FACULTY OF **PHARMACEUTICAL SCIENCES**

The Quest For True Post-Translational Protein Modifications Through Label-Based Quantitative Mass Spectrometry

Pieter Glibert

Ghent University
Faculty of Pharmaceutical Sciences
Laboratory of Pharmaceutical Biotechnology

Promotor: Dieter Deforce
Co-promotor: Maarten Dhaenens

Thesis submitted in the fulfillment of the requirements for the degree of
Doctor in Pharmaceutical Sciences
2014

Members of the Jury:

Chairman:

Prof. Dr. Apr. Christophe Stove
Laboratory of Toxicology, Faculty of Pharmaceutical Sciences, Ghent University

Members of the examination committee:

Prof. Dr. Geert Baggerman
Department of Biology, Faculty of Sciences, University of Antwerp

Prof. Dr. Sebastien Carpentier
Division of Crop Biotechnics, Department of Biosystems, Faculty of Bioscience Engineering, KU Leuven

Prof. Dr. Apr. Dieter Deforce
Laboratory of Pharmaceutical Biotechnology, Faculty of Pharmaceutical Sciences, Ghent University

Prof. Dr. Bart Devreese
Department of Biochemistry and microbiology, Faculty of Sciences, Ghent University

Prof. Dr. Lennart Martens
Department of Medical Protein Research, VIB
Department of Biochemistry, Faculty of Medicine and Health Sciences, Ghent University

Prof. Dr. Jean-Paul Noben
Biomedical Research Institute (BIOMED), Faculty of Medicine and Life Sciences, Hasselt University

Dr. Maarten Dhaenens
Laboratory of Pharmaceutical Biotechnology, Faculty of Pharmaceutical Sciences, Ghent University

Table of contents

Chapter ❸

List of abbreviations

1D-PAGE	One-dimensional polyacrylamide gel electrophoresis
2D-PAGE	Two-dimensional polyacrylamide gel electrophoresis
Ab	Antibody
ACN	Acetonitrile
AQUA	Absolute Quantitation
BSA	Bovine serum albumin
CD	Cluster of differentiation
Chip	Chromatin immunoprecipitation
CID	Collision induced fragmentation
CIP	Calf intestinal phosphatase
CL	Cathepsin L
CLL	Chronic lymphocytic leukemia
Da	Dalton
DB	Database
DDA	Data-directed acquisition
DIA	Data-independent acquisition
ECD	Electron capture dissociation
EGF	Epidermal growth factor
ERK	Extracellular signal regulated protein kinases
ESI	Electrospray ionization
ETD	Electron transfer dissociation
FA	Formic acid
FDR	False discovery rate
FITS	Fluorescein isothiocyanate
FlexiQuant	Full-length expressed stable isotope-labeled proteins for quantification
FTICR	Fourier transform ion cyclotron resonance
GO	Gene Ontology
H2A	Histone H2A
H2Asp	H2A specific protease
HCD	Higher energy C-trap dissociation
HE	Histone extract

ICAT	Isobaric code affinity tags
IGHV	Immunoglobulin heavy chain variable
IMS	Ion mobility separation
IMAC	Immobilized metal affinity chromatography
IS	Internal standard
iTRAQ	Isobaric tags for relative and absolute quantitation
LC	Liquid chromatography
Leu	Leukocytes
M	Mutated *IGVH* genes
m/z	Mass over charge
MALDI	Matrix assisted laser desorption/ionization
MAPK	Mitogen activated protein kinase
mcp	Microchannel plate
mESC	Mouse embryonic stem cells
miRNA	MicroRNA
MOAC	Metal oxide affinity chromatography
mRNA	Messenger RNA
MS	Mass spectrometry / mass spectrometer
MSMS	Tandem MS
MW	Molecular weight
NE	Neutrophil Elastase
NET	Neutrophil extracellular traps
PAI	Protein abundance index
PBMC	Peripheral blood mononuclear cells
PECy	Phycoerythrin / cyanine dye
PEG	Polyethylene glycol
PICS	Proteomic identification of protease cleavage sites
PMA	Phorbol 12-myristate 13-acetate
PMSF	Phenylmethylsulfonyl fluoride
PSAQ	Protein standard for absolute quantification
PSM	Peptide spectrum match
PTM	Post-translational modification
Q	Quadrupole

QconCAT	Quantification concatemer
RA	Retinoic acid
RP-HPLC	Reversed-phase high-pressure/performance liquid chromatography
rpm	Rotations per minute
RT	Retention time
SCX	Strong-cation exchange
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SILAC	Stable isotope labeling by amino acids in cell culture
SR	SYPRO Ruby
SRM	Selected reaction monitoring
TBP	Tributylphosphine
TEABC	Triethylammonium bicarbonate
TFA	Trifluoroacetic acid
TIC	Total ion current
TMT	Tandem mass tag
TOF	Time-of-flight
UM	Unmutated *IGVH* genes
UPLC	Ultra-high pressure/performance liquid chromatography
V_{114}	Valine 114
WT	Wild type
XIC	Extracted-ion chromatogram
ZAP-70	Zeta-chain-associated protein kinase 70

Chapter 1

OUTLINE & RESEARCH SCOPE

Proteomics, the study of proteins, is a branch in biochemistry in continuous development. At the outset of this PhD project at the laboratory of Pharmaceutical Biotechnology, protein samples of different origin were typically separated by gel-electrophoresis and visualized by blotting or excised for identification by mass spectrometry (MS). The mass spectrometer has gradually matured and is today's main workhorse in proteomics. In **chapter 2 - part 1**, a general framework and some existing capabilities of MS-based proteomics are presented. Although merely a snapshot of the ever-ongoing progression, some technological expectations within the field are briefly touched to outline future directions. As proteomic studies and applications are jeopardized by biological and technical variation we point up the main elements that also challenged the research exerted for this dissertation.

With recent advancements in MS instruments the amount of protein identifications is growing multifold, so quantitative methods are becoming the center of most proteomics studies to interpret the myriad of identifications. I too contributed to this evolution in the lab and witnessed how research projects having a scattered (biological) focus, now extensively rely on quantitative mass spectrometry methods. Many specialized methodologies have been developed, each with their caveats and hurdles. In **chapter 2 - part 2** the basic principles of the applied MS instrument are presented, together with an overview of different quantitative methods.

Proteomics comprises more than protein identification and measuring expression differences as protein maturation is accompanied by so-called post-translational modifications (PTMs). Exploration of the complex PTMs changes is important, as PTMs are the wings that determine the protein activity and interactions, thereby regulating the biological effects. In **chapter 2 - part 3** we discuss proteolysis and phosphorylation, the two PTMs studied during this PhD. Since each project had a technical focus to scrutinize these PTMs, common technical approaches and main pitfalls are described aside from the general biological context. For completeness an overview of the biological models under investigation is given.

This thesis presents the main achievements and technical strategies of the research projects conducted in this PhD. To characterize the proteins a variety of established MS-techniques typically rely on a database search whereby PTMs are detected based on the predictable mass shift introduced at the peptide level. Quantitative strategies further enhance the proteome knowledge, measuring the protein presence and PTM alterations. Proteins and peptides can also be modified artificially to address the technical challenges in quantitative proteomics e.g. with iTRAQ (described in more detail in chapter 2 - part 2), a quantitative technique applied in multiple studies in this dissertation. As a central theme, iTRAQ is used in multiple approaches for answering specific questions, thereby validating the "true" extent of a PTM. The results are categorized into four subchapters:

1. Quantitative proteomics iTRAQ study revealed a specific proteolytic PTM of histone H2A.

   A "standard" iTRAQ analysis was set up to compare the protein lysates of different leukemia cells, derived from patients with a different predicted outcome. This biomarker study lead to the serendipitous discovery of a proteolytic product of histone H2A. The specific H2A truncation was already described in the past and linked to leukemia and had more recently been investigated in some leukemia cell lines in the context of differentiation.

   In a first part of this thesis we further pursued this remarkable modification by quantitative MS. Since both literature and our iTRAQ results could not unambiguously confirm the actual presence of clipped H2A a more sensitive quantitative label technique was developed to screen a larger patient population and the leukemia cell lines. The results of this investigation, described in detail in **chapter 3 - part 1**, mainly refute previous findings which emphasizes the importance of critically revising literature and preliminary results.

2. Identification of the enzyme, responsible for H2A clipping.

   The efficiency of the isotopically labeled AQUA peptides for high-throughput screening of clipped H2A V_{114} was validated on an increasing amount of histone extracts. A second focus of research was the localization of the H2A clipping event in samples of different haematopoietic origin. The results, presented in **chapter 3 - part 2**, surfaced the myeloid characteristic of H2A clipping. *In vitro*-assays and *null* mice ultimately led to the identification of the enzyme responsible for H2A clipping, previously known as H2A-specific protease. In literature, histone clipping is presented in both the context of epigenetics and histone degradation so we briefly discussed the possible biological consequences of H2A clipping in line with our findings.

3. Assessing isobaric labels for the large-scale study of phosphopeptide stoichiometry.

Phosphorylation is a well-studied PTM, which plays an important role in cell signaling, contributing to the regulation of a vast amount of cellular processes. Although numerous strategies exist for the identification of phosphorylations, quantitative information on the stoichiometry is mandatory to draw relevant conclusions. We developed a strategy whereby a peptide sample is briefly split in two identical parts and differentially labeled preceding the phosphatase treatment of one part. Our protocol marks the first time that isobaric tags such as iTRAQ were applied for the large-scale analysis of phosphopeptides on a complete lysate. The strategy was extended by adding gel fractionation, presenting a symbiosis between old and more recent proteomics techniques. After thorough optimization of the protocol, the method was tested on high-end instruments. The results, given in **chapter 3 - part 3**, display the efficiency of the approach but equally demonstrate the pitfalls which occur when analyzing iTRAQ-data at the peptide level.

4. Adapting iTRAQ for method optimization.

The iTRAQ method, developed for the comparison of protein expression, has known limitations, as also shown in chapter 3 – part 3. In **chapter 3 - part 4** we provide an alternative use of the iTRAQ labels wherein these iTRAQ biases are of minor relevance. When exploring the average reporter ratios of the isobaric labels, differences in protein yield can be measured, making iTRAQ a tool suited for method optimization. As a proof of principle, we validated the different steps of the in-gel digestion protocol, as continuous optimization of well-established methods is key in proteome research.

Chapter 2

INTRODUCTION

PART 1: PREFACE - CONCEPTUAL FRAMEWORK OF MASS SPECTROMETRY BASED PROTEOMICS

1. History and fundamentals of a growing branch in biochemistry

1.1. Introduction

The discovery of DNA in the fifties is considered to be one of the major milestones in life science. Although Watson and Crick presented the famous double-helix structure, many scientists had together paved the way for the exploration of the DNA. Almost one century before, the Swiss Friedrich Miescher encountered a substance in the nuclei of white blood cells with specific properties unlike the proteins he was searching for. He described its accidental discovery as "nuclein", the word that led to the term "nucleic acid" in DNA and RNA [1]. This short historical anecdote brings forth two important messages: (i) in science, great achievements and established landmarks are rarely the work of one person but a result of many and (ii) fortunate discoveries are often made by accident, a process known as serendipity.

With the characterization of DNA, a new way of analyzing and manipulating biological processes emerged, called molecular biology. The human genome is composed of 3 billion base pairs, which corresponds to a stretch of approximately one meter of DNA, present in each cell of our body. Yet, each cell is different and cell groups differ from each other, having different structures and functions as they make different proteins. DNA, the template of life, is transcribed into RNA which is translated by the ribosomal machinery into amino acid chains, the proteins' building blocks [3]. The expression of the genes is influenced by the environment of the cell and is coordinated by a variety of cellular mechanisms such as transcription factors and epigenetic processes (cf. infra).

Figure 1: Starting from the DNA template, present in each cell, the expression and biological function of a protein is determined at various levels during its development.

Humans have an estimated 20.500 protein-coding genes although this number is revised downwards [4]. As RNA can have alternative splicing products and proteins can be modified in many different ways, over a million different proteoforms exist [5]. Altogether, the diversity of proteins and protein variants makes the proteome far more complex than the genome. As the balance between proteins and the protein interactions has a profound influence on the biological system (Figure 1),

proteome research is appealing for numerous reasons. A detailed analysis of the structure, function and control of the proteins aims to help understand the biology in health and disease states [6].

1.2. Proteomics

The name “proteomics” became established in the mid-nineties following the introduction of the term “proteome” by Wilkins *et al*. to describe the protein pool analogous to the “genome” [7]. Proteomics is often defined as: “The analysis of the expression, localization, function, and interactions of proteins expressed by the genetic material of an organism”. In the beginning, researchers focused mainly on protein chemistry and attempted to link the observed phenotype to the relevant genes [6]. The first big wave of classical proteomics-like studies however, preceded the introduction of the name, in the 1970’s and 1980’s where mainly gel-based electrophoresis techniques were applied, followed by Edman degradation. With a growing amount of DNA-sequence data, mainly derived from mRNA, around 1990 mass spectrometry (MS) research became the main technique in proteomics [8]. Molecular weight information provided by the MS instrument combined with database information became a hallmark in protein research. Two decades later, the discipline has evolved exponentially but MS remains the method of choice [2].

A classical proteome study starts with the isolation of proteins from a biological sample. Patient samples and *in vitro* cultured cell lines are the most prominent examined. One can focus on the complete proteome of a sample or focus on a subset, obtained after specific enrichment of a protein family or modification. In gel-based proteomics, proteins are separated by 1D or 2D gel electrophoresis not uncommonly followed by immunodetection. Alternatively, LC-MS driven proteomics uses liquid chromatography (LC) and mass spectrometry to resolve the complexity of a sample. Both techniques are however complementary and are regularly applied in the same study [8]. In typical LC-MS driven workflows, protein samples are sequentially digested into peptides, followed by chromatographic separation and MS analysis (Figure 2). Generated data files are then processed and specialized data analysis tools are applied to identify the proteins [2].

Multiple strategies are available to analyze proteins at the different levels of their functionality. Depending on the research question, techniques can be applied to focus on quantification of expression and/or protein modification [9]. Similarly, degradomics and terminomics study proteolysis events and “interactomics” tries to define how proteins interact [8]. This naturally implies understanding the three-dimensional shape of proteins and complexes. The role of mass spectrometry in structural biology is also gradually growing, not in the least by the implementation of ion mobility separation (cfr. infra) into mass

spectrometers, emerging into a new field called “Structural MS”. The proteomics approaches "interactomics" and "structural MS" are fields in their own right and will not be discussed here [10].

Figure 2: Scheme of a MS-based proteomics workflow. Fractionation and enrichment is possible on the protein and peptide level, hence digestion on multiple levels is possible.

1.3. Mass Spectrometry and Proteomics

In his book “Rays of Positive Electricity and Their Application to Chemical Analyses” J.J. Thomson described several lines on a photographic plate when combining electric and magnetic fields. When changing the gas, the position of the observed lines moved. Although Thomson acknowledged the work of his predecessors, he is considered to be the inventor of the MS as his apparatus was one of the first to manipulate the “rays of positive electricity” in this manner (1907) [11]. A.F. Dempster improved the Thomson model by applying a different mass analyzer and at the same time, one of Thomson’s protégés F.W. Aston persisted in the study and design of different mass spectrographs. With his third model, Aston was able to obtain a resolving power of 2000. It was illustrated that the position of the lines on the plate depends on the e/m value of the ions, where the “e” is numerically equal to that of an electron, so the mass could be determined by the position of the lines.

Figure 3: A) Picture of an early mass spectrometer as developed by Aston. **B)** One of the first mass spectra (of carbon monoxide) by Thomson [11].

The next decades, multiple discoveries, e.g. the Nobel prize winning work on the ion trap technique, contributed to the development of known concepts of today’s mass spectrometers. For years

however, the instrument was primarily applied in analytical chemistry to study small molecules [12]. Large molecules, like proteins and peptides proved to be very difficult to ionize, hampering the MS analyses. In the late eighties, with the introduction of the electrospray ionization (ESI) and matrix assisted laser desorption/ionisation (MALDI) soft ionization techniques that cause a limited amount of internal fragmentation, the mass spectrometer became useful in proteomics [8]. Originally, as a mass analyzer, a triple quadrupole or ion trap was mostly connected to the ESI source and a time-of-flight (TOF) to a MALDI. Comparing the masses of (tryptic) peptides to peptide masses from a database, called peptide mass fingerprinting, was the first widely spread strategy to identify proteins. This approach is often performed in combination with an initial step of 1/2D gel fractionation followed by a MALDI-TOF analysis.

Soon it became clear that additional sequence information was needed. With the development of the Quadrupole (Q) TOF MS/MS spectrometers, also coupled to the ESI source, specific peptide *m/z* could be selected from an ion mixture and fragmented to obtain specific peptide information. Recorded MS/MS spectra contain information of the amino acid composition of a selected peptide since the ion fragmentation is sequence-dependent. Peptide masses in combination with MS/MS sequence information can thus be matched with sequence databases, a strategy that forms the basic principle of what is called “bottom-up” or “shotgun” proteomics, still a standard methodology in the field today.

ESI became increasingly popular since the ionization technique is compatible with liquid chromatography peptide separation, which facilitated the analysis of complex soluble peptide mixtures. Next, the nanospray was introduced increasing the sensitivity and allowing very low sample consumption by applying a very slow flow rate. This leads to an acquisition based on a computer controlled ion selection of complex peptide mixtures coming from an online high-pressure liquid chromatography system: LC-MS/MS. Different principles of acquisition exist and continuous improvement of the instrumentation and both the acquisition and identification algorithms are now the main contributors to the progress in proteome science [2, 6, 8].

2. Possibilities of MS-based proteomics

2.1. Existing Capabilities

Understanding biological systems requires information of the genetic instructions and associated protein expression. Identification of the proteins, for example by the use of the above-described classical bottom-up approach, is only one aspect of mapping the complex proteome. The dynamic nature of protein expression and protein post-translational modifications (PTM) contributes significantly to the diversity of the proteome and illustrates the need for quantification [8]. Various quantification strategies are available, which, together with ever improving technologies, help to accomplish in-depth protein knowledge.

Big steps are made by the technological improvements of the mass spectrometry instrumentation. Innovations in ion inlets, transfer optics and enhanced acquisition speed lead to an enhanced sensitivity on faster MS instruments. An increased amount of annotated peptides is further obtained by the improved identification efficiency, accomplished through the upgraded mass accuracy and resolution but also through the advancing statistics that are continuously adapted to make efficient use of the collected data. The higher quality instruments result in a broad range of opportunities on the different levels of protein research such as the acquisition, quantification and data handling [13].

For the standard shotgun approach, precursors from a survey MS-scan (MS1) are separately selected for fragmentation, giving the peptide sequence information in the MS/MS spectra (MS2). Typically, collision induced fragmentation (CID) is applied, breaking the peptides by interacting with neutral gas molecules. Complementary fragmentation can be accomplished by other techniques such as electron capture dissociation (ECD) or electron transfer dissociation (ETD). These strategies provide an alternative sequence coverage and are useful in PTM research and top-down proteomics [8].

After the acquisition, bioinformatics tools are applied to match the spectra with (genome based) sequence data. Multiple software products with different algorithms have been developed to perform the database searches that lead to protein identification: Mascot (Matrix Science™, based on MOWS algorithm) [14], XTandem [15], Omssa [16], Protein Lynx (Waters™), Proteome Discoverer (Thermo™, Sequest algorithm) and Protein Pilot (AB Sciex™, Paragon algorithm) are some of the well-known software packages. Most tools calculate the false discovery rate by searching a reversed or scrambled database called a decoy database in order to estimate the percentage of false hits. Alternative software algorithms focus on individual amino acid assignments on the MS/MS spectra (*de novo* sequencing) or identify proteins though comparing acquired spectra with a spectral library [17]. To manage the data,

tools are available to create an overview of the subsequent steps of the pre- and post-acquisition pipeline or focus on the quantification or Gene ontology to solve specific problems.

In targeted proteomics experiments, selected reaction monitoring (SRM) acquisition has the ability to increase the sensitivity and specificity of the analysis, also in complex samples. The triple quadrupole mass spectrometer instrument, wherein fragment ions of preselected precursors are monitored, provides enhanced sensitivity, dynamic range and a higher robustness. This strategy is well suited for the verification and validation of candidate biomarkers in a high-throughput manner [18].

Instead of ionizing peptide samples, the mass spectrometry analysis of intact proteins, named top-down based proteomics, is also possible. Knowing the mass of the complete protein is very useful in the identification of protein isoforms, mapping PTMs, monitoring proteolytic processing… Top-down proteomic experiments only recently became popular since the analysis of intact proteins demands greater requirements of the mass instrument in terms of resolution and accuracy and generates relatively complicated computational challenges [19].

Methodological advancements, now and in the future do not only take place at the level of the instrument or data analysis. Improvement of the sample collection and sample preparation can equally contribute significantly to the final output of the MS analysis. This is especially important in light of the day-to-day analysis of (complex) patient samples, which is becoming more and more common practice. Reducing the complexity of the samples can be achieved by orthogonal fractionation techniques and leads to a more in-depth protein knowledge. Popular methods separate protein (or peptide) pools on the size, isoelectric point or polarity [9]. Alternatively, a specific set of proteins is targeted, focusing on one particular class of protein or a specific PTM with an antibody. The reverse phase LC separation of complex peptide samples, generally with a C18 stationary phase, is considered to be a standard fractionation step in proteomics workflows. The chromatographic resolution has been increased by applying larger columns, decreasing the particle size and increasing the pressure [13]. Multiple chromatographic fractionation approaches have been combined to improve the proteome coverage.

2.2. Technological perspectives (based on Angel *et al.*)

The ever-ongoing developments in instrument, sample prep and data analysis lead towards new emerging technologies in the area of LC-MS-based proteomics. In a conventional LC-MS analysis, the retention time, ion intensity and *m/z* values are the defined parameters to distinguish the peptides in a sample. Recently, ion mobility separation (IMS) was introduced in the field as additional separation step. The IMS is based on the different velocities of molecules with different shapes during their migration in a buffer gas under impulse of a weak electric field: ions with a small collisional cross section drift faster than ions with a larger cross section due to the interaction with buffer gas [20].

With the ability of separating peptide ions by shape, IMS contributes to the multidimensional fractionation when coupled to a LC system and a (TOF-) MS platform. First the LC run reduces the complexity of the complete sample and then IMS provides an additional separation step on subsets of co-eluting peptides, thereby reducing background and improving sensitivity of the MS analysis. These complementary techniques are perfectly suited to be combined into one platform since the IMS is much faster than LC [20]. When comparing LC-MS and LC-IMS-MS it has been reported that IMS implementation annotates lower concentrated peptides and additionally significantly increases the confidence of the identification [21]. The enhanced sensitivity and the introduction of the retention (LC) and drift (IMS) time specifications create new challenges on the robustness of the analysis and the informatics tools. Spiked-in peptides with unique sequences whose masses and normalized elution times are known and linked to a database can lead to a higher throughput and identification accuracy [20, 22].

All the improvements described above lead to an increased analytical sensitivity and have paved the way for what seemed impossible until recently: single cell proteomics [23]. Before, a relatively large cell population was required, averaging the measurements' results and ignoring the cell-to-cell heterogeneity. These limitations have great consequences when the cells of interest comprise only a fraction of the sample, e.g. in the event of tumor studies where only a small portion of the cells are cancer stem cells. Latest developments however have led to the exploration of an ever smaller amount of cells and illustrate the capability to identify a set of proteins from 50 pg sample, the estimated equivalent of the amount of proteins in an average sized mammalian cell [20]. The MS instrumental evolution itself is however not sufficient to deal with the analysis of single cells as new approaches on the level of cell selection, lysis and sample preparation are needed. Comparing to conventional analysis methods, sample preparation and transfer events should be minimized as for instance displayed in methods wherein samples are lysed and separated within a capillary. Integrating all prep steps in one device is possible in microfluidic systems due to the ability to manipulate ultra-small volumes in the

range of femto- and nanoliters. Droplet-based microfluidic systems are promising since enclosing picoliter aqueous droplets with a hydrophobic phase creates individual sample vessels. By controlling the dilution one can reach the single cell level and by merging different droplets, reagents from different steps are mixed which allows managing the performed protocol [24]. Subsequent development projects focus on the combination of the microfluidic devices and the LC (-IMS)-MS systems in an integrated platform. Connecting these and other technologies provides new opportunities in both research and the clinical field [20].

2.3. Applications and challenges

2.3.1. Biomarker research

Proteins are omnipresent as hormones, enzymes, antibodies, transporters, receptors… The last decade, mass spectrometry-based proteomics has matured and is becoming increasingly important. Accurate peptide/protein mapping is critical to understand the phenotype and to better interpret the molecular mechanisms that regulate disease development and progression [13, 23].

The discovery and validation of biomarkers with a prognostic or diagnostic value is one of the most prominent applications of MS-based proteomics. Protein biomarker research consists typically out of a multi-staged process starting with the discovery, quantification and verification of a protein candidate. Next, a research assay is developed and optimized before the potential biomarker is validated on a larger patient population. In the final stage, the still successful candidate is commercialized and introduced for clinical use. From a technical MS perspective, the first untargeted/discovery stage aims for a maximum coverage of the proteome usually through the use of extensive fractionation of samples followed by a bottom-up analysis on high-resolution LC-MS platforms. Second, the targeted/verification phase aims for a higher throughput screening of a larger patient cohort, which is often obtained by the sensitive selected reaction monitoring (SRM) approach [18, 20].

Although not discussed here, the MS is also a prominent analytical tool in the field of protein and peptide drugs, a growing class of therapeutics. MS methodologies contribute to R&D processes of protein-based biopharmaceuticals and subsequent quality validation steps in the large-scale commercial production [25].

2.3.2. Addressing challenges

Several biomarkers have been introduced but successes remain modest and applications are mainly limited to a pre-clinical stage. Some of the claimed biomarkers have a limited reliability due to the lack of proper controls and independent validation strategies. Candidate biomarkers are frequently not directly related to the disease mechanism and are often a consequence of underlying cellular responses [13, 20].

2.3.2.1. High dynamic range and limited proteome coverage

The complexity of the proteome and the broad dynamic range of protein concentrations hamper adequate protein mapping and quantification. Current (1D) LC-MS/MS runs have limitations since a range of only 4-6 orders of magnitude is detected, which is in contrast with the biological range, spanning 12 orders of magnitude (from mg/ml to fg/ml) [20]. Biomarker discovery is challenged as the proteome is dominated by a relative small number of proteins, masking the more interesting differential proteins [8].

As for the powerful and most popular bottom-up strategy, complete coverage of a protein is seldomly obtained and information about splice variants, isoforms and PTMs is often not available. Also, database results from shotgun MS runs suffer from “protein inference”: it is not always sure which proteins are actually present in the cell since the same peptide sequence can be an element of different proteins leading to ambiguity in determining the identity of proteins. These problems demonstrate on one hand the need for thoughtful data-interpretation, transparency and a standardized nomenclature [26]. On the other hand, it is obvious that complementary strategies are required, depending on the goal of the study.

More in-depth protein coverage can be obtained by depleting the more abundant proteins and by fractionating the samples on the protein and peptide level prior to the MS analysis. To include more lower content proteins, a higher sample size or sensitive instruments are required [9]. Limiting the number of co-eluting peptides before MS increases the accuracy of both identification and quantification [27].

2.3.2.2. Biological complexity

The molecular mechanism of a disease is extremely complex and hardly ever completely understood. Disease development or therapeutic resistance is rarely caused by a single aberration but is a result of a complex interplay. In order to unambiguously define new biomarkers, elaborately prepared

studies are required to profile both the disease and control groups, a need that can be fulfilled by diverse quantitative proteomics strategies [13, 20]. PTM research equally allows following protein networks across different biological states. Instead of defining a single disease-related protein, protein patterns of serum or a piece of tissue possess disorder-specific characteristics. Protein identification is not always mandatory since the complete pattern of spectra represents the blood proteome, which results in a disease or healthy specific fingerprint, applicable for high-throughput screening [28]. This can allow to cope with disease heterogeneity and will, together with additional technologies, contribute to more personalized treatments for individual patients.

Due to the general variation in biology there is a need for internal standardization, replicate analysis and (from sampling to data handling) reproducible analytical methods. It is thus important to define the research question, choose the best strategy and deliberately plan the experimental procedures, realizing that every technique has advantages and drawbacks [8].

2.3.3. Outlook

With the emerging technologies, the technical limitations are gradually addressed but the step from the bench side to the bedside remains challenging. MS-based proteomics is now an established branch in *panomics* (a combination of genomics, transcriptomics, proteomics and metabolomics) and already has an important role in biological studies [20]. Multiple MS strategies are used these days to understand molecular mechanisms but the applications are mainly limited to the discovery phase. The shortcomings of some claimed biomarker candidates have increased the skepticism in the clinic although several identified biomarkers contributed to successful stories in life science and healthcare [13]. However, MS and proteomics can together play an important role at different stages of the disease management. To stimulate the implementation in the clinic, easy to handle instruments with straightforward data management tools are being developed. MS-based proteomics has the potential to become a vital element in the future clinic if the developed technologies will be able to compete with others in terms of robustness and cost-effectiveness [9].

PART 2: TECHNICAL BACKGROUND

1. Q-TOF Mass spectrometry

In the projects outlined in this dissertation, most of the protein samples were identified by the above-described bottom-up approach on a reversed-phased LC system coupled to an ESI Q-TOF mass spectrometer, also after gel fractionation. In the heart of MS instruments, gas phase ions are separated in a mass analyzer. Additionally, MS experiments require tools to introduce the sample and to generate the gas phase ions, mechanisms to generate fragment ions, a detection system and devices for the computing together with associated software. Below, the fundamental mechanisms of these different instrument components are explained and some of the existing alternatives are briefly touched.

1.1. Sample prep

Mass spectrometry is the leading workhorse in today's proteomics but in the early days, the main strategies were gel-based. For gel methods, protein samples are typically solubilized and separated according to their molecular weight by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). With 2D electrophoresis, an additional electrophoresis step is implemented, often based on the isoelectric point of the proteins. For the detection, proteins are fixed and stained within the gel or alternatively, transferred to a membrane for specific immunoblotting. Gel-based approaches have largely been omitted due to the technical advances made in LC-MS but have still specific applications being an element in sample preparation. Many strategies still rely on gel electrophoresis as molecular weight separation can fractionate complex protein mixtures, thereby increasing the depth of the proteome analysis. As a gel does not retain smaller molecules, the technique simultaneously serves as a purification method. This permits the use of detergents and urea to also include less soluble proteins in the analysis. Therefore gel electrophoresis remains crucial in the study of membrane and other, more hydrophobic proteins. In parallel with the standard digestion, proteins embedded in the acrylamide gel are commonly subjected to a reduction and alkylation steps before the digestion with trypsin. Protocols for the in-gel digestion for MS characterization of proteins are flexible and can be adapted according to the application (chapter 3 - part 4) [29].

1.2. Instrument design

1.2.1. Inlet: Reversed-phase High-pressure Liquid Chromatography

For most of the LC-MS analyses, a dried peptide sample is dissolved in an aqueous buffer and introduced into a reversed-phase high-pressure/performance liquid chromatography (RP-HPLC) system

by an auto sampler device. The separation of the peptides depends on their affinity for the immobilized hydrophobic stationary phase, which consists of porous silica, derivatized by a n-alkyl hydrophobic ligand. For peptide separation, C18 columns are the most popular. Especially for larger polypeptides, shorter, less hydrophobic ligands can be used resulting in an increased recovery of more hydrophobic molecular species. The peptide mixture is gradually eluted by increasing the amount of organic solvent in the mobile phase [30]. Before the final elution, the sample mixture is often trapped on a non-analytical smaller column, providing an additional purification. The most popular buffers for the mobile phase are acidic in order to protonate the peptides (such as TFA or FA), but nonionic detergents supplementation can be used to include more hydrophobic peptides [31]. The most prominent organic solvents are acetonitrile, methanol and 2-propanol. If mandatory, the eluting peptides are monitored by online UV spectroscopy detection at 210-220 nm (absorption wavelengths of the peptide bond) or 280 nm (absorption wavelengths of the aromatic amino acids) [30].

Two-dimensional LC separation techniques are also becoming more and more prominent to increase the depth of the protein analysis and to decrease the matrix interference effects. The second dimension e.g. a cation exchange or a high-pH reversed phase chromatography can be applied as an offline system or directly coupled to the mass spectrometer. LC systems in a Chip-based format are now also available but the advantage in practice still needs to be proven [32].

1.2.2. Ionization: Electrospray ionization

To transform the soluble ions eluting from the LC system into the gas phase, the atmospheric pressure ionization technique called electrospray ionization (ESI) is the most common. Surrounded by a nitrogen flow, a voltage (typically 3-4 kV) is applied on the capillary and the heated electrospray probe. A spray is created from the low flow that arrives from the HPLC tubing connected to the probe. The nitrogen facilitates the evaporation of the solvent and reduces the size of the droplets, leading to an increased density of the charges at the droplet surface. Finally, the repulsion between the charges causes a coulomb explosion generating the gas phase ions, which are then drawn to the vacuum of the MS.

