2683–2692.

95. Giri, S., Khan, M., Rattan, R., Singh, I., and Singh, A.K. (2006) Krabbe disease: psychosine-mediated activation of phospholipase A2 in oligodendrocyte cell death. *J. Lipid Res.* , **47** (7), 1478–1492.

96. Masatoshi, T., Rocío, M., Takashi, K., Toshihisa, T., Masayasu, O., Shinji, T., Takashi, M., Ryota, H., and Ramón, C. (2010) Apolipoprotein E and central nervous system disorders: Reviews of clinical findings. *Psychiatry Clin. Neurosci.*, **64** (6), 592–607.

97. Comley, L.H., Fuller, H.R., Wishart, T.M., Mutsaers, C.A., Thomson, D., Wright,

A.K., Ribchester, R.R., Morris, G.E., Parson, S.H., Horsburgh, K., and Gillingwater, T.H. (2011) ApoE isoform-specific regulation of regeneration in the peripheral nervous system. *Hum. Mol. Genet.*, **20** (12), 2406–2421.

98. Simonovitch, S., Schmukler, E., Bespalko, A., Iram, T., Frenkel, D., Holtzman, D.M., Masliah, E., Michaelson, D.M., and Pinkas-Kramarski, R. (2016) Impaired Autophagy in APOE4 Astrocytes. *J. Alzheimer's Dis.*, **51** (3), 915–927.

99. Masako, T., R., M.J., Toshinori, N., Nobuya, F., Brian, P., Kunihiko, S., and Kinuko, S. (1998) Suppressed UDP- galactose: Ceramide galactosyltransferase and myelin protein mRNA in twitcher mouse brain. *J. Neurosci. Res.*, **51** (4), 536–540.

100. Soichi, K., Hideki, I., A., S.D., and Kunihiko, S. (1982) Glycosylceramide Synthesis in the Developing Spinal Cord and Kidney of the Twitcher Mouse, an Enzymatically Authentic Model of Human Krabbe Disease. *J. Neurochem.*, **39** (5), 1314–1318.

101. Yan, Z., Gibson, S.A., Buckley, J.A., Qin, H., and Benveniste, E.N. (2018) Role of the JAK/STAT signaling pathway in regulation of innate immunity in neuroinflammatory diseases. *Clin. Immunol.*, **189**, 4–13.

102. Liu, Y., Holdbrooks, A.T., De Sarno, P., Rowse, A.L., Yanagisawa, L.L., McFarland, B.C., Harrington, L.E., Raman, C., Sabbaj, S., Benveniste, E.N., and Qin, H. (2014) Therapeutic Efficacy of Suppressing the JAK/STAT Pathway in Multiple Models of Experimental Autoimmune Encephalomyelitis. *J. Immunol.*, **192** (1), 59–72.

103. Milos, P., and Michael, N. (2005) Astrocyte activation and reactive gliosis. *Glia*, **50** (4), 427–434.

104. Pekny, M., Wilhelmsson, U., and Pekna, M. (2014) The dual role of astrocyte activation and reactive gliosis. *Neurosci. Lett.*, **565**, 30–38.

105. Kobayashi, S., Chiu, F.C., Katayama, M., Sacchi, R.S., Suzuki, K., and Suzuki, K. (1986) Expression of glial fibrillary acidic protein in the CNS and PNS of murine globoid cell leukodystrophy, the twitcher. *Am. J. Pathol.*, **125** (2), 227–243.

106. Potter, G.B., Santos, M., Davisson, M.T., Rowitch, D.H., Marks, D.L., Bongarzone, E.R., and Petryniak, M.A. (2013) Missense mutation in mouse GALC mimics human gene defect and offers new insights into Krabbe disease. *Hum. Mol. Genet.*, **22** (17), 3397–3414.

107. A.B., W., F., G., M.I., G., A., L.R., X., Q., R., van B., and E.R., B. (2010) Persistence of psychosine in brain lipid rafts is a limiting factor in the therapeutic recovery of a mouse model for Krabbe disease. *J. Neurosci. Res.*, **89** (3), 352–364.

Chapter 6

Outlook

6.1. The future of MS-based proteomics: space and time resolved proteomics

The recent groundbreaking technological improvements in MS-based proteomics have enabled the in-depth proteome analysis from minute sample amounts (< 1µg) of cell culture samples as well as clinical samples. The possibility to perform ultrasensitive analyses has opened new exciting application areas for MS-based proteomics. Proteomics characterization can now be performed in the space or in the time domains. Spatially-resolved proteomics can be used to study localized pathological features or tissue heterogeneity, while time resolved proteomics can now resolve the translational response of a system to external stimuli. The combination of space and time resolved proteomics represents a unique strategy to study the proteome dynamics in specific cell types.

