NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTAMENT OF CHEMISTRY

Rafael Guerreiro Bento

BSc in Biochemistry

Assessment of a Capsule for Easy Urine Proteome Collection at Home

MASTER IN BIOCHEMISTRY

NOVA University Lisbon

September 2022

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTAMENT OF CHEMISTRY

Assessment of a Capsule for Easy Urine Proteome Collection at Home

Rafael Guerreiro Bento

BSc in Biochemistry

Adviser: Hugo Miguel Baptista Carreira dos Santos
Assistant Researcher, NOVA University Lisbon

Co-advisers: José Luis Capelo
Associate Professor with Habilitation, NOVA University Lisbon

Examination Committee:

Chair: Sofia Rocha Pauleta
Assistant Professor, NOVA University Lisbon

Rapporteurs: Mário Emanuel Campos de Sousa Diniz,
Assistant Professor, *NOVA University Lisbon*

Adviser: Hugo Miguel Baptista Carreira dos Santos
Assistant Researcher, NOVA University Lisbon

MASTER IN BIOCHEMISTRY

NOVA University Lisbon
September, 2022

Assessment of a capsule for easy urine proteome collection at home

ACKNOWLEDGEMENTS

As I look back on this past year, I would like to acknowledge all the people who contributed to my work and supported me throughout this chapter of my life.

I want to start by acknowledging Professor Capelo and Professor Lodeiro, who gave me this fantastic opportunity to integrate with the BIOSCOPE group. I feel incredibly privileged to have been a part of this outstanding research team and wonderful group of people. I take the skills and knowledge I have gained and the amazing friendships I have developed.

I am immensely grateful to Professor Capelo, who has guided me through this journey since the first day and believed in me and my potential.

I would also like to express my enormous gratitude to Dr Hugo Santos for all their time and patience in helping me with my work.

Of course, a special thank you to all the members of the BIOSCOPE team for welcoming me and supporting me in whatever was needed. In particular, Luís Carvalho, Gonçalo Martins, and Joana Galhano for all the help, motivation, and laughs.

To Gustavo Nascimento, Ricardo Roberto, and Gonçalo Sampaio for all the support and laughs.

To Clara Gonçalves, mere words can hardly describe how grateful I am to have had you by my side during this journey.

Finally, to my family for all the support and encouragement. In particular to my mom, João and my cousin Nelson, who were always there for me.

“If I have seen further,

it is by standing on the shoulders of giants.”

(Isaac Newton).

ABSTRACT

Urine as a biological sample contains a wealth of information with great significance for medical diagnosis while also benefiting from an easy and non-invasive collection procedure. Thus, urine has become one of the standard samples for diagnosing multiple conditions and for proteomics-based biomarker discovery studies. In recent years, our group has developed an approach for Bladder cancer (BCa) diagnosis and patient monitoring based on proteomics, using the dysregulation of the urinary proteome to gain information regarding the patient's state, such as tumour diagnosis and stage. Furthermore, analysing multiple urine samples from different time points provides valuable insight into the patient's treatment response and recovery. However, monitoring is impaired by the lack of sample collection at the hospital. Thus, our goal was to develop a new easy-to-use method of urinary proteome collection so that patients could collect their samples at home. Thus, the urine proteome was collected using syringe filters and handled with a standard shotgun proteomics protocol. Multiple filter membranes were tested, from which the molecular weight cut-off membranes were the most efficient for our proposal. A proof of concept was conducted with BCa patient samples. The Filter Aided Sample Preparation (FASP) protocol was used as a control for the results obtained with our new approach. Our results show that the urine proteome can be collected and treated using our filter approach, although further research is needed to increase digestion efficiency to standard protocol levels. We propose syringe filtering as an easy approach to follow-up BCa patients and potentially any other diseases that can be assessed using the urine proteome.

Keywords: Proteomics, Syringe filters, Label-Free Mass Spectrometry

RESUMO

A urina enquanto amostra biológica contém informação extremamente relevante para o diagnóstico de diversas condições médicas, sendo recolhida de um modo fácil e não invasivo, tornando-se uma das amostras biológicas ideais para diagnóstico de condições médicas e para estudos de proteómica baseada em espectrometria de massa. Recentemente, o nosso grupo desenvolveu uma metodologia para diagnóstico e monitorização de pacientes de cancro da bexiga (BCa), onde a desregulação do proteoma urinário permite o diagnóstico e identificação de estadio do tumor. Adicionalmente, ao analisar amostras recolhidas ao longo do tempo, é possível fazer o seguimento do paciente, obtendo informação acerca da sua resposta ao tratamento. Esta monitorização é, contudo, dificultada devido aos grandes intervalos de tempo entre recolhas. Sendo assim, o nosso objetivo foi desenvolver uma nova metodologia para recolha do proteoma urinário, de uso autónomo pelo paciente. Assim, o proteoma urinário foi recolhido usando filtros de seringa e analisado com um protocolo de *shotgun proteomics*. Foram testadas várias membranas, sendo que as de separação por tamanho molecular foram as mais eficientes na recolha do proteoma. Foi ainda feito um *proof of concept* com amostras de pacientes de BCa onde o protocolo de *Filter Aided Sample Preparation* (FASP), foi usando como controlo para os resultados obtidos com a nova metodologia. Os resultados obtidos demonstram que o proteoma urinário pode ser eficazmente recolhido usando estes filtros, contudo a técnica requer otimizações para alcançar níveis de digestão proteica similares aos dos protocolos standard. Propomos filtros de seringa como uma técnica fácil para monitorizar pacientes de BCa, e potencialmente outras doenças a partir do proteoma urinário.

Palavras-Chave: Proteómica, Filtros de Seringa, Espectrometria de Massa

CONTENTS

LIST OF FIGURES

List of Tables

ACRONYMS

ACN	Acetonitrile
AmBic	Ammonium bicarbonate
BCa	Bladder cancer
BSA	Bovine serum albumin
ESI	Electrospray Ionisation
FA	Formic acid
FASP	Filter-aided sample preparation
FDR	False discovery rate
IAA	Iodoacetamide
LC	Liquid chromatography
LC-MS	Liquid chromatography - mass spectrometry
LC-MS/MS	Liquid chromatography - tandem mass spectrometry
LFQ	Label-free quantification
MALDI	Matrix-Assisted Laser Desorption/Ionisation

MS	Mass spectrometry
MWCO	Molecular weight cut-off
m/z	Mass-to-charge ratio
Q	Quadrupole
SILAC	Stable isotope labelling by amino acids in cell culture
TOF	Time-of-Flight

I

INTRODUCTION

I.1 Bladder Cancer

Bladder cancer (BCa) is one of the most common cancers worldwide, with nearly 550 000 new cases diagnosed yearly. BCa has a higher incidence in men, being the fourth and eighth most common cancer in the male and female demographics, respectively [1, 2]. Despite this, men do have slightly higher survival chances since their tumours tend to be diagnosed earlier, which highly facilitates treatment. As for the treatment itself, BCa is the most expensive cancer to treat due to its requirement for surgical interventions and long monitoring periods [2].

Regarding possible risk factors, smoking is the most influential factor, and is responsible for nearly two out of three new diagnoses. Other risk factors include occupational exposure to chemicals or carcinogens, genetic disorders, family history, age, and ethnicity [3].

I.1.1 Classification

The most common type of BCa is urothelial carcinoma (UC) or transitional cell carcinoma (TCC) [1, 2]. This type of BCa originates in the urothelial cells of the urinary tract, which make up the bladder's innermost layer. Because of this, this type of BCa is also often referred to as non-muscle invasive bladder cancer (NMIBC). NMIBC accounts for 75%-80% of all newly diagnosed BCa cases. The other less common types of BCa include squamous cell carcinoma, adenocarcinoma, small cell carcinoma and sarcoma. These are usually associated with more advanced stages of BCa and higher mortality rates [1, 2].

Bladder tumours can also be classified according to their growth patterns. Papillary carcinomas originate in the urothelium and grow toward the center of the bladder. This generally leads to better outcomes because the tumour does not tend to progress into the outer layers of the bladder. On the other hand, flat tumours, or carcinomas in situ (CIS), do not grow into the center of the bladder but instead remain in the urothelium. These tend to have a higher risk than papillary carcinomas of becoming muscle-invasive [4].

If the tumour does progress toward the outer layers of the bladder, its development can be categorised into different stages. Ta is the very initial stage, where the carcinoma is restricted to the urothelium, the bladder's innermost layer. In the T1 stage, the tumour invades the subepithelial connective tissue. Both stages are often referred to in the literature as simply NMIBC. In T2 stage the tumour invades the surrounding muscle, with T3 reaching the surrounding tissues and fat. Lastly, stage T4, where the tumour reaches the blood stream, and can originate metastasis, reaching and invading other organs and tissues. Stages from T2-T4 are often referred to as muscle-invasive bladder cancer (MIBC) [4].

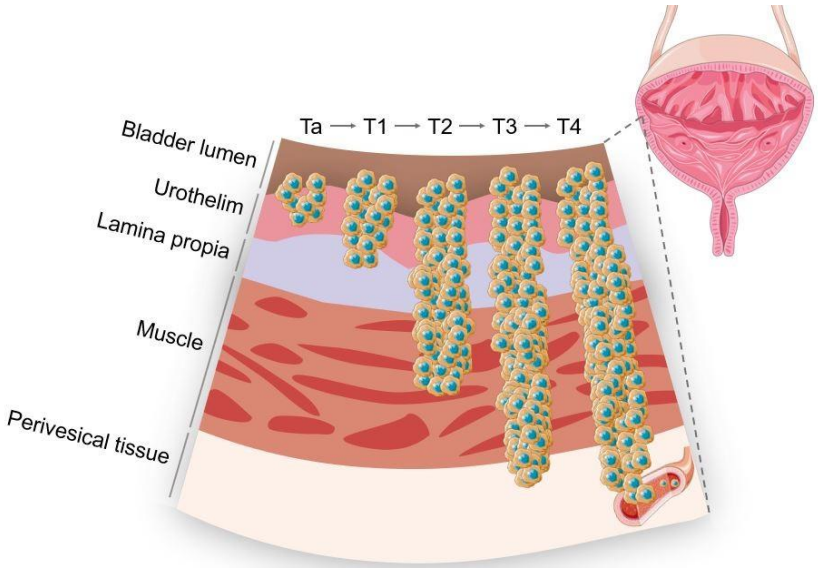

Figure I.1 - Bladder Cancer Stages. Cancer stages are classified according to the tissue layers the tumor has invaded. Figure adapted from Carvalho, L. B. et al poster presentation at International Caparica Conference on Analytical Proteomics 2022.

I.1.2 Diagnostic

The first sign of a bladder tumour is haematuria, an increased number of red blood cells excreted through urine, which may give it a slight red colouration. Haematuria may not always be visible to the naked eye and is not a definite indicator of BCa, although it is the most common symptom [5]. This is usually followed by other urinary symptoms such as pain or burning sensations during urination or augmented urination frequency, to name a few. Other complications in the urinary tract may cause all these symptoms, but there are a series of tests to rule out each possible cause for these symptoms [5].

Once one or more of these symptoms are present, the first thing a physician might ask for is a urine test. The urinalysis will, among other things, assess the number of red blood cells in the urine, evaluating the severity of the haematuria. Because haematuria is random and may be absent for days after its first appearance, this urinalysis may be repeated later. In the case of a positive result for haematuria, the next step is the assessment of the urinary tract to try and rule out other potential causes for symptoms, such as infections or lesions. Depending on this analysis, if the possibility for BCa remains open, the doctor will resort to a cystoscopy [5].

A cystoscopy is a chirurgical intervention where a cystoscope is inserted through the urethra, reaching the bladder [6]. The cystoscope has a camera at the end, allowing the doctor to evaluate the inside of the bladder and look for any irregularities. If a tumour or abnormality is detected, the cystoscope can be used as a biopsy tool, inserting the required equipment through the tube. Due to this procedure's invasive nature, a cystoscopy requires anaesthesia [6]. This is the first factor that raises the costs of BCa treatment. Analysis of the biopsy will reveal if the tissue is cancerous or not. The stage is attributed to how far the tumour has progressed through the bladder layers, according to the classification system mentioned previously.

Around 75%-80% of all new diagnosed BCa are NMIBC, being detected relatively early in their development, which facilitates treatment. However, BCa has a high risk of recurrence

with progression, and for this reason, patients are kept under routine observation for long periods of time. This monitoring of the patients is done through cystoscopies and other adjacent tests every three to six months for two or more years, depending on each case [1, 2].

I.1.3 Treatment

The most effective treatment option for NMIBC is currently a combination of transurethral resection of bladder tumour (TURBT) with Bacille Calmette-Guérin (BCG) treatments. TURBT is performed with the use of a rigid cystoscope, also known as a resectoscope [7]. This device has a wire loop at the end that can scrape the tumour off the bladder wall. The tumour may also be burned with a cystoscope, in a process called fulguration, or using a high-energy laser, to ensure it is completely removed. Depending on tumour stage, one cystoscopy might not be enough, and a second one might be necessary. In some extreme cases, the doctor might also recommend a cystectomy, which is the complete removal of the bladder, to prevent the tumour from spreading to other tissues [7].

Bladder cancer has a high recurrence rate, therefore patients are kept under surveillance in the form of recurrent cystoscopies, to detect possible recurrences as soon as possible. The frequency of these cystoscopies will depend on the stage of the initial tumour when diagnosed, but it can range from one every six months to one every three months over a five-year period [7].

The BCG vaccine was originally intended to treat tuberculosis, and in many countries, it is still used for that purpose. In this case, it is used to elicit the body's immune response and prevent the cancer from returning. The BCG treatments start a few weeks after the first cystoscopy, where BCG is administered once a week for six weeks. Following this initial treatment, maintenance BCG might still be necessary, with monthly administrations for one to three years [7].

BCG is administered directly to the bladder, as to illicit an immune response which will difficult tumour development. In fact, BCG treatment with maintenance does reduce the risk of recurrence with progression in BCa patients [8].

These frequent treatments, and especially the surveillance cystoscopies, all raise the costs of BCa treatment, making it the most expensive cancer to treat [1]. Furthermore, patients are constantly subjected to invasive procedures for years, which is why new, less invasive diagnostic tools, such as the use of molecular biomarkers present in urine, plasma, or serum, have garnered so much interest over the past decade, not only because the costs are reduced, but also because it raises the patient's quality of life.

I.2 Mass Spectrometry

Mass spectrometry (MS), is an analytical technique that has grown tremendously and has established itself as an indispensable tool in various fields, including but not limited to chemistry, biochemistry, toxicology, and medicine [9]. MS measures the mass-to-charge ratio (m/z) of ions in a gas phase. In essence, a sample loaded into a mass spectrometer will first be ionised, resulting in charged molecules or charged molecule fragments. After Ionisation, the ions are separated according to their mass-to-charge ratio (m/z) in the mass analyser. Finally, the ions and their respective abundances are detected, and the results are shown as a mass spectrum [10]. In a mass spectrum, the mass-to-charge ratios of the identified ions are plotted against their intensities, as seen in **Figure I.2**. In other words, each peak in the mass spectrum represents an ion with a specific m/z ratio. In contrast, the peak height usually represents its relative abundance in the sample. To accomplish this, mass spectrometry has three key points, these being ion generation, ion separation, and finally, qualitative, and quantitative ion detection.

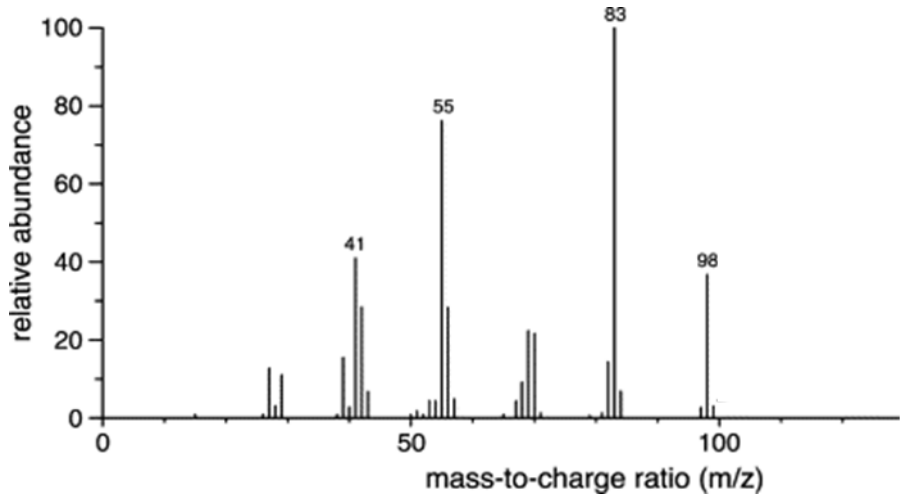

Figure I.2 - Mass spectrum. Adapted from fire Debris Analysis, Chapter 8 [11].