In protein research, compounds are generally positively charged due to the acidic buffer so MS runs are performed in the positive mode. The applied buffers are volatile with preferably low concentrations and with a low salt content in order to limit the salt deposit on the needle and decrease the ion suppression phenomenon. ESI is very attractive in proteomics since it is a soft ionization technique that can generate multiply charged ions. Depending on the manufacturer, the spray is

connected straight towards the cone to assure the sensitivity or a perpendicular “Z-spray” design is applied to protect the cone from non-volatile, non-ionized molecules [33].

Depending on the application, the Matrix-Assisted Laser Desorption/Ionization or MALDI approach is an attractive alternative ionization method. With MALDI, the peptide mixture is incorporated in a matrix of small organic molecules that co-crystallize as spots in a well of a stainless steel or teflon plate. Ionization is achieved by irradiating the spot with a pulsed nitrogen UV laser, creating mostly singly charged ions. MALDI is a very sensitive method and although not easily coupled to an LC instrument, the source is well suited for imaging and other clinically related applications [34].

1.2.3. Mass analyzer: Quadrupole - Time-of-flight

The mass analyzer, the core of the MS instrument, separates the ions. One of the most popular hybrid designs combines two important MS instrument technologies: a quadrupole (Q) coupled to a time-of-flight (TOF) mass analyzer.

1.2.3.1. Quadrupole

A Q consists out of two associated alternating pairs of rods: one with a positive electrical potential and one with a negative potential. On each pair, a combination of direct current (dc) and radio frequency (rf) voltage is applied in order to control the orbit of ions. The field strengths are regulated by the operating software and determine which *m/z* are able to pass through the Q. Changing the dc voltages alters which *m/z* range gets a stable trajectory. Qs are easy-to-use, cover a large mass range and are suited to accurately select ions with a predefined *m/z*, both as a lens, selector or analyzer.

In a hybrid design for tandem MS or a triple quad instrument, different Qs are combined. A first Q serves as a lens or selects an ion precursor and a second Q is able to operate as a collision cell wherein the peptides are fragmented, which allows obtaining detailed sequence information. The fragmentation of the peptides occurs if the electrical potential is elevated and the ions collide with a neutral gas (such as argon) [33].

1.2.3.2. Time-of-flight

In the TOF, the actual mass analyzer, ions are accelerated into the field-free tube. As the distance of the ions from the pusher to the detector is fixed, the velocity is dependent on the molecule *m/z*. Ions with a smaller *m/z* reach the detector faster than ions with a larger *m/z*, so each *m/z* value correlates with a characteristic time-of-flight. It is however important that all the ions possess an identical kinetic energy (= ½ mv^2, where m is the ion mass and v the velocity). Variation in kinetics is

therefore compensated by the reflectron, thus increasing the resolution. The high sensitivity and resolution of the TOF, together with the selectivity of the quadrupoles and capability to fragment preselected ions leads to accurate mass measurements of both the peptide precursors and peptide fragments [2, 33].

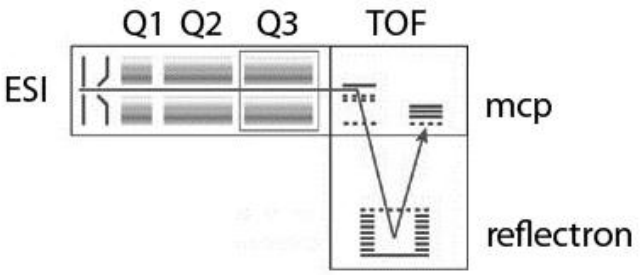


Figure 4: Basic scheme of the Q-TOF mass spectrometer. After the ionization the ions move to the vacuum of the instrument. The first quad acts as a lens, the second as a mass filter and the third as a collision cell. Ions are separated in the TOF mass analyzer before the detector is reached. Ions with a different m/z have a different velocity in the TOF [2].

In proteomics, mass analyzers based on other principles such as the Orbitrap, are becoming increasingly popular. The Orbitrap analyzer consists out of a small (<3 cm) inner coaxial electrode and forms an electrostatic field in which ions oscillate in orbital movements. The instrument can control the ion trapping and has the ability to separate ions since the frequency of oscillation is inversely correlated with the *m/z*. In parallel with the Fourier transform ion cyclotron resonance (FTICR) MS, the ion pathway can be controlled resulting in a very high resolution [35].

1.2.4. Detector: Microchannel plate

The principle of the microchannel plate (mcp) is analogous to the principle of the electron multiplier: In vacuum, under the influence of an electric field, the contact of an electron with emissive material induces emission of more electrons which results in a cascade of electrons propagated through the channels. This stream of electrons starts with the impact of ions and is collected by an anode to create the signal. The multiple channels of the mcp can detect many ions at the same time since a single ion only triggers a few channels simultaneously on a relatively large detection area. After the scan of the ions, the mcp recharges, getting ready for the next scan. An mcp detector has a high sensitivity and a time response smaller than one ns [33].

1.3. From MS acquisition to MS data

Programmed by a predefined set of parameters two kinds of spectra are generated during a typical MS run: one with *m/z* values of intact peptides eluting at a given time and one with *m/z* values of fragments of the selected precursors [36]. Apart from this so-called data-directed analysis (DDA), strategies wherein the acquisition is independent of the selected precursors detected at the MS1-level

are also gaining popularity. These recently developed data-independent acquisitions (DIA) fragment a range or all of the co-eluting peptides at the same time and align the precursor and fragment masses supported by the retention time for peptide identification and quantification [37].

The current mass spectrometry instruments create a high amount of data, which require great demands on the algorithms and biostatistics. Before the identification search, a peak list is created from the raw data by a couple of preprocessing steps such as deisotoping, centroiding and charge deconvolution. The influence of preprocessing is often underestimated since most vendors already couple some of the preprocessing actions automatically to the acquisition software, thereby hampering transparency in data handling and hindering adaptation of the parameters. Inappropriate processing, especially of low quality spectra, endangers the path to identification and increases quantification errors [27, 38].

For the DB algorithms, tandem MS data peak lists are matched to an in-silico digested and fragmented DB. In this dissertation, the Mascot Software (Matrix science) is mainly applied to search the tryptically digested proteins against the SwissProt DB. Swissprot is a Uniprot DB with high quality, manually annotated and non-redundant protein sequences. Depending on a set of parameters (including the taxonomy under investigation, the enzyme to create the peptides and the possible modifications) a DB is converted into a list of sequences with corresponding mass values which are compared to the list of masses obtained from the experimental run according to a predetermined maximum mass error [39]. In case of a hit, the masses of the generated and theoretical fragments are correlated and scored by the algorithm. Different software tools often apply different scoring methods and criteria to include and rank the identified peptide/protein hits.

In practice however, unannotated spectra are the rule rather than the exception. Contaminants, such as PEG are often the subject of undesired ionization and fragmentation, leading to uninformative spectra. Other spectra elude identification due to, (i) unexpected modifications (either endogenously present or *in vitro* induced), (ii) insufficient or unspecific protease activity (either from the protease intentionally applied to digest the proteins or residual proteases from the biological sample), (iii) absence in the DB, (iv) MS/MS spectra that are the product of co-eluting precursors with comparable *m/z*,... The software gives the opportunity to calculate the false discovery rate (FDR) of the identification, mostly by searching the peak lists from the MS-run against a database supplemented with a random, non-existing decoy DB or a DB of a non-related species [40]. In order to increase the relative amount of annotations, preprocessing algorithms often apply spectrum quality filtering wherein only the most informative spectra are withheld for the subsequent analysis. Implementation of filtering early

reduces the need for computational capacity and processing time in the subsequent searches of the data analysis pathway [36].

If a high quality fragmentation spectrum is not identified because for example an unexpected modification occurred, annotation by *de novo* sequencing is still possible. With *de novo* sequencing the amino acid sequence is directly derived from the mass differences between the peaks in the MS2 spectra. Semi manual interpretation of the fragmentation spectra is possible but time consuming and therefore specialized software is available. Some algorithms rely on a tag-based strategy, wherein a subset of a spectrum is *de novo* sequenced and a sequence tag is created and matched with a DB. An increasingly popular strategy matches the query spectra with library spectra. This method circumvents the theoretical formation of an in-silico created DB and thus requires a much smaller search space. The spectral library searches are fast and quite sensitive but do not allow the discovery of previously unobserved peptides [36, 40, 41].

2. Quantitative proteomics

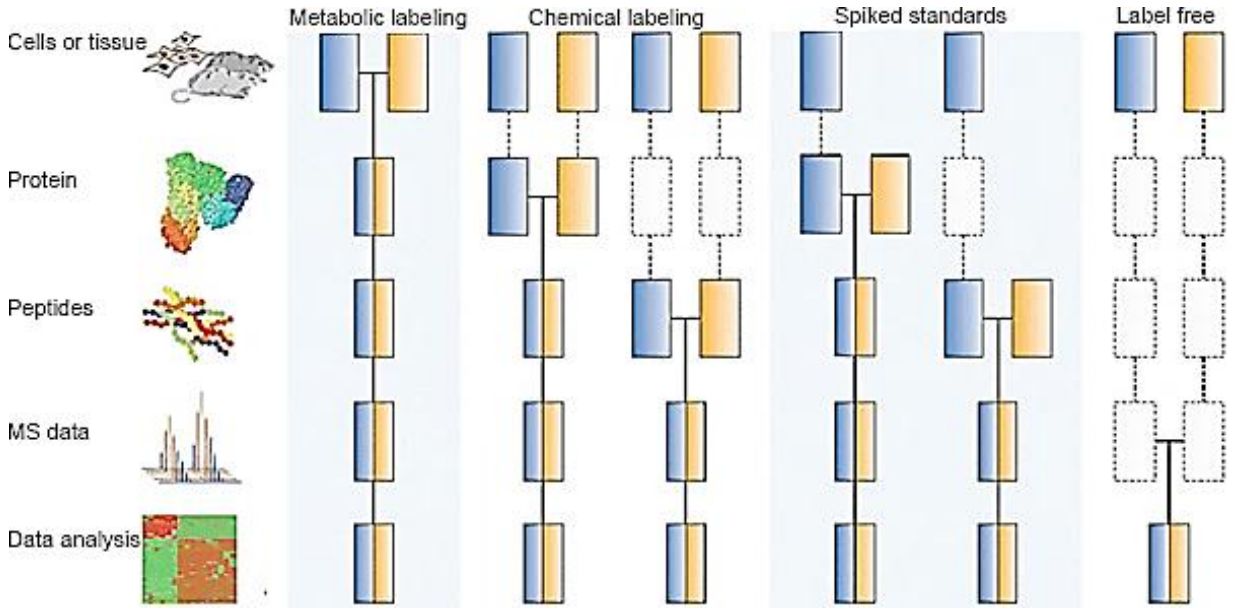


Figure 5: Overview of the different quantification strategies. The blue and yellow boxes symbolize two different experimental conditions and the horizontal lines the level where the samples are combined. Dashed represent where variation can occur (Figure from [37]).

The quantitative analysis is the center of most proteomics studies. Before, protein samples were relatively compared based on the intensities of different gel spots but today, quantification is predominantly performed by MS-based strategies [42]. Here, we give an overview of the different methods that exist in order to compare the protein expression or PTMs between multiple samples. A high robustness is a necessity for quantitative studies which are already gravely challenged by biological and systematic variability, the vast proteome complexity plus its large dynamic concentration range [43]. We more elaborately focus on iTRAQ as this method is applied in multiple studies in this dissertation.

2.1. Metabolic labeling

With metabolic labeling, an isotope label is introduced directly in the medium during the cell growth and replication and hence is present in every protein of the sample. Typically, for the popular SILAC (stable isotope labeling with amino acids in cell culture) method, heavy essential amino acids (^{13}C and ^{15}N) are added to the culture medium as to be incorporated after a couple of cell cycles. Since a second cell population is grown in normal medium, relative quantification of the two populations is possible by comparing the intensities of the peptide isotope clusters in the MS1 level. Although

metabolic labeling methods are limited to cultured samples, they provide a minimal technical variability because the different samples are combined in the beginning of the proteomics workflow. Importantly however, metabolic conversion of amino acids makes the use of SILAC on some cell lines rather impractical [44].

Combining more than two samples is possible but in practice the extension of multiple SILAC labels is in general limited to three samples due to the overlap of the different isotope clusters. Adapting the classic SILAC protocol has led to new applications. For example, protein degradation and turnover can be quantified with the pulsed SILAC method wherein the heavy isotopes are added only for a certain time. The pulsed method has also been modified by the introduction of an additional label in order to measure the influence of a specific cell treatment (e.g. drug incubation) [45]. With the "super" SILAC, a mixture of different SILAC-labeled cell lines serve as a standard in order to compare the protein expression between samples from different origin such as tissue cells [46].

2.2. Chemical labeling

2.2.1. MS1-Quantification

Peptide quantification in the MS1 spectra, similar to SILAC, is also possible by *in vitro* incorporation of chemical labels at the protein or peptide level. As a consequence, quantification of samples from multiple sources is achievable, including those that are not prone to metabolic labeling such as patient samples. For example, supplementation of $H_2{}^{18}O$ at the time of the protein digestion introduces the heavy ^{18}O isotope into the carboxyl termini of the generated peptides. The mass difference with peptides derived from a normal digestion makes quantification possible. The ^{18}O method is easily applicable but the re-introduction of ^{16}O due to the re-accumulation of ^{16}O in the buffer and residual protease activity after a buffer switch in the workflow, often hampers accurate quantification [42].

The Isobaric Code Affinity Tags (ICAT) apply a light or (deuterium containing) heavy linker group in order to compare different samples. The tag also consists of a reactive thiol group in order to react with the reduced cysteines of the peptide mixtures under investigation plus a biotin group that serves as the affinity tag, making purification possible. Despite the promising applications, the technique has never broken through since analysis is limited to cysteine containing peptides. The isotope-code protein label technique on the other hand, works according to the same principle, yet reacts with primary amines and thus includes all peptides in the analysis [47].

An alternative method labels the primary amines by formaldehyde and cyanoborohydride, resulting in a mass shift. With the use of deuterated and ^{13}C-labeled formaldehyde, heavier labels are generated relatively cheap. The combination of isotopomers theoretically allows to multiplex a profound amount of samples although with a maximum since overlap of isotope envelopes should be avoided [48]. Isotope labels are also the basis of miTRAQ, a variant of the known iTRAQ (cfr. infra) method [49].

2.2.1. MS2-Quantification

2.2.1.1. ITRAQ: principle and strength

Reporter based labeling methods where different samples are labeled with different but isobaric labels at the peptide level are widely spread in quantitative proteomics. We mainly focus on the iTRAQ method since this labeling method is used on several occasions in this dissertation. The labels consist of multiple parts: a reactive group that reacts with the peptides, a reporter group inherent to each label and a variable spacer group to ensure that all the labels remain isobaric. For iTRAQ 4plex for example, each label consists out of an N-methyl piperazine reporter group with mass=114.111, 115.108, 116.112, or 117.115 Da, a balance group with mass=30.995, 29.991, 28.986, or 27.987 Da, and an N-hydroxy succinimide ester group that reacts with the primary amines of peptides (Figure 6).

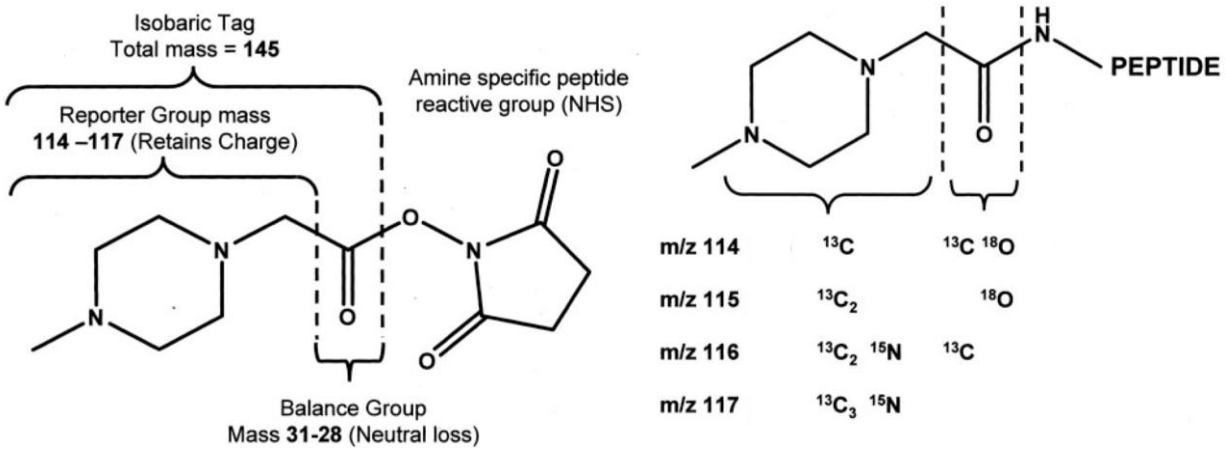


Figure 6: Structure of the iTRAQ label. The NHS ester of the reactive group binds covalently with the primary amines of the peptides. The overall mass of reporter and balancer are kept constant by using differential isotopic enrichment (^{13}C, ^{15}N and ^{18}O). The reporter's mass ranges from 114 to 117 Da which is complementary to the balancer, ranging from 28 to 31 Da. This means that the mass of the whole label remains constant (145 Da) for each of the four reagents.

The iTRAQ 8plex kit, whereby eight samples are simultaneously quantified, is complemented with the 113, 118, 119, 121-reporter masses and corresponding balancer groups in order to reach the total mass of 305 Da instead of 145 (4plex) [27].The isobaric character of the (iTRAQ) labels allows for multiplexing from the peptide level on, so the variability that is associated with every subsequent step i.e. sample purification, LC retention and mass spec analysis, is eliminated. Each peptide precursor has the same mass across the differentially labeled samples, which increases method sensitivity and decreases the susceptibility to background noise (Figure 7). The relative yield of each condition is quantified by the reporter ions in the MS2 spectrum only after fragmentation. By compiling peptide ratios, protein ratios are being defined. An example of a similar isobaric labeling strategy is the tandem mass tag (TMT), available in both duplex, 6plex and 10plex [42].

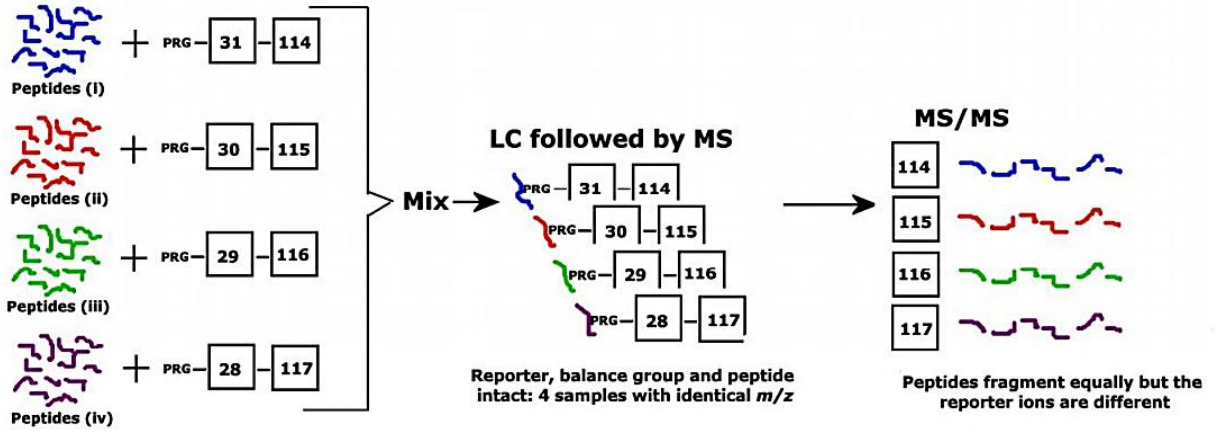


Figure 7: Peptides from different samples are differentially labeled. Due to the isobaric character of the labels, a peptide present in multiple samples is presented as one peak in the MS1 spectrum. After fragmentation, relative quantification is possible in the MS2 spectrum due to the ratio of iTRAQ reporters. [Figure 6 and 7 based on: http://biostat.jhsph.edu/~iruczins/teaching/260.655/misc/iTraq.jpg].

2.2.1.2. ITRAQ: challenges

ITRAQ continues to be a valued technique in proteome research but is losing ground to label free strategies because of its high cost and the increasing discussion about its accuracy and precision [27, 37]. The underestimation of the fold change is one of the main drawbacks especially in the event of a broad dynamical range. This means that calculated iTRAQ ratios are lower than the actual differences between samples. The error in quantification has been evaluated in multiple reports and can be attributed to multiple factors: (i) Cross-label isotopic impurities, mainly M+1 and M+2, as a by-product of the label production, (ii) Near-isobaric fragments that interfere with the iTRAQ reporters in the "silent" iTRAQ region caused by immonium ions or surprisingly, also by mirroring isotopes of the labels

itself. Several near isobaric masses around the reporter labels were detected by high resolution MS. These hamper the iTRAQ quantification on lower resolution instruments since the contamination appears as shoulder or tail peaks (Figure 8), (iii) Mixed MS/MS, arising from co-eluted and near isobaric peptide ions which are co-fragmented, compromises the quantification accuracy in several ways since the reporters of two different precursors contribute to the same MS2 spectrum and compress the ratio [50-52]. This latter source of interference is minimized when one of the precursors is highly abundant and fragmented at the top of its elution. High sample amount and complex peptide mixtures on the other hand increase the degree of this so-called "isobaric elephant" [43, 53].

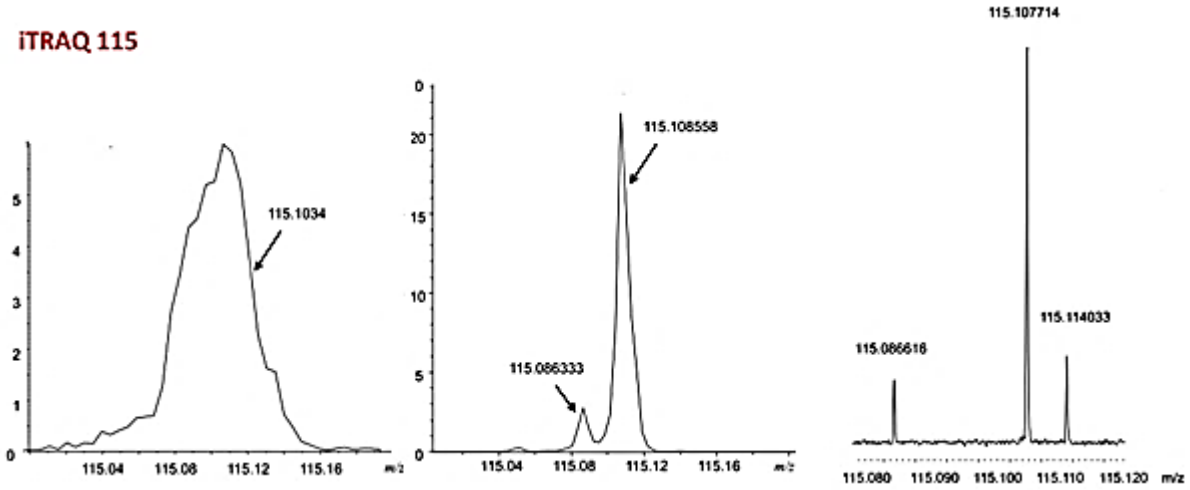


Figure 8: Contaminating peaks in the reporter area hamper accurate quantification especially on low-resolution instruments. MS/MS spectra of reporter label 115 on different mass spectrometry instruments with an enhanced resolution (left to right: R4500/R15.000/R400.000) reveal the presence of near isobaric masses [27].

In terms of precision, iTRAQ heterogeneity between the different measurements could result in high variance, especially in spectra with low intensity (reporter) peaks. Both phenomena, mixed MS/MS and low intensity spectra, omnipresent in complex samples, additionally sabotage the identification. The iTRAQ 8plex also reduces the identification efficiency compared to the 4plex due to its tendency to form multiple charge state precursors as a result of the larger balancer group [27].

2.2.1.3. ITRAQ: addressing the challenges

Because of its great promise in quantitative proteomics, different strategies through adaptation of the software, acquisition or upstream workflow modifications have been developed by the iTRAQ

research community in order to overcome the challenges. Since the manufacturer provides the information of the cross-label isotopic impurities, the available software automatically applies correction filters in the algorithms thus increasing quantification accuracy [27, 54]. Accurate selection of the iTRAQ reporter peaks in the reporter region is essential to distinguish the true reporter masses and the near-isobaric impurities, a process that is facilitated by the advanced technologies of the MS instruments [27, 43].

Supplementary fractionation steps can minimize the amount of mixed MS/MS, which is particularly useful for complex samples. Additional separation is achievable on the different stages of the sample prep, such as extra pre-chromatography steps or ion mobility gas-phase fractionation. To obtain high quality spectra, acquisition parameters can be optimized by e.g. elevating the criteria for fragmentation selection to create the MS/MS close to the elution maximum. By selecting intense fragments in the MS2 spectra for subsequent fragmentation and quantification in an MS3 spectrum, the interference of the co-eluting peptides can also be circumvented [51].

Data derived from iTRAQ experiments has forced data analysts and bioinformaticians to adapt existing software and has inspired the development of specialized bioinformatics tools. Criteria to include peptides for quantification or dataset normalization were already automated from the beginning. In order to respond to the disturbing high variance and low precision, intensity cut offs are still popular to remove the low quality data although the use of this approach is questioned. Other methods propose outlier removal, apply weighted means or stabilize the variance [43, 51, 53]. Another increasingly popular approach adds a known standard mixture or complete lysate in the iTRAQ set-up. This internal control has the ability to measure the error and reproducibility of the quantification and allows to correlate different runs, also in different experiments [55].

2.3. Spiked standard

An important, well established spiking method is the Absolute Quantitation (AQUA) method. This strategy requires prior knowledge of proteotypic peptides of the targeted proteins. For AQUA, a known amount of a known peptide(s) is spiked into a sample, making absolute MS quantification possible (Figure 9). The synthesized AQUA peptides are isotopes (mostly ^{13}C and ^{15}N are incorporated into one of the amino acids) of the peptides under investigation and will thus be indistinguishable until the MS1 level where the quantification becomes possible by correlating the intensity of the spiked and sample peptide peaks [56]. The AQUA strategy is very sensitive, which allows measuring low concentrated peptides and is proven to be very useful in targeted studies such as SRM. Minimizing the

sample prep is however recommended since the peptides are spiked relatively early in the workflow which induces variation.

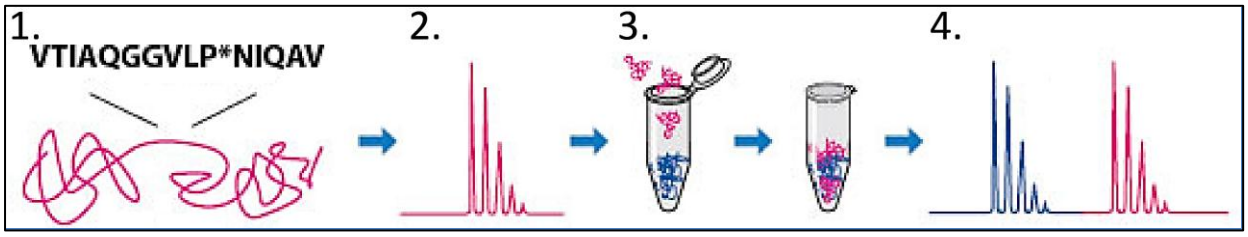


Figure 9: Workflow for AQUA quantification. 1) Peptide selection and production of the isotopically labeled AQUA peptide(s). 2) Optimization of the AQUA LC-MS analysis. 3) Spiking of a known amount of AQUA within the sample and protein digest. 4) LC-MS analysis for quantification (figure based on www.sigmaaldrich.com/life-science/proteomics/).

Alternatively, the Quantification concatemer (QconCAT) technology can simultaneously quantify dozens of proteins. By recombinant technology, polypeptides of interest are synthesized in a heavy isotope enriched medium and provide a sensitive quantification of 100 proteins at the same time, also in biological matrixes. Techniques such as Protein Standard for Absolute Quantification (PSAQ), absolute SILAC and Full-Length Expressed Stable Isotope-labeled Proteins for Quantification (FlexiQuant) apply complete stable isotope-labeled proteins. The multiple peptide standards per protein and spiking before the digestion contributes to the high specificity and measurement robustness of the workflows [37].

2.4. Label free quantification

With label free, no (expensive) reagents are needed and sample protocols can be simplified. Label free approaches have the advantage of comparing larger sample numbers, which also allows introducing replicates more easily. For an identified protein, the number of peptide spectrum matches (PSM) or the number of peptides are both used as a measure of abundance, relying on the intensity or response of a peptide as a quantitative measure. Until recently, signal intensity from the MS1 scans was the main strategy for label free quantification. The eXtracted Ion Chromatogram (XIC), being the area under the curve from the intensity and retention time, is applied to measure peptide abundance. Different methods were developed to estimate protein abundances from these XICs [37].

Most label-free methodologies rely however on MS2-based methodologies. The popular Mascot software uses Protein Abundance Index (PAI)-scores, calculated by dividing the number of identified by the number of observable peptides. Spectrum counts deliver the highest dynamic range and

reproducibility but require filtering of the data i.e. defining which criteria are used to include the PSM in the quantification [42]. Recently, label free quantification has been moving towards independent acquisition strategies. In such an approach, named SWATH, the DIA acquisition is interpreted based on the data of a standard tandem MS run, wherein a retention time and mass tag are attributed to the identified peptides to create a library. Second, during a quantitative independent analysis, all the precursors of small, predefined mass windows are fragmented simultaneously and this repeatedly over the whole mass range. After the acquisition, based on the XICs, the identified parent ions are coupled to the fragment ions and associated intensities of the second run. Another independent approach switches between the acquisition of all the precursors eluting at a given moment and the acquisition of all the fragments from those co-eluting peptides by continuously alternating the collision energy. Afterwards, the data from all the co-eluting precursors and co-eluting fragments are correlated. For both strategies, a robust LC system and a fast instrument with a high resolving power are essential to create the data together with specialized software to handle the identification and quantification [37, 57].

PART 3: BIOLOGICAL CONTEXT

1. PTMs: From the underestimated protein proteolysis to the widely spread phosphorylation

The amino acids of proteins can be modified by hundreds of chemical groups that alter the structure and function of the proteins. PTMs are inherent in protein maturation and allocate the protein activity, including the cellular localization and the formation of protein complexes. The most common modifications in nature are acetylation, formylation, methionine oxidation, phosphorylation, ubiquitination and glycosylation. Apart from the covalent binding of these chemical groups or protein moieties (ubiquitination) to the primary structure, proteolytical cleavage of proteins is also considered as an important PTM (cfr. infra) [58]. This introduction will elaborate only on proteolytic cleavage and phosphorylation as these two modifications were subject to extensive study during the course of this dissertation.

The alteration of proteins by PTMs not only has severe biological consequences but also has implications for the way these proteins can be studied, as modifications could interfere with the sample prep or the subsequent analysis. When a gel-based approach is used, immunodetection of the PTM or even the protein, is often compromised by epitope occlusion or epitope truncation. In LC-MS driven analysis, a PTM causes a mass shift of the peptide molecular weight and consequently modifies the MS/MS spectrum, a transformation that also needs to be taken into account when performing a DB-search. Modifications can also hinder the peptide ionization and fragmentation, resulting in loss of identification, a challenge that is often influenced by the low abundance of particular peptides within the sample. The improved sensitivity of the MS-instruments and specialized methodologies have led to the development of extensive lists of PTM sites, shifting the bottleneck towards the PTM confirmation, quantification and functional interpretation [58-60].

1.1. Proteolysis

1.1.1. Proteolysis in biological systems (based on Rogers *et al.*)

Proteolysis, i.e. the cleavage of proteins into smaller polypeptides by a protease, is often underappreciated as a PTM since the breakdown of proteins is not only irreversible but also omnipresent in biological systems [61]. Proteases regulate a wide range of physiological processes, going from food digestion to more strictly managed blood clotting [62].

The proteolytic processing can activate or deactivate proteins and even be the driving force behind the metamorphosis of a protein function as for example extracellular matrix proteins can be

transformed into growth factors. Also, in protein synthesis the removal of the N-terminal methionine is an essential step in the maturation and secretion of proteins. In addition, the removal of localization sequences, pre- and pro-domains can regulate the protein migration or receptor substrate activity. Due to interaction with protease inhibitors, other PTMs and stimuli originating from its localization, proteases control many cellular mechanisms including DNA replication, cell cycle progression, cell proliferation and cell death [61, 63].

Proteolysis is thus more than a turnover mechanism of proteins that aims to manage the cellular homeostasis. As proteolysis is associated with a variety of pathologies including inflammation, cancer, arthritis and cardiovascular diseases, tracking disruptions of this regulatory PTM is becoming increasingly important. Over 570 genes are predicted to code for proteases, making them one of the largest enzyme families in human [64]. However, although more than 30 drugs in the clinic already target proteases, only a minority of these enzymes has been comprehensively characterized [61, 63].