Tissue heterogeneity occurs at the tissue level and also within the same cell type. Different cell populations in the same tissue exhibit a different phenotype because of the distinct regulation of gene expression and the influence of the microenvironment [1]. The ability to investigate the proteome of specific cell populations is expected to will help understand important biological phenomena which are obscured in bulk measurements. Population distributions can mask the presence of small phenotypic subpopulations of cells, as the response of these cells is diluted by the different populations. Different cell populations may not contribute equally to the biology of the system, for instance the presence of a small population of chemoresistant tumor cells may compromise <1% of all tumor cells but can lead to relapse. Tumor heterogeneity plays a crucial role in cancer onset, progression and resistance to therapy [2]. Differences in the microenvironment and proximity to blood vessels, immune cells, and stromal cells mean even cells of the same type are subject to different regulatory signals. Accordingly these cells will differ in their proteomes, and these differences may also have functional consequences [3][4][5]. In order to detect these differences any analysis will need to be performed at the single cell level.

Genomics and transcriptomics are now established technologies for single-cell analysis, because of the possibility to amplify the copy number using PCR and the availability of high sensitivity next-generation sequencing analysis [6][7]. The absence of an analogous

amplification method for proteins means single-cell MS-based proteomics demands extremely high sensitivity and reproducibity. The first single-cell proteomics study reported the proteome profiling of blastomers isolated from *Xenopus laevis* embryos [8][9]. These analyses were enabled by the high protein content of the blastomers (~10 µg), which was compatible with a down-scaled conventional proteomics workflow (Figure 6.1.). A custom-built single-cell CE-µESI platform allowed the identification of almost a thousand protein groups from just 20 nL of injected sample, and the results exhibited significant cellular heterogeneity at the protein level.

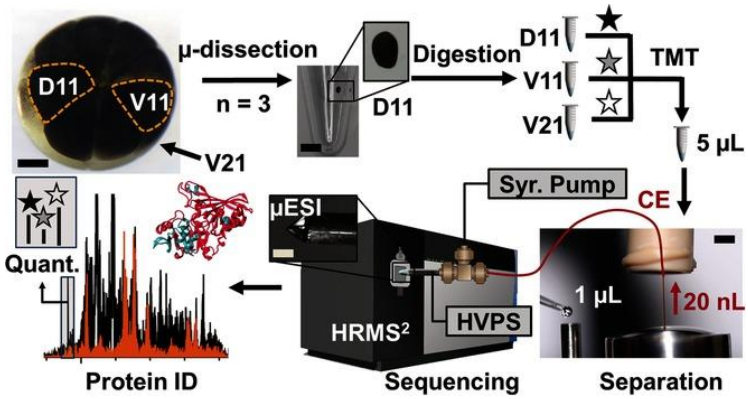


Figure 6.1. *Microanalytical pipeline enabling multiplex proteomics analysis of single cells in the 16-cell Xenopus embryo using microdissection, micro-scale bottom-up proteomics, and a custom-built single-cell CE-µESI platform. Figure adopted from* [8].

In 2018, two different strategies for the proteomics analysis of single mammalian cells were reported [10][11]. The small size of mammalian cells (~10-15 µm diameter) results in a protein content of about 0.1-0.2 ng per cell, which requires not only a highly sensitive instrumental platform but also a miniaturization of the proteomics workflow in order to allow the processing of such minute sample amounts. If the sample amount decreases without a concomitant reduction of the working volume then non-specific adsorption of proteins to the surface of the reaction vessel/pipettes/tubes leads to sample losses and

lower sensitivity. For single cell analysis, such losses quickly lead to the almost complete loss of all available proteins.

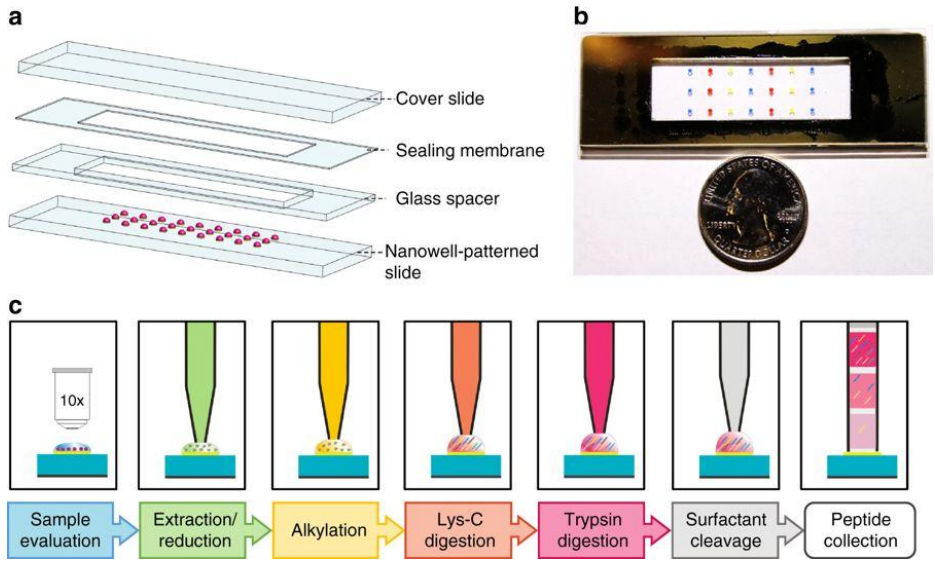


***Figure 6.2**. A) Schematic drawing of the nanoPOTS microfluidic platform structure. B) Photograph showing the nanoPOTS chip with each nanowell filled with 200 nL of colored dye. C) Schematic of the on-chip proteomic sample preparation Figure adopted from* [10].

