Absolute quantification of proteins in complex matrices thanks to a universal standardization kit and differential PTMs evaluation for biomarker discovery studies

a dissertation submitted

by

France Baumans

for the fulfilment of the degree of Doctor of Philosophy in Sciences

Dissertation supervisor:	Gabriel Mazzucchelli
Dissertation co-supervisor:	Loïc Quinton
Institution:	University of Liège
Place:	Mass Spectrometry Laboratory
Academic Year:	2022 - 2023

Jury

Supervisor:	Dr. Gabriel Mazzucchelli	University of Liège, MSLab
Co-supervisor:	Prof. Loïc Quinton	University of Liège, MSLab
President:	Prof. André Matagne	University of Liège, CIP
Secretary:	Prof. Edwin De Pauw	University of Liège, MSLab
Member:	Prof. Marianne Fillet	University of Liège, CIRM
Member:	Prof. Kris Gevaert	University of Ghent, VIB
Member:	Prof. Charles Pineau	University of Rennes, IRSET
Member:	Prof. Ruddy Wattiez	University of Mons, ProtMic

Abstract

In the world of biomarker discovery, up to now, only a few biomarkers have been validated. The validation remains a crucial step in the development of a biomarker but is quite challenging and lengthy. In this context, the development of high-throughput techniques is therefore needed. Mass spectrometry (MS)-based methods such as selected reaction monitoring (SRM) has become a prominent choice for large-scale protein quantification. The difficulty of this approach lies in the standardization of the absolute quantification. Different standardization approaches now exist in proteomics but all of them have their limitation. One example is the use of isotopically-labelled peptides for the absolute quantification of proteins. This method is highly specific of the protein to quantify but does not take into account all the experimental variations introduced at each step of the process applied before SRM analyses. Within this frame, we are developing a comprehensive standardization strategy aiming to control the entire sample preparation process and easing the performance of longitudinal or large-scale studies of biomarker validation. Combined to the isotopically-labelled peptides strategy, we built a chimeric protein containing a structured part associated to an intrinsically disordered domain that will be used, among other things, as a control of the performance of the digestion protocol. The principle is to monitor the accessibility of cleavage sites in function of the structural state of the protein and the digestion protocol applied. Next to this chimeric protein will lies different label-based levels of peptides that are intrinsic of the chimeric protein. These levels of peptides differentially labelled to be distinguishable will allow to control the entire sample preparation process till liquid chromatography-MS (LC-MS) analysis included.

In the context of recombinant proteins, considerable efforts are dedicated to the quality control of the final product. Post-translational modifications (PTMs) are known to alter the structural integrity and stability of the proteins. Among the common existing protein modifications, methionine oxidation is commonly observed either as PTM directly generated in the cell (in vivo modification) or as an artefactual modification induced by the sample preparation process (in vitro modification). Thanks to LC-MS, our work focused on the LC separation that introduces undesired artefactual on-column methionine oxidation. Although in vivo modifications of methionine are relevant for biological purposes, artefactual in vitro modifications should be carefully minimized and evaluated in order to avoid misinterpretation of the results.

For biomarker discovery purposes, methionine oxidation is not the main modification to target, quite the contrary. A multitude of PTMs could help diagnose diseases. They are, at the moment, still underestimated and not used due to the limitation of the techniques used for their quantification. A new methodology based on the synergic use of tryptic digestion and multi-

enzymatic limited digestion (MELD) was developed to improve the quantification of PTMs.

Altogether, this thesis describes absolute quantification of proteins and relative quantification of PTMs as well as the development of a quality control system for any proteomic workflow applicable in any proteomic laboratories.

Acknowledgements

I would like to start by thanking the head of the laboratory at the beginning of my research, Prof. Edwin De Pauw, for providing me with the opportunity to conduct my research in ULiège and for his support and guidance throughout the duration of my PhD. Many thanks also to Prof. Gauthier Eppe for carrying on the work after he took office.

I would like to express my gratitude to my supervisor, Dr. Gabriel Mazzucchelli, for his invaluable guidance and support throughout my PhD journey. His expertise and experience in the field of proteomics were a contributing factor to the success of this thesis. I would also like to thank my co-supervisor, Prof. Loïc Quinton, for his valuable insights, suggestions and for being a critical reader of my thesis.

I am also grateful to the members of the thesis committee for dedicating their time to reviewing my work and sharing their experience with me.

I would like to extend my thanks to the staff and researchers at the MSLab of the University of Liège for providing me with the necessary resources and facilities to conduct my research. More especially to Dr. Dominique Baiwir and Dr. Maximilien Fléron for their expertise as well as their moral support.

I am also grateful to my colleagues and friends in the lab for their support during my time at ULiège. They offered me joy but also comfort in the moments of doubts: Emeline, Nancy, Caroline, Sophie and others I certainly forgot.

Finally, I would like to acknowledge my family and friends for their unwavering support and encouragement throughout the duration of my PhD. Without them, I would not have reached this important milestone of my professional life.

I am deeply appreciative of all the help and support I have received!

Je tiens également à remercier Xavier Baumans pour m'avoir donné le déclic pour développer mon article, qui a ensuite donné lieu à cette thèse. Merci pour ce coup de pouce duquel tout ceci découle !

Je suis également extrêmement reconnaissante envers Emeline Hanozin pour son aide inestimable tout au long de ma thèse. Ta manière incroyable d'agencer les résultats m'a per-

mis d'arriver au bout de cet article malgré toutes les difficultés. Merci pour toutes les heures de discussions, de corrections, de soutien moral, de compassion et tout cela malgré un océan qui nous sépare. Ton travail m'a été d'une aide inégalable et ton amitié tout autant.

Je tiens également à remercier Pépite d'avoir été un chat si gentil pendant mes heures d'écriture, ronronnant sur mes genoux et me donnant le courage nécessaire pour avancer.

Enfin, je remercie Monsieur DePauw d'avoir lancé ce doctorat avec une école de biophysique qui s'est déroulée lors d'une croisière en Croatie. Commencer un doctorat de cette façon a été une véritable chance !

Je souhaite également remercier ma comparse de toujours, Sophie. Nous avons commencé cette thèse ensemble en baignant dans notre jus (toi seule comprendra) et la terminons (presque) ensemble. Presque, parce qu'évidemment, on garde le meilleur pour la fin :) Toutes ces années de thèse auraient été bien fades et (encore plus) compliquées sans toi.

Merci à Sophie, Justine, Caroline, Raphaël pour les fous rires, les crises de nerfs, de panique, les défoulements et les conversations qui font du bien.

Je tiens à remercier Nancy et Caroline pour leur soutien inconditionnel envers la petite pleureuse du labo. Merci d'avoir rempli les journées moroses avec plein de paillettes. Vous avez été des collègues en or !

Merci également à Maxou pour ces fameux jeudi frites qui me manquent cruellement maintenant.

Merci à Clovis et Laurence, qui ont osé être mes compagnons de bureau pendant un certain temps. Je n'aurais pas pu rêver mieux comme collègues.

Je suis également reconnaissante envers ma famille et ma belle-famille pour leur soutien bienveillant.

Un grand merci à mes potes de toujours, d'avoir été là pour me changer les idées et me motiver : Alice, Nora, Audrey, Pauline, Vanessa, Marie et Florine. Parfois les petites phrases anodines sont celles qui motivent le plus : "dis-toi que tu travailles ton anglais" m'a redonné le sourire pour au moins quelques jours.

Merci également à Jérôme, Noémie et Thomas pour les fous rires et les longues discussions au début de nos doctorats. Enfin, merci à cette belle brochette de chimistes tous plus fous et géniaux les uns que les autres : Matthieu, Alice, Chloé, Sophie, Laura, Yasmina, Thomas et Noémie.

Je tiens à remercier mes nouvelles collègues, Chloé et France, de m'avoir appris à me "détacher". Il y a encore un peu de boulot pour y arriver mais on est sur la bonne voie.

Merci enfin à Pierre Wyzen. Lui, je lui en ai fait voir de toutes les couleurs pendant toutes ces années et malgré cela, il a quand même eu l'audace de me demander en mariage. Je dédie cette thèse à la personne formidable que tu es et l'époux parfait que tu es devenu. Je te donne le p de mon phd et pas seulement parce que tu t'appelles Pierre avec un p mais qu'à force de me soutenir, de passer des soirées interminables à m'aider sur cette thèse, tu mérites amplement une partie des lauriers :)

Contents

Nomenclature

Abbreviations

Abbreviation	Definition
AAA	Amino Acid Analysis
ACN	Acetonitrile
AUC	Area Under the Curve
BMI	Body Mass Index
CD	Circular Dichroism
CSF	CerebroSpinal Fluid
cICAT	Cleavable Isotope-Coded Affinity Tag
DDA	Data-Dependent Acquisition
DIA	Data-Independent Acquisition
DTT	DiThioThreitol
ELISA	Enzyme-Linked ImmunoSorbent Assay
emPAI	Exponential Modified Protein Abundance Index
EPINOX	Erythrocyte Proteomics for ImmuNOdetection of OXidation
FLEXIQuant	Full-Length EXpressed stable Isotope-labelled proteins for QUANTification
FS	Flanking Sequence
HbNO	Nitrosylated haemoglobin
IAA	IodoAcetAmide
ICAT	Isotope-Coded Affinity Tag
iTRAQ	Isobaric Tag for Relative and Absolute Quantification
LC	Liquid Chromatography
m/z	mass-to-charge
MassIVE	Mass Spectrometry Interactive Virtual Environment
MRM	Multiple Reaction Monitoring
MS	Mass Spectrometry
MS/MS or MS^2	Tandem Mass Spectrometry
MELD	Multi-Enzymatic Limited Digestion
NHS	N-Hydroxysuccinimide
NO	Nitric Oxide
PCS	Peptide-Concatenated Standard

Abbreviation	Definition
PolySIS	POLYprotein Stable Isotope-labelled internal Standard
PRIDE	Proteomics Identification Database
PRM	Parallel Reaction Monitoring
PSAQ	Protein Standard for Absolute Quantification
PSM	Peptide Spectrum Match
PTM	Post-Translational Modification
QconCAT	Quantification of conCATenated peptides
QPrEST	Quantification of PRotein Epitope Signature Tag
RBC	Red Blood Cell
RISQ	Recombinant Isotope Labeled and Selenium Quantified
RNS	Reactive Nitrogen Species
ROS	Reactive Oxygen Species
RP	Reversed-Phase
RT	Room Temperature
SILAC	Stable Isotope Labelling by Amino Acids in Cell Culture
SRM	Selected Reaction Monitoring
SSRCalc	Sequence-Specific Retention Calculator
SWATH	Sequential Windowed Acquisition of All Theoretical Fragment ions
TMT	Tandem Mass Tag

Symbols

Symbol	Definition	Unit
$[\theta]_{MRE}$	Mean residue ellipticity	[deg.cm^2.dmol^{-1}.residue^{-1}]

1 Introduction

Biomarkers are biological markers that can be used to indicate the state of an individual. They can be used to measure the presence or concentration of a specific molecule, such as a protein or a metabolite, in a biological sample like blood, urine, or tissue. The healthcare industry makes use of them to monitor health, diagnose disease or track the progress or impact of a treatment. This work sheds light to the complexity of the selection and the analysis of proteins by means of proteomics used as biomarkers.

1.1. Background information

Only a limited set of biomarkers reaches the clinical application phase. Several factors impede their validation beyond research units and are introduced hereunder. This section introduces some of them.

First are the technical limitations of the currently employed methods in the biomarker discovery phases. Immunoassay techniques furnish reliable and reproducible data but are limited in their capacity of multiplexing (i.e. the simultaneous evaluation of multiple targets in a sample). Mass spectrometry (MS)-based proteomics methods offer high multiplexing capacity and high throughput but are difficult to standardize from one analysis unit to another. Lack of standardized protocols makes the comparison of results and the establishment of consensus difficult.

Furthermore, in the validation phase, large-scale datasets are required to produce statistically significant results, which can make the process both expensive and time-consuming. This involves the recruitment and monitoring of thousands of patients, generating an enormous amount of raw data that must be analysed. Additionally, the analysis of such data is

complex and requires specialized bioinformatic resources, adding to the overall cost and time required for biomarker validation.

Another limitation in biomarker discovery is that differences in the quantity of a protein may not always be the defining factor, as the presence or absence of post-translational modifications (PTMs) can also be significant. PTMs have been shown to regulate various cellular processes and have been implicated in the pathogenesis of several diseases. Therefore, identifying and quantifying them in biological samples could uncover novel biomarkers and therapeutic targets for disease diagnosis and treatment. Biomarker discovery on post-translationally modified proteins can be however challenging due to several reasons explained in detail in this manuscript.

1.2. Problem statement

The context described in the previous section highlights two obstacles for the validation of biomarker that are addressed in this document.

1. Despite technological progress, the lack of standardized proteins quantification protocols prevents reliable identification of good biomarkers candidates to be further validated.
2. Developing sensitive and specific assays for PTM proteins is crucial for successful biomarker discovery due to the complex and diverse nature of PTMs and their dynamic regulation.

Addressing these challenges will contribute to the strengthening of biomarker reliability and ultimately, their use for diagnosis and therapeutic follow-up purposes.

1.3. Objectives

The objective of this work is thus twofold. On one hand, we aim to set up an innovative standardization methodology for proteomic experiments to be applied either in discovery phase or for MS-based validation phase of biomarkers while on the other hand, we endeavour to develop a new strategy to analyse post-translationally modified proteins in large cohorts.

We aim to quantify proteins with a more efficient and reliable process than existing protocols using liquid chromatography (LC) coupled with MS. Current methodologies can be time-consuming and prone to biases, which limits their effectiveness for identifying and quantifying proteins. The solution should take into account the experimental variations occurring during the sample processing, the control of the LC-MS separation and the reproducibility of measure among other considerations. By achieving this goal, we would universalize the protocol materialized into a toolkit in order to be usable by any laboratory, regardless of the preparation

processes and instrumentations.

The second goal of this work is to develop a new strategy for the analysis of proteins with PTMs in large cohorts. The new strategy would consider the relative quantification of all potential post-translational modified proteins by combining the advantages of the multi-enzymatic limited digestion (MELD) innovative approach with the quantification obtained through tryptic digestion. The confidence level and the time efficiency of the results obtained with this methodology should be the indicators of the objective achievement.

1.4. Overview

The context of biomarker discovery and its diverse techniques of quantification are elaborated in chapter 2. The chapter will discuss the most commonly used methods to quantify proteins such as label-based and label-free approaches but also antibody quantitation, targeted MS techniques and PTMs. The discussions also include a review of the pitfalls experienced by the listed methodologies.

The first problem statement is addressed in chapter 3 where the development of a comprehensive standardization strategy aiming to control the entire sample preparation process of biomarker validation is elaborated.

While investigating the first problem statement of this research, it was discovered thanks to LC-MS that undesired artefactual on-column methionine oxidation was introduced by the LC separation. An article which provides information about the evaluation of this oxidation was published as part of this research and is the subject of chapter 4.

A new methodology to take PTMs into consideration when searching for biomarkers is reported in chapter 5. It is based on the synergic use of tryptic digestion and MELD.

2

Biomarker discovery

2.1. Definitions

The term biomarker, or biological marker, is defined in 1998 by the National Institutes of Health Biomarkers Definitions Working Group as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” [1]. A biomarker is therefore a molecule describing a specific biological state that can be accurately and repeatedly detected and/or quantified to better apprehend patient’s health condition. To date, biomarkers are used for many applications such as a predictive (prediction of clinical outcomes, early detection of predisposed state), a diagnostic (identification of a disease) or a prognostic (expectation of the evolution of the disease) tool. The recent development in high-throughput powerful analytical proteomic technologies for the characterization of biomarkers now generates “big data”. Combined with improved or new bioinformatics tools for data processing, it has opened the era of “personalized medicine”. This concept makes it possible to prevent a disease or to apply a treatment based on the variability of an individual’s biological proteome. Even though RNA and DNA biomarkers gain more and more insight due to their high accuracy and specificity and also their lower cost [2], they are not available for every pathology. Therefore, protein biomarkers are often preferred and recognized as reliable biomarkers because they reflect the overall proteome variability against a disease. Lots of work have been dedicated to research on biomarkers these last decades (figure 2.1). In 2020, 55,977 paper concerning biomarkers were published whereas only 6% of them were about validated biomarkers.

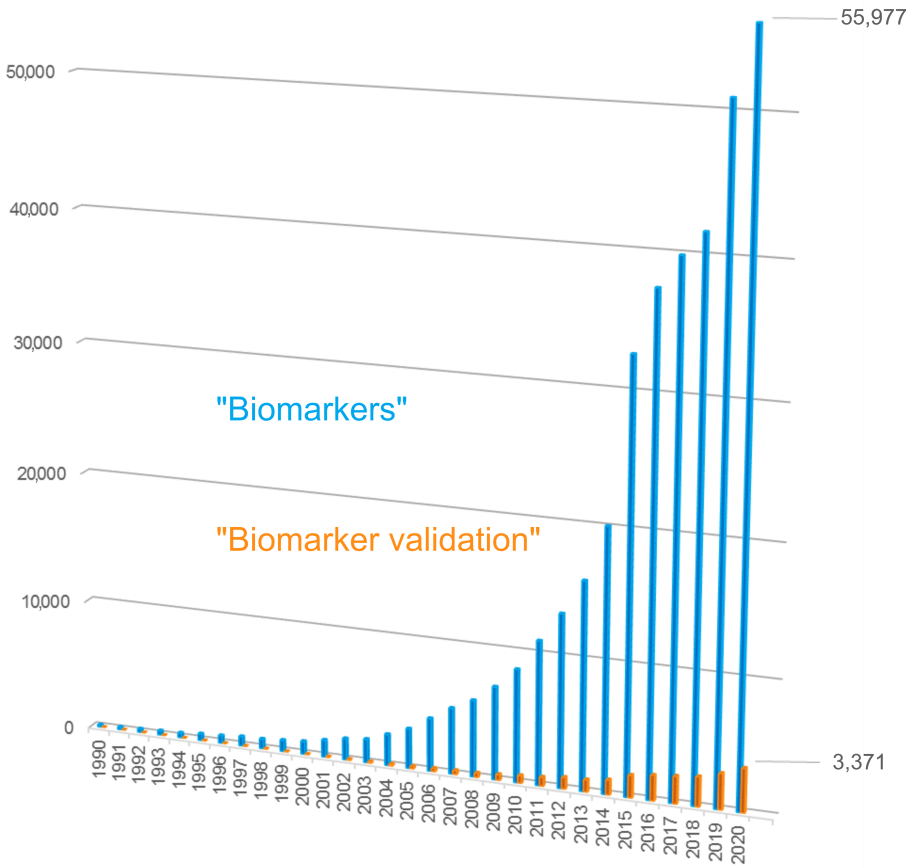


Figure 2.1: *Number of documents published per year in Scopus about biomarkers (in blue) and biomarker validation (in orange) based on a Scopus search with **"Biomarkers"** or **"Biomarker validation"** [Article title, Abstract, Keywords] AND Article [document type] published each year since 1990.*

When comparing figure 2.1 with figure 2.2, which includes the term ”proteomics” in the search, we can observe that in 2020, there were 1,353 published papers on biomarkers, but only 0.4% of them were focused on validated biomarkers in the proteomics area. It is important to note that this low number of validated biomarkers in proteomics does not imply that this field is incapable of producing high-quality biomarkers during the discovery phase. In fact, there could be a significant number of biomarkers discovered through proteomics that are validated using other techniques, such as immunoassay-based methods. This highlights the fact that the validation step remains a significant bottleneck for biomarkers to progress from the laboratory stage to clinical applications.

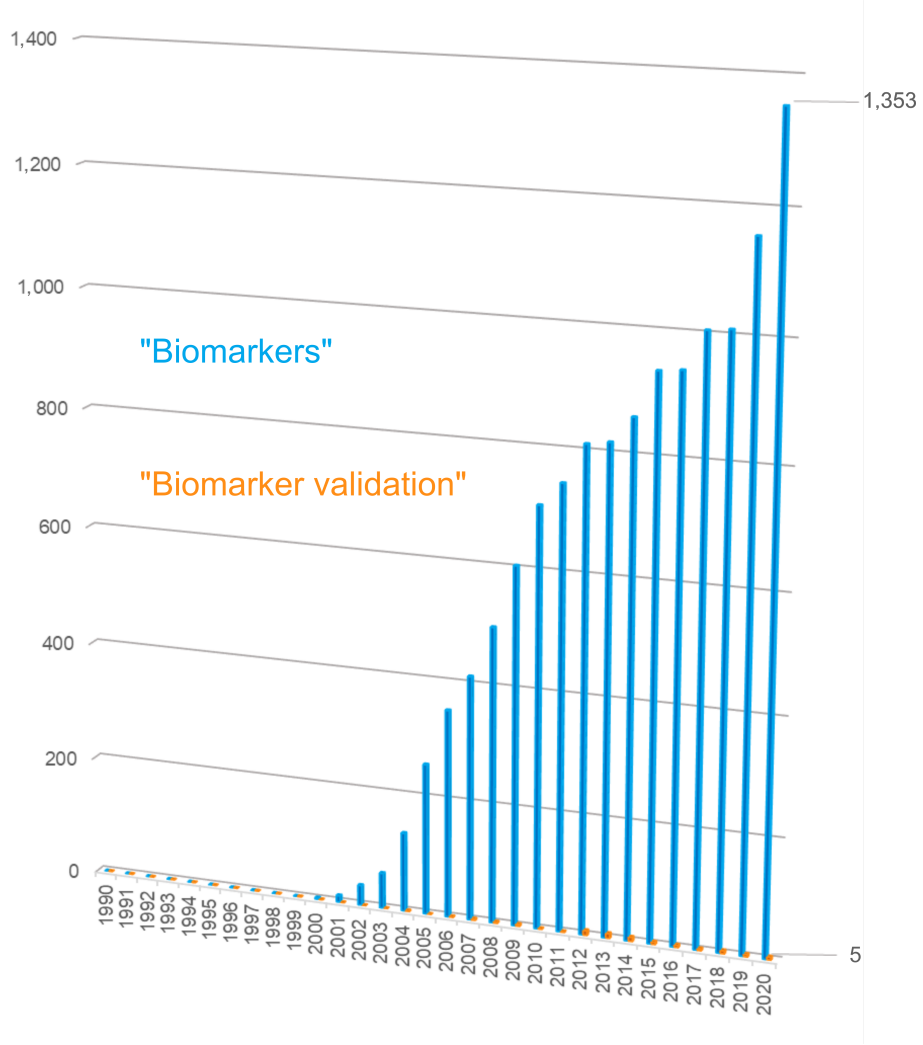


Figure 2.2: *Number of documents published per year in Scopus about biomarkers (in blue) and biomarker validation (in orange) in the filed of proteomics based on a Scopus search with **"Biomarkers"** AND **"Proteomics"** or **"Biomarker validation"** AND **"Proteomics"** [Article title, Abstract, Keywords] AND Article [document type] published each year since 1990.*

Different phases are necessary for the implementation of any biomarker in clinical applications. The first phase is the discovery process, in which standardization and robustness issues in relation to the experimental procedure (from collection to injection of the samples) can arise, leading to non-convincing results or false positives. A verification of the biomarker specificity is then needed to ensure the significant difference of the potential biomarkers between disease and control groups and their relevance for use as therapeutic, diagnostic or prognostic follow-up. The few biomarker candidates slipping through this verification phase enter then into the validation step where larger cohort of individuals are investigated to confirm the biomarkers on an extensive scale. This last step before rigorous clinical steps and final commercialization is a long and costly juncture because of the currently employed method namely immunoassay based technologies. In this latest approach, specific antibodies against the targeted biomarkers are needed and require production and adaptation of the experimental procedure for each

antibody beforehand. The problem lies in the number of samples required for this last step to succeed. Indeed, throughout the whole biomarker discovery workflow, from the discovery phase to the validation one, while the number of analytes to be evaluated decreases, the number of samples must constantly increase to obtain clear statistical differential expression. Even though immunoassay methods offer high throughput, the capacity of multiplexing is low and this technique is also limited by the availability of specific antibody.

2.2. Considerations for clinical applications

Before diving into the biomarker discovery phase, one has to pay attention to the nature of the cohorts that will be under study but also the number of individuals to take into considerations for successful proteomic experiments. This section will discuss these matters along with some technical considerations for clinical applications.

Drawing a good study design and gathering reliable samples are the prerequisites for a successful study [3, 4]. The main difficulty resides in the sampling choice, the perfect one being a cohort where the patients are their own controls [5]. This means that a patient's sample collected at a certain time point (*e.g.* before treatment) should be compared to a sample collected from the same patient at another time point (*e.g.* after treatment). The rationale behind this approach is that individual variations in protein expression levels can be minimized, and any changes in protein expression that are observed can be attributed more confidently to the treatment or disease status. Unfortunately, this method is not always feasible in practice. The main assumption is to state that controls and patients only differ in the condition of interest and that they are otherwise identical. To reach this ideal cohort, one has to pay attention to balance as much as possible the control vs the patient groups and therefore avoid underpowered study. Factors such as age, gender, intake of alcohol, caffeine, vitamin, tobacco and drug consumptions, body mass index (BMI), physical activity and medication use must therefore well balanced in each group. Zhou *et al.* [6] highlighted that six samples per group (controls vs patients) were necessary for robust biomarker identification, *i.e.* sufficient statistical power to detect proteins with a fold change (ratio between protein quantities from both groups) > 2.

Another critical consideration is the pre-analytical step such as the collection or the storage of the samples that must be done in a rigorous and standardized manner, not to induce bias in any of the samples. For logistical reasons, samples are not necessarily drawn off at the same time. Therefore, detailed protocol must be followed to ensure the quality and reliability of the biological content to analyse. The same reasoning must be considered for the sample preparation and the LC-MS process. The reproducibility of procedures is of critical importance.

The choice of the sample type is also a great concern and is ideally directly linked to the place where the disease process takes place. In consequence, tissues are often preferred

over serum/plasma samples as the dynamic range of the proteome of this latter is quite complex [4]. Complexity of such samples reduces the signal-to-noise ratio due to the wide dynamic range of protein abundance of these matrices. Despite its advantages, the use of tissue samples in biomarker discovery also presents some drawbacks, such as the invasive nature of the procedure, tissue heterogeneity, the need for careful storage and transport, and the requirement for detailed histopathological evaluation by experts [7].

Sample pooling should not be considered unless if it exists an incompatibility with the analytical resources, for budget consideration or for material availability purposes [4]. Pooling the samples does not necessary increase the level of information and can even produce a degradation of its quality by averaging the detected proteome variations. For example, if one sample of the pool does not express a protein, this protein could be switched to the detection threshold in the pool, resulting in a false negative. On the contrary, a protein highly expressed in one sample but not in the other samples of the group could lead to a false positive by detecting this protein in the pool [8]. Yet, a biomarker could be identified from pooled samples if its variation is important and stable enough in all the samples of one condition.

When searching for new biomarker candidates, it is rather easy to consider a protein as a good candidate if this protein is systematically detected in the majority of the patient samples and statistically differentially expressed from the pathological sample sets to the healthy ones. The opposite can also be acceptable; if a protein is present in the control sample series and absent in the patients. Indeed, in some cases, disease factors can negatively influence the expression of a protein and reduce it in comparison to healthy individuals. This balance between presence and absence must be clearly defined using lower limit of detection for the protein of interest. Qualitative differences are thus easy to handle in opposition to quantitative differences which are somewhat more challenging. Indeed, quantitative MS-based approaches must be applied to accurately quantify the protein in each sample. Relative quantitation methods are applied in the discovery biomarker workflow at the very first phase of the process, namely the discovery phase. Throughout the biomarker discovery phases, the number of samples is limited whereas the proteome coverage to be analysed is as large as possible. The workflow goes from 1) multiple analysed proteins in a relative quantification manner to 2) the analysis of specific proteins using integration and absolute quantification and finally to 3) absolute and multiplexed methods on a set of restricted proteins of interest. The increasing sample size needed, with the subsequent phases 1), 2) and 3) of the development study to increase the statistical power, renders the task difficult.

2.3. State of the art

A multitude of technologies and diverse methods currently exist to quantify proteins ([9–12]. The absolute quantification of proteins or differential approaches are a complex and tech-

nical subject and a currently uphill project in proteomics. This section aims at approaching and discussing the most commonly used methods up to this day and will focus especially on MS-based proteomics technology, although antibody-based methods are well established and will therefore also be discussed.

MS-based proteomics can broadly be divided into different categories: either relative *vs* absolute quantification, either label-free approaches *vs* label-based approaches, and either discovery approaches *vs* targeted ones. The MS acquisition methods have also their importance and drives the sample preparation methods that can be applied: data-dependent acquisition (DDA) *vs* data-independent acquisition (DIA).

For proteomics studies, a reversed-phase (RP) liquid chromatography system hyphenated to a mass spectrometer through an electrospray source is often employed. While discovery approaches using either DDA or DIA seek the global proteome of a sample by identifying the maximum number of peptides in a certain time frame, targeted acquisitions focus on specific ions formerly chosen [13, 14].

Different quantitative strategies can be implemented such as metabolic labelling (SILAC), enzymatic labelling (with ^{18}O incorporation), chemical labelling (ICAT, cICAT, iTRAQ, TMT), isotopic dilution strategies that use addition of isotopic labelled peptides (AQUA), concatenated peptide standards (QconCAT, PCS, polySIS, QPrEST) or labelled protein standard (PSAQ, FLEXIQuant, RISQ). Metabolic, enzymatic and chemical labelling offer relative quantification whereas the use of spiked peptides (i.e. peptides added in the sample) and heavy protein standard enable absolute quantification. These label-based approaches are in opposition to the label-free approaches that sit in the relative quantification methods. However, regardless of the method used for protein quantitation, LC-MS/MS is and always will be an inherently non absolute quantitative technique if no reference is used, due to the difference in ionization efficiencies [15]. Current methods aim to achieve higher quantitative accuracy by addressing bias arising from both sample preparation and instrumentation. These approaches work towards reducing bias or acknowledging its presence, and in some cases, establish a threshold beyond which the data must be discarded. A schematic description of all the quantitative proteomics methods that will be described in this thesis is represented in figure 2.3.

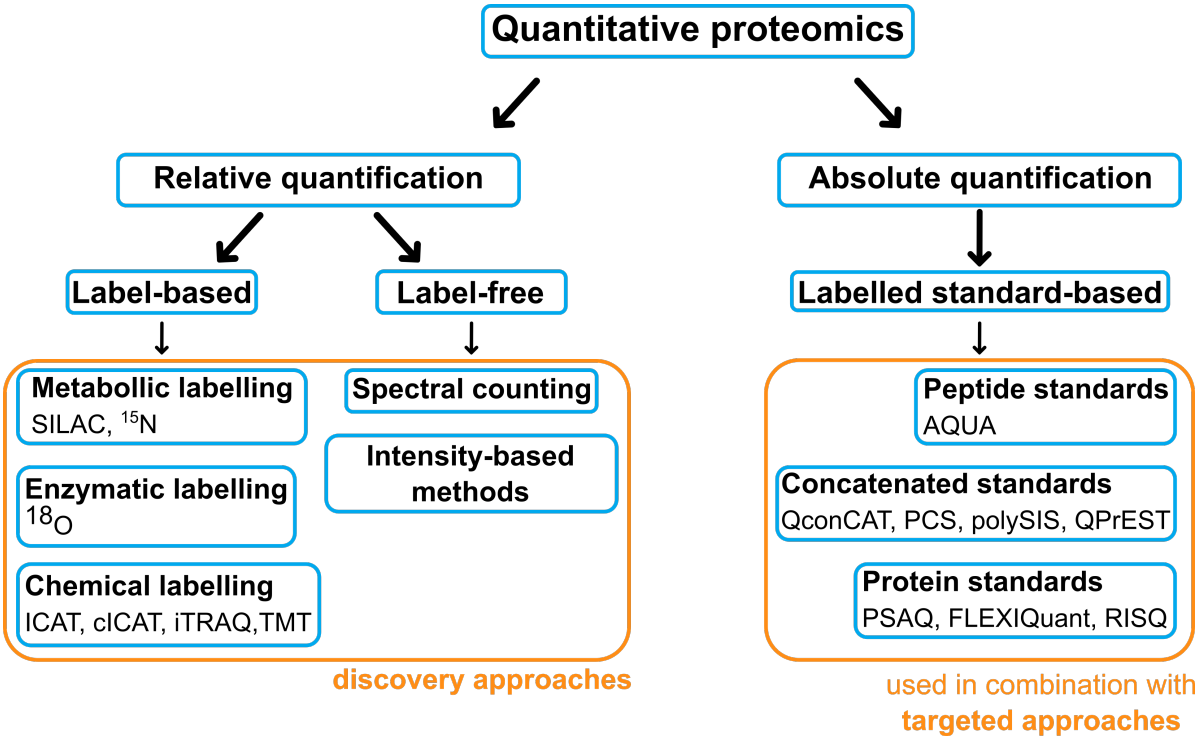


Figure 2.3: *Representation of the common quantitative proteomics approaches divided in relative and absolute quantification techniques.*

2.3.1. Relative vs absolute quantification

For most biological and biomedical study, only relative quantification is desired and offers a differential analysis, *i.e.* the protein intensities of one state in relation to the other biological state. The amount of given protein is often expressed in terms of a fold change for two conditions.

While relative quantification allows acquiring quantitative information for a large number of proteins present in a sample, absolute quantification only allows quantification information of a limited number of proteins of interest. This is explained by the accuracy of the given information in either method. In absolute quantification, the amount of the protein of interest is determined by quantification process based on an isotopic dilution approach whereas in relative quantification, only a relative change in protein amount between two states or more is obtained. In principle, absolute quantification outperforms relative quantification: if the protein amount of two samples is precisely known, their relative ratio is automatically calculated.

The gold standard method to determine protein concentrations is via the use of synthetic standards labelled with stable isotopes. The standard is the same peptide or protein than the

one to be quantified in the matrix, but labelled to allow the distinction between the standard and the endogenous peptide. It is inserted in a known amount at a specific step of the sample preparation process (depending of the type of standards). The intensity of the endogenous peptides and the standard are then compared. This direct comparison of intensities can be used for identical peptides because they have the same ionization efficiency and their elution profiles are the same. Practically, the AUC of both the standard and the target peptide is determined and the ratio target peptide/standard is multiplied by the quantity of standard added to the sample (see equation 2.1). The final protein concentration ensues from this value.

$$\text{Quantity of analyte} = \frac{\text{AUC}_{analyte}}{\text{AUC}_{standard}} * \text{quantity of standard} \tag{2.1}$$

First, we will detail the label-based approaches, followed by the label-free approaches. We will then discuss the targeted approaches, which lead to the standard use for absolute quantitative proteomics. Finally, enzyme-linked immunosorbent assay (ELISA) and PTMs characterization will be described.

2.3.2. Label-based approaches

Label-based approaches are, as the name suggests, methods that rely on different stable isotope labelling to produce specific mass-tagged peptides detected by MS and that are advantageous for quantitation in a multiplexed analysis. These tagged peptides, referred as heavy peptides, could therefore be differentiated from their unlabelled counterpart, namely the light peptides. These tags can be inserted in proteins or peptides in a (i) metabolic, (ii) enzymatic, (iii) chemical or (iv) isotopologue way. These four methods will be detailed below but general requirements can already be drawn. Despite the labelling method chosen, one has to be careful to avoid overlaps between the isotopic distributions of the light and the heavy peptides by ensuring enough mass difference between the two [16]. An appropriate mass shift, with clearly distinguishable isotopic distributions will lead to more accurate quantification. ^{13}C, ^{15}N and ^{18}O labelling are often preferred over ^{2}H labelling that creates retention time shifts between the labelled and the unlabelled counterpart. This particularity complicates the data interpretation and may create a bias when integrating the two peptide species.

Metabolic labelling

A class of labelling is the metabolic incorporation of the stable isotope. The method consists in introducing the wanted stable isotopes in the cell culture medium in order that it becomes metabolically incorporated into the proteome through cell replication. Beside this heavy media another light media is used in parallel to further compare the two cell populations thanks

to MS. The main advantage of this method is that the labels are metabolically incorporated at the cell level and therefore there is no need for production of synthetic labelled peptides. Also, as the labelling is introduced at the earliest step of the process, *i.e.* during cell growth, metabolic labelling offers an unbiased sample preparation process when samples from light and heavy states are mixed together before beginning the digestion protocol required for MS. Indeed, as one sample, they will be affected in the same way by the preparation process leading to exact same quantification bias.

Initially used as ^{15}N labelling for total labelling of bacteria, Ong and Mann [17] rendered the method appealing for the proteomics community by introducing the stable isotope labelling by amino acids in cell culture or otherwise abbreviated SILAC. As the name implies, stable isotope amino acids are supplemented in the media instead of natural amino acids that gives rise to nearly full metabolic protein labelling after several cell replications. While one cell population is grown with light amino acids, the other is grown with stable isotopes of essential amino acids. Originally [17], leucine residue was first tested as the heavy stable amino acid to prove the concept of the method as it is the most abundant residue in proteins. Generally, arginine and lysine residues are picked when tryptic digestion follows to ensure that all the tryptic peptides carry at least one tag. Relative quantification is obtained by comparing the intensities of the co-eluting heavy and light peptides from the MS spectrum. A maximum of three different experimental conditions can be multiplexed in a single MS analysis (unlabelled, ^{13}C, ^{13}C ^{15}N labelling).

Enzymatic labelling

In vitro labelling approaches differ from metabolic labelling by being inserted post-biosynthetic production. Enzymatic labelling is an in vitro labelling method where the stable isotopes are introduced into the peptides via the protease that will digest the protein content of the sample. Labelling is performed during proteolytic digestion or during an incubation step after proteolysis with ^{18}O-rich water hence the incorporation of ^{18}O-atoms in the sample [18, 19]. For relative quantification, samples are simultaneously subjected to digestion using either ^{16}O water or ^{18}O water, followed by combination and joint analysis using LC-MS/MS. The next step involves comparing the response of both labelled and unlabelled peptides [20]. Despite its ease of use, this method suffers from some drawbacks such as an incomplete ^{18}O incorporation, a susceptibility of instability of the label induced by back-exchange and an inability to compare multiple samples within a single experiment.

Chemical labelling

Another in vitro labelling is the chemical labelling that consists in adding a tag on the reactive sites of a protein/peptide by inserting in the sample a specific chemical reagent that contains stable isotope. Once again, this labelling method offers relative quantification by chemically modifying the protein content from two distinct samples with light and heavy chemical reagents.

One of the very first reagent commercialized was the isotope-coded affinity tag (ICAT) [21] which employs sulfhydryl group-oriented reagents that will therefore specifically bind to cysteine residues. The cysteine residues of two proteomes are respectively derivatized with the light and the heavy ICAT reagent after reduction of the proteins of the samples. The two samples are then combined and enzymatically digested to generate peptides, some with ICAT tag. These tagged peptides are picked out using avidin affinity chromatography and finally analysed by LC-MS/MS. Because of the isolation of the tagged cysteine-containing peptides, the complexity of the mix is lowered but information about peptides that do not contain cysteine is lost. Enhancement of the method gave rise to cICAT for cleavable isotope-coded affinity tag where the biotin moiety of the ICAT reagent can be removed and therefore allows better quality of the fragmentation spectra [22, 23]. The cICAT reagent also contains ^{13}C atoms instead of deuterium in the heavy form to get co-elution from the heavy and light modified peptides and a more reliable quantification than with deuterium/hydrogen that do not co-elute as mentioned in the beginning of section 2.3.2.

Taking advantages of the numerous amino acids present in a protein and therefore the possible reactions of their functional groups with various tags, lots of other chemical labelling came out after the ICAT applications. The methods described in the following paragraphs are among the most used in proteomics applications.

A very popular method is the isobaric tag for relative and absolute quantification (iTRAQ) [24]. The tags used in this technique are covalently bond to the N-termini and lysine side chains of peptides and take the form of isobaric mass labels once they are linked to the peptides. This simplifies the mass spectra while allowing the identification differences via the differentially labelled mass tags which, upon fragmentation, release mass varying reporter ions. iTRAQ offers multiplexing up to 8 samples in one single MS experiment; the main current reagents are the 4-plex and the 8-plex, respectively used for the determination of protein levels in 4 and 8 different experimental conditions. This method is mainly used for relative quantification. Absolute quantification can be achieved by inserting a known quantity of synthetic peptide labelled with one of the reporter isobaric reagent and compare it to the intensity of the same labelled peptide in the sample of interest. An extension of the iTRAQ isobaric tags is the mTRAQ (mass differential tags for relative and absolute quantification) [25] which relies on nonisobaric labelling. Chemically identical to the iTRAQ reagents except for the iso-

tope content, precursors with different masses can be directly quantified in MS^1 (MS^1 refers to the sample ionized and analysed by mass spectrometry in opposition to MS2 or MS/MS experiment where the sample from MS^1 is fragmented and these fragments are analysed by MS). Another widely used isobaric reagent is the tandem mass tag (TMT) [26]. Identification and quantification of proteins are provided by tandem mass spectrometry (MS/MS) hence the name tandem mass tag. As in the iTRAQ method, in the TMT approach, tags are designed in a way that upon fragmentation they will lose a fragment that will make way for a specific mass ion detected by MS. TMT reagents also target the peptide N-termini and amino group of lysines with an N-hydroxysuccinimide (NHS) ester-based reactive group. The molecular mass of the reporter ions lies between 126 and 131 m/z for TMT and 113 and 121 for iTRAQ. Different numbers and combinations of ^{15}N and ^{13}C isotopes in the mass reporter are required to obtain similar structure of the reagents. TMT is available in up to 16-plex format and therefore allow high multiplexing with the possibility of analysing lots of different conditions at the same time. Drawback of these two methods, that apply a labelling at the peptide level, is the possibility of bias results arising from the sample preparation process between the parallel analysed samples. Another limitation is linked to the inner molecular structure of the reagents as they are identical in structure, they are chromatographically indistinguishable and also therefore indiscernible in a first MS spectrum. This disadvantage implies a quantification at the MS^2 level (therefore no quantification based on the area under the curve (AUC) but rather on a single point from the MS/MS data) and prevents from direct targeted methods. Distinct labelled peptide ions with close m/z values can also co-elute within the selected isolation window leads and therefore co-fragmented. This phenomenon is called the “ratio compression” and leads to bias in the measured reporter ions.

2.3.3. Label-free approaches

Label-free approaches are either based on spectral counting or on peak intensity for comparison of the relative abundance of proteins between groups. Spectral counting is often used when large cohorts of samples must be analysed as this relative quantification is rather straightforward and suitable for large scale proteomic analysis. Spectral counting method relies on the observation that a protein amount is directly proportional to the number of its peptides and therefore to the number of peptide spectrum matches (PSMs) generated during acquisition. Spectral count does not account for protein size as larger proteins contributes to more peptides. This method offers a global quantitation regarding two different conditions in a sample data set. Several methods based on spectral counting are available nowadays such as peptide count, spectrum count or protein abundance index. A well explained overview of these spectral counting approaches is reported in the review of Bantscheff *et al.* [10]. Although label-free quantitation technique is known as a relative quantification method, normalized exponentially modified protein abundance index (normalized emPAI) techniques (derived from spectral counting methods, see [27] for further details) allows “absolute quantification of pro-

teins". An accurate approach of label-free MS quantitation is the integrated measurement of chromatographic AUC obtain by LC-MS runs for any given peptide and, by inference, quantitation of the protein of interest. In this configuration, MS spectra allow the quantification and MS/MS spectra the identification. A linear response is obtained between peak intensity and protein concentration. In differential experiments, the AUC of peptides is compared between LC-MS runs of all the samples. Variations can arise from biological distinctions including elution time and background noise. MS data analysis software are thus necessary to normalize the data acquired at different time in multiple MS analyses [28]. The main advantages of these two techniques are that they do not require laborious and costly label-based workflows. The limitation resides in the low throughput as multiplexing approaches cannot be applied.

There are two main acquisition technologies used to generate MS proteomic data in an untargeted manner: the DDA and the DIA [29]. These methods rely on label-free differential proteomics that offer relative quantification without the use of any labelled or modified protein or peptide assisting the quantification process. In DDA and DIA mode, acquisition time is shared between MS^1 and MS^2. These two methods are depicted in figure 2.4 and explained in detail in the next sections.

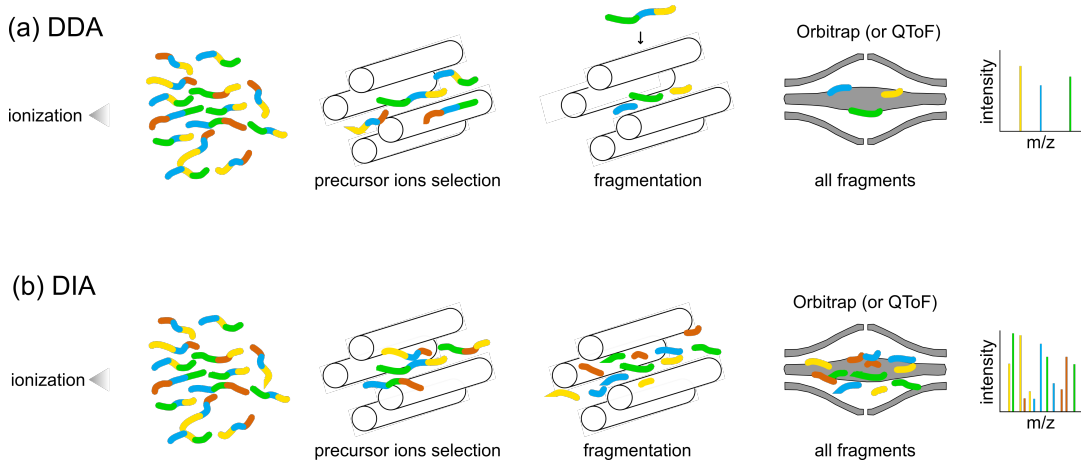


Figure 2.4: *MS label-free acquisition methods. (a) By data-dependent analysis (DDA), ions are submitted to a full scan analysis and the TopN precursor ions are subsequently sequentially selected for fragmentation. (b) By DIA, a group of precursor ions is simultaneously selected and fragmented leading to more MS/MS complex spectra.*

Data-dependent acquisition

In DDA mode, peptides coming from "bottom-up" proteomics are introduced in the mass spectrometer to undergo a first MS^1 full-scan. Then, precursor ions, typically those of highest abundances, are automatically selected for fragmentation leading to the tandem mass analysis. This acquisition mode often refers as "TopN" analysis for the number of precursors detected during the first stage of MS and selected for the second MS/MS stage. The N most

intense ions are generally selected based on their intensity (above a configurable threshold) and their charge states (z > +1) and fragmented sequentially. This cycle of MS and MS/MS acquisitions progresses throughout the whole chromatographic separation. A dynamic exclusion parameter ensures that a same precursor is not selected multiple times over a pre-set period of time, depending on the chromatographic resolution. A balance must be found between short dynamic exclusion time that matches at least the average peak width and long dynamic exclusion time that leads to less redundant spectra. Generally, between 10 and 15 peptides are picked per cycle time. The cycle time is the ratio of the average peak width and the minimum number of full scan MS detection points needed per chromatographic peak:

$$\text{cycle time [sec]} = \frac{\text{average peak width [sec]}}{\text{minimum number of points across peak}} \tag{2.2}$$

For accurate quantification purpose, 10 to 15 data points is the ideal number of points needed across the resulting MS^1 peak of the ion of interest [30]. Peptides are identified thanks to MS^2 but quantified thanks to MS extracted signals. Less data points will result in a poor generated peak shape and therefore an imprecise quantification. This is even more true for low abundant proteins. Because of the stochastic nature of DDA, *i.e.* the selection of the most abundant precursors from the MS level for further fragmentation, it is not a reproducible method as repeated analyses will identify additional peptides and high abundance peptides will always be favoured in comparison to low-abundance peptides that are being neglected. This generates missing data that can eventually be filled thanks to bioinformatics using the match-between-runs algorithm [31]. The generated data are easy to handle as the tandem mass spectra can be matched to spectra from existing databases.