1.1.2. Challenges and considerations: histone proteolysis as a peculiar example

1.1.2.1. General considerations

Since half of the known human proteases still have unknown substrates, approaches to track proteolytic products are still emerging. In contrast to other PTMs like phosphorylation, proteolyzed proteins have no unique chemical signature that can be targeted for a specific enrichment by affinity chromatography [65]. Most techniques rely on the formation of newly formed N-termini but these procedures have limitations since approximately 85% of soluble proteins are N-terminally acetylated in eukaryotic cells [66]. These strategies are also challenged since primary amines are present at both N-termini and lysines which results in the contamination of abundant internal fragments when blocking is incomplete [61]. Nevertheless, several methodologies have been developed successfully for large-scale screening, which led to the identification of hundreds of cleavage sites in one run. To distinguish the relevant cleavage sites from the background, it is however mandatory that protease research evolves from qualitative to quantitative data [62]. Quantitative technologies can help complete the biological picture and can differentiate between substrates prone to *in vivo* protease activity and substrates only observed *in vitro*. Obviously, the experimental conditions need to be controlled to avoid *in vitro* proteolysis as this severally hampers the accurate understanding of the biology.

1.1.2.2. Histone proteolysis

In line with chapters 3 - parts 1 and 2, we here discuss in more detail some of the considerations that should be taken into account when studying (histone) proteolysis. The focus of the literature

regarding this subject is quite scattered and was only recently reviewed for the first time, illustrating that drawing biological conclusions on proteolysis is a hard task, even when certain cleaving sites are unambiguously identified [67, 68].

Histone proteins are associated with the DNA, together forming the nucleosome. Both DNA and the histone units are subjected to reversible PTMs that act as transcriptional regulators, called epigenetic modifications [69]. Proteolytical activities on histones were already described in the seventies and not surprisingly, these reports alluded on the epigenetic potential of such irreversible PTM from the beginning [70]. While evidence of the epigenetic role is rather gradually emerging, especially with the increased focus on histone PTMs the last decade, it is also becoming clear that histones are being degraded at high rates [71]. As histones are accumulating at the time of DNA synthesis, proteolytic processing can remove the excess of histones. Histone degradation is also described as a repair mechanism for damaged DNA-histone crosslinks or for the removal of wrongly incorporated histone-variants, followed by the re-establishing of new histones with fresh epigenetic marks [72]. However, histone proteolysis is equally presented as a reaction upon infection [73], in spermatogenesis [74], differentiation [75]… in a variety of species and as a result of different enzymes .

When overviewing the literature, we discriminate two main classes: “histone degradation” and “histone clipping” (framework adapted from [68]). The incoherence of all those reports that causes the lack of a consensus is in part due to the fact that (i) many reports only briefly mention detecting histone fragments, but never pursued their origination, (ii) many reports may potentially have observed *in vitro* side effects, blurring the biological context, (iii) the same enzymes can mediate both histone degradation and histone clipping and (iv) most reports struggle to elucidate the biology of the proteolysis event under investigation.

1.1.3. Technical approaches (based on Rogers *et al.*)

Due to the biological implication of proteolytical PTMs, degradomics and terminomics have now emerged as specialized fields in proteomics focusing on the analysis of proteolytical products and the N- and C-termini [61]. Although none of the specialized approaches were used in this dissertation, a brief overview of common methodologies is provided here for completeness.

1D or 2D electrophoresis of control and proteolyzed samples will show different spots after staining, which can be excised and identified by LC-MS/MS. The sensitivity of gel-based methods is however limited and not suited to detect small cleavage fragments. Alternative SDS-PAGE approaches have been developed, for example based on in-gel proteolysis after a first separation, followed by an

additional step of electrophoresis in another dimension. Proteins that are migrated off the diagonal are likely targeted by the used protease [65].

To characterize the active site specificity several array- and library-based methods can be applied. For microarrays, peptide libraries are incubated with a test protease, resulting in the release of a fluorophore or fluorescent quencher [76]. However, most methods rely on MS to identify the new N- and C-termini, formed as a result of proteolytic processing. Typically, an equal amount of protease treated and control sample are subjected to isotopic (SILAC) or isobaric (iTRAQ) labeling resulting in detection of neo-N-terminal peptides only present in the protease sample. Multiple variations of this strategy are applied, based on the negative or positive selection of N-terminal peptides [62]. E.g. with PICS (Proteomic identification of protease cleavage sites) the primary amines of peptides derived from the digestion of a proteome sample are blocked by reductive methylation, followed by incubation with a test protease. The newly generated N-termini are the result of the proteases and are tagged by biotin to be selected by an additional purification step prior to identification [77].

As most methods deplete the internal fragments the proteins are often represented by only a few (terminal) peptides, making the sample less complex but coordinately challenging identification. Not all N- or C-terminal peptides are easy to identify by LC-MS/MS due to their length or physicochemical properties hampering ionization or fragmentation. These problems are addressed, for instance by adding an extra derivatization step that increases the mass of the peptide or by employing alternative fragmentation strategies such as ETD [78, 79]. As the peptides of interest in large-scale proteolytic studies are often labeled and blocked, parameters for DB searches need to be well considered. To provide an overview of the high-throughput studies, public databases are available. MEROPS (http://merops.sanger.ac.uk) is best known and classifies information about the proteases, cleavage sites and protease inhibitors [80].

1.2. Phosphorylation

1.2.1. Phosphorylation in biological systems

Protein phosphorylation, the addition of a phosphate group to the proteins, is one of the most widely spread PTMs. Phosphorylation is catalyzed by the enzymes of the kinase family and removed by phosphatases [81]. The most common targets are the amino acids serine, threonine or tyrosine (respectively phosphorylated according to a suggested 88-11-1% ratio, Figure 10), but phosphorylation of histidine and aspartate are also known. Histidine is thought to account for up to 6% of all protein phosphorylations but is labile in an acidic environment and disappears from the radar in most phosphoproteomic studies [82].

Phosphorylation is transient and induces conformational changes in the target protein or blocks active sites causing alteration of the protein activity and protein interaction. The reversible character allows to regulate protein activation / deactivation by changing the phosphorylation / dephosphorylation status within a short time frame as a reaction to cellular or environmental stimuli, making this PTM one of the fundamentals of intracellular signaling.

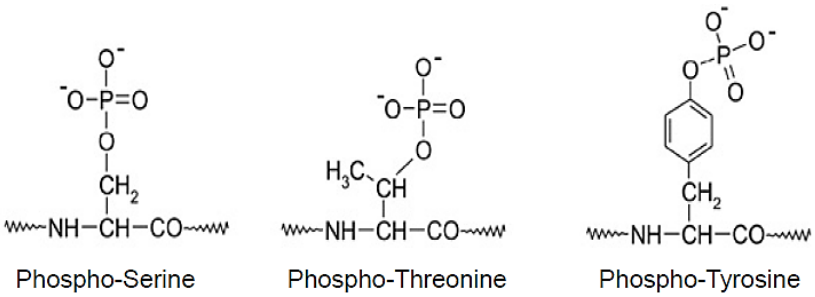


Figure 10: Chemical scheme (www.piercenet.com) of the most commonly phosphorylated amino acids**.**

Since phosphorylation is essential for biological communication, many cellular processes such as the cell metabolism, growth, differentiation and proliferation are regulated by phosphorylation pathways. It has been predicted that over 40% of all the proteins are being phosphorylated and that over 100 000 phosphosites exist in the human proteome. Evenly, 2-3% of all genes code for kinases and more than a hundred for phosphatases [81, 83]. This ubiquitous role of phosphorylation makes it one of the most studied modifications, reflected in the vast amount of papers on “phosphorylation” published each year (Figure 11, based on http://www.ncbi.nlm.nih.gov/pubmed).

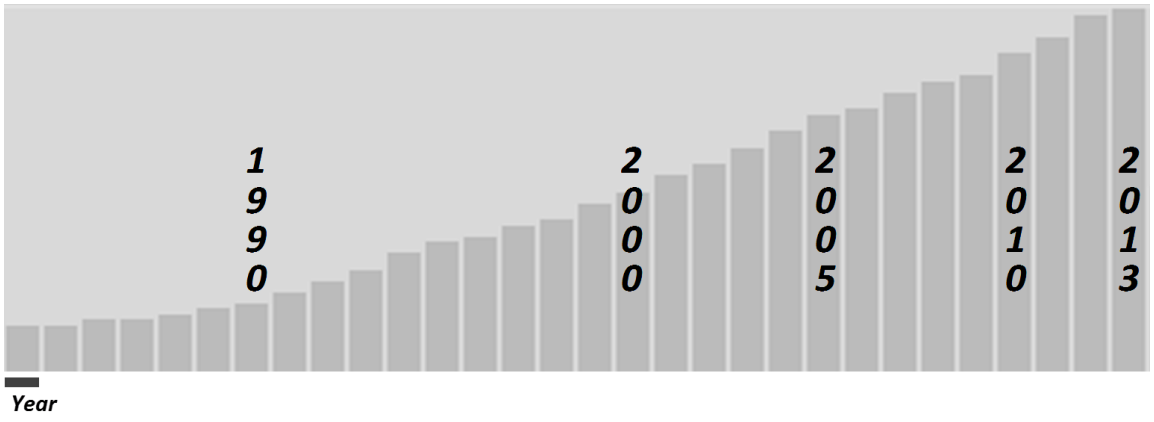


Figure 11: Relative presentation of the number of PubMed results by year for “phosphorylation”. *1990*; 2802, *2000*; 6974, *2005*; 9985, *2010*; 12443, *2013*; 14279. *2014* is still ongoing at the time of publication.

Mutations or overexpression of the enzyme families controlling the signal transduction pathways can perturb the regulation of dynamic phosphorylation, a process that is often observed in developmental disorders and cancer [84]. This is for instance illustrated by the mitogen activated protein kinase (MAPK) pathways, which are deregulated in many malignancies. For example, extracellular signal regulated protein kinases (ERK) of the MAPK family phosphorylate cytoplasmic proteins that migrate to the nucleus and trigger transcription factors involved in cellular proliferation. Irregular signaling of the MAPK/ERK pathway has been described in prostate, breast and cervical cancer [85]. In addition, erratic phosphorylation patterns have been observed in inflammation as well as in immunodeficiencies (e.g. in leukemias) [86]. Assessing the phosphorylation landscape and the modulating kinases is thus of major importance to understand pathologies and to develop new therapies. This is illustrated by the increasing amount of kinase inhibitors entering the pharmaceutical market.

1.2.2. Challenges and considerations

To map protein phosphorylation dynamics the main analytical workhorse is once more the mass spectrometer, although a number of issues complicate the phosphosite identification and data interpretation. First, the number of phosphopeptides is relatively low compared to the peptides that are not phosphorylated. As many phosphorylated peptides are of low stoichiometry, i.e. only a small fraction of a particular peptide is phosphorylated, it is difficult to discover the phosphorylated sequences within the bulk of the unmodified peptides [87]. The phosphate group is also relatively unstable, a characteristic that must be taken into account during the manipulations prior to the LC-MS/MS analysis (e.g. during protein extraction from a biological sample). At the time of fragmentation, the labile phosphate moiety is often released from its peptide precursor at the expense of the fragmentation of the backbone. This results in MS/MS spectra containing less information on the peptide sequence, leading to less confident annotations. Even if a phosphopeptide is detected, it is still challenging to identify the exact site of the peptide that is phosphorylated [88]. Peptides with the same sequence, phosphorylated on a different residue, have often many fragments with the same mass, as only fragments derived from breaking points between two sites are phosphosite-specific. This phenomenon becomes especially relevant in the case of neighboring or multi-phosphorylated peptides. Moreover, bias is introduced as the fragmentation favors certain breaking points over others. To map sites in all the regions of the protein it is advised to digest with different enzymes to aim for maximum protein sequence coverage [87].

Both the broad large-scale mapping and specific phosphosite analysis help to unravel the primordial questions in phosphoproteomics: What is the biochemical, cellular and overall biological

function of the phospho-events? Experiments designed to detect a maximum of phosphosites in a single analysis have proven their value for mapping the signal networks [82]. However, these studies often provide a minimal confidence of the peptide annotations and lack detailed information of the unique phosphosites and the responsible kinases. Frequently, the phosphorylation of a protein is regulated by multiple kinases, which are often interconnected as being part of the same or an alternating signaling cascade. To interpret this tangled web, a combination of *in vivo* and *in vitro* data is required to read the activity of each individual kinase. In the experimental assay the relationship between the *in vitro* and *in vivo* conditions must be controlled as high concentrations of a purified kinase *in vitro*, prolonged incubation times and the absence of endogenous phosphatases can result in low-level background phosphorylation [87].

Although for large-scale studies thousands of phosphosites are described, it is unclear which of these display actual biological functions. Especially since optimized methods for phosphopeptide enrichment together with adapted MS acquisition workflows have enabled the identification of phosphosites of such a low stoichiometry that may be meaningless. Information of the stoichiometry is thus important and can help to distinguish the relevant phosphorylation sites from the trivial identifications [87, 89]. Further, knowledge of the protein abundance is complementary to the stoichiometry measurements and is key to distinguish the promiscuous from the functional phosphorylation [90]. All these factors stress the importance to cultivate quantitative methodologies to obtain information on multiple levels: protein, phosphopeptide and non-phosphopeptide.

1.2.3. Technical approaches

A variety of (complementary) techniques have been developed to address the analytical difficulties in phosphoproteomics. Here, we present a short overview of the most established experimental designs, based on [81-85, 87, 88, 91]. For phosphopeptide quantification, a pillar becoming progressively important in the field, most strategies are a direct application of the “standard” methods in quantitative mass spectrometry already describe above.

1.2.3.1. Sample prep and phospho-enrichment

To preserve the integrity of the phosphosites, appropriate protease and phosphatase inhibitors must be used during the cell lysis and protein extraction. Lysis must be performed at low temperatures or cell pellets can be snap-frozen or heat inactivated [92]. Due to the low abundance of phosphopeptides (<5%), phosphoproteomic research requires 50-100 times more material than standard LC-MS/MS proteomics workflows [84, 85].

As a first step, several gel-based techniques can be applied to select phosphoproteins out of the protein pool before the MS identification e.g. by an electrophoresis mobility shift, although not all phosphorylation events cause a change in mobility. To achieve a better separation between phosphorylated and non-phosphorylated proteins within the gel, phosphate-binding molecules can be incorporated into the polyacrylamide gel (Phos-Tag) as these tags decrease the migration speed of phosphorylated proteins [87].

To cope with the low proportion of phosphopeptides in a protein digest, most workflows rely on the specific enrichment of phosphopeptides prior to the MS-analysis. LC-enrichment based on ion exchange or hydrophobic interactions are being used to increase the selectivity of the phospho detection. The most popular enrichment methods apply immobilized metal affinity chromatography (IMAC) with Fe^{3+} or oxide affinity chromatography (MOAC) with TiO_2. Numerous variations of these methods are described e.g. focusing on the tyrosine purification or combing IMAC and MOAC, as reviewed in [93].

1.2.3.2. Mass spectrometry of phosphopeptides

Multiple MS acquisition strategies have been developed to facilitate the identification of phosphorylations. Compared to CID, higher energy C-trap dissociation (HCD) of the orbitrap analyzer provides more extensive fragmentation and results in better quality spectra with a higher resolution and lower *m/z* cutoff. The advantage of HCD over CID for the large-scale phospho-analysis is still quite controversial as Mann *et al.* were able to detect more phosphopetides with HCD as with CID as opposed to Gygi and co-workers who identified twice the amount of peptides with CID compared to HCD. Despite the higher scoring peptides obtained with HCD, CID was superior due to the faster acquisition [88].

The ion of the peptide backbone that is generated after the neutral loss of the phosphate group can also be selected for an additional fragmentation in the MS3 mode. If the selected peptide backbone is abundant, the extra energy results in a useful fragmentation spectrum. With MultiStage Activation, a pseudo MS3 approach, the MS2 and MS3 data are combined in a hybrid spectrum containing confident information of the peptide sequence and phosphosite. Alternatively, phosphopeptides are fragmented by ECD or ETD. With ECD, thermal electrons and with ETD, radical ions disrupt the backbone of multiply protonated peptide to produces c- and z-type ions with fragments still containing the phosphate group. This data is different compared to CID, making peptide matching and scoring with the classic search engines difficult. This has led to the development of adapted informatics procedures [88]. As ECD and ETD are less sensitive than CID, it is recommended to apply CID for exploratory studies and ETD for validating exact phosphosites [85]. The combination of both fragmentation strategies is becoming increasingly popular although algorithms that are able to cope with the merging spectra of CID and ETD

still need to improve [88].

For the interpretation of the data, multiple computational tools have been developed to address a variety of specific questions in phosphoproteomics. To cope with localizing phosphate groups, e.g. Ascore or Phosphoscore have been developed to assign a probability score to the possible phosphosites. Publicly accessible databases with earlier identified phosphopeptides (like PHOSIDA) are available for multiple species. Other software recognizes sequence patterns, the so-called “motifs” that are correlated with phosphorylation databases. To further interpret results of a phosphoproteomic study, other tools modulate the relationship between kinases and substrates, visualize interacting networks or classify lists of proteins [82].

2. Cancer cells: a common model in molecular biology

2.1. Cell lines

In proteomics studies, either aiming at optimizing specific methodologies or focusing on specific biological questions such as particular protein functions or protein-protein interactions, a substantial amount of work is done on cultured cell lines. Most cell lines are derived from human tumors, extracted from different tissues and adapted for growth in culture. Mammalian cell lines not only help to generate insights in the cellular processes of the tissue/cancer of origin but also contribute to answer more fundamental biological questions [94]. From a pragmatic point of view, cell lines are convenient as conditions for culturing and harvesting can be controlled, which is extremely valuable for comparative studies where only the factor of interest is varied between cells that are grown completely similar. Moreover, the population of cultured cells is relatively pure, increasing the experimental signal to noise. Theoretically, cell lines provide an unlimited source of "working material", convenient for protein studies which require a lot of material compared to genetic studies. Typically, new or freshly adapted methodologies are first optimized on protein extracts derived from cell lines before patient material is being tested [9, 95, 96].

In this dissertation, a variety of cell lines were applied across the results (chapter 3). Some were picked randomly, others were chosen as popular models in that particular niche of proteomics. In chapter 3 - part 1, THP-1 cells are used, which are derived from acute monocytic leukemia. These cells were already described in the context of H2A clipping which we pursued in a variety of cells [97, 98]. In part 2, the Raji and Jurkat cell lines originate from B-lymphocytes of a Burkitt's lymphoma and immortalized T cell leukemia, respectively. These cells are a valuable and "pure" alternative for the hematopoietic B- and T-cells under investigation [99]. In chapter 3 - part 3, HeLa cells are used, one of the oldest and popular cell lines and very commonly used in phosphoproteomic research. The cells, originated from a cervical cancer, are frequently stimulated with the epidermal growth factor (EGF), a reagent that binds to the EGF-receptor and triggers downstream phosphorylation pathways [100]. Since the phosphopeptides identified in these cells are the basis of the Phosida DB, we applied the HeLa cell line to optimize and validate the technique [101]. In chapter 3 - part 4, HepG2 cells from a hepatocellular carcinoma were chosen as these adhesive cells were likely to have a higher amount of membrane proteins than cells in suspension, a favorable characteristic for the technical approach that was optimized [102].

2.2. Chronic lymphocytic leukemia (based on Rodríguez-Vincente *et al.* and Pekarsky *et al.*)

The most common adult hematopoietic cancer in the Western world is Chronic Lymphocytic Leukemia (CLL), affecting patients >50 years of age. Morphologically, the disease is characterized by the accumulation of B-lymphocytes in the blood, bone marrow and lymph nodes although patients are often diagnosed accidentally. With survival rates varying from a couple of months to decades, CLL is very unpredictable. Initially no symptoms occur but if the disease progresses, patients suffer from hypogammaglobulinemia and immunodeficiency [103]. 3-10% of the patients develop an aggressive lymphoma (Richters syndrome) and CLL may also transform into prolymphocytic leukemia, which has a worse outcome. Previously it was accepted that an inherent defect made the CLL cells insusceptible to apoptosis and causes the accumulation, leading to the high cell count. The last years, this consensus has been challenged since simultaneously active proliferation of CLL cells was observed in the bone marrow, spleen and lymph nodes [104].

For understanding the remarkable diversity of the disease and for the discovery of accurate diagnostic and therapeutic markers the molecular basis of CLL is extensively studied. The discovery that immunoglobulin heavy chain (*IGVH*) genes encoding for the B-cell antigen-binding domain in CLL can be either mutated or unmutated contributed to the understanding of CLL and its prognosis. CLL patients who have B-cells with unmutated (UM, >98% germ line identity) *IGVH* genes have an unfavorable outcome, whereas mutated (M) *IGVH* genes predict a more indolent course. It is hypothesized that CLL B-cells with the more conserved UM *IGVH* sequences are more susceptible to the repetitive interaction between (auto) antigens and the polyreactive B-cell receptor than B-cells with M *IGVH* genes. This antigen stimulation promotes clonal growth and contributes to the inferior clinical outcome of patients with UM *IGVH* genes. An UM *IGVH* gene status is commonly accompanied by elevated levels of Zeta-chain-associated protein kinase 70 (ZAP-70). ZAP-70, normally expressed in T- and NK-cells, plays a role in initiating T-cell signaling but also in signaling through the B-cell receptor of CLL cells. As ZAP-70 promotes the B-cell receptor signaling, the level of ZAP-70 is predictive for a poorer outcome and serves as an independent biomarker in CLL [103-105].

Together with other serum markers, the IGHV status and ZAP-70 expression are popular in the clinic for the staging of the patients [105]. Alternative prognostic and biological markers remain however needed to understand the biomolecular heterogeneity. Cytogenetic studies have shown that chromosomal abnormalities are also a hallmark in CLL and serve as independent biomarkers, although 10-20% of the cases do not display common aberrations. Mutations in tumor suppressor genes such as Ataxia telangiectasia mutated (ATM), are also associated with an aggressive outcome and resistance to therapeutic agents. The relationship between the cytogenetic abnormalities and miRNA (microRNA,

being non-coding RNAs) is an interesting field in CLL research as deregulated miRNAs correlate with the established IGHV gene and ZAP-70 prognostic markers. More recently, the role of the epigenetic regulation in the progression in the disease pathology became more established. In CLL patients a global hypomethylation is observed compared to healthy but on the other hand, tumor suppressor genes are silenced by local hypermethylation [104, 106]. These epigenetic alterations have the potential to become popular therapeutic targets e.g. several histone deacetylase inhibitors for CLL treatment are tested in clinical studies [107].

References

[1] Pray, L. A., Discovery of DNA Structure and Function: Watson and Crick. *Nature Education* 2008, *1*, 100.

[2] Aebersold, R., Mann, M., Mass spectrometry-based proteomics. *Nature* 2003, *422*, 198-207.

[3] Mathew, C. K., Van Holde, K. E., Ahern, K. G., *Biochemistry* Heyden, Robin 2000.

[4] Ezkurdia, I. J., D.; Rodriguez, J.M.; Frankish, A.; Diekhans, M.; Harrow, J.; Vazquez J.; Valencia, A.; Tress, M.L., The shrinking human protein coding complement: are there now fewer than 20,000 genes? . *arXiv.org* 2014.

[5] Smith, L. M., Kelleher, N. L., Consortium for Top Down, P., Proteoform: a single term describing protein complexity. *Nature methods* 2013, *10*, 186-187.

[6] Patterson, S. D., Aebersold, R. H., Proteomics: the first decade and beyond. *Nat Genet* 2003, *33 Suppl*, 311-323.

[7] Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D*., et al.*, Progress with proteome projects: Why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng* 1996, *13*, 19-50.

[8] Roepstorff, P., Mass spectrometry based proteomics, background, status and future needs. *Protein & cell* 2012, *3*, 641-647.

[9] Mallick, P., Kuster, B., Proteomics: a pragmatic perspective. *Nat Biotechnol* 2010, *28*, 695-709.

[10] Zhong, Y., Hyung, S. J., Ruotolo, B. T., Ion mobility-mass spectrometry for structural proteomics. *Expert review of proteomics* 2012, *9*, 47-58.

[11] Thomson, J. J., Rays of positive electricity. *Proceedings of the Royal Society* 1913, 1-20.

[12] Grayson, M. A., *Measuring Mass: From Positive Rays to Proteins*, Chemical Heritage Foundation 2002.

[13] Altelaar, A. F., Munoz, J., Heck, A. J., Next-generation proteomics: towards an integrative view of proteome dynamics. *Nature reviews. Genetics* 2013, *14*, 35-48.

[14] Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, *20*, 3551-3567.

[15] Craig, R., Beavis, R. C., TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 2004, *20*, 1466-1467.

[16] Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L*., et al.*, Open mass spectrometry search algorithm. *J Proteome Res* 2004, *3*, 958-964.

[17] Gonzalez-Galarza, F. F., Lawless, C., Hubbard, S. J., Fan, J*., et al.*, A critical appraisal of techniques, software packages, and standards for quantitative proteomic analysis. *Omics : a journal of integrative biology* 2012, *16*, 431-442.

[18] Picotti, P., Aebersold, R., Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature methods* 2012, *9*, 555-566.

[19] Armirotti, A., Damonte, G., Achievements and perspectives of top-down proteomics. *Proteomics* 2010, *10*, 3566-3576.

[20] Angel, T. E., Aryal, U. K., Hengel, S. M., Baker, E. S*., et al.*, Mass spectrometry-based proteomics: existing capabilities and future directions. *Chemical Society reviews* 2012, *41*, 3912-3928.

[21] Belov, M. E., Clowers, B. H., Prior, D. C., Danielson, W. F., 3rd*, et al.*, Dynamically multiplexed ion mobility time-of-flight mass spectrometry. *Analytical chemistry* 2008, *80*, 5873-5883.

[22] Shah, A. R., Agarwal, K., Baker, E. S., Singhal, M*., et al.*, Machine learning based prediction for peptide drift times in ion mobility spectrometry. *Bioinformatics* 2010, *26*, 1601-1607.

[23] Shapiro, E., Biezuner, T., Linnarsson, S., Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nature reviews. Genetics* 2013, *14*, 618-630.

[24] Chiu, D. T., Lorenz, R. M., Jeffries, G. D., Droplets for ultrasmall-volume analysis. *Analytical chemistry* 2009, *81*, 5111-5118.

[25] van den Broek, I., Niessen, W. M., van Dongen, W. D., Bioanalytical LC-MS/MS of protein-based biopharmaceuticals. *J Chromatogr B Analyt Technol Biomed Life Sci* 2013, *929*, 161-179.

[26] Nesvizhskii, A. I., Aebersold, R., Interpretation of shotgun proteomic data: the protein inference problem. *Molecular & cellular proteomics : MCP* 2005, *4*, 1419-1440.
[27] Evans, C., Noirel, J., Ow, S. Y., Salim, M.*, et al.*, An insight into iTRAQ: where do we stand now? *Anal Bioanal Chem* 2012, *404*, 1011-1027.
[28] Petricoin, E. F., Zoon, K. C., Kohn, E. C., Barrett, J. C., Liotta, L. A., Clinical proteomics: translating benchside promise into bedside reality. *Nature reviews. Drug discovery* 2002, *1*, 683-695.
[29] Shevchenko, A., Loboda, A., Ens, W., Schraven, B.*, et al.*, Archived polyacrylamide gels as a resource for proteome characterization by mass spectrometry. *Electrophoresis* 2001, *22*, 1194-1203.
[30] Aguilar, M.-I., Humana Press Inc. 2004.
[31] Kawashima, Y., Takahashi, N., Satoh, M., Saito, T.*, et al.*, Enhanced recovery of lyophilized peptides in shotgun proteomics by using an LC-ESI-MS compatible surfactant. *Proteomics* 2013, *13*, 751-755.
[32] Gumustas, M., Kurbanoglu, S., Uslu, B., Ozkan, S. A., UPLC versus HPLC on Drug Analysis: Advantageous, Applications and Their Validation Parameters. *Chromatographia* 2013, *76*, 1365-1427.
[33] Hoffmann, D. S., V. , *Mass Spectrometry: Principles and Applications*, 2007.
[34] Cho, Y. T., Su, H., Huang, T. L., Chen, H. C.*, et al.*, Matrix-assisted laser desorption ionization/time-of-flight mass spectrometry for clinical diagnosis. *Clinica chimica acta; international journal of clinical chemistry* 2013, *415*, 266-275.
[35] Hu, Q., Noll, R. J., Li, H., Makarov, A.*, et al.*, The Orbitrap: a new mass spectrometer. *Journal of mass spectrometry : JMS* 2005, *40*, 430-443.
[36] Reinders J., S. A., Martens L., Apweiler R., *Proteomics, Methods in Molecular Biology, Vol 564*, 2009.
[37] Bantscheff, M., Lemeer, S., Savitski, M. M., Kuster, B., Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* 2012, *404*, 939-965.
[38] Martens, L., Nesvizhskii, A. I., Hermjakob, H., Adamski, M.*, et al.*, Do we want our data raw? Including binary mass spectrometry data in public proteomics data repositories. *Proteomics* 2005, *5*, 3501-3505.
[39] Magrane, M., Consortium, U., UniProt Knowledgebase: a hub of integrated protein data. *Database : the journal of biological databases and curation* 2011, *2011*, bar009.
[40] Nesvizhskii, A. I., Vitek, O., Aebersold, R., Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature methods* 2007, *4*, 787-797.
[41] Matthiesen, R., Methods, algorithms and tools in computational proteomics: a practical point of view. *Proteomics* 2007, *7*, 2815-2832.
[42] Vaudel, M., Sickmann, A., Martens, L., Peptide and protein quantification: a map of the minefield. *Proteomics* 2010, *10*, 650-670.
[43] Hultin-Rosenberg, L., Forshed, J., Branca, R. M., Lehtio, J., Johansson, H. J., Defining, comparing, and improving iTRAQ quantification in mass spectrometry proteomics data. *Molecular & cellular proteomics : MCP* 2013, *12*, 2021-2031.
[44] Ong, S. E., Mann, M., A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat Protoc* 2006, *1*, 2650-2660.
[45] Schwanhausser, B., Gossen, M., Dittmar, G., Selbach, M., Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* 2009, *9*, 205-209.
[46] Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R., Mann, M., Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature methods* 2010, *7*, 383-385.
[47] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F.*, et al.*, Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 1999, *17*, 994-999.
[48] Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S., Heck, A. J., Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protoc* 2009, *4*, 484-494.
[49] Mertins, P., Udeshi, N. D., Clauser, K. R., Mani, D. R.*, et al.*, iTRAQ labeling is superior to mTRAQ for quantitative global proteomics and phosphoproteomics. *Molecular & cellular proteomics : MCP* 2012, *11*, M111 014423.
[50] Casado-Vela, J., Martinez-Esteso, M. J., Rodriguez, E., Borras, E.*, et al.*, iTRAQ-based quantitative analysis of protein mixtures with large fold change and dynamic range. *Proteomics* 2010, *10*, 343-347.