Zhu et al. developed the nanoPOTS (nanodroplet processing in one pot for trace analysis) platform for the proteomics analysis of single-cells (Figure 6.2.). The key advantage of this platform is the miniaturization of the working volume to about 200 nL, thus minimizing adsorption based sample losses. The nanoPOTS platform consists of a glass slide microfabricated with photolithographically patterned hydrophilic pedestals surrounded by a hydrophobic surface to serve as nanodroplet reaction vessels. The patterned glass slide is assembled with a glass spacer sealed to a membrane-coated glass slide in order to minimize evaporation during the various incubation steps (Figure 6.2.A). A robotic platform is used to dispense cells and reagents with high accuracy into each nanodroplet reaction vessel, and to retrieve samples for subsequent nanoLC-MS/MS analysis. Cells are lysed using an acid-cleavable surfactant (RapiGest) and the proteins reduced, alkylated and

digested with typsin. Peptide TMT labeling can also be easily performed after protein digestion for accurate protein quantification [12]. The solution is then acidified to inactivate the surfactant by precipitation, and the peptide solution retrieved from the nanoPOT well and finally injected into the nanoLC-MS/MS system.

The efficiency of the nanoPOTS protocol was assessed using different amount of HeLa cells (from 139 down to 12 cells). The lysis efficiency, evaluated in terms of the percentage of zero missed cleavages, was found to be higher than 97% and comparable to that obtained in bulk experiments. Pairwise analysis of any two samples with similar sample amounts showed a Pearson correlation coefficient greater than 0.98 at the protein level and a coefficient of variation lower than 13.1%, demonstrating high reproducibility.

The open structure of the nanoPOTS platform is suitable for hyphenation with upstream sample isolation procedures (such laser microdissection [10][13] or FACS [12][14]) and downstream transfer to LC-MS/MS systems. The possibility to integrate nanoPOTS with LCM (Figure 6.3.) makes this technology particularly powerful for tissue proteomics.

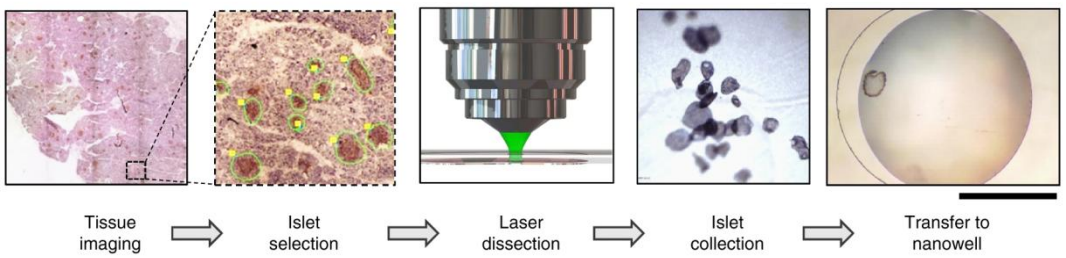


Figure 6.3*. Schematic workflow showing the identification, laser microdissection, collection and transfer of small tissue sections into nanowells. Figure modified from* [10].

The nanoPOTS platform was used to analyze 10-µm-thick cross-sections of individual human islets that were isolated by laser microdissection from clinical pancreatic tissue slices. The proteomic profile of nine single pancreatic islets from non-diabetic donors (control) were compared to nine islets from type 1 diabetic patients (T1D). The highly

sensitive nanoPOTS platform provided the identification of more than 2600 proteins, of which more than 300 exhibited altered expression between T1D patients and controls.

The nanoPOTS platform has proven to be a promising tool for proteomics analysis of single cells but requires access to a nanostructuring facility. The fabrication of the device is relatively straightforward but requires specific expertise and facilities for its realization.

In 2018 Budnik et al. developed Single Cell ProtEomics by Mass Spectromety (ScoPE-MS) to study the proteome of single mammalian cells [11]. The major limitation of single cell proteomics is the low abundance of peptides that can be retrieved from single cells. If a peptide's abundance is too low, the ion current will not be sufficient and the peptides cannot be identified. SCoPE-MS increases the effective abundance of peptides analyzed by the mass spectrometer by multiplexing the analysis. Specifically Tandem Mass Tags (TMT) are used to label peptides from single cells and a so-called carrier sample (see Paragraph 1.2.5.2. for more details about the TMT labeling mechanism). 10-plex TMT allows the simultaneous analysis of up to ten different samples, as ten tags are available for the peptide labeling. In SCoPE-MS eight single cell samples are analyzed as well as a carrier sample containing 200 cells to increase the available peptide ion current (Figure 6.4.). The key element of this strategy is the carrier sample, which serves to increase the available peptide ion current and to reduce the loss of protein from the single cell samples due to adhesion on the surface of the equipment. Individual cells are isolated under a microscope and transferred into glass vials. Cell lysis by ultrasonication, protein denaturation and tryptic digestion are performed in the same glass vial. The resulting peptides are labeled with a 10-plex TMT tag, mixed and analyzed by LC-MS/MS. No sample clean-up is required before analysis as no detergents or contaminating agents are introduced into the samples.