I.2.1 Ionisation Sources

As mentioned, mass spectrometry is an analytical tool that measures the m/z ratios of ions in a gas phase. This stems from the fact that the separation and later detection of the analytes require these species to be charged. Thus, once the sample enters the mass spectrometer, the first step will be its ionisation. This is simply the process through which a neutral analyte will either gain or lose electrons, becoming an anionic or cationic species, respectively. Typically, in proteomics, mass spectrometers are set up to work with positively charged analytes (cations), although the spectrometers can also be tuned to work with negatively charged ones (anions) [12]. Ionisation can be achieved through different methods, which can be classified into two categories (I) hard ionisation methods and (II) soft ionisation sources [9]. Hard ionisation methods deliver excessive energy, leading to analyte fragmentation. On the other hand, soft ionisation methods provide much smaller amounts of energy to the sample, producing ions without analyte fragmentation. Thus, soft ionisation methods reduce the amount of analyte fragmentation when working with larger molecules, such as peptides, proteins or other biomolecules [9].

Among the most widely used ionisation methods in proteomics is Matrix-Assisted Laser Desorption/Ionisation (MALDI). This soft ionisation source has become a staple in a clinical microbiology setting for fast identification of bacteria and other microorganisms when in a MALDI-TOF setup, which will later be further explained [13]. Briefly, in MALDI, the sample is first crystalised onto a plate with a matrix solution. This plate is then inserted into the mass spectrometer, where the ionisation occurs. In MALDI, ionisation is done by irradiating the crystalised sample with a laser beam. This causes a rapid temperature rise, leading to the evaporation and formation of protonated matrix ions [14]. As the neutrally charged sample analytes are contained within the matrix structure, the gas phase formed by the evaporation will contain these neutral analytes along will positively charged matrix ions, resulting in overall positively charged molecules that can be analysed [14].

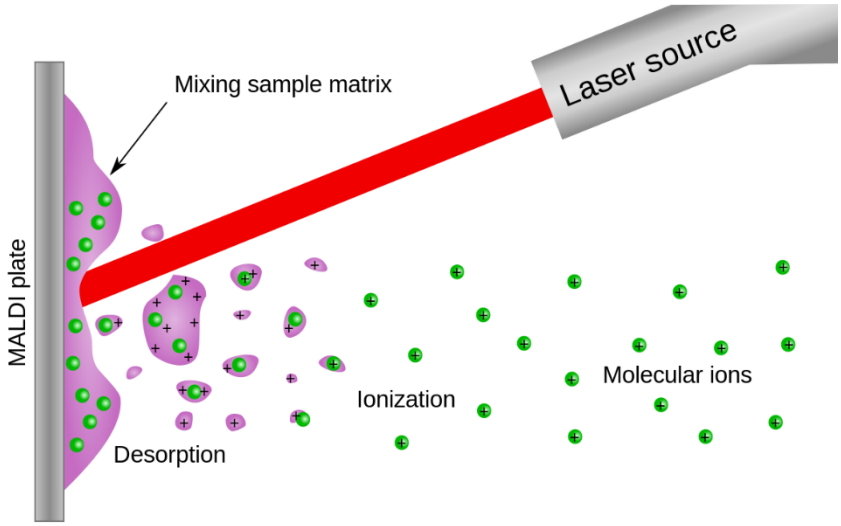

Figure I.3 - Matrix-Assisted Laser Desorption/Ionisation (MALDI). Schematic representation of the ionisation process caused by the irradiation of the matrix containing the analyte by a laser. Adapted from Wikipedia Webpage (MALDI) [15].

Another widely used soft ionisation method in mass spectrometry-based proteomics is Electrospray Ionisation (ESI). In this approach, the sample is not crystallised as it is in MALDI but instead dissolved in a conductive solvent and introduced into the mass spectrometer in the liquid phase [16]. This allows for the implementation of separation techniques, such as Liquid Chromatography (LC) or High-Performance Liquid Chromatography (HPLC), that reduce sample complexity before reaching the mass spectrometer [17]. This is an important aspect that increases the number of species being detected. As the analytes are detected, and the mass spectrum is produced, analytes with similar m/z ratios may have drastically different abundancies. This may lead to the camouflage of low abundant analytes in the sample. This is only enhanced by the fact that proteomes have vast dynamic ranges regarding protein abundance [18]. In practical terms, this means that lower abundant proteins run the risk of not being detected in the analysis, despite being present in the sample. Thus, a separation technique such as the one mentioned before can vastly reduce this risk by allowing small fractions of the

sample to be introduced in the mass spectrometer gradually and therefore increase the number of identified proteins, increasing the depth of proteome coverage.

Regarding the ESI method itself, however, and as mentioned, the sample is first dissolved in a conductive solvent and introduced in the mass spectrometer in a liquid phase [19]. The sample is passed through a needle where high voltage is applied to the solution, creating an aerosol containing charged droplets. These droplets are then evaporated with the help of a flow of heated gas until the eventual release of ionised analytes in a gas phase, which can then advance into the mass analyser.

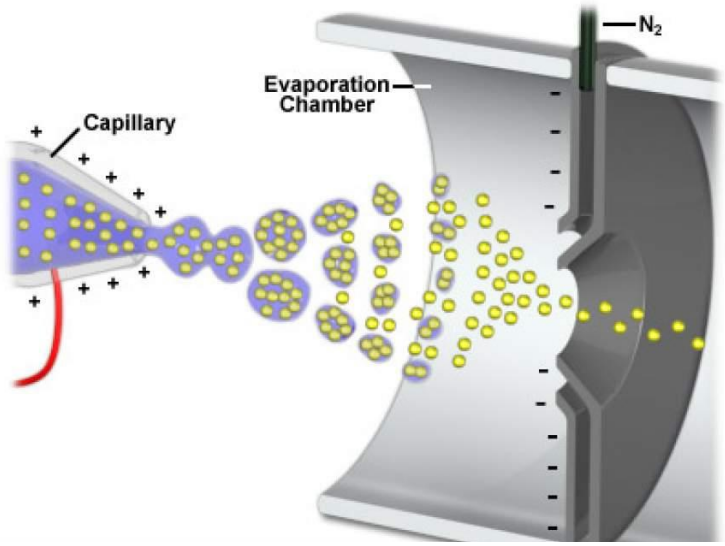

Figure I.4 - Electrospary Ionisation (ESI). It shows the formation of the aerosol, and the evaporation which leads to the formation of the analyte ions. Adapted from National High Magnetic Field Laboratory webpage [19].

I.2.2 Mass Analysers

As mentioned, mass analysers serve the purpose of separating the analyte ions according to their m/z before they reach the detector. There is a variety of different mass analysers, each of them with its methods for separating ions.

Quadrupole (Q) is one such mass analyser, consisting of four electrically charged metal rods, as shown in **Figure I.5**. The separation in Q mass analysers is done by generating an electric field using these metal rods. As mentioned, each analyte will have a specific m/z ratio, and different electrical fields can alter the oscillating trajectory of these analytes through the metal rods based on their m/z ratios [20]. This way, the Q analyser can generate an electrical field that can sustain the stable oscillating trajectory of a specific m/z ratio analyte through the metal rods, allowing it to reach the detector. Other analytes with different m/z ratios will not be able to maintain a stable oscillating trajectory through the metal rods and will eventually collide with them. By varying the electrical field being generated, the Q analysers can isolate one m/z ratio at a time to reach the detector [20].

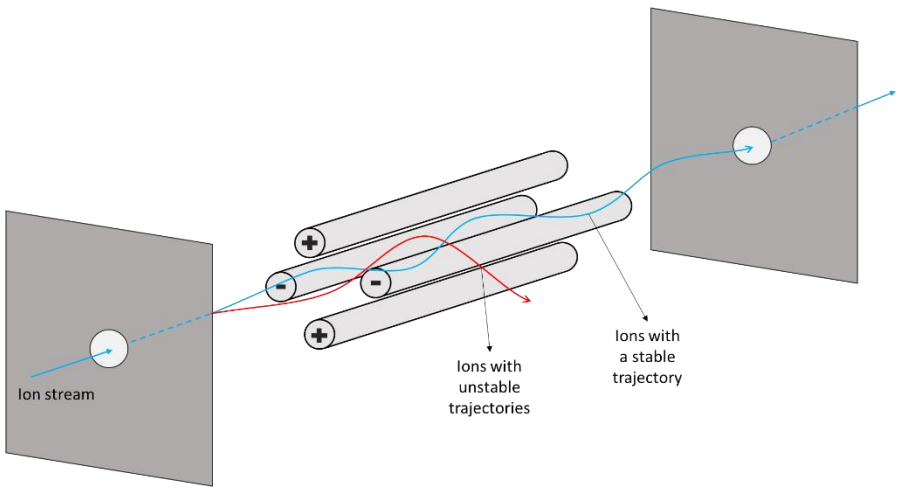

Figure I.5 - Quadrupole (Q) Mass Analyser. Shows the charged metal bars, and how the separation of the ions is achieved.

Ion traps (IT) also referred to as quadrupole ion traps, trap and store ions in an orbital motion within the ion trap and eject ions for detection. Storage is performed by collecting ions in a potential energy well, and ejection provides energy for ions to escape the potential energy well [20].

Another type of mass analyser is the Orbitrap. These mass analysers consist of three electrodes, a central spindle-shaped one and two outer ones. These analysers require an additional ion trap [21]. Having been filtered by the ion trap, the ions enter the orbitrap and start to oscillate around the central electrode due to their charge and the electrostatic fields. As different ions will have different oscillation frequencies, they will separate. The mass spectrum is then achieved by measuring the oscillation frequencies of each ion [20, 21].

Time of Flight (TOF) mass analysers have previously been mentioned when explaining MALDI but can also be used in ESI systems [21,22]. The TOF analysers are based on a very easy to understand principle that equally accelerated particles will take different times to travel the same distance based on their mass [21,22]. This way, TOF mass analysers are composed of an acceleration grid, and a flight tube. Briefly, the ions enter the TOF analyser, where they are accelerated with the same kinetic energy. As the ions have different masses, they will achieve different velocities to travel the flight tube. Thus, higher mass ions will achieve lower speeds than lower mass ones. Because all the ions must travel the same distance to reach the detector, lower mass ions, which obtain faster speeds, will arrive earlier than higher mass ones. These differences allow for the determination of each ion's mass [21]. Therefore, the longer the travel distance, the bigger the separation between the different ion masses will be. Consequently, the bigger separation between these masses, the higher the resolution of the system. Because the flight tube is physically dependent on the size of the TOF instrument, a way to increase the travel distance is with the use of reflectors, as seen in **Figure I.6**. By enabling the possibility of a curved trajectory to the detector, it is possible to increase the travel distance without necessarily increasing the physical size of the TOF instrument [21,22].

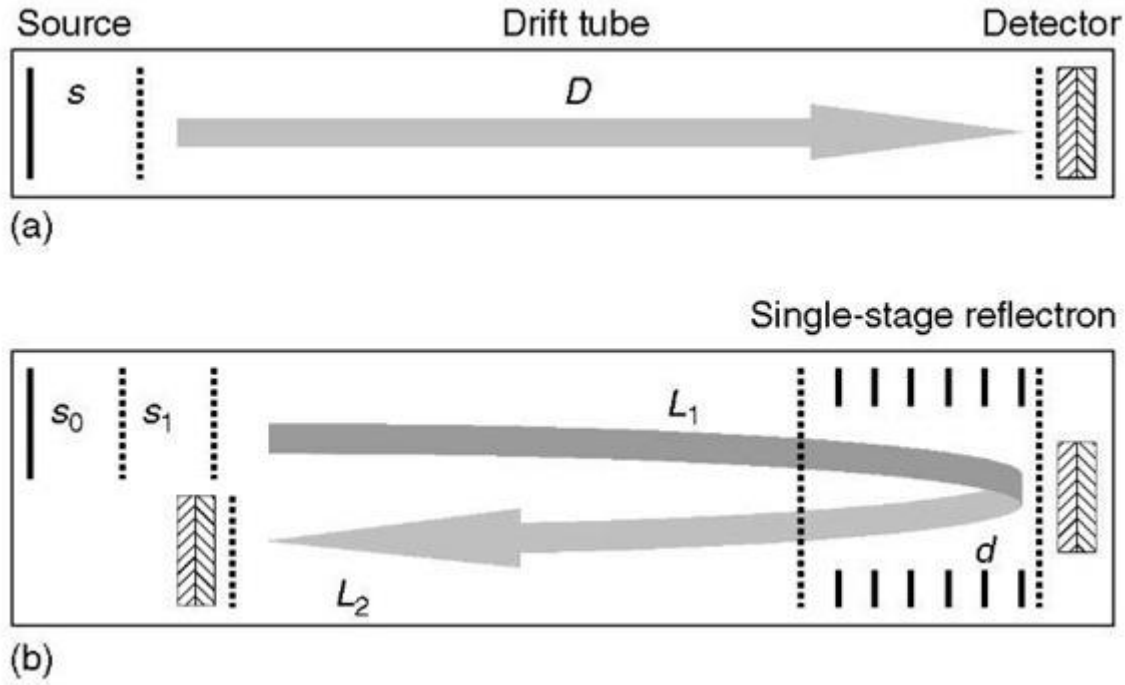

Figure I.6 - Time-of-Flight (TOF) mass analyser. A - Linear TOF analyser. **B** - TOF analyser with reflector. Adapted from (Lovric, 2011) [23].

I.2.3 Quantitative Proteomics

Modern Mass spectrometry-based proteomics is devoted to protein quantification. This is one of the most crucial factors in biomarker discovery, as proteins must be quantified in samples with different physiological conditions to evaluate their potential to be a biomarker for a certain condition. To this end, there are two main approaches in mass spectrometry for protein quantification, these being stable isotope labelling (I) and Label-free quantification (II) [24].

In stable isotope labelling methods, mass tags (stable isotopes) are incorporated in the proteins or peptides, depending on the proteomics approach. This labelling will produce a light and a heavy form of the analyte, which will appear in the mass spectrum as a pair of peaks separated by the mass of the tag. The tags used are usually 4-8 Da, as a smaller difference in

mass can cause interference between the isotopic peaks of the light and heavy peptide forms [25]. One of the most well-known stable isotope labelling methods is stable isotope labelling by amino acids in cell culture (SILAC). In this approach, an amino acid's light and heavy forms will be incorporated into the proteins of two cell cultures, respectively [25]. The proteins will then be harvested from the cultures and mixed into one single sample. This sample, now containing proteins with both heavy and light labels, will be digested and analysed by mass spectrometry.

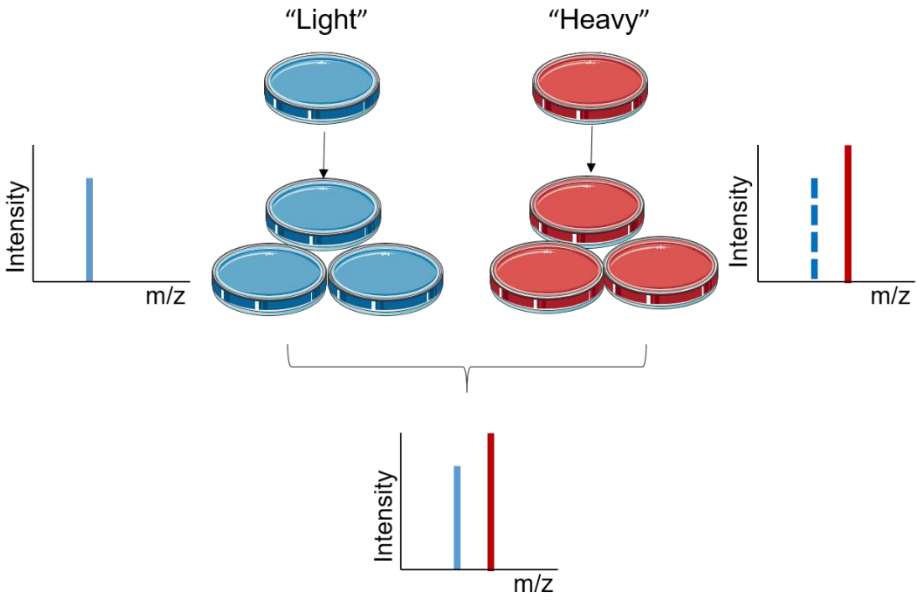

Figure I.7 - Stable isotope labeling by amino acids in cell culture (SILAC). The proteins with light and heavy amino acid labels are mixed into one sample, digested, and analyzed. This produces a pair of peaks in the MS spectrum, where the difference between the peaks corresponds to the mass difference between the light and the heavy forms. Adapted from Creative Proteomics Blog Webpage [26].