Data-independent acquisition

DIA strategies have been implemented to circumvent the variability and missing values problems observed with the DDA method. In DIA mode, all eluting precursors are isolated in broader m/z window and subjected to fragmentation in the second MS/MS step altogether. Unlike DDA, MS^2 will automatically be activated regardless of peptide intensities. The resulting generated fragmentation spectra deriving from multiple precursors ions leads to much more complex MS/MS spectra compared to DDA. Indeed, as several precursor ions are fragmented at the same time, the direct link between a precursor and its fragments is lost.

It exists different DIA approaches according to the vendor solutions. DIA methods are referred as SWATH (for Sequential Windowed Acquisition of All Theoretical Fragment ions, Sciex) [32], MS^E (Waters), DIA-PASEF (Bruker) or the DIA solution from Thermo. In the SWATH approach, a wide isolation window between 10 to 25 m/z units is required and moves forward along the mass range during the run time. The mass range is divided into smaller mass windows and the mass spectrometer fragments all precursors from this isolation window

to end up with an overall and precise acquisition of all the selected precursors. Identification is based on spectral libraries previously recorded thanks to DDA shotgun proteomics. The fact that comprehensive DDA-based libraries must be generated before interpreting the DIA data is the major limitations of this technique [30]. Alternatively, a new methodology called DIA-NN came to light and exploits deep neural networks and signal correction strategies for interpretation of DIA data acquisition [33]. SWATH-MS is used for identification but also for quantification but this approach is still 3- to 10-fold less sensitive than current gold methods for quantification such as selected reaction monitoring (SRM) or parallel reaction monitoring (PRM) (described later in section 2.3.4) [32]. In MS^E acquisition, the whole mass range is covered and either low and high collision energy alternates to respectively obtain precursor ion mass spectra and fragment ions information in a single analytical run [34, 35]. This latter acquisition technique was also coupled to ion mobility and help in the signal deconvolution [36]. Other specific acquisition methods included in the DIA approach exist but will not be discussed in this thesis.

2.3.4. Targeted approaches

Targeted MS techniques are essentially used for the detection and quantification of predefined analytes in complex samples [13, 37]. They differ from shotgun proteomics by selecting chosen precursors from proteins of interest, removing any other information from the complex sample but improving the selectivity and specificity [38].

In tandem MS, a two-step technique is necessary with the addition of a dissociation process or a chemical reaction to induce the fragmentation of an ion. In a common MS/MS experiment, the ionized molecules enter the first analyser, in general a quadrupole mass analyser, and are being separated by their mass-to-charge (m/z) ratio. One of them, called the precursor ion, is isolated and fragmented into product ions and neutral fragments in a collision chamber either spontaneously or by activation. A second analyser separates then the fragment ions by their m/z ratio and detects them [39].

Selected reaction monitoring integrates in this type of analysis by being one of the main scan modes of the tandem MS [40, 41]. The SRM principle consists in detecting a particular ion thanks to one or more transitions (precursor/fragment ion pair) of this ion. Both the first and the third analyser act as filters for the selection of specific precursor mass and specific fragment mass. In a complex sample, molecules are ionized and brought to the gas phase to be separated by their m/z ratio. The targeted ion is selected upon its m/z ratio in the first analyser, then fragmented by collision with a neutral gas to result in product ions in the second analyser (acting as a collision chamber) and finally one or several of these fragments are selected thanks to the third analyser. This scan mode is achievable within a triple quadrupole mass spectrometer [41] and is represented in figure 2.5 (a). SRM is also called multiple reac-

tion monitoring (MRM) and is the application of SRM: one precursor ion dissociates in multiple fragments and each of them are sequentially selected and detected. Within complex samples such as plasma or urine, the m/z ratio of precursor ion does not ensure on its own the specificity of a ion as interferences from isobars are too high. Selection of multiple specific fragments of the precursor ion guarantees a higher specificity of the approach by filtering out all the fragments coming from isobaric and/or co-eluting compounds. Although there is very little probability that different molecules co-elute, have the same m/z ratio and give the same fragment, the specificity of the technique can still be increased by using more transitions, *i.e.* selecting several fragments characteristic of the precursor ion. The non-scanning nature of this mode allows to increase the sensitivity by focusing on the precursor and fragment ions over longer times than conventional full scan techniques. Furthermore, the linear response of this approach makes it an excellent acquisition mode for quantitative analysis over a wide dynamic range. Besides, the multiplexing capacity of this method allows to target several peptides from the same protein and therefore reach a high statistical significance in term of results. Indeed, each targeted peptide in an SRM analysis is independently measured from other targeted peptides of the same protein to quantify.

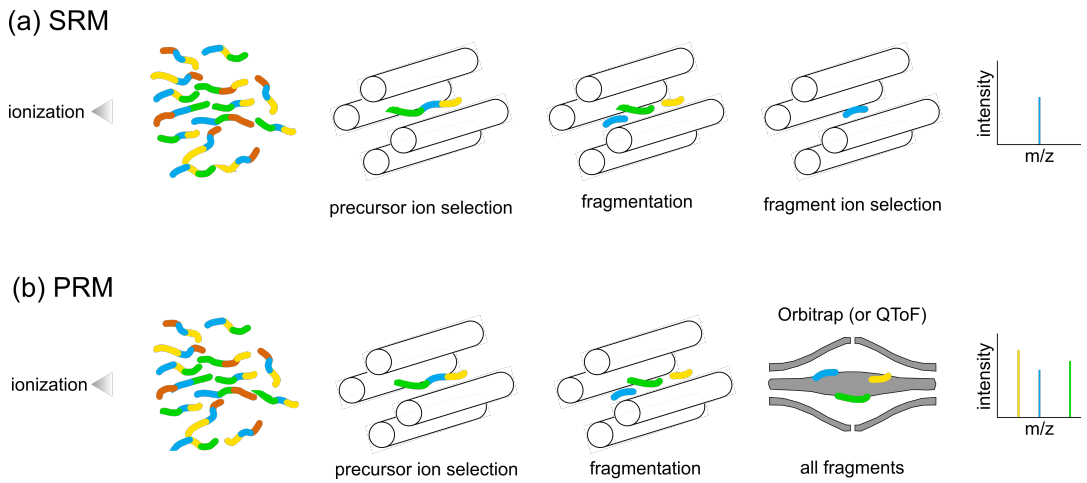


Figure 2.5: *Two approaches in MS-based proteomics (a) selected reaction monitoring (SRM) where each fragment ion from a precursor ion is monitored at a time and (b) parallel reaction monitoring (PRM) where all fragment ions from a precursor are analysed in concert.*

The initial step in developing a targeted approach is the careful selection of peptides to be targeted. They need to be unique to the protein of interest and easily detectable by LC-MS, which is called proteotypic peptide. They preferably should also be quantotypic, which means that they should accurately represent the level of the protein to consider. Quantotypic peptides are difficult to obtain as, for now, there is no complete understanding of all the factors that influence the peptide quantotypic behaviour. Amongst those factors, the need for having a good response in LC-MS, the requirement for no missed cleavage sites and absence of frequently modified amino acid residues (*i.e.* oxidation of methionine, deamidation of asparagine and

glutamine, carbamidomethylation of cysteine and phosphorylation of histidine) are the most commonly cited [42, 43]. Criteria such as peptide size and hydrophobicity must also be taken into account and wisely chosen. The fragments ensuing from the selected peptides should also be selected based on their intensity and low level of interfering signal [40].

Its high specificity, its high-throughput and its multiplexing capability make the SRM the method of choice to quantify protein in complex samples. This technique confers precise measurements with excellent analytical reproducibility and very low coefficients of variation [37]. This combination of advantages is perfectly adapted to the needs of biomarker validation and outperformed the multiplexing capacity of traditional immunological methods.

Another targeted approach has emerged from the SRM technique: the parallel reaction monitoring (PRM). PRM differs from the SRM by substituting the third quadrupole used in SRM by a high resolution mass analyser such as an Orbitrap or a QToF [44, 45]. Its implementation is shown in figure 2.5 (b). A precursor ion is selected in the first quadrupole, fragmented in the second and all these fragments are analysed in the last analyser to record the full picture of the selected ion. The method development is simplified in comparison to SRM because no transition must be chosen prior the acquisition. It gains attention in the field of quantitative proteomics as it combines both sensitivity and selectivity. Parameters such as fill time in an Orbitrap instrument or accumulation time in a Q-TOF instrument can be controlled to improve sensitivity while the selectivity is enhanced with the increase resolving power of the analyser in both MS and MS/MS mode [45]. PRM is therefore especially suitable for characterizing high complex biological samples and allows low limits of detection and quantification.

2.3.5. Standard for absolute quantitative proteomics

Different types of standards may be used for absolute quantification and are divided in three categories detailed hereunder: peptide standards, concatenated standards and entire protein standards.

Peptide standards

Peptide standards can be spiked either before or after the enzymatic digestion in the sample preparation process. Prior to the insertion in the samples, these standards must be chemically synthesized with the wanted labelled amino acids and further accurately quantified using amino acid analysis (AAA). Although introduced earlier, it is Gerber *et al.* [46] that apply for the first time a strategy for the absolute quantification of proteins termed AQUA. Today, AQUA peptides have become a trademark and defines heavy peptides that can be synthesized at the custom of the client.

As these standard peptides are introduced during the sample preparation process, some losses arising at the beginning of the protocol may not be taken into account. Indeed, handling, storage, yield of the enzymatic protein digestion, purification steps or fractionation steps may have impacted the real quantity of endogenous protein to target and consequently biased the true expression levels of this protein in vivo. Furthermore, the limitation of this method relies in the cost and time of synthesis of each specific needed standard peptide.

Concatenated standards

To circumvent the problem of tryptic digestion bias, some laboratories developed standards inserted in a chimeric protein that will help evaluating the digestion efficiency. Beynon *et al.* [47] first designed in 2005 the QconCAT (for Quantification of conCATenated peptides) protein composed of standard peptides specifically chosen to meet the need of the proteins of interest to quantify. These peptides originate from these targeted proteins allowing the concomitant quantification of many independent proteins. The protein is artificially created by gene expression in *E. coli.* with an added His-tag for further purification. In the work of Beynon, $^{15}NH_4Cl$ was used as the nitrogen source in the expression medium to provide the labelling and therefore differentiate the endogenous proteins from the peptides originating from the QconCAT artificial protein. The QconCAT protein is added to the sample under study before the enzymatic digestion and produces the labelled peptides whose known amount can be compared with the intensities of the unlabelled endogenous peptides for quantification. The advantages of the introduction of the peptides within a protein is that they are protected from adsorption to the tube sides, removing the bias from this potential peptides adsorption. Additionally, peptides are released in the same identical quantity as they come from a single protein. Furthermore, handling a single protein for the quantification of multiple intrinsic proteins reduces the pipetting bias in comparison to the handling of several synthetic labelled peptides. Finally, as it is introduced at the beginning of the sample procedure, it takes into account the losses from the purification.

Another peptide-concatenated standard called PCS was simultaneously developed by the Ito group [48] using a similar approach as QconCAT. Anderson and Hunter [49] described another comparable method with their polyprotein stable isotope-labelled internal standard (polySIS) approach which lies in the implementation of isotopically labelled lysine to obtain a consistent mass difference between the labelled and the natural peptides and hence facilitating the interpretation of the acquired data. Zeiler *et al.* and Boström [50, 51] developed QPrEST (Quantification of PRotein Epitope Signature Tag) standards which relies on protein epitope signature tag (PrEST) expressed in *E. coli*. These protein fragments are sequences incorporated with heavy isotope-labelled amino acids and are identical to short region of the human protein of interest. This standard is used in combination with SILAC to precisely quan-

tify the PrEST peptides.

Some limitations remain from these techniques such as the inconsistent mass shift between the labelled and unlabelled counterparts when uniform labelling is used. Incorporation of specific isotopic amino acids is possible and desired when a consistent mass shift between unlabelled and labelled pairs is required for the ease of data analysis. Another drawback is the fact that, upon proteolysis, the standard peptides do not act as the endogenous peptides from the sample because of the fold of the QconCAT protein. Even if the peptides are surrounded by natural flanking sequences (FS) (i.e. the amino acid residues that immediately precede and follow the peptide of interest within a larger protein sequence) to mimic the native protein, the digestion efficiency will not behave 100% similarly to the synthetic protein.

Protein standards

Among all the standards used for absolute quantification, protein standards are certainly the most ideal ones. Indeed, they exactly behave as the protein to quantify over the sample preparation process and they are spiked at the beginning of the protocol therefore taking into account the experimental induced variations impacting the quantification. Comparison between synthetic peptide standards with intact protein standards was investigated by Brun *et al.* [52] and demonstrated that peptides standards resulted in smaller quantified amount than with protein standards, probably due to the digestion efficiency and showed better accuracy with the protein standards.

Several groups developed protein standards, among the most popular is the PSAQ method, the protein standard for absolute quantification, now protected by a trademark [52, 53]. The main advantage of a full-length isotopically labelled protein standard is that it behaves exactly as the endogenous protein of interest and is therefore the gold standard. Another asset is that all peptides belonging to the protein are available for quantification in comparison to AQUA strategy where the number of peptides is restricted due to the cost and time to synthesize the wanted labelled peptide standards.

Limitations of this technique are essentially focused on protein production. High molecular weight proteins will be difficult or even impossible to synthesize and purification steps can be a laborious task as well as re-solubilization of the protein. Knowing that the fold of the synthetically produced protein could be different from the native targeted protein, digestion efficiency could be also influenced in that way. Intact protein standard are limited to non-modified proteins when produced in cell culture, preventing the applicability of the method to PTMs. In an attempt to allow the quantification of PTMs using protein standards, Singh *et. al* [54] developed a novel tool consisting of a protein standard tagged with a peptide that will be further used for the quantification of the produced protein. This labelled protein containing a

labelled FLEX-peptide comes in complement to the addition of unlabelled FLEX-peptides that allows interpretation of biological events (i.e. PTMs) by comparing the light peptide intensities with the heavy ones. Indeed, as the standard protein is not modified, heavy peptides from the standard protein act as references. Therefore any decrease in light peptide intensities can be considered as a modification. It was named FLEXIQuant for full-length expressed stable isotope-labelled protein for quantification. In a similar way, the RISQ protein (Recombinant Isotope Labeled and Selenium Quantified) was developed by Zinn *et. al* [55] in which the methionine residues are replaced by selenomethionine to allow quantification by ICP-MS. As in the FLEXIQuant project, the standard protein does not contain any modification, allowing the quantification of protein modifications.

2.3.6. Enzyme-linked immunosorbent assay

Enzyme-linked immunosorbent assay (ELISA) is a method allowing the detection and quantification of a substance, commonly a protein, in a liquid sample with the help of an antibody. The principle relies on the binding between an antigen from a sample to be studied and a specific antibody. Typically, the antigen to be tested is attached to particular surface. Then, its antibody counterpart is added so that it binds to the protein of interest. This antibody is either already linked to an enzyme or this latter is added as a second enzyme-conjugated antibody that recognises the primary antibody. Finally, the addition of the enzyme's substrate generates a colour that is further detected. The presence of the analyte is confirmed either qualitatively, by generating a colour when the complex antigen/antibody is created, or quantitatively, by measuring the absorbance of the sample and comparing it to the concentration of a standard curve of known concentration. There exist several types of ELISA tests that differ by the order of introduction of the antibodies and antigens and how they are coupled together (for further explanations see [56]), with advantages and drawbacks to either option.

In the biomarker validation phase, ELISA is the method of choice because of its sensitivity (up to picogram depending on the characteristics of the antibody-antigen interaction [5]) and specificity, that is highly attributed to the antibody-antigen binding approach. One of the primary advantages is its ability to perform multiple analyses in parallel. Nevertheless, the process cannot be scaled up or fully automated as it demands antibodies that are specific to the protein of interest. These antibodies are not necessary immediately available on the market and must often be generated which is a long and costly project. Furthermore, and once the specific antibody is obtained, even though the approach is known and well documented, the procedure requires methodological focus and improvement when a new antibody is used. The methodology has to be validated for each new target. Expertise is required to conduct this labour-intensive assay. Moreover, the multiplexing capacity of the ELISA is limited. Even though, there has been remarkable advances in the multiplex era of the immunoassays [57], they lag behind other strategies essentially because of a lack of automatization.

2.3.7. PTMs characterization

Post-translational modifications (PTMs) are covalent chemical modifications occurring in the cell after synthesis or throughout the life of the protein. They play a key role in the protein function through folding regulation, cellular localization or protein interactions and alter the structural integrity and stability of proteins. The study of these important mechanisms can help in diagnostic or prevention of disease and are therefore of great interest in the cell biology field. In 2019, Sharma *et al.* summarized some of the clinically relevant PTMs associated with proteins that are overexpressed in patients suffering from cancer [58]. MS-based techniques are considered as mainstays for PTM analysis because of their high level of sensitivity and specificity [59] Identification of PTMs can be challenging because of the stability of the modification, the efficiency of the detection tools, their low stoichiometry in complex samples and also the database searching algorithm. These obstacles will be successively explained in the next paragraphs.

The dynamic nature of proteins, temporally or spatially, as well as their changing modifications that add a layer of complexity in the proteome, makes the identification of PTMs a tedious work. Glycosylation, phosphorylation, acetylation, methylation, ubiquitination are all well-known modifications with precise functions in the cell. Depending on the nature of the modification, some PTMs can be irreversible or reversible, this latter one being therefore even more difficult to grasp as it is changing over time. Some modifications could occur in the cell and reverse or be less stable during the sample preparation process. Phosphorylation of tyrosines, for example, is a modification rapidly replaced by protein phosphatases [60]. Such modifications are then often underestimated because of their rapid kinetics. One has to consider kinetics and stability of modifications to be able to understand the full picture. Other chemical modifications could arise during the sample procedure. Consequently the biological relevance of such modifications becomes insignificant. Careful assessment of such modifications is worthwhile to allow the true measurement of an in vivo protein modification state.

PTMs are usually studied using “bottom-up” proteomics approach where peptides are first cleaved by a protease prior to MS analysis. The major limitation of this approach is the loss of link between the analysed peptides and their respective protein. Typically, trypsin is used as the enzyme to digest proteins. This choice limits the identification of several PTMs by removing peptides with extremely short or long sequences and therefore a reduced sequence coverage. The employment of other proteases or a combination of multiple proteases can help reaching higher digestion efficiencies and could increase the sequence coverage of proteins in complex mixture. Hence resulting in higher sensitivity of peptide analysis and their modifications. The majority of PTMs studies are acquired using the widespread DDA strategy. The major limitation of this acquisition method is the loss of information due to the selection of the TopN most intense ions. On the contrary, DIA can overcome this limitation by analysing all the detected fragments simultaneously and therefore achieve deep proteome coverage. However,

the use of DIA for PTMs analyses is still in its initial stage of development, even though some promising attempts are being submitted [61, 62]. While shotgun proteomics is well established for the PTMs analyses, the “top-down” approach begins to gain some popularity. By analysing proteins instead of peptides, one gets rid of the digestion process and provide a global view of protein isoforms and their PTMs [63, 64].

In a complex biological samples, peptides carrying a PTM are often drowned in the vast amount of unmodified peptides. Their low abundance affects their detection and special enrichment methods are thus necessary to detect them. By separating the modified and unmodified proteins or peptides thanks to immunoprecipitation-based techniques or chromatographic methods, the complexity of the samples decreases and allows for better identification of the modified peptides.

The identification of PTMs is generally done using database searching. Experimental MS/MS spectra are matched to theoretical ones using database search tools in which the user has included the modifications to target. As the proteome under study is potentially full of PTMs that are not specified by the user, either known or unknown modifications, a large number of spectra are being sidelined and therefore remained unexploited. New types of search have been developed to expand the search space. The “open search” strategy relies on the identification on peptides with any mass shift, corresponding to one or more modifications [65]. This prevents from specifying the modifications in the search and allows for the discovery of new modifications. A drawback of the open search method is the identity and the localization of the modification within the peptide which is unknown and must be recovered using other computational characterizations. Another strategy is the “spectral pair” search where the MS/MS spectrum of an unmodified peptide (previously identified in an initial search) is compared to unassigned MS/MS spectra with similar retention time to that of the unmodified peptide to provide mass shift differences corresponding to PTMs present in the sample [66].

2.4. Quantitative proteomics pitfalls

In section 2.3, various standardization approaches were discussed, each with its own advantages and limitations, and the choice of which to use depends on the specific type of quantification needed. Despite the emergence of new proteomics strategies over the years that aim to improve accuracy, reproducibility, throughput, and minimize bias, there is still no perfect quantitation method that is both fast and effective, with minimal bias throughout the entire sample procedure. As shown in figure 2.6, each strategy is incorporated at a specific stage in the protocol. The earlier the quantitation step is incorporated, the more accurate the quantification. Additionally, the shorter the protocol, the less in vitro bias is introduced.

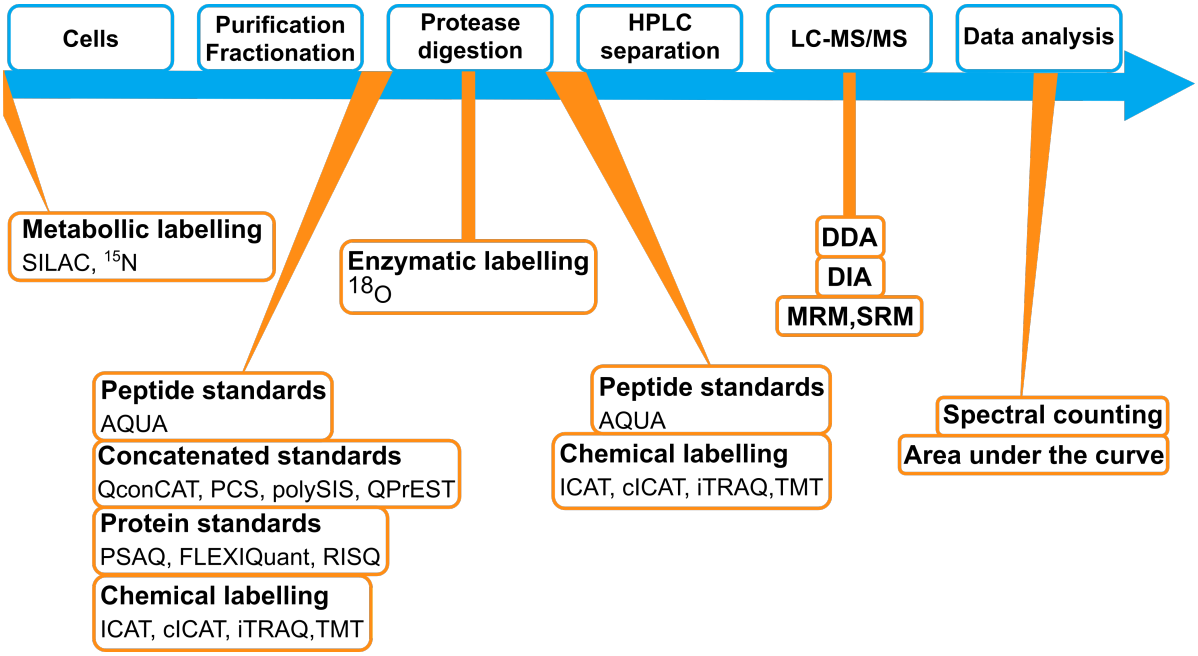


Figure 2.6: *Overview of the incorporation of the quantitation strategies into a typical sample preparation workflow.*

Some strategies are limited to cell culture such as SILAC, where the mixing of the samples under study is performed right after labelling leading to less in vitro variability between the two samples and with a result depicting the true in vivo difference of the two conditions. Isobaric and isotopic labelling referred as chemical labelling in figure 2.6 can be introduced before or after protein digestion, but combining differentially labelled samples at the protein or peptide level introduces errors during the previous stages of purification/fractionation and/or enzymatic digestion. Peptides, concatenated and protein standards also encounter the same problematic. Enzymatic labelling inserted during the protease digestion can also suffers from quantitation bias and incomplete labelling. Post-processing techniques, such as feature detection, retention time alignment, noise reduction, peak picking, and normalization of MS intensities, are often applied in combination with different strategies to correct for differences in sample preparation and instrumentation variations. Normalization methods vary and include summing the intensities of all peptides, taking the mean of all peptides, using internal standards or constant proteins, among others. The choice of strategy depends on the type of samples and biological information required.

Absolute quantification is essential for biomarker validation and routine analyses. Targeted label-based approaches are the most reliable techniques for clinical applications of biomarkers. Standards are spiked in known amounts during the experimental procedure and must undergo AAA before utilization. In a typical proteomic workflow, proteins are submitted to a purification/precipitation step to concentrate them and remove any contaminants. Then the disulphide bonds are reduced, which serves to destabilize or denature the protein structure. To prevent the disulphide bonds from spontaneously reforming, an alkylation step is carried

out. The protein is subsequently digested into peptides using enzymes. A purification step is sometimes performed on the peptides to clean the sample after digestion. The peptides are then separated by liquid chromatography before entering the mass spectrometer, with standards for absolute quantification inserted at certain points in the protocol. Protein standards, considered the gold standard, are spiked before the purification step to take into account all the variability of the protocol. However, synthesizing the protein of interest is costly and time-consuming. PCSs are cheaper, and multiple endogenous proteins can be quantitated using only one artificial protein standard. Although this standard is introduced before enzymatic digestion to remove any systematic errors associated with the digestion step, the cleavage efficiency between the analyte and its standard is not always perfectly equal, resulting in quantification bias. It has been demonstrated that the amino acids surrounding a trypsin cleavage site directly affects the digestion efficiency [67–70]. The initial use of QconCAT did not include the natural flanking sequences (FS) of the inserted peptides in the standard protein. However, during this time, the peptide-concatenated standard (PCS) was developed, which included the natural FS of the tryptic peptides. Different studies demonstrated the importance of the incorporation of natural FS into the PCS to accurately quantify proteins [48, 52, 71]. Standard peptides that can be spiked either before or after the protease digestion cannot consider the bias from the efficacy of tryptic cleavage. Nevertheless, they help in diagnosing losses of peptides during the digestion and the potential purification step that follows, or act as references for quantification if injected before LC separation. However, caution should be taken as LC separation may also induce losses that the inserted standards do not consider.

In the realm of quantitative proteomics, researchers have sought to standardize various steps in the sample treatment process. Some groups have focused on optimizing i) the efficiency of enzymatic digestion, ii) the reproducibility of measurements, and iii) the control of LC-MS separation. As far as the enzymatic digestion yield is concerned, Burkhart *et al.* [72] did not use stable isotope labelled peptides as standard but rather a systematic approach for optimal and robust digestion of proteins. A four-steps procedure was applied for the determination of digestion yield, reproducibility and quantitation and is recommended to optimize digestion conditions. Proc *et al.* [73] showed that the yield obtained with stable isotope peptides is highly dependent on digestion conditions, which must be well-tuned, although there is no one-size-fits-all solution since protein-dependent factors are at play. This statement leads to the conclusion that the research of the best proteomic protocol is an utopia and that optimization must go through the understanding and control of the limiting steps. This, of course, must stand in the more reproducible manner possible, *i.e.* with the use of perfectly standardized and reproducible protocols. Regarding the reproducibility of measure, 13 laboratories collaborated on inter-laboratory studies aimed at developing proteomic standards [74] to improve inter-laboratory reproducibility. Concerning the control of the chromatographic separation, quality of the raw data is directly linked to the HPLC function. Highly stable and reproducible LC separations are required for accurate quantification. Burkhart *et al.* [75] developed a quality control composed of a mixture of synthetic peptides that cover the entire

standard LC gradient and allow for the monitoring of the parameters and performance of the LC separation. Finally, Escher *et al.* [76] developed a robust method of RT prediction that helps minimize RT variance between runs and improve robustness.

The proteomic community has exerted much effort in the research of a perfect standardized method for precise protein quantification. Researchers have made progress by targeting specific stages of sample treatment over the years [77–79]. However, none of the current approaches account for all the potential biases that may arise during the sample preparation process, as highlighted in this thesis.

3

Kit Quanta

3.1. Description and Strategy

In the context of the research of a gold standard for protein quantification in complex matrices, we developed a toolkit that takes into account the experimental variations occurring during the sample treatment namely the digestion efficiency of any proteomic workflow, the control of the LC-MS separation, the reproducibility of measure, etc. We invested great efforts in the design of the kit to universalize it. Therefore, any proteomic laboratory may use it, regardless of the preparation processes and instrumentations.

The kit is composed of four different elements added at specific moment of the sample preparation process and its representation is depicted in figure 3.1:

1. A protein that, after digestion, generates unique peptides, not existing in any biological human matrices and that are subsequently monitored. Working with an entire protein allows to spike the sample at the beginning of the process and therefore evaluate the bias at the dawn of the experiment.

   As the sample preparation process comprises several steps, it is important to assess the bias of these different steps. For that purpose, three forms of labelled unique peptides (the same as in the chimeric protein) are introduced in the kit:

2. Tryptic peptides with FS inserted at the same time as the chimeric protein. They will help seek differences in digestion efficiencies whether the peptides are flanked on either side by a certain number of amino acids or embedded in the protein.

3. Tryptic peptides without FS also inserted before digestion to track bias unlinked to the enzymatic digestion.

4. Tryptic peptides without FS introduced just before the LC-MS injection that act as references.

This four-elements composition kit is supplemented by labelled specific peptides of the protein to quantify when used in real life-size experiment. This last element ensures an exact quantification of the protein to quantify whereas the kit allows to precisely point out the failure of the experimental procedure if any.

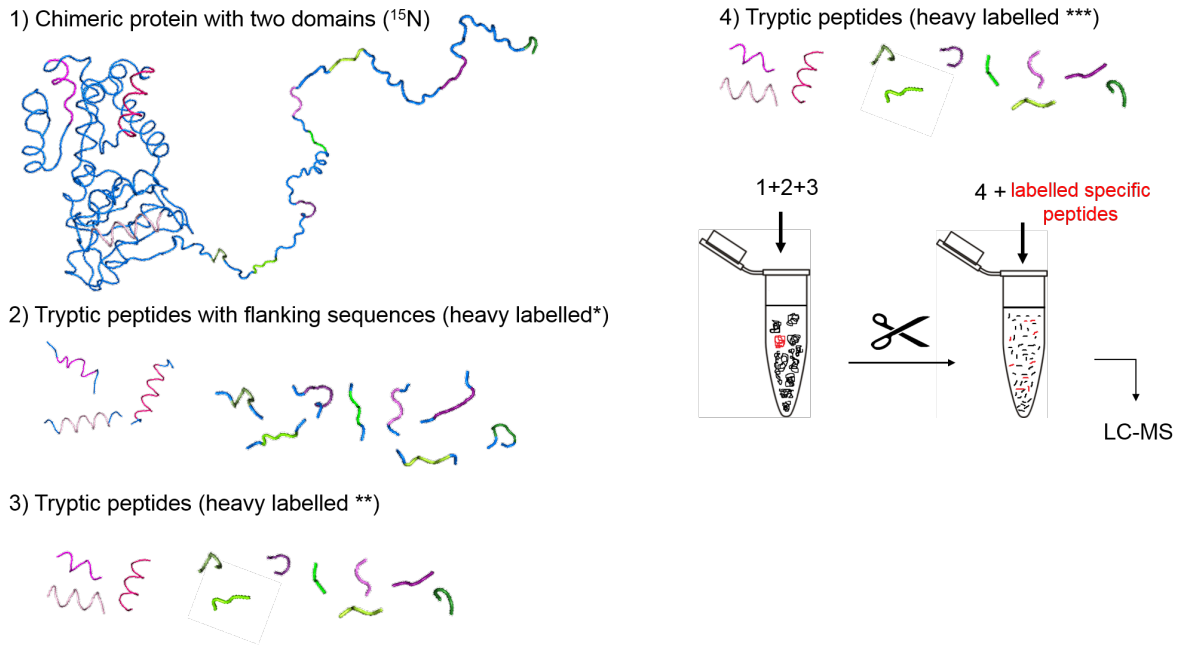


Figure 3.1: *Schematic representation of all the elements of the kit. The different levels of peptides are differentially labelled. Labelled specific peptides coming from the biomarker to quantify is also introduced at the same time than the fourth level and is a complement to the kit that is discussed in section 3.1.6.*

The purpose and choice of these elements is the principal subject of the next paragraphs.

3.1.1. Impact of the structure on the protein digestion yield

We created a standard protein integrating two different structural configurations: a structured part associated to an intrinsically disordered domain. The goal of this chimeric protein is to evaluate the digestion efficiency of the experimental protocol applied and therefore monitor the accessibility of cleavage sites in function of two structural states belonging to two different domains of the synthetic protein. To evaluate the digestion efficiency, a comparison between the digestion yield from a peptide belonging to the entire protein and the one generated from

the same peptide sequence having in addition short FS at both ends is considered. If clear differences arise from the two domains, one can assume the propensity of the structural domain of the protein to play a role in this difference which imposes a modification of the digestion protocol to ensure an optimal digestion. The efficiency of digestion can be influenced by various factors, including the quaternary structure (which refers to the three-dimensional arrangement of the protein), the tertiary structure (which involves the interactions between the protein and its environment such as hydrogen bonds, hydrophobic and hydrophilic interactions, and ionic bonds), the secondary structure (which is the specific arrangement of the peptide within the protein, including α-helices, β-sheets, turn, coil, etc.), and the primary structure (which is the sequence of amino acids in the polypeptide chain). All these levels of protein structure are interconnected and influence the enzymatic digestion. The specific conformation of a protein's three-dimensional shape is determined by the sequence of amino acids. The enzyme preferably reaches solvent exposed and accessible amino acids, while cleavage sites that are buried deeper in the protein structure are inaccessible. Additionally, the effectiveness of the enzyme is also influenced by the amino acids surrounding the cleavage site.

A fusion or chimeric protein is the assembly of two different proteins or part of proteins to generate a final product with the desired properties of each separated protein in a single construct. In the case of this project, the characteristic of each domain is based on its conformation in the protein which must be structured in its first part and completely disordered in a second part. The chimeric protein that we have obtained was synthesized by our collaborators at the University of Mons using two proteins, namely Q0KC03 and Q9F3S0, which they have extensively studied in the past. The model protein for the structured domain is the TRAP-type transporter (Uniprot ID: Q0KC03) from *Cupriavidus necator* (a Gram negative bacteria) of 342 amino acids. The secondary structure of this protein has been studied in detail using X-ray diffraction [80] (PDB: 4P8B) and is composed of various arrangements such as β-sheets, α-helices or turns. It is represented in a 3D view in figure 3.2.

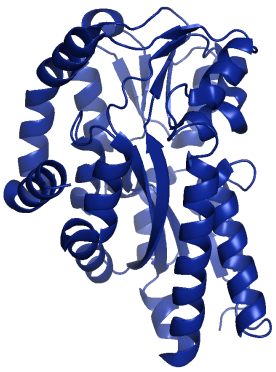


Figure 3.2: *3D view of the PDB structure (4P8B) of model protein for the structured part of the chimeric protein.*

The CopB protein (Uniprot ID: Q9F3S0) from *C. metallidurans* is used as the model protein for the disordered part, consisting of 462 amino acids. Apart its extraordinary capacity to resist

to a heavy metal environment, this protein is known to have a disordered sequence region from amino acid 1 to 194 as mentioned in Uniprot from the MobiDB-lite automatic assertion method [81]. The motif KMQGMDQGSMQGMDHS, beginning at amino acid 42 and constituting the disordered area, repeats itself and assembles together eight times with a longer motif close to the repeated one KMQGMDQGSMQGMDQGSMQGMDHS to give a resulting 153 amino acids sequence (figure 3.3) used to make the disordered part of our protein.

MVRRQTKQRRIHAPRRRHPTAKKQGSTPSAGQGSMQGMDHSKMQGMDQGSMQGMDHSKMQGMD
QGSMQGMDHSKMQGMDQGSMQGMDHSKMQGMDQGSMQGMDQGSMQGMDHSKMQGMDQGSM
QGMDHSKMQGMDQGSMQGMDHSKMQGMDQGSMQGMDHSKMQGMDQGSMQGMDHSKMQGMD
QGSMQGMDHSKMQGMDQGSMQGMDHGAMQMQGGSRPPDARDPDAYSDGAKLGVGQYAIGPSRS
LHMADEHIFASVLVDRLEWTKGNESNAASYEAQAWIGNAFNKFVIKAEGEVEKGKVPEARTELLWGHAI
ATYWDTQLGVRNDAGYGRPARNWLAFGVQGLAPYWFEVEATGYVGTEGRTAVRLSGEYDLLLTQRLIL
QPRLEASLYGKNDPEIGNRLSGLSSGAVGVRLRYEFSRQFAPYIGIERSQSFGNTANMVRASGGRSGET
RFVAGIRMWF

Figure 3.3: *Sequence of the CopB protein with the repeated motif of 153 amino acids highlighted.*

Although it is feasible to connect the two domains directly, we opted to use a linker between them to enhance protein folding and production yield. The GGGGS linker was chosen specifically for its small amino acid size, which promotes flexibility and optimal mobility of the domains. The protein conformation should be thoroughly and experimentally analysed to demonstrate its utility and further performance. For this purpose, a cleavage site will be introduced between the two parts of the protein, besides the linker. Once synthesized, the protein could therefore be structurally analysed in its entire form as well as in the two parts of the protein separately, to render the conformational analysis easier. This cleavage site will only be inserted for development purposes.

As cysteines were absent from both model proteins chosen, no cysteine residue is present in the final chimeric construct. Even though disulphide bridges can stabilize protein structures, cysteines were avoided as they are known to complicate recombinant protein production in bacteria by misfolding or aggregation through mis-pairing of cysteine residues.

The chimeric protein was produced using a bacterial expression system, namely *E. coli*. After optimization of the codons for the expression in *E. coli*, the DNA coding for our protein was inserted into a plasmid expression vector pET, designed by ATUM (Newark, CA, USA). The corresponding plasmid was introduced into a bacterial cell to produce the protein of interest. Protein production and purification has been realised by the Proteomics and Microbiology Laboratory of the University of Mons, our collaborators. For the sake of purification, a polyhistidine-tag (6× His tag sequence) was added at the C-terminus of the protein. The critical steps here are the possibility to obtain a protein which is insoluble or forming inclusion bodies or else difficult to purify.

The proteins will not be utilized in their current state, as certain peptide sequences will be substituted with more functional sequences, which will be discussed in the following section. The difficulty of this stage is to retain the two desired structures.

3.1.2. How to choose the peptide sequences integrated into the synthetic protein standard

In addition to the chimeric protein releasing its peptides during digestion, three other sets of peptides are added at specific points during the sample handling process, as shown in figure 3.1. The 1^{st}, 2^{nd} and 3^{rd} levels are added prior to the enzymatic digestion step, while the 4^{th} level is introduced just before the LC-MS analysis. Each level contains the same set of peptides and can be differentiated by their respective isotopic labelling, monitored by MS. By incorporating each set of peptides at specific stages of the process, any issues with the sample handling procedure or specific samples can be precisely identified. Additionally, any deviations during an experiment can be localized using this comprehensive approach.

Some of 1^{st} level peptides (digested from the chimeric protein) are associated with the structured portion, while others are associated with the disordered portion. They allow to assess the obtained bias from the very beginning of the sample process namely the likely purification of the complex matrices and its overall digestion procedure. As some peptides are endogenous to the matrices of interest and other are exogenous (explained further), the first ones are flanked with their three natural amino acids on either side of the peptide, to perfectly reflect what happens in the endogenous protein.

The 2^{nd} level peptides are made of peptides including three flanking amino acids on either side of the peptide. These are useful for the determination of digestion efficiency, without reflecting the influence of the protein structure, but only the primary structure of the peptide chain. It is currently accepted that, six amino acid residues on either side of the peptide are necessary for providing reliable quantification of a protein [69], *i.e.* to obtain identical rates of proteolysis from the flanked peptides and the chimeric protein. In that context, the insertion of twelve supplementary amino acids for each peptide would have greatly affected the structure of the chimeric protein and considerably increased its size as well. Therefore, a good compromise to obtain an equilibrium between proteolysis differences (from the flanked peptides and the chimeric protein) and the 3D structure of the chimeric protein was to add three amino acids on both sides of the standard peptides, only. Keeping in mind that, depending on the tested protocol containing or not a denaturing step before digestion, the variation in trypsin cleavage efficiency will only be attributed to the difference in the primary amino acid flanked sequence, or allocated to this flanked sequence difference including structural effects of the protein.

The 3^{rd} set of peptides, composed of tryptic peptides, allows to evaluate the loss of peptides due to the digestion experimental procedure (degradation, peptide modification, adsorption due to lyophilisation…) but not the digestion itself, as the peptides are already tryptic.

The 4^{th} level peptides act as reference peptide set, used for quantification.

Thus, each step of the sample process will be scrutinized to assess the efficiency of the whole sample preparation protocol starting from the gathering of the complex biological sample to its LC-MS analysis.

The set of standard peptides introduced in the kit at different levels were carefully chosen according to several criteria related to the quantification, the chemical synthesis and their intrinsic properties (see hereafter). As far as quantification is concerned, a good standard molecule must ionize well, be detectable and gives a stable intense signal. The kit developed in the frame of this work wants to be “universal”, meaning that the inserted peptides must be artificial to serve as standards for any kind of biological samples. They must also be categorized as “proteotypic peptides”, meaning that they are found in only a single known protein and therefore serves to identify and correctly quantify this protein. Being artificial implies that they already meet the proteotypicity requirement as they uniquely identify to a chimeric protein. They must accurately represent the protein to quantify in a stoichiometric manner and therefore should be free of PTMs and must not contain cleavage site. Peptides containing amino acid residues prone to modifications during the sample preparation (*i.e.* oxidation of methionine, deamidation of asparagine and glutamine, carbamidomethylation of cysteine and phosphorylation of histidine) were systematically discarded. The peptide length, directly related to the m/z value, must be controlled to be ideally around 20 amino acids long, to provide classical m/z values, corresponding to the mass range of the instrument. This property also provides peptides representative of the average length of classical tryptic peptides, obtained from natural proteins. To be suitable as a good set of standards, it is preferred that the chosen peptides evenly distribute to a typical proteomic reverse-phase gradient to cover the wide range of possible retention time in a complex matrix. For easier peptide synthesis, it is recommended to reach a maximum of 60% hydrophobic amino acid residues.

As explained, the set of peptides is globally composed of exogenous peptides (exogenous to the biological matrices), but endogenous peptides are also added to the set to bring a supplementary and unique value to the kit. These endogenous peptides are carefully chosen from an endogenous protein (to the matrices of interest) that is commonly clinically measured: the complement component 4 (C4) protein. Figure 3.4 shows the theoretical intensity of each level of peptide that we can obtain after an experimental procedure. In an ideal world, the intensity of each level will be the same as no bias would arise from the protocol. Nevertheless, in reality, loss occurs at each step of the process and is depicted in figure 3.4. For a same quantity of protein/peptides injected, peptides coming from the ^{15}N protein (1^{st} level of peptides) are less

intense than peptides coming from the 2^{nd} level of peptides (*), themselves less intense than the 3^{rd} level of peptides (**), themselves again less intense than the 4^{th} level of peptides (***) acting as the reference set. Colours from figures 3.4 refers to the colours used in figure 3.1: in green are the endogenous peptides and in pink the exogenous ones. Based on the reference 4^{th} set of peptides (***) in pink (exogenous ones) and the 1^{st} level of peptides (^{15}N), both introduced in the same known quantity, the yield of the all sample preparation process can be evaluated. Based on the reference 4^{th} set of peptides (***) in green (endogenous ones), a precise quantification of the complement C4 protein can be reached. The other levels (* and **) serve as quality control to precisely highlight the bias in the middle of the sample preparation and allow to improve and change the protocol to lose less.

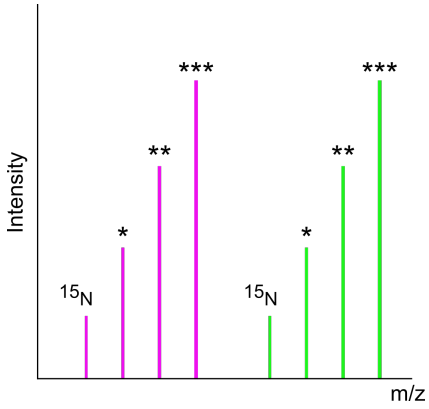


Figure 3.4: *Theoretical intensity of each level of peptides: ^{15}N = 1^{st} level of peptides, * = 2^{nd} level of peptides, ** = 3^{rd} level of peptides, *** = 4^{th} level of peptides. Colours from figures 3.4 refers to the colours used in figure 3.1: in green are the endogenous peptides and in pink the exogenous ones.*

Let us assume that the value obtained from the clinical assay measurement for the complement C4 protein is the true value. If we compare this value and the one obtained from the LC-MS analysis described above, there will be a difference between these two values which is due to the bias from the sample preparation process. With the calculation of the bias obtained on the chimeric protein, can we transfer the bias obtained from the chimeric protein to obtain the true value of the complement C4 using LC-MS ? Unfortunately, the answer is no, because the conformation of the chimeric protein and the endogenous one, although similar, are not identical. This is why the true value obtained by the clinical assay is useful: it allows us to calculate a yield for the complement C4 thanks to 4^{th} set of peptides (***) in green. This enables the comparison of yield of purification between an intrinsic protein of the studied matrices and the one from the chimeric protein.

To differentiate peptides belonging to the chimeric protein and the endogenous ones, the chimeric protein is expressed in a media containing a ^{15}N nitrogen source, typically ^{15}N-amonium chloride to yield a fully ^{15}N isotopically labelled protein. To be able to use the kit properly, we must be able to distinguish all levels of peptides from each other. As we also want to evaluate the digestion efficiency thanks to the second level of peptides (the ones with the FS) and as their proteolysis may be incomplete, the mass of the peptides with and without

FS must also be distinguishable by mass measurement. While the 1^{st} level peptides will be fully ^{15}N labelled, the other peptides, chemically synthesized, will be doubly labelled with ^{13}C and ^{15}N isotopes. The main challenge is to wisely choose which amino acids are doubly labelled to avoid overlap in the isotopic distribution of each peptide. To facilitate the visualization of these sets of peptides in the mass spectrum, the peptides are decreasing in mass from the 2^{nd} level to the 4^{th} level, the last level consequently being the lighter one. One level of the non-natural peptides can be entirely not labelled as by definition they will not be present in the complex matrices.

For a large range of unique peptides, we will obtain details of yield and recovery of each peptide at precise incorporation time point of the sample preparation method. They will allow to monitor and optimize the biases introduced during the different steps of the sample procedure.

3.1.3. Quantification in complex matrices

Standardization of quantitative measurements of proteins in biological matrices is a hard task as the complexity or concentration of the fluids are different from matrix to matrix. Thus the adoption of one specific protocol for all matrices seems inappropriate. The kit has been specifically designed to be extremely useful for biological dosage in plasma, urine and CSF but can also be applied on several other fluids in evaluating the efficiency of the sample handling. The endogenous peptides included in the chimeric protein actually are present in the three most common human biological fluids studied nowadays: plasma, urine and cerebrospinal fluid (CSF). These matrices are of major interests for the research of protein molecular signatures for prognostic, diagnostic and therapeutic purposes. Their collection is non-(to minimally)invasive.