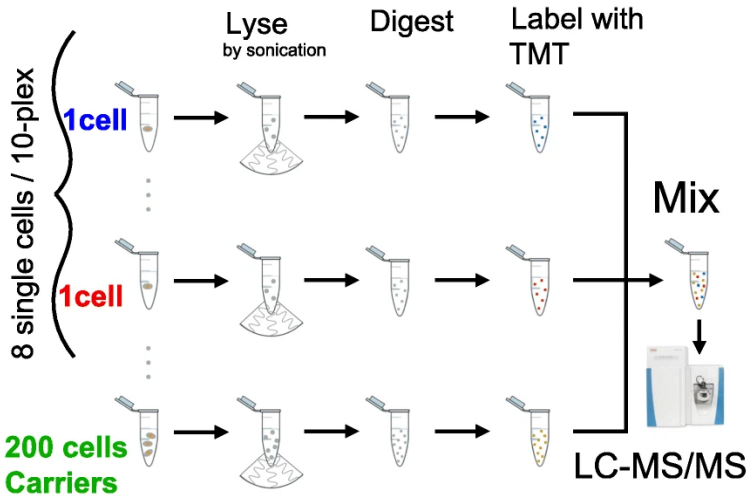


***Figure 6.4**. Conceptual diagram and workflow of SCoPE-MS. Figure adopted from* [11].

The SCoPE-MS strategy has been successfully applied to study the differentiation of embryonic stem cells, providing the identification of more than 1000 proteins [11]. Principal Component Analysis and Gene Onthology analysis showed variation of the single-cell proteome during differentiation and were used to identify the molecular functions and biological processes involved in the cell differentiation process. SCoPE-MS represents a promising strategy for single-cell proteomics because it can be implemented in a proteomics lab with little extra costs or equipment. However, particular attention should be paid to the selection of the carrier cells, as it defines which proteins can be identified and quantified in the single cell samples.

The recent advances in ultrasensitive proteomics have enabled tissue heterogeneity to be studied by comparing the proteome of minute tissue samples and even single microdissected cells. But what about proteome dynamics? Can proteomics be used to study the time evolution of the proteome of cells?

The proteome is dynamic and tightly regulated. The primary function of protein turnover is to alter the levels of specific proteins in response to physiological changes, hormonal state or diet [15]. One of the main challenges in the study of protein dynamics is the turnover

rate, which indicates the lifetime of a given protein in a cell [16]. If protein synthesis is linked to transcriptional control, protein degradation is more coupled to the metabolic activity [17]. In a steady-state condition the concentration of the protein pool is unchanged because protein synthesis and degradation are balanced (Figure 6.5.).

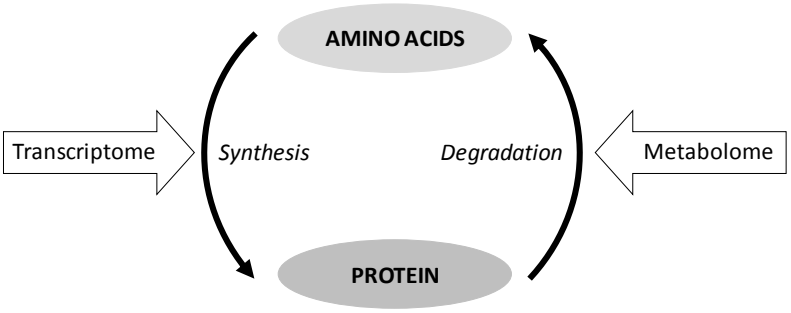


Figure 6.5. *Schematic of protein turnover regulation.*

When the available protein pool needs to be increased this can be achieved by increasing the rate of protein synthesis (translation) or decreasing the degradation rate (metabolic activity). Conversely, a reduction in the available protein pool results from decreased synthesis and/or increased degradation. It has been noted that high abundance proteins exhibit the lowest turnover rate, as they are involved in "housekeeping" roles to maintain cell homeostasis. Regulatory proteins, which are more informative about cellular processes, are expected to have higher turnover rates [18] to enable more rapid adaption.

The study of proteome dynamics can be achieved via incorporation of a label into nascent proteins. Radioactive isotopes have been used to detect newly synthesized proteins with a pulse-probe-methodology [19]. If an amino acid bearing a radioactive isotope, e.g. [^{35}S] methionine, is administrated to a cell culture for a given period of time (amino acid pulse), all the proteins synthesized during this period will contain a radioactive methionine. The labeled proteins can then be detected and distinguished from pre-existing (unlabeled) proteins by electrophoretic separation followed by radiographic detection. More recently, mass spectrometry has enabled the use of stable isotopes to study proteome dynamics [20].