Label-free quantification (LFQ), however, is another protein quantification approach in mass spectrometry that does not require the use of labels. This is a clear advantage over the stable isotope labelling methods, requiring no extra sample handling steps and the possibility

to be used with any sample type. There are two main strategies in label-free quantification, (I) peak intensity and (II) spectral counting [24].

In spectral counting, approach quantification is done by counting the number of MS/MS spectra corresponding to the same peptides and comparing it across multiple LC-MS/MS runs. The more abundant a peptide is, the longer it will take to elute from the chromatographic column, providing it with more possibilities of being selected for MS/MS analysis [24]. Therefore, the more MS/MS spectra of a specific peptide, the more abundant it was in the sample. The comparison between different LC-MS/MS runs allows for a relative quantification, as illustrated in **Figure I.X**. This relative quantification provides only a fold change value for each peptide or protein between different samples. However, absolute quantification can also be achieved using different methods, such as exponentially modified protein abundance index (emPAI), which uses mathematical systems to estimate the total protein content in the sample [26].

In the ion peak intensity approach, the quantification is based on the intensity of the peptide signals in the mass spectrum. Briefly, the eluted peptides will produce an MS signal or peak, and quantification is done by integrating this signal. In other words, the quantification is done by calculating the area under the curve of each peak [27]. Once again, comparing these intensities between different samples allows for relative quantification. However, absolute protein quantification can also be achieved using informatic tools, such as the total protein approach (TPA) [28].

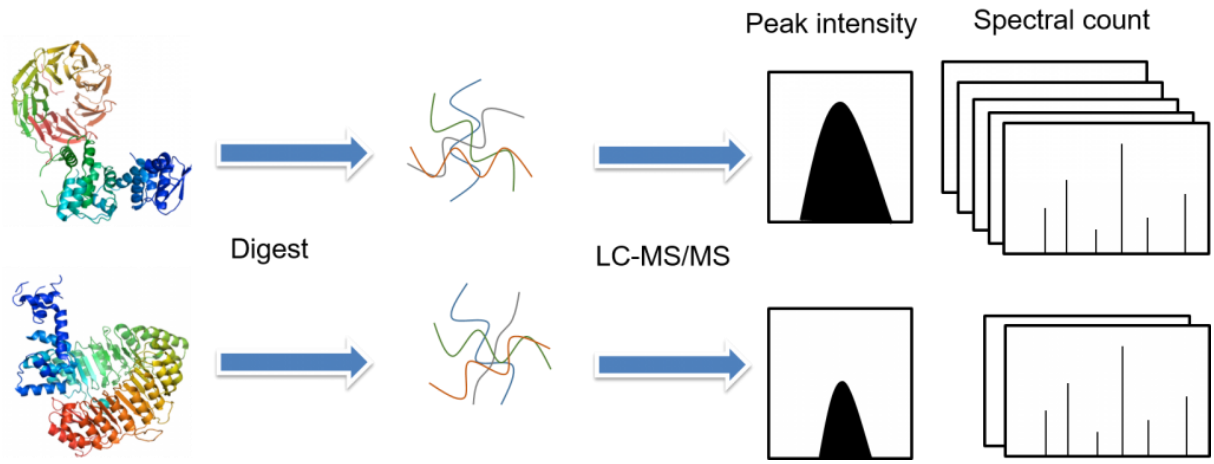

Figure I.8 - Label-Free Quantification (LFQ). Shows the principle of both peak intensity and spectral counting strategies for LFQ. Adapted from Creative Proteomics Blog Webpage [29].

I.3 Proteomics

I.3.1 What is Proteomics?

Proteomics is the large-scale study of proteins, consisting in the analysis of complex proteomes. the term proteome refers to all the proteins being expressed by cell, tissue, organ, or organism at a given moment [30]. Proteomes change over time due to numerous factors, such as age or disease. This allows for a complex analysis of how protein expression changes during the disease course, for instance, and how such changes affect the biological pathways involved in disease progression.

Proteins are a class of macromolecules essential for life, possessing various functions from participating in biochemical reactions to serving as structural components or messengers. Proteomics is, therefore, a highly complex field, not only due to the large number of proteins being expressed by an organism at any given moment but also because alternative splicing and post-translational modifications (PTMs) introduce new levels of complexity affecting protein function and diversity [30].

Proteomics research has a variety of applications, but by far, and especially regarding cancer, most proteomics studies are aimed at biomarker discovery. As mentioned, the proteome is a dynamic complex system which changes during cancer progression [31]. It is precisely these changes in protein expression that are useful for the identification of potential biomarkers. Furthermore, proteomics studies can assess protein expression, PTMs, protein-protein interactions, and even the actual pathways these proteins are involved in [30].

I.3.2 Mass Spectrometry-based Top-down and Bottom-up Proteomics

There are two main approaches for identifying and characterising proteins through mass spectrometry, these being top-down and bottom-up proteomics [32]. Top-down proteomics involves the analysis of intact proteins by the mass spectrometer, while bottom-up proteomics focuses on the study of the peptides that form the proteins.

In a Top-down proteomics approach, whole proteins are used as a starting point, and only during the mass spectrometric analysis will they be fragmented. It is from comparing masses of these fragments with pre-existing databases that information about the proteins is obtained. For instance, in a top-down proteomics approach, by comparing the masses obtained from the protein fragments with the expected values present in the databases, it is possible to identify and localise post-translational modifications and mutations in the protein structure [32].

In a Bottom-up proteomics approach, however, proteins are digested into peptides before the mass spectrometric analysis. This digestion is commonly done with Trypsin, which has specific cleavage sites in the protein structure at the carboxyl side of arginine and lysine residues [33]. Protein identification in this approach is made by comparing the masses of the obtained peptides with pre-existing peptide mass databases (Peptide Mass Fingerprint (PMF)), revealing a possible match. This can be used for single isolated proteins or small groups of proteins; however, for more complex samples, such as proteomes, the match possibilities would be too great, rendering this approach ineffective [33]. For such complex samples, measuring peptide mass is insufficient, and tandem mass spectrometric approaches are required.

In tandem mass spectrometry, proteins are digested with one or more enzymes, and the resulting peptides are analysed by high-resolution nano-Liquid Chromatography-Mass Spec (nLC-MS/MS) [34]. The peptides are then fragmented inside the mass spectrometer, allowing the amino acid sequence to be obtained. Using bioinformatics tools, these amino acid sequences are then used to identify the proteins.

I.3.3 FDA-Approved Biomarkers

The Food and Drug Administration (FDA) already has a list of approved protein biomarkers for a variety of carcinomas (**Table I.1**) [35]. Among these, nuclear mitotic apparatus protein 1 (NMP22, or NUMA as it is also referred to) and complement factor-H related protein (hCFHrp) are two protein biomarkers approved for bladder cancer diagnosis [36]. There are other approved tests for bladder cancer diagnosis which do not target protein biomarkers; however, as is the case with NMP22 and hCFHrg, their sensitivities and specificities are not yet sufficient to completely replace the standard method of cystoscopy, being instead used as complementary analysis [37].

NMP22 is a microtubule-binding nuclear matrix protein that plays a vital role in the alignment and segregation of the chromosomes during mitosis by forming the spindle nodes [36]. In urine samples, NMP22 functions well as a urothelial cell death biomarker, is frequently elevated in urine samples of BCa patients. However, due to NMP22 being released by apoptotic cells, which are not exclusive to cancer and may also occur during other processes such as inflammation, infection or haematuria, its specificity as a BCa biomarker is somewhat impaired by the common occurrence of false positives. Test kits such as the NMP22®BC test and NMP-22 BladderChek® have sensitivities of 69% and 58% and specificities of 77% and 88%, respectively [37].

Complement factor-H-related protein, or hCFHrg, is present in BCa cell lines, and has been approved for the follow-up of NMIBC patients. It can be detected in urine using either BTA stat® or BTA TRAK® tests, which have sensitivities of 74% and 65%, and specificities of 77% and 64%, respectively [37]. Once again, the specificity of both tests is affected by the presence of this protein in the blood, which results in common false positives [37].

Table I.1 - List of FDA-approved cancer biomarkers. Adapted from Füzéry, A. K., et al. [35].

Biomarker	Clinical use	Cancer type	Sample
Pro2PSA	Discriminating cancer from benign disease	Prostate	Serum
ROMA (HE4+CA-125)	Prediction of malignancy	Ovarian	Serum
OVA1 (multiple proteins)	Prediction of malignancy	Ovarian	Serum
HE4	Monitoring recurrence or progression of disease	Ovarian	Serum
Fibrin/ fibrinogen degradation product (DR-70)	Monitoring progression of disease	Colorectal	Serum
AFP-L3%	Risk assessment for development of disease	Hepatocellular	Serum
Circulating Tumor Cells (EpCAM, CD45, cytokeratins 8, 18+, 19+)	Prediction of cancer progression and survival	Breast	Whole blood
p63 protein	Aid in differential diagnosis	Prostate	FFPE tissue
c-Kit	Detection of tumors, aid in selection of patients	Gastrointestinal stromal tumors	FFPE tissue
CA19-9	Monitoring disease status	Pancreatic	Serum, plasma
Estrogen receptor (ER)	Prognosis, response to therapy	Breast	FFPE tissue
Progesterone receptor (PR)	Prognosis, response to therapy	Breast	FFPE tissue
HER-2/neu	Assessment for therapy	Breast	FFPE tissue
CA-125	Monitoring disease progression, response to therapy	Ovarian	Serum, plasma
CA15-3	Monitoring disease response to therapy	Breast	Serum, plasma
CA27.29	Monitoring disease response to therapy	Breast	Serum
Free PSA	Discriminating cancer from benign disease	Prostate	Serum
Thyroglobulin	Aid in monitoring	Thyroid	Serum, plasma
Nuclear Mitotic Apparatus protein (NuMA, NMP22)	Diagnosis and monitoring of disease (professional and home use)	Bladder	Urine

I.4 Urine Proteomics

I.4.1 Urine as a Biofluid

Urine is produced in the kidneys as one of the primary forms of excretion of metabolic wastes and toxins from the body [38]. As one of the most important excretory organs in the human body, the kidneys are responsible for blood plasma filtration. Due to cellular metabolism, a lot of by-products are created, including nitrogen-rich species such as urea and creatinine, which must be filtered out of the bloodstream and expelled from the body [38].

As stated previously, the main focus of clinical proteomics studies has mostly been biomarker discovery, with the intent of establishing a protein or a group of proteins as an early indicator of a particular disease [39]. To this end, urine is an ideal candidate as a biological sample, with one of its main advantages being its non-invasive collection. Because of its nature, urine is much easier to collection than blood and can be collected in a non-invasive manner, at any moment of the day in large volumes. The main inconvenience for urine samples is that some protein biomarkers that can be present in the tissues or blood may not be present in urine.

I.4.2 Urine Sample Handling in Proteomics-Based Mass Spectrometry

I.4.2.1 Sample Collection

Sample handling is a crucial part of any proteomics study, with different samples requiring different methods to isolate and store the proteome in a stable manner, so it can later be analysed. Sample handling must be a straightforward process to minimize systematic and random errors between analysis.

Sample handling starts at the moment of sample collection. In the case of urine samples, current urinalysis may call for different urine collection types, including, but not limited to, the first morning urine (first morning void), the second-morning urine, and the 24-hour urine collection [40]. Each urine type provides different information and can be useful for a certain analysis. However, for biomarker discovery in proteomics, the second void morning urine is becoming the standard sample collected [41]. By collecting the second void urine of the morning, most of the contaminants accumulated during the night in the urine will be gone, such as dead epithelial cells and overgrown bacteria. This way, the second-morning urine provides a more accurate look at the standard proteome of the patient.

Once collected, the sample must be preserved until it can be treated and analysed. For urine samples, the main concern is bacterial growth, as this can alter the proteome that will be analyzed. To prevent this, bacterial growth inhibitors are usually added to the samples [42].

I.4.2.2 Sample Storage

Currently, a standard urine sample storage protocol for proteomics consists of an initial 5000x g centrifugation to remove any epithelial cells that may be present in urine, all the while avoiding the risk of cell lysis, which would undoubtedly affect the proteome and freezing the samples at a temperature range from -20°C to -80°C [43]. This methodology, however, presents some evident drawbacks, namely the power usage requirements to keep the freezers working non-stop, and the necessity for backup power supplies in case of power failures. Another drawback, and potentially the most important one, is that freeze, and thaw cycles influence the proteome stability, resulting in noticeable changes after 5 cycles [44].

I.4.2.3 Sample Treatment for Mass Spectrometry

A typical bottom-up proteomics workflow will generally involve the following steps: (I) Sample collection, (II) Sample preparation, which entails the isolation and purification of the proteome, and its reduction, alkylation, and enzymatic digestion, and lastly, (III) the analysis of the proteome with the mass spectrometer [33]. Having already covered the sample collection methods and the steps taken to preserve the proteome until its analysis, the next step is the sample preparation.

I.4.2.3.1 Protein Isolation and Purification

Another methodology that can be used to separate and purify the proteome from the urine is ultrafiltration. Ultrafiltration uses a membrane to separate proteins from all the other components of the sample. One of these ultrafiltration protocols is Filter Aided Sample Preparation (FASP) which has been developed for bottom-up proteomics [46]. FASP uses a membrane with a molecular weight cut-off ranging from 10 kDa to 30 kDa. This membrane separates the proteins from metabolites and other smaller components present in the already centrifuged urine through multiple washing steps.
Furthermore, it is also possible to perform the reduction, alkylation, and enzymatic digestion of the proteins, all in the same membrane, resulting in the collection of the purified peptides at the end. This methodology has been further enhanced using ultrasounds to speed up the entire process, reducing the total time from 1 day to just 3 hours in total [43]. This is mainly due to the acceleration of the protein digestion step, in which trypsin is most commonly used, which has reduced its time from overnight to 4 minutes [43].

I.4.2.3.1 Protein Quantification

The last step in preparing the samples for analysis in the mass spectrometer is peptide quantification. This is also an important step that must be considered, as it is crucial for label-free

quantification later on. In essence, the samples being analysed have the same concentration, so the obtained mass spectra may be compared quantitatively. Furthermore, in the protein digestion step, it is also essential to know how much protein is in the sample to add the appropriate trypsin quantity. Various methods can be used for protein quantification, which we can split into two categories: colourimetric and UV absorbance methods [47]. Starting with the later ones, the simplest of the UV absorbance methods is the measurement of UV absorbance at 280 nm. This method is based on the characteristic absorbance peak displayed by proteins in the 280 nm region. This is due to amino acids with aromatic rings, which have a maximum absorbance at 280 nm. Tyrosine and Tryptophan are the main contributors to this peak. Here lies the first drawback of this method, as it relies on the presence of these amino acids in the composition of the protein. However, if the protein does not contain any of these residues in its primary structure, this method cannot quantify it accurately [47].

As for colourimetric methods, the Bradford, Lowrey, or BCA assays are some of the most widely used for protein quantification. Taking, for instance, the Bradford assay as an example. This method relies on the Coomassie Brilliant Blue G-250 dye binding to the protein and the shift in absorbance peak that ensues. , the dye binds to the proteins through electrostatic interactions with basic amino acids and through hydrophobic interactions [47]. This bond stabilises the anionic form of the dye, shifting the absorbance maximum from 465 nm to 595 nm. However, this method has some disadvantages, such as many interfering compounds that may be present in the samples, altering the results. Another major drawback common to all dye-binding colourimetric assays is the loss of the sample used in the quantification. This is more problematic when working with scarce or valuable samples, for instance, biological samples from patients, which reflect the patient's proteome at the time of collection.