3.1.4. Different instrumentations

Transfer an MS-based quantification method from one laboratory to another is a difficult task that involves different operators at different times and conditions. One can rapidly struggle to standardize the whole method when it comes to biomarker validation. To overcome this obstacle, we decided to assess the reproducibility of our kit in two different laboratories equipped with different instrumentations. This multisite evaluation was not limited to the validation of all the components of the kit in a real quantification assay but took place from the outset of the development of this tool kit. Having two laboratories throughout the whole elaborating process is a real asset as issues arising from one laboratory will not necessarily appear in the other lab and inversely. The complementarity of both sites will help anticipate and precociously detect any disruption of the elements of the kit.

Distinct setups were used on both sites. An ACQUITY UPLC M-Class System (Waters, Milford, MA, USA) composed of two successive columns: a nanoEase M/Z Symmetry C18 Trap Column (100 Å, 5 µm, 180 µm ×20 mm) and a nanoEase M/Z HSS T3 Column (100 Å, 1.8 µm, 75 µm ×250 mm), both commercialized by Waters, was used in Liège. An Ekspert nanoLC 425 with a YMC Triart C18 trap column (120 Å, 3 µm, 0.3 mm x 5mm) followed by a YMC Triart C18 column (120 Å, 3 µm, 0.3 mm x 150 mm) commercialized by YMC Europe GmbH (Dinslaken, Düsseldorf, Germany) was used in UMons. These sample separation systems were respectively hyphenated to a Q Exactive™ Hybrid Quadrupole-Orbitrap™ mass spectrometer (Thermo Fisher Scientific, Pittsburgh, PA, USA) and a TripleTof 6600+ (AB Sciex part of Danaher, Washington, WA, USA). The development phases essentially focused on the Q Exactive™ Hybrid Quadrupole-Orbitrap™ mass spectrometer for the MS analysis in Liège. The choice of this high resolution mass spectrometer for the conception phase was driven by the need to evaluate the impurities degradation products, purity of labelling of the kit quanta peptides with a high mass accuracy and dynamic range. Known for its excellent selectivity and sensitivity, selected reaction monitoring (SRM) analysis executed on a Xevo TQ-S Triple Quadrupole mass spectrometer (Waters, Milford, MA, USA) will be later used during the final phase of quantification assay.

3.1.5. PTMs and quantitative studies

Peptides easily subject to modifications are generally discarded from accurate protein quantification study. They will effectively introduce bias in the quantification because their intensities will not reflect the injected quantity in the sample but a reducing one due to a certain percentage of this standard being modified during the sample preparation. Even if the non-quantotypic peptide is added as standard at the end of the sample preparation process before the LC-MS analysis, error in quantification can still rise during this short time period. In any cases, the percentage of modification may not necessarily be equal between the standard and the analyte to quantify.

Despite these obstacles, it seemed necessary to create a kit that will allow to assess the level of frequent chemical modifications occurring in peptides. For this purpose, we introduced a supplementary set of non-quantotypic peptides dedicated to assess this phenomenon. Two frequent modifications will be taken into account by the kit and will therefore take the form of two peptides, each for one modification. These should be chosen among the endogenous peptides to have the possibility to be quantitated before the addition of the standard to the sample. Three endogenous peptides will be chosen for the accurate quantification of the endogenous protein present in plasma, urine and CSF. Thanks to this quantification, we could inversely found the intensity of the potentially modified peptides coming from the endogenous protein and deduce the percentage of modification for these peptides. The condition being that, at the time of the introduction of the peptide standard (2^{nd}, 3^{rd} or 4^{th} level), their percentage of

modification must be perfectly controlled. One possibility is to lyophilize these peptides to prevent them from modifications and solubilize them right before their introduction in the sample. Taking into consideration some modifications is an asset in comparison to other contestants who develop standardization kit of quantification.

It is generally admitted that a minimum of two peptides is required for accurate quantification of proteins. For even better accuracy, we opt for three endogenous peptides. Two non-quantotypic peptides containing one frequently encountered modifiable amino acid each and belonging to the same endogenous analyte are also added to the kit. A total of five endogenous standard peptides coming from a protein shared by three frequent biological fluids are thus inserted in the kit.

3.1.6. Complement to the specific labelled peptides of targeted proteins

The aim of the kit is to act as a quality control of the whole sample workflow starting from the sample storage to the LC-MS analysis. Any variability of the process will be highlighted by the different sets of standard peptides or protein appended in the kit. In addition, complement C4-A protein could be absolutely quantified thanks to the endogenous standard peptides. For biomarker quantitation, specific labelled peptides of the targeted proteins are required and can be inserted at the end of the experimental protocol with the fourth level of standards as depicted in figure 3.1. These two complementary elements will help us to i) select the best protocol that allows to reduce the bias at any step of the process, ii) to control the performances during a longitudinal study and iii) to quantify the protein of interest thanks to the use of their heavy synthesized peptide counterparts.

3.2. Results and discussion

3.2.1. Theoretical and experimental establishment of endogenous and exogenous target peptides

Nowadays, selection of peptide candidates for targeted MS-based proteomics quantitative assays has been largely facilitated by data sharing of laboratories through public repositories. Limiting the peptide choice to its own empirical data only offers a list of peptides from one single type of instrument and therefore reduces the panel to a specific ionization technique, mass spectrometer and often same biological matrices. ProteomeXchange [82] is a consortium gathering MS proteomics data around the world and contain repositories such as Proteomics Identification Database (PRIDE) [83], PeptideAtlas [84] or Mass Spectrometry Interactive Virtual Environment (MassIVE) [85] among others. The access to these open-source

public empirical databases greatly increases the success rate of an experiment containing standard peptide analytes. Indeed, if they are present in a database, it means that they have already been detected by MS and so that they ionize correctly. Theoretical approaches based on bioinformatics tools can predict peptides detectability, retention time, proteolysis but are generally less reliable than empirical methods. Nevertheless, in some cases where information is missing, bioinformatics can be of great help.

For our purpose, PeptideAtlas seemed the best data repository as it comprises data from a large set of proteomics experiments collected over several organisms and allows to apply various research criteria. For the selection of the exogenous peptides, we selected different human databases present in the PeptideAtlas repository and applied some filters to obtain proteotypic peptides, namely a uniquely mapping peptide ($N_{protein}$ mappings = 1), multiple observations (number of spectra identified to this peptide N_{obs} > 1), allowing to detect a protein via this peptide (relative to other peptides from the same protein) thanks to an empirical observability score > 0.3 and with no missed cleavage. Only peptides with a peptide length between 6 and 20 amino acid residues were kept and those containing methionine, asparagine, glutamine, cysteine and histidine were systematically removed for better quantotypic behaviour and hence better quantification. To facilitate chemical synthesis, peptides with successive amino acid repetition were discarded. In order to select the exogenous peptides, one has to change at least one amino acid in the sequences from the human database. Consequently, the endogenous sequences become exogenous as desired. A comparison between these new modified peptide sequences against a database is essential to ensure the non-equivalence in the human proteome and is done thanks to the Basic Local Alignement Search Tool (https://blast.ncbi.nlm.nih.gov). These selected peptides were then classified according to their hydrophobicity index using the Sequence-Specific Retention Calculator (SSRCalc) [86] to efficiently spread the set of peptides along the elution window.

The same criteria were used in PeptideAtlas for the selection of the endogenous peptides with the addition of specific residues prone to modifications. The choice of the modifications to be targeted in the kit was evaluated thanks to an inter-laboratory statistical analysis that has been conducted between ULiège and UMons on the same plasma sample. We evaluated the extent of frequent modifications on our respective instrumentations. Data treatment was homogenized to be able to compare the results. Methionine oxidation and deamidation of asparagine or glutamine are common modifications both observed *in vivo* or *in vitro* experiments and we therefore focused our analyses on these modifications. Oxidation of methionine-containing peptides of 26.7% and 24.0% were detected with the instrumentations of Liège and Mons respectively. Deamidation of asparagine and glutamine-containing peptides represents 12.2% and 6.0% in ULiège datasets and 9.0% and 2.5% in UMons datasets. These similar results between our two laboratories shows that methionine oxidation and asparagine deamidation are the most prevalent modifications and have therefore an interest to be considered in the standardization kit. One peptide with one methionine and another one with an asparagine

were then inserted in the set of five endogenous peptides. The three other peptides being bereft of modifiable amino acids. We first focused our research on abundant proteins present in the three matrices of interest (plasma, urine and CSF) such as albumin (Uniprot ID: Q56G89) or lysozyme (Uniprot ID: P61626). But these proteins are often depleted from complex matrices before any other sample preparation therefore drastically reducing the number of laboratories that could use the full potential of the kit. On the basis of the Kit Seppro® IgY14 commercialized by Merck (Darmstadt, Germany), we discarded these 14 proteins as potential endogenous peptide's candidates because of their likely possibility of being removed in many proteomic laboratories worldwide where depletion is often part of a typical proteomics protocol. Lists of proteins from human plasma, urine and CSF were merged and correlated with a reference source of clinically dosed proteins from the University Hospital Centre of Liège accessible online (https://www.chu.ulg.ac.be/jcms/c_355424/fr/referentiel-des-examens-realises-par-l-unilab-lg) to have an external reference point for the quantification. Only one protein meets all these criteria: the complement C4-A (Uniprot ID: P0C0L4). Discovery analyses were then realized on plasma, urine and CSF to select common peptides from P0C0L4 between the three matrices. Plasma sample from Belgian blood donors was obtained from the Belgian Red Cross; Urine sample was a pool obtained from 6 healthy donors from the MSLab; CSF sample from pooled donors was purchased from Tebu-Bio. Normal color of CSF should be clear and colorless, the sample from Tebu-Bio is yellow indicating a probable blood contamination. A protein assay of the CSF sample indicated a 10- to 20-fold higher protein concentration than normal. Blood contamination of CSF is known to be difficult to avoid during lumbar puncture for collection of CSF [87]. However, PeptideAtlas search within the repositories showed the presence of the complement C4-A in the three matrices. We matched the lists of complement C4-A peptides from the three matrices and only one peptide was found in common. The urine list was the limiting factor. We picked the common peptides between plasma and CSF and run a target analysis on the urine sample. We succeeded to select five peptides with proteotypic and quantotypic criteria, having between 15-20 amino acids in length and which are detectable in the three matrices. Two of them containing one methionine or one asparagine residue to assess the methionine oxidation and the asparagine deamidation.

The structures of the chosen peptides were theoretically predicted using online available resources (namely the SOPMA [88] and the CFSSP [89] prediction servers) and mutually correlated. The purpose was to replace an intrinsic peptide in the two model proteins by a peptide with the same secondary structure to avoid structural change in the final chimeric protein. As far as the disordered protein is concerned, inserted peptides must be characterized by non-ordered structures region to keep this part of the chimeric protein unstructured.

A total of 11 peptides were inserted in the chimeric protein by replacing an existing peptide sequence of the structured domain and by adding a non-ordered structure peptide in the unstructured region. There was a balance between the number of peptides to add in each region to not disturb too much the structure but also to have a large panel of peptides available

as standards and not having a protein with a mass higher than 100 kDa. Adding six exogenous and 5 endogenous peptides was a good compromise. Three exogenous peptides were replaced in the structured region and three others were added to the disordered one. The exogenous peptides will serve as standard indicators of the digestion kinetics in terms of embedded or dislodged peptides in the protein. For better release of the endogenous peptides and to ensure reliable quantification of the endogenous protein, the 5 endogenous peptides were included in the disordered region. The list of the 11 chosen peptides is shown in table 3.1. With the insertion of the standard peptides in the model proteins, double cleavage sites appeared (*i.e.* the succession of K and R, R and K, R and R or K and K). Some of them were already present in the sequence of the two model proteins and were removed by the replacement of one K or R by a histidine residue. Only one double cleavage site (K450 and K451) was kept in the disordered region to evaluate the influence of different digestion on this peptide.

Table 3.1: *List of the 11 peptides sequences inserted or replaced in the chimeric protein and their characteristics. The monoisotopic mass, the peptide length, the hydrophobic index calculated with the Sequence Specific Retention Calculator [90] choosing a 100Å, C18 column with 0.1% formic acid separation system to correlate with our system, the structure of the peptide as it was in its intrinsic protein obtain with the SOPMA online resource, the protein domain in which each peptide is included and their endogenous or exogenous nature.*

Sequence	Mono. MM (Da)	Peptide length	H.I.	Peptide structure	Protein domain	Endo or exo
GGVTGSPDASISGSK	1318.6365	15	N/A	coil	disordered	exogenous
AGVTGSPEASISGSK	1346.6678	15	18.35	coil	disordered	exogenous
VHDPTEEATPTPFGK	1624.7733	15	21.94	coil	disordered	exogenous
LELSVDGAK	930.5022	9	20.46	coil	disordered	endogenous
EMSGSPASGIPVK	1258.6227	13	21.59	coil	disordered	endogenous
VDFTLSSER	1052.5138	9	24.57	coil	disordered	endogenous
FGLLDEDGK	992.4815	9	27.62	coil	disordered	endogenous
SFFPENWLWR	1380.6620	10	50.58	coil	disordered	endogenous
LDSDLYLELR	1235.6398	10	41.34	α-helix	structured	exogenous
TILWIGLAGEEGSR	1500.7636	14	44.31	β-sheet	structured	exogenous
ESLLVSAWSEELISK	1689.8825	15	47.51	α-helix	structured	exogenous

3.2.2. LC-MS response of the Kit Quanta peptides

The 11 peptides belonging to the list established to be part of the kit was synthesized before pursuing the development of the kit to ensure the reliability of all peptides. Eurogentec provided 1 mg of each peptide in a lyophilised form. The first critical point was the solubilisation of the peptides. Several parameters affect their solubility such as their isoelectric point, the number of charges, the nature of their amino acids or the length of the peptides. To ensure a good solubility of the peptides and facilitate the work of the future users, the set of peptides was solubilized in the same solvent: H_2O/ACN 50/50. All the peptides were soluble in this solvent and samples of 2 mg/ml per peptide were aliquoted and stored at -80°C. The standards were first injected in direct infusion on a Q Exactive™ Hybrid Quadrupole-Orbitrap™ mass Spectrometer to check for impurities, exact mass and ionization. They were then injected in LC-MS alone, altogether and then altogether in a complex matrix to evaluate the effect of the mix and the

matrix on the retention time and on the potential ionization suppression.

The peptides were mixed together to reach 200 fmoles of each peptide in the sample for an injection volume of 9 µL. The sample was injected using LC-MS/MS with a 150 minutes elution gradient. As can be seen from figure 3.5, the 11 peptides ionize well and are evenly distributed along the gradient. With the same quantity injected for each peptide, ionization efficiency is clearly different from one peptide to another as their intensity greatly varies. Ionization efficiency is dependent on multiple factors such as the sample matrix, the solvent and the physicochemical properties of the analyte. In our case, SFFPENWLWR, LDSDLYLELR, TILWIGLAGEEGSR and ESLLVSAWSEELISK are the peptides with less response in MS.

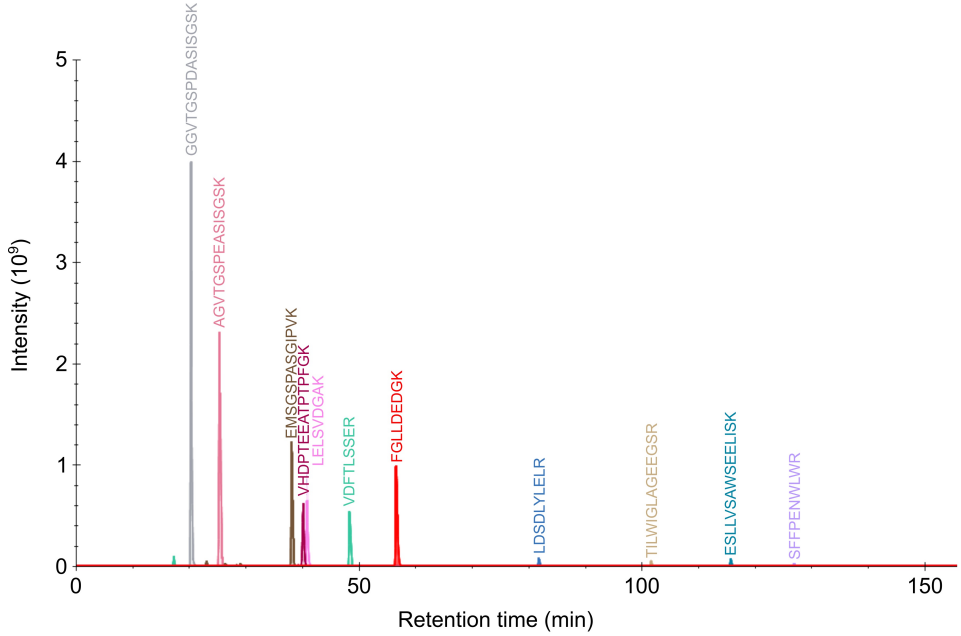


Figure 3.5: *Chromatogram of the 11 kit quanta peptides analysed using LC-MS/MS on a Q Exactive™ Hybrid Quadrupole-Orbitrap™ mass spectrometer equipped with a nanoLC MClass system using a 150 minutes elution gradient. Peptides were injected at 200 fmoles in 9 µL each.*

The peptides with FS were also synthesised and analysed in their current form by LC-MS. Even though this form of the peptide will not be present as these peptides are introduced before the enzymatic digestion in the sample preparation process, they should be detectable with their FS to evaluate their digestion efficacy. Figure 3.6 resumes the peptides with and without their FS injected altogether with or without matrix. In blue, the peptides are mixed altogether and injected at a concentration of 22.2 nM each (injection of 9 µL = 200 fmoles) and in red, the mix of peptides is spiked into a plasma digest injected at a concentration of 0.075 µg/µL (injection of 9 µL with 0.675 µg of plasma and 200 fmoles of each peptide). Peptides with FS as well as peptides without FS are detectable and we observe that in the presence of a matrix peptide intensity is higher for the majority of peptides probably due to the lower propensity of peptides adsorption on the vials due to the coating of the matrix. Signal suppression linked to a matrix

effect is not encountered with these peptides.

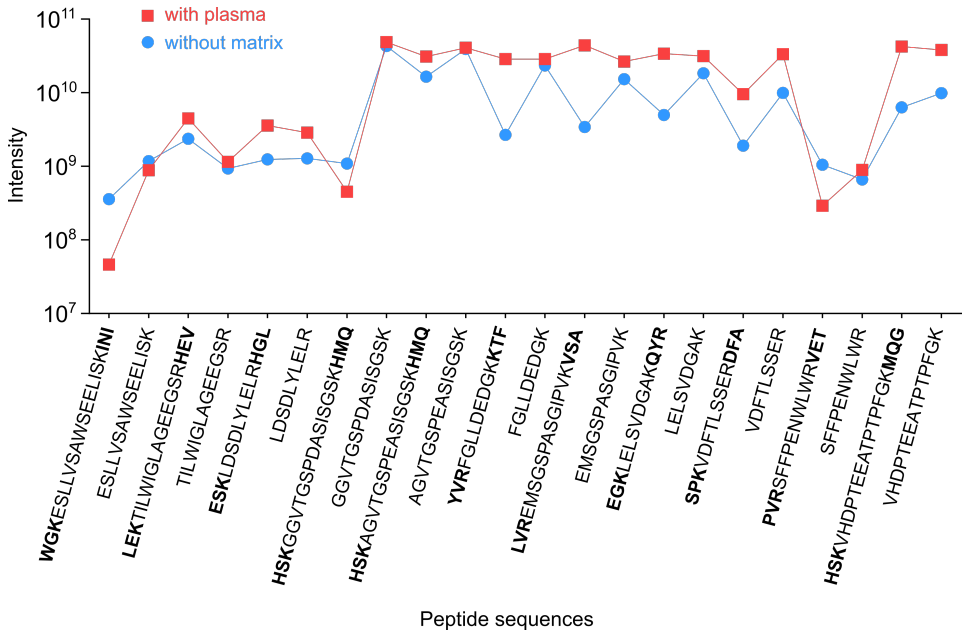


Figure 3.6: *Integrated area of the 11 kit quanta peptides with and without their FS. The peptides are sorted as they are in the sequence of the chimeric protein. These peptides were injected altogether without matrix (blue dots) or with plasma (red square).*

In conclusion, the whole set of peptides (with and without FS) is easily detectable without matrix and even more when spiked in a plasma digest. Furthermore, they span the whole gradient range of the chromatographic separation and are thus a good set of standards for the purpose of standardization quantification.

3.2.3. Purity of isotope-labelled internal standard

Production of the fully ^{15}N isotopically labelled protein was obtained by expressing the protein in a ^{15}N media nitrogen source whereas the labelled peptides where chemically synthesized using ^{13}C and ^{15}N labelled amino acid. The ^{15}N chimeric protein has an average mass of 71,008.5202 Da and a monoisotopic mass of 70,966.5215 Da; the monoisotopic mass of its peptides can be observed in table 3.2. Several factors were taken into considerations to yield a labelling strategy. The labelling must only cover amino acids on the short peptides (not the FS) as after digestion the short version appears and must therefore be distinguishable from the other levels. In table 3.2, both the 2^{nd} level with and without FS were added because the monoisotopic mass of the peptide without FS must be taken into account for the labelling strategy but it is the entire peptide (with FS) that will be synthesized. Nevertheless, the digestion process may not be complete and peptides with FS could be detected as well but their retention time will be different from the other levels and will be distinguishable. The exogenous peptide could have a level without labelling because they do not exist in the complex matrices studied.

On the contrary, all levels of the endogenous peptides must be labelled. To avoid overlap in their isotopic distribution between two peptides, a difference in mass of 5 Da between the monoisotopic mass of the peptides is a strict minimum. Below that the overlapping can be critical by representing a certain percentage of the isotopic distribution of the other peptide. The theoretical isotopic distribution of each peptide was examined to ensure this non-overlap. The maximum overlap was obtained for the VDFTLSSER peptide between the 1^{st} level and the 3^{rd} level where 0.08% of the isotopic distribution of the 1^{st} level peptide is common to the isotopic distribution of the 3^{rd} level peptide. The majority of the overlap percentage was below 0.001% without reaching more than 0.08%. The difference in mass of 5 Da is a compulsory restriction for the same peptides with different levels of labelling. Some peptides (distinct ones) are closer in mass (for example the 1^{st} level of EMSGSPASGIPVK and the 2^{nd} level without FS of LDSDLYLELR with only 1.14 Da of difference) but they will be distinguishable by their retention time. Figure A.1 exhibits the mass spectra of two experimental sets of peptides, chosen arbitrarily.

Table 3.2: *Sequences of the 11 kit quanta peptides listed as they are ordered in the chimeric protein and their labelling. In purple, the ^{15}N isotopically labelled amnio acid (1^{st} level peptides coming from the chimeric protein) and in blue, the ^{13}C and ^{15}N labelled amino acid (2^{nd}, 3^{rd} and 4^{th} level peptides). The monoisotopic mass of each labelled peptides was also added to have a quick overview of the succession of the peptides in the mass spectrum: the 4^{th} level being the lighter one, followed by the 3^{rd} level and the 2^{nd} level. The 1^{st} level of peptides being randomly placed between levels because of its fixed mass determined by the complete ^{15}N labelling.*

Sequence	Peptide level	Mono. MM (Da)
ESLLVSAWSEELISK	1^{st} level	1710.8203
WGKESLLVSAWSEELISKINI	2^{nd} level with FS	2430.3550
ESLLVSAWSEELISK	2^{nd} level without FS	1718.9482
ESLLVSAWSEELISK	3^{rd} level	1697.8967
ESLLVSAWSEELISK	4^{th} level	1689.8825
TILWIGLAGEEGSR	1^{st} level	1516.7162
LEKTILWIGLAGEEGSRHEV	2^{nd} level with FS	2264.2538
TILWIGLAGEEGSR	2^{nd} level without FS	1528.8323
TILWIGLAGEEGSR	3^{rd} level	1510.7719
TILWIGLAGEEGSR	4^{th} level	1500.7636
LDSDLYLELR	1^{st} level	1253.5864
ESKLDSDLYLELRHGL	2^{nd} level with FS	1925.0507
LDSDLYLELR	2^{nd} level without FS	1273.7167
LDSDLYLELR	3^{rd} level	1263.7084
LDSDLYLELR	4^{th} level	1235.6398
GGVTGSPDASISGSK	1^{st} level	1335.5861
HSKGGVTGSPDASISGSKHMQ	2^{nd} level with FS	2094.0393
GGVTGSPDASISGSK	2^{nd} level without FS	1345.6955
GGVTGSPDASISGSK	3^{rd} level	1326.6507
GGVTGSPDASISGSK	4^{th} level	1318.6365
AGVTGSPEASISGSK	1^{st} level	1364.6144
HSKAGVTGSPEASISGSKHMQ	2^{nd} level with FS	2122.0706
AGVTGSPEASISGSK	2^{nd} level without FS	1373.7268
AGVTGSPEASISGSK	3^{rd} level	1354.6820
AGVTGSPEASISGSK	4^{th} level	1346.6678
FGLLDEDGK	1^{st} level	1002.4518
YVRFGLLDEDGKKTF	2^{nd} level with FS	1835.0359
FGLLDEDGK	2^{nd} level without FS	1030.5647
FGLLDEDGK	3^{rd} level	1020.5375

FGLLDEDGK	4^{th} level	1012.5233
EMSGSPASGIPVK	1^{st} level	1272.5812
LVREMSGSPASGIPVKVSA	2^{nd} level with FS	1922.0971
EMSGSPASGIPVK	2^{nd} level without FS	1284.6784
EMSGSPASGIPVK	3^{rd} level	1278.6646
EMSGSPASGIPVK	4^{th} level	1266.6369
LELSVDGAK	1^{st} level	940.4726
EGKLELSVDGAKQYR	2^{nd} level with FS	1726.9574
LELSVDGAK	2^{nd} level without FS	965.5754
LELSVDGAK	3^{rd} level	955.5545
LELSVDGAK	4^{th} level	947.5340
VDFTLSSER	1^{st} level	1064.4783
SPKVDFTLSSERDFA	2^{nd} level with FS	1720.8653
VDFTLSSER	2^{nd} level without FS	1075.5531
VDFTLSSER	3^{rd} level	1069.5393
VDFTLSSER	4^{th} level	1058.5276
SFFPENWLWR	1^{st} level	1396.6141
PVRSFFPENWLWRVET	2^{nd} level with FS	2109.1433
SFFPENWLWR	2^{nd} level without FS	1427.7623
SFFPENWLWR	3^{rd} level	1407.7079
SFFPENWLWR	4^{th} level	1387.6787
VHDPTEEATPTPFGK	1^{st} level	1637.7348
HSKVHDPTEEATPTPFGKMQG	2^{nd} level with FS	2319.1354
VHDPTEEATPTPFGK	2^{nd} level without FS	1650.8289
VHDPTEEATPTPFGK	3^{rd} level	1631.7841
VHDPTEEATPTPFGK	4^{th} level	1624.7733

Quantification of these standards must be as accurate as possible to ensure the reliability of the method. Exact knowledge of peptide quantity is required for the utilisation of the kit. To date, AAA is the method of choice to quantify peptides or proteins [91]. After synthesis, all the peptides were therefore quantified with this technique. The amino acid used for the quantification were either unlabelled either labelled amino acid and even for some a mix of both. For the peptides uniquely quantified using labelled amino acid, the quantification only takes into account the labelled amino acid therefore any contribution of unlabelled peptide will not be considered. So, for those peptides, even if the rate of isotope incorporation is not 100%, the unlabelled percentage will not be comprised in the quantification and will not introduce bias. On the contrary, peptides quantified using unlabelled amino acids or a mix of unlabelled and labelled amino acid need to be analysed to evaluate the unlabelled contribution in the quantification. Measuring the rate of isotope contribution is rather straightforward with peptides containing only one labelled amino acid by comparing the area of the labelled peptide and its labelled counterpart. In our case, the unlabelled counterpart greatly differs in mass as multiple amino acid are labelled in a peptide. The probability to detect the completely unlabelled peptide is low. This assumption is verified in each of the spectrum where no unlabelled peptide could be observed. In addition, the isotopic purity of the ^{13}C and ^{15}N labelled amino acids used for the labelled peptides synthesis is 99%, leading to a very limited bias in the quantification.

A solubilisation protocol was established between ULiège and UMons. Thanks to the AAA quantification, the exact quantity of peptide in each tube was known. Each peptide was first

solubilized at 0.5 mM in H_2O/ACN 50/50, then aliquoted by 5 µL and 100 µL before being finally stored at -80°C. For easier use of the kit, mix of peptides from same levels were also realised then aliquoted and stored at -80°C.

To verify the AAA quantification as well as the solubilisation protocol, 3^{rd} level and 4^{th} level peptides were injected together at the same concentration. These peptides sharing the same amino acid sequences and having therefore the same ionization efficiency should have the same intensity if they are injected at the same concentration. Pool of 3^{rd} level and 4^{th} level peptides were mixed together and diluted to a final concentration of 11,11 nM with H_2O/TFA 0.1 % to inject 9 µL (100 fmoles of each peptide) in LC-MS/MS. The dilution was done in triplicates. Data were processed in the Skyline software [92] (MacCoss Lab Software, University of Washington, WA, USA) and mean of the integrated areas are reported in figure 3.7. Comparing the 3^{rd} and 4^{th} level, they all have the same intensity regarding the error bars given for the three replicates of the experiments meaning that the AAA quantification seems correct and that the solubilisation and dilution were done properly for all the peptides. There is, however, an exception for the peptide AGVTGSPEASISGSK. Indeed, the error bars do not overlap between the 3^{rd} and 4^{th} levels for this peptide. Looking at table 3.3 containing the mean, the standard deviation and the coefficient of variation of each peptide for both the 3^{rd} and 4^{th} level, we observe tiny coefficient of variation for AGVTGSPEASISGSK. This explains the non-overlapping error bars although the means are relatively similar between the 3^{rd} and the 4^{th} level.

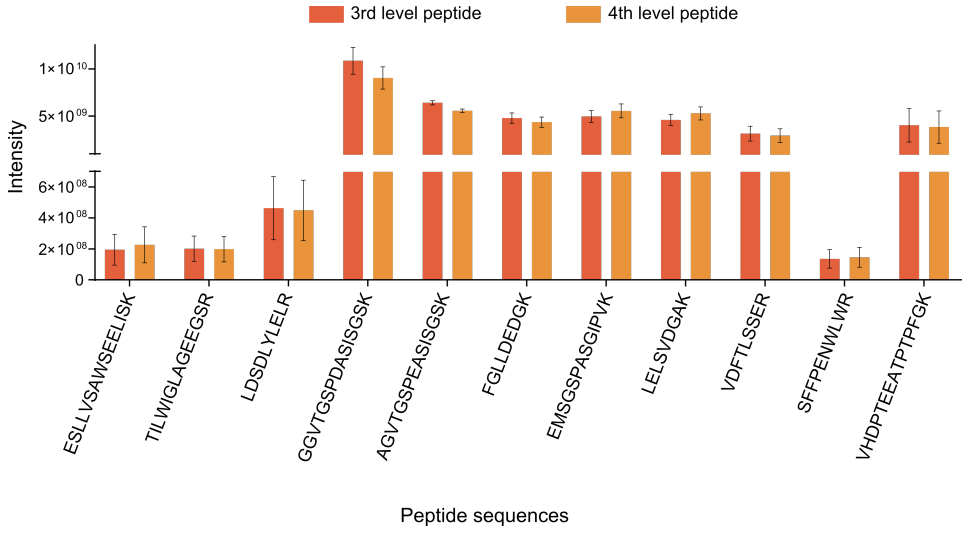


Figure 3.7: *Integrated area of the 3^{rd} and 4^{th} level peptides of Kit Quanta injected at the same concentration in triplicates. Peptides are ordered as they are in the protein.*

Table 3.3: *Mean, standard deviation (SD) and coefficient of variation (CV) of the kit quanta peptides from the 3^{rd} and 4^{th} levels*

Peptides	3^{rd} level			4^{th} level		
	Mean	SD	CV [%]	Mean	SD	CV [%]
ESLLVSAWSEELISK	$1.95 \cdot 10^8$	$9.97 \cdot 10^7$	51	$2.26 \cdot 10^8$	$1.17 \cdot 10^8$	51.5
TILWIGLAGEEGSR	$2.01 \cdot 10^8$	$8.20 \cdot 10^7$	40.9	$1.98 \cdot 10^8$	$8.20 \cdot 10^7$	41.3
LDSDLYLELR	$4.63 \cdot 10^8$	$2.04 \cdot 10^8$	43.9	$4.49 \cdot 10^8$	$1.95 \cdot 10^8$	43.4
GGVTGSPDASISGSK	$1.09 \cdot 10^{10}$	$1.43 \cdot 10^9$	13.1	$9.05 \cdot 10^9$	$1.19 \cdot 10^9$	13.1
AGVTGSPEASISGSK	$6.41 \cdot 10^9$	$2.15 \cdot 10^8$	3.4	$5.57 \cdot 10^9$	$1.88 \cdot 10^8$	3.4
FGLLDEDGK	$4.79 \cdot 10^9$	$5.41 \cdot 10^8$	11.3	$4.37 \cdot 10^9$	$5.49 \cdot 10^8$	12.6
EMSGSPASGIPVK	$4.96 \cdot 10^9$	$6.26 \cdot 10^8$	12.6	$5.55 \cdot 10^9$	$7.34 \cdot 10^8$	13.2
LELSVDGAK	$4.60 \cdot 10^9$	$6.02 \cdot 10^8$	13.1	$5.29 \cdot 10^9$	$6.89 \cdot 10^8$	13
VDFTLSSER	$3.14 \cdot 10^9$	$7.86 \cdot 10^8$	25	$2.94 \cdot 10^9$	$7.20 \cdot 10^8$	24.5
SFFPENWLWR	$1.36 \cdot 10^8$	$6.00 \cdot 10^7$	44.1	$1.46 \cdot 10^8$	$6.48 \cdot 10^7$	44.5
VHDPTEEATPTPFGK	$4.03 \cdot 10^9$	$1.77 \cdot 10^9$	44	$3.84 \cdot 10^9$	$1.71 \cdot 10^9$	44.4

The ^{15}N chimeric protein standard was carefully checked for purity using SDS-PAGE. The stain was a Coomassie brilliant blue R-250 for which the detection limit is between 8 ng and 10 ng [93]. 5 µg of the chimeric protein were spotted on the gel and only one band appeared, consequently resulting in 99.8% minimum purity. In figure 3.8, we observe a band slightly above 70kDa which is the expected mass of the protein (MM = 71,008.5202 Da).

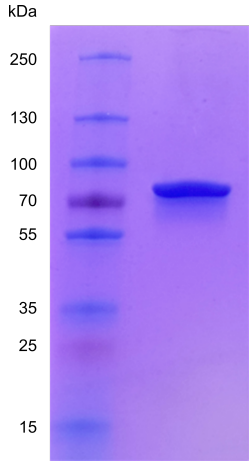


Figure 3.8: *SDS-PAGE of the ^{15}N chimeric protein.*

3.2.4. Conformation analysis of the chimeric protein

The spatial assembly of secondary structure elements may not be similar as the domains are attached together or free. Although the polypeptide backbone plays a critical role in determining the conformation of a protein, interactions can also occur between elements of the structured and disordered regions due to tertiary interactions. As a result, analysing the entire

protein is essential to fully understand its structure. Many different approaches have been used to comprehend the tridimensional structure of the chimeric protein and will be discussed in the next sections.

Theoretical approaches

Based on the SOPMA and the CFSSP online software, the chimeric protein secondary structure was theoretically predicted and can be seen in figures 3.9 and 3.10. H or h, E or e, T or t, C or c (upper cases correspond to the CFSSP predictor, lower cases to the SOPMA one) respectively correspond to α-helices, extended strands, β turns and non-ordered structures. According to the SOPMA predictor, 14.91% of the chimeric protein is in an extended strand conformational state, 30.43% in α-helices, 41.30% in non-ordered structures and 13.35% in β turns. The non-ordered structure state is essentially found in the so-called disordered part of the chimeric protein and the majority of the 8 standard peptides in this domain also contain a large proportion of non-ordered structures as we build it. Concerning the three peptides in the structured region, LDSDLYLELR and ESLLVSAWSEELISK are predicted to be in an α-helix conformational state as expected (see table 3.1. TILWIGLAGEEGSR, even though containing a one-half proportion of non-ordered structure, is categorized in β-sheet conformation. As for the CFSSP prediction, it matches less to the conformational domain we imagined: the repeated motif from the CopB protein known to have no apparent structure is not recognized as a non-ordered structure using this algorithm. Therefore reducing the proportion of non-ordered structure to 12.42%. The extended strand conformational state reaches 22.05%, the β turns 9.32% and the α-helices rises to 56.21%. Only half of the chosen coil peptides are foreseen as non-ordered structure sequences (GGVTGSPDASISGSK, AGVTGSPEASISGSK, EMSGSPASGIPVK and VHDPTEEATPTPFGK).

MQTYKSEYKMSLVLGPAFPWGK**ESLLVSAWSEELISK**INIKLYPGTSLVAGDQTREFSAIRQGVIDMAVG

hhhccccceeeeccccccttccchhhhhhhhhhttttceeeeeecccceettcccchheeeehttccceehc

STINWSPQVRELNLFSLPFLMPDYKALDALTQGEVGKSIFATLEK**TILWIGLAGEEGSR**HEVSNSKHEIR

ccccccccccccceeecceeeccchhhhhhhhttcchhhhhhhhhhtttceeeeeccccchhhhhhccccccc

HPEDLKGMKLRVVGSPLYIETFNALGANPTQMSWADAQPAMASGAVDGQENPQSVFAAAKLYTVGQKFVT

ccccchhhceeeettcccehhhhhhttcccceechhhhhhhhhhtteecttccccceeeeeeeeehhhhhhh

TWGYVADPLIFVVNKQIWESWTPADREIVKAHEESK**LDSDLYLELR**HGLAEPGAPAWKDMEAHGVKVTHL

hheeecccceeeeectthhtcccthhhhhhhhhhhhhcchhhhhhhtccccccccchhhhhhttceeeec

TPAEHDAFRHATAKVYDKWHKQIGTDLVTKAEGAIAKHGGGGSDDDDKGGGGSKMQGMDQGSMQGMDHSK

cchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhttcccccccccttcccceehhccttchecccctt

GGVTGSPDASISGSKHMQGMDQGSMQGMDQGSMQGMDHSKMQGMDQGSMQGMDHSK**AGVTGSPEASISGS**

ccccccccccccccchhhhccttthhhhccttthhhhcchhhhhtccttchhccctttcccccccccccccc

KHMQGMDQGSMQGMDHSKYVR**FGLLDEDGK**KTFMQGMDQGSMQGMDHSKLVR**EMSGSPASGIPVK**VSAMQ

chhhhccttchttccttceeeeeeecttttcheeeehccttccttcchhhhhhhhtcccccccceeeehhh

GMDQGSMQGMDHSKEGK**LELSVDGAK**QYRMQGMDQGSMQGMDHSKSPK**VDFTLSSER**DFAMQGMDQGSMQ

hccttchhhccccttcceeeeecccceeeeecccttchhhcccccccceeeeeeechhhhhhtccttchh

GMDHSKPVR**SFFPENWLWR**VETMQGMDQGSMQGMDHSK**VHDPTEEATPTPFGK**MQGMDQGSMQGMDHSKG

hccccccccccctttheeehhhhttccttchttcccccccccccccccccchhcccttthhhtcccttt

GGGSHHHHHHHHH

ccccccccccccc

Figure 3.9: *Sequence of the chimeric protein and its theoretically structure predicted using SOPMA. h, e, t, c are for α-helices, extended strands, β turns and non-ordered structures. Bold and highlighted amino acids in the sequence are the 11 standard peptides.*

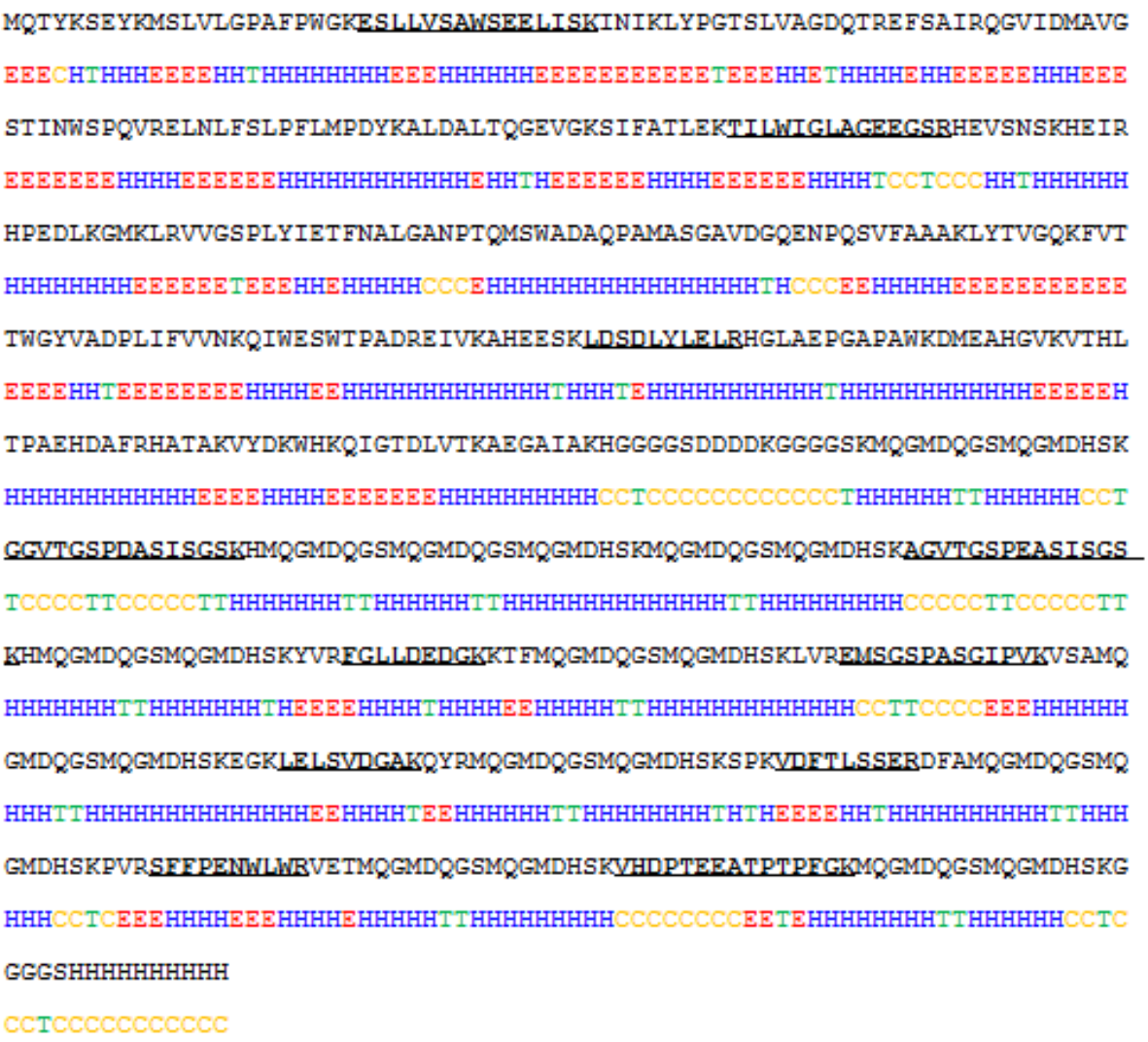


Figure 3.10: *Sequence of the chimeric protein and its theoretically structure predicted using CFSSP. H, E, T, C are for α-helices, extended strands, β turns and non-ordered structures. Bold and highlighted amino acids in the sequence are the 11 standard peptides.*

Circular dichroism

We first evaluated the secondary structure of the chimeric protein using far-UV circular dichroism (CD) spectroscopy. CD measurements can be easily used to see at a glance the presence of α-helices (at 222nm and 208nm) [94]. Even though CD cannot point out the exact localization of α-helices, β strands, turns or non-ordered structures, the content of protein secondary structure can be estimated based on different algorithms. Different solvents were tested for CD measurements, physiological as well as denatured solvents. Tris 50 mM NaCl 150 mM pH 7.5 (solvent in which the chimeric protein is dialysed after its production), NaH_2PO_4/Na_2HPO_4 20 mM pH 7.5 (typical solvent used in CD measurements) and NH_4HCO_3

50 mM pH 7.5 (tested because this solvent is the one used in the digestion protocols in our laboratory). Urea was also tested as a denaturant in different concentrations: 1 M, 3 M and 6 M.

Measurements were recorded on a Jasco J-810 spectropolarimeter using a 1mm path length quartz Suprasil cell (Hellma) and scanning from 260 to 190 nm. A concentration of 0.1 mg/ml is ideal for CD recordings but due to the limited amount of chimeric protein we had, we did not reach this concentration for all the different solvents. Data were therefore normalized with their respective concentration afterwards to allow the comparison between the solvents. The protein concentration was determined by measuring the absorbance of the samples at 280 nm. The CD parameters were set as follows: scan accumulation = 4 times, digital integration time (DIT) = 1s, bandwidth = 1 nm, data pitch = 0.1 nm, scanning speed = 50 nm/min, measure range = 260 - 190 nm, acquisition temperature = 20°C. Three different algorithms were used to calculate the secondary structure content of the protein: SELCON3 [95], CDSSTR [96] and CONTINLL [97] using the CDPro software [98]. The reference data set was chosen based on the range of the input data and the nature of the reference proteins. Our datasets going from 260 to 190 nm brought us to the choice of the SMP56 references sets (including the SP43 datasets) containing 43 soluble proteins and 13 membrane proteins. This reference data set well corresponds to the experimental parameters and the nature of the targeted protein. Other reference datasets containing denatured proteins were also tested.

Tris NaCl and urea are solvents commonly not used in CD spectroscopy as it absorbs in the far-UV region. CD spectra were recorded with these solvents and the tension signal was checked to remove any signal arriving at saturation (tension > 600 V). In the following graph, Tris NaCl curve begin at 195 nm because below the tension signal was saturated due to the absorption of the solvent. The signal saturated at 204 nm, 207 nm and 209 nm respectively for urea 1 M, 3 M and 6 M. Figure 3.11 shows the evolution of the mean residue ellipticity of the chimeric protein in different solvents in function of the wavelength. All the solvent conditions exhibit similar CD profiles characterized by very low ellipticity above 210 nm which is typically observed in disordered protein. A protein containing 100% α-helices has a $[\theta]_{MRE}$ roughly equals to -35,000 deg.cm^2.dmol^{-1}.residue^{-1} at 222 nm. Looking at 208 and 222 nm wavelength, $[\theta]_{MRE}$ goes as follows: Tris NaCl < NH_4HCO_3 < NaH_2PO_4/Na_2HPO_4 < urea 1 M < urea 3 M < urea 6 M suggesting a decrease in α-helices content as the ionic force decreases and as urea concentration increases.