In 2006 Dieterich et al. introduced Bio-Orthogonal Non Canonical Amino acid Tagging (BONCAT) to study nascent proteins by mass spectrometry [21]. Alternatively, nascent proteins can be visualized by conventional fluorescence microscopy with FlUorescent Non Canonical Amino acid Tagging (FUNCAT) [22]. A pulse of a non-canonical aminoacid (ncAA) is administrated to a cell culture in order to incorporate the ncAA in the newly-synthesized proteins (Figure 6.6.).

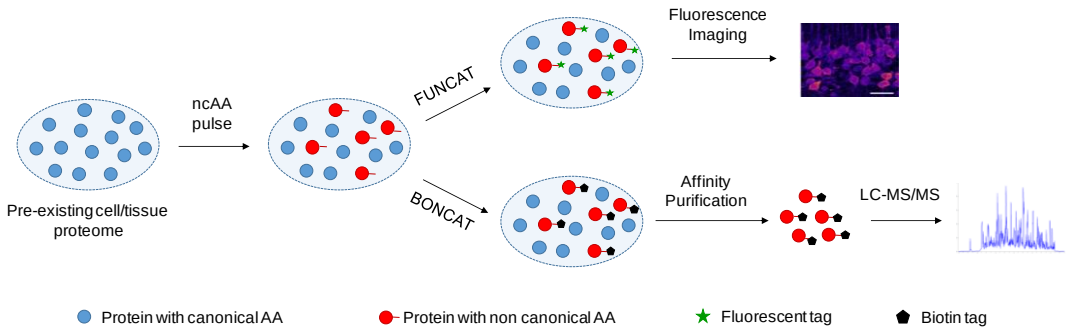


***Figure 6.6**. Schematics of BONCAT and FUNCAT strategies.*

The use of non canonical amino acids (ncAA) bearing a reactive functional group, such as an alkyne or azide moiety, allows the newly synthesized proteins to be visualized/isolated by click chemistry [23]. In FUNCAT the ncAA are labeled with a fluorescent tag, while in BONCAT the ncAA are labeled with biotin in order to isolate the nascent proteins by affinity purification techniques. In the last years BONCAT has been widely used both *in vitro* and *in vivo* to study proteome dynamics [24][25][21][26][27].

A groundbreaking development in tissue proteomics came in 2017 with work that combined space and time resolved proteomics. Alvarez-Castelao et al. developed a strategy to study proteome dynamics in specific cell types *in vivo* [28]. Using a modified BONCAT procedure and a transgenic mouse model it was possible to study nascent proteins in excitatory primary neurons and Purkinje neurons *in vitro* (brain slices) and *in vivo*.

The incorporation of ncAA requires processing by the endogenous cell machinery. A mutant form of methionyl tRNA synthetase (metRS) was used to allow the binding of azidonorleucine (ANL) instead of methionine, and thus the incorporation of ANL into nascent proteins. In the mouse model reported by Alvarez-Castelao et al. MetRS* expression is under the control of Cre recombinase. Thus, MetRS* is expressed only if Cre recombinase is also expressed. In order to promote cell-type specific incorporation, the authors engineered two other mouse models in which the expression of Cre recombinase is under the control of a cell type specific promoter: (i) CamK2a-Cre in which Cre recombinase is expressed only in excitatory neurons and (ii) GAD2-Cre in which Cre recombinase is expressed only in inhibitory neurons. By crossing the mouse model MetRS* with CamK2a-Cre they obtained a mouse model in which MetRS* is expressed only in excitatory neurons. This means that in this mouse model the ANL incorporation can be achieved only in excitatory neurons. Similarly by crossing the mouse model MetRS* with GAD2-Cre incorporation can occur only in inhibitory neurons. In the *in vivo* experiment mice were fed with ANL for 21 days and after sacrifice the brain slices were used for FUNCAT and BONCAT (Figure 6.7. A).