Recently, a new method for proteome quantification was proposed using tryptophan emission, which avoids some of these drawbacks [48]. Briefly, this method is based on exiting the proteins with UV light and the measurement of the Tryptophan fluorescence. This fluorescence can then be correlated to the protein abundance in the sample. To accurately quantify one purified protein in a solution with this method, it is necessary to know the protein's structure to know precisely how many tryptophan residues it has in its primary structure. This way, it can

accurately correlate the tryptophan fluorescence with the protein abundance [48]. However, in a complex proteome, it is impossible to know precisely which proteins are present in the sample before the mass spectrometry analysis. So the protein's structures are also unknown.

Interestingly, it is described that in a complex proteome composed of hundreds of different proteins, the differences in tryptophan residue abundance from protein to protein tend to even themselves out. Thus, it is possible to calculate a ratio between Tryptophan fluorescence and protein abundance for a specific sample type. Without adding dyes, the sample can be recovered after the quantification. This methodology already avoids some of the pitfalls of previously mentioned methods, such as sample loss.

II

OBJECTIVES

In this work our goal was to address some of the challenges currently faced with the current approach to proteomics-based mass spectrometry, more specifically when it comes to protein quantification methods and urine sample collection. To this end, some of the goals of this work are as follows:

(I) - Test and adapt the Tryptophan Fluorescence Assay developed by Wiśniewski et al. 2015 [48] for protein quantification in urine samples.

(II) - Assess the protein loading and extraction capabilities of different filter types.

(III) - Integrate this new proteome collection method in our laboratory's current protocol for proteomics-based mass spectrometry.

III

MATERIALS AND METHODS

III.1 Experimental Workflow Design

The Tryptophan fluorescence assay in urine samples was optimized using the experimental design as depicted in **Figure III.1**. Urine samples were collected from healthy volunteers and underwent initial processing to isolate and quantify the proteome using the Bradford assay. The tryptophan fluorescence was then measured and the ratio of tryptophan fluorescence to protein abundance was calculated to evaluate the effectiveness of the assay.

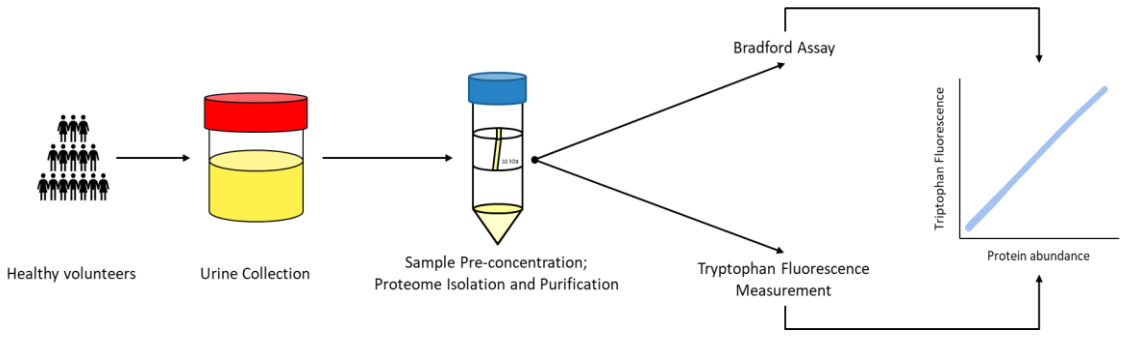

Figure II.1 – Tryptophan Assay experimental design. The proteomes from healthy volunteers were collected and treated. Protein abundance was determined through Bradford Assay, while the tryptophan Fluorescence was also measured.

The proteome collection method depicted in **Figure III.2** involved the use of a syringe filter. Urine samples were passed through the filter, which served to isolate the proteome within

the membrane. The extracted proteome was then subjected to FASP digestion, and the effectiveness of proteome digestion directly on the filter was also tested. These procedures were conducted to optimize the proteome collection and digestion process.

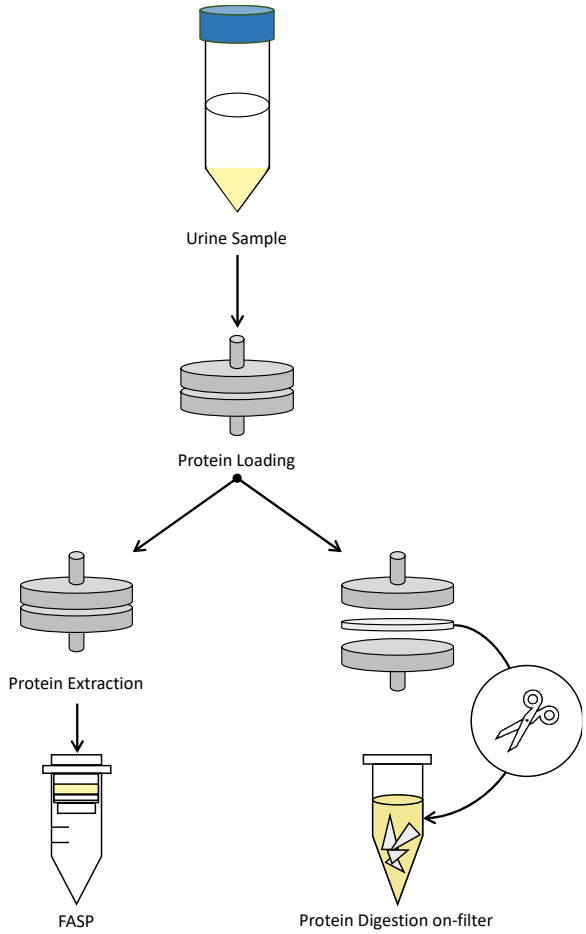

Figure III.2 – Experimental design for filter proteome collection and digestion. Urine proteome was loaded onto the filters using a syringe. Tests were done to try to extract the proteome form the filter membrane and also to digest the proteome directly in the membrane.

III.2 Urine Collection and Storage

Reagents: Boric Acid (Panreac AppliChem)

Apparatus: 100 mL urine collection tube; 10 mL serological pipette; 50 mL Falcon (Abdos); 10 mL Falcon (Abdos); Med. Instruments Centrifuge MPW-352 (Rotor 11457); FastPette Plus (Labnet).

Procedure: Second-morning urine samples were collected from volunteers upon receiving a signed consent form for using their samples in the present study. The second-morning urine was collected mid-stream into a tube with boric acid 20 mM. The mid-stream collection helps to reduce the bacterial and epithelial cell presence in the samples. Furthermore, boric acid is a preservative agent to prevent bacterial growth. Upon arrival, the urine samples were split between two 50 mL Falcons, which were then centrifuged at 5000x g for 20 minutes to remove cellular debris. After centrifugation, the supernatants were collected, split into 10 mL aliquots, and stored at -60 ºC.

III.3 Protein Quantification (Bradford Assay)

Reagents: Bradford Reagent (Sigma-Aldrich); BSA (Sigma-Aldrich); MQH2O.

Apparatus: 96-well plate; Centrifuge and vortex (Sky-Line); ClarioStar Spectrophotometer (BGM LABTECH).

Procedure: A BSA stock working solution was prepared with a concentration of 100 ug/mL of BSA. This solution was then stored in 500 mL aliquots at -20 ˚C. For the assay, a calibration curve was done in duplicate, with the standards prepared according to Table III.1. 150 µL of each standard was loaded onto the 96-well plate, and 150 µL of Bradford reagent was added to each. The urine samples were diluted to ensure their protein levels were inside the linear range of the calibration curve (5-20 µg/mL). After this, they were loaded onto the 96-well plate following the same steps taken with the BSA standards. Lastly, the absorbance was measured at 596 nm using the CalrioStar spectrophotometer.

Table III.1 - Preparation of BSA standards for calibration curve to be used in the Bradford Assay.

Concentration (µg/mL)	*Volume of BSA 100 µg/mL (µL)*	*Volume of MQH2O (µL)*
0	0	400
1	4	396
5	20	380
10	40	360
15	60	340
20	80	320
25	100	300

III.4 Protein Quantification (Tryptophan Fluorescence Assay)

Reagents: L-Tryptophan 99% (Alfa Aesar); Ammonium bicarbonate (AmBic)(Fluka); Urea (Sigma); MQH2O.

Apparatus: Vivacon 500 (10 000 Da MWCO) Hydrosart (Sartorius); Med. Instruments Centrifuge MPW-352 (Rotor 11457); Centrifuge and vortex (Sky-Line); black base quartz 96-well plate ; ClarioStar Spectrophotometer (BGM LABTECH).

Procedure: A stock working solution was prepared with a concentration of 0.010205 µg/µL of tryptophan in Urea 8M and AmBic 25 mM. This solution was stored in 500 µL aliquots at -20 ˚C. A calibration curve was done using the stock tryptophan solution, according to the Table III.2. The standards were done using successive dilutions. For this assay, urine was pre-concentrated using a Vivacon 500 (10 kDa MWCO). A 10 mL aliquot was transferred to the top portion of the Vivacon 500 and centrifuged at 5000x G for 20 minutes. After this circa 300 µL remain in the top portion of the Vivacon tube. At this point 2 sample washes consisting of the addition of 1 mL of MQH2O and centrifugation for 5000x G for 5 minutes each were done. At the end, around 300 µL remained in the top of the Vivacon tube, which were then collected for analysis. This pre-concentration step serves as a way to concentrate the protein form a 10 mL sample into a small volume, as well as a way to eliminate possible contaminants in the sample, namely the free tryptophan present in urine. Urine samples were then diluted in order to ensure that their protein contents were inside of the linear portion of the calibration curve. These dilutions were done with Urea 8M and AmBic 25 mM. Both the calibration curve standards and the samples were loaded onto the quartz-based 96-well plate, adding 75 µL to each well. Fluorescence was measured at 280 nm using the ClarioStar spectrophotometer.

Table III.2 - Preparation of Tryptophan standards for calibration curve to be used in the Tryptophan Fluorescence Assay.

Tube	*Tryptophan concentration (µg/µL)*	*Volume of Tryptophan Stock (µL)*	*Volume of Urea solution (µL)*
A	0	0	200
B	1.99E-05	200 µL of tube C solution	200
C	3.99E-05	200 µL of tube D solution	200
D	7.97E-05	200 µL of tube E solution	200
E	1.59E-04	200 µL of tube F solution	200
F	3.19E-04	200 µL of tube G solution	200
G	6.38E-04	200 µL of tube H solution	200
H	1.28E-03	200 µL of tube I solution	200
I	2.55E-03	200 µL of tube J solution	200
J	5.10E-03	200	200

III.5 Protein Filter Loading

Reagents: Sodium citrate tribasic dihydrate (Sigma-Aldrich); Citric Acid (Panreac); Trizma base (Sigma-Aldrich); MQH2O.

Apparatus: 5 mL syringe (Braun); Chromafil GF/RC-20/25 disposable syringe filters (Macherey-Nagel); Chromafil GF/RC-45/25 disposable syringe filters (Macherey-Nagel); Nylon membrane filter, 25 mm (CHMLAB GROUP); HYDROSART membrane, 10 000 Da MWCO, 25 mm (Sartorius); Ultracel Membrane, 10 000 Da MWCO, 25 mm (Milipore); RE-usable Syringe Filter Holders, 25 mm (Satrorius)

III.5.1 Chromafil GF/RC and Nylon membranes

Procedure: A buffer solution of citric acid was prepared at a concentration of 0.1M, and a pH of 3, as described in Jesus, J. R. et al. 2018 [49]. Before the urine sample, 1 mL of MQH2O was passed through the disposable syringe filter, using a 5 mL syringe. 0.5 mL of citric acid buffer 0.1 M, pH 3, was added to 1 mL of urine sample. This 1.5 mL solution was passed through the disposable syringe filter 5 times, using a 5 mL syringe. After these 5 passes, the syringe is used to pass air through the filter, as a way to remove as much volume as possible. The flowthrough obtained after the 5 passes through the filter was then quantified through the Bradford assay, to assess protein retention in the filter.

III.5.2 Hydrosart and Ultracel membranes

Procedure: The filter was assembled by loading 1 membrane with 2 O-rings into the reusable membrane capsule. The membrane was placed inside the capsule with the shiny side faced upwards, as indicated in the user manual. 1 mL of MQH2O was passed through the filter, using a 5 mL syringe. Afterwards, 1 mL of urine was passed once through the filter using a 5 mL syringe. Several passes of air through the filter with the syringe were required to remove as

much of the liquid volume from the filter. The flowthrough was collected and then quantified through Bradford in order to assess the protein retention of the membranes.

III.6 Protein Extraction from filter

Reagents: Trizma base (Sigma-Aldrich); Ammonium bicarbonate (AmBic)(Fluka); MQH2O.

Apparatus: Peristaltic pump ISM828 (ISMATEC); Chromafil GF/RC-20/25 disposable syringe filters (Macherey-Nagel); Chromafil GF/RC-45/25 disposable syringe filters (Macherey-Nagel); MNY Nylon membrane filter (chm by CHMLAB GROUP); HYDROSART© membrane, 10 000 Da MWCO, 25 mm (Sartorius); Ultracel Membrane, 10 000 Da MWCO, 25 mm (Milipore); Re-usable Syringe Filter Holder, for 25 mm membrane filters (Sartorius).

III.6.1 Chromafil GF/RC and Nylon membranes

Procedure: A buffer solution of Tris-HCL 50 mM was prepared with a pH of 11, in accordance with the methodology firstly developed by Jesus, J. R. et al. 2018 [49]. In order to extract the protein, 0.6 mL of Tris-HCL 50 mM, pH11 was passed through the filter using a peristaltic pump. This process was repeated 4 times for each filter, resulting in 4 extraction fractions from each one. The extraction fractions were all quantified through Bradford to assess protein extraction from the filters.

III.6.2 Hydrosart and Ultracel membranes

First approach: The filter capsule was opened, and the membrane was inverted. 1 mL of AmBic 12.5 mM was passed once through the filter, this time passing through the membrane in the opposite direction. The flowthrough was then quantified by Bradford assay to assess protein extraction from the filters.

Second approach: The filter capsule was opened, and the membrane was transferred into a 1.5 mL Eppendorf. 1.2 mL of AmBic 12.5 mM was then added to the Eppendorf to completely cover

the membrane inside. The Eppendorf was shaken and vortexed for 5 minutes and then mounted on a cup-horn ultrasound system for 2 minutes at 25% amplitude. The flowthrough was then quantified by Bradford assay to assess protein extraction from the filters.

III.7 Protein Digestion in filter

Reagents: Ammonium bicarbonate (AmBic)(Fluka); MQH2O; Ditiotreitol (DTT) (Alfa Aesar); Iodoacetamide (IAA) (Sigma); Trypsin Protease MS-Grade (Thermo Scientific); MS-Grade acetonitrile (Carlo Erba Reagents); Formic acid LC-MS Grade (Fisher Chemical)

Apparatus: Refrigerated Micro-Centrifuge (LAB-NET Prism); Speed Vac with vacuum concentrator centrifuge (Univapo 150 ECH) with a refrigerated aspirator vacuum pump (Unijet II); Mini incubator (Labnet).

Procedure: The filter capsule was opened, and the membrane was carefully removed. The membrane was then cut into small pieces using a pair of scissors and tweezers. The pieces were transferred to a 1.5 mL Eppendorf, where the digestion process took place. The first step was the reduction of the proteins, by adding 500 µL of DTT 50 mM AmBic 12.5 mM, and incubating the tube at 37˚C, for 1 hour. The volume was set at 500 µL to ensure the membrane was completely submerged. After this time, the DDT was removed from the tube. The next step was protein alkylation, for which 500 µL of IAA 50 mM AmBic 12.5 mM were added into the tube. The tube was then incubated at room temperature for 45 minutes, shielded from light. Lastly, the IAA was removed from the tube, and 500 µL of Trypsin 1:30 (Trypsin: Protein) 12.5 mM AmBic were added into the tube. Next, cup-horn ultrasounds were applied to the sample for 2 minutes (30 seconds On, 15 seconds Off, cycles) at 25% amplitude. At this point, another 500 µL of Trypsin 1:30 12.5 mM AmBic were added to the sample, and the 2 minutes of ultrasound cycles were repeated. After the ultrasound cycles the sample was centrifuged for 20 minutes at 14 000x g. The supernatant was collected to another tube, and 500 µL of 3% (V/V) acetonitrile (ACN) containing 0.1% (V/V) formic acid (FA) were added to the membrane pieces, which were centrifuged once again for 20 minutes at 14 000x g. After this, the 500 µL were collected and added to the previously collected supernatant. At this point, small debris may still be present in the tube, due to the membrane being cut, and the ultrasonic energy applied. To avoid this, the supernatant was once again centrifuged at 14 000x g for 20 minutes, allowing for its

collection and discarding of the debris. Once collected, the peptide solution was lyophilised using the speed vac, and then stored at - 20˚C, awaiting LC-MS analysis.