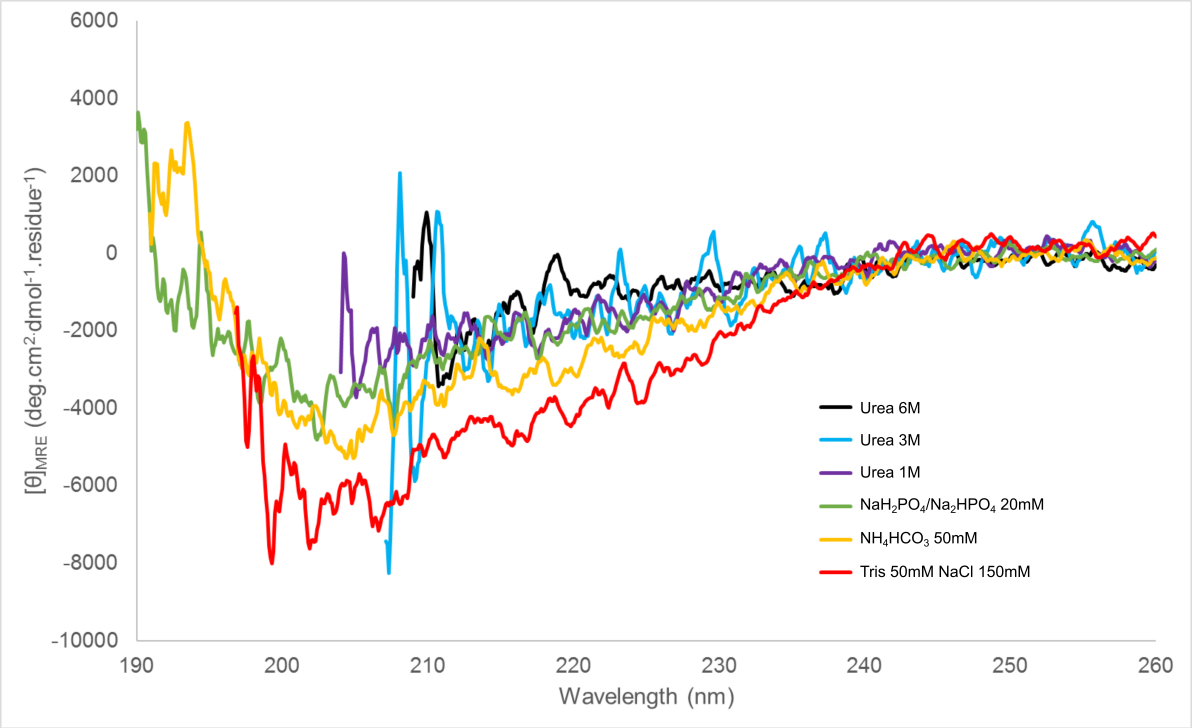


Figure 3.11: *Far-UV spectra of the chimeric protein showing the mean residue ellipticity $[\theta]_{MRE}$ as a function of the wavelength for different solvents: Tris 50 mM NaCl 150 mM, NH_4HCO_3 50 mM, NaH_2PO_4/Na_2HPO_4 20 mM, urea 1 M, urea 3 M and urea 6 M.*

These observations were compared with the calculated secondary structure content shown in table 3.4. The calculations were only possible for the physiological solvents as no reference datasets is accurate enough with narrow wavelength ranges as obtained in the denatured conditions. The high proportion of unordered residue correlates with the low ellipticity value of the chimeric protein and it matches to the unordered proportion obtained with the SOPMA theoretical approach.

Table 3.4: *Content of secondary structures calculated thanks to CDpro using SELCON3, CDSSTR and CONTINLL algorithms for three different solvents. The data presented in this table are the mean of each algorithm result.*

solvent	α-helix	β-strand	Turn	Unordered
Tris NaCl	9,47 ± 5,33 %	31,83 ± 6,36 %	22,10 ± 1,87 %	34,87 ± 1,62 %
NH_4HCO_3	17,17 ± 0,97 %	29,8 ± 7,01 %	22,80 ± 1,04 %	30,53 ± 6,96 %
NaH_2PO_4/Na_2HPO_4	3.55 ± 1.77 %	39.95 ± 0.49 %	23.80 ± 2.40 %	33.05 ± 0.35 %

For each algorithm, the secondary structure content is associated with a “goodness-of-fit” parameter, namely the normalized standard deviation (NRMSD), that reflects the correspondence between the experimental and the reference data. A NMRSD < 0.2 shows good agreement between the experimental and the reference data. In our case, NMRSD lays between 0.1 and 0.5 depending on the algorithm, indicating a not so good correspondence between the calculated and experimental spectra. This is probably because the proteins included in the chosen reference datasets did not encompassed the range of structure of the chimeric

protein. Other reference datasets did not result in better NMRSD and were not consistent with the algorithm used. The results presented in table 3.4 must therefore be taken with caution and correlated with other techniques.

Thermal denaturation

Thermal denaturation involves subjecting a protein in solution to heat, causing it to denature at a specific temperature. In the case of the chimeric protein, CD spectroscopy was used to monitor the thermal denaturation of the structured portion. The denaturation temperature is characterized by a sudden jump in the curve, which indicates the loss of a 3D structure.

For thermal denaturation, protein must be in a solvent that does not absorb at the work wavelength and that is not highly volatile. Water is the best choice as it meets these two criteria. But to be in its native conformation, a protein needs some ionic forces therefore not just water. Chloride ions strongly absorbs in the far-UV preventing the use of Tris NaCl solvent. Sodium bicarbonate would be a good solvent for further comparison with the digestion made in this solvent but rise in temperature using this solvent will decompose and is already volatile at 40°C to reject CO2 and NH3. Phosphate buffer is the solvent of choice, prepared by mixing mono- and di-sodium phosphate together to achieve the proper pH and to avoid the addition of chloride ions for the pH adjustment. For comparison with the digestion performed in NH_4HCO_3 50 mM, we decided to opt for a 20 mM phosphate buffer so that similar ionic forces are present in both solvents.

Protein concentration was 0.1 mg/ml and measurements were recorded on the Jasco J-810 spectropolarimeter. The DIT was set to 4s, the bandwith to 1 nm and the wavelength fixed to 222 nm. The solution was heated from 30°C to 97°C and then cooled back to 30°C at a ramp rate of 0.5°C/min. Figure 3.12 depicts the evolution of the ellipticity measured on the chimeric protein according to temperature. Both heating and cooling measurements show a flat line evidencing (1) a strong thermostability of the protein or (2) a lack of structure in the protein. The low ellipticity value obtained in the CD measurements with a range of wavelength already indicated that a large proportion of the protein is in an unordered conformation.

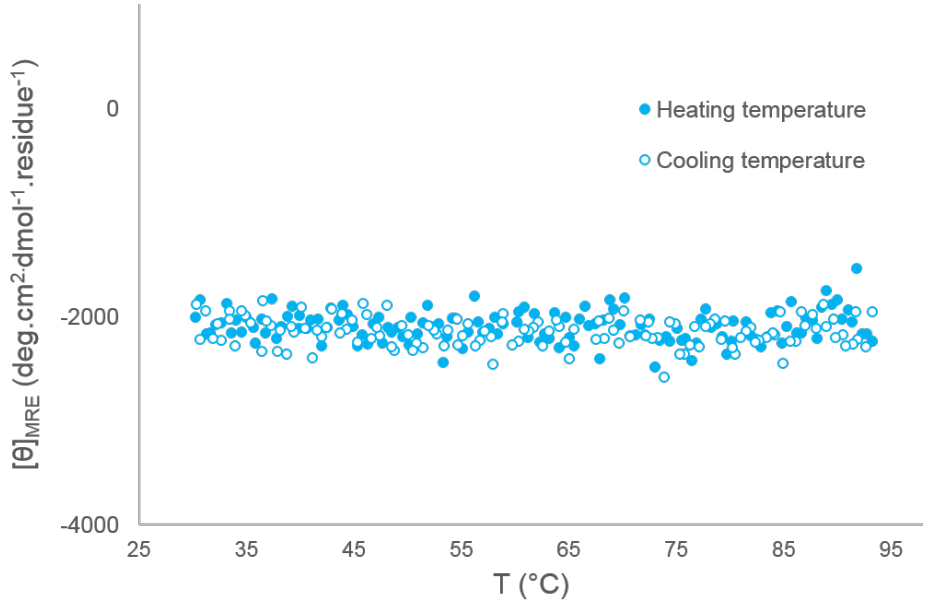


Figure 3.12: *Monitoring of $[\theta]_{MRE}$ in function of the temperature at 222 nm for the chimeric protein in solution in NaH_2PO_4/Na_2HPO_4 20 mM.*

Other techniques were then implemented to obtain concrete evidence of the structured portion of the protein.

Tryptophan fluorescence

While CD allows to evaluate the secondary structure of proteins, fluorescence spectroscopy helps to track down their tertiary structure. By applying a certain energy to a protein in solution, molecules known as fluorophores can re-emit this energy after excitation. In proteins, phenylalanine, tyrosine and tryptophan have natural fluorescent properties. Tryptophan fluorescence emission generally dominates because: 1) phenylalanine does not absorb at 280 nm and hence is merely not excited; 2) tyrosine i) has a lower extinction coefficient and ii) is quenched via energy transfer to tryptophan [99]. Tryptophan fluorescence is a good probe of folding or unfolding of a protein. Its emission spectrum changes in function of the conformational environment around it. In other words, a tryptophan buried in the core of a protein (*i.e.* native state) would have a different emission spectrum than the same tryptophan exposed to the solvent in a unfolded state.

The limitation for a protein to be studied by tryptophan fluorescence is obviously to contain tryptophan. The chimeric protein holds a total of 12 tryptophans in its design: 10 are present in the structured part and 2 in the disordered part. Tryptophan fluorescence spectra were monitored using a Jasco FP6300 spectrofluorometer. An excitation wavelength of 280

nm was applied while emission spectra were recorded between 300 and 450 nm. Protein concentration was 0.05 µg/µL in NaH_2PO_4/Na_2HPO_4 20 mM. Protein solution was brought to different temperatures to monitor the unfolding process. The emission spectra are displayed in figure 3.13. The emission profiles show a slight red shift upon thermal unfolding. Indeed, the emission maximum of the 97°C profile is shifted to the right in comparison to the emission maximum from the room temperature (RT) profile. This increase in the wavelength with increasing temperature indicates a change in the direct environment of the tryptophan residues. As a reference point, tryptophan surrounded by aqueous solvent exhibits an emission maximum around 355 nm while this value is significantly blue shifted when buried into a structure [100]. The emission peak observed at RT has a high wavelength, indicating that tryptophan residues, which were originally exposed to the solvent in the native protein state, contribute more to the signal even though there are more tryptophan residues located in the structured part. The recorded spectra represent an average of all tryptophan residues in the protein, regardless of their local environment. This emission maximum suggests that, in the native state, tryptophan residues are already solvent-exposed, and that their exposure increases at higher temperatures.

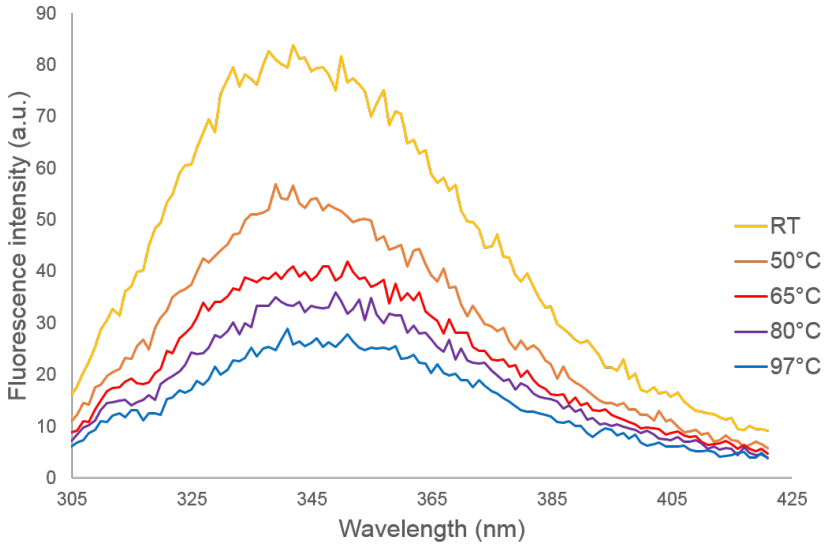


Figure 3.13: *Fluorescence emission spectra recorded at RT, 50°C, 65°C, 80°C and 97°C with an excitation wavelength of 280 nm.*

Our findings indicate that certain regions of the protein are more deeply embedded in the structure compared to others. Additionally, we observed that as the temperature is raised, these buried regions become more exposed to the solvent, suggesting a protein unfolding process. This implies that the protein has a three-dimensional structure. With the use of digestion kinetics, we aim to achieve more distinct variations in the structure of both protein domains.

Digestion kinetics

Protease digestion kinetics were done either in native and denatured state. Both kinetic experiments were done in triplicates in the exact same conditions each and fresh solution of urea was prepared for each replicate. Protein aliquots were stored at -80°C in Tris 50 mM NaCl 150 mM pH 7.5 and thawed on ice for further digestion. 15 µg of protein were dissolved in water or in a final concentration of 6 M urea respectively for the native and the denatured kinetics and incubated at RT for 1 hour. Then, right before the beginning of digestion, samples were diluted with NH_4HCO_3 50 mM to reach an urea concentration of 1 M (and to similarly dilute the protein in native state). High concentration of urea and salts such as NaCl reduces or even inhibits trypsin activity. It is therefore important to obtain a concentration of salt < 100 mM and urea concentration of maximum 1 M before tryptic digestion. Samples were digested at a protease-protein ratio of 1/1500 and digestion was stopped at ten time points with a final concentration of water/TFA 0.5 % to evaluate the release of peptide through this kinetics: 1 min, 2 min, 3 min, 5 min, 10 min, 15 min, 25 min, 40 min, 60 min and 120 min. Each digest was then purified using C18 tip with a 35 % ACN elution (with 65 % water/TFA 0.1 %) preventing the more hydrophobic peptides (the longer and miss-cleaved peptides) to further clog the LC column. They were finally diluted to a final concentration of 0.055 µg/µL with water/TFA 0.1 % (v/v) and 9 µL (0.5 µg) were injected on the ACQUITY UPLC M-Class System equipped with a C18 trap and a nanoEase HSS T3 columns already described in section 3.1.4. For the sake of comparison, a complete digestion was done in parallel in the exact same conditions as the native and denatured kinetics except that the first digestion was done overnight with a 1/50 protease/protein ratio) and the second digestion was realised in 80 % ACN with a 1/100 protease/protein ratio for three hours. Proteome Discoverer 2.1.1.21 [101] was used for database searches on the chimeric protein digests with a non-human database (yeast) containing 6,721 sequences (Uniprot, Swiss-Prot reviewed, Cerevisiae, downloaded on 28/05/2018) supplemented with the sequence of the chimeric protein. Deamidation of asparagine and glutamine and oxidation of methionine were set as variable modifications. Trypsin was selected as the enzyme with a maximum of 10 missed cleavage sites allowed. These data were processed using an in-house R script [102] developed by Raphaël La Rocca.

Results are presented in figure 3.14 in a heatmap format. The intensities of the tryptic peptides coming from the chimeric protein are represented in function of the digestion time in the native (N-xxmin) and denatured (D-xxmin) state (y-axis) and the x-axis corresponds to the amino acid residue number of the chimeric protein. The peptides released from the complete digestion are also shown on the map as well as the theoretical tryptic peptides pattern with each of the 59 R or K localisations. The horizontal line dividing the graph between residue number 318 and 334 shows the linker separating the structural part (on the left) and the disordered part (on the right) of the protein. For a sake of clarity, only one replicate is shown in figure 3.14, the other two show similar results. Figure 3.14 depicts clear differences between the two domain of the protein. In both the native and the denatured kinetics, peptides are

gradually released with the increase of the digestion time and the amount of peptides is higher in the disordered domain (right part of the heatmap) than in the structured domain (left part of the heatmap) evidencing a difference of conformation in the protein. Indeed, after a digestion time of 120 minutes, the percentage of cleavage site reached in the structured part is 19.79 $\pm$ 4.77 % and 18.75 $\pm$ 3.13 % respectively for the native and the denatured kinetics whereas the percentage of cleavage sites reached in the unstructured domain is 70.37 $\pm$ 9.80 % and 65.43 $\pm$ 4.28 % respectively for the native and the denatured kinetics (values are mean of the three replicates). After the complete digestion, these values increase to 65.63 % (structured) and 85.19 % (unstructured). It has been demonstrated that dibasic sites (KK, RR, KR or RK) are difficult to cleave [103] and Rodriguez and its co-workers [104] added cysteine and proline following a lysine or an arginine to the list of residues with the least cleavage site efficiency. However, none of these paired residues are present in the structured part of the protein. Only one KP and one KK are found in the disordered part of the protein. These observations prove that the primary structure of the chimeric protein cannot explain the digestion efficiency and that the 3D structure should therefore have an influence. This affirmation is counterbalanced by the fact that no particular discrepancy is observed between the native and the denatured kinetics (see figure 3.14). One would have expected a higher number of peptides released from the structured part in the early times of digestion, evidencing a loss of conformation of this domain and therefore an improved trypsin efficiency due to the better accessibility of the cleavage sites. However, urea does not seem to unfold the protein. The incubation time of 1h with 6 M urea may not be sufficient for the unfolding to occur. In comparison with the complete digestion where the second digestion takes place in ACN solvent, the percentage of cleavage sites reached is only 65.63 % demonstrating the difficulty of cleavage in this structured part.

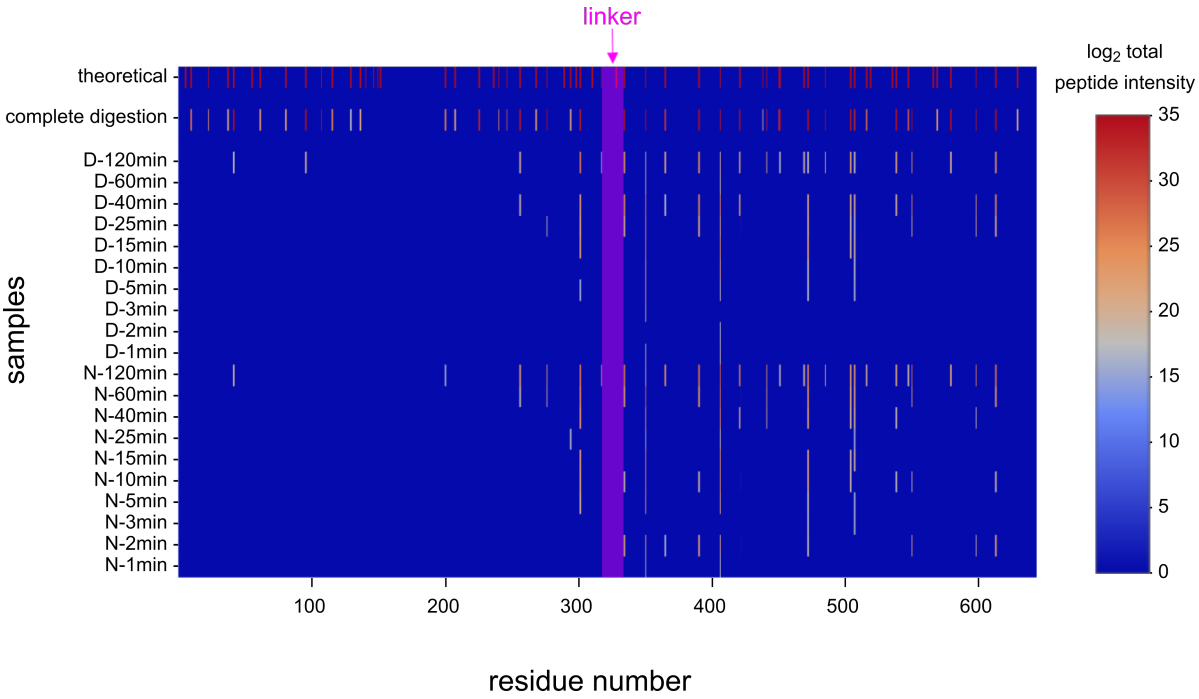


Figure 3.14: *Heatmap view of peptides identified in the chimeric protein in function of their release digestion time. The vertical axis from bottom to top indicates the digestion time in the native (N-XXmin) or the denatured (D-XXmin) state as well as the complete and theoretical digestion. The horizontal axis indicates the amino acid residue number in the order of the chimeric protein. An horizontal line divides the heatmap in two parts: the structural domain on the left and the disordered one on the right. The legend represents the peptide intensities expressed in binary logarithm.*

Conclusions

The tridimensional structure of the chimeric protein was evaluated using complementary approaches: theoretical, circular dichroism, thermal denaturation, tryptophan fluorescence and digestion kinetics followed by LC-MS analyses. The intrinsic nature of the chimeric protein being both structured and unstructured complicated the analyses. Even though there is little evidence of a structure domain, we know that peptides are more readily released from the disordered part than from the structured one thanks to the digestion kinetics.

In light of these results, trigger the release of the free protein domains to evaluate their conformation separately is recommended. The linker introduced between the two connecting functional domains is a protease-sensitive linker composed of a cleavage site receptive to a specific protease, the enterokinase, of sequence DDDDK flanked by the linker GGGGS (mentioned in section 3.1) on either side of this cleavage site to bring flexibility. Different protocols were tested to cleave the protein at this specific site but the resulting 2D gel electrophoresis showed several bands at non-expected mass. Other attempts of separation should be tested to understand more profoundly the structure of these two parts.

Taking the kit in its entirety, having a structured domain and a disordered one helps to experimentally choose a sample processing protocol that will be ideal for both configurations. To the least structured protein from the sample under study to the most structured one, any analytes could be digested in a controlled manner thanks to these two extreme tridimensional configurations.

3.2.5. Analysis of all the elements on the two instrumentations

The purpose of the Kit Quanta project was to precisely evaluate the bias introduced during the sample handling procedure before the detection by MS and to limit these bias by choosing to the greatest extent a digestion procedure and a separation system that reduce these bias. To evaluate the usefulness of the kit, the fully isotopically labelled chimeric protein standard containing the 11 standard peptides and three levels of isotopically labelled peptides are inserted at specific moments of the digestion process in a biological sample. Four different digestion procedures were tested and are reported in figure 3.15. The test 1 is the two-step in-solution digestion protocol supply by Promega (Madison, WI, USA) and consists in reduction with a final concentration of 5 mM DTT and incubation at 37°C during 30 minutes followed by alkylation with a final concentration of 15 mM of IAA and incubation at RT during 30 minutes. The enzymatic reaction is then carried out in denaturing conditions (8 M urea) using trypsin and lys-c at a 1/25 enzyme/protein ratio and incubated at 37°C during 4h. Dilution of the sample to reach a final concentration of 1 M urea is achieved and incubation at 37°C overnight follows. The digestion is stopped using a final concentration of 0.5% TFA and then purified by zip-tip (C18 tips 10µL from Pierce) with a 50 % ACN elution (with 50 % water/TFA 0.1 %). The test 2 is our in-house digestion protocol beginning with a reduction with a final concentration of 10 mM DTT and incubation at 56°C during 40 minutes followed by alkylation with a final concentration of 20 mM IAA and incubated at RT for 30 minutes. Then a second reduction is performed with a final concentration of 21 mM DTT at RT during 10 minutes. The first digestion takes place in NH_4HCO_3 50 mM with an enzyme/protein ratio of 1/50 at 37°C overnight. Then a second digestion in denaturing conditions (80% volume of ACN) is performed at 37°C during 3h. The reaction is quenched with a final concentration of 0.1% TFA and then purified by zip-tip with a 50 % ACN elution (with 50 % water/TFA 0.1 %). The test 3 protocol is the same as the test 2 with the addition of a purification step carried out using the 2D Clean-Up kit from GE Healthcare before the first digestion. And finally, the test 4 is similar as the test 2 but the two-step in-solution digestion is replaced by a single step of only 2h digestion with trypsin. All tests were carried out similarly in terms of quantity of plasma and protein digested and quantity of the different spiked peptides and were performed in triplicates. At the end of the procedure, 100 fmoles of each spiked peptide, 7.1 ng of protein (corresponding to 100 fmoles) and 0.71 µg of plasma are injected in LC-MS.

The samples were processed in triplicates and each divided in two at the end of the sample procedure to inject the same samples on both instrumentations from Liège and Mons (already described in section 3.1.4): an ACQUITY UPLC M-Class System coupled to a Q Exactive™ Hybrid Quadrupole-Orbitrap™ mass spectrometer in Liège and an Ekspert nanoLC 425 system coupled to a TripleTof 6600+ in Mons. The gradient method, instrumentations parameters and injection volume were set up according to the general procedure followed in the respective laboratories. The gradient method was fixed to 180 minutes for the first setup and 30 minutes for the second setup. The eluted peptides were mass detected according to a data-dependent acquisition method in positive electrospray ionization (ESI) mode in both setups. For same quantity injected, the injected volume was set to 9 µL in Liège and 5 µL in Mons. The Skyline software was used for data processing and manual integration of the standards peptides. Data are presented in figures 3.16 and 3.17 respectively for the samples injected on the instrumentations of Liège and Mons

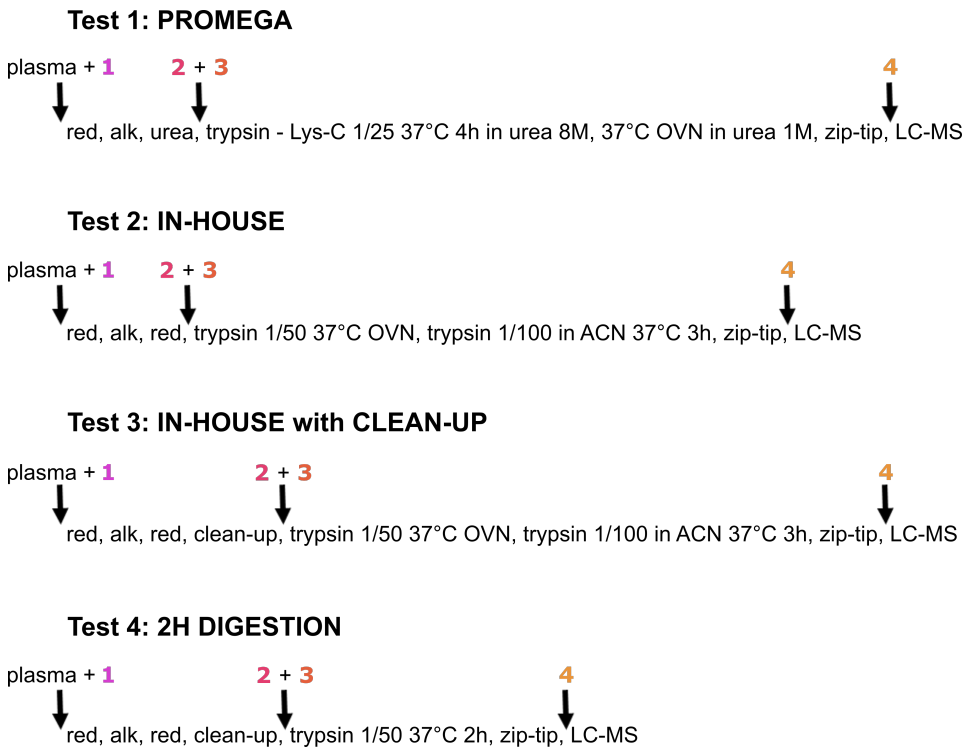
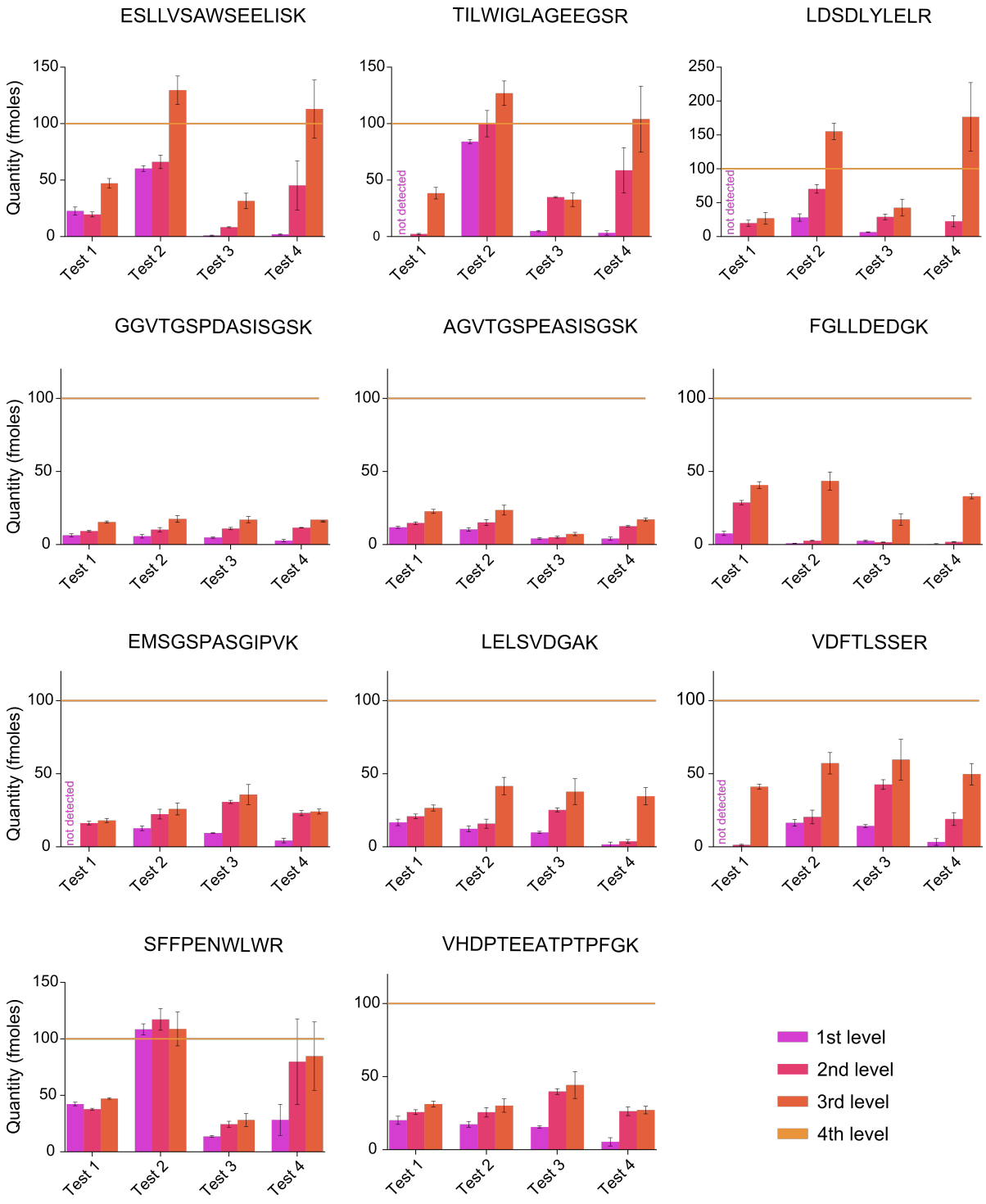


Figure 3.15: *Four digestion procedures tested with all the elements of the kit in plasma. The moment of the spiked of the standards are depicted by arrows and are similar in each protocol. Level 1 (the chimeric protein) is first spiked in plasma at the beginning of the sample process. Level 2 (peptides with FS) and 3 are inserted just before the enzymatic digestion and level 4 is spiked as the reference standard set right before the LC-MS injection.*

For all the 11 peptides, the quantity of each level is represented in fmoles in comparison to the 4^{th} level (yellow line at 100 fmoles in each graph of figure 3.16) which is the reference level spiked just before the LC-MS injection. The peptides are ordered as they are in the sequence of the chimeric protein, *i.e.* the three first graph correspond to the three peptides from the structured part of the chimeric protein and the following ones come from the disordered part.

A general and clear observation that can be picked from figure 3.16 is the stairs profiles: the quantity of peptides decreases from fourth level to 1^{st} level. This general remark was expected as the sooner the peptides are introduced in the sample preparation process, the more they can be subjected to losses during the process. Moreover, peptides embedded in the chimeric protein (level 1) will recover less than peptides with FS (level 2) and these latter from peptides without FS (level 3 and 4). For four peptides (ESLLVSAWSEELISK, TILWIGLAGEESR, LDSDLYLELR, SFFPENWLWR), the quantity of some levels exceeds the 100 fmoles of the reference level, which is not consistent with the spiked quantity. In most cases, the threshold is surpassed for the 3^{rd} level and therefore cannot be a quantification error as the quantification of level 3 and 4 was compared in figure 3.7 and that same injection quantities gave same intensities (within the error bars). The fact that these peptides have a smaller relative intensity (10^{9} order of magnitude) than the other peptides (10^{10-11} order of magnitude) (see figure 3.6) may explain their peculiar response. Another general trend pulled out from these graphs is the loss of the 1^{st} level (peptides coming from the chimeric protein inserted at the beginning of the sample preparation) more important in the 2h digestion conditions (test 4) than in any other reaction conditions. The in-house conditions with the addition of clean-up (test 3) also reduced the release of the 1^{st} level peptides in comparison to the in-house condition without clean-up (test 2) or the trypsin - lys-C digestion (test 1). These experiments showed that none of them allows the complete release of the standard peptides but also that depending on the experiments this release fluctuates and is affected by a short trypsin digestion procedure (test 4) and a purification step (clean-up from test 3). The Promega protocol therefore seems to be the best suited to recover a maximum of peptides from the chimeric protein. In general, the Promega (test 1) and the in-house without purification step (test 2) generally behave similarly for all peptides and levels with the exception of the peptide FGLLDEDGK. This peptide possesses a double cleavage site kept in the chimeric protein to evaluate the influence of different digestion reactions. With a digestion procedure only involving trypsin as the enzyme, a dibasic cleavage site, as the one encounter for this peptide, is known to be difficult to split. These four experiments confirm this statement. Indeed, the 1^{st} and 2^{nd} level of peptides, respectively the peptide within the chimeric protein and the one with FS (YVRFGLLDEDGKKTF), have a poor release when the digestion depends on trypsin. This quantity largely increases with the use of the mix of enzymes from the Promega protocol. Levels 2 and 3 should not differ between the two in-house protocols (test 2 and 3) because these peptides are introduced before the enzymatic digestion and that at this point of the protocol the following steps are exactly the same. But depending on the peptides, either they increase from test 2 to test 3 (EMSGSPASGIPVK, VHDPTEEATPTPFGK) or they decrease (ESLLVSAWSEELISK, TILWIGLAGEEGSR, LDSDLYLELR, AGVTGSPEASISGSK, FGLLDEDGK, SFFPENWLWR) or some stay constant (GGVTGSPDASISGSK, LELSVDGAK, VDFTLSSER). The zip-tip at the end of the protocol could be a potential explanation to this behaviour as it could not be 100% reproducible. GGVTGSPDASISGSK does not seem to be a good sensor of the variation obtained with the different experiments as the signal is constant throughout all the four experiments. Conclusion regarding the structured and the disordered part of the protein will be

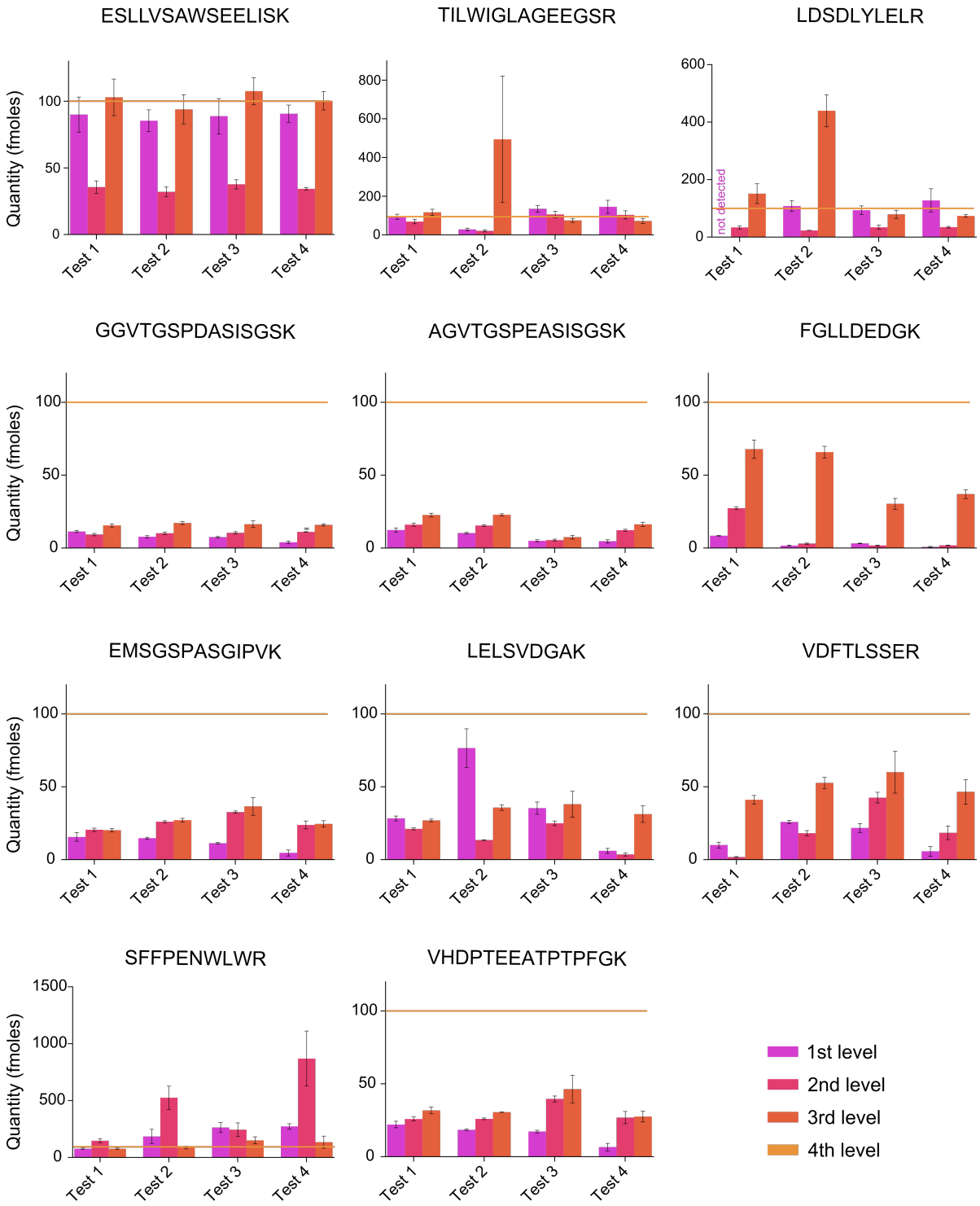
difficult to draw because the three peptides from the structured part behave quite incoherently (levels with a quantity > 100 fmoles). One would have expected a less recovery of the protein from peptides of the structured domain.



Test	Digestion conditions
1	Trypsin - Lys-C 1/25 37°C 4h in urea 8M, 37°C OVN in urea 1M
2	Trypsin 1/50 37°C OVN, trypsin 1/100 in ACN 37°C 3h
3	Clean-up, trypsin 1/50 37°C OVN, trypsin 1/100 in ACN 37°C 3h
4	Clean-up, trypsin 1/50 37°C 2h

Figure 3.16: *Proteolysis in a plasma matrix with the whole set of standards from the kit. Plasma samples were spiked with the standards at specific moments of the sample handling (see figure 3.15 for detailed spiked procedure) and were submitted to four different digestion protocols (test 1, 2, 3 and 4). The 11 standard peptides are each separately presented in function of the test and in function of the level where they come from. The fourth level of peptides was spiked at 100 fmoles and is used as the reference level to measure the amount of detected standard peptides from the other levels. Peptides are sorted as they are in the protein, the three first one being the peptides from the structured part and the eight following peptides being from the disordered part of the protein. Samples presented here were separated on an ACQUITY UPLC M-Class System and then analysed with a Q Exactive™ Hybrid Quadrupole-Orbitrap™ mass spectrometer, the Liège setup. Each bar corresponds to the mean ± the standard deviation of the triplicate.*

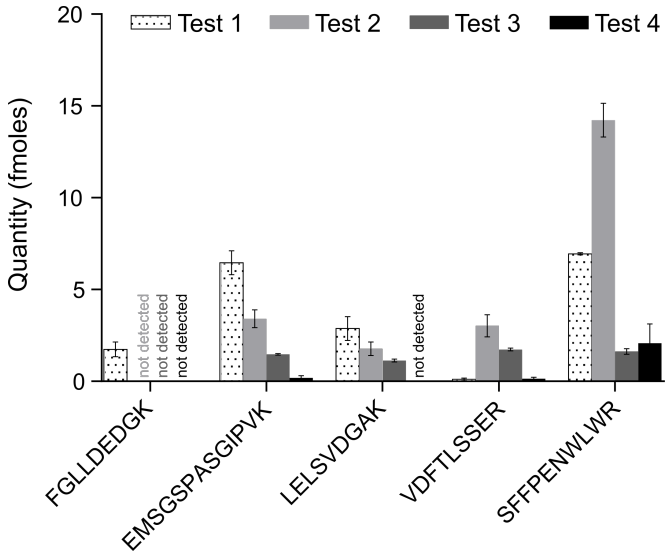
The same samples injected on the setup from Mons are depicted in figure 3.17. They share similar behaviour with the peptides injected on the setup from Liège. ESLLVSAWSEELISK, TILWIGLAGEESR, LDSDLYLELR and SFFPENWLWR peptides have incoherent profiles that can be explained by the relative intensity of these peptides. With very few exceptions, the other peptides from the four levels have identical profiles than those of Liège. The inter-laboratory injection shows that the same sample injected on different instrumentations gives consistent and similar results. The same procedure applied in both laboratories by two different experimenters should be done to evaluate the robustness of the kit.



Test	Digestion conditions
1	Trypsin - Lys-C 1/25 37°C 4h in urea 8M, 37°C OVN in urea 1M
2	Trypsin 1/50 37°C OVN, trypsin 1/100 in ACN 37°C 3h
3	Clean-up, trypsin 1/50 37°C OVN, trypsin 1/100 in ACN 37°C 3h
4	Clean-up, trypsin 1/50 37°C 2h

Figure 3.17: *Proteolysis in a plasma matrix with the whole set of standards from the kit. Plasma samples were spiked with the standards at specific moments of the sample handling (see figure 3.15 for detailed spiked procedure) and were submitted to four different digestion protocols (test 1, 2, 3 and 4). The 11 standard peptides are each separately presented in function of the test and in function of the level where they come from. The fourth level of peptides was spiked at 100 fmoles and is used as the reference level to measure the amount of detected standard peptides from the other levels. Peptides are sorted as they are in the protein, the three first one being the peptides from the structured part and the eight following peptides being from the disordered part of the protein. Samples presented here were separated on an an Ekspert nanoLC 425 System and then analysed with a TripleTof 6600+ mass spectrometer, the Mons setup. Each bar corresponds to the mean ± the standard deviation of the triplicate.*

As previously mentioned, the kit was tested in these four experiments in a biological matrix, *i.e.* plasma. The endogenous peptides from plasma could therefore be integrated and quantified using the reference level of the endogenous standard peptides. From the 11 standard peptides, 6 are artificial and 5 are endogenous to some biological matrices but are labelled to be distinguishable from the intrinsic endogenous peptides. These fives endogenous peptides are represented in figure 3.18 in function of the experiments (test 1 to 4). Globally, digestion efficiency goes as follows: test 1 > test 2 > test 3 > test 4. The peptide with a double cleavage site is not detected in any of the experiment using only trypsin as the digestion enzyme, which was expected. Those results demonstrate the ability of the kit to identify the ideal proteolysis protocols for intrinsic peptides and proves its usefulness for the evaluation of the overall sample preparation process for any biological protein biomarkers. These peptides could not be detected in the data acquired in the setup from Mons, possibly because of the short length of the gradient used.



Test	Digestion conditions
1	Trypsin - Lys-C 1/25 37°C 4h in urea 8M, 37°C OVN in urea 1M
2	Trypsin 1/50 37°C OVN, trypsin 1/100 in ACN 37°C 3h
3	Clean-up, trypsin 1/50 37°C OVN, trypsin 1/100 in ACN 37°C 3h
4	Clean-up, trypsin 1/50 37°C 2h

Figure 3.18: *Endogenous peptides from plasma detected in the four experiments detailed previously. The data were acquired on the Liège setup.*

3.3. Conclusions

The purpose of this chapter was to construct a comprehensive toolkit that addresses all possible experimental variations in the proteomic workflow, from the initial collection and storage of samples to the LC-MS analysis considering 3 major biological fluids of clinical interest (plasma, urine and CSF). The resulting kit comprises meticulously selected proteins and peptides that meet the ideal standard. We created a protein, whose structure allows, upon digestion, the asynchronous release of its peptides. This feature enables the evaluation of digestion

efficiency and the selection of the most appropriate protocol accordingly. The kit also includes a set of standard peptides that meet the requirements of proteotypicity, quantotypicity, and cover a wide range of retention times to represent all future evaluated peptides. These peptides are clearly distinguishable from each other, given the criteria of mass difference between the different levels of labelling. We also prove the accurate AAA quantification of the 3^{rd} and 4^{th} levels of peptides.

All these elements were thoughtfully designed and tested to provide the highest quality set of standard peptides and proteins to be used as a quality control for any sample processing procedure. The efficacy of the kit was demonstrated through the testing of various protocols, which showed different recovery rates of the levels of peptides, highlighting the main issue in the proteomic workflow. Based on this, one can choose the best-suited protocol, such as the PROMEGA protocol in our case, and adapt or improve it to minimize experimental bias. The capacity of the kit to be used with different instrumentations was also demonstrated by obtaining consistent results between two laboratories using distinct setups.

Although the usefulness of the two domains of the chimeric protein could not be clearly demonstrated, our protocol brings a new dimension to the control and standardization of the sample processing steps required in MS-based proteomic analysis. The combination use of exogenous and endogenous peptides makes the kit powerful to compare quantification obtained from our MS-based sample process with the one obtained from the clinical assay, thanks to the complement C4 protein.

Despite our efforts, we were not able to clearly demonstrate the impact of the two domains of the chimeric protein due to the dependency of digestion efficiency on various factors such as the protein's structure, the location of its buried and exposed parts, and the amino acid composition surrounding the kit quanta peptides. Additionally, we observed that the peptides inserted in the disordered domain displayed lower ionization efficiency compared to the rest of the set of peptides, potentially impeding their applicability.

The field of MS-based proteomics has witnessed tremendous efforts towards controlling and standardizing the sample processing steps required for accurate protein quantification. Several methods have been developed to monitor specific steps of the process individually and have been already reported in section 2.4. Lebert *et al.* [105] developed a universal standard for the control of bottom-up proteomic experiments containing a soluble recombinant protein flanking with a set of 11 artificial peptides. These peptides have been carefully chosen such that their flanking amino acids promote or inhibit the digestion efficiency. This standard allows to assess the quality of the sample preparation process in its entirety and can evaluate the efficacy of a digestion protocol and the performance of different laboratories, instruments, and software tools. In this context, our toolkit brings a novel approach to this field. By enabling precise tracking of specific set of peptides through various labelling, we can identify the loss

at each stage of the workflow. Furthermore, the use of exogenous and endogenous peptides makes our kit powerful for comparing MS-based sample process quantification with clinical assay measurement thanks to the complement C4 protein.

Future work will be dedicated to implementing the kit in a real cohort and validating it for reproducibility and comparability of results across different laboratories and experiments, including different MS platforms and operators. The kit should also be tested on different sample types, such as plasma, urine, or CSF, for which it has been designed.

The quantotypicity of all the standard peptides was not entirely maintained due to the introduction of a modified peptide in the set of peptides to evaluate the level of methionine oxidation. The assessment of this peptide and its added value in the kit are the subjects of the following chapter.

Overall, we believe that Kit Quanta holds significant promise as a universal standardization kit for evaluating and improving any proteomic workflow across laboratories worldwide. Its comprehensive design and ability to assess sample preparation quality in its entirety make it a valuable tool for the proteomics community.