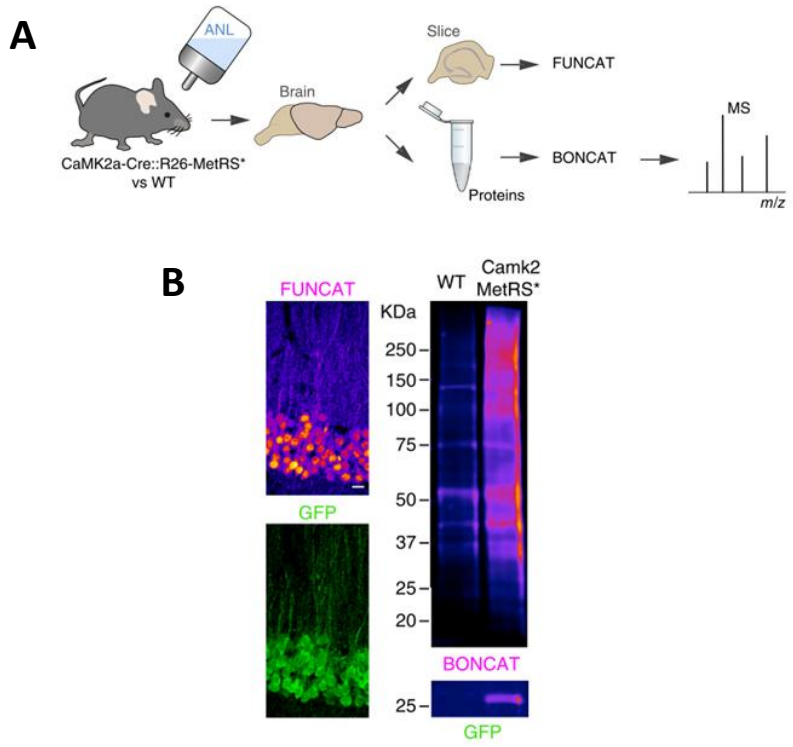


Figure 6.7*. A) Schematic of the in vivo experiment; B) FUNCAT signal in MetRS* mouse and BONCAT in WT and MetRS* mice. Figure modified from* [28].

Nascent-labeled proteins, detected using FUNCAT, were clearly visible in hippocampal neurons, indicating that ANL incorporation occurred only in excitatory neurons (Figure 6.7.B). Moreover, Western blot analysis revealed an abundance of biotinylated nascent proteins, spanning all molecular weights, in hippocampal tissue from the CaMK2a-Cre MetRS* mouse and much lower background levels of biotinylated proteins obtained from the WT mice, indicating the high selectivity of the BONCAT procedure. MS-based proteomics was also used to compare the proteomes obtained from the CaMK2a-Cre MetRS* mouse (or GAD2-Cre MetRS* mouse) with WT mouse. More than 1000 proteins were found to be enriched in the CaMK2a-Cre MetRS* mouse and Gene Onthoogy analysis found a significant enrichment in processes related to synaptic transmission and plasticity. The protein expression profiles obtained from the CaMK2a-Cre MetRS* and GAD2-Cre MetRS* mice were also compared from protein expression profiles obtained

from whole hippocampal, cerebellar and glial cell populations. Principal Component Analysis (PCA) showed a net segregation between the different cell types (Figure 6.8.), confirming the high selectivity of the experimental procedure.

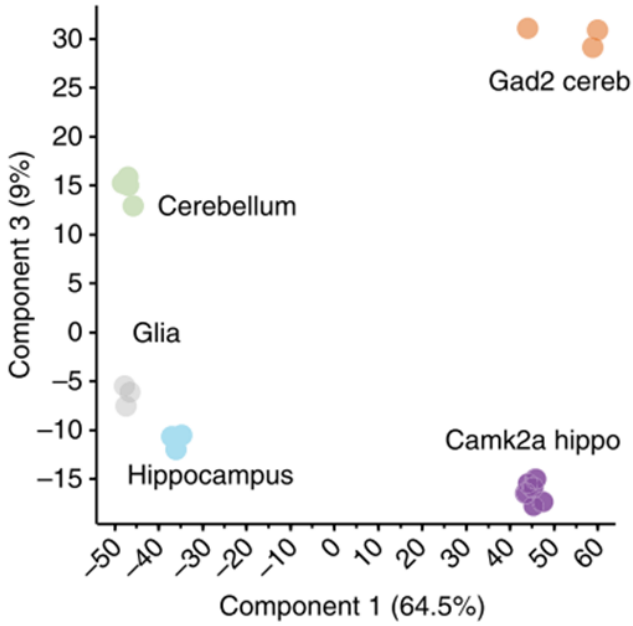


Figure 6.8*. Principal Component Analysis performed on protein expression levels of different cell types. Figure adopted from* [28].

The results demonstrated that BONCAT enabled the identification of cell-type specific proteomes labeled *in vivo* in complex tissues without mechanical dissociation and isolation of specific cell types**.** The authors studied also the proteome dynamics induced by external stimuli. Since the sensory environment have been proven to modify brain circuits and synapses, they compared the proteomes of mice exposed to an enriched environment with those housed in standard conditions. The BONCAT strategy provided the identification of more than 200 proteins that exhibited significant differences in expression between the two groups of mice.

In conclusion, all these works show that MS-based tissue proteomics is still improving. The technological and experimental developments achieved in the last years are opening exciting new application fields, from single cell analysis to personalized medicine. In the future MS-based proteomics will be useful to add a spatial and temporal dimension to our knowledge of cellular metabolism in both normal and diseased tissues.

6.2. Personal considerations

When I started to work in the proteomics field at the beginning of my PhD in 2014, there were only a few studies about ultrasensitive proteomics. The vast majority of the proteomics studies were focused on whole organs or whole tissue sections. In that scenario, the study of the proteome changes associated with disease was challenging because the technologies available were not able to characterize heterogeneous tissues with sufficient cellular/phenotypic specificity. The development of high sensitivity sample preparation protocols and the coupling of these procedures with laser capture microdissection have helped revolutionize the proteomics field, as it became finally possible to focus the analysis on specific tissue areas/cells of interest.