III.8 Filter Aided Sample Preparation (FASP)

Reagents: Ammonium bicarbonate (AmBic)(Fluka); Urea (Sigma); MQH2O; Ditiotreitol (DTT) (Alfa Aesar); Iodoacetamide (IAA) (Sigma); Trypsin Protease MS-Grade (Thermo Scientific).

Apparatus: Refrigerated Micro-Centrifuge (LAB-NET Prism); Speed Vac with vacuum concentrator centrifuge (Univapo 150 ECH) with a refrigerated aspirator vacuum pump (Unijet II); Mini incubator (Labnet); ULTRASONS.

Procedure: For Filter Aided Sample Preparation, the protein was first loaded onto the membrane. To this end, 1 mL of urine was passed through the membrane in three separated volumes of 400 ul, 400 ul and 200 uL. After protein loading, the sample was washed by adding 200 µL of 8M Urea 25 mM AmBic solution and centrifuged at 14 000x G for 20 minutes. Following this, the proteins were reduced by adding 200 µL of 50 mM DTT 8M Urea 25 mM AmBic solution and applying ultrasounds for 4 minutes at 25% amplitude. The FASP system was then centrifuged for 20 minutes at 14 000x g. Then a washing step was done to ensure all the excess DTT was removed. This was done by adding 200 µL of 25mM AmBic, and centrifugating for 20 minutes at 14 000x g once again. Next, the proteins were alkylated by adding 100 µL of 50 mM IAA 8M Urea 25 mM AmBic, and ultrasounds were also applied for 4 minutes at 25% amplitude. Once again, the FASP was then centrifuged at 14 000x g for 20 minutes. The FASP was then washed twice, by adding 200 µL of 25mM AmBic and centrifuged at 14 000x g. Finally, the proteins were digested by adding 100 µL 1:30 trypsin in 12.5 mM AmBic solution, and the reaction was sped up by applying, once again, ultrasounds for 4 minutes at 25% amplitude. After digestion, the samples were centrifuged for 20 minutes at 14 000x g. This was followed by adding 100 µL of 12.5mM AmBic solution and a 14 000x g centrifugation for 20 minutes.

This previous step was then repeated. Lastly, the peptides were collected to a new tube and were then lyophilized, using the speed vac. Once lyophilized, the peptides were stored at -20 ℃, while awaiting LC-MS analysis.

III.9 Peptide Quantification

Reagents: Quantitative Colorimetric Peptide Assay (Thermo Scientific); MQH2O; MS-Grade acetonitrile (Carlo Erba Reagents); MS-Grade formic acid (Fluka); 96-well plate.

Apparatus: ClarioStar Spectrophotometer (BMG LABTECH); 96-well plate; Centrifuge and vortex (Sky-Line); Ultrasonic bath (Elma).

Procedure: For peptide quantification, the previously lyophilized samples must first be resuspended. For this sample resuspension, 200 mL of 3% (V/V) acetonitrile (ACN) containing 0.1% (V/V) formic acid (FA) were added to the tubes, which were then vortexed for 5 minutes and another 10 minutes of ultrasonic bath at 100% amplitude. At this stage the resuspended peptides were quantified through the quantitative colorimetric peptide assay [50]. A calibration curve was prepared using the peptide standard included in the kit. This curve was done with serial dilutions of the peptide standard, as seen in the **Table III.3**. For the assay, the samples and calibration curve were loaded onto the 96-well plate in duplicate, by adding 20 µL of solution in each well. Following this, 180 µL of the reagent provided in the quantification kit, was added to each well, and the plate was incubated at 37 ºC for 15 minutes. Lastly, the asorbance at 480 nm was measured using the ClarioStar spectrophotometer.

Table III.3 - Preparation of peptide standards for calibration curve to be used in the Quantitative Colorimetric Peptide Assay.

Tube	concentration (µg/µL)	Volume Peptide Stock (1 mg/mL) (µL)	Volume of H2O (µL)
A	0	0	150
B	15.625	75 mL of tube C solution	75
C	31.25	75 mL of tube D solution	75
D	62.5	75 mL of tube E solution	75
E	125	75 mL of tube F solution	75
F	250	75 mL of tube G solution	75
G	500	75	75

III.10 LC-MS/MS Analysis

Reagents: MS-Grade acetonitrile (Carlo Erba Reagents); MS-Grade formic acid (Fluka).

Apparatus: Analytical column µPAC™ C18, 50 cm (Thermo Fisher Scientific); UHR-QqTOF IMPACT HD (Bruker); trap-column µPAC™ C18 (Thermo Fisher Scientific).

Procedure: 400 ng of peptides were loaded onto a Trap column µPAC™ C18, desalted for 5 min at 3% B (B: 100% (v/v) acetonitrile0.1% (v/v) FA) at a flow rate of 15 µL min-1. Then the peptides were separated using an analytical column µPAC™ C18, 50 cm, with a linear gradient at 500 nL min-1 (mobile phase A: aqueous FA 0.1% (v/v); mobile phase B 100% (v/v) acetonitrile and 0.1% (v/v) FA) as shown in Figure III.1.Chromatographic separation was carried out at 35 ºC. MS acquisition was set to cycles of MS (2 Hz), followed by MS/MS (8–32Hz), cycle time 3.0 seconds, active exclusion, exclude after one spectrum, release after 0.5 min. Reconsider precursor if current intensity, previous intensity 3.0 an intensity threshold for fragmentation of 2500 counts. All spectra were acquired in the range 150–2200 Da.

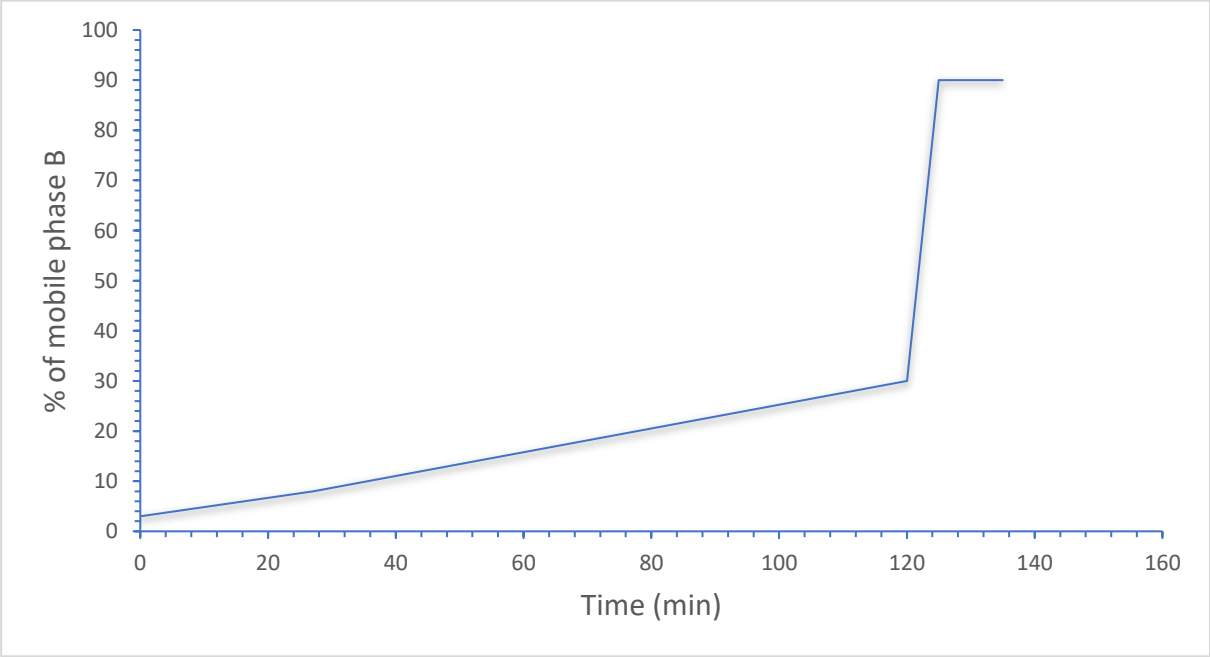

Figure III.1 - Chromatographic gradient. Chromatographic gradient used for the separation of the peptides.

III.11 Bioinformatical Analysis

Label-free quantification was carried out using MaxQuant software V1.6.10.43. All raw files were processed in a single run with default parameters [50]. Database searches are performed using the Andromeda search engine with the UniProt-SwissProt Human Uniprot Proteome database as a reference and a contaminants database of common contaminants. Statistical analysis was performed using Perseus (version 2.0.6.0). In brief, protein group LFQ intensities were log2-transformed to reduce the effect of outliers. To overcome the obstacle of missing LFQ values, missing values were imputed before fitting the models. Log ratios were calculated as the difference in average log2 LFQ intensity values between the two digestion methods tested (two-tailed, Student's t test). A protein was considered statistically significant if its fold change was ≥ 1.5 and FDR ≤ 0.05.

IV

RESULTS AND DISCUSSION

IV.1 Protein Quantification in urine samples (Tryptophan Assay)

Protein quantification is one of the crucial steps in proteomics-based mass spectrometry. It is particularly important in the protein digestion step, where trypsin must be added to the sample at a ratio of 1:30 (trypsin: protein). Bradford assay, which is currently one of the standard methods to quantify protein, is not ideal, as being this a colourimetric assay, all the samples used in the quantification will be lost.

The tryptophan fluorescence assay for protein quantification was recently updated by Wiśniewski et al. 2015, as a way to overcome the problems mentioned above [48]. In brief, it is possible to determine the protein content of a sample based on its tryptophan fluorescence. Proteins owe their fluorescence to three amino acids: tryptophan, tyrosine, and phenylalanine. Among these, tryptophan is the most significant contributor to this fluorescence and has been mostly used for protein structure determination and function. However, Wiśniewski et al. 2015 [48] demonstrated that protein quantification was also accomplished in tissue lysates.

To adapt this method to proteome quantification in urine, samples collected from healthy volunteers were treated and measured with both the Bradford assay and Tryptophan fluorescence. The idea was to determine the ratio between the protein content and the tryptophan fluorescence for urine proteomes. As proteins vary significantly in primary structure,

not all will have the same amount of tryptophan residues, and therefore, not all proteins will contribute equally to the sample fluorescence. However, as shown by Wiśniewski et al. 2015 [48], in complex proteomes, with thousands of proteins, these differences tend to even out, allowing for the determination of the ratio between tryptophan fluorescence and protein content.

Once the urine samples were collected, the cellular debris was removed through a 5000x g, 20-minute centrifugation step. Following this, 10 mL aliquots were pre-concentrated using a ten kDa molecular weight cut-off membrane. In addition to pre-concentrating the sample, it also serves as a washing step to guarantee that only proteins remain in the sample. The Bradford assay was used to quantify the protein, as it is currently our laboratory's standard optimized method for protein quantification.

The samples were diluted in urea 8M, AmBic 25 mM for Tryptophan fluorescence measurements. This ensures that the proteins are denatured, essential as tryptophan fluorescence is highly sensitive to its surroundings. The addition of urea for the quantification does not affect the rest of the proteomics protocol, as it can be easily removed later, with no extra steps added to the current workflow.

Once both assays were concluded and protein and tryptophan contents were known for each sample, a plot was charted relating both measurements. The resulting plot can be seen in Figure IV.1.

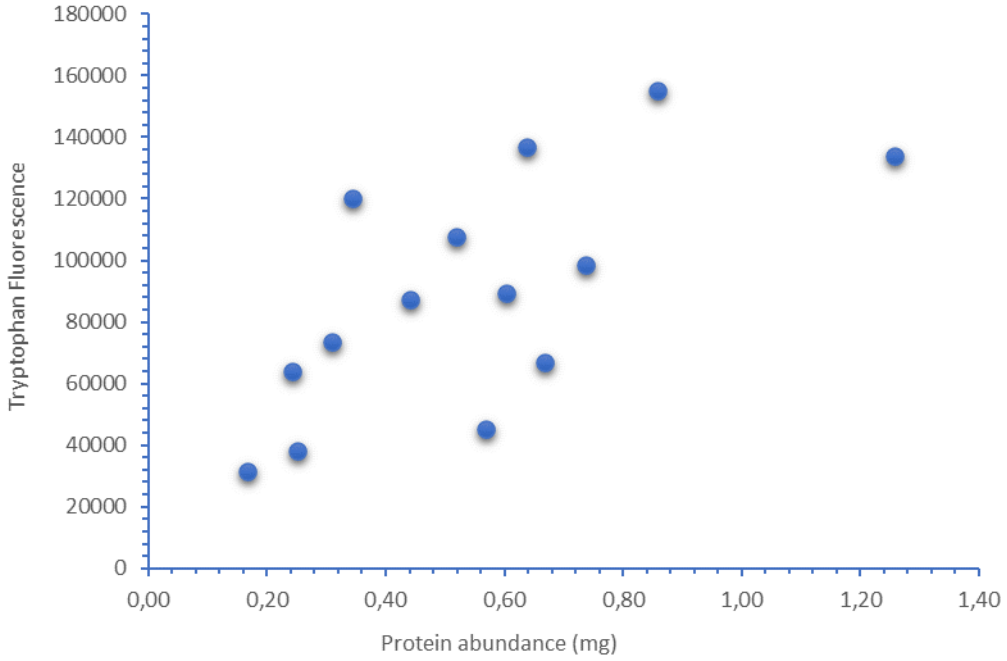

Figure IV.1 - Ratio between Tryptophan fluorescence and protein abundance. The measured tryptophan fluorescence of each sample is plotted against protein abundance of the respective sample, as measured with Bradford assay.

From this analysis, it is possible to observe a tendency of increased tryptophan fluorescence with the increase in protein concentration, however, as shown in **Figure IV.1**, this trend is non-linear or even predictable.

A hypothesis was proposed to explain these results: the free tryptophan in urine must be thoroughly washed from the samples. To test this hypothesis, a urine sample was pre-concentrated using the same ten kDa weight cut-off membrane as before, with another washing step. The disregard-ed washing solution was collected after each sample wash. So, 1 mL of ultrapure water to clean the proteome was added to the top of the membrane, and the sample was centrifuged at 5000x g for 5 minutes. This step was repeated twice for a total of 3 sample

washes. Tryptophan fluorescence was measured in each disregarded washing solution to evaluate how much free tryptophan remained in the isolated proteome. Data presented in **Figure IV.2** shows that tryptophan fluorescence was still one interference after the third washing step. Thus, it was concluded that the experimental protocol using tryptophan in urine requires further optimisation. So it was decided to set the Bradford method as the reference method to quantify proteins for further experiments.

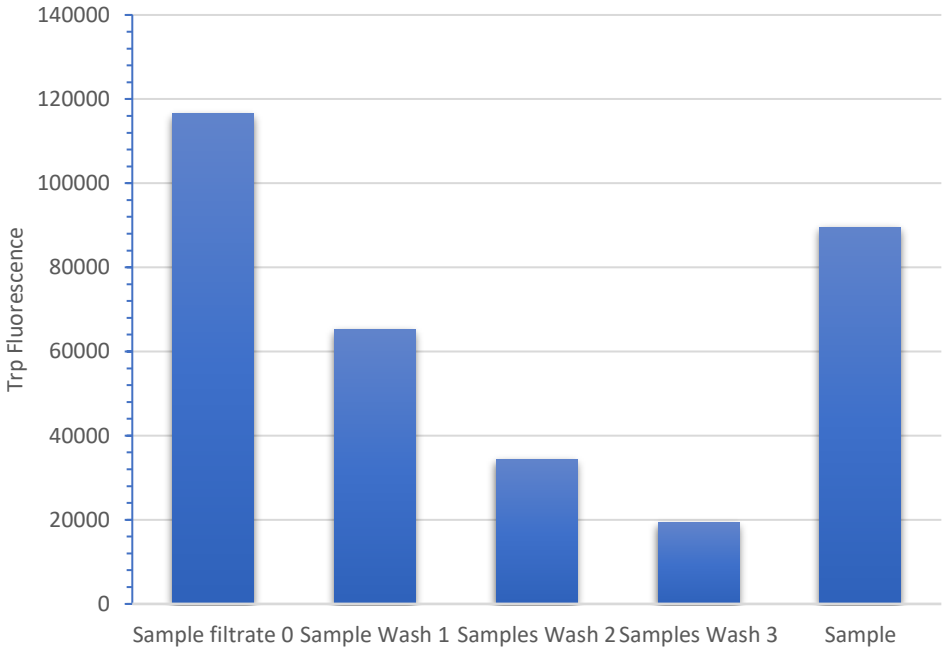

Figure IV.2 - Tryptophan fluorescence in sample washes. Tryptophan fluorescence was measured in the discarded portions of each sample washing step to evaluate how much free tryptophan was still being removed from the sample.