4

Methionine oxidation

The methionine-containing peptide was introduced in the chimeric protein and also as a standard peptide in the kit to be able to quantitate this particular PTM. The behaviour of this peptide obliged us to extensively study this standard for absolute quantitation. As a reminder, 5 peptides coming from the “C4” protein (the endogenous protein present in plasma, urine and CSF) were inserted in the chimeric protein to allow comparison between the absolute quantification obtained with mass-spectrometry analysis and the clinical quantitation of the “C4” protein (see figure 3.1 and table 3.1 for the sake of completeness). In fact, only three of them can help for the quantitation as the other two contain modifiable amino acids. The methionine-containing peptide, the peptide under discussion in this chapter, is part of these two peptides. Taking into account the bias of this non-quantotypic peptide, either during storage, sample handling or injection, it could be used as absolute quantitative standard. Conversely, the quantitation of the protein “C4” can help us determine the concentration of this modifiable peptide and deduce its actual oxidation proportion. This chapter stems from the article “Liquid chromatography setup-dependent artefactual methionine oxidation of peptides: the importance of an adapted quality control process” published in the *Journal of Chromatography A* in July 2021 and focuses on the assessment of the oxidation of the EMSGSPASGIPVK peptide used in the kit. Several strategies to minimize the oxidation are shared throughout this chapter.

4.1. Original scientific article

Liquid chromatography setup-dependent artefactual methionine oxidation of peptides: the importance of an adapted quality control process

France Baumans[a], Emeline Hanozin[a], Dominique Baiwir[b], Corentin Decroo[c], Ruddy Wattiez[c], Edwin De Pauw[a], Gauthier Eppe[a], and Gabriel Mazzucchelli[a,*]

[a] Mass Spectrometry Laboratory, MolSys Research Unit, University of Liege, Liege 4000, Belgium
[b] GIGA Proteomics Facility, University of Liege, Liege 4000, Belgium
[c] Proteomics and Microbiology Laboratory, University of Mons, Mons 7000, Belgium
* Corresponding author. E-mail address: gabriel.mazzucchelli@uliege.be; tel: +32 4 366 34 11

4.1.1. Abstract

In both biologics quality control experiments and protein post-translational modification studies, the analytical system used is not supposed to bring any artefactual modifications which could impair the results. In this work, we investigated oxidation of methionine-containing peptides during reversed-phase (RP) chromatographic separation. We first used a synthetic methionine-containing peptide to evaluate this artefactual phenomenon and then considered more complex samples (*i.e.*, plasma and HeLa protein digests). The methionine oxidation levels of the peptides were systematically assessed and compared for the long-term use of the analytical column, the sample trapping time, the gradient length, the sample load and the nature of the stationary phase (HSS T3 from Waters, YMC Triart C18 from YMC Europe GmbH and BEH130 C18 from Waters). In addition to the oxidation of methionine in solution, we observed on the HSS T3 and the BEH130 stationary phases an additional broad peak corresponding to an on-column oxidised species. Considering the HSS T3 phase, our results highlight that the on-column oxidation level significantly increases with the age of the analytical column and the gradient length and reaches 56 % when a 1-year-old column set is used with a 180 min-long LC method. These levels go to 0 % and 18 % for the YMC Triart C18 and the BEH130 C18 phases respectively. Interestingly, the on-column oxidation proportion decreases as the injected sample load increases suggesting the presence of a discrete number of oxidation sites within the stationary phase of the analytical column. Those findings observed in different laboratories using distinct set of columns, albeit to varying degrees, strengthen the need for a standard of methionine-containing peptide that could be used as a quality control to appraise the status of the liquid chromatographic columns.

4.1.2. Introduction

The production of recombinant proteins for use as therapeutic agents and vaccine candidates has become a major market and is expected to continue its rapid expansion in the years to come [106, 107]. In this context, considerable efforts are made on the quality control of the final product to carefully evaluate the properties of the proteins, including their structural integrity and their stability. These crucial aspects are of major concerns as they directly impact the efficacy and the safety of the therapeutic agents.

Liquid chromatography (LC) coupled with mass spectrometry (MS) is today the gold standard of analytical tools to confirm the protein primary structure, and to identify and localize PTMs [108]. To perform an exhaustive characterization and cover the whole protein sequence, different proteomics methods are available, ranging from top-down [109], middle-down [110] and bottom-up [111] strategies to combinatorial workflows [112]. As such, we recently developed a Multi-Enzymatic Limited Digestion (MELD) procedure that merges bottom-up and middle-down ideologies, resulting in the generation of overlapping peptides that enables refined and reliable characterization of proteins [113].

Nevertheless, a comprehensive protein characterization analysis relies on the adequate combination of both sample preparation and analytical methods to produce and detect signals free of artefactual modifications. Among the common protein modifications, methionine oxidation is frequently observed either as PTM directly generated in the cell (*in vivo* modification) or as an artefactual modification induced by the sample preparation process (*in vitro* modification). Methionine can be oxidised to methionine sulfoxide (MetO) and, albeit less frequently observed, to methionine sulfone ($MetO_2$). Other amino acids can undergo *in vitro* oxidation but are less common.

Although *in vivo* modifications of methionine are relevant for biological purposes, artefactual *in vitro* modifications should be carefully minimized and evaluated in order to avoid misinterpretation of the results. Indeed, protein storage and purification, reduction and alkylation of cysteines, proteolysis, liquid chromatography separation as well as the ionization process [114–116] are all steps that can introduce undesired artefactual methionine oxidation and therefore impair the quality of the analytical results. This last decade, several new methodologies have been developed to reliably quantify or minimize methionine oxidation during the sample preparation process. Stable-isotope labelled reporter peptide and antioxidants that protect methionine residues from oxidation were used to monitor the artefactual modifications emerging during desalting procedures and the LC-MS separation [117]. Liu *et al.* [118] uses stable-isotope labelling to generate completely oxidised methionine residues in proteins before the sample preparation. Proteins with fully oxidised methionine residues are composed of a mix of ^{16}O and ^{18}O atoms, the ^{16}O atoms originally being associated with the *in vivo* oxidation level. Shipman *et al.* [119] enhanced this latter work by comparing theoretical and experimental MS spectra, eliminating the need for control samples and therefore reducing by half the protein consumption and analysis time. Rougemont [120] described the application of a triple calibration with differently labelled standards to control the oxidation of the targeted peptide as well as the oxidation of the labelled standards. While several articles refer to the control of methionine oxidation on targeted peptides during sample preparation [116–127], only few of them take the LC separation in consideration as a potential source of oxidation [117, 121] and none of these studies systematically evaluate the potential oxidation induced by the LC system.

Here, we focused on the methionine oxidation arising on our specific LC system (see materials and methods for more details). We first evidenced a significant oxidation phenomenon on methionine-containing peptides during the separation process and measured the evolution of this oxidation over a 1-year use of the column set. We then evaluated and compared the oxidation levels of methionine-containing peptides with different trapping times and gradient lengths which allows us to frame the exact localization of the LC separation bias. In addition, these results were also observed on publicly available data (PRIDE) where another RP column was used, indicating that this phenomenon is not limited to our setup. Subsequently, we showed that sample loads affect the oxidation response. Finally, we provide recommendations to evaluate, monitor and control this system-dependent artefact.

4.1.3. Materials and methods

Chemicals

HeLa whole-cell extract was purchased from Antibodies-online (Aachen, Cologne, Germany). Standard of HeLa protein digest, MS grade trypsin protease >95% and C18 tips 10µL were obtained from Pierce (Thermo Fisher Scientific, Pittsburgh, PA, USA). Extra pure grade trifluoroacetic acid (TFA) 99.5% was purchased from Acros Organics (Thermo Fisher Scientific, Pittsburgh, PA, USA). Plasma sample from Belgian blood donors was obtained from the Belgian Red Cross. The reference methionine-containing peptide was synthesized by Kaneka Eurogentec (Seraing, Liège, Belgium). Acetonitrile (ACN) was obtained from Biosolve (”ULC-MS´´ grade) (Dieuze, Moselle, France). Water was obtained from a Milli-Q purification system from Millipore (Burlington, MA, USA). Ammonium bicarbonate (NH_4HCO_3) >99.5% and iodoacetamide (IAA) >99% were supplied by Sigma-Aldrich (St. Louis, MO, USA). Reducing agent and detergent compatible protein assay (RC-DC Protein Assay) was purchased from BIO-RAD (Hercules, CA, USA). 2D Clean-Up Kit was supplied by GE Healthcare (Chicago, IL, USA). Ultra-pure grade >99.5% dithiothreitol (DTT) was obtained from Affymetrix (Thermo Fisher Scientific, Pittsburgh, PA, USA).

Synthetic peptide EM(M_{ox})SGSPASGIPVK solution preparation

Stock solution of the synthetic methionine-containing peptide was prepared by resuspending the lyophilized powder in water/ACN 1:1 (v/v) to a final concentration of 2 mg/mL. The stock solution was separated into 10 µL aliquots and stored at -80°C. Before each LC-MS analysis, the aliquots were freshly thawed on ice and diluted to a final concentration of 55.5 nM with water/TFA 0.1 % (v/v). For each experiment, 9µL (500 fmoles, 0.63 ng) were injected on the column.

Tryptic digestion of human plasma samples

Lyophilized plasma sample was resuspended with 1mL of Milli-Q water and the total protein content was evaluated based on the RC DC protein assay kit from Bio-Rad. 20 µg of plasma sample were diluted to 1 mg/ml with 50 mM NH_4HCO_3 and further processed into peptides following a standardized procedure (reduction-alkylation, clean-up, overnight tryptic digestion, C18 tip purification). The purified sample was divided in 0.825 µg protein aliquots and stored at -80°C. Before each LC-MS analysis, the aliquots were freshly thawed on ice, diluted to a final concentration of 0.075 µg/µL with water/TFA 0.1 % (v/v) and 9µL (0.675 µg) were injected on the column.

Tryptic digestion of HeLa whole cell extract

The same digestion protocol was applied to the HeLa whole protein extract. Aliquots of in-house HeLa digests were stored at -80°C. Before each LC-MS analysis, the aliquots were freshly thawed on ice and diluted with water/TFA 0.1 % (v/v) to 0.056 µg/µl and 9 µL (0.5µg) were injected on the column.

Commercial standard of HeLa protein digest preparation

Commercial HeLa protein digests were used as standards and compared with our in-house HeLa protein digests. Before each LC-MS analysis, the aliquots were freshly thawed on ice and diluted with water/TFA 0.1 % (v/v) to 0.056 µg/µl and 9 µL (0.5µg) were injected on the column.

Liquid chromatography – mass spectrometry analysis

Samples were analysed on an ACQUITY UPLC M-Class System (Waters, Milford, MA, USA) hyphenated to a Q Exactive™ Hybrid Quadrupole-Orbitrap™ Mass Spectrometer equipped with a Nanospray Flex™ ion source (Thermo Fisher Scientific, Pittsburgh, PA, USA). The UPLC system is composed of two successive columns: a nanoEase M/Z Symmetry C18 Trap Column (100 Å, 5 µm, 180 µm × 20 mm) and a nanoEase M/Z HSS T3 Column (100 Å, 1.8 µm, 75 µm × 250 mm), both commercialized by Waters. The temperature of the LC column was set to 40°C (trap column is at room temperature). Two sets of trap and analytical columns were tested. The first set is composed of brand-new trap and analytical columns. The columns were first conditioned with two injections of plasma sample and a quality control of the system was performed with a commercial HeLa protein digest standard to validate its performance

thanks to a threshold in the number of proteins detected. The second set is composed of 1-year old trap and analytical columns used for proteomics analysis. The sample is first loaded in a 10 µL sample loop, sent to the trap column and desalted with an eluent composition of water/ACN 98:2 (v/v) with 0.1 % formic acid (FA) at a flow rate of 20 µL/min and subsequently eluted thanks to an ACN gradient at a flow rate of 0.6µL/min. The trapping time was fixed to either 3 min or 15 min, while the length of the gradient elution was fixed to either 30 min or 150 min (appendix, table S1 and S2). Three combinations of trapping-gradient times were used: (1) a 60 min-long UPLC method made of 3 min trapping time, 30 min gradient elution and 27 min washing and re-equilibration (2) a 72 min-long UPLC method made of 15 min trapping time, 30 min gradient elution and 27 min washing and re-equilibration, and (3) a 180 min-long UPLC method made of 3 min trapping time, 150 min gradient elution and 27 min washing and re-equilibration. The eluted peptides were mass detected according to a data-dependent acquisition (DDA) method in positive electrospray ionization (ESI) mode. The mass spectrometer settings were set as follows: capillary voltage and temperature are 2.30 kV and 270°C respectively, S-lens RF level = 50.0, scan range was set from 400 to 1600 m/z. The MS scans were acquired with resolution = 70 000 (at 200 m/z), automatic gain control (AGC) target = 1×10^6, maximum injection time (IT) = 50 ms. The DDA acquisition was Top12 with resolution of 17 500 (at 200 m/z), AGC target = 1×10^5, maximum IT = 50 ms, isolation window = 2 m/z, a normalized collision energy NCE = 28 and a lock mass at 445.12003 m/z.

Analyses generated with two other instrumentations were investigated for comparison with the data created with our LC-MS setup: an Ekspert nanoLC 425 hyphenated to a TripleTOF 6600 system both from Sciex (Washington, D.C, USA) and a nanoACQUITY UPLC System from Waters coupled to a Q Exactive Plus™ Hybrid Quadrupole-Orbitrap™ Mass Spectrometer from Thermo Fisher Scientific (PRIDE data repository). The set of columns in use with the Sciex system is as follows: a YMC Triart C18 trap column (120 Å, 3 µm, 0.3 mm x 5mm) followed by a YMC Triart C18 column (120 Å, 3 µm, 0.3 mm x 150 mm) commercialized by YMC Europe GmbH (Dinslaken, Düsseldorf, Germany). The set employed with the Waters/Thermo system is a nanoEase M/Z Symmetry C18 trap column (100 Å, 5 µm, 180 µm x 20 mm) followed by a nanoEase M/Z BEH130 C18 column (130 Å, 1.7 µm, 75 µm x 250 mm) both from Waters.

Database search

PEAKS Studio software v.10.5 (Bioinformatics Solutions Inc., Waterloo, ON, Canada) was used for database searches on the HeLa protein digests and the plasma samples both with a human database containing 20 365 sequences (Uniprot, Swiss-Prot reviewed, Homo Sapiens 9606, downloaded on 22/10/2019). The tolerance on the precursor and fragment masses were fixed to 5.0 ppm and 0.015 Da respectively. Carbamidomethylation of cysteines was set as a fixed modification, whereas deamidation of asparagine and glutamine and oxidation of

methionine were set as variable modifications. A maximum of 3 variable PTMs per peptide were allowed. Trypsin was selected according to a specific digest mode with a maximum of 2 missed cleavages allowed. A false discovery rate (FDR) < 0.1 % was applied to MS/MS peptide identification.

Due to the complexity of the chromatographic profiles of methionine-containing peptides, the data were reprocessed in the Skyline software v.20.1.0.76 [128] (MacCoss Lab Software, University of Washington, WA, USA) following their identification via the PEAKS Studio database search. This step allows to manually select and integrate multi-peaks chromatographic profiles associated with a given methionine-containing identified peptide. Peptides identified in the 180 min-long UPLC run of both HeLa protein digest and plasma samples were selected to generate two starting lists of peptides (generated from the Peaks analysis, one from the HeLa protein digest sample and the other one from the plasma sample) as this method showed the highest number of identified proteins/peptides. These lists were then filtered to exclusively keep peptides that are characterized by a single methionine residue and imported with their respective PEAKS results. Out of all the manually integrated peptides in the 180 min-long UPLC run, the ten most intense peptides were considered for integration in the other elution time conditions. The peak areas associated with the synthetic peptide EM(M_{ox})SGSPASGIPVK (peptide injected alone) were only integrated using the Skyline software.

4.1.4. Results and discussion

Oxidation of synthetic peptide

Based on a 13 residue-long synthetic peptide, EMSGSPASGIPVK, we first evaluated the level of methionine oxidation in solution $M_{ox,sol}$ by performing a direct infusion of this standard in positive ionization mode on a Q Exactive (Thermo Fisher Scientific, Pittsburgh, PA, USA) mass spectrometer (appendix, figure A.2). Based on the m/z intensity ratio of the oxidised (+15.9949 m/z) to the native forms of the standard, we evaluated the oxidation level to about 2 % of the sum of the two forms. The behaviour of the standard peptide when subjected to different LC-MS analysis setups was then probed. Otherwise stated, results are reported and discussed for the HSS T3 column. Figure 4.1 shows the retention time (RT) profiles of the synthetic peptide on both a brand-new set of trap and analytical columns (figure 4.1a) and a 1-year old columns set (figure 4.1c). Figure 4.1a essentially presents a single chromatographic peak (RT = 17.8 min) corresponding to the native form of the methionine-containing peptide M_{native}. Two smaller contributions at M_{native}+15.9949 m/z are observed (relative abundance < 1 %) with RRT = 0.93 and RRT = 1 (RRT = relative retention time with the native form taken as reference). The latter species (figure 4.1b) is characterized by the exact same elution time as M_{native} and corresponds to in-source oxidation $M_{ox,ESI}$ occurring during the electrospray process [114]. The second contribution corresponds to $M_{ox,sol}$. As expected, on RP columns,

the more hydrophilic side chain of the oxidised methionine makes this modified peptide less retained compared to M_{native}. Similar experiments performed on an old columns set (figure 4.1c) highlight an additional broad chromatographic contribution (RRT = 0.96) lying in between the retention times of $M_{ox,sol}$ and M_{native} and corresponding to an increment of 15.9949 m/z. To identify the origin of this contribution, we postulated that the physical location of the oxidation process may differ from both $M_{ox,sol}$ and $M_{ox,ESI}$. First, the retention time lies between RT($M_{ox,sol}$) and RT($M_{ox,ESI}$) and suggests that the oxidation occurs after the injection of the sample and before the ESI process, therefore into the LC system. Second, the chromatographic peak is broader (elution time difference ΔRT = 50 sec) than the peaks associated with $M_{ox,sol}$ and M_{native} (ΔRT = 15-25 sec) which is compatible with a modification arising at the stage of the chromatographic analysis. Altogether, both observations are consistent with an on-column oxidation of methionine residues $M_{ox,LC}$ taking place along the chromatographic separation. To further evidence this phenomenon, we exploited the oxidised synthetic counterpart (EM_{ox}SGSPASGIPVK) of our standard peptide by spiking it with the unoxidised synthetic peptide at a 1:2 (w:w) ratio before the UPLC separation (figure 4.1d). We observed a 25-fold increase in the intensity of $M_{ox,sol}$, while the intensity of the broad chromatographic peak is barely affected by the addition of the oxidised standard. This suggests that, as EM_{ox}SGSPASGIPVK is oxidised before entering the LC system, only little extra methionine oxidation can happen within the LC system. These observations support our initial hypothesis stating that the additional artefactual oxidation of methionine residue is located within the LC system.

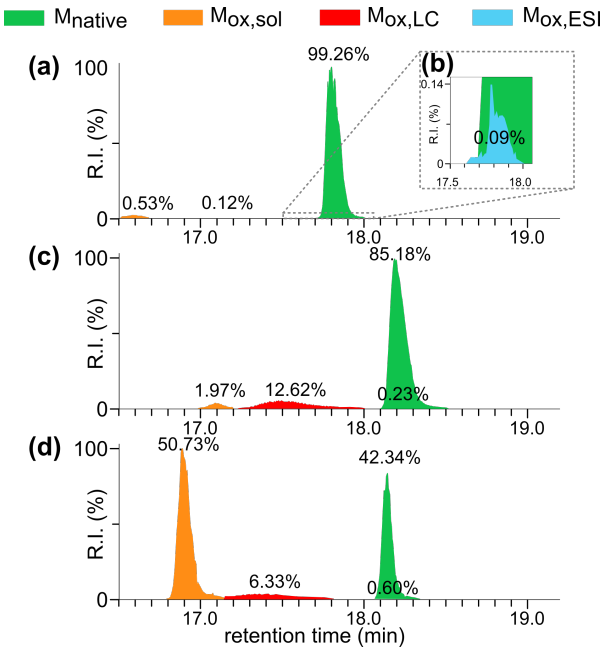


Figure 4.1: *Extracted ion chromatograms corresponding to the methionine-containing peptides EMSGSPASGIPVK and EM_{ox}SGSPASGIPVK analysed with a 60 min-long UPLC method. The peptide EMSGSPASGIPVK was analysed on (a) a brand-new set of trap and analytical columns and (c) a 1-year old trap and analytical columns set. (b) At the same retention time of M_{native}, small contribution of the in-source oxidised peptide $M_{ox,ESI}$ with an increment of 15.9949 m/z. (d) Separation of EMSGSPASGIPVK spiked with EM_{ox}SGSPASGIPVK on a 1-year old trap and analytical columns set.*

Oxidation of complex samples

Methionine residues present in cells are referred as oxidative stress protectors [129] being easily oxidised in the cells. Specific biological molecules, including catalases and superoxide dismutases protect methionine residues by scavenging the free radicals responsible for oxidation [127]. In plasma, the methionine sulfoxide reductases a and b are reported to inhibit oxidation phenomenon by reducing oxidised methionine back to methionine [123, 130–132]. In parallel, previous works have demonstrated that metal-containing proteins such as haemoglobin are well-known protagonists of methionine oxidation [132–135]. In biological samples, oxidative degradation also occurs via ROS such as free radicals or peroxides [129, 136, 137]. In contrast to pure peptide samples, complex matrix environments may therefore present antioxidant properties that can be exploited to potentially prevent oxidation phenomenon during both the sample preparation steps and, to a lesser extent, the chromatographic separation. Such behaviours have to be considered and assessed when evaluating the methionine oxidation level of $M_{ox,sol}$ in more complex samples.

Figure 4.2 shows the proportions of M_{native}, $M_{ox,sol}$, $M_{ox,LC}$ and $M_{ox,ESI}$ for the 10 most abundant methionine-containing peptides of a plasma sample (Figure 4.2a), an in-house HeLa protein digest (Figure 4.2b) and the synthetic methionine-containing peptide EMSGSPASGIPVK as a reference point (Figure 4.2c). The retention time profiles of the selected methionine-containing peptides are similar to the synthetic peptide presented in the previous section and display a clear broad chromatographic peak intermediate to RT($M_{ox,sol}$) and RT($M_{ox,ESI}$). We observed similar levels of $M_{ox,sol}$ and $M_{ox,LC}$ in the plasma ($M_{ox,sol}$ = 9.7 ± 5.7 %, $M_{ox,LC}$ = 55.7 ± 16.5 %) and the HeLa ($M_{ox,sol}$ = 12.8 % ± 27.1 %, $M_{ox,LC}$ = 57.9 ± 26.61 %) samples. In similar LC conditions, the synthetic peptide has a $M_{ox,LC}$ of 51.4 ± 1.3 % and a low formation of $M_{ox,sol}$ as this peptide is not subjected to digestion. Moreover, the peptides presented in figure 4.2a and 4.2b are sorted according to increasing retention time which shows no particular trend in $M_{ox,LC}$ either in the plasma or the HeLa samples. These similar levels of $M_{ox,sol}$ and $M_{ox,LC}$ in the complex samples and for the synthetic peptide unequivocally implies that no potential molecules naturally present in the complex samples inhibits the methionine oxidation phenomenon. We observed that while the so-called protector effect of a complex matrix plays a role *in vivo*, it does not provide oxidation protection *in vitro*.

Comparing the in-house HeLa digest samples with the commercial HeLa protein digest standard (appendix, figure A.3), we observed that levels of $M_{ox,sol}$ and $M_{ox,LC}$ are higher in the former samples than in the later. In particular, a significantly higher level of $M_{ox,LC}$. is observed for the in-house HeLa digest samples which may be attributed to our sample preparation protocol that involves 19 h tryptic digestion. As a matter of fact, the digestion incubation

time has been proven to have a significant impact on the methionine oxidation rate [116, 125, 138]. The commercial HeLa protein digest standard was digested with LysC and trypsin to ensure less than 10 % methionine oxidation in solution according to the manufacturer. Although the digestion time was not mentioned, we expect a shorter digestion time than in our protocol. As for the significant difference of $M_{ox,LC}$, it is not excluded that, to reach a peptide quality with less than 10 % methionine oxidation, some oxidant scavengers are used in the HeLa protein digest formulation. Scavengers which potentially protect methionine residues from important on-column oxidation. Nevertheless, the levels of $M_{ox,LC}$ drops to zero in both complex matrices when the LC separation is performed on a brand-new set of trap and analytical columns (appendix, figure A.4).

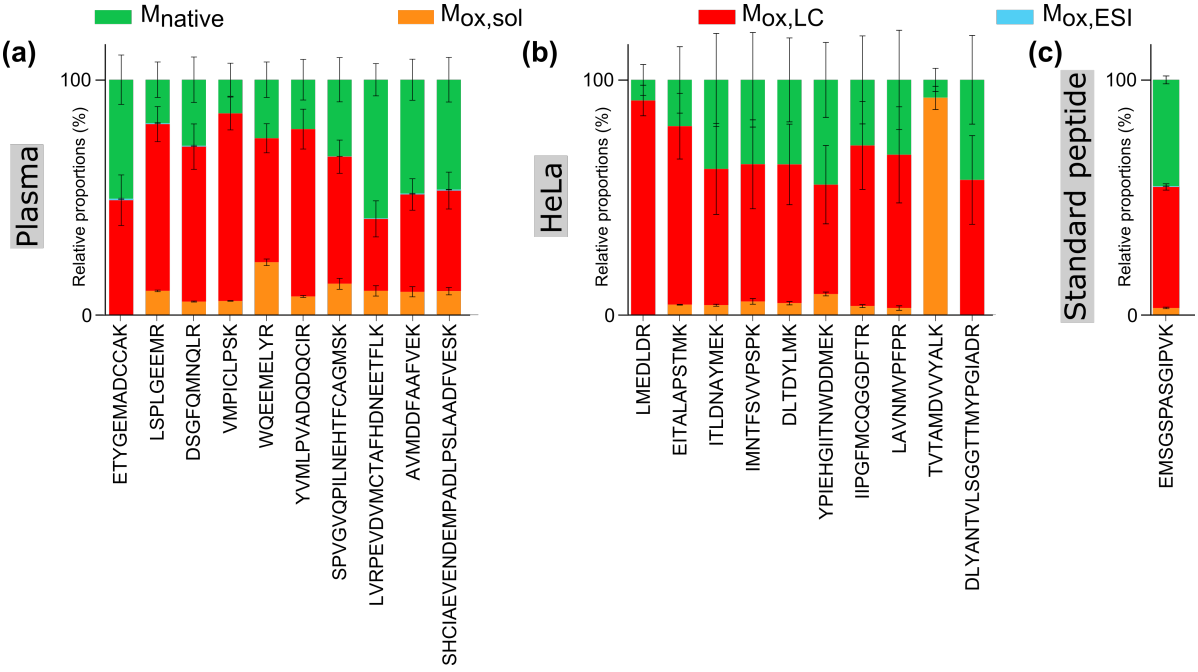


Figure 4.2: *Relative proportions of M_{native}, $M_{ox,sol}$, $M_{ox,LC}$ and $M_{ox,ESI}$ for the 10 most intense methionine-containing peptides identified in (a) plasma samples, (b) in-house HeLa protein digest and (c) the synthetic methionine-containing peptide EMSGSPASGIPVK following a 180 min-long UPLC method performed on a 1 year-old columns set with the HSS T3 phase. Contribution from $M_{ox,ESI}$ is minor (< 1 %). Peptides are sorted according to increasing retention time and error bars are estimated based on technical triplicates. Methionine-containing peptides issued from complex matrix are significantly oxidised along the chromatographic separation process (red contribution).*

How do the oxidation levels evolve with the age of the columns set?

To evidence the substantial differences on the measured oxidation levels when using either a brand-new set of columns or a 1-year old one, we studied the evolution of M_{native}, $M_{ox,sol}$, $M_{ox,LC}$ and $M_{ox,ESI}$ throughout the age of the columns set by routinely injecting a commercial HeLa protein digest standard during a whole year on the exact same set of columns. Figure

4.3 shows the evolution of the different forms of DLTDYLMK peptide with time. Additional trends are provided in appendix figure A.5. We observed that $M_{ox,sol}$ remains constant over a year of routine injection (2.8 ± 0.6 %), while $M_{ox,LC}$ increases progressively from 0 % at day-0 up to 15 % at day-365. After 10 months of injections, $M_{ox,LC}$ reaches a plateau value. These results demonstrate that the use of periodic control samples to assess the quality of the chromatographic separation should also allow the measurement of the oxidation levels in order to operate in optimal conditions. One can therefore systematically inject a QC sample which should include a methionine-containing peptide as an oxidation sensor.

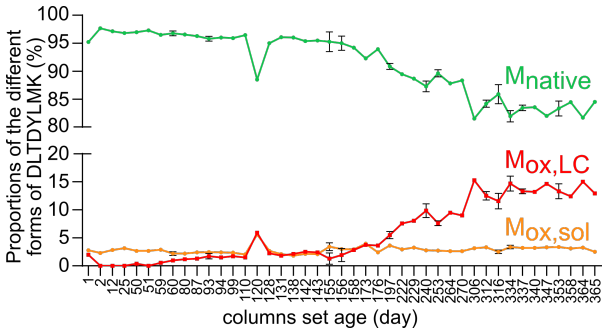


Figure 4.3: *Evolution of the percentage of M_{native}, $M_{ox,sol}$ and $M_{ox,LC}$ of a HeLa protein digest peptide (DLTDYLMK) over a 1-year UPLC use. Brand-new trap and analytical columns set (HSS T3 phase) were fitted on an ACQUITY UPLC M-Class system routinely used for proteomics analysis on day-0 and were removed on day-365. Error bars are given when several HeLa were injected the same day. $M_{ox,ESI}$ was not detected for this peptide. $M_{ox,sol}$ remains constant over the year, while $M_{ox,LC}$ progressively increases with the age of the column.*

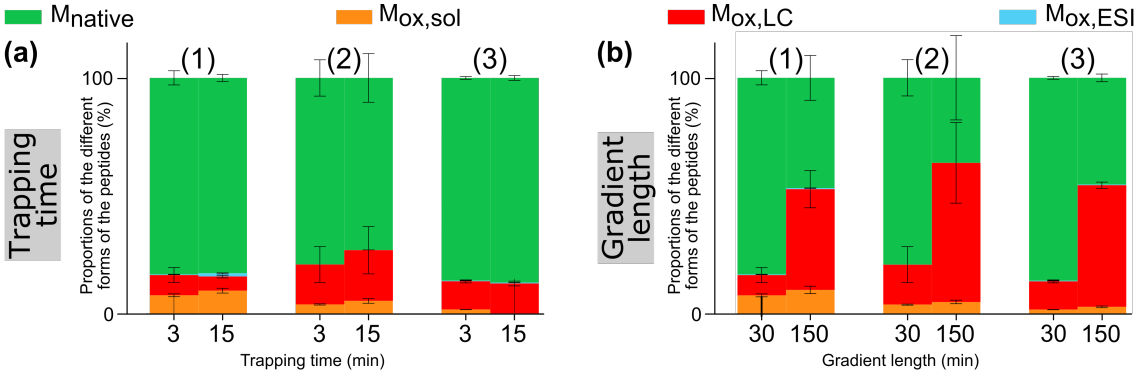
In the following sections, we focus our work on the methionine oxidation process taking place within the chromatographic system and more precisely investigate how the trapping time, the elution time, the sample load and the nature of the RP stationary phase influence the oxidation level during a standard chromatographic analysis.

Where does the oxidation take place within the chromatographic system?

Once injected to the chromatographic system, a sample successively flows along a sampling needle, an injection port, a sample loop, a trap column and an analytical column before reaching the ESI stage, each part being coupled together by various capillary connections. All these components may potentially contribute to oxidation during the chromatographic analysis of a sample. Capillary connections are made of polyetheretherketone (PEEK), an inert material known to not react with biological samples, therefore ruling out these parts as significant oxidation places. The needle, the injection port and the sample loop are all made of stainless steel, a material that has been proposed as a source of oxidation for methionine-containing peptides in the LC system [121]. Only one single injection per vial was performed with a strong needle/injection port wash (ACN with FA 0.1 %) between each injection to avoid the introduction of iron into the LC system. Although the trap column has been previously suggested as

a potential place for methionine oxidation [117], little is known on how the residence times in both the trap and the analytical columns influence the methionine oxidation level $M_{ox,LC}$ during the chromatographic separation.

Figure 4.4a and 4.4b respectively highlight the impact of the trapping time and the gradient length on the relative proportions of M_{native}, $M_{ox,sol}$ and $M_{ox,LC}$ measured on the HSS T3 1-year old columns set. Data are presented for the synthetic methionine-containing peptide (EMSGSPASGIPVK), together with two arbitrarily selected methionine-containing peptides from plasma and in-house HeLa protein digests. Additional profiles are reported in appendix figure A.6. The trapping time was varied from 3 minutes to 15 minutes with a constant 57-min long UPLC analytical run (figure 4.4a). We observed that the different oxidation levels of all the methionine-containing peptides are barely affected by an increase in the trapping time, with $M_{ox,LC}$ of about 7 ± 2 %, 19 ± 8 % and 12 ± 1 % for respectively the plasma peptide, the HeLa peptide and the standard peptide. The gradient length was varied from 30 min to 150 min with a fixed trapping time of 3 min (figure 4.4b). Longer UPLC gradients significantly affect the on-column oxidation level with a 5-fold increase in $M_{ox,LC}$ for the plasma peptide (43 ± 8 %), a 3 fold-increase for the HeLa peptide (59 ± 17 %) and up to 4-fold for the standard peptide (51 ± 1 %). Although there is a clear tendency to favour the on-column oxidation with longer UPLC gradient, we observe no significant correlation between the retention time RT and $M_{ox,LC}$ (see additional profiles sorted by RT in the appendix figure A.6). An identical procedure was applied to a brand-new set of trap and analytical columns for sake of comparison. The results are reported in appendix figure A.7. In contrast to aging columns, we observed that neither the trapping time nor the UPLC gradient length affect $M_{ox,sol}$ and $M_{ox,LC}$.



(1) plasma peptide = SHCIAEVENDEMPADLPSLAADFVESK (2) HeLa peptide = DLTDYLMK (3) standard peptide = EMSGSPASGIPVK

Figure 4.4: *Effect of (a) the trapping time (3 or 15 min on a 30 min-long UPLC gradient) and (b) the gradient length (30 or 150 min with a 3 min trapping time) on the relative proportions of M_{native}, $M_{ox,sol}$ and $M_{ox,LC}$ associated with (1) a plasma peptide (SHCIAEVENDEMPADLPSLAADFVESK), (2) an in-house HeLa protein digest peptide (DLTDYLMK), and (3) the synthetic methionine-containing peptide (EMSGSPASGIPVK). The chromatographic analysis was performed on a HSS T3 1-year old columns set. An increase in the trapping time does not induce significant changes in M_{native}, $M_{ox,sol}$ and $M_{ox,LC}$ proportions, while an increase in the gradient length drastically enhanced the on-column oxidation $M_{ox,LC}$.*

Altogether, these results highlight that the trap column is not the place of methionine oxi-

dation, while significant oxidation of methionine-containing peptides is observed on the 1-year old analytical column. Long UPLC gradients enhance the methionine oxidation level whereas short UPLC gradients minimize $M_{ox,LC}$. Such observations may be impeded to the nature of the chemical function used for the stationary phase of the column. The bonding technology of the column is called High Strength Silica (HSS) T3 and is derived from C18 ligands and end-capping. Although the exact surface structure of the column has not been released by the manufacturer, we may consider that the methionine residues react with free oxidant sites within the column and that the availability of these sites increases with the column aging. Additionally, free silanols, present in all silica columns, are the place of hydrogen-bonding interactions as well as ion-exchange interactions [139] that may play a role in methionine oxidation inside the analytical column. In the following section, we compare the HSS T3 stationary phase with other available technologies in order to better understand the origin of the on-column oxidation process.

Does the on-column oxidation occur on other RP columns?

Two additional RP columns, relying on different technologies, were tested to evaluate the exclusiveness of the artefactual oxidation of methionine to our 1-year old HSS T3 C18 column and UPLC systems. Similarly to the previous experiments performed on a 1-year old HSS T3 C18 column, we injected a plasma sample in triplicates on a 1-year old YMC Triart C18 column with a 180 min-long method. Alongside this analysis, we considered another dataset accessible on the Proteomics Identification Database (dataset PXD006882 from PRIDE archive https://www.ebi.ac.uk/pride/archive/). It consists of skeletal muscle biopsy samples injected on a BEH130 C18 column following a 175-min long method, the age of the analytical column is not reported. These data are biological replicates from three different donors and not technical replicates as the others (for more details, see [140]). Figure 4.5 displays the proportions of M_{native}, $M_{ox,sol}$, $M_{ox,LC}$ and $M_{ox,ESI}$ for the 10 most abundant methionine-containing peptides of the samples injected on a HSS T3 C18 column (figure 4.5a), a YMC Triart C18 column (figure 4.5b) and a BEH130 C18 column (figure 4.5c). We observe similar broad chromatographic peaks corresponding to $M_{ox,LC}$ for peptides separated on a HSS T3 C18 column ($M_{ox,LC}$ = 55.7 ± 16.5 %) and a BEH130 C18 column ($M_{ox,LC}$ = 18.1 ± 11.4 %), whereas no broad peak is identified on the YMC Triart column ($M_{ox,LC}$ = 0.0 ± 0.0 %). Although the age of the BEH130 C18 column is unknown, we expect that the column is not brand new in light of the $M_{ox,LC}$ levels. The three RP columns are based on trifunctionally C18 alkyl phase bonding and end-capping, the T3 bonding of the HSS T3 column being an enhancement of the trifunctional bonding characterized by a lower carbon coverage than a normal C18 trifunctional column. They however differentiate by their particle technology: ethylene bridged hybrid for the BEH column, high strength silica for the HSS T3 and silica-organic hybrid stationary phase for the YMC Triart. This similarity of bonding phase for the latter two may explain their analogous chromatographic profiles pattern. Furthermore, the YMC Triart columns are part of

a so-called biocompatible system in which all the components that may come into contact with the samples or the solvents are made of PEEK. This is different from the two other columns where, as described before for the HSS T3, samples and solvent are in contact with some stainless-steel parts. Even if the age and wear is unknown for the BEH column, the chromatographic profile anomaly is definitely present. This phenomenon observed in two different laboratories using two distinct instrumentation setups could therefore be witnessed on other setups.

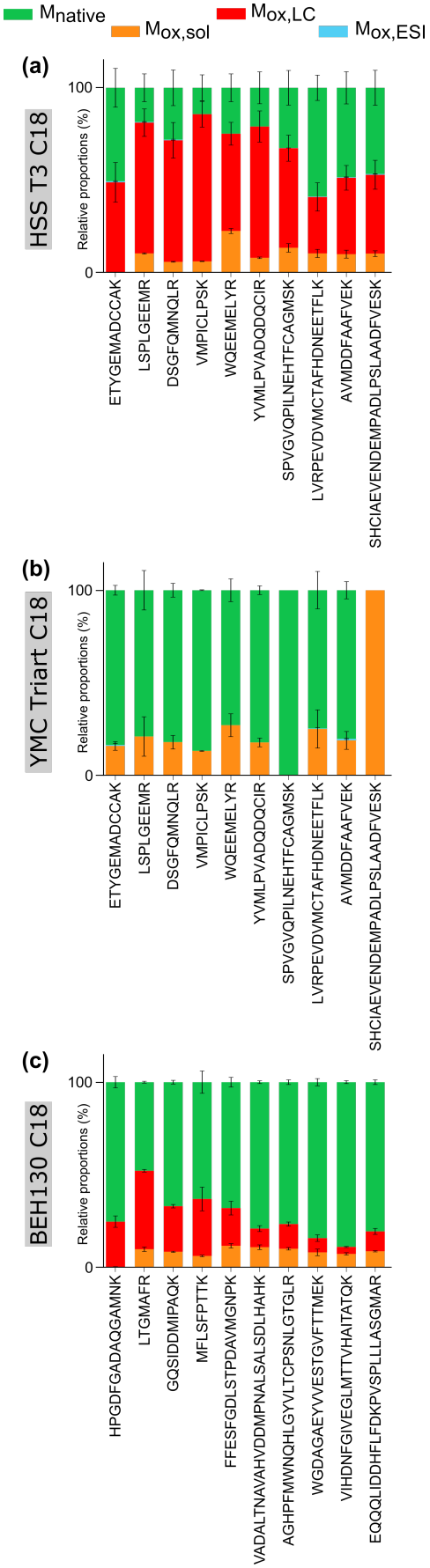


Figure 4.5: *Relative proportions of M_{native}, $M_{ox,sol}$, $M_{ox,LC}$ and $M_{ox,ESI}$ for the 10 most intense methionine-containing peptides identified in complex matrices using three different RP columns: (a) plasma sample injected on a HSS T3 C18 column following a 180 min-long method, (b) the same plasma sample injected on a YMC Triart C18 column following a 180 min-long method, and (c) skeletal muscle biopsy samples injected on a BEH130 C18 following a 175 min-long method. Peptides are sorted according to increasing retention time and error bars are estimated based on technical triplicates for (a) and (b) and biological triplicates for (c).*

Does the oxidation depend on the sample load?

To assess the impact of the sample load on $M_{ox,LC}$, four different quantity of commercial HeLa protein digests were injected in technical triplicate on the 1-year old columns set: 0.10 µg/9 µL, 0.25 µg/9 µL, 0.50 µg/9 µL and 1.00 µg/9 µL. Figure 4.6 shows the evolution of $M_{ox,LC}$ for a selected methionine-containing peptide (see appendix figure A.8 for more peptides). We observed that $M_{ox,LC}$ decreases as the injected protein load increases. Consequently, as $M_{ox,sol}$ stays constant, the evolution of M_{native} increases as the injected protein load increases. The data are fitted with a one-phase exponential decay model exhibiting a R^2 of 0.9746 and 0.9752 for $M_{ox,LC}$ and M_{native} respectively. Based on these results, we hypothesized that there is a limited number of available oxidation sites embodied in the stationary phase of the analytical column and that may be related to the increase in free silanols availability with the lifetime of the column. As such, a low number of methionine-containing peptides, associated with low sample load, results in a complete oxidation of these peptides as there are enough oxidation sites available within the analytical column to oxidise the entire population of these peptides, subsequently leading to higher $M_{ox,LC}$. On the contrary, a high number of methionine residues, associated with high sample load, cannot be entirely oxidised as the oxidation sites within the analytical column are saturated. This leads to a high percentage of methionine residues in their native form and, consequently, to the decrease of the proportion of oxidised residues.

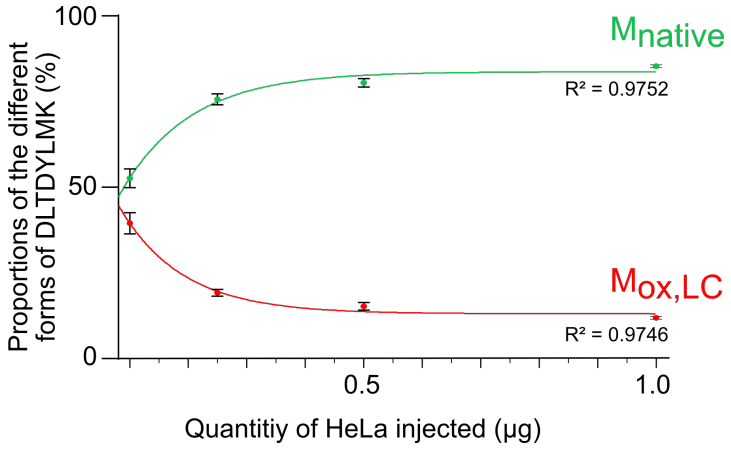


Figure 4.6: *Evolution of $M_{ox,LC}$ and M_{native} for DLTDYLMK peptide as function of the quantity of HeLa protein digest injected in the LC system. Four different quantities 0.10 µg, 0.25 µg, 0.50 µg and 1.00 µg were injected in triplicates on a 1-year old columns set using a 180 min UPLC method (3 min of trapping time and gradient length of 150 min). The data points are fitted by a one-phase exponential decay model with a R^2 of 0.9746 and 0.9752 for $M_{ox,LC}$ and M_{native} respectively. The oxidation level reaches in the analytical column increases as the sample load decreases evidencing the presence of limited available oxidation sites within the analytical column.*

4.1.5. Conclusions

In this work, we focused our research on the identification and the understanding of the parameters that influence the artefactual oxidation of methionine residues during the analytical separation of biomolecules and complex samples on reversed-phase chromatographic systems. We observed that, in addition to solution and electrospray oxidations, methionine-containing peptides can be significantly oxidised inside the analytical column, along the chromatographic separation. Such phenomenon is identified based on the retention time profile of the analytes: methionine-oxidised peptides are systematically characterized by a broad chromatographic peak having an intermediate retention time between methionine-containing peptide oxidised in solution and their native counterpart. Similar observation was noticed on another setup from another laboratory based on publicly available data. With our instrumental setup, we observed a progressive increase in on-column oxidation proportion with the age of the columns set, the gradient length and a reduced sample load. With regards to the bonding technology of the column (HSS T3), the observations were rationalized by the presence of a limited number of oxidation sites within the column stationary phase, being more and more exposed to the mobile phase and to the sample over time. Residual metal ions were pointed out as a source of variability in methionine oxidation measurement. Traces of metal ions could indeed potentially catalyse the oxidation of methionine residues in the chromatography. However, the fact that discrepancies are observed between old and new columns undermined this assumption. The difference between both configurations exclusively lies in the age of the trap and the analytical columns, while the buffers and the injection system were kept identical. Nevertheless, an accumulation of metal ions in the analytical column is not excluded and could damage the stationary phase by creating available oxidant sites in the column. The complete on-column oxidation suppression observed on the YMC Triart C18 column, being part of a biocompatible system, reinforces the metal ions accumulation hypothesis and suggests that the nature of the stationary phase may not be the issue. In the future, this assumption will be evaluated by using HSS T3 and BEH130 columns on biocompatible LC system over a year of routine injections.

Altogether, this work highlights the necessity to implement a procedure to follow the liquid chromatography instrumental setup status which induces undesired artefactual oxidation both in routine analysis and in general proteomics studies. At this stage, general guidelines can be drawn:
(1) A periodic monitoring of the on-column oxidation level based on standard methionine-containing peptides should help to diagnose LC system failures or an early disruption of the analytical column.
(2) Fast digestion workflows and short gradient lengths should be favoured in order to reduce undesired methionine oxidation in solution and during the LC separation respectively.
(3) For targeted therapeutic proteins studies, labelled internal standard will allow to correct for LC induced bias.

These guidelines are of particular interests for quantification studies in which database searches expect single chromatographic peak for oxidised methionine peptides. As demonstrated, several peaks are associated with a given oxidised analyte, such multiple chromatographic signatures can possibly lead to erroneous assignment and quantification.

Finally, complementary studies dedicated to the stabilization of the oxidation during the chromatographic separation will help to prevent and to further understand the origin of on-column methionine oxidation. Indeed, with the emergence of downscaling proteomics, analysis of peptide modifications when low quantity is injected need to be rigorously monitored and controlled. Other experiments will be conducted to draw conclusions with regards to the bonding technology of the stationary phase and the structure of the surface of the RP columns.

4.2. Strategies to understand the origin of on-column methionine oxidation

The results presented above helped us to precisely localized the $M_{ox,LC}$ phenomenon. Additional experiments were carried out to reach a better comprehension of this event and to provide strategies aiming at minimizing undesired LC oxidation.

4.2.1. Comprehension of the $M_{ox,LC}$ phenomenon

To evaluate the capacity of dissolved metal ions to oxidise methionine-containing peptides, the synthetic EMSGSPASGIPVK peptide was incubated with metal ions carrying different oxidation states, *i.e.* Fe^{2+}, Fe^{3+} and Mn^{2+}. We observed that the direct addition of metal ions to the peptide solution did not promote the oxidation of the methionine-containing peptide. To further enhanced the oxidation of methionine, the standard peptide together with metal cations was also incubated in the presence of hydrogen peroxide. The objective of that experiment was to generate free radicals that could potentially engage in secondary reactions such as the oxidation of methionine residues. The experimental details and results are summarized in the following section.