[51] Christoforou, A., Lilley, K. S., Taming the isobaric tagging elephant in the room in quantitative proteomics. *Nature methods* 2011, *8*, 911-+.
[52] Ow, S. Y., Salim, M., Noirel, J., Evans, C.*, et al.*, iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". *J Proteome Res* 2009, *8*, 5347-5355.
[53] Karp, N. A., Huber, W., Sadowski, P. G., Charles, P. D.*, et al.*, Addressing Accuracy and Precision Issues in iTRAQ Quantitation. *Molecular & Cellular Proteomics* 2010, *9*, 1885-1897.
[54] Shadforth, I. P., Dunkley, T. P., Lilley, K. S., Bessant, C., i-Tracker: for quantitative proteomics using iTRAQ. *BMC Genomics* 2005, *6*, 145.
[55] Vaudel, M., Burkhart, J. M., Radau, S., Zahedi, R. P.*, et al.*, Integral quantification accuracy estimation for reporter ion-based quantitative proteomics (iQuARI). *J Proteome Res* 2012, *11*, 5072-5080.
[56] Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., Gygi, S. P., Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* 2003, *100*, 6940-6945.
[57] Collins, B. C., Gillet, L. C., Rosenberger, G., Rost, H. L.*, et al.*, Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nature methods* 2013, *10*, 1246-1253.
[58] Parker, C. E. M., V.; Mocanu, M.; Dicheva, N.; Warren., M.R., *Mass Spectrometry for Post-Translational Modifications*, 2010.
[59] Jensen, O. N., Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Current opinion in chemical biology* 2004, *8*, 33-41.
[60] Witze, E. S., Old, W. M., Resing, K. A., Ahn, N. G., Mapping protein post-translational modifications with mass spectrometry. *Nature methods* 2007, *4*, 798-806.
[61] Rogers, L. D., Overall, C. M., Proteolytic post-translational modification of proteins: proteomic tools and methodology. *Molecular & cellular proteomics : MCP* 2013, *12*, 3532-3542.
[62] Impens, F., Colaert, N., Helsens, K., Plasman, K.*, et al.*, MS-driven protease substrate degradomics. *Proteomics* 2010, *10*, 1284-1296.
[63] Turk, B., Targeting proteases: successes, failures and future prospects. *Nature reviews. Drug discovery* 2006, *5*, 785-799.
[64] Quesada, V., Ordonez, G. R., Sanchez, L. M., Puente, X. S., Lopez-Otin, C., The Degradome database: mammalian proteases and diseases of proteolysis. *Nucleic Acids Res* 2009, *37*, D239-243.
[65] Agard, N. J., Wells, J. A., Methods for the proteomic identification of protease substrates. *Current opinion in chemical biology* 2009, *13*, 503-509.
[66] Polevoda, B., Sherman, F., Nalpha -terminal acetylation of eukaryotic proteins. *J Biol Chem* 2000, *275*, 36479-36482.
[67] Azad, G. K., Tomar, R. S., Proteolytic clipping of histone tails: the emerging role of histone proteases in regulation of various biological processes. *Mol Biol Rep* 2014, *41*, 2717-2730.
[68] Dhaenens, M., Glibert, P., Meert, P., Vossaert, L., Deforce, D., Histone proteolysis: framing a fuzzy picture. *manuscript 'in preparation'* 2014.
[69] Jenuwein, T., Allis, C. D., Translating the histone code. *Science* 2001, *293*, 1074-1080.
[70] Eickbush, T. H., Watson, D. K., Moudrianakis, E. N., A chromatin-bound proteolytic activity with unique specificity for histone H2A. *Cell* 1976, *9*, 785-792.
[71] Deal, R. B., Henikoff, J. G., Henikoff, S., Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* 2010, *328*, 1161-1164.
[72] Gunjan, A., Paik, J., Verreault, A., The emergence of regulated histone proteolysis. *Curr Opin Genet Dev* 2006, *16*, 112-118.
[73] Brinkmann, V., Reichard, U., Goosmann, C., Fauler, B.*, et al.*, Neutrophil extracellular traps kill bacteria. *Science* 2004, *303*, 1532-1535.
[74] Qian, M. X., Pang, Y., Liu, C. H., Haratake, K.*, et al.*, Acetylation-mediated proteasomal degradation of core histones during DNA repair and spermatogenesis. *Cell* 2013, *153*, 1012-1024.
[75] Vossaert, L., Meert, P., Scheerlinck, E., Glibert, P.*, et al.*, Identification of histone H3 clipping activity in human embryonic stem cells. *Stem cell research* 2014, *13*, 123-134.

[76] Salisbury, C. M., Maly, D. J., Ellman, J. A., Peptide microarrays for the determination of protease substrate specificity. *J Am Chem Soc* 2002, *124*, 14868-14870.
[77] Schilling, O., Overall, C. M., Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat Biotechnol* 2008, *26*, 685-694.
[78] Frese, C. K., Altelaar, A. F., Hennrich, M. L., Nolting, D*., et al.*, Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J Proteome Res* 2011, *10*, 2377-2388.
[79] Gevaert, K., Goethals, M., Martens, L., Van Damme, J*., et al.*, Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat Biotechnol* 2003, *21*, 566-569.
[80] Rawlings, N. D., Morton, F. R., Kok, C. Y., Kong, J., Barrett, A. J., MEROPS: the peptidase database. *Nucleic Acids Res* 2008, *36*, D320-325.
[81] Thingholm, T. E., Jensen, O. N., Larsen, M. R., Analytical strategies for phosphoproteomics. *Proteomics* 2009, *9*, 1451-1468.
[82] Jünger, M. A. A., R., Mass spectrometry-driven phosphoproteomics: patterning the systems biology mosaic. *Dev Biol* 2014, *3*, 83-112.
[83] Tichy, A., Salovska, B., Rehulka, P., Klimentova, J*., et al.*, Phosphoproteomics: searching for a needle in a haystack. *Journal of proteomics* 2011, *74*, 2786-2797.
[84] Roux, P. P., Thibault, P., The coming of age of phosphoproteomics--from large data sets to inference of protein functions. *Molecular & cellular proteomics : MCP* 2013, *12*, 3453-3464.
[85] Lopez, E., Lopez, I., Ferreira, A., Sequi, J., Clinical and technical phosphoproteomic research. *Proteome Sci* 2011, *9*, 27.
[86] Patterson, H., Nibbs, R., McInnes, I., Siebert, S., Protein kinase inhibitors in the treatment of inflammatory and autoimmune diseases. *Clin Exp Immunol* 2014, *176*, 1-10.
[87] Dephoure, N., Gould, K. L., Gygi, S. P., Kellogg, D. R., Mapping and analysis of phosphorylation sites: a quick guide for cell biologists. *Mol Biol Cell* 2013, *24*, 535-542.
[88] Nilsson, C. L., Advances in quantitative phosphoproteomics. *Analytical chemistry* 2012, *84*, 735-746.
[89] Harvey, S. L., Enciso, G., Dephoure, N., Gygi, S. P*., et al.*, A phosphatase threshold sets the level of Cdk1 activity in early mitosis in budding yeast. *Mol Biol Cell* 2011, *22*, 3595-3608.
[90] Levy, E. D., Michnick, S. W., Landry, C. R., Protein abundance is key to distinguish promiscuous from functional phosphorylation based on evolutionary information. *Philos T R Soc B* 2012, *367*, 2594-2606.
[91] Olsen, J. V., Mann, M., Status of large-scale analysis of post-translational modifications by mass spectrometry. *Molecular & cellular proteomics : MCP* 2013, *12*, 3444-3452.
[92] Kofanova, O. A., Fack, F., Niclou, S. P., Betsou, F., Combined Effect of Tissue Stabilization and Protein Extraction Methods on Phosphoprotein Analysis. *Biopreserv Biobank* 2013, *11*, 161-165.
[93] Beltran, L., Cutillas, P. R., Advances in phosphopeptide enrichment techniques for phosphoproteomics. *Amino Acids* 2012, *43*, 1009-1024.
[94] Geiger, T., Wehner, A., Schaab, C., Cox, J., Mann, M., Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Molecular & cellular proteomics : MCP* 2012, *11*, M111 014050.
[95] Boyd, R. S., Dyer, M. J., Cain, K., Proteomic analysis of B-cell malignancies. *Journal of proteomics* 2010, *73*, 1804-1822.
[96] Zamo, A., Cecconi, D., Proteomic analysis of lymphoid and haematopoietic neoplasms: there's more than biomarker discovery. *Journal of proteomics* 2010, *73*, 508-520.
[97] Glibert, P., Vossaert, L., Van Steendam, K., Lambrecht, S*., et al.*, Quantitative Proteomics to Characterize Specific Histone H2A Proteolysis in Chronic Lymphocytic Leukemia and the Myeloid THP-1 Cell Line. *International journal of molecular sciences* 2014, *15*, 9407-9421.
[98] Minami, J., Takada, K., Aoki, K., Shimada, Y*., et al.*, Purification and characterization of C-terminal truncated forms of histone H2A in monocytic THP-1 cells. *Int J Biochem Cell Biol* 2007, *39*, 171-180.
[99] Dhaenens, M., Glibert, P., Lambrecht, S., Vossaert, L*., et al.*, Neutrophil Elastase in the capacity of the "H2A-specific protease". *Int J Biochem Cell Biol* 2014, *51*, 39-44.

[100] Olsen, J. V., Blagoev, B., Gnad, F., Macek, B.*, et al.*, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 2006, *127*, 635-648.
[101] Gnad, F., Gunawardena, J., Mann, M., PHOSIDA 2011: the posttranslational modification database. *Nucleic Acids Res* 2011, *39*, D253-260.
[102] Glibert, P., Van Steendam, K., Dhaenens, M., Deforce, D., iTRAQ as a method for optimization: enhancing peptide recovery after gel fractionation. *Proteomics* 2014, *14*, 680-684.
[103] Chiorazzi, N., Rai, K. R., Ferrarini, M., Chronic lymphocytic leukemia. *N Engl J Med* 2005, *352*, 804-815.
[104] Pekarsky, Y., Zanesi, N., Croce, C. M., Molecular basis of CLL. *Seminars in cancer biology* 2010, *20*, 370-376.
[105] Chiorazzi, N., Implications of new prognostic markers in chronic lymphocytic leukemia. *Hematology Am Soc Hematol Educ Program* 2012, *2012*, 76-87.
[106] Rodriguez-Vicente, A. E., Diaz, M. G., Hernandez-Rivas, J. M., Chronic lymphocytic leukemia: a clinical and molecular heterogenous disease. *Cancer genetics* 2013, *206*, 49-62.
[107] Wendtner, C. M., [Chemotherapy-free treatment of chronic lymphocytic leukemia?]. *Deutsche medizinische Wochenschrift* 2013, *138*, 2104-2106.

Chapter 3

RESULTS

PART 1: QUANTITATIVE PROTEOMICS TO CHARACTERIZE SPECIFIC HISTONE H2A PROTEOLYSIS IN CHRONIC LYMPHOCYTIC LEUKEMIA AND THE MYELOID THP-1 CELL LINE

Based on:

Quantitative Proteomics to Characterize Specific Histone H2A Proteolysis in Chronic Lymphocytic Leukemia and the Myeloid THP-1 Cell Line

Pieter Glibert [1], Liesbeth Vossaert [1], Katleen Van Steendam [1], Stijn Lambrecht [2], Filip Van Nieuwerburgh [1], Fritz Offner [3], Thomas Kipps [4], Maarten Dhaenens [1,†] and Dieter Deforce [1,†,*]

[1] Laboratory of Pharmaceutical Biotechnology, Ghent University, 460 Ottergemsesteenweg, B-9000 Ghent, Belgium; E-Mails: pieter.glibert@ugent.be (P.G.); liesbeth.vossaert@ugent.be (L.V.); katleen.vansteendam@ugent.be (K.V.S.); filip.vannieuwerburgh@ugent.be (F.V.N.); maarten.dhaenens@ugent.be (M.D.)

[2] Department of Rheumatology, Ghent University Hospital, 185 1P7 De Pintelaan, B-9000 Ghent, Belgium; E-Mail: stijn.lambrecht@hotmail.com

[3] Department of Hematology, Ghent University Hospital, 185 1P7 De Pintelaan, B-9000 Ghent, Belgium; E-Mail: fritz.offner@ugent.be

[4] Department of Medicine, Moores Cancer Center, University of California at San Diego (UCSD), 3855 Health Sciences Drive, La Jolla, CA 92093, USA; E-Mail: tkipps@ucsd.edu

† These authors contributed equally to this work.

* Author to whom correspondence should be addressed; E-Mail: dieter.deforce@ugent.be; Tel.: +32-9-264-8067; Fax: +32-9-220-6688.

Accepted for publication:

Int J Mol Sci. 2014 May 27;15(6):9407-21. doi: 10.3390/ijms15069407.

1. Abstract

Proteome studies on hematological malignancies contribute to the understanding of the disease mechanism and to the identification of new biomarker candidates. With the isobaric tag for relative and absolute quantitation (iTRAQ) method we analyzed the protein expression between B-cells of healthy people and chronic lymphocytic leukemia (CLL) B-cells. CLL is the most common lymphoid cancer of the blood and is characterized by a variable clinical course. By comparing samples of patients with an aggressive *vs.* indolent disease, we identified a limited list of differentially regulated proteins. The enhanced sensitivity attributed to the iTRAQ labels led to the discovery of a previously reported but still not clarified proteolytic product of histone H2A (cH2A) which we further investigated in light of the suggested functional properties of this modification. In the exploratory proteome study the Histone H2A peptide was up-regulated in CLL samples but a more specific and sensitive screening of a larger patient cohort indicated that cH2A is of myeloid origin. Our subsequent quantitative analysis led to a more profound characterization of the clipping in acute monocytic leukemia THP-1 cells subjected to induced differentiation.

2. Introduction

Proteomics approaches are often trailing genetic studies but are essential in the multi-disciplinary field of hematological research. As opposed to, e.g., mRNA microarray data, there is a better understanding of which proteins are actually expressed, although seeing the forest for the trees in long lists of protein identifications remains challenging [1,2]. In neoplastic hematology, protein studies have contributed to the elucidation of the disease mechanism, defined prognostic or therapeutic biomarkers and clarified previously reported uncharacterized phenomena [2]. By analyzing body fluids, cell lines, and tissues with quantitative high throughput mass spectrometry techniques complementary biological insights in hematopoietic malignancies can be generated. Uncovering relevant posttranslational modifications (PTMs), such as phosphorylations and proteolytical cleavages, might be associated with specific disease stages and could, hence, be informative on the biology of the disease [3].

The most common adult hematopoietic malignancy is chronic lymphocytic leukemia (CLL), a disease characterized by a widely variable median survival. After the initial staging of the patient, the risk of progression is defined by a set of genetic - and protein-based laboratory assays. A well-established prognostic marker is the mutational status of the immunoglobulin heavy chain (*IGVH*) genes encoding for the B-cell antigen-binding domain: CLL patients who have B-cells with unmutated (UM, >98% germ line identity) *IGVH* genes have an unfavorable outcome, whereas mutated (M) *IGHV* genes predict a more indolent course [4]. Surrogate markers on the protein level, such as the Zeta-chain-associated protein kinase (ZAP) 70 and CD38, are more easily applicable in the clinical practice although CD38 is considered to have less predictive value since discordancy with gene status is commonly observed and problems with standardization occur. ZAP70 expression however, is still used in the research on CLL pathogenesis and is considered as an independent biomarker [5]. Hence, for determining disease progression and survival of CLL patients, specific genetic markers, e.g., chromosomal aberrations, such as 13q14 deletion, are of increasing importance [6]. More recently, also genome sequencing, miRNA expression profiling and methylome studies are starting to offer new insights in the disease onset and progression. By the same token, specific epigenetic modifications together with protein alterations became valuable targets in leukemia research due to their reversible character and thus potential in therapy [7].

To extend the knowledge on CLL pathology and to identify new biomarker candidates, we applied quantitative mass spectrometry strategies to target the lower abundant proteins and peptides on patient and control samples [3]. The expressional differences between isolated age-matched healthy B-cells and CLL B-cells clearly showed that the morphological differences inherent to cancerous cells challenge disease marker discovery. Our comparative proteome analysis of UM and M CLL B-cells however, revealed that remarkably, only a limited amount of the identified proteins was differentially

expressed between patients with a different outcome. For both the UM and M patient group, known up-regulations of proteins contributing to cell proliferation were corroborated [8,9].

Analysis of the iTRAQ (isobaric tag for relative and absolute quantitation) data at the peptide level surfaced an interesting aberrant proteolytic product of a histone protein: clipping of the histone H2A *C*-tail. The specific clipping of histone H2A after V_{114} (cH2A) was previously reported in leukemia and leukemia cell lines and is catalyzed by the so-called "H2A specific protease" (H2Asp) [10–14]. Recently, we identified the enzyme Neutrophil Elastase (NE) as an important candidate for the identity of the H2Asp [15]. Even though proteolysis is often not considered as a regulated PTM and the importance of protein degradation in biological functions is frequently unclear, new technologies have started to unravel the critical role of clipping in cellular homeostasis and disease [16,17]. In some reports, histone clipping has even been suggested to be a functional modification with epigenetic potential [18,19]. More specifically, cH2A caught our attention as H2A is the only histone with a *C*-terminus protruding out of the nucleosomal core and the clipping site is localized at the entry and exit points of the DNA [20]. In general, histone modifications help in determining the heritable transcriptional state and lineage commitment development in normal B-cells [21].

Consequently, we persisted in the investigation of the H2A clipping in CLL as disruption of the histone code is suggested to drive hematopoietic cells in lymphomagenesis [22]. Here, we initially observed that H2A clipping was more abundant in CLL patients compared to healthy controls while no differences were found between M and UM. However, we showed this was not due to the disease itself, since this clipping seemed to be associated with the amount of myeloid cells present in the predominantly lymphoid samples. To further unfold the actual role of histone H2A proteolysis, we examined this cH2A clipping during induced differentiation of myeloid THP-1 cells into macrophages through quantitative mass spectrometry [12]. We concluded that synchronization of the THP-1 cells before the stimulation abrogates the temporally uprise of H2A clipping which has initially been observed at the onset of the differentiation.

3. Results

3.1. Differential Protein Expression between Healthy B-Cells and Chronic Lymphocytic Leukemia (CLL) B-Cells of M^- and UM^+ Staged Patients

The difference in protein expression between the B-cells of patients with indolent mutated $ZAP70^-$ (M^-) CLL, the aggressive unmutated $ZAP70^+$ (UM^+) CLL and healthy 50+ donors was quantified using iTRAQ. The six runs were merged into a dataset which comprised 18,014 MS/MS (mass spectrometry) queries, yielding 7473 annotated peptides which were derived from 536 unique proteins (Table S1a). For the ratios included in the analysis, in total 107 proteins differed significantly between groups (Table S1b). Among the leukemia samples, only 22 proteins were distinct between M^- and UM^+ as opposed to 70 between healthy and both CLL samples. One label in each run comprised the same pool of all CLL samples, which was included to simplify inter-run comparison and to increase sensitivity.

Functional clustering annotation of the protein lists from differentially expressed proteins between healthy and leukemia B-cells, showed that most up-regulated proteins in the leukemia samples (both UM^+ and M^-) are involved in mRNA processing, implying an increased transcriptional activity in cancer cells (Table S1c). Compared to healthy, most of the proteins that are down-regulated in the leukemia cells are predominantly connected to actin binding and cellular localization [23]. A cellular component analysis confirmed that 64% of the proteins up-regulated in CLL are categorized as nuclear whereas down-regulated proteins are primarily located in the cytosol (60%) (A.3.1. Figure 1 & Table S1c) [24]. These results suggest morphological differences between the healthy and cancerous cells rather than true molecular aberrancies.

In the limited list of proteins specifically overexpressed in the aggressive UM^+ compared to M^- two proteins are involved in ATP (Adenosine triphosphate)-binding, one is a ribosomal subunit and the H2B and Prohibitin proteins are involved in chromosomal organization (Table 1). For proteins significantly down-regulated in UM^+/M^-, the Interleukin enhancer-binding factor 3 and Bcl-2-associated transcription factor 1 are associated with DNA but most proteins are involved in metabolic mechanisms or are cytoskeletal.

Table 1. Proteins which are differentially expressed between UM^+ and M^- Chronic Lymphocytic Leukemia (CLL) B-cells. The table lists the protein name, Uniprot_ID, average UM^+/M^- ratio, the number of iTRAQ samples where the protein was identified (*n*) and the *p*-value.

UM^+ *vs.* M^-	Protein Name	Identified Protein	UM^+/M^- Average	*n*	*p*-Value
Down-regulated in UM^+/M^-	Serine/arginine-rich splicing factor	SRSF1_HUMAN	0.61	4	2.77×10^{-4}
		SRSF3_HUMAN	0.77	5	1.30×10^{-2}
		SRSF7_HUMAN	0.90	3	2.70×10^{-2}
	Bcl-2-associated transcription factor	BCLF1_HUMAN	0.69	3	3.84×10^{-2}
	40S ribosomal protein	RS21_HUMAN	0.70	3	3.64×10^{-2}
		RS28_HUMAN	0.67	4	1.47×10^{-2}
	Stress-induced-phosphoprotein	STIP1_HUMAN	0.73	3	4.68×10^{-2}
	Protein S100-A6	S10A6_HUMAN	0.75	5	2.24×10^{-2}
	Heat shock cognate 71 kDa protein	HSP7C_HUMAN	0.79	5	1.44×10^{-2}
	Matrin-3	MATR3_HUMAN	0.80	3	1.85×10^{-2}
	Ezrin	EZRI_HUMAN	0.84	4	3.32×10^{-2}
	Transitional endoplasmic reticulum ATPase	TERA_HUMAN	0.84	5	1.13×10^{-2}
	Interleukin enhancer-binding factor	ILF3_HUMAN	0.87	3	5.81×10^{-3}
	Heterogeneous nuclear ribonucleoprotein	HNRPK_HUMAN	0.88	6	1.38×10^{-3}
	Heterogeneous nuclear ribonucleoprotein	ROA1_HUMAN	0.89	5	1.15×10^{-2}
Up-regulated in UM^+/M^-	40S ribosomal protein	RS13_HUMAN	1.27	5	4.81×10^{-2}
	Prohibitin	PHB_HUMAN	1.35	4	3.72×10^{-2}
	Endoplasmin	ENPL_HUMAN	1.80	3	2.05×10^{-2}
	ATP synthase subunit beta, mitochondrial	ATPB_HUMAN	1.93	3	3.72×10^{-2}
	Histone H2B	H2B1C_HUMAN	2.59	3	1.24×10^{-2}
		H2B1D_HUMAN	1.10	3	1.84×10^{-2}
		H2B1O_HUMAN	1.18	3	2.97×10^{-2}

3.2. Quantitative Mass Spectrometry and Western Blot Analysis on CLL Samples Endorsed the Myeloid Origin of H2A Clipping

Additional in-depth analysis at the peptide level, surfaced one remarkably aberrant modification in all the samples of the described iTRAQ analysis: clipping of histone H2A after V_{114} (Figure 1). The peptide VTIAQGGVLPNIQAV (*m/z* 740.4, charge 2+) was the only semi-tryptic peptide out of the >7400 annotated MS/MS spectra that was identified in all six runs. The annotation of the peptide was confirmed by *de novo* sequencing on the MS/MS spectrum (A.3.1. Figure 2). These results highlight how iTRAQ chemistry contributes to a better annotation of semi-tryptic peptides by enhancing the sensitivity due to the multiplexing and increased *b*-ion formation.

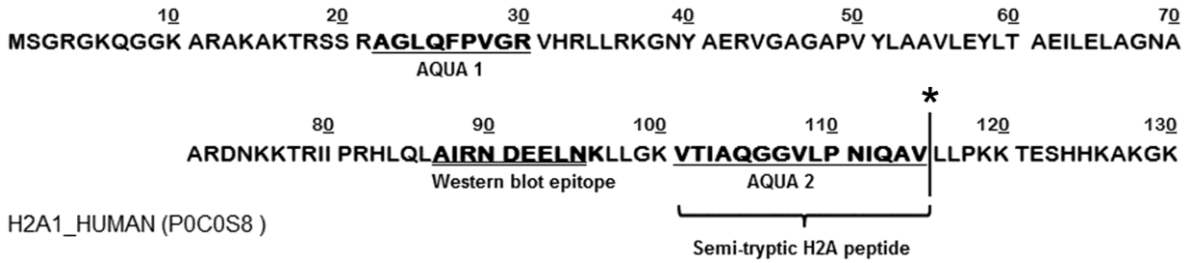


Figure 1. Sequence of histone H2A. Sequence from Uniprot [25]. The VTIAQGGVLPNIQAV peptide (*m/z* 740.4, charge 2+) was the only semi-tryptic peptide identified in all six runs. ***** Indicates the clipping site after V_{114} (here described as amino acid 115 since methionine, generated by the start codon, is the first amino acid in the Uniprot sequence); Bold underlined sequences are the same as the isotopically labeled absolute quantification (AQUA) peptides, used for the subsequent specific cH2A quantification. The bold double underlined sequence is the Western blot epitope.

For quantification of this fragment, we compensated for abovementioned morphological differences by normalizing the reporter intensities of the clipping fragments to the average of the whole H2A protein. The relative amount of H2A that is cleaved after V_{114} was found to be on average higher in leukemia cells compared to healthy B-cells as seen by the log-transformed ratios. Particularly, all six ratios of the pool of UM^+ and M^- samples were consistently positive opposed to healthy (Figure 2A).

The consequences of proteolytic PTMs are ubiquitously underappreciated so we persisted in the more specific investigation of cH2A in a larger patient population due to the biomarker potential of this previously reported modification (Scheme of workflow: A.3.1. Figure 3). Immunodetection of H2A on histone extracts visualized cH2A only in one sample (Figure 2B, SYPRO staining: A.3.1. Figure 4).

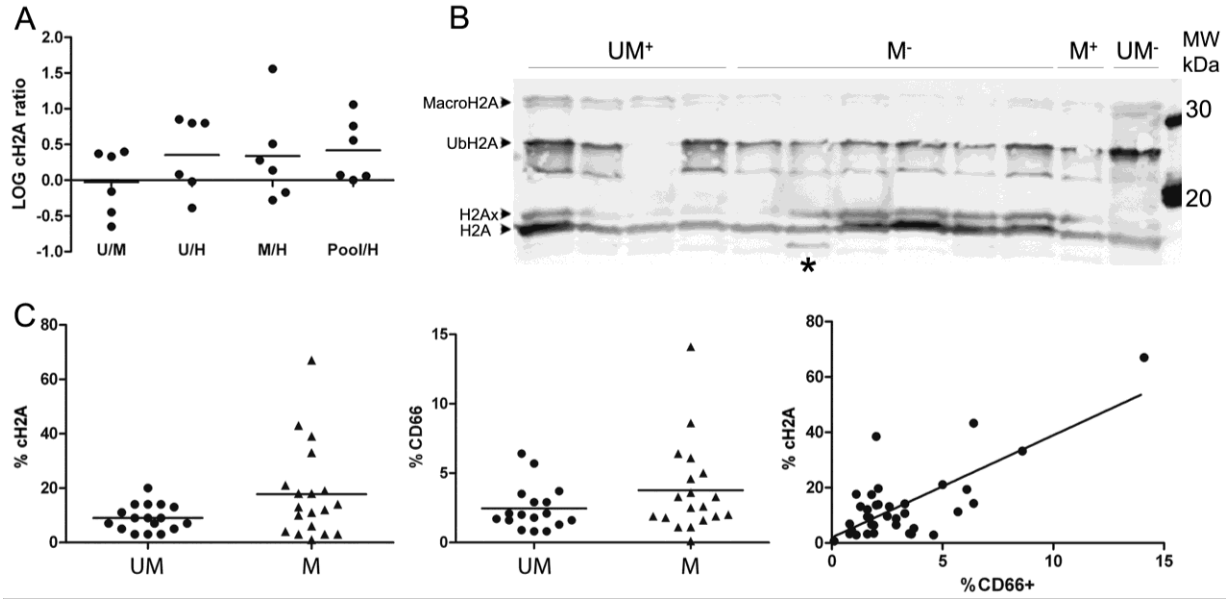


Figure 2. Quantitative mass spectrometry and Western blot analysis on CLL samples revealed that histone H2A clipping is from myeloid origin. (**A**) The iTRAQ ratios of the cH2A peptide (normalized to the average of the H2A protein to compensate for morphological differences) hint towards an increased abundance in CLL compared to healthy B-cells. The individual log ratios of all six runs are presented as dots, the average ratios as the horizontal bars. Histone H2A (cH2A) is up-regulated in the samples of the leukemia pool compared to the samples of the healthy B-cells in all six runs; (**B**) The histone extracts from 12 of the 36 samples (From left to right: patient samples UM^+ 8, 9, 10 and 11; M^- 9, 10, 11, 12, 13 and 14; M^+ 5 and UM^- 6). Western blot with an H2A antibody against the epitope depicted in Figure 1 detects the H2A variants H2Ax, Ubiquitinated H2A and macroH2A. The band under ubiquitinated H2A could not be identified. cH2A was only faintly detected, except for one sample (*****); (**C**) Specific cH2A screening with AQUA peptides and flow cytometry data revealed the myeloid characteristic of the clipping. **Left panel**: %cH2A differs significantly (p = 0.049) between CLL patients with a distinct mutational status (UM: 11× UM^+ & 6× UM^-, M: 14× M^- & 5× M^+); **Middle panel**: Although not significant (p = 0.15), the %CD66b suggested a similar correlation with the mutational status; **Right panel**: Relation between %cH2a and %CD66b. Spearman's Rho correlation between $CD66b^+$ and %cH2A was significant at the 0.01 level (Spearman's Rho correlation coefficient: 0.439; p = 0.007; Data: Table S2b).

The flow cytometry data showed that this particular sample had the highest amount of granulocytes (expressed as $\%CD66b^+$ cells) in the screened patient population. In line with this, direct comparison of %cH2A with the more sensitive and specific AQUA (absolute quantification) approach suggested higher clipping in the M compared to UM CLL B-cells (Figure 2C, left panel) where a similar

relationship between the amount of granulocytes cells and the respective mutational status was observed (Figure 2C, middle panel). Indeed, a statistical analysis of the possible correlation between the amounts of V_{114} clipped H2A and all cellular markers investigated for each sample confirmed a significant correlation between the amount of $CD66b^+$ cells present in the samples and %cH2A (Figure 2C, right panel) (Details patient screening are listed in A.3.1.Table 1.). Even though all steps were performed at 4 °C in the presence of protease inhibitors, we observed that H2A clipping can actually be caused *in vitro* to some extent as we observed some residual clipping activity on spiked-in biotinylated H2A (data not shown).

3.3. Characterization of cH2A in THP-1 Cells

To further corroborate the myeloid nature of cH2A we cultured human leukemia cell lines of different origin and prepared histone extracts when a density of approximately 1×10^6/mL was achieved. In all the investigated lymphoid cell lines, no clipping of H2A was detected. For the myeloid cells, ±10% cH2A was measured in the monocyte-like THP-1 and U-937 cell lines and, although not unambiguously, ±2% in the promyelocytic HL-60 cells (Figure 3A).

Next, we applied our sensitive and specific AQUA approach to specifically quantify $H2AV_{114}$ clipping in a previously reported model in which Ohkawa *et al.* describe a temporal increase of truncated H2A proteins during stimulated differentiation of acute monocytic leukemia THP-1 cells into macrophages using both phorbol 12-myristate 13-acetate (PMA) and retinoic acid (RA) [13]. For three biological replicates, %cH2A was specifically quantified in histone extracts, isolated at different time-points shortly after PMA supplementation (Figure 3B). The clipping at T_{PMA0} was significantly different than at $T_{PMA\ 10}$ ($p = 0.045$) and T_{PMA60} ($p = 0.011$). Equally as to T_{PMA60}, %cH2A was significantly different than at $T_{PMA\ 5}$ ($p = 0.041$), $T_{PMA\ 10}$ ($p = 0.010$) and $T_{PMA\ 30}$ ($p = 0.0075$). These results confirm that H2A clipping at V_{114} indeed ascends and sequentially decreases upon THP-1 differentiation.

Finally, since we hypothesized that the variation might be due to differences in cell cycle synchronization, THP-1 cells were synchronized by double thymidine block, as the more open chromatin structure during cell division could very well be more susceptible to proteolytical activity during cell lysis. After synchronization, cells were either left untreated or were subjected to differentiation and cH2A was subsequently quantified by AQUA screening on histone extracts, obtained from different time points (Figure 3B). %cH2A was significantly lower in both the control ($p = 6.09 \times 10^{-4}$) and stimulated ($p = 9.18 \times 10^{-5}$) THP-1 cells compared to the unsynchronized cells. Further, no difference was observed between the synchronized stimulated and control cells ($p = 0.22$). Sampling 48 h after the start of the

differentiation indicated that the amount of clipped H2A was further reduced in both cell lines (data not shown).