IV.2 Filter Handling

Another significant challenge in mass spectrometry-based proteomics is sample handling and storage. Sample handling must be straightforward to avoid random and systematic errors, and sample storage must be optimised to preserve the proteome. For urine proteomics, the samples must be collected with a preserving agent to avoid bacterial growth in the urine. Furthermore, the samples must be transported to the lab in refrigerated compartments and, after some preparation, stored at -60 ºC or – 80ºC, for further use. This is not ideal, as it increases the costs of sample processing. The freezing also affects the analysis, as the freeze-thaw cycles have been shown to have a detrimental effect on the proteome [44].

A new method for collecting the urinary proteome is proposed to overcome the challenges mentioned above. Syringe filters to such an end were first described in Jesus, J. R. et al. 2018 [49], where the urinary proteome was collected using nitrocellulose membranes. Based on this previous work developed in our lab, our goal was to optimize this collection method with the current mass spectrometry workflow. To this end, various membrane filters were tested to see which would be more appropriate for proteome collection.

The first type of membrane tested was composed of Regenerated Cellulose and Glass Fiber, with a pore of 22 µm and a 25 mm diameter. Before protein loading, 1 mL of ultrapure water was passed through the filters. Following this, 1 mL of urine was collected from a 10 mL sample after removing cellular debris through centrifugation. To the 1 mL of urine, 0.5 mL of 0.1 M citric acid buffer at pH three was added. This 1.5 mL sample was then passed through the filter five times. Protein loading was assessed by quantifying the protein content of the urine sample before and after passing through the filter. To ensure the maximum amount of protein was loaded onto the filters, the same sample was filtered using two filters. In this approach, the sample was passed five times through each filter.

The same protocol for protein loading was also tested in filters with membranes composed of RC/GF, with a 25 mm diameter, but a 0.45 µm pore size, instead of the 0.22 mm, used previously.

Furthermore, Nylon (NY) membranes were also tested, as NY was described as having a high protein binding capacity [50]. As no specific buffer was mentioned to achieve maximum protein binding, different buffers were prepared with the following pHs: 3, 5, 7, and 9. Protein loading was done following the same protocol as before, with the only alteration being that the urine sample was only passed once through each filter.

Lastly, Hydrosart membranes with a ten kDa MWCO and a 25 mm diameter were also used, as these are the same type of membranes used in the FASP protocol. Protein loading was done with no buffer, as Hydrosart membranes do not bind to protein, simply retaining the proteome based on molecular weight. The urine sample was only passed through the filter one time. Protein retention was calculated as described above, and results are shown in **Figure IV.3**.

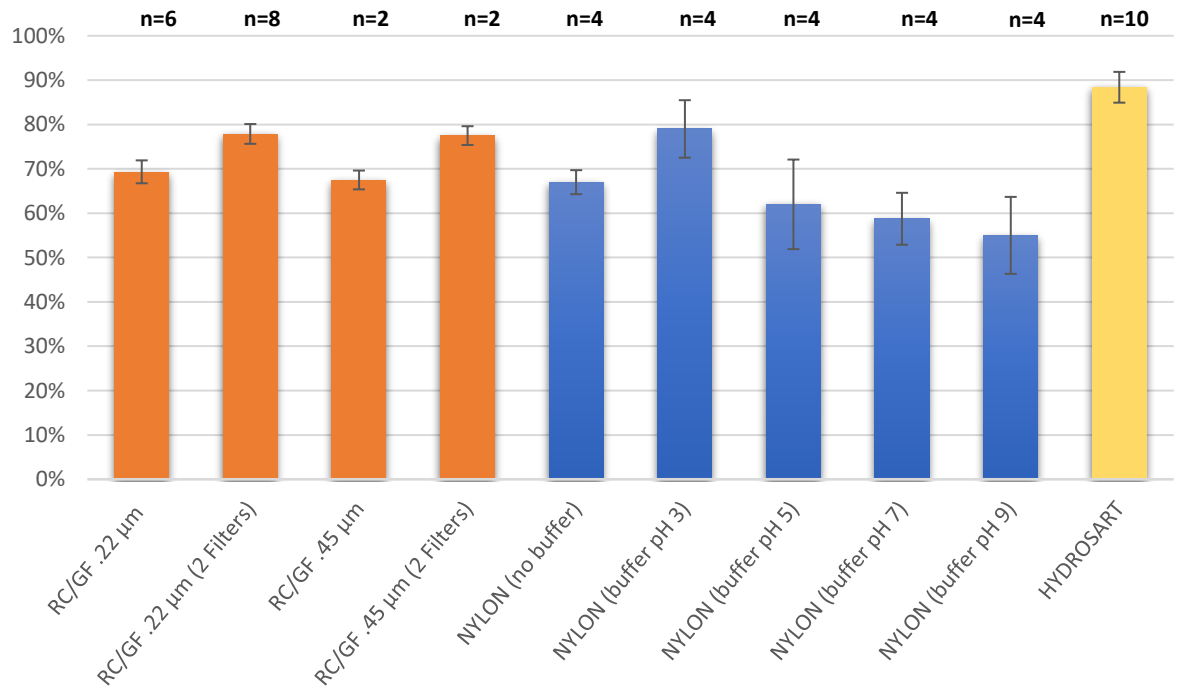

Figure IV.3 - Protein retention across all filters. Protein retention is displayed according to the percentage of total protein in the sample that was retained in the filter..

Between the RC/GF filters, both 0.22 and 0.45 µm pore sizes had similar performances. On average, the 0.45 µm pore membranes had a slightly lower protein retention capacity. However, this difference is negligible, with the filter of 0.22 µm pore diameter retaining 69% of the total protein in the sample, against 68% for the filter with 0.45 µm pore. When using two filters consecutively, however, protein retention increases to 78% of the initial protein concentration in both cases.

As for the NY membranes, the buffer pH was highly influential in protein retention, as expected. Protein retention in the membrane increased at lower pHs, with the best retention obtained at a pH of 3, a 79%. Lower pHs values were not tested as the membrane manufacturer set the lower pH limit to avoid membrane degradation at 3.

However, the highest protein retention was achieved with the Hydrosart membranes, with an average protein retention of 88%. These were the membranes selected for further experiments, as the higher the protein retention is, the more complete the data of the patient´s urinary proteome. Also, the sample treatment was easier as a step of filter conditioning with buffer before protein loading was unnecessary.

IV.3 Protein Filter Extraction

Since the goal was to develop a method to extract the urinary proteome at home by the own patient, the next key aspect to be considered after protein separation and retention was protein extraction. Ensuring the highest protein retention in the filter is obviously a crucial aspect, as it will provide the most complete overview of the patients' urinary proteome. However, good protein retention is linked to good protein extraction.

To this end, protein extraction was tested with different protocols for all the membrane types tested previously apart from nylon, as it was already described in the literature as being very poor [50].

For RC/GF membranes, the extraction was done by passing 0.6 mL of 50 mM Tris-HCL at pH11, with a peristaltic pump. The flow was set so that the total volume took 5 minutes to pass through the filter. For each filter 4 total extractions were done, and the results shown in Figure IV.4, where the total protein extracted from each filter is presented by the sum the total protein of all extracts. Extraction in the RC/GF filters with a 0.45 µm pore size was better than in the 0.22 µm filters. For the 0.45 µm pore filters, was 75% of the loaded protein with one filter, and 78% when using two filters in the separation procedure, while for the 0.22 µm pore filters these numbers were 58% and 60%, respectively. Since these percentages are based on the amount of protein that was retained in the filters, a 75% extraction on the 0.45 µm filters, that had 68% protein retention, represent roughly 51% of the proteome present in the urine. Taking this into account, the best result was a recovery of the 60% of the original urinary proteome, which was obtained when using 2 filter of 0.45 µm pore size during the extraction procedure.

For the Hydrosart© membranes, extraction was first attempted by opening the filter capsule, inverting the membrane, and passing 1 mL of 12.5 mM AmBic. Since these membranes are described as retaining protein solely based on their molecular weight, we hypothesized that the equivalent and opposite force required to load the proteins into the membrane should be enough to extract them. This was not the case, as can be seen in **Figure IV.4**, this extraction method only retrieved 37% of the loaded protein.

Another method for extraction was also tested, where the membrane was removed from its capsule, and placed in a clean tube with 1.2 mL of 12.5 mM AmBic. Then, 4 minutes of cup-horn ultrasound at 25% amplitude was applied, and then a 5-minute vortex. This method proved even less effective than the last one, being only able to extract 9% of the retained protein.

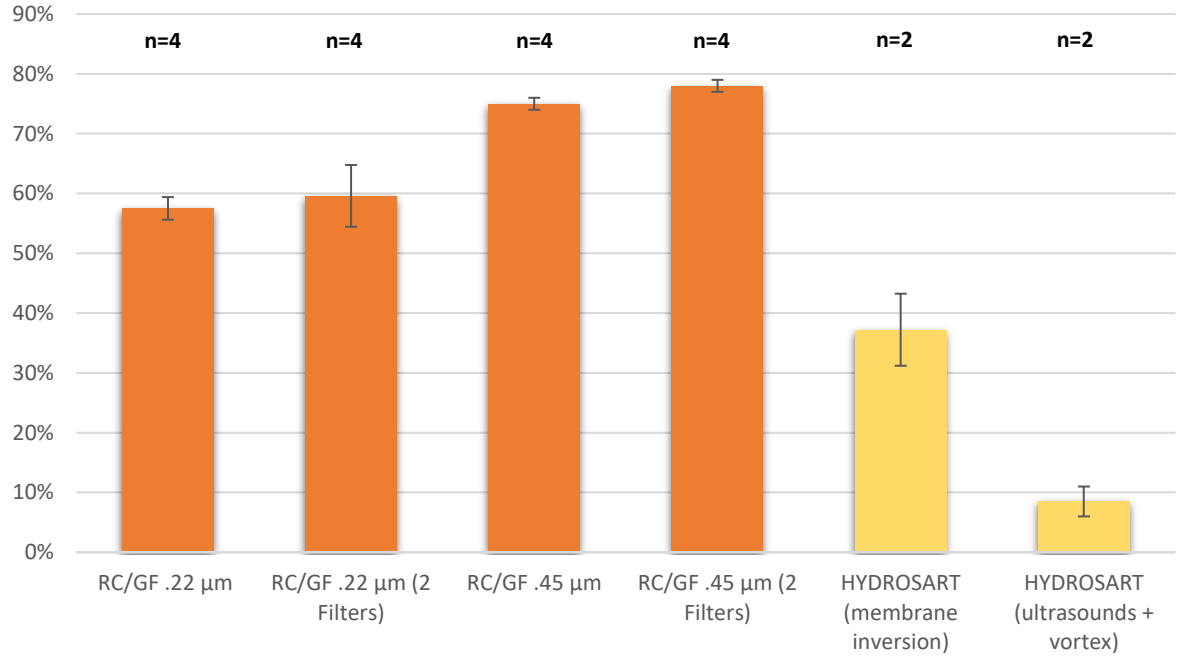

Figure IV.4 - Protein Extraction across different filters. The percentage of protein extracted is based on the total protein retained in each filter, and not the total amount of protein in the urine sample.

IV.4 Optimization of On-filter Protein Digestion

Since quantitative protein extraction from the filter was troublesome, with most of the protein content remaining in the membrane, it was decided to cleavage the proteome on the filter. To this end, the following variables were identified and optimized: Influence of filter type, reproducibility of the method, effect of trypsin concentration and effect of Ultrasound amplitude.

IV.4.1 Effect of filter type (Millipore versus Hydrosart)

As mentioned, the Hydrosart© membranes with a 10 kDa MWCO showed the highest protein retention and therefore were chosen for the following experiments. For the further experiments, both Hydrosart© from Sartorius and Millipore filter membranes with 10 kDa MWCO were acquired as both companies are well known in the filtering market.

Once the urinary proteome was separated, cleavage of proteins was done with the same ultrasonic conditions and the same trypsin-to-protein ratio, 1:6, for both types of filter. The data presented in **Figure IV.5** shows that the number of total proteins identified was almost the same and that, from the common proteins, the majority (higher than 90%) presented with a quantification that was not statistically significantly different. It was concluded that the methodology has similar performance regardless of the filter type, so we selected to the cheaper one for further experiments.

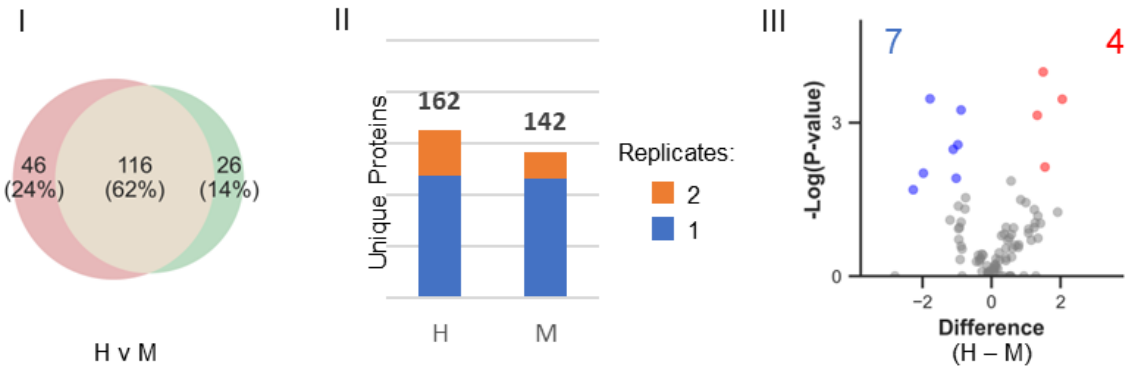

Figure IV.5 - Effect of filter type in analytical performance. Hydrosart© (H) versus milllipore (M) filter types, both with a 10 kDa MWCO. Ultrasonic conditions: Cup-Horn, 25% ultrasonic amplitude, 4 min ultrasonics time (pulses 30 s on-off). 1:6 trypsin to protein ratio. **(I)** Venn diagrams of common and non-common unique proteins. **(II)** Total proteins identified using two technical replicates in each protocol. **(III)** Volcano plot showing variations in protein levels between protocols. Volcano plots are generated using the proteins present in at least 70% of the technical replicates (4 in total, 2 for H and 2 for M). Red and Blue dots and numbers are proteins statistically significantly different while grey dots represent non- statistically significantly different ones. A protein was considered statistically significant if its fold change was ≥ 1.5 and FDR ≤ 0.05.

IV.4.2 Reproducibility: FASP method versus Filter method

In our laboratory, the FASP protocol is regularly used to handle samples for proteomics purposes [43, 51]. This method can be considered universal, and we use it as a reference methodology to compare the reproducibility of our data. For the FASP protocol, **Figure IV.6 (I)** shows that about 80% of the proteins were the same for the inter-day technical replicates and **Figure IV.6 (II)** shows no statistically significant difference between the levels of the proteins present at least 70% of the replicates.

On the other hand, the filtering protocol showed a lower performance. Thus, the number of shared proteins identified was 61%, and the number of proteins found statistically different was 40, over a total of 142. We noted at this stage that the trypsin-to-protein ratio (1:6) used in the protocol presented some problems with the chromato-graphic resolution as many intense trypsin peaks were present in the chromatograms. Therefore, the subsequent experiment was designed to assess this ratio.