The set of high sensitivity and high specificity protocols developed during my PhD provides a powerful tool for the proteomics community. The main strength of the procedures is not only the increased depth of coverage that can be achieved but also their versatility. Indeed, these procedures can be easily implemented in every proteomics laboratory and are compatible with a wide variety of applications. In this thesis I showed the application of these procedures to mouse kidney substructures, to the characterization of the brain tumor microenvironment and to the study of a neurodegenerative disorder. But the same set of protocols can be adjusted to study FACS sorted cells and to characterize every type of tissue.

The advances reported here and developed by partner proteomics laboratories now the enable the analysis of the proteome of small numbers of mammalian cells. In my opinion the key strategy for the further development of single-cell proteomics is the coupling between microfluidics and proteomics. The possibility to use microfluidic technologies to manipulate sample volumes in the order of nL is crucial to reduce sample losses and to increase sensitivity. Indeed, the best results on single-cell proteomics were obtained using a microfluidic device (the nanoPOTS platform). On the other hand, I believe that microfluidics represents also the main limitation for the rapid establishment of single-cell proteomics. The use of microfluidics in proteomics is not a new idea; in the last decade many publications have reported different methods for the coupling of microfluidics and

proteomics. However only two years ago the combination of these two technologies allowed the successful analysis of the proteome of a single cell. This partially reflects the fact that proteomics and microfluidics are two very distinct fields that require distinct expertise and technologies. Thus, a very close collaboration and cooperation between these two fields must be established in order to develop a powerful technology that can be widely implemented and used by the proteomics community. It is expected that the encouraging results obtained in these first studies will help promote such cooperation and that in the coming years numerous platforms will be developed for single-cell proteomics.

Another aspect that should not be underestimated is the careful experimental design required for a single-cell study. In single cell RNAseq a large number of cells can be simultaneously analyzed, using genomic bar-codes to distinguish the contributions from different cells. In this manner RNAseq can examine the molecular variability of a cell population, as well as investigate the characteristics of specific subpopulations. The low throughput of single cell proteomics precludes such an approach and single cells must be selected from a larger number of cells prior to the proteomics experiment. The selection of the single cell is crucial and great care should be taken to ensure that the cell is representative of the biological process under investigation.

In conclusion, I believe that the combination of microfluidics, single-cell genomics and single-cell proteomics will represent the winning strategy to understand the biological cellular networks that occur in cancer, neuroscience and stem cell biology and will provide new methods to diagnose and treat diseases.

References

1. Altschuler, S.J., and Wu, L.F. (2010) Cellular Heterogeneity: Do Differences Make a Difference? *Cell*, **141** (4), 559–563.

2. Dagogo-Jack, I., and Shaw, A.T. (2017) Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.*, **15**, 81.

3. Raj, A., and van Oudenaarden, A. (2008) Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell*, **135** (2), 216–226.

4. Levy, E., and Slavov, N. (2018) Single cell protein analysis for systems biology. *Essays Biochem.*, **62** (4), 595–605.

5. Rubakhin, S.S., Romanova, E. V, Nemes, P., and Sweedler, J. V (2011) Profiling metabolites and peptides in single cells. *Nat. Methods*, **8**, S20.

6. Achim, K., Pettit, J.-B., Saraiva, L.R., Gavriouchkina, D., Larsson, T., Arendt, D., and Marioni, J.C. (2015) High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.*, **33**, 503.

7. Shapiro, E., Biezuner, T., and Linnarsson, S. (2013) Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.*, **14**, 618.

8. Lombard-Banek, C., Moody, S.A., and Nemes, P. (2016) Single-Cell Mass Spectrometry for Discovery Proteomics: Quantifying Translational Cell Heterogeneity in the 16-Cell Frog (Xenopus) Embryo. *Angew. Chemie Int. Ed.*, **55** (7), 2454–2458.

9. Sun, L., Dubiak, K.M., Peuchen, E.H., Zhang, Z., Zhu, G., Huber, P.W., and Dovichi, N.J. (2016) Single Cell Proteomics Using Frog (Xenopus laevis) Blastomeres Isolated from Early Stage Embryos, Which Form a Geometric Progression in Protein Content. *Anal. Chem.*, **88** (13), 6653–6657.

10. Zhu, Y., Piehowski, P.D., Zhao, R., Chen, J., Shen, Y., Moore, R.J., Shukla, A.K., Petyuk, V.A., Campbell-Thompson, M., Mathews, C.E., Smith, R.D., Qian, W.-J.,

and Kelly, R.T. (2018) Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nat. Commun.*, **9** (1), 882.

11. Budnik, B., Levy, E., Harmange, G., and Slavov, N. (2018) SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.*, **19** (1), 161.

12. Dou, M., Clair, G., Tsai, C.-F., Xu, K., Chrisler, W.B., Sontag, R.L., Zhao, R., Moore, R.J., Liu, T., Pasa-Tolic, L., Smith, R.D., Shi, T., Adkins, J.N., Qian, W.-J., Kelly, R.T., Ansong, C., and Zhu, Y. (2019) High-Throughput Single Cell Proteomics Enabled by Multiplex Isobaric Labeling in a Nanodroplet Sample Preparation Platform. *Anal. Chem.*, **91** (20), 13119–13127.