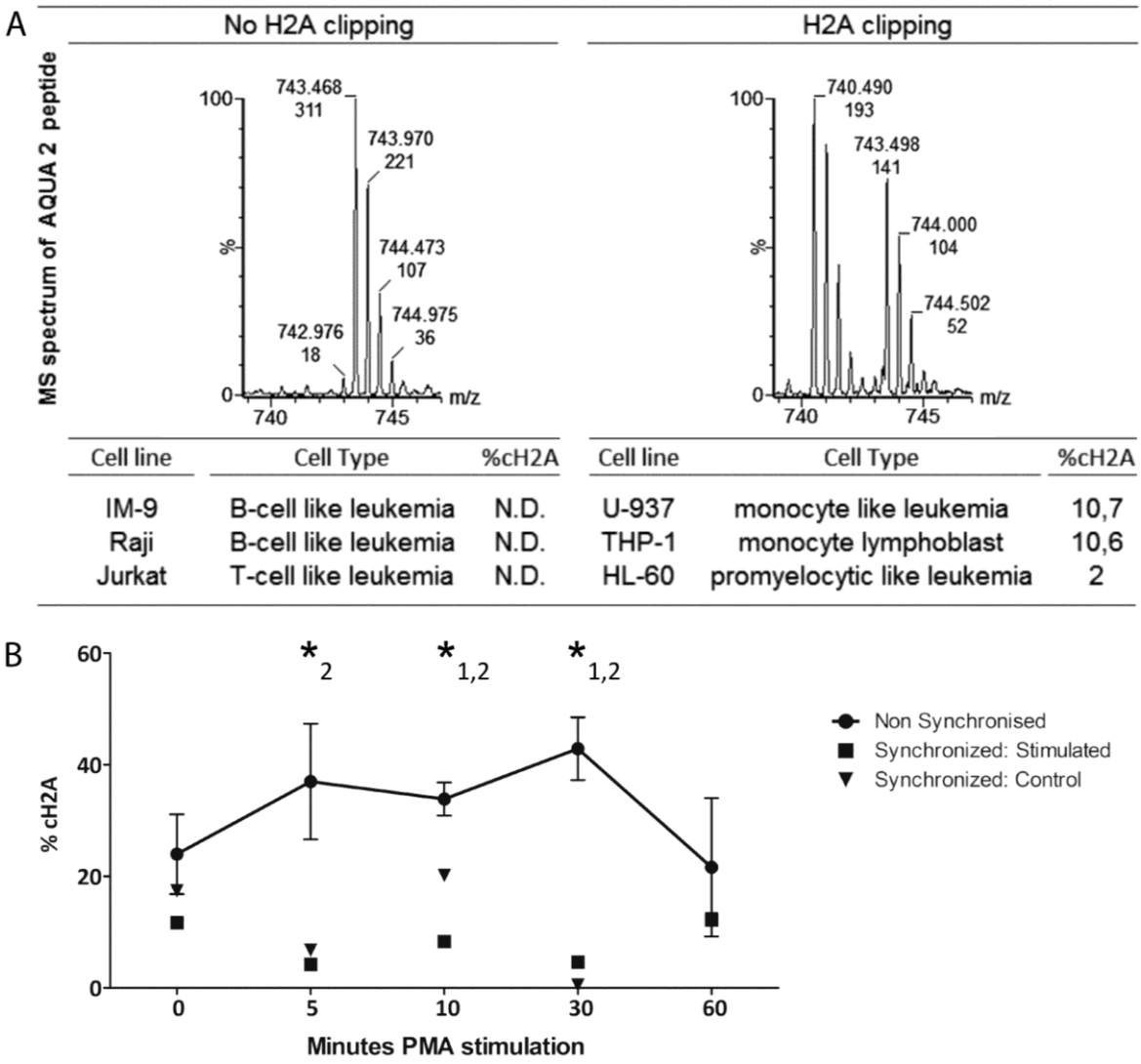


Figure 3. Specific cH2A quantification in myeloid cell lines. (**A**) Example of an MS spectrum of the AQUA 2 peptide if H2A V_{114} clipping is absent (**left**) or present (**right**). The analysis of different cell lines confirms the myeloid characteristic of cH2A in the U-937, THP-1 and HL-60 cells. H2A clipping is not present in the investigated lymphoid cell lines and could not be unambiguously detected in the promyelocytic HL-60 cell line; (**B**) THP-1 cells stimulated with PMA show a transient H2A clipping pattern. Three biological replicates display the high variance. *****1: significant differences of %cH2A between T_{PMA0} and the other data points; *****2: the equivalent for T_{PMA60}. The %cH2A of synchronized stimulated and control THP-1 cells, presented respectively as squares and triangles, is lower than the %cH2A of non-synchronized cells.

4. Discussion

Quantitative proteomics on patient samples and on leukemia cell lines can help to define new biomarkers and insights into the pathogenesis of lymphoid and hematopoietic neoplasms [2,3]. In our comparative proteome study between healthy and CLL B-cells of patients with different prognosis, we implemented two quantitative label-based mass spectrometry methods: iTRAQ and AQUA. The isobaric character of the iTRAQ labels allows multiplexing different samples, resulting in an increased signal and lower sample complexity [26]. We included both CLL samples of patients with a different disease prognosis and healthy B-cells in the analysis to obtain deeper insights in disease biology. A substantial part of the observed significant differences in relative protein expression between healthy and CLL B-cells is primarily due to differences in morphology between healthy and neoplastic cells [27]. Cancerous B-cells are indeed characterized by a larger nucleus and a denser cytoplasm, which we affirmed here by functional grouping and cellular component analysis of the up- and down-regulated proteins. We caution for the interpretation of proteomics data when healthy and malignant cells differ strongly in their morphology, an important restriction that is not always validated prior to relative comparison in proteomics approaches. From a clinical perspective however, aberrant ratios between patient groups with UM^+, who have a bad prognosis and poor overall survival range and the less aggressive M^-, are more relevant. Notably, less than 5% of the identified proteins were significantly different between these two groups. In this list of proteins with a log ratio significantly different from zero, defining distinctions between M^- and UM^+, one of the most remarkable candidates is the Bcl-2-associated transcription factor 1 ($p = 3.84 \times 10^{-2}$). The Bcl-2 family regulates apoptosis and is an established hallmark in CLL as aberrant expression of Bcl-2 proteins causes apoptosis resistance of CLL B-lymphocytes. Although not all the Bcl-2 proteins correlate consistently with known CLL biomarkers, several Bcl-2-antagonists are in clinical trials for CLL treatment [8].

The use of the iTRAQ label and the subsequent manual analysis of the results at the peptide level together surfaced another fascinating finding: a semi-tryptic peptide derived from histone H2A clipped at V_{114}. Of all the trypsin-based mass spectrometry experiments uploaded in the PRIDE database, this fragment was only annotated once (accession: 10,528) [28]. Remarkably, we identified the peptide in all six runs. This could be explained by the contribution of the iTRAQ labels which are covalently bound to the peptide *N*-termini, generating intense b-series and consequently contributing to an increased peptide score [26]. cH2A is generated by the removal of fifteen amino acids from the carboxy-terminal end of the intact H2A molecule after V_{114}, coordinately removing K119 which is an important site of mono-ubiquitination [29–33]. cH2A had been described over 35 years ago as a product of H2A specific protease (H2Asp) activity [14] and shortly afterward, other groups observed similar cleavage

patterns in extracts from both myeloid and lymphoid leukemias [10,11]. Since then, this clipping event was only referenced on a few occasions.

Although the biological function of protein degradation is largely unknown, proteolysis is an important category of PTMs, e.g., 5%–10% of all drug targets are proteases [16]. Truncation of histone tails has already been linked to cell differentiation and the *C*-tail of H2A is known to be important in cellular homeostasis and chromatin biology [18,20]. We recently identified the myeloid enzyme NE as being a prime candidate to fulfill the reaction mediated by the H2Asp but could not clarify if cH2A formation is involved in an epigenetic process or is rather a consequence of NET formation [15]. In healthy hematopoietic cells we only observed H2A clipping in cells of myeloid origin but the references to clipped H2A found in literature are mainly in the context of leukemia [10–13]. We thus persisted in a more detailed analysis of this modification and examined if cH2A formation is an epigenetic hallmark of CLL.

The preliminary iTRAQ results indicated a greater average abundance of the cH2A peptide in leukemia samples compared to the healthy controls. As validation is required, we specifically quantified cH2A in histone extracts derived from CLL B-cells of 36 clinically staged patients by applying the isotopic synthesized AQUA peptides that allow to target specific known H2A peptides, present in low concentrations [26]. We could this time not define any direct connection between %cH2A and any known disease marker. However, the amount of granulocytes in the samples did correlate with %cH2A, which corroborates the myeloid character of cH2A and is in line with the identification of NE. As complete inhibition of proteases is known to be very challenging, as also seen by us in a spike-in experiment of biotinylated H2A (not shown), it is difficult to define how much, if any, cH2A is endogenously generated [34].

On the other hand, the reported transient clipping of the H2A *C*-tail during the induced differentiation of THP-1 promonocytes, implies a potential biological role of H2A processing in the hematopoietic development of cells from myeloid origin [13]. Cells may for instance have mechanisms to control histone degradation for re-establishing the epigenetic marks on their tails in the proliferating state [35]. However, although our AQUA results demonstrate a brief uprise in specific H2A V_{114} clipping upon PMA or RA stimulation, synchronization of the THP-1 cells before the stimulation abrogated such trend. Histone clipping seems to correlate with the myeloid cell cycle and as our results suggest, cH2A fluctuation is more likely caused by possible differences in cell cycle stage synchronization, rather than PMA or RA induced differentiation. Instead of being a regulated mechanism, we hypothesize that the high degree of variation found in these experiments probably is due to the more open chromatin structure during cell division rendering histones more susceptible to proteolytical activity.

5. Experimental Section

5.1. Cells and Reagents

Phosphate buffered saline (PBS), media, L-glutamine, Fetal bovine serum (FBS), penicillin/ streptomycin, Dynabeads and SYPRO Ruby were from Life Technologies (San Diego, CA, USA), ammonium bicarbonate (ABC), sodium dodecyl sulfate (SDS), *N*-cyclohexyl-3-aminopropanesulfonic acid (CAPS) and Tween-20 from Millipore (Billerica, MA, USA). ReadyPrep sequential extraction kit was from BioRad (Hercules, CA, USA) and Vivaspin-2 columns with a molecular weight cut-off of 3000 Da from Sartorius (Göttingen, Germany). The Recombinant human H2A (M2502S) was obtained from New England Biolabs (Ipswich, MA, USA) and bovine histone extract (cat. no. 223565) from Roche (Basel, Switzerland). All other reagents were purchased from Sigma Aldrich (St. Louis, MO, USA) unless described otherwise. Raji, Jurkat and HL-60 cells were cultured in Dulbecco's Modified Eagle Medium and IM-9, U-937 and THP-1 cells in RPMI-1640 medium, both enriched with 2% (*w*/*v*) L-glutamine, 10% (*w*/*v*) FBS and 50 IU/mL penicillin/streptomycin. To achieve synchronization at the late G1− early S phase, 2 mM thymidine was added over two intervals of 12–16 h, with an incubation in non-thymidine containing RPMI-1640 medium for 8 h in between 50 ng/mL PMA was added to the THP-1 medium for the differentiation of the non-synchronized cells and 1 μM RA for the differentiation of the synchronized cells.

5.2. Patient Samples

For the iTRAQ analysis, whole blood of six UM^+ and six M^- clinically staged CLL patients was obtained from Ghent University Hospital, Department of Hematology. Informed consent was given according to the requirements of the Ethics Committee of the Ghent University Hospital. Peripheral blood mononuclear cells (PBMCs) were isolated from a Ficoll-Paque (GE Healthcare, Waukesha, WI, USA) density gradient, B-cells were purified using Dynabeads Untouched Human B (Life Technologies, San Diego, CA, USA) [36]. Correspondingly, control samples were isolated from healthy volunteers aged 50+. For the high throughput screening of cH2A, samples were obtained from the UCSD CLL Research Consortium (CRC). Immediately after thawing, cells were washed twice with cold PBS containing 1 mM phenylmethanesulfonyl fluoride (PMSF) and protease inhibitor cocktail (Roche, Basel, Switzerland). $CD5^+CD19^+$ cells consistently made out more than 85% of the lymphocyte population as seen by flow cytometry [36].

5.3. Flow Cytometry

For each measurement 2 × 10^5 cells were washed twice at 4 °C with PBS 1% Bovine Serum Albumin (BSA) and analyzed using a Cytomics FC500 flow cytometer (Becton Dickinson Immunocytometry Systems, San Jose, CA, USA) with monoclonal antibodies (mAbs) antibodies from BD-Biosciences (Franklin Lakes, NJ, USA): Isotype controls, anti-CD5 (PECy5), anti-CD19 (FITC), anti-CD33 (PECy5), anti-CD66b (FITC) and anti-Annexin V (FITC). The synchronization of THP-1 cells was monitored with Propidium Iodide staining [36].

5.4. Cell Lysis and Histone Isolation

All steps were performed at 4 °C. To obtain a complete cell lysate for the iTRAQ analysis, cells were washed twice with PBS, pelleted and resuspended in the ReadyPrep sequential extraction buffer 1 at 5 × 10^6 cells/mL, supplemented with protease inhibitor cocktail (Roche), 1 mM PMSF, 10 µL 200 mM Tributylphosphine (Biorad), 20 µL phosphatase inhibitor cocktail 1 and 2 and 1 µL 250-units/µL benzonase. After sonication and centrifugation at 1500 rpm for 5 min, the proteins in the supernatant were transferred to a new eppendorf. The obtained pellet was resuspended in buffer 3 from the extraction kit and sonicated for 10 min. After centrifugation at 1500 rpm for 5 min the supernatant was pooled with the previous extract. Detergents, inhibitors and urea were removed by washing twice with Milli-Q water on a Vivaspin-2 column [37].

For the histone extracts, harvested cells were washed twice in PBS containing 1 mM PMSF, and protease inhibitor cocktail. 10^7 cells/mL were resuspended in Triton extraction buffer (PBS containing 0.5% (*v*/*v*) Triton X-100, 1 mM PMSF and protease inhibitor cocktail) and lysed by gentle stirring. Pelleted nuclei were subsequently washed in PBS containing 1 mM PMSF and proteinase inhibitor cocktail. Histones were extracted overnight after benzonase treatment of the sonicated nuclei by acid extraction: incubation in 250 µL 0.2 M HCl at 4 °C with gentle stirring. Precipitated proteins were pelleted and the supernatant containing the histones was dried and stored at −20 °C until further use [38]. The Bradford Coomassie Assay determined the protein content of all samples.

5.5. Western Blot Analysis

An amount of 3 µg of a dried histone extract was suspended in Laemmli buffer and run on a 15% PreCast Gel (Biorad) for 30 min at 150 V and 60 min at 200 V and subsequently transferred to a nitrocellulose membrane in a 10 mM CAPS buffer with 20% MeOH (Merck, New York, NY, USA). The remaining proteins in the gel were visualized after overnight SYPRO Ruby staining. For Western blot, all steps were performed at room temperature with intermediate washing steps in 0.3% Tween-20 in PBS.

The Histone H2A (LS-C24265, LifeSpan BioScience, Seattle, WA, USA) antibody incubation was performed overnight in a 1:1000 dilution in PBS 1% BSA, followed by 1 h incubation in the same buffer, with a stabilized horseradish-peroxidase (HRP)-conjugated goat anti-rabbit immunoglobulin G (Pierce, Rockford, IL, USA). The Supersignal West Dura Extended Duration Substrate (Pierce) was applied to perform the chemiluminescence and both the gel and Western blot membrane were visualized with a VersaDoc Imaging System [39].

5.6. Quantitative Mass Spectrometry Analysis

All digests were performed according to the iTRAQ (ABSciex, Foster City, CA, USA) reagent Kit guidelines, as was the given labeling itself. For the iTRAQ analysis, six 4plex runs each encompassed 4 × 100 µg of the protein lysates from CLL B-cells from UM^+, M^- and B-cells from healthy donors aged 50^+. The fourth label was used for an additional pool of all 12 CLL samples for inter-run comparison and to increase the number of identifications. Technical variation was also minimized by reversing the labeling order. All six samples were first fractionated off line into 12 fractions on a Poros 10S strong-cation exchange (SCX) column (300 µm i.d. × 15 cm, ABSciex, Foster City, CA, USA) for subsequent nano reversed phase liquid chromatography (Dionex U3000, Dionex, Chelmsford, MA, USA) separation on a gradient specifically optimized for sample content of each fraction [37]. All other samples were separated on a 70 min organic gradient from 4%–100% buffer B (80% (*v*/*v*) acetonitrile (Millipore, Billerica, MA, USA) in 0.1% (*v*/*v*) FA). All samples were analyzed on an Electrospray ionization—Q-TOF Premier (Waters, Wilmslow, UK). Data was searched against the SwissProt (541,954 sequence entries) database using Mascot 2.3 and additionally manually interpreted with Mascot Distiller (Matrix Science, London, UK). Proteins were included for the analysis if a proteotypic peptide had a Mascot score above 45 in at least 3 runs. For iTRAQ quantification, ratios were normalized based on summed intensities and only proteins recurring in at least three out of six different runs were withheld by an automated in-house approach. Gene ontology (GO) analysis was performed on the significantly up- or down-regulated proteins with Uniprot, DAVID and the WEB-based GEne SeT AnaLysis Toolkit (WEBGESTAT) [23–25].

For specific cH2A screening, a mix of the isobaric peptides (AQUA1, Thermo, Waltham, MA, USA) AQUA2, Sigma Aldrich (St. Louis, MO, USA) was added right before MS analysis. The tryptic H2A *N*-terminal peptide R.AGLQFPVGR.V (*m/z* 475.7 was used to quantify the total amount of histone H2A present in the sample (AQUA 1). Specific clipping is quantified by means of the semi-tryptic peptide K.VTIAQGGVLPNIQAV.L (*m/z* 743.4) (AQUA2) [40]. This is the same sequence as was found during the iTRAQ analysis. To 1 µg of HE, 10 pmol of AQUA1 and 1 pmol of AQUA2 was spiked right before the MS analysis [15]. For the quantification, the total ion current (TIC) from both AQUA peptides was obtained from a single MS-scan acquired on the top of the two extracted-ion chromatograms. An example of the MS

spectra of the two AQUA peptides is given at the right side of the accolade in A.3.1. Figure 3. To confirm the efficiency of the histone extraction, the raw data of the data-directed analysis was equally searched against the SwissProt database using Mascot 2.3 (Matrix Sciences, London, UK).

5.7. Statistical Analysis

For the iTRAQ analysis, ratios were log transformed and averaged. Statistical analysis was performed by means of a homoscedastic two-tailed *t*-test to inspect which log ratios were significantly different from zero ($p = 0.05$). The average ratio of a protein indicated whether a protein is up- or down-regulated between two samples. The same *t*-tests were performed to test if %cH2A and %CD66b was different between the M and UM patients. For the comprehensive analysis of the AQUA screening, the relationship between cH2A and the other variables was determined by the nonparametric Spearman's Rho correlation test wherein the ZAP and the mutational status were analyzed as dichotomized values (+ or −) with the SPSS Statistics 20 software (Endicott, NY, USA). To corroborate if the clipping temporally increased during the differentiation of the THP-1 cells, one-tailed *t*-tests were applied to compare each data point with the %cH2A at $T_{PMA\ 0}$ and $T_{PMA\ 60}$. With the same test, all the data points of the non-synchronized experiment were examined against all the data points of the synchronized (stimulated or control) experiment. Similarly, a *t*-test was used to validate if the %cH2A between the synchronized stimulated and control cells was analogous.

6. Conclusions

We particularly emphasize both the pitfalls and the benefits of applying quantitative proteomic strategies in hematological research. Although morphological differences hamper proteome-wide comparison of healthy and leukemia B-cells, we resurfaced a previously described clipping product of the H2A *C*-tail. A more profound characterization in CLL and THP-1 samples indicates that H2A V_{114} clipping occurs in hematopoietic cells of the myeloid lineage. Although the process of histone clipping has considerable epigenetic potential, many questions about the relevance of specific histone proteolysis remain.

ACKNOWLEDGMENTS

We thank Koji Takada from the Jikei University School of Medicine for the vials with the THP-1 cells.

AUTHOR CONTRIBUTIONS

P.G. wrote the manuscript and performed the bulk of the described laboratory and mass spectrometry experiments together with the subsequent data analysis. L.V. performed the THP-1 synchronization and associated flow monitoring and shared her expertise on cell culturing and overall histone research. K.V.S. helped with/optimized the electrophoresis and Western blot analyses and reviewed the writing. S.L. assisted with the design of the iTRAQ analysis and performed the statistics of the CLL screening. F.V.N. wrote the semi-automated data analysis tool and contributed to the interpretation of the data. F.O. provided the healthy and CLL patient samples for the iTRAQ analysis and supervised the determination of the mutational status of the *IGVH* genes and the ZAP70 expression. T.K. provided the CLL patient samples for the AQUA analysis and supervised the determination of the mutational status of the *IGVH* genes and the ZAP70 expression. M.D. performed the iTRAQ analysis and assisted with the subsequent experiments, data analysis and the writing of the manuscript. D.D. supervised this work.

CONFLICTS OF INTEREST

The authors declared no conflict of interest.

SUPPLEMENTARY FILES IN ADDENDUM AND ONLINE

http://www.mdpi.com/1422-0067/15/6/9407

References

[1] Unwin, R. D., Whetton, A. D., How will haematologists use proteomics? Blood Rev 2007, 21, 315-326.
[2] Boyd, R. S., Dyer, M. J., Cain, K., Proteomic analysis of B-cell malignancies. Journal of proteomics 2010, 73, 1804-1822.
[3] Zamo, A., Cecconi, D., Proteomic analysis of lymphoid and haematopoietic neoplasms: there's more than biomarker discovery. Journal of proteomics 2010, 73, 508-520.
[4] Chiorazzi, N., Rai, K. R., Ferrarini, M., Chronic lymphocytic leukemia. N Engl J Med 2005, 352, 804-815.
[5] Chiorazzi, N., Implications of new prognostic markers in chronic lymphocytic leukemia. Hematology Am Soc Hematol Educ Program 2012, 2012, 76-87.
[6] Smolewski, P., Witkowska, M., Korycka-Wolowiec, A., New Insights into Biology, Prognostic Factors, and Current Therapeutic Strategies in Chronic Lymphocytic Leukemia. ISRN oncology 2013, 2013, 740615.
[7] Rodriguez-Vicente, A. E., Diaz, M. G., Hernandez-Rivas, J. M., Chronic lymphocytic leukemia: a clinical and molecular heterogenous disease. Cancer genetics 2013, 206, 49-62.
[8] Scarfo, L., Ghia, P., Reprogramming cell death: BCL2 family inhibition in hematological malignancies. Immunology letters 2013.
[9] Woodlock, T. J., Bethlendy, G., Segel, G. B., Prohibitin expression is increased in phorbol ester-treated chronic leukemic B-lymphocytes. Blood cells, molecules & diseases 2001, 27, 27-34.
[10] Simpkins, H., Mahon, K., The histone content of chromatin preparations from leukaemic cells. Br J Haematol 1977, 37, 467-473.
[11] Pantazis, P., Sarin, P. S., Gallo, R. C., Detection of the histone-2A related polypeptide in differentiated human myeloid cells (HL-60) and its distribution in human acute leukemia. Int J Cancer 1981, 27, 585-592.
[12] Okawa, Y., Takada, K., Minami, J., Aoki, K., et al., Purification of N-terminally truncated histone H2A-monoubiquitin conjugates from leukemic cell nuclei: probable proteolytic products of ubiquitinated H2A. Int J Biochem Cell Biol 2003, 35, 1588-1600.
[13] Minami, J., Takada, K., Aoki, K., Shimada, Y., et al., Purification and characterization of C-terminal truncated forms of histone H2A in monocytic THP-1 cells. Int J Biochem Cell Biol 2007, 39, 171-180.
[14] Eickbush, T. H., Watson, D. K., Moudrianakis, E. N., A chromatin-bound proteolytic activity with unique specificity for histone H2A. Cell 1976, 9, 785-792.
[15] Dhaenens, M., Glibert, P., Lambrecht, S., Vossaert, L., et al., Neutrophil Elastase in the capacity of the "H2A-specific protease". Int J Biochem Cell Biol 2014, 51C, 39-44.
[16] Doucet, A., Butler, G. S., Rodriguez, D., Prudova, A., Overall, C. M., Metadegradomics: toward in vivo quantitative degradomics of proteolytic post-translational modifications of the cancer proteome. Molecular & cellular proteomics : MCP 2008, 7, 1925-1951.
[17] Rogers, L., Overall, C. M., Proteolytic post translational modification ofproteins: proteomic tools and methodology. Molecular & cellular proteomics : MCP 2013.
[18] Osley, M. A., Epigenetics: how to lose a tail. Nature 2008, 456, 885-886.
[19] Azad, G. K., Tomar, R. S., Proteolytic clipping of histone tails: the emerging role of histone proteases in regulation of various biological processes. Mol Biol Rep 2014.
[20] Vogler, C., Huber, C., Waldmann, T., Ettig, R., et al., Histone H2A C-terminus regulates chromatin dynamics, remodeling, and histone H1 binding. PLoS genetics 2010, 6, e1001234.
[21] Parra, M., Epigenetic events during B lymphocyte development. Epigenetics 2009, 4, 462-468.
[22] Taylor, K. H., Briley, A., Wang, Z., Cheng, J., et al., Aberrant epigenetic gene regulation in lymphoid malignancies. Seminars in hematology 2013, 50, 38-47.
[23] Huang da, W., Sherman, B. T., Lempicki, R. A., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009, 4, 44-57.
[24] Wang, J., Duncan, D., Shi, Z., Zhang, B., WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res 2013, 41, W77-83.

[25] Magrane, M., Consortium, U., UniProt Knowledgebase: a hub of integrated protein data. Database : the journal of biological databases and curation 2011, 2011, bar009.
[26] Vaudel, M., Sickmann, A., Martens, L., Peptide and protein quantification: a map of the minefield. Proteomics 2010, 10, 650-670.
[27] Baba A.I., C. C., Chapter 3, TUMOR CELL MORPHOLOGY, Comparative Oncology 2007.
[28] Vizcaino, J. A., Cote, R. G., Csordas, A., Dianes, J. A., et al., The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res 2013, 41, D1063-1069.
[29] Eickbush, T. H., Moudrianakis, E. N., The histone core complex: an octamer assembled by two sets of protein-protein interactions. Biochemistry 1978, 17, 4955-4964.
[30] Eickbush, T. H., Godfrey, J. E., Elia, M. C., Moudrianakis, E. N., H2a-specific proteolysis as a unique probe in the analysis of the histone octamer. J Biol Chem 1988, 263, 18972-18978.
[31] Davie, J. R., Numerow, L., Delcuve, G. P., The nonhistone chromosomal protein, H2A-specific protease, is selectively associated with nucleosomes containing histone H1. J Biol Chem 1986, 261, 10410-10416.
[32] Elia, M. C., Moudrianakis, E. N., Regulation of H2a-specific proteolysis by the histone H3:H4 tetramer. J Biol Chem 1988, 263, 9958-9964.
[33] Watson, D. K., Moudrianakis, E. N., Histone-dependent reconstitution and nucleosomal localization of a nonhistone chromosomal protein: the H2A-specific protease. Biochemistry 1982, 21, 248-256.
[34] Groth, I., Alban, S., Elastase inhibition assay with peptide substrates - an example for the limited comparability of *in vitro* results. Planta medica 2008, 74, 852-858.
[35] Gunjan, A., Paik, J., Verreault, A., The emergence of regulated histone proteolysis. Curr Opin Genet Dev 2006, 16, 112-118.
[36] Geeraerts, B., Vanhoecke, B., Vanden Berghe, W., Philippe, J., et al., Deguelin inhibits expression of IkappaBalpha protein and induces apoptosis of B-CLL cells in vitro. Leukemia 2007, 21, 1610-1618.
[37] Lambrecht, S., Dhaenens, M., Almqvist, F., Verdonk, P., et al., Proteome characterization of human articular chondrocytes leads to novel insights in the function of small heat-shock proteins in chondrocyte homeostasis. Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society 2010, 18, 440-446.
[38] Vossaert, L., Meert, P., Scheerlinck, E., Glibert, P., et al., Identification of histone H3 clipping activity in human embryonic stem cells. Stem cell research 2014, 13, 123-134.
[39] Vermoote, M., Van Steendam, K., Flahou, B., Smet, A., et al., Immunization with the immunodominant Helicobacter suis urease subunit B induces partial protection against H. suis infection in a mouse model. Veterinary research 2012, 43, 72.
[40] Stemmann, O., Zou, H., Gerber, S. A., Gygi, S. P., Kirschner, M. W., Dual inhibition of sister chromatid separation at metaphase. Cell 2001, 107, 715-726.

PART 2: NEUTROPHIL ELASTASE IN THE CAPACITY OF THE "H2A-SPECIFIC PROTEASE"

Based on:

Neutrophil Elastase in the capacity of the "H2A-specific protease"

Dhaenens M.[1†], Glibert P.[1†], Lambrecht S.[2], Vossaert L. [1],

Van Steendam K. [1], Elewaut D. [2], Deforce D. [1*]

[1] Laboratory of Pharmaceutical Biotechnology, Ghent University, 460 Ottergemsesteenweg, B-9000 Ghent, Belgium; E-Mails: maarten.dhaenens@ugent.be (M.D.); pieter.glibert@ugent.be (P.G.); liesbeth.vossaert@ugent.be (L.V.); katleen.vansteendam@ugent.be (K.V.S.)

[2] Department of Rheumatology, Ghent University Hospital, 185 1P7 De Pintelaan, B-9000 Ghent, Belgium; E-Mail: stijn.lambrecht@hotmail.com (S.L.); dirk.elewaut@UGent.be (D.E.)

[†] These authors contributed equally to this work.

[*] Author to whom correspondence should be addressed; E-Mail: dieter.deforce@ugent.be; Tel.: +32-9-264-8067; Fax: +32-9-220-6688.

Accepted for publication:

Int J Biochem Cell Biol. 2014 Jun;51:39-44. doi: 10.1016/j.biocel.2014.03.017.

1. Abstract

The amino-terminal tail of histones and the carboxy-tail of histone H2A protrude from the nucleosome and can become modified by many different posttranslational modifications (PTM). During a mass spectrometric proteome analysis on haematopoietic cells we encountered a histone PTM that has received only little attention since its discovery over 35 years ago: truncation of the histone H2A C-tail at V_{114} which is mediated by the "H2A specific protease" (H2Asp). This enzyme is still referenced today but it was never identified. We first developed a sensitive AQUA approach for specific quantification of the $H2AV_{114}$ clipping. This clipping was found only in myeloid cells and further cellular fractionation led to the annotation of the H2Asp as Neutrophil Elastase (NE). Ultimate proof was provided by NE incubation experiments and by studying histone extracts from NE *Null* mice. The annotation of the H2Asp not only is an indispensable first step in elucidating the potential biological role of this enzymatic interaction but equally provides the necessary background to critically revise earlier reports of H2A clipping.

2. Introduction

The evolutionary conserved histone proteins are intimately associated with DNA to form the chromatin. Their amino-terminal tail and the carboxy-tail of histone H2A protrude from the nucleosome and can become modified by many different PTM. One somewhat dramatic PTM is the enzymatic clipping of histone tails, which was recently shown to play an epigenetic role in differentiation [1, 2]. Although many previous reports on histone clipping have speculated on its potential to influence transcription [3], histone truncation has equally been described in the context of other biological processes, such as neutrophil extracellular trap or NET formation [4].

During a proteome analysis on haematopoietic cells, we came across a histone clipping event that was first described over 35 years ago in calf thymus histone extracts: truncation of the C-tail of histone H2A at V_{114} [5]. The responsible enzyme was named the 'H2A specific protease' (H2Asp) and its involvement in transcriptional regulation was soon hypothesized but it was never sequenced nor identified [6-10]. Structurally, this truncation has been suggested to modulate chromatin dynamics and it was shown to be induced during macrophage differentiation [11, 12]. Even today, references to this histone H2A C-tail clipping and the H2Asp still recur, yet no reports have thus far questioned the identity of the enzyme or its involvement in epigenetics [13-16].

Here we show that the H2Asp actually is Neutrophil Elastase (NE). While we continue to search for the biological potential of this truncation in health or disease, we emphasize that the clipping of the histone H2A C-tail shows remarkable parallels to the more epigenetically established clipping of the H3 N-tail, but could equally be involved in other biological processes such as NET formation. Importantly, we caution that most experimental approaches still do not anticipate on these clipping events while high enzyme kinetics and strong association between nucleosomes could greatly hamper efficient enzyme inhibition in any protocol.

3. Materials and Methods

3.1. Cells and reagents

Phosphate buffered saline (PBS), media, L-glutamine, Fetal bovine serum (FBS), penicillin/streptomycine, Dynabeads and SYPRO Ruby were from Invitrogen (San Diego, CA), ammonium bicarbonate (ABC), sodium dodecyl sulfate (SDS), N-cyclohexyl-3-aminopropanesulfonic acid (CAPS) and Tween-20 from Millipore (Billerica, MA). All other reagents were purchased from Sigma Aldrich (St. Louis, MI, USA). Bovine histones (cat.no. 223565) were from Roche (Basel, Switzerland), recombinant human H2A (M2502S) from New England Biolabs (Ipswich, MA) and purified Neutrophil Elastase from

Abcam (Ab80475, Cambridge, UK). Raji and Jurkat cells were obtained from ATCC. Total leukocytes were obtained by red blood cell lysis on whole blood (Qiagen, Venlo, Netherlands). Lymphocytes were isolated from healthy volunteers using Ficoll-Paque. T-cells were purified by means of the RosetteSep® Human T Cell Enrichment Cocktail (Stem cell Technologies, Grenoble, France). Cells were cultured in Dulbecco's Modified Eagle medium supplemented with 2% (w/v) L–glutamine, 10% (w/v) FBS and 50 IU/ml penicillin/streptomycine at 37°C. Apoptosis in Raji's was induced by overnight incubation in 0.2µg/ml MHCII antibody or 100 ng/ml PMA at 37°C or by 15min UV irradiation.

3.2. Mice

Null mice *stock# 006112 strain# B6.129X1-Elanetm1Sds/J* and wild type control mice *stock# 005304 Strain# C57BL/6NJ* were purchased from The Jackson Laboratory (Bar Harbor, Ma, USA) and bred at the animal facility of Ghent University Hospital. All experimental procedures were approved by the local ethics committee according to national animal welfare legislation. All mice were genotyped as instructed by the supplier [17]. Blood samples were obtained by cardiac puncture of mice under terminal anaesthesia. Leukocytes were recovered by red blood cell lysis [18] and histones were isolated as described below.