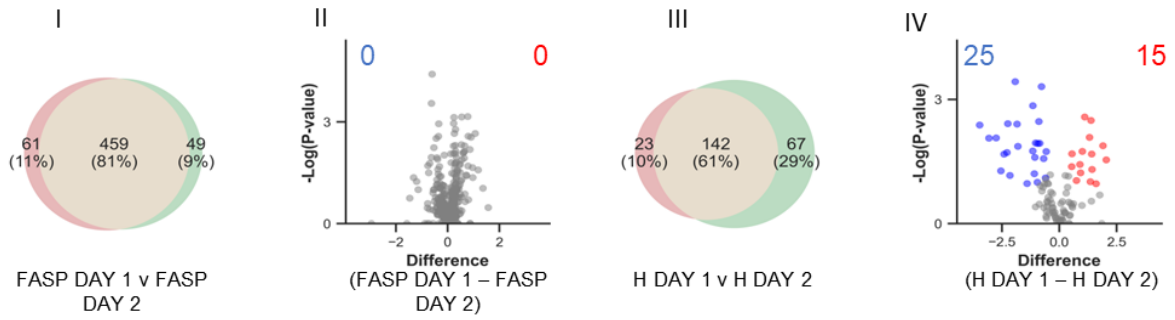

Figure IV.6 - Inter-batch variability of the FASP and Filter protocols. FASP protocol: **(I)**: Venn diagrams of common and non-common unique identified proteins in batch 1 against batch 2, each one made in a different day (n=2 technical replicates for each batch). **(II)**: Volcano plots were generated using the proteins present in at least 70% of the technical replicates. A protein was considered statistically significant if its fold change was ≥ 1.5 and FDR ≤ 0.05. Hydrosart Filter protocol: **(III)**: Venn diagrams of common and non-common unique identified proteins in batch 1 against batch 2, each one made in a different day (n=2 technical replicates for each bath). **(IV)**: Volcano plots were generated using the proteins present in at least 70% of the technical replicates. Red and Blue dots represent statistically significant proteins, while grey dots represent non-statistically significant ones. The numbers in blue and red at the top of the plots indicate the number of statistically significant proteins in that condition. A protein was considered statistically significant if its fold change was ≥ 1.5 and FDR ≤ 0.05. A 10 kDa MWCO Hydrosart filter was used in the filter protocol. Ultrasonic conditions: Cup-Horn, 25% ultrasonic amplitude, 4 min ultrasonics time (pulses 30 s on-off). 1:30 trypsin to protein ratio for the FASP protocol and 1:6 trypsin to protein ratio for the Hydrosart protocol.

IV.4.3 Effect of trypsin amount on protein digestion

It is well known that the amount of trypsin used to perform digestion critically influences quantitative proteomics data [43]. In our previous work with filters, it was shown that an amount of 1:6 was adequate to identify many proteins Jesus, J. R. et al., 2018 [49]. On the other hand, it is also essential to ensure that the amount of trypsin used will not interfere with the quantification process. Therefore, based on our experience, we tested 1:30 and 1:15 trypsin-to-protein ratios. **Figure IV.7** shows that the total numbers of proteins identified are similar with either ratio. Furthermore, quantifying the common proteins, as shown in the volcano plot, rendered a performance in digestion similar for both trypsin-to-protein ratios. Therefore, we decided to use the 1:30 trypsin-to-protein ratio for further experiments.

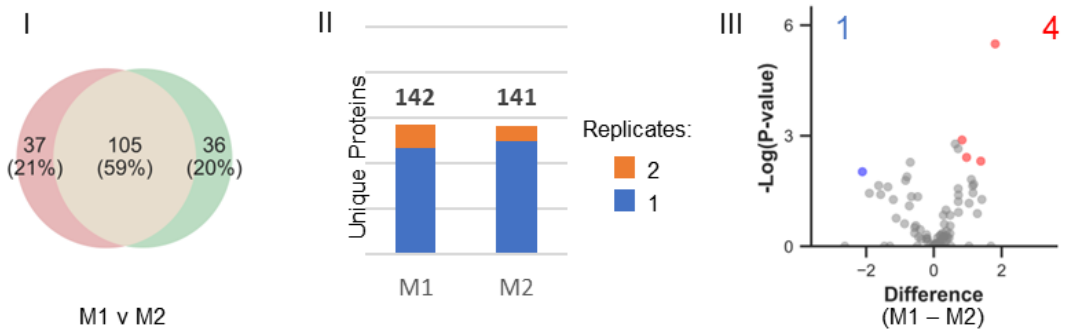

Figure IV.7 - Effect of trypsin amount in filter protein digestion. Milipore filter membranes with a 10 kDa MWCO were used for this experiment. Ultrasonic conditions: Cup-Horn, 25% ultrasonic amplitude, 4 min ultrasonics time (pulses 30 s on-off). M1: 1 shot of trypsin (1:30 trypsin to protein ratio). M2: 2 shots of trypsin (1:30 trypsin to protein ratio added twice. Final ratio 1:15). **(I)** Venn diagrams of common and non-common M1 vs M2 unique proteins. **(II)** Total proteins identified using two technical replicates in each protocol. **(III)** Volcano plot showing variations in protein levels between each protocol. Volcano plots were generated using the proteins present in at least 70% of the technical replicates (4 in total, 2 for M1 and 2 for M2). Red and Blue dots and numbers are proteins statistically significantly different while grey dots represent non-statistically significantly different ones. A protein was considered statistically significant if its fold change was ≥ 1.5 and FDR ≤ 0.05.

IV.4.4 Effect of Ultrasound Amplitude (10% versus 25%)

The ultrasonic amplitude is another parameter that must be carefully selected, as it can lead to the degradation of the filter membranes, resulting in small debris that can block the chromatographic column used in the LC-MS/MS analysis.

The results suggest that amplitudes higher than 25% lead to the fast deterioration of the membranes. Therefore, with selected 25% and 10% ultrasonic amplitudes, **Figures IV.8-(I)** and **IV.8-(II)** show similar digestion performance, as the number of total proteins identified was similar. No differences were found when the digestion levels of common proteins (62%) were assessed using a volcano plot (**Figure IV.8-(III)**). Precisely, our expertise also suggests using amplitudes below 25% only when a clear advantage in terms of sample treatment benefits from such a procedure. This is because future changes in the methodology promoted by sample type modification or modification of filter type are prone to alter digestion performance when low amplitudes are used. Thus, it was decided to maintain 25% as the standard ultrasonic amplitude in further experiments.

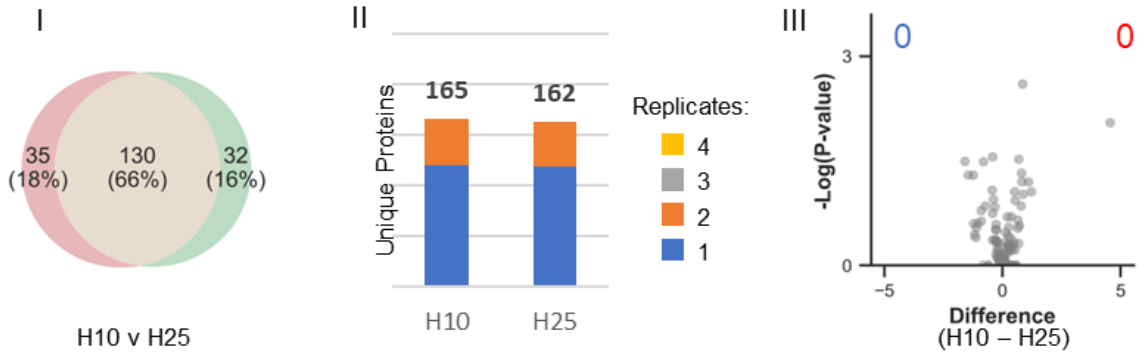

Figure IV.8 - Effect of ultrasonic amplitude in protein digestion. Hydrosart © filter membranes with a 10 kDa MWCO were used for this experiment. Ultrasonic conditions: Cup-Horn, 25% (H25) or 10% (H10) ultrasonic amplitude, 4 min ultrasonics time (pulses 30 s on-off). 1:6 trypsin to protein ratio. **(I)** Venn diagrams of common and non-common H10 vs H25 unique proteins. **(II)** Total proteins identified using two technical replicates in each protocol. **(III)** Volcano plot showing variations in protein levels between each protocol. Volcano plots are generated using the proteins present in at least 70% of the technical replicates (4 in total, 2 for H10 and 2 for H25). Red and Blue dots and numbers are proteins statistically significantly different while grey dots represent non- statistically significantly different ones. A protein was considered statistically significant if its fold change was ≥ 1.5 and FDR ≤ 0.05.

IV.4.5 FASP Method versus Filter Methods.

All the proteomics results obtained with the different conditions using filters assessed in this work were compared with the FASP protocol, as depicted in **Figure IV.9**. These results confirm the validity of the final conditions chosen as optimum, namely: cup-horn device, 4 minutes ultrasonic time (pulses 30 s on-off), 1:30 trypsin to protein ratio, and 25% ultrasonic amplitude.

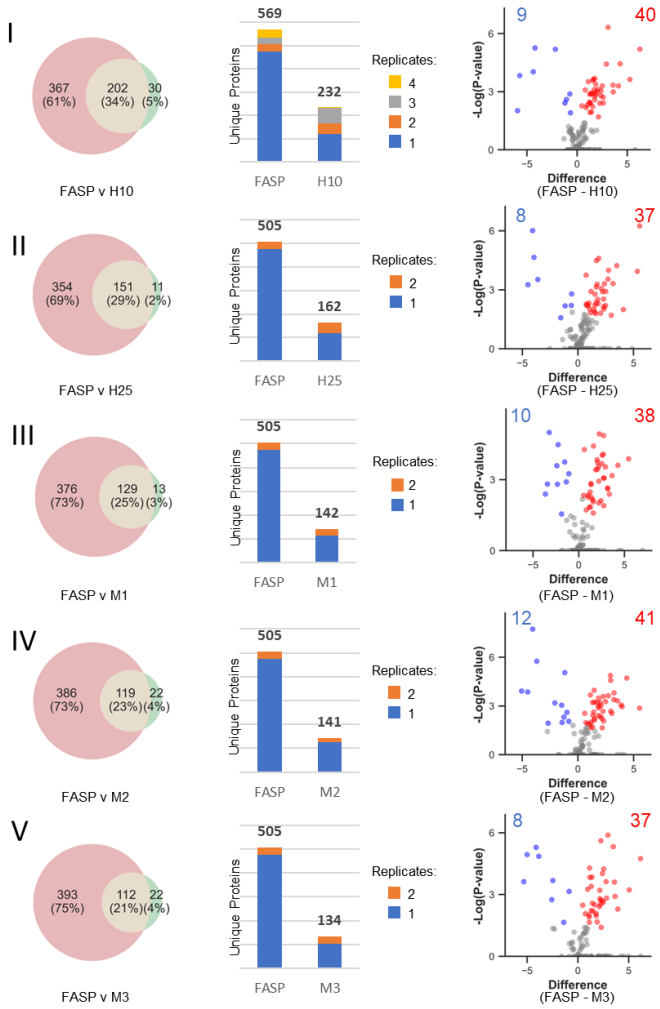

Figure IV.9 - FASP protocol versus Filter protocols. Cup-Horn and 4 min ultrasonic time (pulses 30 s on-off) were used in all the filter protocols tested. **(I)**: H10, Hydrosart membrane, 1:6 trypsin to protein ratio, 10% ultrasonic amplitude. **(II)**: H25, Hydrosart membrane, 25% ultrasonic amplitude, 1:6 trypsin to protein ratio. **(III)**: M1, Millipore filter, 25% ultrasonic amplitude, 1:30 trypsin to protein ratio. **(IV)**: M2, Millipore filter, 25%. ultrasonic amplitude, 1:15 trypsin to protein ratio (2 shots of 1:30). **(V)**: M3, 1:15 trypsin to protein ratio (2 shots of 1:30), overnight digestion at 37 ºC.

IV.5 Proof of Concept: Comparison between non-recurrence and recurrence T1 stage bladder cancer patients

The analytical performance of the filtering protocol was compared with a previous FASP standard protocol developed by our team. To this end, the urine of two groups of patients was used. One group comprised five patients that experienced bladder cancer recurrence, whilst the other group was constituted of 5 patients that have not experienced it. **Figure IV.10** shows the comparison of the results obtained with both protocols. The hierarchical clustering shows a clear separation between both treatments (**Figure IV.10 - (I)**). Also, the grey dots denote many unidentified proteins in the groups of samples treated using the filters protocol when compared with the FASP one. The number of total proteins identified was two times higher in the FASP standard protocol than in the one using filters (**Figure IV.10 - (VI)**), whilst the number of peptides was almost 3 (2.86) times higher (**Figure IV.10 - (V)**).

Furthermore, when a grouping of patients as a function of recurrence versus non-recurrence is done, the FASP protocol can group 4 out of 5 of the patients with no recurrence. Also, all individual replicates cluster together. On the other hand, the clustering obtained with the filters protocol can only group 5 in 10 of the individual replicates correctly, failing in grouping patients with no recurrence. Also, the Pearson correlation diagram shows a low correlation among the sample protocols when all samples are compared.

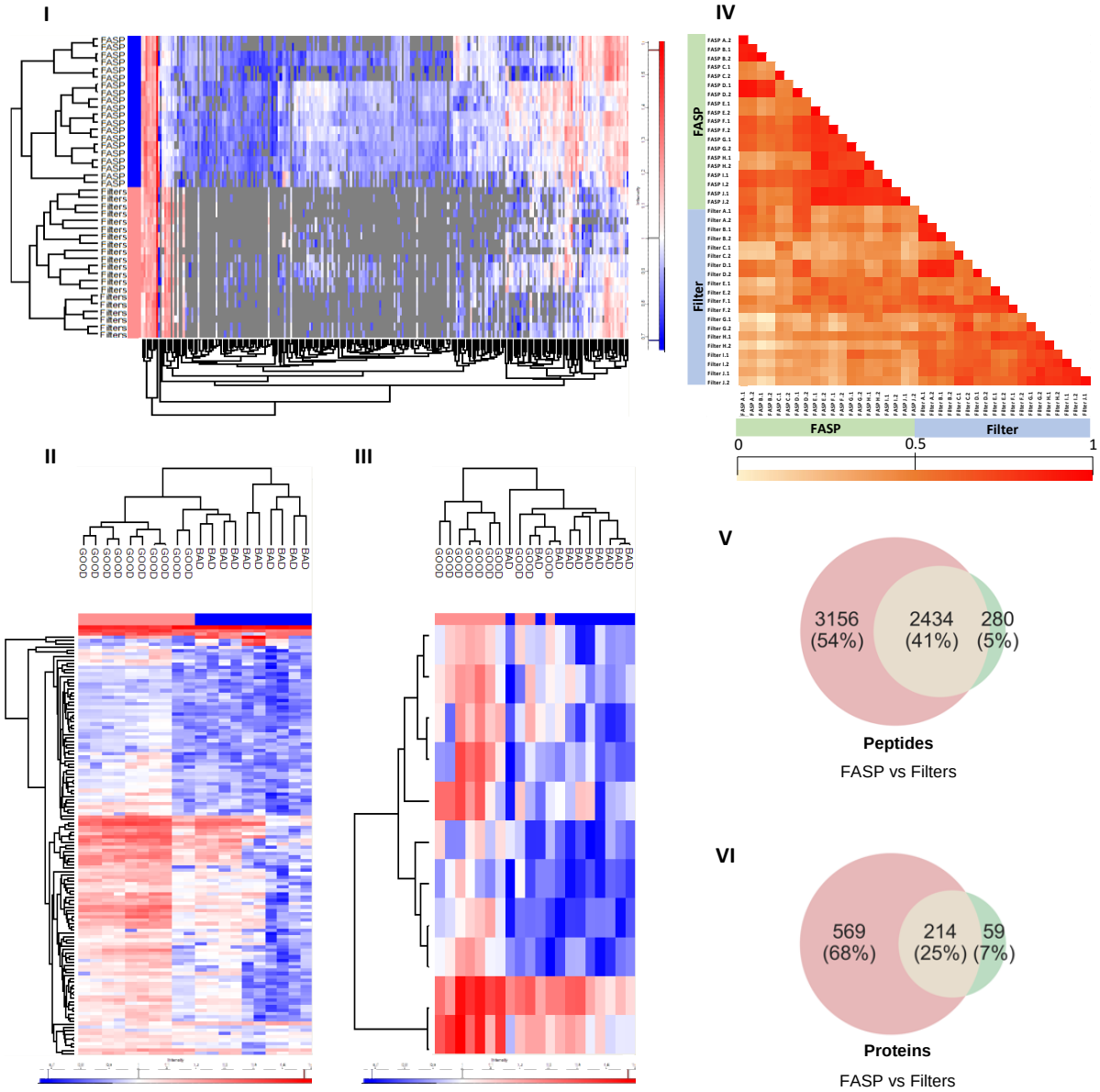

Figure IV.10 - Proof of concept: Comparison between non-recurrence and recurrence T1 stage bladder cancer patients. **(I)** Hierarchical Cluster comparing the filter and FASP protocols. The cluster was generated using proteins identified in at least 70% of replicates, and with no imputation. Red and blue colors represent higher protein expression and lower protein expression, respectively. Grey color represents proteins that were not identified. **(II)** Hierarchical cluster comparing patients with recurrence (BAD) and non-recurrence (GOOD) using the FASP protocol. The cluster was generated using proteins identified in at least 70% of replicates, and with imputation. Red color represents higher protein expression, while blue color represents lower protein expression. **(III)** Hierarchical cluster comparing patients with recurrence (BAD) and non-recurrence (GOOD) using the Filter protocol. The cluster was generated using proteins identified in at least 70% of replicates, and with imputation. Red color represents higher protein expression, while blue color represents lower protein expression. **(IV)** Pearson correlation between all the samples treated using both FASP and Filter methods. The Pearson correlation was done considering all proteins identified in each sample. Red color represents two samples that are highly similar, while more pale colors represent samples that are less similar. **(V)** Venn Diagram of common and non-common peptides identified in FASP and Filter methods. **(VI)** Venn Diagram of common and non-common proteins identified in FASP and Filter methods.