To account for residual metal ion traces released from the LC system to the mobile phase, the standard peptide at a concentration of 5 µM was incubated with metal ions of µM to nM concentrations. Solutions of 100 µM and 100 nM of Fe^{2+}, Fe^{3+} and Mn^{2+} were done using respectively iron (II) sulfate heptahydrate, amonium iron (III) sulfate dodecahydrate purchased from Sigma-Aldrich, and manganese sulfate monohydrate from VWR. To mimic the eluent composition as the peptide elutes from the LC column, the peptide and metal ions were dissolved in water/ACN 75/25 (v/v) with 0.1% of formic acid. Finally, H_2O_2 was added to the sample in a

final concentration of 0.1 mM. The incubation was carried out during 1h at room temperature before direct infusion on the Q Exactive™ mass spectrometer. Each condition was repeated in triplicates. Direct infusion experiments were acquired using the following parameters: scan range from 400 to 1750 m/z, resolution = 70,000, polarity = positive, lock mass = 445.12003, AGC target = 1×10^6. The parameters of the ESI source were set as follows: spray voltage = 4 kV, capillary temperature = 270°C, S-lens RF level = 50.0.

Figure 4.7 depicts the proportions of M_{native} and $M_{ox,sol}$ resulting from the incubation of EMSGSPASGIPVK methionine-containing peptide with H_2O_2 and Fe^{2+} or Fe^{3+} or Mn^{2+} in different concentrations. Two control experiments were run to asses our method and are labelled as "control" and "H_2O_2" in figure 4.7. "Control" refers to the incubation of the standard peptide with the solvent without metal cations or peroxide, while "H_2O_2" refers to the incubation of the standard peptide with the solvent and peroxide. The incubation of the peptide with H_2O_2 increases the $M_{ox,sol}$ proportion from 1,77 % (no addition of hydrogen peroxide) to 9.49 ± 2.71 %. The introduction of metal ions, regardless of their nature or their quantity, gives similar $M_{ox,sol}$ proportions than the hydrogen peroxide control: 7.01 ± 2.77 %, 10.46 ± 6.82 %, 9.55 ± 2.62 %, 10.45 ± 0.22 %, 8.97 ± 0.90 %, 9.37 ± 2.12 % respectively for Fe^{2+} 1 nM, Fe^{2+} 1 µM, Fe^{3+} 1 nM, Fe^{3+} 1 µM, Mn^{2+} 1 nM, Mn^{2+} 1 µM. In light of these results, metal ions did not seem to catalyse the oxidation reaction of the peptide.

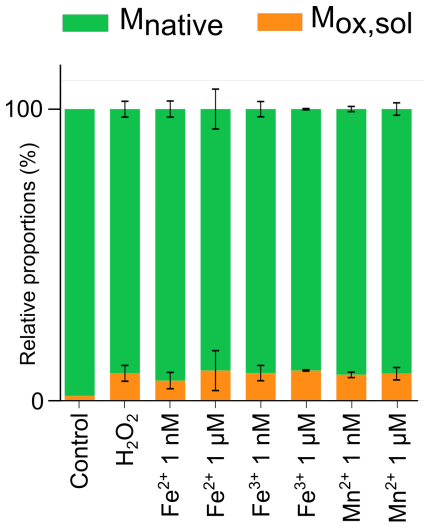


Figure 4.7: *Relative proportions of M_{native} and $M_{ox,sol}$ of EMSGSPASGIPVK peptide following different experimental conditions. Control is the standard peptide diluted like the other conditions without the addition of any metals and hydrogen peroxide. The following condition is the peptide with only the addition of hydrogen peroxide. This sample is used as a control of the efficacy of the metal ions. The next conditions are the peptide with hydrogen peroxide with the different sort and quantity of metal ions. $M_{ox,LC}$ being absent here because sample is infused in the mass spectrometer without previous LC separation.*

Although we did not verify our former hypothesis on the methionine oxidation phenomenon

occuring in the LC system, these experiments did not entirely ruled out the role of metal cations in the oxidation process. Indeed, several experimental parameters remain different between our test experiments and the LC system. Among them, the temperature of the oven in which lies the analytical column was set at 40°C whereas the 1h incubation was done at room temperature. Similarly, the pressure at which the sample is submitted in the column, whereas only the atmospheric pressure influenced the reaction in this experiment. And finally, the main difference is the contact of the sample with the stationary phase when flowing through the column. Other molecules, originating from the stationary phase itself or trapped on the column from previously injected samples, might also interact with the methionine residue further leading to their oxidation.

4.2.2. How to reduce $M_{ox,LC}$?

In this section, two strategies were implemented to reduce but also prevent the oxidation of methionine during LC separation. The first one relies on the use of metal ion binding chemicals to clean the column before LC-MS analysis, while the second one capitalizes on the use of antioxidant scavengers in the sample preparation step. The outcomes of both strategies will be presented in this section.

Column cleaning

To remove the possible metal ions from the LC column, the column was washed with a solution able to capture metal ions and further elute them from the column. EDTA was chosen to form water-soluble complexes with metal ions which were potentially present in the column. A solution stock of EDTA 20 mM pH=8 was prepared and 9 µL of this solution were injected as a sample on the column three times in a row. First, the synthetic methionine-containing peptide, prepared as described in section 4.1.3, was injected in triplicate before washing the column with EDTA. The column was then washed with the EDTA solution and the synthetic peptide was injected again in triplicate under the same instrumental conditions. The percentages of oxidation of the methionine-containing peptide were then compared between the two experiments. The proportions of $M_{ox,LC}$ were 47.84 ± 0.95 % and 48.99 ± 3.17 % respectively in samples injected before and after the injection of EDTA. These similar levels of $M_{ox,LC}$ indicated that the cleaning with EDTA as a sample did not have an effect on the oxidation level of injected peptides. Injected as a sample, EDTA may have not reached the analytical column and was probably sent to waste after going through the trap column.

To overcome this problem, the entire tubing and columns system was then washed using EDTA. Before washing the column and the tubing with EDTA, methionine-containing peptide was injected in triplicates onto the column. Then, the system was flushed with water/ACN

90:10 (v/v) for 30 minutes to remove any acid that could precipitate with the EDTA solution. Then the column was cleaned with a solution of ACN/20 mM EDTA pH=8 10/90 (v/v) for one hour. The system was reconditioned using water/ACN 90:10 (v/v) for 30 minutes. Finally, the methionine-containing peptide was injected into the washed column for comparison with the samples injected before the wash. This experiment resulted in damage of the column set leading to unexploitable data.

Scavenger treatment

Pure methionine inserted in a sample can act as a scavenger that can be oxidised in place of methionine-containing peptides present in a sample.

Tryptic digestion of human plasma samples was processed as described in section 4.1.3. Aliquots of digested plasma were then stored at -80°C. To be comparable to the plasma digest samples discussed in figure 4.2, samples of this section were processed in the exact same way (in terms of concentration, use of ice for thawing, quantity injected onto the columns, column set and experimental timing). The only difference is the addition of pure methionine to the plasma digest after dilution and prior to the subsequent injection. Pure methionine was added at a 1/5 ratio protein/methionine. In the 9 µL injected, there was 0.675 µg of digested plasma and 3,375 µg of pure methionine. L-methionine was purchased from Sigma-Aldrich. The acquisition method and the data processing were implemented likewise the peptides from plasma samples injected alone.

Figure 4.8(a) compares the proportions of M_{native}, $M_{ox,sol}$, $M_{ox,LC}$ and $M_{ox,ESI}$ for the 10 most abundant methionine-containing peptides of a digested plasma sample (plain bars) and the same plasma digest in which pure methionine has been spiked prior to the injection (dotted bars). Figure 4.8(b) is the proportions of only $M_{ox,LC}$ pulled out of from (a) to allow quick evaluation of oxidation occurring during the liquid chromatography separation. In both experimental conditions, with or without pure methionine, the proportions of $M_{ox,LC}$ is relatively high. Nevertheless, when looking at figure 4.8(b), $M_{ox,LC}$ is lower in the methionine-enriched condition for nearly all the represented peptides. This global trend encourages us to think that pure methionine effectively act as a scavenger of the oxidation inside the analytical column. Another observation from figure 4.8 is that $M_{ox,sol}$ slightly increases in the pure methionine enriched sample. This could be explained by the storage time differences in the freezer. Given that the pool of plasma has been digested in one pot and frozen at -80°C in aliquots and that plasma without methionine has been injected one week after being frozen at -80°C while plasma in which we added pure methionine has been injected 8 months later, this time past in the freezer could be the explanation to this slight increase of $M_{ox,sol}$.

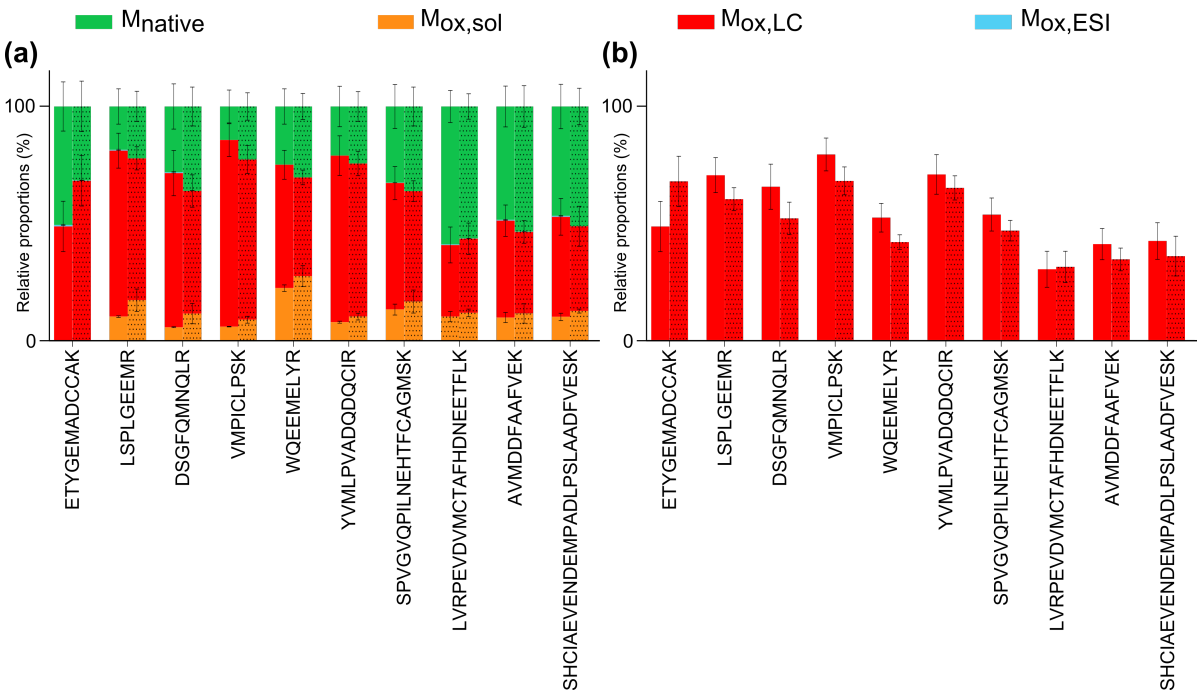


Figure 4.8: *Relative proportions of M_{native}, $M_{ox,sol}$, $M_{ox,LC}$ and $M_{ox,ESI}$ for the 10 most intense methionine-containing peptides identified in plasma samples following a 180 min-long UPLC method performed on a 1 year-old columns set with the HSS T3 phase. For each peptide, two columns are depicted: on the left plasma sample already presented in figure 4.2; on the right (bars patterned with dots) plasma samples with the addition of pure methionine before injection onto the mass spectrometer. (a) represents all forms of the peptides, non-oxidised and all the oxidised forms and (b) only represents the contribution of $M_{ox,LC}$ to understand at a glance the influence of pure methionine on $M_{ox,LC}$. Contribution from $M_{ox,ESI}$ is minor (< 1 %). Peptides are sorted according to increasing retention time and error bars are estimated based on technical triplicates.*

Digestion of plasma samples with an initial addition of pure methionine could help reducing $M_{ox,sol}$ as well as $M_{ox,LC}$.

4.3. Conclusions

We were able to find a scavenger that can reduce $M_{ox,LC}$, but it did not significantly decrease it. The most effective way to deal with high percentages of oxidation in analytical columns is to replace them with a new one. The experimenter should determine the $M_{ox,LC}$ thresholds at which the column must be replaced based on their laboratory standards. Methionine-containing peptides can be used as a quality control. In our laboratory, the commercial HeLa protein digest standard is already used to validate the performance of the LC-MS system based on the number of proteins detected. To assess the quality of the LC system, we evaluated the evolution of in-column oxidation in four peptides over one year (see figures 4.3 and A.4). The threshold for $M_{ox,LC}$ could be set at 5%, above which $M_{ox,LC}$ tends to increase more rapidly, and the LC system deteriorates quickly. Therefore, when Mox,LC reaches this threshold, the LC columns should be changed.

Is the in-column oxidation due to metal ions? Is it due to digestion residues that accumulate in the columns throughout the injections? Investigating this phenomenon by washing out the stationary phase of an old column set and analysing the molecules that are released could provide insight. Comparing the molecules found in the stationary phases of an old and new column set could also be valuable. While this phenomenon is not yet fully understood, this discovery can be used to prevent incorrect integration of methionine oxidation and ensure accurate quantitation.

A detailed study of EMSGSPASGIPVK has led us to understand the needs of a good standard peptide for the quantitation of potentially modified peptides. First, at the moment of the spike, the percentage of oxidation of the methionine-containing standard must be precisely known. Quantification of oxidised analyte peptides would be easier with completely unmodified standard peptide but this ideal view is unfortunately rarely achieved. In most cases, a small percentage of oxidation is detected in the standard peptide. This problem does not exclude the use of methionine-containing peptides as standards as long as the fraction of oxidised molecules in solution is known. The best way of including such standard in the kit is to keep it lyophilised and solubilized it at the very last moment of the sample preparation procedure. From batches to batches, a quality control will be conducted, and $M_{ox,sol}$ will be accurately measured. The resulting lyophilized powder will be stored under controlled conditions (which need to be further investigated) to prevent any undesired reactions and enable accurate quantification of methionine-oxidised analytes. Second, integration of methionine oxidised peptide via automated software must be checked and corrected if needed. Indeed, we have identified multiple chromatographic peaks associated with oxidised forms of the peptide that are not all recognised by automatic integration, leading to imprecise quantification.

By carefully following the requirements outlined above, it is possible to achieve absolute quantification of proteins using methionine-containing peptides. However, this method is limited to targeted proteomics with a small number of samples to analyse. For cohorts with a large number of samples that need to be compared, relative quantification is the preferred approach. The next chapter presents a new method for evaluating differential PTMs for biomarker discovery studies.

5

Differential post-translational modifications

In addition to the specific quantification of methionine-containing peptides introduced in the previous chapter, the relative quantification of post-translational modifications (PTMs) in general is of major interest for biomarker discovery studies performed on large cohorts. As a consequence, various methodologies such as ”bottom-up” proteomics approach, the use of DIA acquisition, the ”top-down” approach, enrichment methods were developed in recent years to improve PTM characterization. This chapter introduces a new strategy that aims to improve digestion efficiency and increase protein sequence coverage to further enhance PTM characterization.

5.1. Workflow

In this context, we here describe a novel innovative strategy that combines multi-enzymatic limited digestion (MELD) to tryptic digestion to reinforce the characterization of PTMs and further allow their quantification.

The MELD method has already been described in 2019 by Morsa *et al.* [113]. It is an enzymatic digestion that lies between the bottom-up and the middle-down strategies. Indeed, thanks to a mix of several enzymes, the MELD can generate more various peptides than in classic tryptic digestion due to the multiplicity of cleavages sites. This multi-enzymatic digestion is limited thanks to a dilution of the enzymes used and a shorter digestion time (only 2h of digestion). It produces many missed cleavage sites and therefore results in a broad range of unique peptides covering common sequences (overlapping peptides). The sequence coverage is then improved as well as the confidence for the detection and localization of PTMs.

Thanks to overlapping peptides, if a modification site is observed multiple time in several extended residues, its confidence certainly increases due to redundancy observations and the multiple PSMs. This method synergistically combined with the tryptic digestion will offer robust and improved detection and quantification of modified peptides. Comparatively to other digestion methods, it offers an improved sequence coverage and characterization accuracy of peptides and proteins. It also enables the accurate identification and localization of PTMs and enhances the diversity of peptides in terms of length and cleavage site.

Figure 5.1 presents the schematic representation of the method's analysis flow. The method involves two differential analyses, one on the protein level and the other on the peptide level, to distinguish differentially expressed modified peptides from those originating from an already overexpressed protein. To analyse a cohort of samples using this workflow, disease patient samples and control patient samples are pooled separately, and two MELDs are required: one for the pool of control patients and one for the pool of disease patients. These pools undergo technical triplicates for analysis. The MELD analyses, which are processed separately for each group, provide an overview of the number and nature of modifications in both groups. The differential analysis of the MELD analyses generates a list of protein modification types and localizations, detected in all the samples, based on multiple peptide detections that result in high confidence. Next, all samples are subjected to a typical tryptic digestion for quantification of the modifications after comparison with the list obtained with the MELD experiment. This step is crucial as it provides a final list of confident peptides for which quantification results are available. Differential peptide analysis is performed on this final list of confident peptides using PERSEUS, thanks to the label-free quantification performed separately on the different tryptic digest samples.

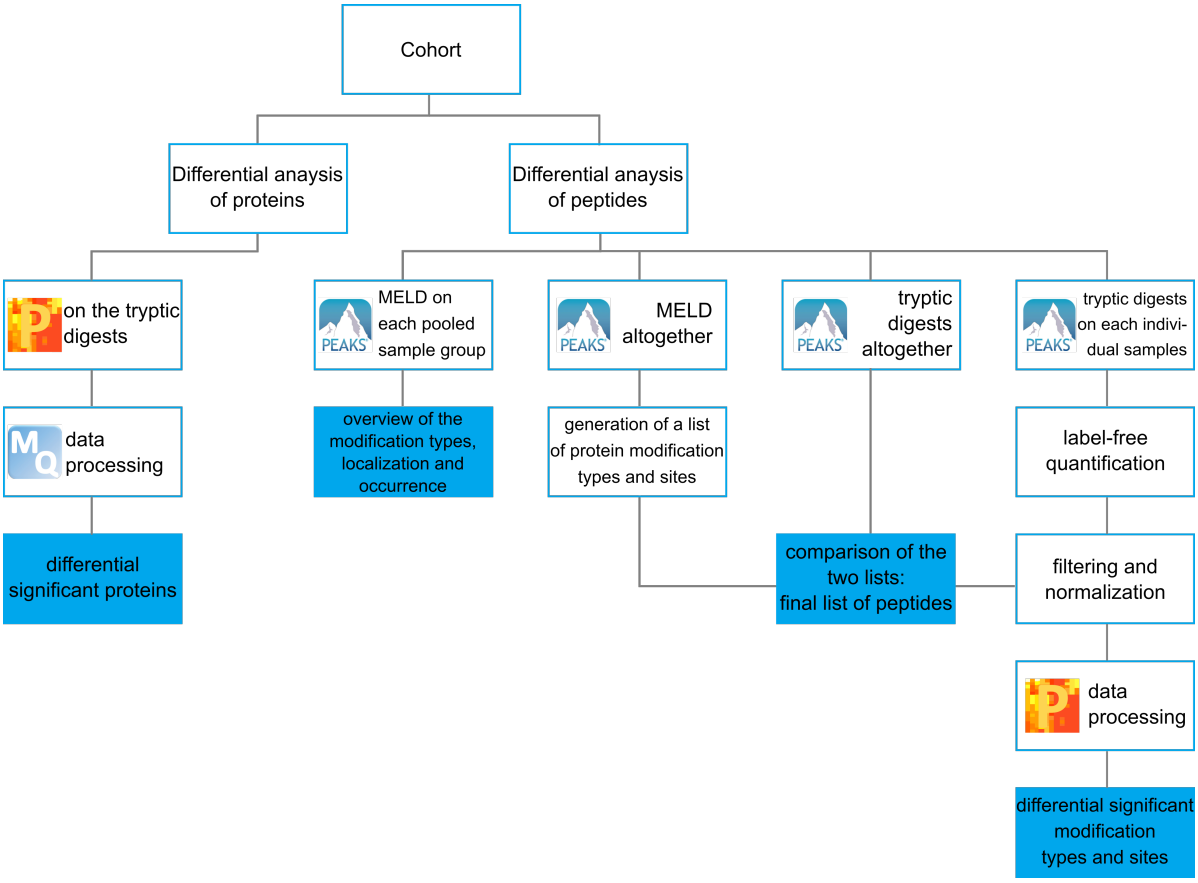


Figure 5.1: *Workflow for the analysis of PTMs in a new approach.*

5.2. Applications

This new PTMs strategy analysis was applied on the samples of the Win2Wal EPINOX (Wallonia research SPW funding) project. EPINOX stands for Erythrocyte Proteomics for ImmuNOdetection of OXidation and aims at the identification of new epitopes from erythrocyte modified proteins that are representative of an in vivo oxidation stress and NO metabolism dysregulation. These epitopes will further be used to produce immunological tests for cardiovascular disease prevention. These biomarkers will therefore be highlighted by the newly developed proteomic approach described above.

5.2.1. Introduction

Endothelial cells form the inner layer (the endothelium) of a blood vessel and play critical role in the cardiovascular homeostasis. The endothelium helps in producing and controlling the synthesis of different substances that maintain cell capacity. This functional capacity may

be altered and therefore referred as the so-called endothelial dysfunction. It is evaluated based on the amount of nitric oxide (NO) produced and the vasodilatation effect. The increase of the inner diameter of the blood vessels, called vasodilatation, that helps the transport of the blood flow, is facilitated by vasodilators such as NO. The impairment of this NO-dependent vasodilatation leads to inflammatory processes and vascular diseases [141]. Under normal conditions, concentration of NO in the vascular system is generally in the order of the nM [142]. An imbalance between reactive oxygen species (ROS) or reactive nitrogen species (RNS) and antioxidants is the primary cause of endothelial dysfunction. This dysfunction leads to an increase of NO (in the order of µM) that favours its reaction with ROS and RNS. The bioavailability of NO is therefore drastically reduced by the creation of oxidant molecules taking part in the oxidative stress. NO is also involved in another process: it binds to the haemoglobin of circulating red blood cells (RBCs) leading to the formation of nitrosylated haemoglobin (HbNO). In 2013, Lobysheva *et al.* [143] showed that HbNO levels directly correlates with vascular endothelial function. These observations highlight that endothelial dysfunction is associated with i) a decrease of NO bioavailability, ii) a decrease concentration of HbNO and iii) an increase of the oxidative stress. The origin of the oxidative stress is still to be discovered and could end up to new targeted treatments. The dosage of HbNO levels has been patented and can be used as a sensor of oxidative stress.

Apart from the drop in HbNO levels, oxidative stress can cause modifications at the proteome level that can be classified in two categories: reversible and irreversible modifications [144]. These include cysteine modification products as well as nitration and carbonylation, respectively. Nitration is the proliferation of 3-nitrotyrosine by the addition of a nitro functional group onto a tyrosine residue (or to a lesser extent a tryptophan residue). The formation of 3-nitrotyrosine can be explained by two mechanisms involving ROS, RNS and NO. In the first mechanism, the tyrosine residue reacts with ROS to form a tyrosyl radical quickly combining with NO to yield the desired product. In the second one, RNS or ROS directly react with NO to give an unstable peroxynitrite intermediate that further react with the tyrosine residue [145]. Review of the literature on this subject help us to draw a list of the most cited and common protein carbonylation which refers to the introduction of carbonyl groups (C=O) through a variety of oxidative pathways involving ROS [146–153]. In biological samples, the generation of protein carbonylation is often metal-catalysed oxidation. At first, an extended list of modifications was tested on our samples but resulted in the detection of only 9 of them, which were kept for the following analyses. The nitration and the carbonylation modification types are summarised in table 5.1.

Table 5.1: *List of the nitration and the 9 carbonylations most encountered in our data sets with the name of the modification, its composition and its mass difference.*

Amino acid	Modification	Composition	△mass (Da)
Tyrosine/Tryptophan Y/W	nitration	-H, +N, +2O	+44.98
Proline P	5-oxoproline= pyroglutamic acid	-2H, +1O	+13.98
Proline P	glutamic semialdehyde	+O	+15.99
Threonine T	2-amino-3-ketobutyric acid	-2H	-2.02
Lysine K	aminoadipic semialdehyde	3H, -1N, +1O	-1.03
Arginine R	glutamic semialdehyde	-5H, -1C, -3N, +1O	-43.05
Methionine M	aspartate 4 semi-aldehyde	-4H, -1C, -1S, +1O	-32.01
Glutamic acid E	4-aminobutyraldehyde	-2H, -1C, -1O	-30.1
Aspartic acid D	decarboxylation	-2H, -1C, -1O	-30.1
Leucine L	leucine carbonyl	-2H, +1O	+13.98

Oxidative stress is difficult to measure because quick intermediates or fast-degrading molecules come into play. It can however promote protein modifications on RBCs protein constituent, that can be detectable and quantitated. Looking at RBCs, instead of commonly analysed plasma biomarkers, is a choice dictated by the lifetime of erythrocytes as well as their absence of nucleus. Indeed, a RBC approximately circulates during 120 days and, because of the absence of nucleus, it does not regenerate itself [154, 155]. Presence of carbonylated or nitrated proteins at the surface of erythrocytes is therefore regarded as an excess of oxidative stress and will be, in our hypothesis, the first sign of an endothelial dysfunction. It reflects, in a more dynamic way, relatively new and active oxidation processes, better than the plasma products. Other risk factors such as hypertension, hypercholesterolaemia, diabetes, inflammation, hyperglycaemia, alcohol, tobacco as well as ageing may increase the oxidative stress. Endothelial dysfunction leads to cardiovascular diseases, the first cause of death worldwide and a socio-economic burden.

Several cohorts of patients were recruited in this project, all of them correlated to control patients: contraceptive pills consumers, diabetes patients, COVID patients, patients with high or intermediate risk of cardiovascular diseases. Only the latter one is presented in this thesis because the cohort contains the highest number of patients (in comparison to the other cohorts) and is therefore statistically more relevant than the others. At first only washed erythrocytes were analysed. A washing protocol was established to separate the membranes of the erythrocytes from the matrix to produce ghost erythrocytes. Throughout the analyses of these cohorts, we struggled finding differential modified peptides on the cohorts and we therefore focused our research on complete RBCs. For the sake of completeness, plasma samples of the same patients were also analysed in the same fashion. The cohort is made of 63 RBC samples (41 high or intermediate risk and 22 control) and 61 plasma samples (39 high or intermediate risk and 22 control). The difficulty lies in the recruitment of control patients with the same biological characteristics (*i.e.* age, BMI, etc.) without clinical risk of cardiovascular disease. The control group contains individuals with a mean age of 49 ± 17 % and a mean BMI of 24.3 ± 2.3 % whereas the mean age of the patient group is 65 ± 12.1 % and a mean BMI of 27.0 ± 6.5 %. The same proportion of men and women constitutes the two groups:

59 % women and 41 % men. None of the samples comes from tobacco consumers. Other biological and clinical characteristics of the patients can be found in table 5.2.

Table 5.2: *Clinical and biological characteristics of the study subjects. HC = hypercholesterolaemia; DT2 = type 2 diabetes; HTN = hypertension; CVA = cerebrovascular accident; CVD = cardiovascular disease; AF = atrial fibrillation; CKD = chronic kidney disease. Medical treatment of each individual is known but is not shown in this table for sake of clarity*

Risk	RBC	Plasma	Peroxides	Age	Weight (kg)	Height (m)	BMI	Gender	Chol	TG	HbNO (nM)	Diseases	Tobacco consumption
control	v	v	705.35	69	NA	NA	22.6	Woman	235	89	87	NA	no
control	v	v	601.1	62	NA	NA	21.9	Woman	180	134	197	no	no
control	v	v	863.66	62	NA	NA	22.3	Woman	133	228	NA	NA	no
control	v	v	487.69	62	NA	NA	26.5	Man	127	188	344	no	no
control	v	v	407.24	62	NA	NA	20.4	Woman	NA	NA	NA	no	no
control	v	v	447.44	61	NA	NA	25.7	Man	201	92	NA	no	no
control	v	v	332.07	69	NA	NA	22.2	Woman	0	0	NA	no	no
control	v	v	467.59	65	NA	NA	25.5	Man	187	186	NA	no	no
control	v	v	1221.1	59	64	1.64	23.8	Woman	NA	NA	NA	rheumatoid arthritis	no
control	v	v	396.06	60	NA	NA	24.9	Woman	NA	NA	NA	no	no
control	v	v	311.1	65	85	1.77	27.1	Man	178	103	NA	sleep apnoea	no
control	v	v	535.28	75	NA	NA	25.5	Woman	NA	NA	NA	no	no
control	v	v	459.47	28	NA	NA	21.8	Woman	NA	NA	NA	no	no
control	v	v	1459.42	30	NA	NA	20.9	Woman	NA	NA	NA	no	no
control	v	v	1726.08	26	NA	NA	21.6	Woman	NA	NA	NA	no	no
control	v	v	192.96	37	NA	NA	28.3	Man	NA	NA	NA	no	no
control	v	v	817.51	34	NA	NA	26.2	Woman	NA	NA	NA	no	no
control	v	v	417.14	29	NA	NA	26.8	Woman	NA	NA	NA	no	no
control	v	v	334.87	42	NA	NA	23.2	Man	NA	NA	NA	no	no
control	v	v	183.44	31	NA	NA	27.6	Man	NA	NA	NA	no	no
control	v	v	442.91	29	NA	NA	25.8	Man	NA	NA	NA	no	no
control	v	v	416.3	29	NA	NA	23.8	Man	NA	NA	NA	no	no
HR+Interm	v	v	796.98	64	NA	NA	30.1	Woman	139	223	40	HTN, hypothyroidism	no
HR+Interm	v	v	753.54	81	NA	NA	29.5	Man	NA	NA	33	HTN, infarctus, AF, renal failure, transient ischemic attack	no
HR+Interm	v	v	-15.98	77	58	1.75	18.9	Man	NA	NA	NA	HC, heart failure	no
HR+Interm	v	v	579.74	78	NA	NA	27.1	Man	175	49	46	IRC, HTN	no
HR+Interm	v	v	364.79	60	86	1.9	23.8	Man	236	176	NA	HC, pancreatic tumor	no
HR+Interm	v	v	413.91	43	58	1.58	23.2	Woman	138	49	NA	HC	no
HR+Interm	v	v	279.95	39	84	1.78	26.5	Man	147	134	NA	HC, HTN, CVA	oui
HR+Interm	v	v	219.31	55	91	1.76	29.4	Man	133	228	NA	HC, HTN	no
HR+Interm	v	v	343.39	82	63	1.7	21.8	Man	77	166	NA	HC, HTN, heart failure	no
HR+Interm	v	v	891.43	55	NA	NA	23.7	Woman	217	72	46	acute renal failure	no
HR+Interm	v	v	283.17	67	80	1.8	24.7	Man	221	164	NA	HC	no
HR+Interm	v		NA	74	57	1.42	28.3	Man	NA	NA	NA	HC, CVA	no
HR+Interm	v	v	372.22	64	72	1.72	24.3	Man	195	157	NA	HC	no
HR+Interm	v	v	586.14	77	NA	NA	19.0	Man	147	114	NA	infarctus AF	no
HR+Interm	v	v	735.38	36	NA	NA	30.9	Man	205	157	NA	HC	no
HR+Interm	v	v	1594.81	67	NA	NA	23.0	Woman	177	81	NA	NA	no
HR+Interm	v		NA	64	78	1.52	33.8	Woman	241	342	NA	HC	no
HR+Interm	v	v	897.72	59	82	1.82	24.8	Man	228	237	NA	HC	no
HR+Interm	v	v	719.5	71	NA	NA	26.4	Woman	280	143	NA	CVA, HC	no
HR+Interm	v	v	751.26	62	NA	NA	22.9	Woman	224	74	NA	HC	no
HR+Interm	v	v	NA	67	83	1.79	25.9	Man	168	160	NA	HTN	no

HR+Interm	v	v	1347.75	58	54.5	1.58	21.8	Woman	239	87	NA	HC	no
HR+Interm	v	v	716.43	56	58	1.5	25.8	Woman	225	146	NA	HC	no
HR+Interm	v	v	642.36	67	67	1.6	26.2	Woman	248	63	NA	HC, HTN	no
HR+Interm	v	v	1339.26	83	138	1.55	57.4	Woman	139	208	NA	HTN	no
HR+Interm	v	v	733.84	73	NA	NA	26.6	Woman	NA	NA	NA	HC, sleep apnoea	no
HR+Interm	v	v	564.5	56	75	1.64	27.9	Woman	NA	NA	NA	HTN	no
HR+Interm	v	v	676.95	61	61	1.47	28.0	Woman	209	254	NA	HTN	no
HR+Interm	v	v	726.46	82	NA	NA	25.9	Man	123	79	NA	HTN, cardiac stimulator, AF	no
HR+Interm	v	v	2469.26	42	NA	NA	41.1	Woman	213	241	NA	NA	no
HR+Interm	v	v	233.77	66	61	1.73	20.4	Man	254	67	NA	HC	no
HR+Interm	v	v	900.98	71	NA	NA	17.3	Woman	252	115	NA	Infarctus, HC	no
HR+Interm	v	v	694.6	55	NA	NA	25.5	Woman	257	61	NA	HC	no
HR+Interm	v	v	147.42	83	79	1.71	27.0	Man	NA	NA	NA	HC	no
HR+Interm	v	v	714.36	74	68	1.57	27.6	Woman	269	141	NA	HC	no
HR+Interm	v	v	387.53	78	86	1.68	30.5	Man	180	80	NA	HTN	no
HR+Interm	v	v	987.57	58	64	1.64	23.8	Woman	199	120	NA	DT2	no
HR+Interm	v	v	540.13	66	70	1.66	25.4	Woman	231	182	NA	HC	no
HR+Interm	v	v	1121.32	59	82	1.75	26.8	Man	196	192	NA	HC, sarcoma soft tissue	no
HR+Interm	v	v	208.25	54	84	1.72	28.4	Man	206	230	NA	DT2, HC	no
HR+Interm	v	v	818.23	66	NA	NA	22.5	Man	215	136	NA	CVD family history	no

5.2.2. Materials and methods

Sample preparation

Patients were recruited by the FATH research group in the University of UCLouvain. Blood sampling were collected via venipuncture and transferred in tubes with EDTA and centrifuged. Plasma was separated in 500 µL tubes. Platelets and white blood cells were removed through pipetting and RBCs were finally placed in 500 µL tubes. All aliquots were stored at -80°C and send to ULiège for further analyses. To avoid any sample preparation bias, all the samples, either RBC or plasma were treated similarly. Each sample was thawed on ice. 950 µl of a solution of Tris-HCl 50 mM pH=7.4 with 1 % SDS, PhosSTOP and cOmplete were added to the 50 µL of sample as a lysis buffer. The sample was then diluted twice with Tris-HCl 50 mM pH=7.4 with 1 % SDS. The total protein content was evaluated based on the RC DC protein assay kit from Bio-Rad and resuspended with Tris-HCl 50 mM pH=7.4 to 1 mg/ml.

Tryptic digestion of red blood cell and plasma samples

Each sample was processed separately using the same protocol as described for the human plasma samples of the chapter 4.1.3 on 20 µg of the plasma and RBC samples. After tryptic digestion, samples are resuspended in water/TFA 0.1 % (v/v) to a final concentration of 0.11 µg/µL and 9 µL were injected on the column (1 µg).

Multi-enzymatic limited digestion of red blood cell and plasma samples

Each sample of the two groups was equimolarly pooled together, i.e the 41 or 39 (respectively for the RBC or for the plasma) patients constitutes a pool and the 22 controls formed another pool. 20 µg of each pool were treated in triplicate and submitted to reduction-alkylation followed by a clean-up. The dried pellets were solubilized in 40 µL of NH_4HCO_3 50 mM and further divided in two equal volumes, each of the tube containing 10 µg in 20 µL. Two multi-enzymatic solutions containing different ratio of diluted enzymes were prepared and immediately added to the samples. The first one was a mix of trypsin, chymotrypsin and Glu-c inserted at a final enzyme-protein ratio of 1/85, 1/55 and 1/85 respectively. The second one was a 9-fold dilution of the first mix. $CaCl_2$ was also added to the enzymatic mix to reach a concentration of 5 mM during digestion. One of the samples was digested with the high-ratio mixture and the other one with the low-ratio mixture for 2 h at 37°C under stirring at 600 rpm before quenching the digestion with a final concentration of 0.5 % TFA. The two digests are then pooled together in equal volume. They are finally dried under vacuum and resuspended in water/TFA 0.1 % (v/v) to a final concentration of 0.17 µg/µL and 9 µL were injected on the column (1.5 µg).

An overview of the treatment of each sample is represented in figure 5.2.

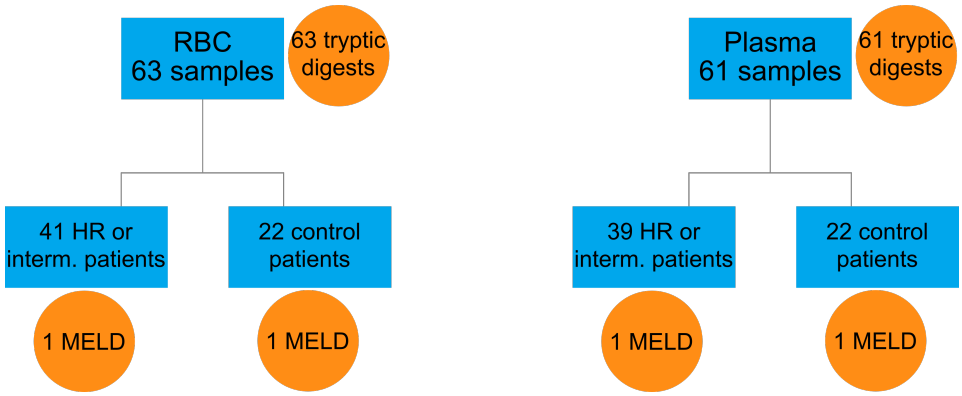


Figure 5.2: *Overview of the samples, their group (in blue) and their preparation process applied (in orange).*

Liquid chromatography - mass spectrometry analysis

Samples were injected on an ACQUITY UPLC M-Class System hyphenated to a Q Exactive™ Hybrid Quadrupole-Orbitrap™ Mass Spectrometer equipped with a Nanospray Flex™ ion source already described in paragraph 4.1.3. The gradient used was a 180 min-long UPLC method (3 min trapping time and 177 min analytical gradient).

Database search

A differential analysis at the protein level (on the tryptic digests) as well as the peptide level was conducted (see figure 5.1). For the differential analysis of proteins, MaxQuant software v.1.6.14 [156] was used for database searches on the tryptic digests of the RBC and plasma samples both with a human database containing 20 365 sequences (Uniprot, Swiss-Prot reviewed, Homo Sapiens 9606, downloaded on 22/10/2019). Carbamidomethylation of cysteines was set as a fixed modification, whereas deamidation of asparagine and glutamine, and oxidation of methionine were set as variable modifications. A maximum of 5 PTMs per peptide was allowed. Trypsin was selected according to a specific digest mode with a maximum of 2 missed cleavages allowed. Label-free quantification was selected based on a classic normalization type included in the software.

PEAKS Studio software v.10.5 was used for database searches of the differential analysis of peptides. A first global search was done on the two separated MELD to obtain modification categories in function of the group. The two MELD were then treated separately with the following parameters: human database containing 20,365 sequences (Uniprot, Swiss-Prot reviewed,

Homo Sapiens 9606, downloaded on 22/10/2019), unspecific digestion mode, tolerance on the precursor and fragment masses = 5.0 ppm and 0.015 Da respectively, fixed modifications = carbamidomethylation of cysteines, variable modifications = 35 commonly found modifications (including the list presented in table 5.1), FDR < 0.1 %, maximum variable PTMs per peptide = 5. Then, a first list of modified peptides was created thanks to the MELD analysis to obtain a list of robust peptides. For this purpose, the MELD from the patient and the control groups were combined for a single search with the same parameters. A second list of peptides was obtained from the analyses of all the tryptic digests for comparison with the list of peptides from the MELD. The parameters were the same as those described hereinabove, except for the digestion enzyme which was trypsin in this search. Another search was done on all the tryptic digests separately to obtain a quantification for each sample. All the results were then grouped in one file thanks to the label-free quantification option in PEAKS that allows grouping of pre-treated samples. For this analysis, the mass error tolerance was fixed to 5 ppm, the retention time shift tolerance to 5 min and the FDR threshold to less than 0.1 %.

Perseus software v.1.6.10.0 [157] was used for the statistical analysis of patient groups after database searches carried out in MaxQuant or PEAKS Studio.

5.2.3. Differential analysis on proteins

The analysis of the RBC displays a contamination by keratin. Indeed, out of the 241 detected proteins, two keratins are part of the 20 most abundant proteins of the sample with a high sequence coverage. This contamination is not encountered in the plasma samples.

241 proteins were detected in the RBC datasets, whereas 264 were detected in the plasma datasets. After filtering to only keep the proteins that are present in each sample of the cohorts, a Student's T-test was performed with a threshold p-value of 0.05. The final lists of proteins differentially expressed in one group or the other are presented in tables 5.3 and 5.4 respectively for the RBC and the plasma datasets. In the RBC samples, 16 proteins are differentially expressed and all overexpressed in the control group. Concerning the plasma samples, the 41 proteins differentially expressed are either overexpressed in one group or the other (see overexpression factor in tables 5.3 and 5.4). The factors are barely higher than one for every protein, indicating slight variations between the controls and the patients. These subtle variations in factors among the two groups imply that a clear distinction between the groups cannot be established at the protein level.

Table 5.3: *List of the 16 proteins differentially expressed in the RBC datasets. Proteins are ordered by p-value.*

Majority protein IDs	Protein names	Gene names	p-value	overexpressed in CVD	overexpressed in CONTROL
P16157	Ankyrin-1	ANK1	3.50E-03		1.09
P30043	Flavin reductase (NADPH)	BLVRB	4.86E-03		1.16
P00352	Retinal dehydrogenase 1	ALDH1A1	7.50E-03		1.15
P11171	Protein 4.1	EPB41	8.14E-03		1.12
P32119	Peroxiredoxin-2	PRDX2	9.63E-03		1.15
Q99497	Protein deglycase DJ-1	PARK7	1.08E-02		1.15
P02652	Apolipoprotein A-II	APOA2	1.18E-02		1.33
P00441	Superoxide dismutase [Cu-Zn]	SOD1	1.72E-02		1.11
P02549	Spectrin alpha chain, erythrocytic 1	SPTA1	2.46E-02		1.08
Q06323	Proteasome activator complex subunit 1	PSME1	2.58E-02		1.13
P11277	Spectrin beta chain, erythrocytic	SPTB	2.63E-02		1.07
P00491	Purine nucleoside phosphorylase	PNP	2.73E-02		1.11
P23528	Cofilin-1	CFL1	3.03E-02		1.14
P04040	Catalase	CAT	3.89E-02		1.08
P30041	Peroxiredoxin-6	PRDX6	3.95E-02		1.11
P00915	Carbonic anhydrase 1	CA1	4.85E-02		1.13

Table 5.4: *List of the 41 proteins differentially expressed in the plasma datasets. Proteins are ordered by p-value.*

Majority protein IDs	Protein names	Gene names	p-value	overexpressed in CVD	overexpressed in CONTROL
P68871	Hemoglobin subunit beta	HBB	1.15E-04		1.98
P69905	Hemoglobin subunit alpha	HBA1	1.68E-04		1.95
P02749	Beta-2-glycoprotein 1	APOH	4.14E-04	1.32	
P12259	Coagulation factor V	F5	1.79E-03	1.27	
P04003	C4b-binding protein alpha chain	C4BPA	1.85E-03	1.20	
P63261	Actin, cytoplasmic 2	ACTG1	2.82E-03		1.83
P01024	Complement C3	C3	2.88E-03	1.11	
P0DJI8	Serum amyloid A-1 protein	SAA1	3.13E-03	3.12	
P02787	Serotransferrin	TF	3.68E-03		1.19
P02647	Apolipoprotein A-I	APOA1	5.44E-03		1.24
P0DOX7	Immunoglobulin kappa light chain	IGK	5.67E-03		1.19
P0C0L5	Complement C4-B	C4B	8.00E-03	1.27	
P02748	Complement component C9	C9	8.01E-03	1.27	
P06681	Complement C2	C2	8.51E-03	1.11	
P02675	Fibrinogen beta chain	FGB	9.72E-03	1.20	
P01624	Ig kappa chain V-III region POM		1.07E-02	1.35	
P02763	Alpha-1-acid glycoprotein 1	ORM1	1.10E-02	1.27	
P00488	Coagulation factor XIII A chain	F13A1	1.11E-02		1.37
P02765	Alpha-2-HS-glycoprotein	AHSG	1.13E-02		1.19
P02760	Protein AMBP	AMBP	1.15E-02	1.36	
P80108	Phosphatidylinositol-glycan-specific phospholipase D	GPLD1	1.22E-02		1.26
P06312	Ig kappa chain V-IV region	IGKV4-1	1.41E-02		1.34
P02743	Serum amyloid P-component	APCS	1.48E-02	1.25	
P08603	Complement factor H	CFH	1.48E-02	1.12	
P01008	Antithrombin-III	SERPINC1	1.64E-02		1.10
P05156	Complement factor I	CFI	1.67E-02	1.15	
P0DOX5	Ig gamma-1 chain C region	IGHG1	1.69E-02		1.18
P00748	Coagulation factor XII	F12	1.84E-02		1.30
P04406	Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	1.86E-02		1.56
P02679	Fibrinogen gamma chain	FGG	1.92E-02	1.18	
P01619	Ig kappa chain V-III region B6		1.95E-02	1.25	
P01031	Complement C5	C5	2.10E-02	1.11	
P02652	Apolipoprotein A-II	APOA2	2.13E-02		1.31
P02750	Leucine-rich alpha-2-glycoprotein	LRG1	2.33E-02	1.25	
P01834	Ig kappa chain C region	IGKC	2.52E-02		1.20
P08185	Corticosteroid-binding globulin	SERPINA6	3.32E-02		1.18
P01780	Ig heavy chain V-III region JON	IGHV3-21	3.46E-02		1.24
P02671	Fibrinogen alpha chain	FGA	4.08E-02	1.14	
P04433	Ig kappa chain V-III region VG	IGKV3D-11	4.13E-02		1.29
P06396	Gelsolin	GSN	4.23E-02		1.11
P35542	Serum amyloid A-4 protein	SAA4	4.33E-02	1.19	

We introduced these differentially expressed proteins in the STRING database [158] to evaluate the protein-protein interactions and obtained biological pathway for RBC and plasma respectively observed in figure 5.3(a) and (b). In RBC, a biological process associated with the hydrogen peroxide metabolic process is observed ((a) in red) and a cellular component: the spectrin-associated skeleton ((a) in blue). In plasma, biological process linked to a response to stress ((b) in green) is monitored as well as the KEGG pathway linked to the complement and coagulation cascades ((b) in yellow).