3.3. Histone isolation

All steps were done at 4°C. Harvested cells were washed twice in PBS containing 1mM PMSF and protease inhibitor cocktail. 10^7 cells/ml were resuspended in Triton extraction buffer (PBS containing 0.5% (v/v) Triton X 100, 1mM PMSF and protease inhibitor cocktail) and lysed by gentle stirring. Pelleted nuclei were washed in PBS containing 1mM PMSF and proteinase inhibitor cocktail. Histones were extracted overnight after benzonase treatment of the sonicated nuclei by acid extraction: incubation in 250µl 0.2M HCl at 4°C with gentle stirring [19]. Precipitated proteins were pelleted and the supernatant containing the histones was dried and stored at -80°C until further use. Protein quantification was done by Bradford Coomassie Assay from Thermo (St Waltham, MA, USA).

3.4. Western blot analysis

3µg of dried histone extract was resuspended in Laemmli buffer, incubated at 99°C and run on a 15% TrisHCl Gel (Biorad, Hercules, CA) and transferred to nitrocellulose membrane in a 10 mM CAPS buffer with 20% MeOH (Merck) in 100 min at 120V. Histone H2A antibody (LS-C24265, LifeSpan BioScience, Seattle, WA) was used at a 1:1000 dilution in 0.3% Tween20, 1%BSA [20]. For the detection of biotinylated H2A 1:10 000 Avidin-HRP (18-4100-94) (eBioscience, San Diego, CA) was used [19].

3.5. Mass spectrometry analysis

All tryptic digests were performed after reduction in freshly prepared 50mM TCEP 25mM TEABC solution for 2h at 56°C and alkylation in 200mM MMTS solution for 1h at room temperature. The samples were separated on an U3000 (Dionex) in a 70min organic gradient from 4-100% buffer B (80% (v/v) ACN in 0.1% (v/v) FA) and analyzed on an ESI Q-TOF Premier (Waters, Wilmslow, UK). Data was searched against SwissProt database using Mascot 2.3 (Matrix Science, London, UK). For the specific analysis of the H2A V_{114} clipping, the isobaric peptides (AQUA1; AQUA2) were combined in a stock solution in a ten (AQUA1, Thermo) to one (AQUA2, Sigma Aldrich) ratio and added to 1µg of histone extract right before MS analysis [21].

3.6. Sucrose gradient ultracentrifugation

Mechanical isolation of nuclei was done by lysing the cells in a cell plunger after 5 minutes of incubation in 10 mM Hepes pH 8, 10 mM KCl, 1,5 mM MgCl2, 0,5 mM DTT at 4°C. After 10 strokes, nuclear isolation efficiency was found to be at least 90% by microscopy [22]. The pelleted nuclei were resuspended in 3ml buffer S1 (0,25 M Sucrose, 10 mM MgCl2) and brought on top of 3ml buffer S2 (0,35 M Sucrose, 0,5 mM MgCl2). After 5 min at 572xg at 4 °C, the pellet was resuspended in S2 for a second wash and the nuclei were lysed by sonication in 50mM Tris-HCL pH7.5 supplemented with 250U of benzonases for 10 min for DNA degradation. The linear sucrose gradient (5ml) 10-50% was prepared in 0.01M Tris-HCL pH7.5 with 0.001M Na_2EDTA. The gradient was placed at 4°C for 2h. 200µl of the nuclear extract was applied on top of the gradients and they were spun at 190000g for 1040 min at 4°C in a Centrikon T1080 ultracentrifuge [23]. After centrifugation, 10 fractions of 500µl were collected manually.

4. Results and Discussion

4.1. High Throughput Quantification of H2AV_{114} clipping by AQUA

During a mass spectrometric proteome analysis on haematopietic cells, we detected only one semi-tryptic peptide out of >7400 annotated MS/MS spectra that recurred in all 6 replicate runs: VTIAQGGVLPNIQAV, a fragment derived from histone H2A ending at V_{114} (clipped H2A or cH2A, Figure 1A). Remarkably, this specific fragment was already described in calf thymus in 1976 but the responsible enzyme, named H2A specific protease (H2Asp) was never identified [5]. To be able to specifically quantify the V_{114} clipping, a sensitive mass spectrometry approach based on the AQUA (Absolute Quantitation) principle was developed using two isotopically labeled synthetic peptides (Figure 1A and 1B). To our knowledge, this is the first description of a technique that can specifically

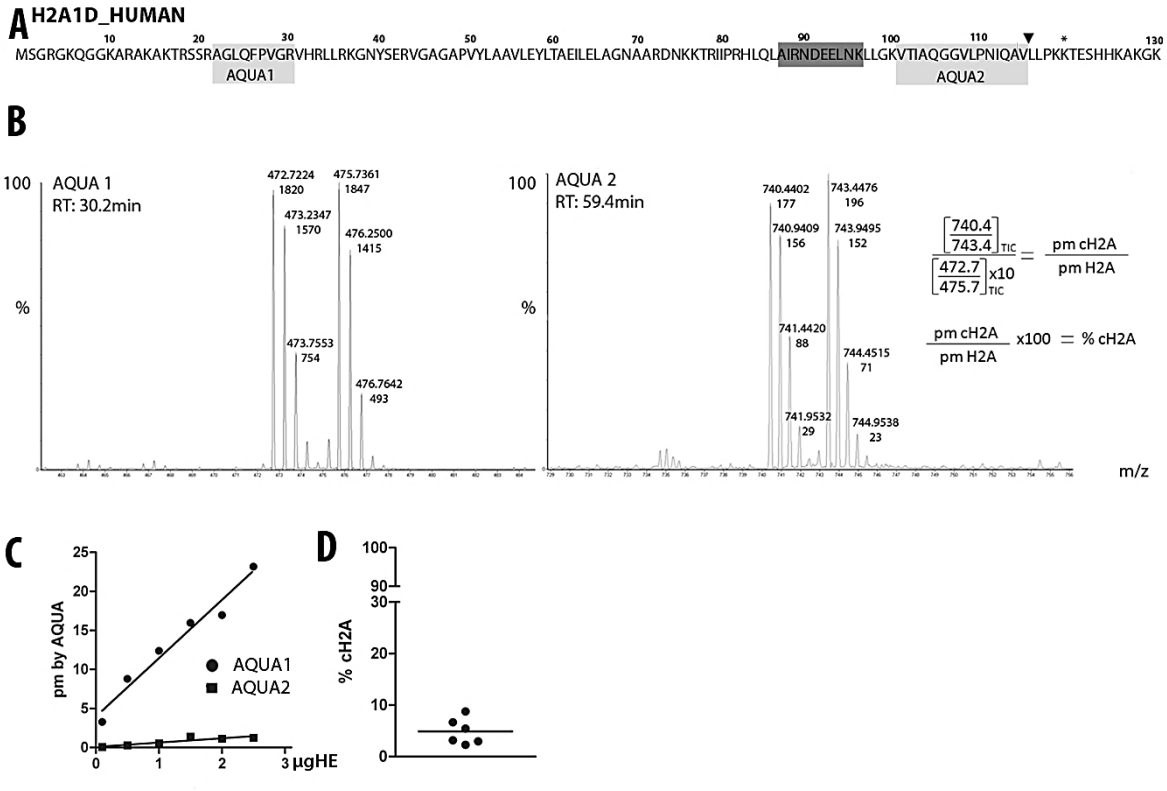


Figure 1. Accurate quantification of the $H2AV_{114}$ clipping event. (A) Histone H2A sequence. Epitope for Western blot detection is highlighted in dark grey (cH2A = 12 kDa). Sequences boxed in light grey are used for the AQUA approach. Asterisk indicates K119 ubiquitination site. Arrowhead: $H2AV_{114}$ clipping. (B) MS spectrum of a histone extract spiked with both AQUA peptide sequences indicated in light grey in (A). Both synthetic peptides are isotopically labelled with 13C and 15N at the proline increasing their molecular weight with 6 Dalton, which results in a 3 *m/z* (2+) shift. 10 pmol of the N-terminal tryptic AQUA1 peptide AGLQFPVGR (*m/z* 475.7) and 1 pmol of the specific semi-tryptic C-terminal AQUA2 peptide VTIAQGGVLPNIQAV_{114} (*m/z* 743.4), allow to quantify the relative amount of H2A that is clipped (cH2A) in a given sample according to the formula given on the right. This 10-fold difference in loading compensates for differences in ionization efficiency and in-source decay. Peaks are annotated with the *m/z* on the top and the TIC (Total Ion Count) below. RT, retention time. (C) The AQUA1 peptide quantifies the total amount of H2A present and can thus be used to eliminate all sample variation due to differential protein composition of the histone extract. To illustrate the need for the AQUA1 peptide 0.1–2.5 µg of bovine histone extract was loaded with 10 pmol of AQUA1 and 1 pmol AQUA2. The quantity of both total H2A (determined by AQUA1) and cH2A (determined by AQUA2) can be absolutely determined (left axis, absolute quantity). (D) The formula was used to calculate the % cH2A in all samples analyzed in (B) and predicted the clipping within a 6% range despite the 25-fold loading difference.

quantify a histone clipping event in high throughput. To validate the efficiency of this approach, we quantified the amount of clipped H2A (cH2A) in 0.1 to 2.5 µg of bovine histones and consistently found it to be between 3-9%, despite the 25-fold loading difference (Figure 1C and 1D). Note that this cH2A was detected in commercial calf thymus histones, the source from which the H2A specific protease was first isolated [5].

4.2. The H2A specific protease is Neutrophil Elastase

Screening larger cohorts of histone extracts from different haematopoietic origins indicated that cH2A prevalence was strongly related to myeloid content since PBMCs, leukocytes and the pellet from Ficoll-Paque separated haematopoietic cells showed the presence of cleaved H2A (Figure 2A, left panel). In T-cells, Raji's and Jurkat cells only intact H2A was detected. Lymphoid Raji cells were subjected to different forms of stress to exclude any aspecific degradation event (Figure 2A, middle panel). No cH2A could be detected after overnight incubation with MHC II antibodies or PMA, nor after 15min of UV irradiation, while all these treatments induced cell death as seen by trypan blue staining. For all abovementioned samples the presence or absence of the specific V_{114} clipping site was analyzed by AQUA. Eickbush et al. (1976) described that the H2Asp itself is co-extracted during an acid histone extract. Histones were thus isolated from samples with low myeloid content (PBMC) without protease inhibitors to obtain samples with active H2Asp. Incubation for 30 minutes at 37°C corroborated the presence of active H2Asp, which also clipped K_{119} ubiquitinated H2A (Figure 2A, right panel). Next, a constant level of bovine histones (10 µg) was incubated by an increasing amount of this extract (1 ng to 1 µg) for 2 hours at 37°C (0.05–1 µg are shown in Figure 2B). Bovine histone H2A was the histone being clipped most at the lowest amount of extract, confirming that $H2AV_{114}$ is a very sensitive clipping site.

When these histone extracts containing H2Asp activity were analyzed by mass spectrometry only one serine protease was identified in these histone extracts: Neutrophil Elastase (NE), previously known as medullasin. This finding agrees perfectly with all previous reports on the H2Asp [5-10]: (i) NE is a serine protease [24], (ii) which is inactivated at low pH, (iii) activated by high salt content [25], (iv) is inactivated by (calf thymus) DNA [26], (v) it is transcribed in the thymus, one of the few tissues where NE mRNA can be found (Hs.99863 in EST Profile Viewer, NCBI), (vi) is extremely basic (pI 9.9) making it very susceptible to co-extraction during acid histone isolation, (vii) it has known valine preference [27] and a predicted cleavage site at $H2AV_{114}$ with >99% specificity in SitePrediction [28], (viii) ubiquitinated H2A is a substrate, and finally (ix) we too found it to associate strongly with nucleosomes in our sucrose gradient ultracentrifugation experiments on proteins extracted from isolated nuclei (*see section 3.4*).

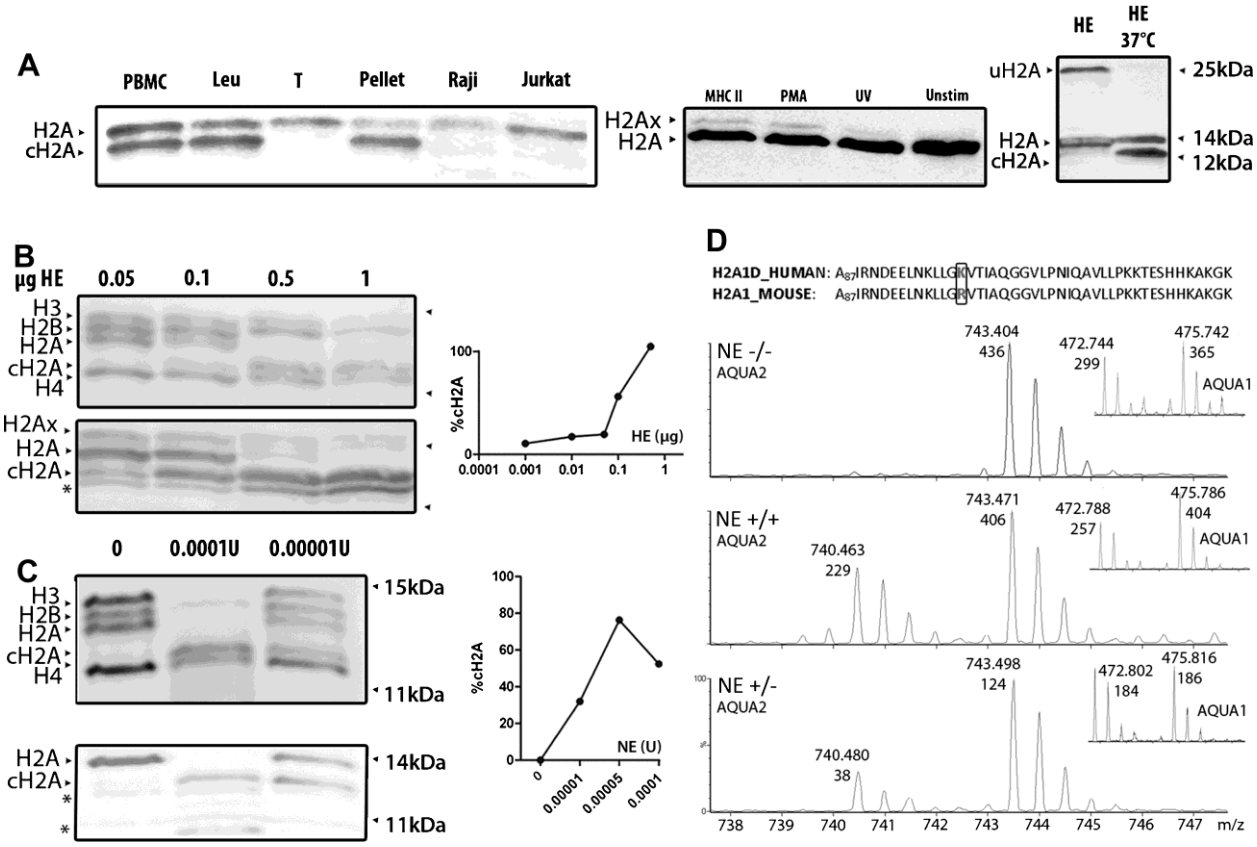


Figure 2. Neutrophil Elastase in the capacity of the H2A specific protease. (A) cH2A is present in the total leukocytes (Leu), buffy coat (PBMCs) and pellet from Ficoll-Paque separated haematopoietic cells (left panel). Purified T-cells (T), nor lymphoid Raji or Jurkat cell lines showed any cH2A. Raji cells incubated overnight with MHCII antibodies or PMA, or irradiated for 15 min with UV lacked cH2A, excluding aspecific degradation (middle panel). Histones extracted in the absence of inhibitors (i. e. PMSF) show H2Asp activity coordinately clipping ubiquitinated H2A (uH2A, identified by MS) (right panel). (B) 1 ng to 1 µg of leucocyte histone extract was incubated with 10 µg of buffered bovine histones for 2 h at 37 °C (0.05–1 µg are shown). SYPRO stain (upper panel) and H2A western blot (lower panel) illustrate the presence of active H2Asp in these histone extracts and show the susceptibility of histone H2A. AQUA quantification confirms the V_{114} clipping site (1 ng to 1 µg is shown in right panel). Asterisks: additional degradation fragment. (C) Commercially purified NE clips all bovine histones when incubated at only 0.0001 U (upper panel: SYPRO-stain). Histone H2A is the only substrate at 0.00001 U (lower panel: H2A blot). cH2A AQUA quantification of Raji histones (RH) incubated with 0–0.005 U of NE. A High NE concentration impairs accurate quantification by AQUA (right panel). (D) Histone extracts from leukocytes from five mice per phenotype were pooled and analyzed by AQUA. The only mutated amino acid compared to human is indicated by the box. Both K and R are tryptic cleavage sites. cH2A is found in histone extracts from leukocytes from both wild type and heterozygous mice (NE+/+ & +/- lower two

panels) but is undetectable in NE Null mice (NE -/- upper panel). Upper right corner: AQUA 1 peptide (475.7 *m/z*) used to quantify the total amount of H2A.

4.3. Confirmation of NE in its capacity of the H2Asp

Next, both histones from Raji cells and bovine histones were incubated with commercially available purified NE. This results in cleavage of all histones in a dose dependent manner (Figure 2C). The specific V_{114} fragment was the only detected fragment in the lowest concentration range (0.00001U), illustrating the susceptibility of the substrate. Ultimate proof was generated by comparing isolated leukocytes from WT, heterozygous $NE^{+/-}$ and NE *Null* mice using the AQUA approach (Figure 2D). The only mutated amino acid compared to human is indicated by the box. Both K and R are tryptic cleavage sites and when samples are prepared by this enzymatic digestion, the peptide sequences corresponding to the AQUA peptides will be formed in both. As in human leukocytes, the clipping was present in control and heterozygous mouse leukocytes. However, the clipping of H2A at V_{114} could not be detected in the NE *Null* mice.

4.4. NE strongly interacts with nucleosomes in vitro

For histone H2A clipping to occur *in vivo*, requires both enzyme and substrate to physically interact. A strong association of the H2Asp to chromatin and nucleosomes was reported during earlier studies [7, 8]. Equally, nuclear substrates of NE have been reported after NE migrates to the nucleus *in vivo* [29]. To verify if NE from leukocytes is present in the nucleus and associates with nucleosomes, we isolated nuclei using mechanical disruption after red blood cell lysis and fractionated the proteins by differential ultracentrifugation on a sucrose gradient from 10 to 50%. Ten different fractions were collected and partially loaded onto a 1D PAGE to visualize the fractionation (Figure 3). The remaining part of each fraction was digested with trypsin to identify the main proteins present. Remarkably, the only protein that was identified solely in all fractions containing the nucleosomes in the gradient was NE. This illustrates that NE strongly associates with and is being “pulled down” by the nucleosomes.

Although this finding supports a potential physical interaction in the nucleus, we cannot exclude that the presence of the NE, and H2Asp in other reports, is in fact a cytoplasmic contaminant, as we could also detect other cytoplasmic proteins in the nuclear extracts. On the other hand, it can be concluded that NE will strongly attach to the nucleosomes when they are colocalized, since the other proteins did not co-migrate with the nucleosomes in the gradient.

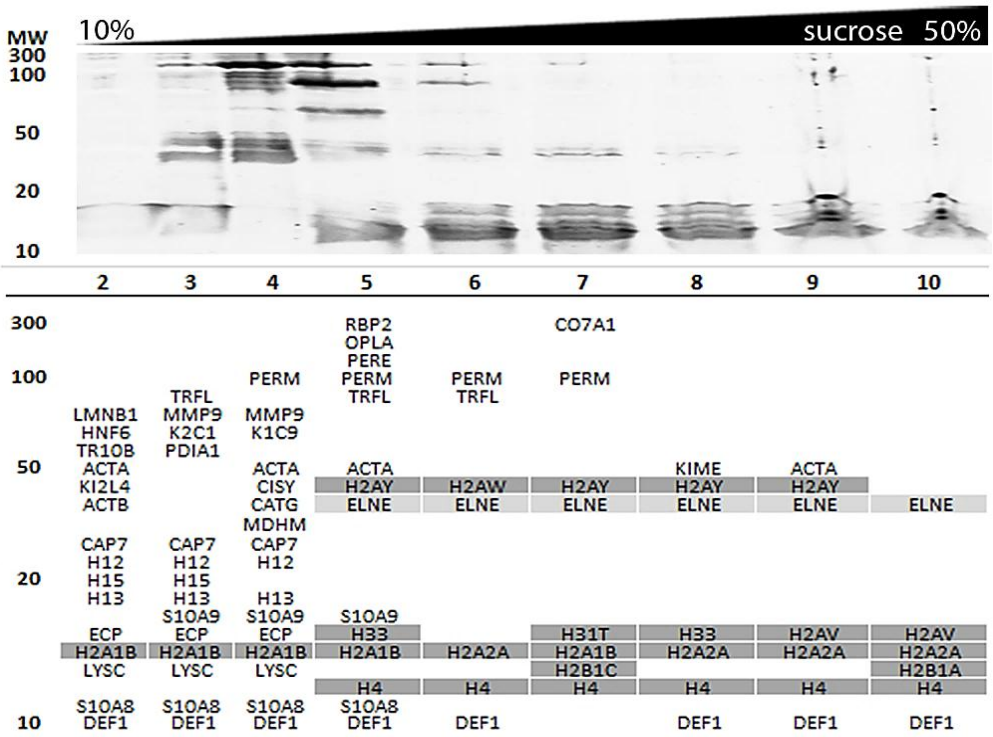


Figure 3. Strong association between NE and nucleosomes in differential sucrose gradient ultracentrifugation. Protein extracts from mechanically isolated nuclei were fractionated on a sucrose gradient to simultaneously verify direct associations between nucleosomes and NE. Ten different fractions were collected from 10% to 50% linear sucrose gradient after overnight ultra-centrifugation at 190,000 × g. These fractions were loaded onto a 1D PAGE and stained with SYPRO staining to visualize the efficiency of the fractionation (upper panel, image vertically compressed). Note that the increasing sucrose content of these samples greatly reduces band pattern resolution towards the 50% sucrose fractions. All fractions were coordinately analyzed by mass spectrometry (lower panel). Proteins are annotated by their Uniprot accession name where "*_HUMAN" is discarded. Core histone isoforms are not indicated. Proteins are roughly organized by their molecular weight (MW) for comparison to the gel image in the upper panel. All higher density lanes (5–10) are enriched for nucleosomes, as seen by the presence of the core histones highlighted in dark grey and as seen by their characteristic band pattern in the gel. NE (here defined as ELNE) was the only protein to have specifically co-migrated with the nucleosomes towards higher densities

Because the DNA was enzymatically degraded in these extracts, this finding is in line with the notion that NE might directly bind the histone H3 N-tail as was suggested for the H2Asp earlier [9]. Very high enzyme activity to specific substrate sequences, as seen here for NE-mediated $H2AV_{114}$ clipping, has also been suggested to be a possible indicator of functional relationship [30]. Taken together however, the very strong association with the nucleosomes and the high enzyme kinetics towards histone H2A make complete inactivation of this enzymatic reaction extremely difficult. Notably, by spiking biotinylated recombinant H2A during the histone extraction procedure and staining the subsequent western blot with avidin we found that not all *in vitro* activity could be inhibited in isolated cell populations with high myeloid content in our hands (*data not shown*). Because this is not the first report on pitfalls during *in vitro* inhibition of NE [31], we caution that unintentional histone clipping during cell disruption can strongly influence many epigenetic experimental approaches such as ChIP seq and western blot.

The annotation of the H2Asp as NE thus is not only an essential first step towards elucidating any potential biological role of this enzymatic interaction, but equally provides the necessary background to critically revise earlier reports of $H2AV_{114}$ clipping and coordinately increases awareness on histone clipping.

4.5. The H2Asp is Neutrophil Elastase: broadening the biological perspective

Nearly all previous reports on $H2AV_{114}$ clipping and the H2Asp have speculated on the potential transcriptional implications of such reaction [5-9, 11, 13, 16]. Although we do not provide evidence of epigenetic implications here, by expressing the V_{114} truncated cH2A directly, Vogler et al. have at least demonstrated the viability of cells with this truncated H2A [12]. Even more, they also showed that this led to increased histone exchange kinetics and nucleosome mobility *in vivo* and *in vitro,* in line with what was found for H2Asp-mediated clipping of H2A *in vitro* [7]. An interesting parallel thus emerges when considering that retinoic acid (RA)-induced differentiation of THP-1 promonocytes into macrophages is briefly accompanied by histone $H2AV_{114}$ clipping [11], just as H3 clipping by cathepsin L (CL) is induced while mouse embryonic stem cells (mESC) are differentiated with RA [1]. Both the H2A carboxy-tail (V_{114}) and the H3 amino-tail (A_{21}) clipping sites are located adjacent at the entry and exit points of the DNA. Of note, while the clipping of histone H3 is considered essential for early mESC differentiation, Duncan et al. argued that the viability of CL *Null* mice might be due to functional redundancy [1]. We found that histone H2A isolated from NE *Null* mice showed additional bands on western blot that are similar in molecular weight, but are not derived from V_{114} clipping as seen by our specific AQUA approach (*data not shown*). Just as for CL *Null* mice, NE *Null* mice are viable and do not show intrinsic deficiencies in hematopoietic differentiation [17]. Yet, this could equally be seen as an argument against its

involvement in epigenetics. Notably, the NE *Null* mice do show a reduced capacity to form immunological structures called neutrophil extracellular traps or NET [4]. This process involves the active migration of NE and myeloperoxidase to the nucleus in neutrophils followed by chromatin decondensation and externalization of the DNA to entrap microbial pathogens. It would thus be of great interest to verify the prevalence of the $H2AV_{114}$ clipping site in NET.

Taken together, the fact that the H2Asp actually is NE now surfaces an alternative biological process apart from the epigenetic context in which it has been primarily described up to now. However, it will remain challenging to uncouple any histone clipping event from continuous histone turnover and degradation [32-34].

References

[1] Duncan, E. M., Muratore-Schroeder, T. L., Cook, R. G., Garcia, B. A*., et al.*, Cathepsin L proteolytically processes histone H3 during mouse embryonic stem cell differentiation. *Cell* 2008, *135*, 284-294.
[2] Osley, M. A., Epigenetics: how to lose a tail. *Nature* 2008, *456*, 885-886.
[3] Duncan, E. M., Allis, C. D., Errors in erasure: links between histone lysine methylation removal and disease. *Progress in drug research. Fortschritte der Arzneimittelforschung. Progres des recherches pharmaceutiques* 2011, *67*, 69-90.
[4] Papayannopoulos, V., Metzler, K. D., Hakkim, A., Zychlinsky, A., Neutrophil elastase and myeloperoxidase regulate the formation of neutrophil extracellular traps. *J Cell Biol* 2010, *191*, 677-691.
[5] Eickbush, T. H., Watson, D. K., Moudrianakis, E. N., A chromatin-bound proteolytic activity with unique specificity for histone H2A. *Cell* 1976, *9*, 785-792.
[6] Eickbush, T. H., Moudrianakis, E. N., The histone core complex: an octamer assembled by two sets of protein-protein interactions. *Biochemistry* 1978, *17*, 4955-4964.
[7] Eickbush, T. H., Godfrey, J. E., Elia, M. C., Moudrianakis, E. N., H2a-specific proteolysis as a unique probe in the analysis of the histone octamer. *J Biol Chem* 1988, *263*, 18972-18978.
[8] Davie, J. R., Numerow, L., Delcuve, G. P., The nonhistone chromosomal protein, H2A-specific protease, is selectively associated with nucleosomes containing histone H1. *J Biol Chem* 1986, *261*, 10410-10416.
[9] Elia, M. C., Moudrianakis, E. N., Regulation of H2a-specific proteolysis by the histone H3:H4 tetramer. *J Biol Chem* 1988, *263*, 9958-9964.
[10] Watson, D. K., Moudrianakis, E. N., Histone-dependent reconstitution and nucleosomal localization of a nonhistone chromosomal protein: the H2A-specific protease. *Biochemistry* 1982, *21*, 248-256.
[11] Minami, J., Takada, K., Aoki, K., Shimada, Y*., et al.*, Purification and characterization of C-terminal truncated forms of histone H2A in monocytic THP-1 cells. *Int J Biochem Cell Biol* 2007, *39*, 171-180.
[12] Vogler, C., Huber, C., Waldmann, T., Ettig, R*., et al.*, Histone H2A C-terminus regulates chromatin dynamics, remodeling, and histone H1 binding. *PLoS genetics* 2010, *6*, e1001234.
[13] Okawa, Y., Takada, K., Minami, J., Aoki, K*., et al.*, Purification of N-terminally truncated histone H2A-monoubiquitin conjugates from leukemic cell nuclei: probable proteolytic products of ubiquitinated H2A. *Int J Biochem Cell Biol* 2003, *35*, 1588-1600.
[14] Santos-Rosa, H., Kirmizis, A., Nelson, C., Bartke, T*., et al.*, Histone H3 tail clipping regulates gene expression. *Nat Struct Mol Biol* 2009, *16*, 17-22.
[15] Mandal, P., Azad, G. K., Tomar, R. S., Identification of a novel histone H3 specific protease activity in nuclei of chicken liver. *Biochem Biophys Res Commun* 2012, *421*, 261-267.
[16] Azad, G. K., Tomar, R. S., Proteolytic clipping of histone tails: the emerging role of histone proteases in regulation of various biological processes. *Mol Biol Rep* 2014.
[17] Belaaouaj, A., McCarthy, R., Baumann, M., Gao, Z*., et al.*, Mice lacking neutrophil elastase reveal impaired host defense against gram negative bacterial sepsis. *Nat Med* 1998, *4*, 615-618.
[18] Cameron, I. L., Fullerton, G. D., A model to explain the osmotic pressure behavior of hemoglobin and serum albumin. *Biochem Cell Biol* 1990, *68*, 894-898.
[19] Vossaert, L., Meert, P., Scheerlinck, E., Glibert, P*., et al.*, Identification of histone H3 clipping activity in human embryonic stem cells. *Stem cell research* 2014, *13*, 123-134.
[20] Vermoote, M., Van Steendam, K., Flahou, B., Smet, A*., et al.*, Immunization with the immunodominant Helicobacter suis urease subunit B induces partial protection against H. suis infection in a mouse model. *Veterinary research* 2012, *43*, 72.
[21] Stemmann, O., Zou, H., Gerber, S. A., Gygi, S. P., Kirschner, M. W., Dual inhibition of sister chromatid separation at metaphase. *Cell* 2001, *107*, 715-726.
[22] Lodish H, B. A., Zipursky SL, et al., *Molecular Cell Biology*.
[23] Norton, W. T., Poduslo, S. E., Myelination in rat brain: method of myelin isolation. *Journal of neurochemistry* 1973, *21*, 749-757.
[24] Korkmaz, B., Moreau, T., Gauthier, F., Neutrophil elastase, proteinase 3 and cathepsin G: physicochemical properties, activity and physiopathological functions. *Biochimie* 2008, *90*, 227-242.

[25] Lestienne, P., Bieth, J. G., Activation of human leukocyte elastase activity by excess substrate, hydrophobic solvents, and ionic strength. *J Biol Chem* 1980, *255*, 9289-9294.
[26] Lestienne, P., Bieth, J. G., Inhibition of human leucocyte elastase by polynucleotides. *Biochimie* 1983, *65*, 49-52.
[27] Schilling, O., Barre, O., Huesgen, P. F., Overall, C. M., Proteome-wide analysis of protein carboxy termini: C terminomics. *Nat Methods* 2010, *7*, 508-511.
[28] Verspurten, J., Gevaert, K., Declercq, W., Vandenabeele, P., SitePredicting the cleavage of proteinase substrates. *Trends Biochem Sci* 2009, *34*, 319-323.
[29] Lane, A. A., Ley, T. J., Neutrophil elastase cleaves PML-RARalpha and is important for the development of acute promyelocytic leukemia in mice. *Cell* 2003, *115*, 305-318.
[30] Timmer, J. C., Zhu, W., Pop, C., Regan, T.*, et al.*, Structural and kinetic determinants of protease substrates. *Nat Struct Mol Biol* 2009, *16*, 1101-1108.
[31] Groth, I., Alban, S., Elastase inhibition assay with peptide substrates - an example for the limited comparability of in vitro results. *Planta medica* 2008, *74*, 852-858.
[32] Ivanov, A., Pawlikowski, J., Manoharan, I., van Tuyn, J.*, et al.*, Lysosome-mediated processing of chromatin in senescence. *J Cell Biol* 2013, *202*, 129-143.
[33] Qian, M. X., Pang, Y., Liu, C. H., Haratake, K.*, et al.*, Acetylation-mediated proteasomal degradation of core histones during DNA repair and spermatogenesis. *Cell* 2013, *153*, 1012-1024.
[34] Deal, R. B., Henikoff, J. G., Henikoff, S., Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* 2010, *328*, 1161-1164.