13. Zhu, Y., Dou, M., Piehowski, P.D., Liang, Y., Wang, F., Chu, R.K., Chrisler, W.B., Smith, J.N., Schwarz, K.C., Shen, Y., Shukla, A.K., Moore, R.J., Smith, R.D., Qian, W.-J., and Kelly, R.T. (2018) Spatially Resolved Proteome Mapping of Laser Capture Microdissected Tissue with Automated Sample Transfer to Nanodroplets. *Mol. & Cell. Proteomics*, **17** (9), 1864 LP – 1874.

14. Zhu, Y., Clair, G., Chrisler, W.B., Shen, Y., Zhao, R., Shukla, A.K., Moore, R.J., Misra, R.S., Pryhuber, G.S., Smith, R.D., Ansong, C., and Kelly, R.T. (2018) Proteomic Analysis of Single Mammalian Cells Enabled by Microfluidic Nanodroplet Sample Preparation and Ultrasensitive NanoLC-MS. *Angew. Chemie Int. Ed.*, **57** (38), 12370–12374.

15. Claydon, A.J., and Beynon, R. (2012) Proteome Dynamics: Revisiting Turnover with a Global Perspective. *Mol. Cell. Proteomics*, **11** (12), 1551–1565.

16. King, R.W., Deshaies, R.J., Peters, J.M., and Kirschner, M.W. (1996) How Proteolysis Drives the Cell Cycle. *Science*, **274** (5293), 1652–1659.

17. Doherty, M.K., and Beynon, R.J. (2006) Protein turnover on the scale of the proteome. *Expert Rev. Proteomics*, **3** (1), 97–110.

18. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011) Global quantification of mammalian gene expression

control. *Nature*, **473** (7347), 337–342.

19. Beynon, R.J., and Pratt, J.M. (2005) Metabolic Labeling of Proteins for Proteomics. *Mol. Cell. Proteomics*, **4** (7), 857–872.

20. Mann, M. (2006) Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.*, **7** (12), 952–958.

21. Dieterich, D.C., Link, A.J., Graumann, J., Tirrell, D.A., and Schuman, E.M. (2006) Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). **103** (25), 9482–9487.

22. Dieterich, D.C., Hodas, J.J.L., Gouzer, G., Shadrin, I.Y., Ngo, J.T., Triller, A., Tirrell, D.A., and Schuman, E.M. (2010) In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons. *Nat. Neurosci.*, **13** (7), 897–905.

23. Szychowski, J., Mahdavi, A., Hodas, J.J.L., Bagert, J.D., Ngo, J.T., Landgraf, P., Dieterich, D.C., Schuman, E.M., and Tirrell, D.A. (2010) Cleavable Biotin Probes for Labeling of Biomolecules via Azide - Alkyne Cycloaddition. *J. Am. Chem. Soc.*, (15), 18351–18360.

24. Bagert, J.D., Xie, Y.J., Sweredoski, M.J., Qi, Y., Hess, S., Schuman, E.M., Tirrell, D.A., Hess, S., and Tirrell, D.A. (2014) Quantitative , Time-Resolved Proteomic Analysis by Combining Bioorthogonal Noncanonical Amino Acid Tagging and Pulsed Stable Isotope Labeling by Amino Acids in Cell Culture. *Mol. Cell. Proteomics*, **13** (5), 1–19.

25. Daniel, A., Huber, M., Begley, J., Peter, C., Forest, M., Rothenberg, D.A., Taliaferro, J.M., Huber, S.M., Begley, T.J., and Dedon, P.C. (2018) Article A Proteomics Approach to Profiling the Temporal Translational Response to Stress and Growth A Proteomics Approach to Profiling the Temporal Translational Response to Stress and Growth. *ISCIENCE*, **9**, 367–381.

26. Hodas, J.J.L., Nehring, A., Nicole, H., Sweredoski, M.J., Pielot, R., Hess, S.,

Tirrell, D.A., Dieterich, D.C., and Schuman, E.M. (2012) Dopaminergic modulation of the hippocampal neuropil proteome identified by bioorthogonal noncanonical amino acid tagging (BONCAT). *Proteomics*, **15** (16), 2464–2476.

27. Shen, W., Liu, H., Schiapparelli, L., Mcclatchy, D., He, H., and Iii, J.R.Y. (2014) Article Acute Synthesis of CPEB Is Required for Plasticity of Visual Avoidance Behavior in Xenopus. *CellReports*, **6** (4), 737–747.

28. Alvarez-castelao, B., Schanzenbächer, C.T., Hanus, C., Glock, C., Dieck, S., Dörrbaum, A.R., Bartnik, I., Nassim-assir, B., Ciirdaeva, E., Mueller, A., Dieterich, D.C., Tirrell, D.A., Langer, J.D., and Schuman, E.M. (2017) Cell-type-specific metabolic labeling of nascent proteomes in vivo. *Nat. Biotechnol.*, **35** (12), 1196–1201.