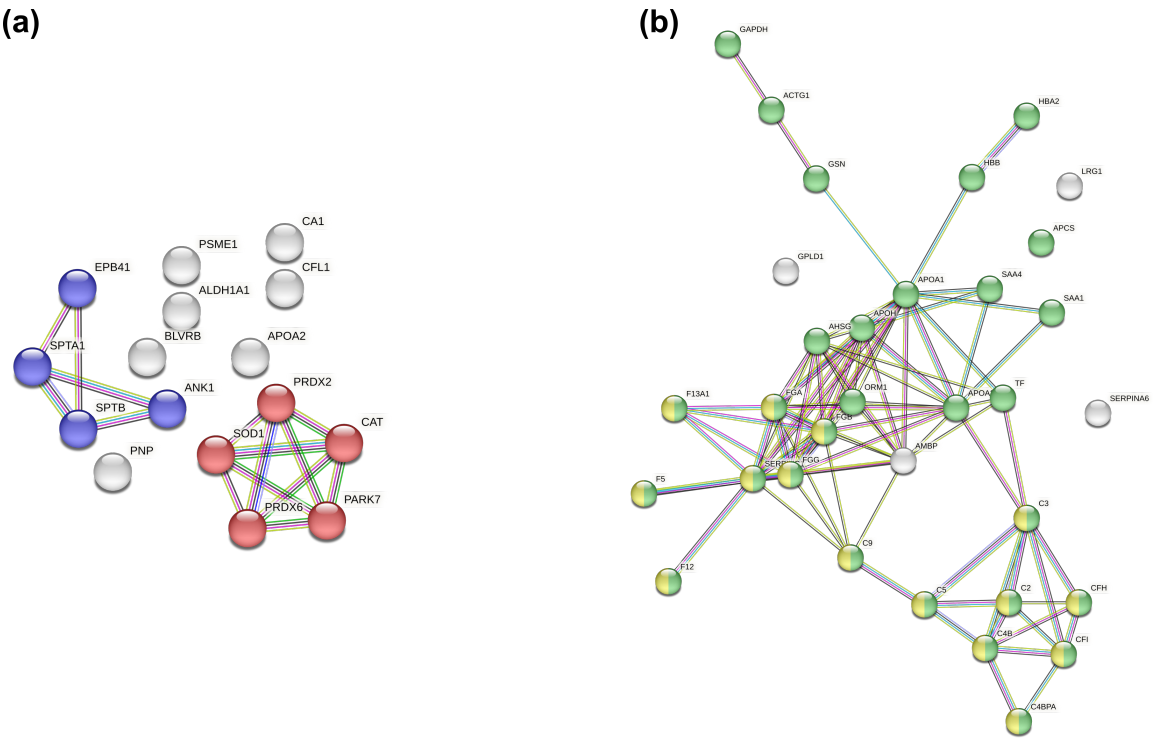


Figure 5.3: *Protein-protein interactions obtained thanks to the STRING database. Differentially expressed proteins found in (a) RBC and (b) plasma. In red, biological process associated with hydrogen peroxide metabolic process, in blue a cellular component: the spectrin-associated skeleton, in green biological process linked to a response to stress and in yellow KEGG pathway linked to the complement and coagulation cascades.*

In both RBC and plasma samples we found protein-protein interactions linked to the oxidative stress. Indeed, hydrogen peroxide metabolic process and the response to stress are directly linked to the oxidative stress. Plasma is known to contain proteins involved in defence mechanisms against pathogens (the complement and coagulation cascades). Spectrin-based skeleton are a set of proteins that maintain organization inside membrane protein domains [159]. The proteins highlighted in yellow in figure 5.3b were already detected as biomarkers for bladder cancer by Gomez *et al* [160]. Others were discovered as involved in the complement and coagulation cascade [161] or else in the response and resistance to therapy of breast cancer [162]. Proteins highlighted in figure 5.3a are associated with damage and remodelling in RBC of COVID-19 patients [163].

5.2.4. Differential analysis on peptides

The first analysis consists in comparing the two MELD separately to obtain an overview of the most encountered modifications in function of the group: disease or control. By comparing the number of PSMs for each modification type in each group for the RBC and the plasma samples, similar number of PSM were observed in each group for all detected modifications. This global search did not allow to highlight modification differences from the disease or the

control group from either the RBC or the plasma samples. A deeper analysis is then needed to eventually observe differences between the groups.

The lists of modified peptides obtained from the MELD and tryptic digest analyses were filtered to only keep peptides with assured modifications (Ascore = 1000, which means that the confidence of the modification site is 100%). Looking at the number of total modified peptides without duplicate from table 5.5, we observe that it is higher in the tryptic digest than in the MELD for both RBC and plasma samples. Indeed, MELD generates a more diverse set of peptides with varying sizes and abundances, which can result in a narrower dynamic range compared to tryptic digestion and may lead to the identification of fewer diverse sequences, primarily consisting of highly abundant proteins. With reduced sequence diversity, the probability of identifying modifications is also lowered. While we can achieve better coverage of major proteins, our analysis is less in-depth with the MELD. Even though the number of identified modifications is lower in the MELD, they are more repeated. Indeed, if we compare the number of total modifications with the total modifications without duplicate, we observe a 3.49 and 1.93 fold decrease in the MELD and a 2.38 and a 1.62 fold decrease with the tryptic digest respectively for RBC and plasma samples, meaning that a unique modification site is observed on several overlapping peptides in the MELD and to a lesser extent in the tryptic digest. In definitive, MELD allows us to say that the detection of modified peptides is robust.

Table 5.5: *Number of modified peptides detected in the MELD and in the tryptic digest in the RBC and plasma samples.*

		Number of total modified peptides	Number of total modified peptides without duplicate	Number of peptides common to MELD or TD without duplicate
RBC	Tryptic digest	3941	1657	460
	MELD	4384	1256	
Plasma	Tryptic digest	3080	1895	427
	MELD	2086	1078	

Taking advantage of the quantification obtained for each sample thanks to PEAKS, all modified peptides highlighted in the comparison of the MELD and the tryptic digest were label-free quantitated. A total intensity (sum) normalization was necessary to make samples more comparable between the different runs. The intensities of all the detected peptides in the different samples were summed up samples by samples and a factor of normalization was obtained for each sample by dividing the sum of a sample by the sum of another sample (always the same) taken arbitrarily as reference. This normalization factor was then used to correct all the intensities of each modified peptide highlighted by the method. A statistical analysis of this list of peptides was therefore made using Perseus. From this analysis, 7 peptides were found differentially expressed in the RBC samples and 11 in the plasma samples (see tables 5.6 and 5.7). Proteins from which these peptides originate were checked for their eventual overexpression in the disease or the control group. Indeed, a modified peptide arising from a protein which is overexpressed in a group may be differentially expressed because its

protein is. In the case of the highlighted peptides, none of them came from a protein already overexpressed in a group, therefore being well an overexpression that appears at the peptide level and not because the protein itself is differentially expressed. Two carbonylations and methionine, histidine and tryptophan oxidations are modified peptides differentially expressed in the RBC samples whereas only methionine oxidation are observed in the plasma samples. The chosen peptides are derived from the most prevalent proteins in both matrices. This is because, since no sample enrichment was performed, modifications on high-abundance proteins are more easily identified than those on less abundant proteins. These results are statistically significant, however, for the development of new epitopes used as cardiovascular biomarkers, calculated fold changes are probably too low.

Table 5.6: *Lists of modified peptides differentially expressed from the RBC samples.*

Peptide	p-value	overexpressed in CVD	overexpressed in CONTROL	Accession	PTM
GTFATLSELH(+15.99)CDK	1.65E-03	2.14		P68871|HBB	Oxidation (HW)
VIHDNFGIVEGLM(+15.99)TTVHAITATQK	2.87E-03	1.53		P04406|G3P	Oxidation (M)
EFTPQM(+31.99)QAAYQK	3.39E-03	1.71		P02042|HBD	Dioxidation (M)
SAVTALW(+15.99)GKVNVDEVGGEALGR	7.97E-03	1.46		P68871|HBB	Oxidation (HW)
LLGNVLVC(+57.02)VLAHHFGKEFT(-2.02)PPVQAAYQK	2.40E-02		1.35	P68871|HBB	Carbamidomethylation;threonine to 2-amino-3-ketobutyric acid
DLYANTVLSGGTTM(+15.99)YPGIADR	3.17E-02	1.81		P60709|ACTB	Oxidation (M)
VVAGVANAL(+13.98)AHKYH	3.18E-02	1.32		P68871|HBB	Leucine carbonyl

Table 5.7: *Lists of modified peptides differentially expressed from the plasma samples.*

Peptide	p-value	overexpressed in CVD	overexpressed in CONTROL	Accession	PTM
SHCIAEVENDEM(+31.99)PADLPSLAADFVESK	7.37E-05		1.75	P02768|ALBU	Dioxidation (M)
VM(+15.99)PICLPSKDYAEVGR	3.76E-03	2.20		P00738|HPT	Oxidation (M)
DYFM(+15.99)PCPGR	4.97E-03	2.48		P02790|HEMO	Oxidation (M)
TVM(+15.99)VNIENPEGIPVK	5.80E-03	1.62		P01024|CO3	Oxidation (M)
YVM(+15.99)LPVADQDQC(+57.02)IR	6.33E-03	2.24		P00738|HPT	Oxidation (M);Carbamidomethylation
M(+15.99)ELERPGGNEITR	7.04E-03	1.59		P02671|FIBA	Oxidation (M)
DIFTGLIGPM(+15.99)K	1.72E-02	1.59		P00450|CERU	Oxidation (M)
IVSSAM(+15.99)EPDR	2.23E-02	1.30		P08603|CFAH	Oxidation (M)
LVRPEVDVM(+15.99)C(+57.02)TAFHDNEETFLK	2.59E-02	1.30		P02768|ALBU	Oxidation (M);Carbamidomethylation
VTM(+15.99)QNLNDR	3.85E-02	1.64			Oxidation (M)
M(+15.99)KPVPDLVPGNFK	4.17E-02	1.27		P02671|FIBA	Oxidation (M)

5.2.5. Conclusions

In this chapter, we presented a new strategy for quantifying post-translationally modified peptides that combines the advantages of the MELD innovative approach with the quantification obtained through tryptic digestion. The MELD provides improved sequence coverage and characterization accuracy in comparison to other monoenzymatic workflow. The tryptic digestion provides reproducible and consistent results used for quantification purposes. This synergistic combination resulted in reliable and confident quantification of modified peptides in a fast and high-performance method. This combination of approaches allowed us to develop a method that is able to highlight biomarkers with PTMs.

Using this approach, we identified modified peptides with differential expression in both plasma and RBC samples. The modifications we considered included oxidative stress-related

modifications like nitration and common carbonylation, as well as oxidation of methionine, tryptophan, and histidine. Among the highlighted peptides in the RBC samples, VIHDNFGIVEGLM(+15.99)TTVHAITATQK and EFTPQM(+31.99)QAAYQK were previously identified in a diabetes patient cohort (although not presented in this study), so re-detecting them in this cohort is an important finding. Concerning the plasma samples, the peptide LVRPEVDVM(+31.99)C(+57.02)TAFHDNEETFLK was previously highlighted in a study by Suzuki *et al.* [164] as highly oxidized in patients with diabetes and renal failure. This is a promising result and suggests the potential for this peptide to be a clinical biomarker.

One of the main limitations of this method is the complexity of the mixtures generated by the MELD technique, which uses multiple enzymes to create a wide variety of peptides. The combination of low enzyme concentration with low incubation time leads to an increase in missed cleavage events, making data treatment difficult and resulting in the loss of valuable information. However, the high volume of sequence redundancies increases the level of reliability.

Another limitation is our database search restricted to known modifications of interest. Using the dependent peptide search on our samples could lead to the discovery of new *in vitro* modifications that could be beneficial as potential biomarkers.

Moreover, it is worth emphasizing that relative quantification has a significant drawback, as it relies on a value relative to another and lacks the use of quantification standards. Thus, the error in the value of a protein from one group and the error in the value of the same protein from the other group are unknown. Unlike absolute quantification, there is no standard to precisely determine the value. Our new method enables the detection of biological variation with only a single replicate, without any possibility of calculating errors in the values. Employing technical replicates of the same samples would improve the statistical significance of the results.

To summarize, the approach proposed in this study for quantifying post-translationally modified peptides appears to be a promising method to improve PTM characterization. The future directions of this work include conducting detailed analyses of the identified modified peptides to determine how the modified amino acids are situated in the protein structure and how they impact the protein's folding and activity. Additionally, there is a need to conduct a more in-depth dependent peptide analysis to identify new modifications and to leverage bioinformatics tools to extract further information from the analysis.

6

Conclusion and perspectives

In conclusion, this research has contributed to the development of an innovative toolkit for the quantification of proteins, as well as specific assays for the analysis of post-translational modifications (PTMs) in large cohorts. By advancing the standardization of protein quantification and enabling the analysis of PTMs at scale, this work has the potential to impact the validation of biomarkers in clinical applications.

6.1. Summary

The research was divided in two parts, one per objective: set up an innovative standardization methodology for the quantification of biomarkers and develop a new strategy to analyse post-translationally modified proteins. A third part was also included in the document, namely the methionine oxidation chapter, to cover a specific bias identified during the development of the standardization kit.

6.1.1. Standard quantification

The efforts of this research first focused on the absolute quantification of protein with the development of a new kit. All along the biomarker development process, spanning from the discovery phase to the validation phase, the kit allows enhancements of the protocol and standardization processes across various laboratories, equipments and operators. To our knowledge, no other laboratory has developed a strategy that encompasses all possible biases in a single proteomic experiment.

The newly developed kit serves as a complementary method to the conventional spiked labelled peptide technique used for protein quantification in a given sample. While the spiked peptide approach can be effective, it is limited by the possibility of loss of material during the sample preparation process. Despite attempts to minimize the duration of the preparation process, multiple steps such as handling, collection, storage, purification, reduction, alkylation, digestion, C18 tip purification, and final LC/MS injection are still necessary. Unfortunately, each of these stages can potentially result in experimental variations, leading to significant deviations between the calculated protein quantity and its actual value. As such, all of these stages can introduce potential biases into the quantification process.

Experimental variations in the studied sample can result from several factors. For instance, during the collection, storage, and handling of a sample, transfers to different vials can lead to adsorption of material onto the container's surface, resulting in losses of analytes, proteins, and peptides throughout the procedure. In addition, non-standardized pipetting can introduce bias into the results. Incomplete digestion can also affect the quantitated result by leading to the loss of missed-cleaved peptides that are thus not accounted for quantification. Incomplete reactions during purification, reduction, and alkylation can also contribute to experimental variations in proteomic workflows. Another factor is the liquid chromatographic separation, which can lead to chemical modification of peptides during ionization or LC separation, resulting in the partial loss of their native forms and potential errors in quantification. All these sources of errors are considered by our kit. It allows their detection and especially their precise localization.

Two primary features of the kit can be highlighted:

1. the kit's capability to target the experimental variations of any proteomic workflows;
2. the ability to compare quantification obtained from MS-based experiments and the one obtained from the clinical assay thanks to the endogenous protein.

The kit was developed as follows: four levels of peptide are integrated in the experiment workflow. All levels provide insight in the variations of the experiment. The 1^{st} level of peptides introduced at the beginning of the sample process considers the purification of the sample and the whole digestion procedure. The 2^{nd} level of peptide helps in elucidating the digestion efficiency (only considering the amino acids succession) while the 3^{rd} level evaluates the loss around the digestion procedure: degradation and adsorption for example. Finally, the 4^{th} level of peptides acts as a reference peptide set and enables the quantification of the three first levels.

Additionally, the elements of the kit can serve as a quality control tool for the laboratory instrumentations. A methionine-containing peptide allows to measure the degree of wear of the columns. We showed that methionine-containing peptides were characterized by a broad chromatographic peak on old column setups, having an intermediate retention time between

the native and the oxidized peptide. This observation was understood as the peptide being oxidized inside the analytical column. This oxidation occurring in the column, deteriorates the analysis. In summary, by regularly analysing the chromatographic profiles of a methionine-containing peptide standard, researchers can identify any changes in the peak shape or intensity that indicates column deterioration.

6.1.2. Differential post-translational modifications

A new strategy for quantifying post-translationally modified peptides that combines the advantages of the MELD technique with tryptic digestion was presented in chapter 5. The MELD approach provides improved sequence coverage and characterization accuracy, while tryptic digestion provides reproducible and consistent results for quantification purposes. This combination resulted in a reliable and fast method for quantifying modified peptides, which allowed us to identify modified peptides with differential expression in both plasma and RBC samples.

We highlighted modified peptides in the RBC samples that were identified in a diabetes patient cohort (previously studied in our laboratory), suggesting their potential as biomarkers. In the plasma samples, we also identified a highly oxidized peptide that has been previously linked to diabetes and renal failure as well. While we identified fewer modified peptides using the MELD method compared to tryptic digestion, the number of unique modifications observed on different overlapping peptides is greater in the MELD approach than in the tryptic digestion for both plasma and RBC samples. This demonstrates how the MELD method can effectively detect multiple sites susceptible to modification in a robust fashion.

In summary, this study contributes to the progress of PTM research and provides a foundation for further exploration of the identified modified peptides using the new method proposed.

6.2. Significance

The significance of this thesis lies in the development of a new kit that enables absolute quantification of proteins, which is essential for identifying potential biomarkers with improved confidence. The kit addresses the various biases that can arise during the sample preparation process, which can significantly impact the accuracy of the quantification results. The approach considers all possible biases in a single proteomic experiment, which is a significant advancement. The kit allows for the detection and localization of errors, enabling researchers to improve the protocol and minimize losses.

Furthermore, the kit can serve as a quality control tool for laboratory instrumentations, enabling researchers to identify any changes in peak shape or intensity that indicate column

deterioration. The methodology, which involves the integration of four levels of peptides, provides insight into the variations of the experiment. That allows researchers to evaluate the loss around the digestion procedure, elucidate the digestion efficiency, and consider the purification of the sample and the whole digestion procedure. Overall, the development of this new kit can significantly improve the accuracy and reliability of protein quantification, which is critical for advancing proteomic research and identifying potential biomarkers for various diseases.

The development of a new strategy for quantifying post-translationally modified peptides is significant because it allows for the robust identification of potential biomarkers. The combination of the MELD technique and tryptic digestion provides improved sequence coverage and characterization accuracy, while also ensuring reproducible and consistent results for quantification purposes. This new method allowed to highlight promising modified peptides for use as potential biomarkers.

6.3. Limitations & perspectives

There are still some limitations and room for improvement concerning the quantification kit. Currently, it remains a challenge to determine whether the protein structure has an impact on digestion, or if the peptide itself is difficult to cleave or has flanking sequences that affect the digestion process. A possible way to improve the kit would be to design a protein with similar peptides in both the structured and disordered regions. If we observe differences between peptides originating from these two domains, we could conclude that the protein structure is the main factor contributing to these differences, since the cleaved peptides have the same sequences and flanking sequences.

We also acknowledge the limitations of the method introduced in the PTMs chapter: the complexity of the mixtures generated by the MELD technique and the reliance on a database search restricted to known modifications of interest.

The approach for quantifying post-translational modified peptides appears to be a promising method to improve PTMs characterization. Future directions include conducting detailed analyses of the identified modified peptides to determine their impact on protein structure and activity, conducting a more in-depth dependent peptide analysis to identify new modifications, and leveraging bioinformatics tools to extract further information from the analysis. Improving the statistical significance of the results by doing the experiments in triplicates will be a good improvement. Overall, this study provides valuable insights into the quantification of modified peptides and highlights their potential as clinical biomarkers.

The upcoming efforts will focus on deploying the kit in an actual cohort and verifying its reproducibility and consistency of outcomes across diverse laboratories and experiments, as

well as different MS platforms and operators. Based on the results, specifications for the kit could be established, and guidelines for its usage will be developed. Performing the validation of the kit on the cohort mentioned in chapter 5 will create a mutually beneficial scenario, as it would allow the kit to be tested on a genuine cohort and enable the accurate quantification of the modified proteins we identified.

The comprehensive examination of methionine-containing peptides has revealed that on-column oxidation of methionine residues increases when low amounts of samples are injected. This modification is expected to become more significant with the development of new technologies like single-cell proteomics and instruments that use smaller sample sizes. As a result, it will be crucial to rigorously evaluate the LC system using quality controls. Efforts should be invested for the definition of a threshold beyond which the column should be replaced. On-column modifications could potentially impact other types of modifications as well, and further investigations will be necessary to determine the extent of this phenomenon on other modifications.

In conclusion, this report highlighted two techniques for peptide quantification, one of which has been developed into a scientific kit. We also provided recommendations for monitoring and mitigating laboratory instrumentation degradation in chapter 4. Although limitations exist, we hope that this research will pave the way for future advancements in the field of proteins quantification.

6.4. Personal reflection

Reflecting on my researcher journey, it was not a long and quiet river. I faced the challenge of working on two different projects which turned out at the end to be an opportunity. It pushed my scientific mindset to its limits by making bridges where the river was sometimes quite wide.

I am particularly proud about my contribution to the discovery of methionine oxidation in LC columns. It emphasized my scientist mission: exploring and discovering. I was particularly glad and satisfied when I presented these findings to conferences and when peers were intrigued and even surprised by my discovery. I felt my work useful.

At a larger scale, I firmly hope that this work will contribute to progress in medical applications. This is definitely a small step, but it invites others!

In a way, this journey stops here, after six years of university life. But I end this journey with a hold full of new skills and knowledge which will definitely accompany me in many other journeys. I am grateful to those, promoters, colleagues and friends who made me grow.

References

[1] Arthur J. Atkinson et al. “Biomarkers and Surrogate Endpoints: Preferred Definitions and Conceptual Framework”. *Clinical Pharmacology & Therapeutics* vol. 69, no. 3 (2001), pp. 89–95 (cit. on p. 4).

[2] Xiaochen Xi et al. “RNA Biomarkers: Frontier of Precision Medicine for Cancer”. *Non-Coding RNA* vol. 3, no. 1 (Feb. 2017), p. 9 (cit. on p. 4).

[3] Elisabeth Drucker and Kurt Krapfenbauer. “Pitfalls and Limitations in Translation from Biomarker Discovery to Clinical Utility in Predictive and Personalised Medicine”. *EPMA Journal* vol. 4, no. 1 (Feb. 2013), p. 7 (cit. on p. 7).

[4] Werner Zolg. “The Proteomic Search for Diagnostic Biomarkers: Lost in Translation? *”. *Molecular & Cellular Proteomics* vol. 5, no. 10 (Oct. 2006), pp. 1720–1726 (cit. on pp. 7, 8).

[5] Nader Rifai, Michael A. Gillette, and Steven A. Carr. “Protein Biomarker Discovery and Validation: The Long and Uncertain Path to Clinical Utility”. *Nature Biotechnology* vol. 24, no. 8 (Aug. 2006), pp. 971–983 (cit. on pp. 7, 22).

[6] Cong Zhou et al. “Statistical Considerations of Optimal Study Design for Human Plasma Proteomics and Biomarker Discovery”. *Journal of Proteome Research* vol. 11, no. 4 (Apr. 2012), pp. 2103–2113 (cit. on p. 7).

[7] Yoshiyuki Suehara, Daisuke Kubota, and Tsuyoshi Saito. “Tissue Sample Preparation for Biomarker Discovery”. *Proteomics for Biomarker Discovery*. Ed. by Ming Zhou and Timothy Veenstra. Methods in Molecular Biology. Totowa, NJ: Humana Press, 2013, pp. 13–23 (cit. on p. 8).

[8] Angel P. Diz, Manuela Truebano, and David O. F. Skibinski. “The Consequences of Sample Pooling in Proteomics: An Empirical Study”. *Electrophoresis* vol. 30, no. 17 (Sept. 2009), pp. 2967–2975 (cit. on p. 8).

[9] Marcus Bantscheff et al. “Quantitative Mass Spectrometry in Proteomics: A Critical Review”. *Analytical and Bioanalytical Chemistry* vol. 389, no. 4 (Aug. 2007), pp. 1017–1031 (cit. on p. 8).

[10] Marcus Bantscheff et al. “Quantitative Mass Spectrometry in Proteomics: Critical Review Update from 2007 to the Present”. *Analytical and Bioanalytical Chemistry* vol. 404, no. 4 (July 2012), pp. 939–965 (cit. on pp. 8, 14).

[11] Veronika Vidova and Zdenek Spacil. “A Review on Mass Spectrometry-Based Quantitative Proteomics: Targeted and Data Independent Acquisition”. *Analytica Chimica Acta* vol. 964 (Apr. 2017), pp. 7–23 (cit. on p. 8).

[12] Svitlana Rozanova et al. “Quantitative Mass Spectrometry-Based Proteomics: An Overview”. *Methods in Molecular Biology (Clifton, N.J.)* vol. 2228 (2021), pp. 85–116 (cit. on p. 8).

[13] Mirjam van Bentum and Matthias Selbach. “An Introduction to Advanced Targeted Acquisition Methods”. *Molecular & Cellular Proteomics* vol. 20 (Jan. 2021) (cit. on pp. 9, 17).

[14] Claire E. Eyers and Simon J. Gaskell. *Quantitative Proteomics*. Royal Society of Chemistry, 2014 (cit. on p. 9).

[15] John B. Fenn. “Ion Formation from Charged Droplets: Roles of Geometry, Energy, and Time”. *Journal of the American Society for Mass Spectrometry* vol. 4, no. 7 (July 1993), pp. 524–535 (cit. on p. 9).

[16] Shao-En Ong and Matthias Mann. “Mass Spectrometry–Based Proteomics Turns Quantitative”. *Nature Chemical Biology* vol. 1, no. 5 (Oct. 2005), pp. 252–262 (cit. on p. 11).

[17] Shao-En Ong et al. “Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics”. *Molecular & cellular proteomics: MCP* vol. 1, no. 5 (May 2002), pp. 376–386 (cit. on p. 12).

[18] O. A. Mirgorodskaya et al. “Quantitation of Peptides and Proteins by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Using (18)O-labeled Internal Standards”. *Rapid communications in mass spectrometry: RCM* vol. 14, no. 14 (2000), pp. 1226–1232 (cit. on p. 12).

[19] X. Yao et al. “Proteolytic 18O Labeling for Comparative Proteomics: Model Studies with Two Serotypes of Adenovirus”. *Analytical Chemistry* vol. 73, no. 13 (July 2001), pp. 2836–2842 (cit. on p. 12).

[20] Xiaoying Ye et al. “18O Stable Isotope Labeling in MS-based Proteomics”. *Briefings in Functional Genomics and Proteomics* vol. 8, no. 2 (Mar. 2009), pp. 136–144 (cit. on p. 12).

[21] Steven P. Gygi et al. “Quantitative Analysis of Complex Protein Mixtures Using Isotope-Coded Affinity Tags”. *Nature Biotechnology* vol. 17, no. 10 (Oct. 1999), pp. 994–999 (cit. on p. 13).

[22] Kirk C. Hansen et al. “Mass Spectrometric Analysis of Protein Mixtures at Low Levels Using Cleavable 13C-isotope-coded Affinity Tag and Multidimensional Chromatography”. *Molecular & cellular proteomics: MCP* vol. 2, no. 5 (May 2003), pp. 299–314 (cit. on p. 13).

[23] Jiaxu Li, Hanno Steen, and Steven P. Gygi. “Protein Profiling with Cleavable Isotope-coded Affinity Tag (cICAT) Reagents: The Yeast Salinity Stress Response*S”. *Molecular & Cellular Proteomics* vol. 2, no. 11 (Nov. 2003), pp. 1198–1204 (cit. on p. 13).

[24] Philip L. Ross et al. “Multiplexed Protein Quantitation in Saccharomyces Cerevisiae Using Amine-reactive Isobaric Tagging Reagents”. *Molecular & Cellular Proteomics* vol. 3, no. 12 (Jan. 2004), pp. 1154–1169 (cit. on p. 13).

[25] Un-Beom Kang et al. “Quantitative Analysis of mTRAQ-Labeled Proteome Using Full MS Scans”. *Journal of Proteome Research* vol. 9, no. 7 (July 2010), pp. 3750–3758 (cit. on p. 13).

[26] Andrew Thompson et al. “Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS”. *Analytical Chemistry* vol. 75, no. 8 (Apr. 2003), pp. 1895–1904 (cit. on p. 14).

[27] Sushma Anand et al. “Label-Based and Label-Free Strategies for Protein Quantitation”. *Proteome Bioinformatics*. Ed. by Shivakumar Keerthikumar and Suresh Mathivanan. Methods in Molecular Biology. New York, NY: Springer, 2017, pp. 31–43 (cit. on p. 14).

[28] Karlie A. Neilson et al. “Less Label, More Free: Approaches in Label-Free Quantitative Mass Spectrometry”. *PROTEOMICS* vol. 11, no. 4 (2011), pp. 535–553 (cit. on p. 15).

[29] Jian Guo and Tao Huan. “Comparison of Full-Scan, Data-Dependent, and Data-Independent Acquisition Modes in Liquid Chromatography–Mass Spectrometry Based Untargeted Metabolomics”. *Analytical Chemistry* vol. 92, no. 12 (June 2020), pp. 8072–8080 (cit. on p. 15).

[30] Lindsay K. Pino et al. “Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries”. *Molecular & Cellular Proteomics* vol. 19, no. 7 (July 2020), pp. 1088–1103 (cit. on pp. 16, 17).

[31] Matthew Y. Lim, João A. Paulo, and Steven P. Gygi. “Evaluating False Transfer Rates from the Match-between-Runs Algorithm with a Two-Proteome Model”. *Journal of proteome research* vol. 18, no. 11 (Nov. 2019), pp. 4020–4026 (cit. on p. 16).

[32] Christina Ludwig et al. “Data-Independent Acquisition-Based SWATH-MS for Quantitative Proteomics: A Tutorial”. *Molecular Systems Biology* vol. 14, no. 8 (Aug. 2018), e8126 (cit. on pp. 16, 17).

[33] Vadim Demichev et al. “DIA-NN: Neural Networks and Interference Correction Enable Deep Proteome Coverage in High Throughput”. *Nature methods* vol. 17, no. 1 (Jan. 2020), pp. 41–44 (cit. on p. 17).

[34] Robert S. Plumb et al. “UPLC/MS(E); a New Approach for Generating Molecular Fragment Information for Biomarker Structure Elucidation”. *Rapid communications in mass spectrometry: RCM* vol. 20, no. 13 (2006), pp. 1989–1994 (cit. on p. 17).

[35] Jeffrey C. Silva et al. “Absolute Quantification of Proteins by LCMSE A Virtue of Parallel Ms Acquisition”. *Molecular & Cellular Proteomics* vol. 5, no. 1 (Jan. 2006), pp. 144–156 (cit. on p. 17).

[36] Pavel V. Shliaha et al. “Effects of Traveling Wave Ion Mobility Separation on Data Independent Acquisition in Proteomics Studies”. *Journal of Proteome Research* vol. 12, no. 6 (June 2013), pp. 2323–2339 (cit. on p. 17).

[37] Bruno Domon and Ruedi Aebersold. “Options and Considerations When Selecting a Quantitative Proteomics Strategy”. *Nature Biotechnology* vol. 28, no. 7 (July 2010), pp. 710–721 (cit. on pp. 17, 19).

[38] Zhaojing Meng and Timothy D. Veenstra. “Targeted Mass Spectrometry Approaches for Protein Biomarker Verification”. *Journal of Proteomics*. Pharmacoproteomics and Toxicoproteomics vol. 74, no. 12 (Nov. 2011), pp. 2650–2659 (cit. on p. 17).

[39] Edmond de Hoffmann and Vincent Stroobant. *Mass Spectrometry: Principles and Applications*. John Wiley & Sons, Apr. 2013 (cit. on p. 17).

[40] Paola Picotti and Ruedi Aebersold. “Selected Reaction Monitoring-Based Proteomics: Workflows, Potential, Pitfalls and Future Directions”. *Nature Methods* vol. 9, no. 6 (June 2012), pp. 555–566 (cit. on pp. 17, 19).

[41] Vinzenz Lange et al. “Selected Reaction Monitoring for Quantitative Proteomics: A Tutorial”. *Molecular Systems Biology* vol. 4 (Oct. 2008), p. 222 (cit. on p. 17).

[42] Jonathan D. Worboys et al. “Systematic Evaluation of Quantotypic Peptides for Targeted Analysis of the Human Kinome”. *Nature Methods* vol. 11, no. 10 (Oct. 2014), pp. 1041–1044 (cit. on p. 19).

[43] Philip Brownridge et al. “Global Absolute Quantification of a Proteome: Challenges in the Deployment of a QconCAT Strategy”. *PROTEOMICS* vol. 11, no. 15 (2011), pp. 2957–2970 (cit. on p. 19).

[44] Amelia C. Peterson et al. “Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics”. *Molecular & cellular proteomics: MCP* vol. 11, no. 11 (Nov. 2012), pp. 1475–1488 (cit. on p. 19).

[45] Adele Bourmaud, Sebastien Gallien, and Bruno Domon. “Parallel Reaction Monitoring Using Quadrupole-Orbitrap Mass Spectrometer: Principle and Applications”. *PROTEOMICS* vol. 16, no. 15-16 (2016), pp. 2146–2159 (cit. on p. 19).

[46] Scott A. Gerber et al. “Absolute Quantification of Proteins and Phosphoproteins from Cell Lysates by Tandem MS”. *Proceedings of the National Academy of Sciences of the United States of America* vol. 100, no. 12 (June 2003), pp. 6940–6945 (cit. on p. 19).

[47] Robert J. Beynon et al. “Multiplexed Absolute Quantification in Proteomics Using Artificial QCAT Proteins of Concatenated Signature Peptides”. *Nature Methods* vol. 2, no. 8 (Aug. 2005), pp. 587–589 (cit. on p. 20).

[48] Keiji Kito et al. “A Synthetic Protein Approach toward Accurate Mass Spectrometric Quantification of Component Stoichiometry of Multiprotein Complexes”. *Journal of Proteome Research* vol. 6, no. 2 (Feb. 2007), pp. 792–800 (cit. on pp. 20, 26).

[49] Leigh Anderson and Christie L. Hunter. “Quantitative Mass Spectrometric Multiple Reaction Monitoring Assays for Major Plasma Proteins”. *Molecular & Cellular Proteomics* vol. 5, no. 4 (Jan. 2006), pp. 573–588 (cit. on p. 20).

[50] Marlis Zeiler et al. “A Protein Epitope Signature Tag (PrEST) Library Allows SILAC-based Absolute Quantification and Multiplexed Determination of Protein Copy Numbers in Cell Lines”. *Molecular & cellular proteomics: MCP* vol. 11, no. 3 (Mar. 2012), O111.009613 (cit. on p. 20).

[51] Tove Boström. “QPrEST™—Isotope-Labeled Multipeptide Standards for Quantitative Mass Spectrometry–Based Proteomics”. *Nature Methods* vol. 13, no. 3 (Mar. 2016), pp. iv–vi (cit. on p. 20).

[52] Virginie Brun et al. “Isotope-Labeled Protein Standards Toward Absolute Quantitative Proteomics”. *Molecular & Cellular Proteomics* vol. 6, no. 12 (Jan. 2007), pp. 2139–2149 (cit. on pp. 21, 26).

[53] Michel Jaquinod et al. “Mass Spectrometry-Based Absolute Protein Quantification: PSAQ™ Strategy Makes Use of “Noncanonical” Proteotypic Peptides”. *PROTEOMICS* vol. 12, no. 8 (2012), pp. 1217–1221 (cit. on p. 21).

[54] Sasha Singh et al. “FLEXIQuant: A Novel Tool for the Absolute Quantification of Proteins, and the Simultaneous Identification and Quantification of Potentially Modified Peptides”. *Journal of Proteome Research* vol. 8, no. 5 (May 2009), pp. 2201–2210 (cit. on p. 21).

[55] Nico Zinn, Dominic Winter, and Wolf D. Lehmann. “Recombinant Isotope Labeled and Selenium Quantified Proteins for Absolute Protein Quantification”. *Analytical Chemistry* vol. 82, no. 6 (Mar. 2010), pp. 2334–2340 (cit. on p. 22).

[56] Mandy Alhajj and Aisha Farhana. “Enzyme Linked Immunosorbent Assay”. *StatPearls*. Treasure Island (FL): StatPearls Publishing, 2022 (cit. on p. 22).

[57] Patrick J. Tighe et al. “ELISA in the Multiplex Era: Potentials and Pitfalls”. *Proteomics. Clinical Applications* vol. 9, no. 3-4 (Apr. 2015), pp. 406–422 (cit. on p. 22).

[58] B. Sharan Sharma et al. “Post-Translational Modifications (PTMs), from a Cancer Perspective: An Overview”. *Oncogen* vol. 2, no. 3 (May 2019), pp. 1–11 (cit. on p. 23).

[59] Katie Dunphy et al. “Current Methods of Post-Translational Modification Analysis and Their Applications in Blood Cancers”. *Cancers* vol. 13, no. 8 (Apr. 2021), p. 1930 (cit. on p. 23).

[60] José M. Souza, Gonzalo Peluffo, and Rafael Radi. “Protein Tyrosine Nitration—Functional Alteration or Just a Biomarker?” *Free Radical Biology and Medicine* vol. 45, no. 4 (Aug. 2008), pp. 357–366 (cit. on p. 23).

[61] Dorte B. Bekker-Jensen et al. “Rapid and Site-Specific Deep Phosphoproteome Profiling by Data-Independent Acquisition without the Need for Spectral Libraries”. *Nature Communications* vol. 11, no. 1 (Feb. 2020), p. 787 (cit. on p. 24).

[62] Yi Yang and Liang Qiao. “Data-Independent Acquisition Proteomics Methods for Analyzing Post-Translational Modifications”. *PROTEOMICS* vol. n/a, no. n/a (), p. 2200046 (cit. on p. 24).

[63] Wenxuan Cai et al. “Top-down Proteomics: Technology Advancements and Applications to Heart Diseases”. *Expert review of proteomics* vol. 13, no. 8 (Aug. 2016), pp. 717–730 (cit. on p. 24).

[64] Han Zhang and Ying Ge. “Comprehensive Analysis of Protein Modifications by Top-down Mass Spectrometry”. *Circulation. Cardiovascular Genetics* vol. 4, no. 6 (Dec. 2011), p. 711 (cit. on p. 24).

[65] Joel M. Chick et al. “An Ultra-Tolerant Database Search Reveals That a Myriad of Modified Peptides Contributes to Unassigned Spectra in Shotgun Proteomics”. *Nature biotechnology* vol. 33, no. 7 (July 2015), pp. 743–749 (cit. on p. 24).

[66] Mikhail M. Savitski, Michael L. Nielsen, and Roman A. Zubarev. “ModifiComb, a New Proteomic Tool for Mapping Substoichiometric Post-translational Modifications, Finding Novel Types of Modifications, and Fingerprinting Complex Protein Mixtures*”. *Molecular & Cellular Proteomics* vol. 5, no. 5 (May 2006), pp. 935–948 (cit. on p. 24).

[67] Jennifer A. Siepen et al. “Prediction of Missed Cleavage Sites in Tryptic Peptides Aids Protein Identification in Proteomics”. *Journal of Proteome Research* vol. 6, no. 1 (Jan. 2007), pp. 399–408 (cit. on p. 26).

[68] Chia-Yu Yen et al. “Improving Sensitivity in Shotgun Proteomics Using a Peptide-Centric Database with Reduced Complexity: Protease Cleavage and SCX Elution Rules from Data Mining of MS/MS Spectra”. *Analytical Chemistry* vol. 78, no. 4 (Feb. 2006), pp. 1071–1084 (cit. on p. 26).

[69] Crystal S. F. Cheung et al. “Natural Flanking Sequences for Peptides Included in a Quantification Concatamer Internal Standard”. *Analytical Chemistry* vol. 87, no. 2 (Jan. 2015), pp. 1097–1102 (cit. on pp. 26, 32).

[70] Elien Vandermarliere, Michael Mueller, and Lennart Martens. “Getting Intimate with Trypsin, the Leading Protease in Proteomics”. *Mass Spectrometry Reviews* vol. 32, no. 6 (2013), pp. 453–465 (cit. on p. 26).

[71] Dhaval Nanavati et al. “Stoichiometry and Absolute Quantification of Proteins with Mass Spectrometry Using Fluorescent and Isotope-Labeled Concatenated Peptide Standards”. *Molecular & cellular proteomics: MCP* vol. 7, no. 2 (Feb. 2008), pp. 442–447 (cit. on p. 26).

[72] Julia Maria Burkhart et al. “Systematic and Quantitative Comparison of Digest Efficiency and Specificity Reveals the Impact of Trypsin Quality on MS-based Proteomics”. *Journal of Proteomics* vol. 75, no. 4 (Feb. 2012), pp. 1454–1462 (cit. on p. 26).

[73] Jennifer L. Proc et al. “A Quantitative Study of the Effects of Chaotropic Agents, Surfactants, and Solvents on the Digestion Efficiency of Human Plasma Proteins by Trypsin”. *Journal of Proteome Research* vol. 9, no. 10 (Oct. 2010), pp. 5422–5437 (cit. on p. 26).

[74] Alexander R. Ivanov et al. “Interlaboratory Studies and Initiatives Developing Standards for Proteomics”. *PROTEOMICS* vol. 13, no. 6 (Mar. 2013), pp. 904–909 (cit. on p. 26).

[75] Julia Maria Burkhart, Thomas Premsler, and Albert Sickmann. “Quality Control of Nano-LC-MS Systems Using Stable Isotope-Coded Peptides”. *PROTEOMICS* vol. 11, no. 6 (Mar. 2011), pp. 1049–1057 (cit. on p. 26).

[76] Claudia Escher et al. “Using iRT, a Normalized Retention Time for More Targeted Measurement of Peptides”. *PROTEOMICS* vol. 12, no. 8 (Apr. 2012), pp. 1111–1121 (cit. on p. 27).

[77] Wout Bittremieux et al. “Quality Control in Mass Spectrometry-Based Proteomics”. *Mass Spectrometry Reviews* vol. 37, no. 5 (2018), pp. 697–711 (cit. on p. 27).

[78] Daniel Ruderman. “Designing Successful Proteomics Experiments”. *Proteomics: Methods and Protocols*. Ed. by Lucio Comai, Jonathan E. Katz, and Parag Mallick. Methods in Molecular Biology. New York, NY: Springer, 2017, pp. 271–288 (cit. on p. 27).

[79] Teresa Mendes Maia et al. “Simple Peptide Quantification Approach for MS-Based Proteomics Quality Control”. *ACS Omega* vol. 5, no. 12 (Mar. 2020), pp. 6754–6762 (cit. on p. 27).

[80] Matthew W. Vetting et al. “Experimental Strategies for Functional Annotation and Metabolism Discovery: Targeted Screening of Solute Binding Proteins and Unbiased Panning of Metabolomes”. *Biochemistry* vol. 54, no. 3 (Jan. 2015), pp. 909–931 (cit. on p. 30).

[81] Marco Necci et al. “MobiDB-lite: Fast and Highly Specific Consensus Prediction of Intrinsic Disorder in Proteins”. *Bioinformatics* vol. 33, no. 9 (May 2017), pp. 1402–1404 (cit. on p. 31).

[82] Eric W Deutsch et al. “The ProteomeXchange Consortium in 2020: Enabling ‘Big Data’ Approaches in Proteomics”. *Nucleic Acids Research* vol. 48, no. D1 (Jan. 2020), pp. D1145–D1152 (cit. on p. 37).

[83] Philip Jones and Richard Côté. “The PRIDE Proteomics Identifications Database: Data Submission, Query, and Dataset Comparison”. *Functional Proteomics: Methods and Protocols*. Ed. by Julie D. Thompson, Marius Ueffing, and Christine Schaeffer-Reiss. Methods in Molecular Biology. Totowa, NJ: Humana Press, 2008, pp. 287–303 (cit. on p. 37).

[84] Jochen M. Schwenk et al. “The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays”. *Journal of Proteome Research* vol. 16, no. 12 (Dec. 2017), pp. 4299–4310 (cit. on p. 37).

[85] Wojtek James Goscinski et al. “The Multi-Modal Australian ScienceS Imaging and Visualization Environment (MASSIVE) High Performance Computing Infrastructure: Applications in Neuroscience and Neuroinformatics Research”. *Frontiers in Neuroinformatics* vol. 8 (2014) (cit. on p. 37).

[86] O. V. Krokhin et al. “An Improved Model for Prediction of Retention Times of Tryptic Peptides in Ion Pair Reversed-phase HPLC Its Application to Protein Peptide Mapping by Off-Line HPLC-MALDI MS”. *Molecular & Cellular Proteomics* vol. 3, no. 9 (Jan. 2004), pp. 908–919 (cit. on p. 38).

[87] Jin-Sam You et al. “The Impact of Blood Contamination on the Proteome of Cerebrospinal Fluid”. *Proteomics* vol. 5, no. 1 (Jan. 2005), pp. 290–296 (cit. on p. 39).

[88] C. Geourjon and G. Deléage. “SOPMA: Significant Improvements in Protein Secondary Structure Prediction by Consensus Prediction from Multiple Alignments”. *Bioinformatics* vol. 11, no. 6 (Dec. 1995), pp. 681–684 (cit. on p. 39).

[89] Peter Y. Chou and Gerald D. Fasman. “Prediction of Protein Conformation”. *Biochemistry* vol. 13, no. 2 (Jan. 1974), pp. 222–245 (cit. on p. 39).

[90] Oleg V. Krokhin. “Sequence-Specific Retention Calculator. Algorithm for Peptide Retention Prediction in Ion-Pair RP-HPLC: Application to 300- and 100-Å Pore Size C18 Sorbents”. *Analytical Chemistry* vol. 78, no. 22 (Nov. 2006), pp. 7785–7795 (cit. on p. 40).

[91] J. W. Crabb et al. “Amino Acid Analysis”. *Current Protocols in Protein Science* vol. Chapter 11 (May 2001), Unit 11.9 (cit. on p. 44).

[92] Brendan MacLean et al. “Skyline: An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments”. *Bioinformatics* vol. 26, no. 7 (Apr. 2010), pp. 966–968 (cit. on p. 45).

[93] Ming Zhou and Li-Rong Yu. “Proteomic Analysis by Two-Dimensional Polyacrylamide Gel Electrophoresis”. *Advances in Protein Chemistry*. Ed. by Richard D. Smith and Timothy D. Veenstra. Vol. 65. Proteome Characterization and Proteomics. Academic Press, Jan. 2003, pp. 57–84 (cit. on p. 46).

[94] Norma J. Greenfield. “Using Circular Dichroism Spectra to Estimate Protein Secondary Structure”. *Nature protocols* vol. 1, no. 6 (2006), pp. 2876–2890 (cit. on p. 49).

[95] N. Sreerama and R. W. Woody. “A Self-Consistent Method for the Analysis of Protein Secondary Structure from Circular Dichroism”. *Analytical Biochemistry* vol. 209, no. 1 (Feb. 1993), pp. 32–44 (cit. on p. 50).

[96] L. A. Compton and W. C. Johnson. “Analysis of Protein Circular Dichroism Spectra for Secondary Structure Using a Simple Matrix Multiplication”. *Analytical Biochemistry* vol. 155, no. 1 (May 1986), pp. 155–167 (cit. on p. 50).

[97] S. W. Provencher and J. Glöckner. “Estimation of Globular Protein Secondary Structure from Circular Dichroism”. *Biochemistry* vol. 20, no. 1 (Jan. 1981), pp. 33–37 (cit. on p. 50).

[98] N. Sreerama and R. W. Woody. “Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set”. *Analytical Biochemistry* vol. 287, no. 2 (Dec. 2000), pp. 252–260 (cit. on p. 50).

[99] Amar B. T. Ghisaidoobe and Sang J. Chung. “Intrinsic Tryptophan Fluorescence in the Detection and Analysis of Proteins: A Focus on Förster Resonance Energy Transfer Techniques”. *International Journal of Molecular Sciences* vol. 15, no. 12 (Dec. 2014), pp. 22518–22538 (cit. on p. 53).