PART 3: PHOSPHO-ITRAQ, “ASSESSING ISOBARIC LABELS FOR THE LARGE-SCALE STUDY OF PHOSPHOPEPTIDE STOICHIOMETRY”

Based on:

Phospho-iTRAQ: Assessing Isobaric Labels For The Large-Scale Study Of Phosphopeptide Stoichiometry

Pieter Glibert[1], Paulien Meert[1], Katleen Van Steendam[1], Filip Van Nieuwerburgh[1],

Lennart Martens[2], Maarten Dhaenens[1†] and Dieter Deforce[1†*]

[1] Laboratory of Pharmaceutical Biotechnology, Ghent University, 460 Ottergemsesteenweg, B-9000 Ghent, Belgium; E-Mails: pieter.glibert@ugent.be (P.G.); paulien.meert@ugent.be (P.M.); katleen.vansteendam@ugent.be (K.V.S.); filip.vannieuwerburgh@ugent.be (F.V.N.); maarten.dhaenens@ugent.be (M.D.)

[2] Department of Medical Protein Research, VIB, and Department of Biochemistry, Ghent University, 3 Albert Baertsoenkaai , B-9000 Ghent, Belgium; E-Mail: lennart.martens@UGent.be (L.M.)

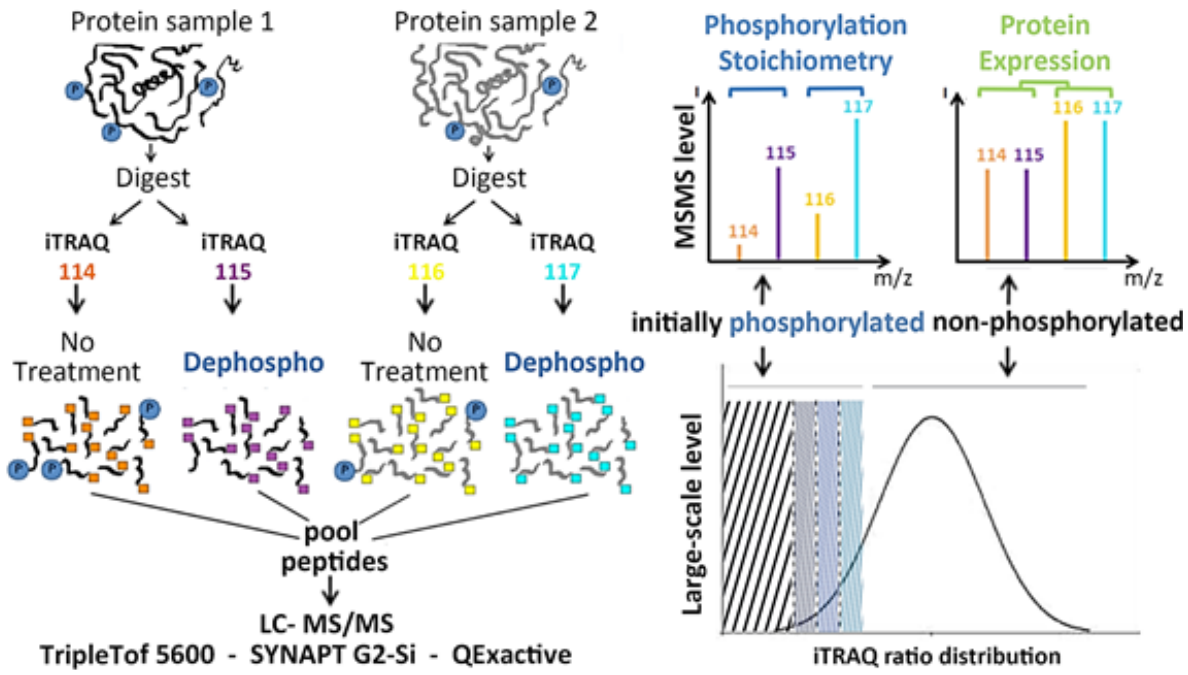
[†] These authors contributed equally to this work.

[*] Author to whom correspondence should be addressed; E-Mail: dieter.deforce@ugent.be; Tel.: +32-9-264-8067; Fax: +32-9-220-6688.

Submitted for publication:

Journal of Proteome Research (on August 2014, under review)

1. Abstract



Distinguishing between phosphopeptides of high and low stoichiometry is essential to discover the true extent of protein phosphorylation. We here extend the strategy whereby a peptide sample is briefly split in two identical parts and differentially labeled preceding the phosphatase treatment of one part. Our use of isobaric tags for relative and absolute quantitation (iTRAQ) marks the first time that isobaric tags have been applied for the large-scale analysis of phosphopeptides. Our Phospho-iTRAQ method focuses on the unmodified counterparts of phosphorylated peptides, thus circumventing the ionization, fragmentation and phospho-enrichment difficulties that hamper quantification of stoichiometry in most common phosphoproteomic methods. As iTRAQ enables multiplexing, simultaneous (phospho)proteome comparison between internal replicates and/or multiple samples are possible. The technique was validated on multiple instrument platforms by adding internal standards of high stoichiometry to a complex lysate of control and EGF-stimulated HeLa cells. To demonstrate the flexibility of Phospho-iTRAQ with regards to the experimental set-up used, the proteome coverage was extended through gel fractionation. We specifically focus on the discovery of high stoichiometry peptides and describe an approach to filter out phosphopeptides from complex datasets.

---

2. Introduction

One of the most extensively studied protein modifications is phosphorylation, a reversible posttranslational modification (PTM) with a central role in a broad range of cellular processes such as cell metabolism, homeostasis, transcription and apoptosis. Due to their transient character, phosphorylations initiate and propagate signal transduction pathways, making these fundamental to inter- and intracellular communication [1, 2]. Over 2% of the genome encodes for kinases, the family of enzymes that catalyze the transfer of a phosphate group to serine, threonine or tyrosine residues. It is estimated that 30% of all proteins are phosphorylated, illustrating the high prevalence of this PTM [3]. Detection of phosphorylation is nevertheless challenging considering the dynamic regulation, low stoichiometry and heterogeneous character of the phosphosites [4]. Numerous dedicated phosphoproteomic workflows have therefore been developed in order to unravel phosphorylation networks and the activity of their modulating enzymes (as extensively reviewed in [2, 5]).

In these workflows, phosphopeptides are typically enriched prior to mass spectrometry analysis to remove the strong background interference caused by the bulk of non-phosphorylated peptides [4]. Apart from requiring thorough optimization and a corresponding high level of expertise of the analyst however, enrichment is also responsible for sample loss and limited quantification accuracy [5]. Subsequent mass spectrometry analysis is furthermore confounded by low ionization efficiency and inferior fragmentation of phosphopeptides. Conventional MS/MS analysis with collision-induced dissociation predominantly results in the neutral loss of the phosphate group, leading to poor sequence coverage due to reduced backbone fragmentation. An additional stage of fragmentation in MS3 or the use of alternative fragmentation technologies such as higher energy C-trap dissociation or electron transfer dissociation can partially facilitate the detection and identification of phosphopeptides [3, 6]. While these methodologies are not yet routinely applied in most laboratories, they have allowed multiple specialized studies to report a grand total of over 100 000 phosphorylation sites [2].

In order to draw accurate biological conclusions it is mandatory to obtain quantitative information on the phospho-modifications. Most quantitative approaches aim to elucidate the relative phosphorylation changes between samples, employing direct label-free comparisons of phospho-enriched fractions, or metabolic or chemical labeling methods such as SILAC, iTRAQ or tandem mass tag (TMT) [5-7]. Such methods however, forego the investigation of phosphorylation stoichiometry, also known as occupancy, since such analysis requires the simultaneous monitoring of the unphosphorylated counterpart [8]. Yet, keeping this complete picture is crucial e.g. for *in vitro* studies of specific kinases, where artificially high enzyme concentrations or prolonged incubation times often result in the annotation of low-level spurious phosphosites that are biologically irrelevant [5, 9]. The stoichiometry

measurement of a single phosphosite is possible by spiking specific isotopically labeled (AQUA) peptides of that particular phosphorylated peptide and its unmodified complement but this strategy is inconvenient for large-scale applications and requires prior insight of the phosphorylation site. Additionally, for comparative studies, the bias of the sample preparation steps cannot be excluded with AQUA and specific knowledge of the relative protein concentrations is not available as with the former mentioned labeling approaches [10].

One interesting labeling strategy that is specifically designed to quantify occupancy circumvents the common difficulties of standard phosphopeptide detection by focusing on the non-phosphorylated counterparts of the phosphopeptides [11]. Herein, a peptide sample is briefly split in two identical parts which are differentially labeled, preceding a phosphatase treatment of one part and mock-treatment of the other. Immediately afterwards the two parts are recombined for the subsequent LC-MS/MS analysis. Dephosphorylation of a phosphopeptide will induce a 79.979 mass shift resulting in a skewed label ratio in the unphosphorylated peptide's reporter region which now specifies the phosphorylation stoichiometry [12]. Although the principle has been described previously [12-19], only one phosphatase-based workflow was adapted specifically for the large-scale study of a complete lysate whereby known phosphopeptides present in a database were quantified by isotopic labels [20]. Although this report is extensively referenced, the method is not yet routinely applied for discovery purposes. As an alternative approach, each half of a peptide mixture can be labeled with isobaric tags prior to the phosphatase treatment of one half. By virtue of the complete multiplexing until the fragmentation stage and subsequent quantification at the MS2 level, isobaric labels have the capacity to analyze up to ten samples simultaneously [21]. Previous reports have already prototyped this strategy with TMT and iTRAQ labels on relatively simple protein samples, but have concentrated mainly on the precision of the stoichiometry measurement [11, 12, 19, 22].

Here we extend for the first time the isobaric tag equivalent of this approach for the large-scale quantitative phospho-analysis of complex mixtures. The technique was validated on different instrument platforms during a comparative analysis of control and epidermal growth factor (EGF)-stimulated HeLa cells [23]. We discuss a data analysis approach for the discovery of high stoichiometry phosphopeptides in complex peptide mixtures and present a way to interpret the sizable dataset without matching the peptide identifications to a phospho-database. To monitor all eluting peptides and not only the phosphorylated fraction of the proteome, our approach does require instruments with high-speed acquisition capabilities. At the same time, our method is very flexible in terms of the experimental set-up, allowing the implementation of extensive fractionation steps such as gel-electrophoresis in the protocol to achieve more in-depth coverage of the (phospho) proteome. Finally,

we also emphasize the need for critical analysis when investigating datasets of such complexity at the peptide level.

3. Materials and methods

3.1. Chemicals and Materials

Cells were obtained from ATCC. The growth medium, epidermal growth factor (EGF), media, supplements, phosphate buffered saline (PBS) and the PeppermintStick internal protein standard were from Invitrogen (Carlsbad, CA, USA). Readyprep sequential extraction kit, tributylphosphine (TBP), Laemmli sample buffer, 2-Mercaptoethanol (BME), Tris-HCl precast polyacrylamide gels and the Precision Plus Protein Marker were obtained from Biorad (Hercules, CA, USA). Triethylammonium bicarbonate (TEABC), sodium dodecyl sulfate (SDS), N-cyclohexyl-3-aminopropanesulfonic acid (CAPS), calcium chloride (CaCl2) and tween-20 were from Millipore (Billerica, MA, USA). Modified Trypsin was acquired from Promega (Fitchburg, WI, USA) and the iTRAQ reagents from ABSciex (Framingham, MA, USA). Acetonitrile and formic acid were from Biosolve, LCMS grade (Dieuze, France). Alkaline phosphatase was obtained from different manufactures: calf intestinal phosphatase was from New England Biolabs (Ipswich, MA, USA) and Sigma-Aldrich (St. Louis, MO, USA), thermosensitive alkaline phosphatase from Promega and Escherichia coli alkaline phosphatase was from Sigma-Aldrich. All other reagents were purchased from Sigma-Aldrich unless stated otherwise.

3.2. Cell culture and lysis

HeLa cells were cultured at 37°C in Dulbecco's Modified Eagle Medium (5% CO2) supplemented with 1% (w/v) L–glutamine, 10% (w/v) FBS and 50 IU/ml penicillin/streptomycin. Cells were starved in FBS-free medium 17 hours before harvesting. For the EGF stimulation, the medium was removed, enriched with 150 ng/ml EGF and immediately re-added to the cells for 10 minutes. Next, the EGF-stimulated and control cells were incubated 15 minutes at 37°C with PBS-based dissociation buffer (Invitrogen) and detached by cell scraping [8]. Cells were washed twice by pre-cooled PBS and centrifugation at 4°C. All subsequent steps were also performed at 4°C unless declared otherwise. Cell lysis was performed by suspending the cell pellet in the R1-buffer from the Readyprep sequential extraction kit at a density of 1ml/10^6 cells. R1 is comprised of 40 mM Tris-buffer which we supplemented with: 2% (v/v) Phosphatase Inhibitor Cocktail 2&3 (P5726 and P0044, Sigma-Aldrich), 2% (v/v) TBP, 0.3% (v/v) benzonase (E1014, Sigma-Aldrich) and 1 tablet/10ml R1 cOmplete Mini EDTA-free Protease Inhibitor Cocktail (118361700001, Roche, Penzberg, Germany). After sonication in an ice-bath for 10 minutes, the samples were split into two parts: 1) one part of the protein extract was mixed with 25:1 volumes pre-cooled acetone for overnight incubation at -20°C. The next day, after a centrifugation

step at maximum speed, the acetone was removed and the pellet was resuspended in 0.5 M TEABC. After an additional spin, the pellet was removed and the supernatant was withheld as the “R1 extract”. 2) the second part of the protein extract was immediately centrifuged at maximum speed. The pellet was resuspended in the R3-buffer (containing urea, thiourea, detergents and ampholytes) from the Readyprep sequential extraction kit and combined with the initial R1 lysate, together forming the “R1+R3 extract” [24]. The protein content of all the extracts was determined by the 2-D Quant Kit (GE Healthcare, Little Chalfont, United Kingdom).

3.3. Gel Fractionation and Digest

20 µg of the “R1 extract” from both the EGF-stimulated and control cells were digested analogous to the iTRAQ protocol (ABSciex). 500 µg of the “R1+R3 extract” derived from the EGF-stimulated cells was combined with 25% (v/v) 4x Laemmli Sample Buffer and reduced with 10% (v/v) BME for 10 minutes at 95°C. Next, 1.5% (w/w) of the PeppermintStick internal protein standard (IS1) was added and the sample was loaded in the central well of the 10% Criterion Tris-HCl Gel, 11cm IPG+ 1well. Electrophoresis of the gel was performed for 30 minutes at 150 V and 60 minutes at 200 V in a Criterion Cell (Biorad) and monitored by the Precision Plus Protein Marker in the outer well [24]. After a fixation and washing step, the gel was cut into four equal pieces comprising different molecular weight fractions which were transferred to glass tubes. The proteins were in-gel digested according to optimized conditions [24]. All the obtained peptides were brought to dryness by the SpeedVac and stored at -20°C until further use.

3.4. Labeling

Peptides of the “R1 extracts” were re-dissolved in 40 µl 50 mM TEABC, while peptides from the molecular weight fractions of the “R1+R3 extract” were re-dissolved in 80 µl. As a second internal standard (IS2), ±7 pmol of each phosphopeptide from the Heavy MS PhosphoMix 1, 2 and 3 (MSP1H, MSP2H, MSP3H Sigma-Aldrich. Composition: see A.3.3. Table 1B) was spiked into each sample. Subsequently, “R1 extracts” were equally divided in two parts (Figure 1), and the fraction of the “R1+R3 extract” was divided in four parts of 20 µl. Before the iTRAQ labeling, each reporter was reconstituted in 250 µl ethanol and combined with the related label of a second 4-plex set. For the ”R1 extracts”, the two halves of the control sample were labeled with 50 µl of the 114 and 115 reporter, the two halves of the EGF-stimulated sample with the 116 and 117 reporter, and each quarter of the four molecular weight fractions of the “R1+R3 extract” with 50 µl of a different 4-plex label (A.3.3. Figure 1). The labeling was performed for two hours at room temperature by continuous shaking before samples were dried out [18].

3.5. Dephosphorylation

The labeled peptides in the 20 vials (both R1 extracts were split in two, and each gel fraction of the R1+R3 extract was split in four) were re-dissolved in 12 µl calf intestinal phosphatase (CIP) buffer (100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl2, 1 mM DTT) and incubated for 45 minutes at room temperature in order to remove the residual free labels [18]. Next, the 114-116 and 115-117 labeled peptides were combined, resulting in 10 vials (114-116 and 115-117 for the "R1 extracts" as well as the four molecular weight fractions of the "R1+R3 extract"). To the 115-117 samples, 30 µl of the dephosphorylating enzyme cocktail (6,25% 0.25U/µl Escherichia coli alkaline phosphatase (P4069, Sigma-Aldrich), 31.25% 10U/µl calf intestinal phosphatase 1 (M0290, New England Biolabs), 31.25% 10U/µl calf intestinal phosphatase 2 (P4978, Sigma-Aldrich)) and 31.25% 1U/µl thermosensitive alkaline phosphatase (M9910, Promega)) was added [18, 20]. For the 114-116 control samples, 30 µl of CIP buffer was supplemented before all samples were incubated at 37°C for the overnight dephosphorylation. Phosphatases were inactivated the next day by heating the samples for 30 minutes at 85°C in the presence of EDTA (50 mM). The 114-116 and 115-117 units were merged and the five samples were cleaned by C18 tips (Sigma-Aldrich) and an offline cleaning system on a PepMap C18 column (Dionex, Thermo) to desalt and remove free labels and excess of phosphatases. The peptide samples which were eluted with 80% acetonitrile (ACN), 20% water, 0.1% formic acid (FA), were split into three equal parts and were dried out immediately.

3.6. LC-MS/MS Analysis

Peptides were resolved in 0.1% FA and approximately 0.5-1 µg was brought on column each run. The nanoLC-MS/MS analysis of the three parts was performed on different ESI mass spectrometry platforms: the first part on a TripleTof 5600 (ABsciex), the second part on a QExactive (Thermo Fisher) and a third part on a SYNAPT G2-Si (Waters Cooperation). 1) For the analysis on the TripleTof 5600, the chromatography was done on an Eksigent ekspert™ nanoLC 400 System with a 120 minutes gradient going from 5% to 90% ACN, 0.1% FA and a 300 nL/min flow. During acquisition, each survey scan accumulated precursors in the range of 400-1250 *m/z* for 250 ms from which the top 20 were fragmented for MS/MS every cycle at a ratio of 180ms/precursor. 2) The QExactive was coupled to a Dionex Ultimate™ 3000 RS nanoLC System. After trapping on a pre-column, peptides were separated on the Acclaim PepMap100, 75 µm x 50 cm by a 90 minute gradient whereby the amount of ACN 0.1% FA increased from 5 to 70%, also through a constant gradient of 300 nl/min. In a scan range of 380-2000 *m/z*, a Full MS & data dependent MS/MS scan was performed on the top 10 by a normalized collision energy (NCE) 30 stepped + collision energy 10% in all cases. 3) For the measurements on the SYNAPT G2-Si, a NanoACQUITY™ system separated the samples with a Waters BEH C18, 75µm x 150mm analytical column after a trapping step of 8 minutes through changing the amount of ACN, 0.1% FA changed from

1 to 40% for 90 minutes. The range of 400 – 1600 *m/z* was inspected during the 0.1 second survey scan from which the top ten of the precursors were fragmented by ramping the collision energy in the "iTRAQ" mode for a 0.1secs/ MS/MS scan if the threshold of 10k counts/spectrum was reached.

3.7. Data Analysis

Data acquisition, recording and pre-processing was carried out with software of the respective mass spectrometer manufacturers. Equally, the raw data was processed by specialized software packages of the vendors: ProteinPilot™ 4.0.8085 for the TripleTof 5600 of ABSciex, Proteome Discoverer 1.4 for the Thermo Scientific QExactive, and PLGS 3.0.1 for the SYNAPT G2-Si of Waters Cooperation. The same software packages were used to export the created peak lists in the MGF format. The database searches of these MGF files were performed with Mascot 2.4.0. (Matrix Science) against the SwissProt Human database (59.084 sequences, supplemented with the sequences of the internal standards). The peptide mass tolerance was 15 ppm (#13C = 1) and the fragment mass tolerance was 0.1 Da. The enzyme was set to trypsin, allowing for a maximum of two missed cleavages. Initially, the prevalence of the most common modifications was explored by an Error Tolerant database search [25]. The cysteine blocking reagent methylthio and the iTRAQ-reagent on amino-termini and lysines were set as fixed modifications, and as a result of the prospective searches iTRAQ on Y, deamidation (NQ), Gln->pyro-Glu (N-term Q and E) and oxidation (M) were set as variable modifications. Additionally, amino-terminal guanidyl, acetyl plus an unspecified +100.02 Da mass shift were set as variable modifications for the "R1" samples, while propionamide (C) and carbamyl (N-term) were added for the "R1+R3" gel samples. The result files of these searches were exported as Mascot-DAT file formats, and each run was then imported separately in the Rover software for individual processing (https://code.google.com/p/compomics-rover/) [26]. After obtaining Rover statistics on the 114/116, 114/115 and 116/117 ratios, the data were filtered in order to export the log2 ratios and z-scores of unique or "razor" peptides for consecutive data validation. Data validation and frequency plot distribution of Z-scores (at a bin size of 0.1) was done in Excel, whereby the Z-scores of the four molecular weight fractions of the "R1+R3 extracts" were combined. Additional database searches were performed by changing the variable modifications in order to annotate the "heavy" internal standard phosphopeptides (A.3.3. Table 2).

4. Results and discussion

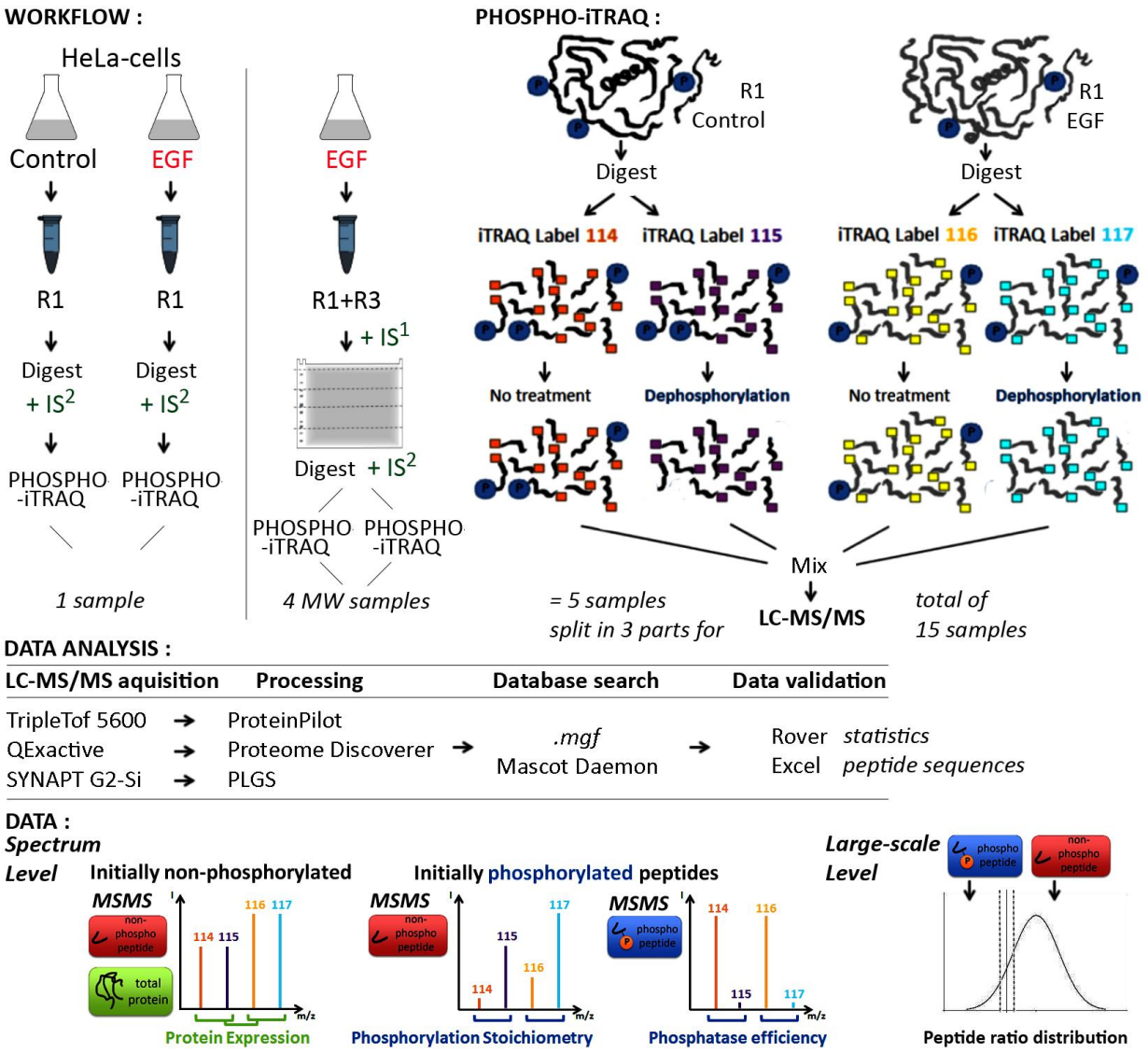


Figure 1. A. Experimental workflow: The soluble protein extracts (R1) of control and EGF-stimulated HeLa cells spiked with an internal standard (IS2) of heavy phosphopeptides were compared by Phospho-iTRAQ. To increase the number of annotated phosphopeptides, the soluble protein extract was complemented with the hydrophobic fraction (R1+R3) of the EGF-stimulated cells and a phosphoprotein internal standard (IS1) and further fractionated on an 1D PAGE into four molecular weight fractions. B. Phospho-iTRAQ protocol: a peptide sample is briefly split in two identical parts and differentially labeled preceding the phosphatase treatment of one part. Afterwards, samples are immediately recombined and split into three parts for the LC-MS/MS analysis on three different instruments. C. Data Analysis: Raw data was processed by the respective vendor's software and MGF-files were searched against the SwissProt Human database using Mascot. Exported DAT-files were imported into Rover for ranking the iTRAQ ratios and further analyzed in Excel. D. Data: Initially phosphorylated peptides have skewed

iTRAQ ratios and arise out of the center of the log-normal distribution of the whole precursor population. The mean of the log-normal ratio distribution is located around zero as the vast majority of the peptides in the data has equal 114/115 or 116/117 reporters.

The large-scale quantitative analysis provides to estimate phospho-stoichiometry which is essential to explore the true extent of phosphorylation as occupancy is associated with functionality [27]. In this report we therefore present for the first time the large-scale exploration of an isobaric tag-based protocol that focuses on the unphosphorylated counterparts of phosphopeptides for its analysis. Details of the optimized protocol are described in the experimental section and a schematic overview of the workflow is given in Figure 1. For the comparative Phospho-iTRAQ analysis, HeLa cells were either treated with EGF for 10 minutes, or left untreated before lysis. The protocol was validated on spiked-in high stoichiometry peptides and was expanded by gel fractionation of the EGF lysate, utilizing the duplication potential of 4-plex iTRAQ.

4.1. Validation of the Phospho-iTRAQ protocol through evaluation of the internal phosphopeptide standards

To validate the effectiveness of the Phospho-iTRAQ method we added a commercially available internal phosphopeptide standard (IS2 in Figure 1) composed of 21 different heavy peptide sequences with at least one phosphorylated amino acid into the sample after digestion. A set of these spiked heavy peptides was identified in all samples (see Supporting information). Not all peptides were annotated however, due to the weak ionization efficiency of some of the sequences, and the technical and inter-run variations characteristic of data dependent acquisition. As observed on raw mass spectra and the processed data of the non-phosphorylated counterparts of the heavy peptides, all of the 114/115 and 116/117 iTRAQ ratios were skewed whereby the phosphatase treated 115 and 117 reporters substantially exceeded the 114 and 116 labels of the control parts, hence characterizing these peptides as phosphopeptides.

In Figure 2, the MS/MS spectrum and the range of the “silent” iTRAQ region of the EVQAEQPSSSSPR peptide (*m/z* 519.2628, 3+) from the R1 sample are given as an example, illustrating that the approach is able to identify phosphopeptides in complex peptide mixtures on all three instruments. Additional proof is provided if the phosphopeptides themselves are annotated (EVQAEQPSSpSSPR *m/z* 818.3734, 2+) as these iTRAQ ratios are inverted compared to their unphosphorylated counterparts (presented in Figure 2, right upper and lower panel). In this manner, the dephosphorylation efficiency can be verified which further allows for the accurate calculation of the

stoichiometry as described elaborately in [12, 19]. If the native phosphorylated peptide is not annotated the method still provides a relative simple way to identify high stoichiometry phosphopeptides on a large scale. As an alternative for the phosphatase treatment, chemical dephosphorylation by means of cerium oxide has also been suggested to ensure the sensitivity of the methodology [22]. While the stoichiometry of the peptides is well defined with this approach, complementary analysis is still required to determine the exact amino acid of the identified phosphopeptide that is modified, especially in case of multi-phosphorylated peptides.

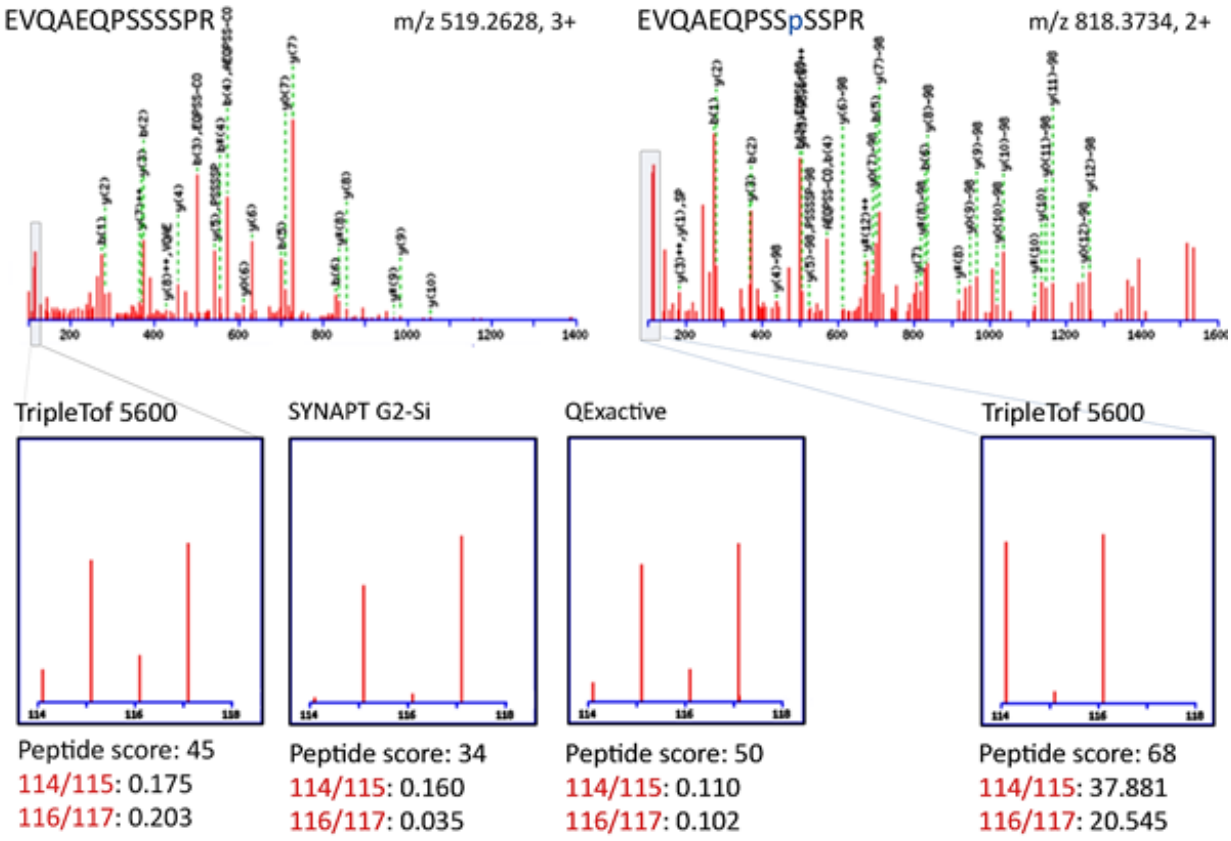


Figure 2. MS/MS spectra *(screenshot of Mascot: low resolution. The text on the picture displays identified ions)* of the EVQAEQPSSSSPR phosphopeptide from the IS2 in the dephosphorylated (left) and phosphorylated (right) form. The range of the iTRAQ reporter region and given ratios of the 114/115 control and 116/117 EGF treated samples show that the Phospho-iTRAQ method is able to identify phosphopeptides in complex peptide mixtures on all three instruments, although variation between the different ratios is observed. The annotated native phosphopeptide has an inverted iTRAQ ratio compared to its unphosphorylated counterpart.