V

Conclusions

Our findings demonstrate that while tryptophan fluorescence measurement has potential as a protein quantification method, further optimization is necessary to address interference from free tryptophan in urine samples.

Additionally, we have shown that molecular weight cut-off Hydrosart membranes can effectively recover up to 90% of the urinary proteome from 1 mL of sample using a simple syringe-based procedure. However, neither the protein extraction protocols nor the digestion in filter protocols we tested was as effective as the established FASP protocol.

While our work highlights the potential for using urine filtering to separate proteins from urine in a convenient and accessible way, further research is required to improve protein extraction recovery from the filter or achieve efficient protein digestion on the filter.

VI

FUTURE WORK

While our results show promise, further optimisation of the protocols is necessary before these approaches can be integrated into current proteomics-based mass spectrometry workflows.

Specifically, the sample washing steps in the tryptophan fluorescence assay for protein quantification must be optimised to minimise the impact of free tryptophan on the measurements and enable accurate protein content quantification in the sample.

For the urinary proteome collection method using syringe filters, optimisation of the protein extraction protocols is needed to maximise proteome recovery and facilitate seamless integration into our current workflow.

Alternatively, optimisation of protein digestion on the filters to achieve results like the FASP method could potentially replace the FASP protocol entirely. Finally, once one of these approaches has been optimised, the stability of the proteome in the membranes at room temperature over time must be evaluated to fully implement this method for collecting and preserving the urinary proteome for mass spectrometry analysis.

Lastly, once either of these routes is optimised, proteome stability over time in the membranes at room temperature must be assessed to truly integrate this new method as a way to collect and preserve the urinary proteome for mass spectrometry analysis.

VI

REFERENCES

[1]. Dobruch, J., & Oszczudłowski, M. (2021). Bladder Cancer: Current Challenges and Future Directions. Medicina, 57(8), 749. https://doi.org/10.3390/medicina57080749

[2]. Teoh, J. Y., Huang, J., Ko, W. Y., Lok, V., Choi, P., Ng, C. F., Sengupta, S., Mostafid, H., Kamat, A. M., Black, P. C., Shariat, S., Babjuk, M., & Wong, M. C. (2020). Global Trends of Bladder Cancer Incidence and Mortality, and Their Associations with Tobacco Use and Gross Domestic Product Per Capita. European urology, 78(6), 893–906. https://doi.org/10.1016/j.eururo.2020.09.006

[3]. Chamie, K., Litwin, M. S., Bassett, J. C., Daskivich, T. J., Lai, J., Hanley, J. M., Konety, B. R., & Saigal, C. S. (2013). Recurrence of high-risk bladder cancer: A population-based analysis. Cancer, 119(17), 3219–3227. https://doi.org/10.1002/cncr.28147

[4]. American Cancer Society. Bladder Cancer Stages. Accessed at https://www.cancer.org/cancer/bladder-cancer/detection-diagnosis-staging/staging.html on 07 of June of 2022.

[5]. DeGeorge, K. C., Holt, H. R., & Hodges, S. C. (2017). Bladder Cancer: Diagnosis and Treatment. American Family Physician, 96(8), 507–514.

[6]. American Cancer Society. Tests for Bladder Cancer. Accessed at https://www.cancer.org/cancer/bladder-cancer/detection-diagnosis-staging/how-diagnosed.html on 09 of June of 2022.

[7]. Patel, V. G., Oh, W. K., & Galsky, M. D. (2020). Treatment of muscle-invasive and advanced bladder cancer in 2020. CA: A Cancer Journal for Clinicians, 70(5), 404–423. https://doi.org/10.3322/caac.21631

[8]. Han, R. F., & Pan, J. G. (2006). Can intravesical bacillus Calmette-Guérin reduce recurrence in patients with superficial bladder cancer? A meta-analysis of randomized trials. Urology, 67(6), 1216–1223. https://doi.org/10.1016/j.urology.2005.12.014

[9]. Lovric, J. (2011). Introducing Proteomics: From Concepts to Sample Separation, Mass Spectrometry and Data Analysis (1st ed.). Wiley.

[10]. Glish, G. L., & Vachet, R. W. (2003). The basics of mass spectrometry in the twenty-first century. Nature Reviews Drug Discovery, 2(2), 140–150. doi:10.1038/nrd1011

[11]. Stauffer, E., Dolan, J. A., & Newman, R. (2008). Gas Chromatography and Gas Chromatography—Mass Spectrometry. Fire Debris Analysis. pp. 235–293.

[12]. Mass Spectrometry Trutotial (Dr. Kamel Harrata) accessed at https://www.cif.iastate.edu/mass-spec/ms-tutorial on 29 of June 2022

[13]. Feucherolles, M., Poppert, S., Utzinger, J. et al. MALDI-TOF mass spectrometry as a diagnostic tool in human and veterinary helminthology: a systematic review. Parasites Vectors 12, 245 (2019). https://doi.org/10.1186/s13071-019-3493-9

[14]. Gross, JH., Matrix-Assisted Laser Desorption/Ionisation. In Mass Spectrometry; Springer International Publishing: Cham, 2017; pp. 651–720.

[15]. Wkipedia.org MALDI. Accessed at https://pt.wikipedia.org/wiki/Ficheiro:Maldi.svg on 1st of July 2022.

[16]. Ho, C. S., Lam, C. W. K., Chan, M. H. M., Cheung, R. C. K., Law, L. K., Lit, L. C. W., Ng, K. F., Suen, M. W. M., & Tai, H. L. (2003). Electrospray ionisation mass spectrometry: principles and clinical applications. The Clinical Biochemist. Reviews / Australian Association of Clinical Biochemists.

[17]. Pitt J. J. (2009). Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. The Clinical biochemist. Reviews, 30(1), 19–34.

[18]. Zubarev R. A. (2013). The challenge of the proteome dynamic range and its implications for in-depth proteomics. Proteomics, 13(5), 723–726. https://doi.org/10.1002/pmic.201200451

[19]. National High Magnetic Field Laboratory. Electrospray Ionisation (ESI). Accessed at https://nationalmaglab.org/user-facilities/icr/techniques/esi on 1st of July 2022.

[20]. Lovric, J. (2011). Introducing Proteomics: From Concepts to Sample Separation, Mass Spectrometry and Data Analysis (1st ed.). Wiley; pp. 116-170

[21]. Eliuk, S., & Makarov, A. (2015). Evolution of Orbitrap Mass Spectrometry Instrumentation. Annual review of analytical chemistry (Palo Alto, Calif.), 8, 61–80. https://doi.org/10.1146/annurev-anchem-071114-040325

[22]. Gross, JH., Instrumentation. In Mass Spectrometry; Springer International Publishing: Cham, 2017; pp. 151–292.

[23]. Aslam, B., Basit, M., Nisar, M. A., Khurshid, M., & Rasool, M. H. (2017). Proteomics: Technologies and Their Applications. Journal of chromatographic science, 55(2), 182–196. https://doi.org/10.1093/chromsci/bmw167

[24]. Bantscheff, M., Schirle, M., Sweetman, G. et al. Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem 389, 1017–1031 (2007). https://doi.org/10.1007/s00216-007-1486-6

[25]. Jiang, H., & English, A. M. (2002). Quantitative analysis of the yeast proteome by incorporation of isotopically labeled leucine. Journal of proteome research, 1(4), 345–350. https://doi.org/10.1021/pr025523f

[26]. Creative Proteomics Blog. Stable Isotope Labeling Using Amino Acids in Cell Culture (SILAC): Principles, Workflow, and Applications. Available at: https://www.creative-proteomics.com/blog/index.php/stable-isotope-labeling-using-amino-acids-in-cell-culture-silac-principles-workflow-and-applications/. (Accessed at 5th of August 2022)

[27]. Lovric, J. (2011). Introducing Proteomics: From Concepts to Sample Separation, Mass Spectrometry and Data Analysis (1st ed.). Wiley. pp. 209-219

[28]. Wiśniewski JR. Label-Free and Standard-Free Absolute Quantitative Proteomics Using the "Total Protein" and "Proteomic Ruler" Approaches. Methods Enzymol. 2017;585:49-60. doi: 10.1016/bs.mie.2016.10.002.

[29]. Creative Proteomics Blog. Label-Free Quantitative Proteomics. Available at: https://www.creative-proteomics.com/blog/index.php/label-free-quantitative-proteomics/. (Accessed at 5th of August 2022)

[30]. Lovric, J. (2011). Introducing Proteomics: From Concepts to Sample Separation, Mass Spectrometry and Data Analysis (1st ed.). Wiley. pp. 1-18

[31]. Hayashi, E., Kuramitsu, Y., Okada, F., Fujimoto, M., Zhang, X., Kobayashi, M., Iizuka, N., Ueyama, Y., & Nakamura, K. (2005). Proteomic profiling for cancer progression: Differential display analysis for the expression of intracellular proteins between regressive and progressive cancer cell lines. Proteomics, 5(4), 1024–1032. https://doi.org/10.1002/pmic.200401132

[32]. Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng. 2009;11:49-79. doi: 10.1146/annurev-bioeng-061008-124934.

[33]. Olsen, J. V., Ong, S.-E., & Mann, M. (2004). Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues. Molecular & Cellular Proteomics, 3(6), 608–614. doi:10.1074/mcp.t400003-mcp200

[34]. Vogeser, M., & Parhofer, K. G. (2007). Liquid chromatography tandem-mass spectrometry (LC-MS/MS)--technique and applications in endocrinology. Experimental and clinical endocrinology & diabetes : official journal, German Society of Endocrinology [and] German Diabetes Association, 115(9), 559–570. https://doi.org/10.1055/s-2007-981458

[35]. Füzéry, A. K., Levin, J., Chan, M. M., & Chan, D. W. (2013). Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges. Clinical proteomics, 10(1), 13. https://doi.org/10.1186/1559-0275-10-13

[36]. Batista, R., Vinagre, N., Meireles, S., Vinagre, J., Prazeres, H., Leão, R., Máximo, V., & Soares, P. (2020). Biomarkers for Bladder Cancer Diagnosis and Surveillance: A Comprehensive Review. Diagnostics (Basel, Switzerland), 10(1), 39. https://doi.org/10.3390/diagnostics10010039

[37]. Oeyen, E., Hoekx, L., De Wachter, S., Baldewijns, M., Ameye, F., & Mertens, I. (2019). Bladder cancer diagnosis and follow-up: The current status and possible role of extracellular vesicles. International Journal of Molecular Sciences, 20(4). https://doi.org/10.3390/ijms20040821

[38]. Zumla A. (2010). Mandell, Douglas, and Bennett's principles and practice of infectious diseases. The Lancet. Infectious Diseases, 10(5), 962-989. https://doi.org/10.1016/S1473-3099(10)70089-X

[39]. Khamis, M. M., Adamko, D. J., & El-Aneed, A. (2017). Mass spectrometric based approaches in urine metabolomics and biomarker discovery. Mass spectrometry reviews, 36(2), 115–134. https://doi.org/10.1002/mas.21455

[40]. Thomas, C. E., Sexton, W., Benson, K., Sutphen, R., & Koomen, J. (2010). Urine collection and processing for protein biomarker discovery and quantification. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 19(4), 953–959. https://doi.org/10.1158/1055-9965.EPI-10-0069

[41]. Lum KT, Meers PD. Boric acid converts urine into an effective bacteriostatic transport medium. J Infect. 1989 Jan;18(1):51-8. doi: 10.1016/s0163-4453(89)93667-0.

[42]. Simerville, J. A., Maxted, W. C., & Pahira, J. J. (2005). Urinalysis: a comprehensive review. American family physician, 71(6), 1153–1162.

[43]. Carvalho, L. B., Capelo-Martínez, J. L., Lodeiro, C., Wiśniewski, J. R., & Santos, H. M. (2020). Ultrasonic-based filter aided sample preparation as the general method to sample preparation in proteomics. Analytical Chemistry. doi: 10.1021/acs.analchem.0c01470

[44]. Zhang, Y., Luo, Y., Lu, H., Wang, N., Shen, Y., Chen, R., … Jia, W. (2015). Effect of Freeze/Thaw Cycles on Several Biomarkers in Urine from Patients with Kidney Disease. Biopreservation and Biobanking, 13(2), 144–146. doi:10.1089/bio.2014.0033

[45]. J. Stone, Chapter 3 - Sample preparation techniques for mass spectrometry in the clinical laboratory, Mass Spectrometry for the Clinical Laboratory, Academic Press, 2017, Pages 37-62, https://doi.org/10.1016/B978-0-12-800871-3.00003-1

[46]. Wiśniewski J. R. (2018). Filter-Aided Sample Preparation for Proteome Analysis. Methods in molecular biology (Clifton, N.J.), 1841, 3–10. https://doi.org/10.1007/978-1-4939-8695-8_1

[47]. Noble, J. E., & Bailey, M. J. (2009). Quantitation of protein. Methods in enzymology, 463, 73–95. https://doi.org/10.1016/S0076-6879(09)63008-1

[48]. Wiśniewski, J. R., & Gaugaz, F. Z. (2015). Fast and Sensitive Total Protein and Peptide Assays for Proteomic Analysis. Analytical Chemistry, 87(8), 4110–4116. doi:10.1021/ac504689z

[49]. Jesus, J. R., Santos, H. M., López-Fernández, H., Lodeiro, C., Arruda, M. A. Z., & Capelo, J. L. (2018). Ultrasonic-based membrane aided sample preparation of urine proteomes. Talanta, 178, 864–869. doi:10.1016/j.talanta.2017.09.078

[50]. Joseph E. Gabriels. Applications Scientist Milipore Corporation. Using A New, Fast Flow, Low Protein Binding Membrane for Sterile Filtration. Accessed at https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/product/documents/428/537/pf2001en00.pdf on 8 July 2022.

[51]. Jorge, S., Capelo, J.L., LaFramboise, W. et al. Absolute quantitative proteomics using the total protein approach to identify novel clinical immunohistochemical markers in renal neoplasms. BMC Med 19, 196 (2021). https://doi.org/10.1186/s12916-021-02071-9

[52]. Tyanova, S. et al. Visualization of LC-MS/MS proteomics data in MaxQuant. Proteomics 15, 1453–1456 (2015).

VII

SUPPLEMENTARY MATERIAL

The data files recovered from MaxQuant, that were then used for all the following analysis can be found in the link below. "ProteinGroups.txt" contains all the data from the different filter protocols as well as the FASP protocol. "ProteinGroups_Proof of concept.txt" contains all the data related to the proof of concept.

(https://www.dropbox.com/scl/fo/vhm74mor8gwjwx8takee9/h?dl=0&rlkey=epht3ihh94yl8f4sojytn6vqk) Last time accessed In 2023.01.16.