[100] Nadja Hellmann and Dirk Schneider. “Hands On: Using Tryptophan Fluorescence Spectroscopy to Study Protein Structure”. *Protein Supersecondary Structures: Methods and Protocols*. Ed. by Alexander E. Kister. Methods in Molecular Biology. New York, NY: Springer, 2019, pp. 379–401 (cit. on p. 54).

[101] Benjamin C. Orsburn. “Proteome Discoverer—A Community Enhanced Data Processing Suite for Protein Informatics”. *Proteomes* vol. 9, no. 1 (Mar. 2021), p. 15 (cit. on p. 55).

[102] R Core Team. *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/. 2020 (cit. on p. 55).

[103] Craig Lawless and Simon J. Hubbard. “Prediction of Missed Proteolytic Cleavages for the Selection of Surrogate Peptides for Quantitative Proteomics”. *OMICS : a Journal of Integrative Biology* vol. 16, no. 9 (Sept. 2012), pp. 449–456 (cit. on p. 56).

[104] Jesse Rodriguez et al. “Does Trypsin Cut Before Proline?” *Journal of Proteome Research* vol. 7, no. 1 (Jan. 2008), pp. 300–305 (cit. on p. 56).

[105] Dorothée Lebert et al. “DIGESTIF: A Universal Quality Standard for the Control of Bottom-Up Proteomics Experiments”. *Journal of Proteome Research* vol. 14, no. 2 (Feb. 2015), pp. 787–803 (cit. on p. 66).

[106] P. Buckel. “Recombinant Proteins for Therapy”. *Trends in Pharmacological Sciences* vol. 17, no. 12 (Dec. 1996), pp. 450–456 (cit. on p. 69).

[107] Phuc V. Pham. “Chapter 19 - Medical Biotechnology: Techniques and Applications”. *Omics Technologies and Bio-Engineering*. Ed. by Debmalya Barh and Vasco Azevedo. Academic Press, Jan. 2018, pp. 449–469 (cit. on p. 69).

[108] James J Pitt. “Principles and Applications of Liquid Chromatography-Mass Spectrometry in Clinical Biochemistry”. *The Clinical Biochemist Reviews* vol. 30, no. 1 (Feb. 2009), pp. 19–34 (cit. on p. 70).

[109] Weidong Cui, Henry W. Rohrs, and Michael L. Gross. “Top-Down Mass Spectrometry: Recent Developments, Applications and Perspectives”. *The Analyst* vol. 136, no. 19 (Oct. 2011), pp. 3854–3864 (cit. on p. 70).

[110] P. Boomathi Pandeswari and Varatharajan Sabareesh. “Middle-down Approach: A Choice to Sequence and Characterize Proteins/Proteomes by Mass Spectrometry”. *RSC Advances* vol. 9, no. 1 (Dec. 2018), pp. 313–344 (cit. on p. 70).

[111] Ravi Amunugama et al. “Bottom-Up Mass Spectrometry–Based Proteomics as an Investigative Analytical Tool for Discovery and Quantification of Proteins in Biological Samples”. *Advances in Wound Care* vol. 2, no. 9 (Nov. 2013), pp. 549–557 (cit. on p. 70).

[112] Frederik Lermyte et al. “Top or Middle? Up or Down? Toward a Standard Lexicon for Protein Top-Down and Allied Mass Spectrometry Approaches”. *Journal of The American Society for Mass Spectrometry* vol. 30, no. 7 (July 2019), pp. 1149–1157 (cit. on p. 70).

[113] Denis Morsa et al. “Multi-Enzymatic Limited Digestion: The Next-Generation Sequencing for Proteomics?” *Journal of Proteome Research* vol. 18, no. 6 (June 2019), pp. 2501–2513 (cit. on pp. 70, 91).

[114] Kenneth Morand, Gert Talbo, and Matthias Mann. “Oxidation of Peptides during Electrospray Ionization”. *Rapid Communications in Mass Spectrometry* vol. 7, no. 8 (1993), pp. 738–743 (cit. on pp. 70, 74).

[115] Maolian Chen and Kelsey D. Cook. “Oxidation Artifacts in the Electrospray Mass Spectrometry of Abeta Peptide”. *Analytical Chemistry* vol. 79, no. 5 (Mar. 2007), pp. 2031–2036 (cit. on p. 70).

[116] K. Ilker Sen, Robert Hepler, and Hirsh Nanda. “Detection and Measurement of Methionine Oxidation in Proteins”. *Current Protocols in Protein Science* vol. 87, no. 1 (2017), pp. 14.16.1–14.16.11 (cit. on pp. 70, 77).

[117] Pilsoo Kang et al. “Use of a Stable-Isotope-Labeled Reporter Peptide and Antioxidants for Reliable Quantification of Methionine Oxidation in a Monoclonal Antibody by Liquid Chromatography/Mass Spectrometry”. *Rapid Communications in Mass Spectrometry* vol. 30, no. 14 (July 2016), pp. 1734–1742 (cit. on pp. 70, 79).

[118] Hongcheng Liu et al. “Accurate Determination of Protein Methionine Oxidation by Stable Isotope Labeling and LC-MS Analysis”. *Analytical Chemistry* vol. 85, no. 24 (Dec. 2013), pp. 11705–11709 (cit. on p. 70).

[119] Joshua T. Shipman, Eden P. Go, and Heather Desaire. “Method for Quantifying Oxidized Methionines and Application to HIV-1 Env”. *Journal of The American Society for Mass Spectrometry* vol. 29, no. 10 (Oct. 2018), pp. 2041–2047 (cit. on p. 70).

[120] Blandine Rougemont. “Quantification de protéines dans des matrices complexes par spectrométrie de masse : nouveaux outils et apllications”. PhD thesis. Université de Lyon, July 2016 (cit. on p. 70).

[121] Li Zang et al. “Residual Metals Cause Variability in Methionine Oxidation Measurements in Protein Pharmaceuticals Using LC-UV/MS Peptide Mapping”. *Journal of Chromatography B* vol. 895–896 (May 2012), pp. 71–76 (cit. on pp. 70, 78).

[122] Vera Timm et al. “Identification and Characterization of Oxidation and Deamidation Sites in Monoclonal Rat/Mouse Hybrid Antibodies”. *Journal of Chromatography B* vol. 878, no. 9–10 (Mar. 2010), pp. 777–784 (cit. on p. 70).

[123] Bart Ghesquière et al. “Redox Proteomics of Protein-bound Methionine Oxidation”. *Molecular & Cellular Proteomics* vol. 10, no. 5 (May 2011), p. M110.006866 (cit. on pp. 70, 76).

[124] Ying W. Lao et al. “Chromatographic Behavior of Peptides Containing Oxidized Methionine Residues in Proteomic LC–MS Experiments: Complex Tale of a Simple Modification”. *Journal of Proteomics* vol. 125 (July 2015), pp. 131–139 (cit. on p. 70).

[125] Yi Wang et al. “Simultaneous Monitoring of Oxidation, Deamidation, Isomerization, and Glycosylation of Monoclonal Antibodies by Liquid Chromatography-Mass Spectrometry Method with Ultrafast Tryptic Digestion”. *mAbs* vol. 8, no. 8 (Nov. 2016), pp. 1477–1486 (cit. on pp. 70, 77).

[126] Ivan Verrastro et al. “Mass Spectrometry-Based Methods for Identifying Oxidized Proteins in Disease: Advances and Challenges”. *Biomolecules* vol. 5, no. 2 (Apr. 2015), pp. 378–411 (cit. on p. 70).

[127] Hendrik C Potgieter et al. “Spontaneous Oxidation of Methionine: Effect on the Quantification of Plasma Methionine Levels”. *Analytical Biochemistry* vol. 248, no. 1 (May 1997), pp. 86–93 (cit. on pp. 70, 76).

[128] Lindsay K. Pino et al. “The Skyline Ecosystem: Informatics for Quantitative Mass Spectrometry Proteomics”. *Mass Spectrometry Reviews* vol. 39, no. 3 (2020), pp. 229–244 (cit. on p. 74).

[129] R. Shyama Prasad Rao et al. “Convergent Signaling Pathways—Interaction between Methionine Oxidation and Serine/Threonine/Tyrosine O-phosphorylation”. *Cell Stress and Chaperones* vol. 20, no. 1 (Jan. 2015), pp. 15–21 (cit. on p. 76).

[130] Dung Tien Le et al. “Analysis of Methionine/Selenomethionine Oxidation and Methionine Sulfoxide Reductase Function Using Methionine-Rich Proteins and Antibodies against Their Oxidized Forms”. *Biochemistry* vol. 47, no. 25 (June 2008), pp. 6685–6694 (cit. on p. 76).

[131] Byung Cheon Lee and Vadim N. Gladyshev. “The Biological Significance of Methionine Sulfoxide Stereochemistry”. *Free radical biology & medicine* vol. 50, no. 2 (Jan. 2011), pp. 221–227 (cit. on p. 76).

[132] Earl R. Stadtman, Jackob Moskovitz, and Rodney L. Levine. “Oxidation of Methionine Residues of Proteins: Biological Consequences”. *Antioxidants & Redox Signaling* vol. 5, no. 5 (Oct. 2003), pp. 577–582 (cit. on p. 76).

[133] M. Asensi et al. “A High-Performance Liquid Chromatography Method for Measurement of Oxidized Glutathione in Biological Samples”. *Analytical Biochemistry* vol. 217, no. 2 (Mar. 1994), pp. 323–328 (cit. on p. 76).

[134] Olivier Mozziconacci et al. “Metal-Catalyzed Oxidation of Protein Methionine Residues in Human Parathyroid Hormone (1-34): Formation of Homocysteine and a Novel Methionine-Dependent Hydrolysis Reaction”. *Molecular pharmaceutics* vol. 10, no. 2 (Feb. 2013), pp. 739–755 (cit. on p. 76).

[135] Xanthe M. Lam, Janet Y. Yang, and Jeffrey L. Cleland. “Antioxidants for Prevention of Methionine Oxidation in Recombinant Monoclonal Antibody HER2”. *Journal of Pharmaceutical Sciences* vol. 86, no. 11 (Nov. 1997), pp. 1250–1255 (cit. on p. 76).

[136] Clare L. Hawkins and Michael J. Davies. “Detection, Identification, and Quantification of Oxidative Protein Modifications”. *Journal of Biological Chemistry* vol. 294, no. 51 (Dec. 2019), pp. 19683–19708 (cit. on p. 76).

[137] Walther Vogt. “Oxidation of Methionyl Residues in Proteins: Tools, Targets, and Reversal”. *Free Radical Biology and Medicine* vol. 18, no. 1 (Jan. 1995), pp. 93–105 (cit. on p. 76).

[138] John Q. Bettinger et al. “Quantitative Analysis of in Vivo Methionine Oxidation of the Human Proteome”. *Journal of proteome research* vol. 19, no. 2 (Feb. 2020), pp. 624–633 (cit. on p. 77).

[139] Endler M. Borges. “Silica, Hybrid Silica, Hydride Silica and Non-Silica Stationary Phases for Liquid Chromatography”. *Journal of Chromatographic Science* vol. 53, no. 4 (Apr. 2015), pp. 580–597 (cit. on p. 80).

[140] Helena C. Kenny et al. “Effectiveness of Resistive Vibration Exercise and Whey Protein Supplementation Plus Alkaline Salt on the Skeletal Muscle Proteome Following 21 Days of Bed Rest in Healthy Males”. *Journal of Proteome Research* vol. 19, no. 8 (Aug. 2020), pp. 3438–3451 (cit. on p. 80).

[141] Ricardo J. Esper et al. “Endothelial Dysfunction: A Comprehensive Appraisal”. *Cardiovascular Diabetology* vol. 5, no. 1 (Feb. 2006), p. 4 (cit. on p. 94).

[142] Kejing Chen, Roland N. Pittman, and Aleksander S. Popel. “Nitric Oxide in the Vasculature: Where Does It Come From and Where Does It Go? A Quantitative Perspective”. *Antioxidants & Redox Signaling* vol. 10, no. 7 (July 2008), pp. 1185–1198 (cit. on p. 94).

[143] Irina Lobysheva et al. “Nitrosylated Hemoglobin Levels in Human Venous Erythrocytes Correlate with Vascular Endothelial Function Measured by Digital Reactive Hyperemia”. *PLoS One* vol. 8, no. 10 (2013) (cit. on p. 94).

[144] Zhiyou Cai and Liang-Jun Yan. “Protein Oxidative Modifications: Beneficial Roles in Disease and Health”. *Journal of biochemical and pharmacological research* vol. 1, no. 1 (Mar. 2013), pp. 15–26 (cit. on p. 94).

[145] Pál Pacher, Joseph S. Beckman, and Lucas Liaudet. “Nitric Oxide and Peroxynitrite in Health and Disease”. *Physiological Reviews* vol. 87, no. 1 (Jan. 2007), pp. 315–424 (cit. on p. 94).

[146] Isabella Dalle-Donne et al. “Protein Carbonyl Groups as Biomarkers of Oxidative Stress”. *Clinica Chimica Acta* vol. 329, no. 1 (Mar. 2003), pp. 23–38 (cit. on p. 94).

[147] E. R. Stadtman and R. L. Levine. “Free Radical-Mediated Oxidation of Free Amino Acids and Amino Acid Residues in Proteins”. *Amino Acids* vol. 25, no. 3 (Dec. 2003), pp. 207–218 (cit. on p. 94).

[148] Etienne Maisonneuve et al. “Rules Governing Selective Protein Carbonylation”. *PLoS ONE* vol. 4, no. 10 (Oct. 2009) (cit. on p. 94).

[149] R. Shyama Prasad Rao and Ian Max Møller. “Pattern of Occurrence and Occupancy of Carbonylation Sites in Proteins”. *PROTEOMICS* vol. 11, no. 21 (2011), pp. 4166–4173 (cit. on p. 94).

[150] Angela Bachi, Isabella Dalle-Donne, and Andrea Scaloni. “Redox Proteomics: Chemical Principles, Methodological Approaches and Biological/Biomedical Promises”. *Chemical Reviews* vol. 113, no. 1 (Jan. 2013), pp. 596–698 (cit. on p. 94).

[151] Ravi Chand Bollineni, Ralf Hoffmann, and Maria Fedorova. “Proteome-Wide Profiling of Carbonylated Proteins and Carbonylation Sites in HeLa Cells under Mild Oxidative Stress Conditions”. *Free Radical Biology and Medicine* vol. 68 (Mar. 2014), pp. 186–195 (cit. on p. 94).

[152] Cattaruzza Marco and Hecker Markus. “Protein Carbonylation and Decarboylation”. *Circulation Research* vol. 102, no. 3 (Feb. 2008), pp. 273–274 (cit. on p. 94).

[153] Sabine Matallana-Surget et al. “Shotgun Redox Proteomics: Identification and Quantitation of Carbonylated Proteins in the UVB-Resistant Marine Bacterium, Photobacterium Angustum S14”. *PLoS ONE* vol. 8, no. 7 (July 2013) (cit. on p. 94).

[154] Nadezhda A. Besedina et al. “Persistent Red Blood Cells Retain Their Ability to Move in Microcapillaries under High Levels of Oxidative Stress”. *Communications Biology* vol. 5, no. 1 (July 2022), pp. 1–9 (cit. on p. 95).

[155] Vani Rajashekaraiah et al. *Reactive Oxygen Species and Antioxidant Interactions in Erythrocytes*. IntechOpen, Sept. 2022 (cit. on p. 95).

[156] Stefka Tyanova, Tikira Temu, and Juergen Cox. “The MaxQuant Computational Platform for Mass Spectrometry-Based Shotgun Proteomics”. *Nature Protocols* vol. 11, no. 12 (Dec. 2016), pp. 2301–2319 (cit. on p. 100).

[157] Stefka Tyanova et al. “The Perseus Computational Platform for Comprehensive Analysis of (Prote)Omics Data”. *Nature Methods* vol. 13, no. 9 (Sept. 2016), pp. 731–740 (cit. on p. 101).

[158] Szklarczyk D et al. “The STRING Database in 2021: Customizable Protein-Protein Networks, and Functional Characterization of User-Uploaded Gene/Measurement Sets”. *Nucleic acids research* vol. 49, no. D1 (Aug. 2021) (cit. on p. 102).

[159] Vann Bennett and Anthony J. Baines. “Spectrin and Ankyrin-Based Pathways: Metazoan Inventions for Integrating Cells Into Tissues”. *Physiological Reviews* vol. 81, no. 3 (July 2001), pp. 1353–1392 (cit. on p. 103).

[160] Benito Blanco Gómez et al. “Detection of Circulating Serum Protein Biomarkers of Non-Muscle Invasive Bladder Cancer after Protein Corona-Silver Nanoparticles Analysis by SWATH-MS”. *Nanomaterials (Basel, Switzerland)* vol. 11, no. 9 (Sept. 2021), p. 2384 (cit. on p. 103).

[161] Somaditya Mukherjee et al. “Proteomic Analysis of Human Plasma in Chronic Rheumatic Mitral Stenosis Reveals Proteins Involved in the Complement and Coagulation Cascade”. *Clinical Proteomics* vol. 11, no. 1 (2014), p. 35 (cit. on p. 103).

[162] María Del Pilar Chantada-Vázquez et al. “Circulating Proteins Associated with Response and Resistance to Neoadjuvant Chemotherapy in HER2-Positive Breast Cancer”. *Cancers* vol. 14, no. 4 (Feb. 2022), p. 1087 (cit. on p. 103).

[163] Tiffany Thomas et al. “Evidence of Structural Protein Damage and Membrane Lipid Remodeling in Red Blood Cells from COVID-19 Patients”. *Journal of Proteome Research* vol. 19, no. 11 (Nov. 2020), pp. 4455–4469 (cit. on p. 103).

[164] Satoko Suzuki et al. “Methionine Sulfoxides in Serum Proteins as Potential Clinical Biomarkers of Oxidative Stress”. *Scientific Reports* vol. 6 (Dec. 2016), p. 38299 (cit. on p. 106).

A

Appendix

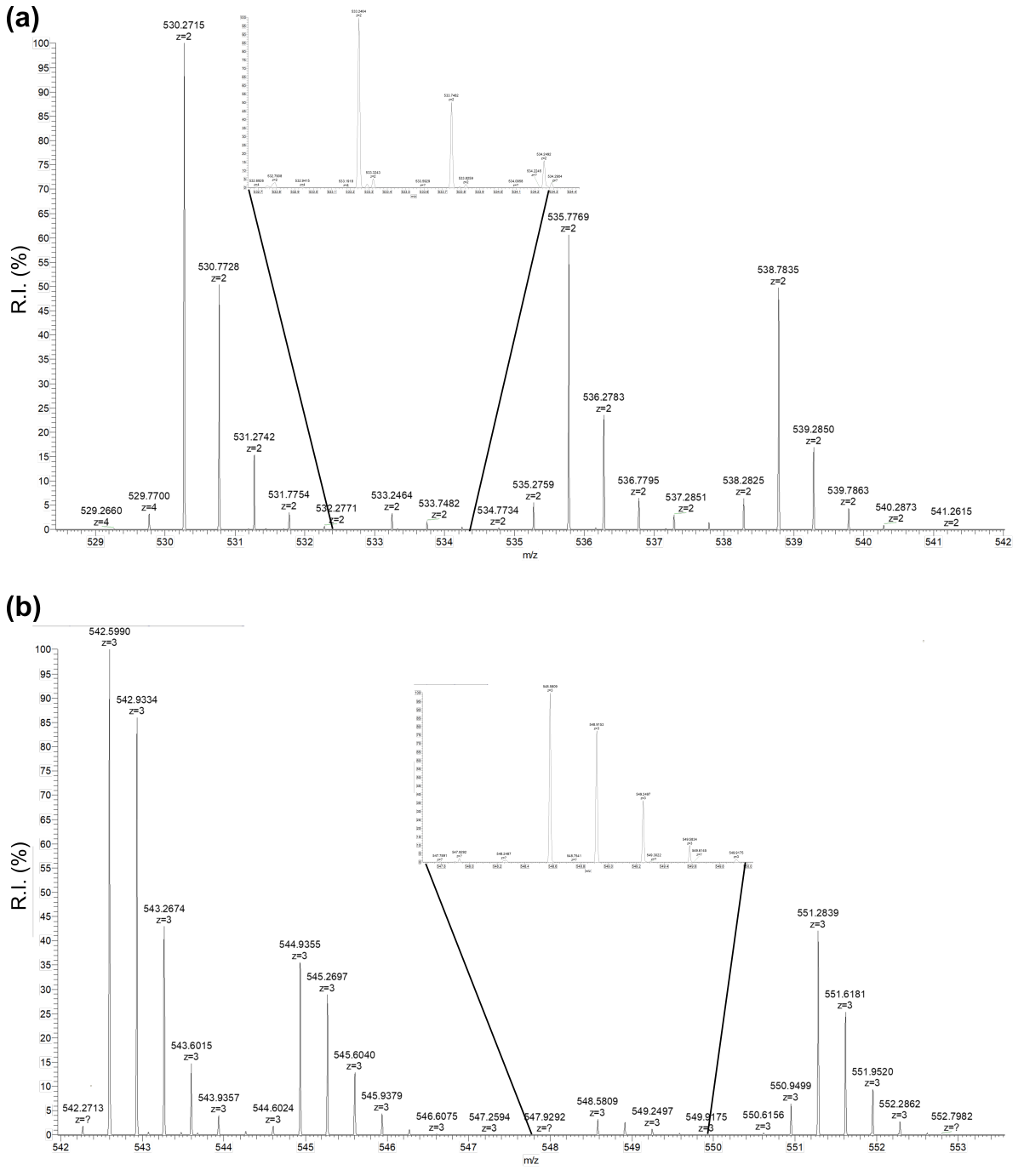


Figure A.1: *Experimental mass spectra of standard peptides (a) VDFTLSSER: 1st level [M+2H]$^{2+}$ = 533,2470; 2nd level [M+2H]$^{2+}$ = 538,7844; 3rd level [M+2H]$^{2+}$ = 535,7775; 4th level [M+2H]$^{2+}$ = 530,2716 and (b) VHDPTEEATPTPFGK: 1st level [M+3H]$^{3+}$ = 548,5811; 2nd level [M+3H]$^{3+}$ = 551,2841; 3rd level [M+3H]$^{3+}$ = 544,9359; 4th level [M+3H]$^{3+}$ = 542,5984.*

Table A.1: *Composition of the elution solvents for the 30 minutes gradient elution.*

Time (min)	Water/0.1 % FA (%)	ACN/0.1 % FA (%)
0	98	2
5	93	7
30	60	40
33	15	85
38	15	85
40	98	2
57	98	2

Table A.2: *Composition of the elution solvent for the 150 minutes gradient elution.*

Time (min)	Water/0.1 % FA (%)	ACN/0.1 % FA (%)
0	98	2
5	93	7
135	70	30
150	60	40
154	10	90
158	10	90
162	98	2
177	98	2

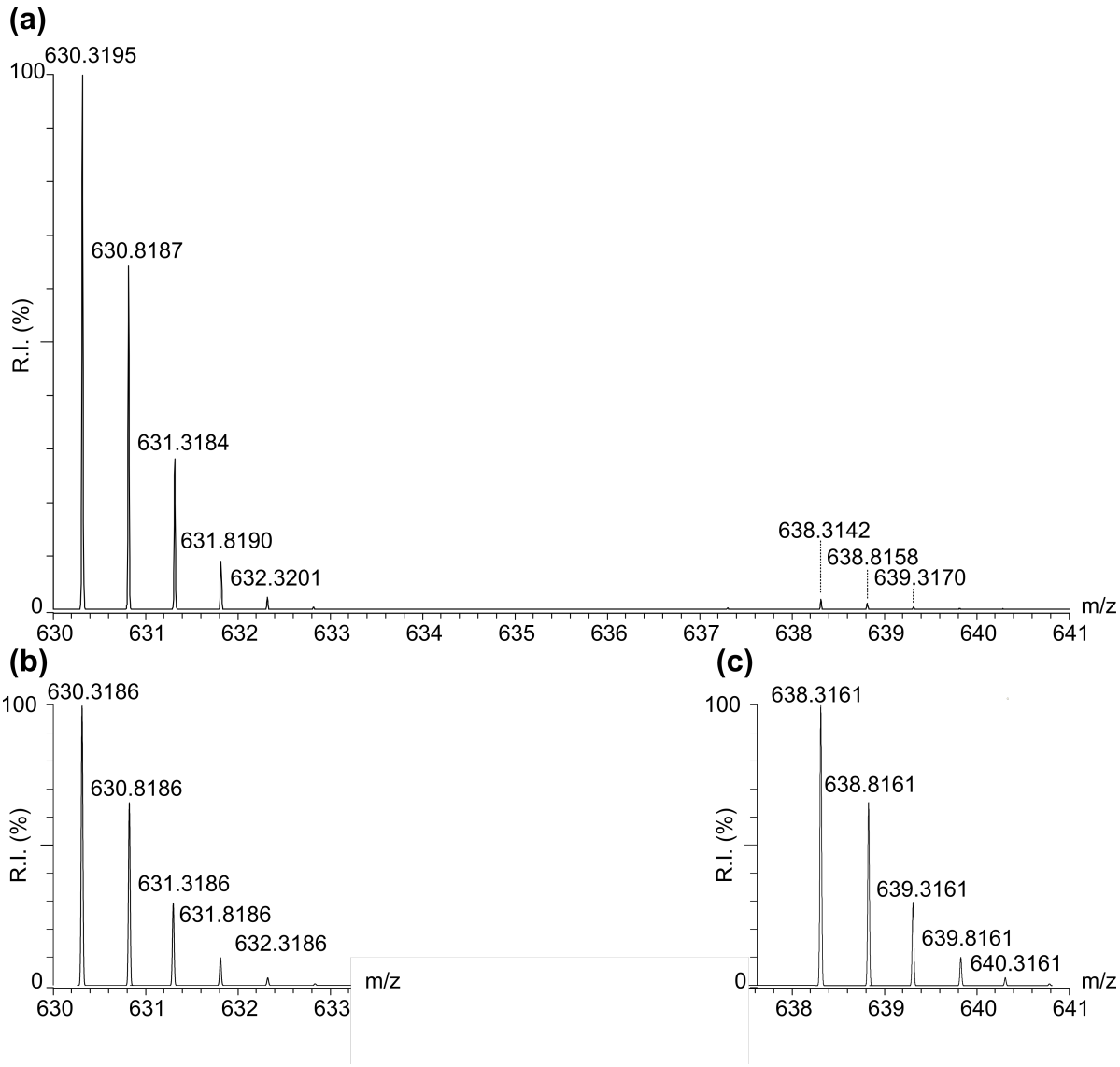


Figure A.2: *(a) Experimental mass spectrum of standard peptide, EMSGSPASGIPVK, infused on a Q Exactive mass spectrometer showing the isotopic distribution of both the native (630.3195 m/z) and the oxidised (638.3142 m/z) form of the peptide. The oxidation level corresponds to 2 % of the total of the two forms. Theoretical isotopic distribution of (b) the native form of the peptide and (c) its oxidised form.*

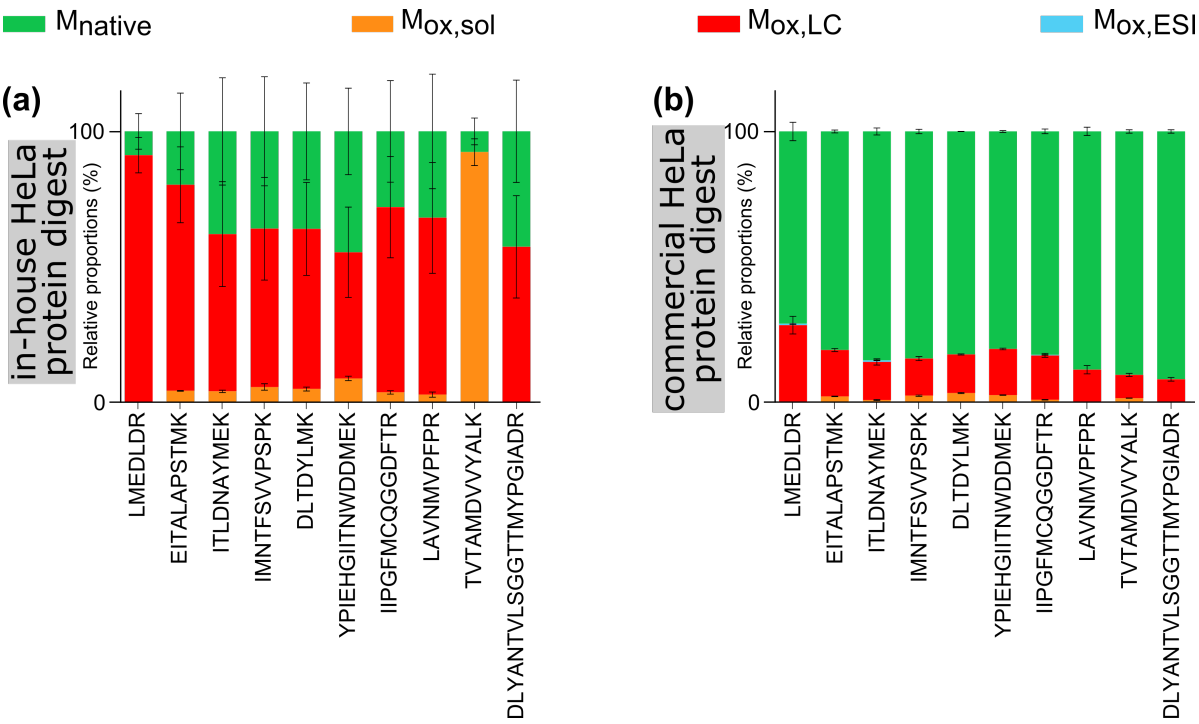


Figure A.3: *Relative proportions of Mnative, $M_{ox,sol}$, $M_{ox,LC}$ and $M_{ox,ESI}$ for the 10 most intense peptides identified in (a) in-house HeLa protein digest and (b) commercial HeLa protein digest standard following a 180 min-long UPLC method performed on a 1 year-old columns set. Peptides are sorted according to their retention time and error bars are estimated based on technical triplicates.*

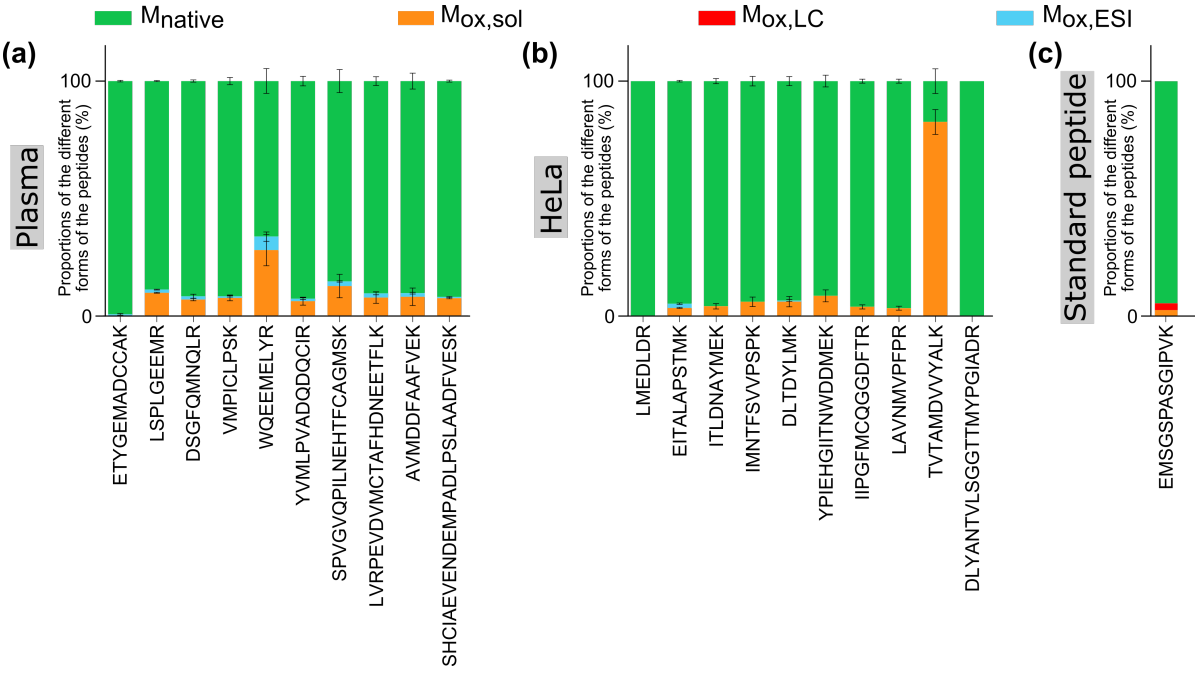


Figure A.4: *Relative proportions of Mnative, $M_{ox,sol}$, $M_{ox,LC}$ and $M_{ox,ESI}$ for the 10 most intense peptides identified in (a) plasma samples, (b) in-house HeLa protein digest and (c) the synthetic methionine-containing peptide EMSGSPASGIPVK following a 180 min-long UPLC method performed on a brand-new set of trap and analytical columns. Peptides are sorted according to their retention time and error bars are estimated based on technical triplicates. Contribution of $M_{ox,LC}$ drops dramatically compared to the use of a 1 year-old columns set.*

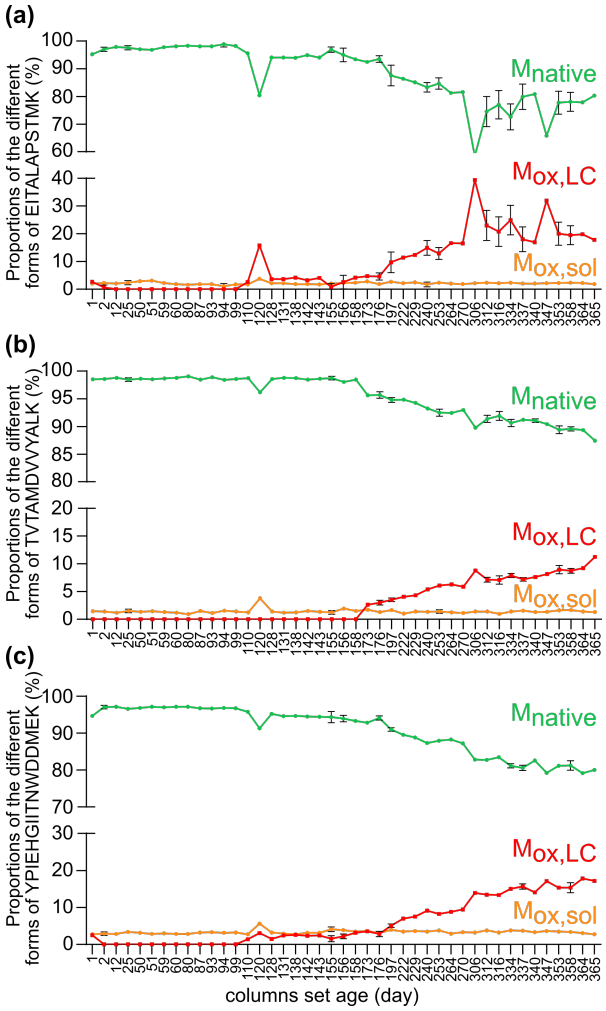
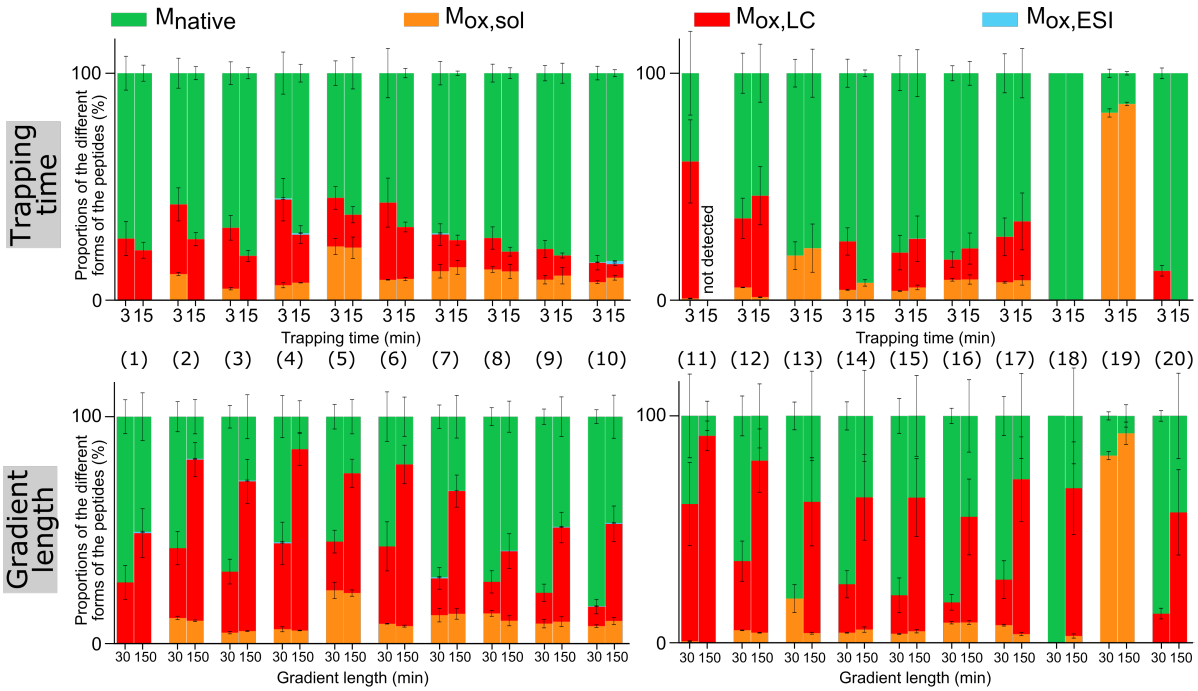


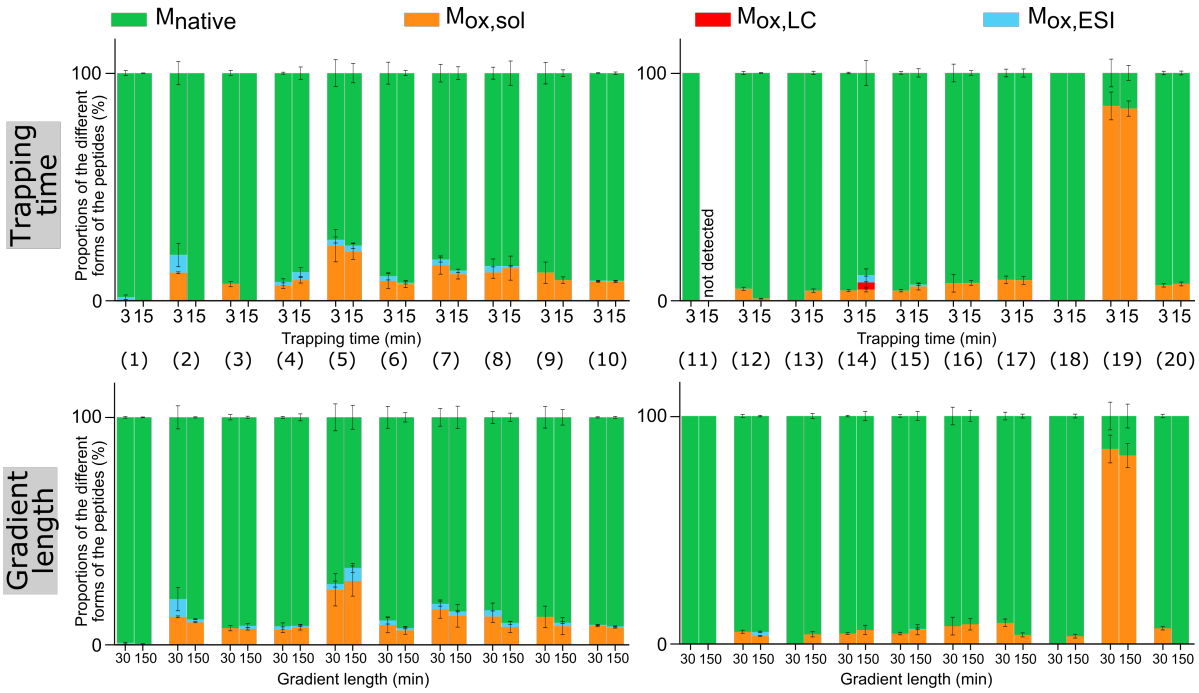
Figure A.5: *Evolution of the percentages of Mnative, $M_{ox,sol}$ and $M_{ox,LC}$ of three commercial HeLa protein digest standard peptides: (a) EITALAPSTMK, (b) TVTAMDVVYALK, (c) YPIEHGIITNWDDMEK over a 1-year UPLC use. Brand-new trap and the analytical columns set were fitted on an ACQUITY UPLC M-Class system routinely used for proteomic analysis on day-0 and were removed on day-365. Error bars are given when several HeLa were injected the same day. $M_{ox,sol}$ remains constant over the year, while $M_{ox,LC}$ progressively increases with column aging.*



(1) ETYGEMADCCAK
(2) LSPLGEEMR
(3) DSGFQMNQLR
(4) VMPICLPSK
(5) WQEEMELYR
(6) YVMLPVADQDQCIR
(7) SPVGVQPILNEHTFCAGMSK
(8) LVRPEVDVMCTAFHDNEETFLK
(9) AVMDDFAAFVEK
(10) SHCIAEVENDEMPADLPSLAADFVESK

(11) LMEDLDR
(12) EITALAPSTMK
(13) ITLDNAYMEK
(14) IMNTFSVVPSPK
(15) DLTDYLMK
(16) YPIEHGIITNWDDMEK
(17) IIPGFMCQGGDFTR
(18) LAVNMVPFPR
(19) TVTAMDVVYALK
(20) DLYANTVLSGGTTMYPGIADR

Figure A.6: *Effect of the trapping time (3 or 15 min on a 30 min-long UPLC gradient) and the gradient length (30 or 150 min with a 3 min trapping time) on the relative proportions of Mnative, $M_{ox,sol}$ and $M_{ox,LC}$ associated with: (1) to (10) plasma peptides; and (11) to (20) in-house HeLa protein digest peptides. The chromatographic analysis was performed on a 1 year-old columns set. An increase in the trapping time does not induce significant changes in Mnative, $M_{ox,sol}$ and $M_{ox,LC}$ proportions, while an increase in the gradient length drastically enhanced on-column oxidation $M_{ox,LC}$.*



(1) ETYGEMADCCAK
(2) LSPLGEEMR
(3) DSGFQMNQLR
(4) VMPICLPSK
(5) WQEEMELYR
(6) YVMLPVADQDQCIR
(7) SPVGVQPILNEHTFCAGMSK
(8) LVRPEVDVMCTAFHDNEETFLK
(9) AVMDDFAAFVEK
(10) SHCIAEVENDEMPADLPSLAADFVESK

(11) LMEDLDR
(12) EITALAPSTMK
(13) ITLDNAYMEK
(14) IMNTFSVVPSPK
(15) DLTDYLMK
(16) YPIEHGIITNWDDMEK
(17) IIPGFMCQGGDFTR
(18) LAVNMVPFPR
(19) TVTAMDVVYALK
(20) DLYANTVLSGGTTMYPGIADR

Figure A.7: *Effect of the trapping time (3 or 15 min on a 30 min-long UPLC gradient) and the gradient length (30 or 150 min with a 3 min trapping time) on the relative proportions of Mnative, $M_{ox,sol}$ and $M_{ox,LC}$ associated with: (1) to (10) plasma peptides and (11) to (20) in-house HeLa protein digest peptides. The chromatographic analysis was performed on a brand-new columns set. Neither the trapping time nor the UPLC gradient length affect $M_{ox,sol}$ and $M_{ox,LC}$.*

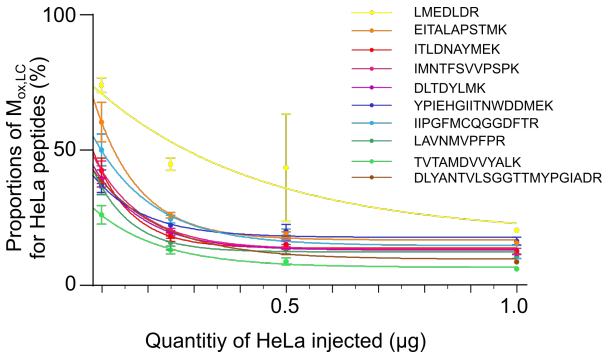


Figure A.8: *Evolution of $M_{ox,LC}$ for ten HeLa peptides as function of the quantity of commercial HeLa protein digest injected. Four different quantities 0.10 µg, 0.25 µg, 0.50 µg and 1.00 µg were injected in triplicates on a 1-year old columns set using a 180 min UPLC method (3 min of trapping time and gradient length of 150 min). The data points are fitted by a one-phase exponential decay model. The oxidation level reaches in the analytical column increases as the sample concentration decreases evidencing the presence of limited available oxidation sites within the analytical column.*

B

Appendix

Publications

- Justine Thiry, Guy Broze, Aude Pestieau, Andrew Tatton, France Baumans, Christian Damblon, Fabrice Krier and Brigitte Evrard. “Investigation of a suitable in vitro dissolution test for itraconazole-based solid dispersions.” *European Journal of Pharmaceutical Sciences* vol. 85 (Mar. 2016), pp. 94-105
- France Baumans, Emeline Hanozin, Dominique Baiwir, Corentin Decroo, Ruddy Wattiez, Edwin De Pauw, Gauthier Eppe, and Gabriel Mazzucchelli. “Liquid chromatography setup-dependent artefactual methionine oxidation of peptides: the importance of an adapted quality control process.” *Journal of Chromatography A* vol. 1654 (Aug. 2021), p. 462449

Posters

- France Baumans et al. “Contribution of cross-linking and ion-mobility for the study of protein and complex structure”
  Croatian Biophysical society and Institute of physics - Sep. 2016
- France Baumans et al. “Kit Quanta - Standardization kit for absolute protein quantitation: study of the digestion kinetic of the chimeric protein standard”
  BePAc - Dec. 2018
- France Baumans et al. “New methodology to monitor the oxidation of Met due to LC separation”
  4th NVMS-BSMS Conference on Mass Spectrometry - Apr. 2019

- France Baumans et al. “Kit Quanta - Standardization kit for absolute protein quantitation: monitoring of methionine oxidation induced by chromatography separation”
  67th ASMS Conference on Mass Spectrometry and Allied Topics - Jun. 2019
- France Baumans et al. “Kit Quanta - Standardization kit for absolute protein quantitation: monitoring of methionine oxidation induced by liquid chromatography separation”
  Department Day of Chemistry - Jul. 2019

Oral Presentations

- **Kit Quanta - Standardization kit for absolute protein quantitation: monitoring of methionine oxidation induced by liquid chromatography separation**
  Rencontres du Club Jeune de la SFSM - Mar. 2019
- **New methodology to monitor the oxidation of Met due to LC separation**
  4th NVMS-BSMS Conference on Mass Spectrometry - Apr. 2019
- **Kit Quanta - Standardization kit for absolute protein quantitation: monitoring of methionine oxidation induced by liquid chromatography separation**
  Department Day of Chemistry - Jul. 2019

Conferences

- **Croatian Biophysical society and Institute of physics**
  Sep. 2016
- **Belgian Society for mass spectrometry**
  Mar. 2018
- **Workshop and Symposium of the Belgian Proteomics Association**
  Dec. 2018
- **Rencontres du Club Jeune de la SFSM (Société française de spectrometrie de masse)**
  Mar. 2019
- **4th NVMS-BSMS Conference on Mass Spectrometry**
  Apr. 2019
- **67th ASMS Conference on Mass Spectrometry and Allied Topics**
  Jun. 2019