4.2. Addressing the variation between annotated peptides with similar stoichiometry

The overall average log2 Mascot ratio of the confidently annotated IS2 peptides (i.e. Mascot peptide score >25) was -3.56 (SD: 2.31, N:42) for the control and -2.84 (SD: 1.46, N:42) for the EGF stimulated sample. The corresponding log2 ratios of the IS2 which were spiked after the in-gel digestion is on average -3.09 (SD: 1.84, N:138) for 114/115 and -3.12 (SD: 1.56, N:138) for 116/117. Although all spiked IS2 peptides were completely phosphorylated (100% stoichiometry), generating strongly skewed "Phospho-iTRAQ ratios", variation between the ratios of different peptides was observed as reflected by the SD. Equally, as also seen in Figure 2, the ratios of peptides that are annotated multiple times in one run or over multiple LC-MS/MS runs are not identical. Although the 114-116 vials were not subjected to dephosphorylation, 114-116 reporter ions of different intensities were observed, causing the variation between the given ratios.

This ratio dilution is attributed to: 1) Underestimation of the fold change which is a known iTRAQ bias and to be expected if phosphopeptides with high stoichiometry are explored by Phospho-iTRAQ; 2) the "isobaric elephant": co-eluting peptides with similar *m/z* values that are co-selected for fragmentation dilute the ratios [28-31]. This results in both a scattering of the ratios of non-phosphopeptides outside the center of the normal ratio distribution of the whole data set, and more importantly, contributes to the underestimation of the phosphopeptide occupancy. Although the "isobaric elephant" challenges the stoichiometry measurements, emerging technologies such as ion-mobility separation (SYNAPT G2-Si) and multi-notch selection (Orbitrap Fusion) hold great promise to reduce this diluting effect and correspondingly increase the accuracy of isobaric labeling techniques in the near future [32, 33]. Furthermore, simply improving separation of the sample prior to analysis has been shown to dramatically lower such interferences as well [31]. However, because the bulk of the peptides in a dataset intrinsically have 1:1 ratios dilution will predominantly cause extreme ratio's to move to the middle of the distribution. Supplemented by the notion that the fold change is generally underestimated in iTRAQ approaches ratios are not expected to move away from the center of the distribution and the smallest ratio of each multiply annotated sequence is arguably the most accurate one.

4.3. A comparative (phospho) proteome analysis of EGF stimulated and non-stimulated HeLa cells on three different mass spectrometry instruments

The database searches with identical parameters resulted in the annotation of 1283 proteins out of the 31562 spectra from the TripleTof 5600 (50% annotation rate), 1567 proteins out of the 34629 spectra from the QExactive (42%) and 905 proteins out of the 21822 spectra from the SYNAPT G2-Si (32%). For large-scale data mining, we applied Rover, a software specifically developed for the visualization, analysis and validation of quantitative proteomics data. The software does not predict the phospho-modification but only performs normalization of selected ratios from mass spectrometry datasets. This allows the ranked iTRAQ ratios and corresponding Z-scores of the unique and "razor" peptides of each run to be exported from Rover and to set cutoffs in order to select peptides with skewed label ratios (Figure 3).

The small Z-scores of the IS2 peptides confirm our previous conclusions on the effectiveness of the approach to discover high stoichiometry peptides in a large dataset (Figure 3). The Z-scores of the spiked phosphopeptides also mark the threshold for defining phosphorylation in each dataset i.e. all the annotated peptides in the data with a Z-score below this value are pinpointed as highly phosphorylated. Alternatively, predefined mixtures of heavy phosphorylated and their non-phosphorylated heavy counterparts could similarly set the thresholds of different sections over the full range of occupancy states [8, 20]. However, even with such an alternative approach, separating the overlapping Z-score ratio distribution of the phosphorylated and non-phosphorylated peptides remains extremely challenging because: 1) only a limited amount of peptides in the non-enriched data are phosphorylated in the bulk of non-phosphorylated peptides; 2) most of the phosphorylated peptides are of low stoichiometry (over 50% of the peptides have a stoichiometry < 30%) [20]; and 3) the abovementioned bias of the iTRAQ technique further leads to overlap between the normal distribution of the ratios of the non-phosphorylated peptides (1:1 ratios) and the distribution of the ratios from initially phosphorylated peptides (deviating ratios).

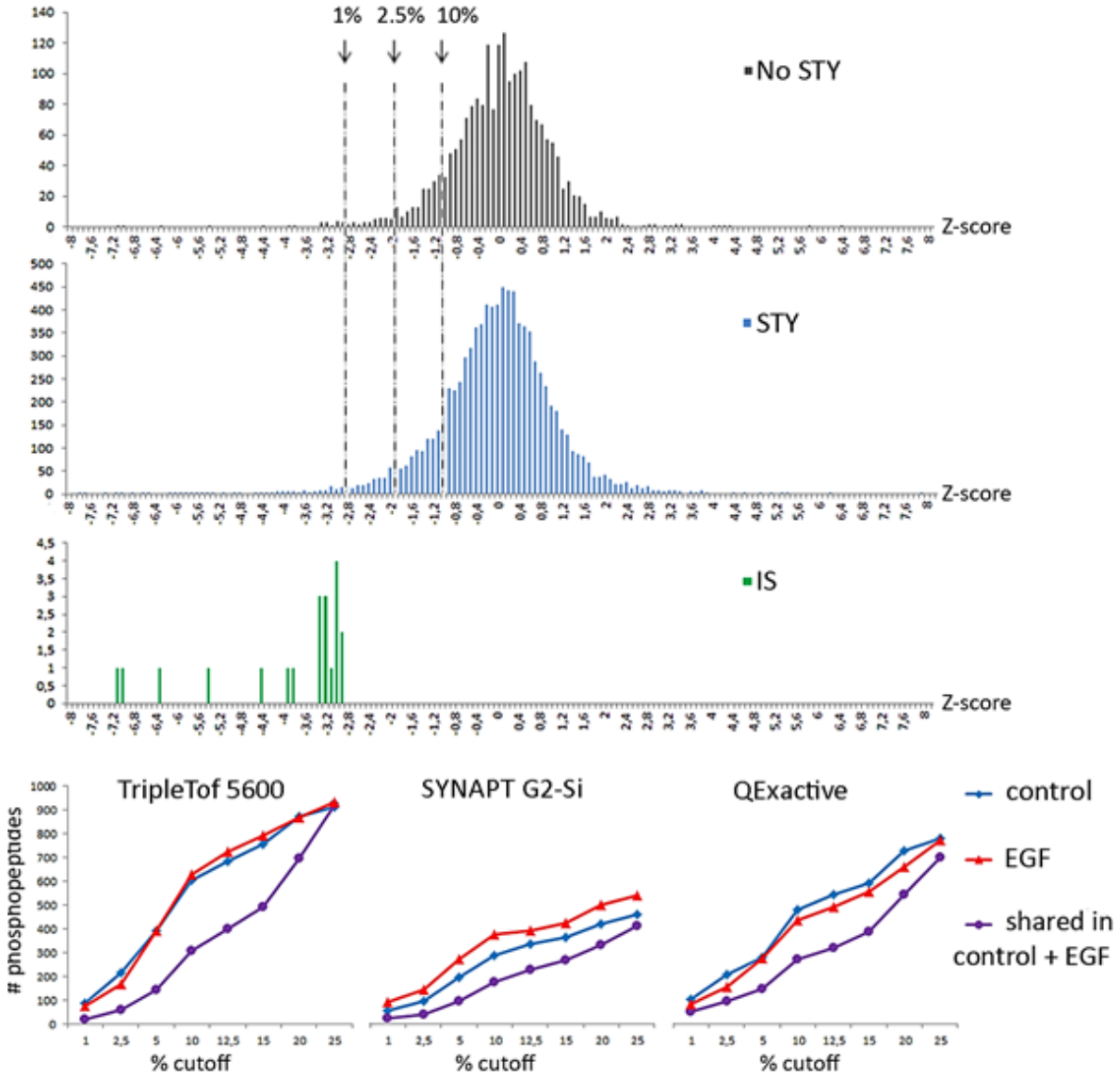


Figure 3. Upper panel: The frequency plot of the Z-scores of all the peptides with no serine, threonine or tyrosine (No STY) in the sequence are used to determine the cutoffs that mark the phosphopeptides. Therefore Z-scores that correspond with the 1% - 2.5% - 5% - 10% - 12.5% - 15% - 20% and 25% subgroup at the left side of the No STY distribution are transposed (only 1, 2.5 and 10% are shown) to the STY dataset. All the STY peptides with a Z-score lower than these exported cutoffs are pinpointed as phosphopeptides. The (low) Z-scores of the identified IS2 peptides in this dataset confirm the effectiveness of Phospho-iTRAQ. Lower panel: Depending on the % cutoff an increasing amount of precursors are classified as phosphopeptides on the different instruments. The multiplexing allows to compare (phospho-) peptides between the control and EGF stimulated sample. On average one-third of the phosphopeptides identified in the lysate of both the control and EGF stimulated cells is shared (purple) between the two conditions.

To assess a threshold for filtering out phosphopeptides, we first extracted the Z-scores of all annotated peptides that cannot be phosphorylated: peptides with no serine, threonine or tyrosine (No STY). On the No STY distribution, Z-scores at different quantile levels (1 to 25% on the left site of the distribution), were chosen as a cutoff value (Figure 3, upper panels). To select the phosphopeptides in the data, these Z-score cutoffs were subsequently transposed to the ranked Z-scores of the peptides that have the ability to be phosphorylated (with an STY in the sequence). Different levels, going from the left percentile to the quartiles of the distribution were applied to set the threshold on the 114/115-control sample and the 116/117-EGF treated sample (see Supporting information for an example of the data analysis workflow and obtained phosphopeptide sequences). The graphs in Figure 3 (lower panel) illustrate the amount of phosphopeptides which are shared or only identified in one of both conditions at each selected threshold. A low cutoff only annotates high stoichiometry peptides while a higher cutoff defines a larger set of phosphopeptides, including those with a lower stoichiometry that could be useful for discovery studies. Interestingly, on average 35% of the identified phosphopeptides were shared between the two independently generated protein lysates from the control and stimulated cells, with the amount of shared peptides increasing towards lower stoichiometry (i.e. higher cutoff).

Depending on the considered cutoff level, an increasing percentage of falsely identified peptides is to be expected. Validating how our Phospho-iTRAQ approach relates to established methods that rely on enrichment for phosphopeptide identification is unachievable, so alternatively, the lists of phosphopeptides were matched with a database of known HeLa phosphosites obtained by different techniques [34]. The results demonstrated that the amount of Phospho-iTRAQ phosphopeptides that are equally present in the database is indeed higher at a lower cutoff. For example, 71 of the 210 phosphopeptides (34%) that were identified in both the control and EGF lysate at the 2.5% cuttof are known, in contrast to 177 of the 818 phosphopeptides (22%) of the selection derived from the 10% cutoff level. Similarly, 130 of the 661 phosphopeptides defined at the 2.5% cutoff level (20%) of the control and 136 of the 608 (22%) of the EGF sample are known, compared to respectively 287 of the 1862 (15%, control) and 284 of the 1898 (15%, EGF) of the selection from the 10% cutoff. Although the Phospho-iTRAQ method does not exactly determine how many phosphopeptides are actually present in the data, the absence of enrichment reduces technical bias in the protocol and provides identification of complementary phosphopeptides.

4.4. The flexibility in experimental set-up provided by Phospho-iTRAQ allows for deeper (phospho) proteome coverage and inclusion of internal replicates.

As samples are only briefly split for the differential dephosphorylation of one half, extensive sample fractionation is possible at both the protein and the peptide level. Additional fractionation prior to the LC-MS/MS acquisition is not only beneficial for more in-depth proteome coverage but is also recommended to minimize the effect of co-eluted and co-fragmented (isobaric-labeled) peptides [29, 31]. To validate this extension of the protocol, the EGF lysate spiked with IS1 was first gel separated in four molecular weight fractions (A.3.3. Figure 2) and after the in-gel digestion split into four to implement an internal replicate analysis. In case of iTRAQ 8-plex or TMT 10-plex, more replicates could be included and since all labels are multiplexed in the workflow, very extensive fractionation at the peptide level is an achievable option. The database searches of the merged MGF-files resulted in the annotation of 2070 proteins out of 105572 spectra from the TripleTof 5600 (37% annotation rate), 2913 proteins out of 155848 spectra from the QExactive (33%) and 1976 proteins out of 82556 spectra from the SYNAPT G2-Si (31%). Figure 4 illustrates the log2 peptide 114/115 ratios of the CASB_BOVIN phosphoprotein of the IS1. Of all 17 annotated peptides, only one peptide of CASB_BOVIN is located outside the overall distribution FQSEEQQQTEDELQDK, a known phosphopeptide [35].

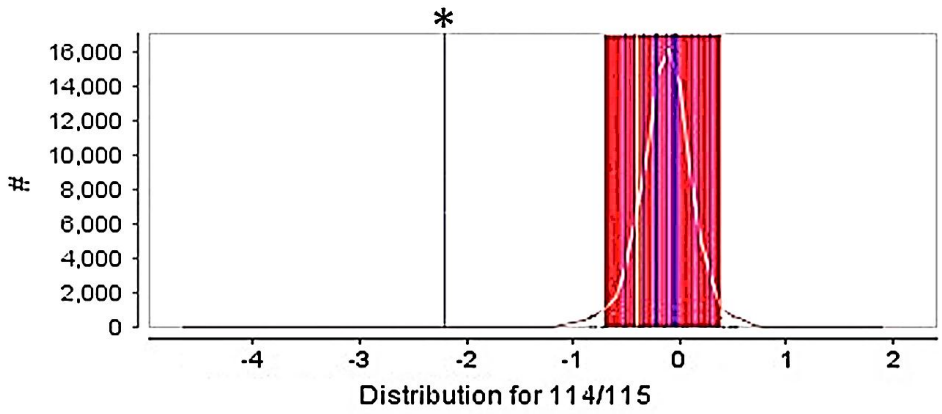


Figure 4. Rover result *(screenshot)* of the CASB_BOVIN protein from IS1 that was spiked in the R1 + R3 protein extract before gel fractionation. The figure visualizes the peptides of CASB_BOVIN (as vertical bars) within the 114/115 distribution of all the iTRAQ ratios. Of all annotated peptides, only one peptide (asterisk) is located outside the overall distribution: FQSEEQQQTEDELQDK, a known phosphopeptide. All other identified peptides of CASB_BOVIN have log2 ratios around 0 and are situated in the center of the distribution.

Analogous to the comparative in-solution Phospho-iTRAQ analysis (cfr. supra), the ranked 114/115 and 116/117 Z-scores of the No STY peptide population allow the setting of the phospho cutoffs at different quantile levels on the STY peptide distribution. Depending on the threshold level, a growing list of phosphopeptide enriched identifications is obtained (Figure 5). Although the overlap between the duplicate analyses might appear quite narrow, this is an intrinsic consequence of setting thresholds. For instance, if for one peptide the 114/115 ratio is 0.3 and the 116/117 ratio is 0.35, one ratio might very well be included at the 1% threshold for 114/115 while the other only appears at the 2.5% threshold for 116/117. For each threshold we compared the percentages of shared peptides to the random chance of being twice in that specific quartile of the distribution. This showed a large enrichment of phosphopeptides at the left side of the distribution; for example, the 10% cutoff yields at least twenty percent of the peptides as shared between replicates as opposed to an expected 0.25% (see graph in Figure 5). Importantly, technical replicates were performed where samples were run completely independently on three different instruments. In the total dataset, a large amount of defined phosphopeptides was shared between these different acquisitions: 1228 of the STY peptides from the left 10% percent of the distribution are present in at least two of the three runs (see Venn diagram in Figure 5). This considerable overlap increases the effectiveness of the Phospho-iTRAQ selection, which could be verified by additional biological replicates. This showed a large enrichment of phosphopeptides at the left side of the distribution; for example, the 10% cutoff yields at least twenty percent of the peptides as shared between replicates as opposed to an expected 0.25% (see graph in Figure 5). Importantly, technical replicates were performed where samples were run completely independently on three different instruments. In the total dataset, a large amount of defined phosphopeptides was shared between these different acquisitions: 1228 of the STY peptides from the left 10% percent of the distribution are present in at least two of the three runs (see Venn diagram in Figure 5). This considerable overlap increases the validity of the Phospho-iTRAQ selection, which could be verified by additional biological replicates.

% cutoff	1	2.5	5	10	12.5	15	20	25
TripleTof 5600	487	1045	1942	3157	3743	4254	5119	5837
SYNAPT G2-Si	265	744	1275	2222	2539	2853	3448	3948
QExactive	553	1102	1745	2877	3307	3686	4417	5142

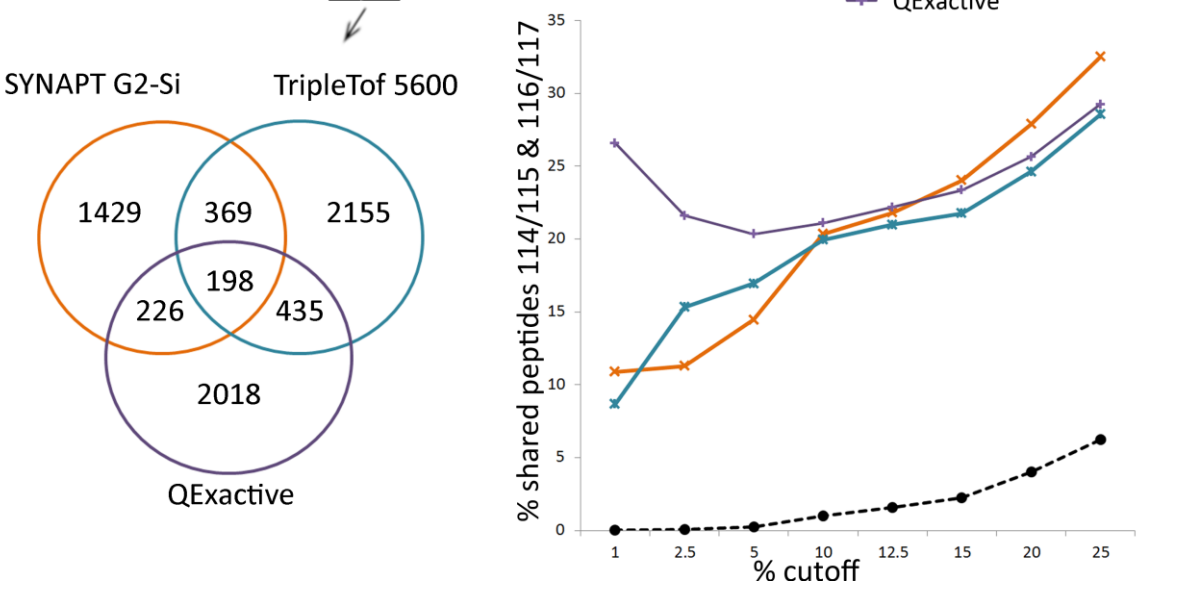


Figure 5. Depending on the threshold level, a growing list of phosphopeptide enriched identifications is acquired from the gel fractionated R1+R3 extract. The total amount of defined (Phospho) peptides from both the 114/115 and the 116/117 distribution are given in the table. The Venn diagram displays the number of phosphopeptides from the left 10% percent of the distribution that are shared between the acquisitions on the different instruments. The right graph demonstrates the percentages of identified phosphopeptides that are shared between the 114/115 and 116/117 replicates at different cutoff levels.

5. Conclusion

We here present and evaluate an isobaric tag based protocol that relies on phosphatase treatment of differentially labeled peptides for large-scale quantitative analysis of phosphopeptides. The Phospho-iTRAQ method focuses only on the non-phosphorylated counterparts of the phosphopeptides, avoiding specialized and relatively imprecise workflows such as phospho-enrichment. Since multiple labels are used, information on relative protein expression can still be obtained in our method.

The application of Phospho-iTRAQ on spiked-in phosphoproteins and phosphopeptides demonstrated its ability to discover high stoichiometry peptides in a large dataset, and this on multiple LC-MS/MS platforms. The iTRAQ data at the spectrum level allows estimation of the phosphorylation stoichiometry, while the ratio distribution of all annotated peptides allows peptides with high stoichiometry to be identified by their deviating Z-scores. The dilution effect of the co-eluted and co-selected peptides still challenges the accuracy of the methodology but can be partially addressed by the multiplexing potential of the isobaric labels which allows extensive fractionation and the inclusion of replicates. Furthermore, newly arising technologies such as ion-mobility separation hold further promise to diminish this interference and thus increase the reproducibility of the method.

Additional confirmation of the identified phosphopeptides with complementary techniques however remains necessary, because our approach is not able to pinpoint the exact phosphorylation site. As with most quantitative analyses, it is up to the researcher to decide where to put the threshold in the ranked ratio distribution of large-scale data that marks the list of the phosphopeptide enriched identifications. Indeed, depending on the goal of the experiment one can focus exclusively on the more confidently annotated high stoichiometry peptides (low cutoff), or one can rather isolate a larger but overall less reliable list of peptides for an exploratory study (high cutoff), thus increasing the need for additional biological replicates to define the common subsets of different runs that will again augment the confidence of the phosphopeptide detection.

ASSOCIATED CONTENT

Supporting Information

A scheme of the labeling of the R1+R3 extract, a boxplot for the validation of the molecular weight separation, the composition of the internal standards and the details of the database search parameters are given in addendum. Excel files are available online: 1) All the annotated peptides of the IS2. 2) An example of the data analysis workflow. 3) List of the phosphopeptides of all samples, defined according to the Phospho-iTRAQ principle.

SUPPLEMENTARY FILES IN ADDENDUM AND ONLINE

https://drive.google.com/folderview?id=0B8-Op8oMG-rCaUljcGpwVFZ3ZEE&usp=sharing

AUTHOR INFORMATION

Corresponding Author

E-mail: dieter.deforce@ugent.be

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Sofie Vande Casteele for her assistance with the maintenance of our Waters Premier LC-MS, the platform applied for the development and optimization of the protocol. We thank Niklaas Colaert for his support with the Rover software and Kristof De Beuf from the FIRE Statistical Consulting of the Ghent University for his help with the statistics. For the LC-MS acquisitions of the final samples we acknowledge Ernst Bouvin and Christian Baumann from ABSciex, Hans Vissers, Chris Hughes and Jan Claereboudt from the Waters Corporation, Wilfried Voorhorst and Heike Schaefer from Thermo Fisher Scientific. This research is funded by a grant from the Fund of Scientific Research Flanders (FWO) and a PhD grant from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen), awarded to P.M.. D.D. and L.M. acknowledge the support of Ghent University (Multidisciplinary Research Partnership “Bioinformatics: from nucleotides to networks”). L.M. also acknowledges support from the PRIME-XS project funded by the European Union seventh Framework Program under grant agreement number 262067.

References

[1] Jünger, M. A. A., R., Mass spectrometry-driven phosphoproteomics: patterning the systems biology mosaic. Dev Biol 2014, 3, 83-112.
[2] Thingholm, T. E., Jensen, O. N., Larsen, M. R., Analytical strategies for phosphoproteomics. Proteomics 2009, 9, 1451-1468.
[3] Tichy, A., Salovska, B., Rehulka, P., Klimentova, J., et al., Phosphoproteomics: searching for a needle in a haystack. Journal of proteomics 2011, 74, 2786-2797.
[4] Roux, P. P., Thibault, P., The coming of age of phosphoproteomics--from large data sets to inference of protein functions. Molecular & cellular proteomics : MCP 2013, 12, 3453-3464.
[5] Dephoure, N., Gould, K. L., Gygi, S. P., Kellogg, D. R., Mapping and analysis of phosphorylation sites: a quick guide for cell biologists. Mol Biol Cell 2013, 24, 535-542.
[6] Nilsson, C. L., Advances in quantitative phosphoproteomics. Analytical chemistry 2012, 84, 735-746.
[7] Linke, D., Hung, C. W., Cassidy, L., Tholey, A., Optimized fragmentation conditions for iTRAQ-labeled phosphopeptides. J Proteome Res 2013, 12, 2755-2763.
[8] Olsen, J. V., Vermeulen, M., Santamaria, A., Kumar, C., et al., Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Science signaling 2010, 3, ra3.
[9] Harvey, S. L., Enciso, G., Dephoure, N., Gygi, S. P., et al., A phosphatase threshold sets the level of Cdk1 activity in early mitosis in budding yeast. Mol Biol Cell 2011, 22, 3595-3608.
[10] Vaudel, M., Sickmann, A., Martens, L., Peptide and protein quantification: a map of the minefield. Proteomics 2010, 10, 650-670.
[11] Zhang, X., Jin, Q. K., Carr, S. A., Annan, R. S., N-Terminal peptide labeling strategy for incorporation of isotopic tags: a method for the determination of site-specific absolute phosphorylation stoichiometry. Rapid Commun Mass Spectrom 2002, 16, 2325-2332.
[12] Bonenfant, D., Schmelzle, T., Jacinto, E., Crespo, J. L., et al., Quantitation of changes in protein phosphorylation: a simple method based on stable isotope labeling and mass spectrometry. Proc Natl Acad Sci U S A 2003, 100, 880-885.
[13] Domanski, D., Murphy, L. C., Borchers, C. H., Assay development for the determination of phosphorylation stoichiometry using multiple reaction monitoring methods with and without phosphatase treatment: application to breast cancer signaling pathways. Analytical chemistry 2010, 82, 5610-5620.
[14] Hegeman, A. D., Harms, A. C., Sussman, M. R., Bunner, A. E., Harper, J. F., An isotope labeling strategy for quantifying the degree of phosphorylation at multiple sites in proteins. Journal of the American Society for Mass Spectrometry 2004, 15, 647-653.
[15] Johnson, H., Eyers, C. E., Eyers, P. A., Beynon, R. J., Gaskell, S. J., Rigorous determination of the stoichiometry of protein phosphorylation using mass spectrometry. Journal of the American Society for Mass Spectrometry 2009, 20, 2211-2220.
[16] Kanshin, E., Wang, S., Ashmarina, L., Fedjaev, M., et al., The stoichiometry of protein phosphorylation in adipocyte lipid droplets: analysis by N-terminal isotope tagging and enzymatic dephosphorylation. Proteomics 2009, 9, 5067-5077.
[17] Previs, M. J., VanBuren, P., Begin, K. J., Vigoreaux, J. O., et al., Quantification of protein phosphorylation by liquid chromatography-mass spectrometry. Analytical chemistry 2008, 80, 5864-5872.
[18] Pflieger, D., Junger, M. A., Muller, M., Rinner, O., et al., Quantitative proteomic analysis of protein complexes: concurrent identification of interactors and their state of phosphorylation. Molecular & cellular proteomics : MCP 2008, 7, 326-346.
[19] Przybylski, C., Junger, M. A., Aubertin, J., Radvanyi, F., et al., Quantitative analysis of protein complex constituents and their phosphorylation states on a LTQ-Orbitrap instrument. J Proteome Res 2010, 9, 5118-5132.
[20] Wu, R., Haas, W., Dephoure, N., Huttlin, E. L., et al., A large-scale method to measure

absolute protein phosphorylation stoichiometries. Nature methods 2011, 8, 677-683.
[21] Hebert, A. S., Merrill, A. E., Stefely, J. A., Bailey, D. J., et al., Amine-reactive neutron-encoded labels for highly plexed proteomic quantitation. Molecular & cellular proteomics : MCP 2013, 12, 3360-3369.
[22] Jia, W., Andaya, A., Leary, J. A., Novel mass spectrometric method for phosphorylation quantification using cerium oxide nanoparticles and tandem mass tags. Analytical chemistry 2012, 84, 2466-2473.
[23] Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., et al., Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 2006, 127, 635-648.
[24] Glibert, P., Van Steendam, K., Dhaenens, M., Deforce, D., iTRAQ as a method for optimization: Enhancing peptide recovery after gel fractionation. Proteomics 2014, 14, 680-684.
[25] Creasy, D. M., Cottrell, J. S., Error tolerant searching of uninterpreted tandem mass spectrometry data. Proteomics 2002, 2, 1426-1434.
[26] Colaert, N., Helsens, K., Impens, F., Vandekerckhove, J., Gevaert, K., Rover: a tool to visualize and validate quantitative proteomics data from different sources. Proteomics 2010, 10, 1226-1229.
[27] Olsen, J. V., Mann, M., Status of large-scale analysis of post-translational modifications by mass spectrometry. Molecular & cellular proteomics : MCP 2013, 12, 3444-3452.
[28] Ow, S. Y., Salim, M., Noirel, J., Evans, C., et al., iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". J Proteome Res 2009, 8, 5347-5355.
[29] Evans, C., Noirel, J., Ow, S. Y., Salim, M., et al., An insight into iTRAQ: where do we stand now? Anal Bioanal Chem 2012, 404, 1011-1027.
[30] Christoforou, A., Lilley, K. S., Taming the isobaric tagging elephant in the room in quantitative proteomics. Nature methods 2011, 8, 911-913.
[31] Vaudel, M., Burkhart, J. M., Radau, S., Zahedi, R. P., et al., Integral quantification accuracy estimation for reporter ion-based quantitative proteomics (iQuARI). J Proteome Res 2012, 11, 5072-5080.
[32] Ting, L., Rad, R., Gygi, S. P., Haas, W., MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. Nature methods 2011, 8, 937-940.
[33] Sturm, R. M., Lietz, C. B., Li, L., Improved isobaric tandem mass tag quantification by ion mobility mass spectrometry. Rapid Commun Mass Spectrom 2014, 28, 1051-1060.
[34] Gnad, F., Gunawardena, J., Mann, M., PHOSIDA 2011: the posttranslational modification database. Nucleic Acids Res 2011, 39, D253-260.
[35] Magrane, M., Consortium, U., UniProt Knowledgebase: a hub of integrated protein data. Database : the journal of biological databases and curation 2011, 2011, bar009.

PART 4: iTRAQ AS A METHOD FOR OPTIMIZATION- ENHANCING PEPTIDE RECOVERY AFTER GEL FRACTIONATION

Based on:

iTRAQ as a method for optimization: enhancing peptide recovery after gel fractionation

Pieter Glibert [1], Katleen Van Steendam [1],

Maarten Dhaenens [1,†] and Dieter Deforce [1,†,*]

[1] Laboratory of Pharmaceutical Biotechnology, Ghent University, 460 Ottergemsesteenweg, B-9000 Ghent, Belgium; E-Mails: (M.D.); pieter.glibert@ugent.be (P.G.); katleen.vansteendam@ugent.be (K.V.S.); maarten.dhaenens@ugent.be (M.D.)

[†] These authors contributed equally to this work.

[*] Author to whom correspondence should be addressed; E-Mail: dieter.deforce@ugent.be; Tel.: +32-9-264-8067; Fax: +32-9-220-6688.

Accepted for publication:

Proteomics. 2014 Mar;14(6):680-4. doi: 10.1002/pmic.201300444.

1. Abstract

At the dawn of a new era in label-free quantification on high-resolution MS instruments, classical methods such as iTRAQ continue to provide very useful insights in comparative proteomics. The potential to multiplex samples makes this reporter-based labeling technique highly suited for method optimization as demonstrated here by a set of standard series. Instead of studying ratios of annotated proteins, we propose an alternative method, based on the analysis of the average reporter ratios of all the spectra from a sample or a large distinct subset herein. This strategy circumvents the bias, associated with the annotation and iTRAQ quantification, leading to increased adequacy in measuring yield differences between workflows. As gel electrophoresis prior to MS analysis is highly beneficial, for example, as a fractionation step, the approach was applied to evaluate the influence of several parameters of the established in-gel digestion protocol. We quantified the negative effect of SYPRO Ruby staining and the positive effect of gel fixation prior to digestion on peptide yield. Finally, we emphasize the benefits of adding $CaCl_2$ and ACN to a tryptic in-gel digest, resulting in an up to tenfold enhanced peptide recovery and fewer trypsin missed cleavages

2. Introduction

Over the years, isobaric labeling techniques have been applied on a large scale in different proteome studies where relative quantification is required. Although label-free methods, such as SWATH (AB SCIEX) and HDMS^E (Waters Corporation) are gaining popularity, label-based strategies remain important. The reporter-based labeling methods still have the ability to give complementary insights, especially in terms of minimizing technical variation by parallel quantification of multiple samples [1, 2].

One of the main challenges in interpreting iTRAQ data is the underestimation of the fold change, partially caused by interfering masses in the silent region of the reporters and mixed MS/MS [3, 4]. Here, we present a new workflow for method optimization by means of isobaric tags (such as iTRAQ) to overcome these challenges without the need of complex data analysis tools. In-depth protocol knowledge can be achieved by dividing a (standard) peptide mixture into equal parts and differentially labeling them for each experimental condition under investigation. In the event of a digestion optimization, a protein sample can be split into equal parts, which are labeled and pooled after digestion [5]. The relative yield of each condition is then defined by all the associated reporter ions in a run instead of only focusing on the identified proteins. Additionally, the effect of different experimental conditions on a specific peptide set with similar physicochemical classes can be evaluated after annotation.

3. Results and Discussion

3.1. Validating iTRAQ for optimization

We first analyzed a standard series of known ratios to test the preciseness of this approach and more specifically validate the influence of contaminating peaks in the reporter region on quantification accuracy on our lower resolution ESI-Q-TOF MS platform. A well-defined protein mixture digest (Dionex no. 161088) was equally split into four, differentially labeled according to the manufacturer's protocol and mixed in different ratios (Table 1). For LC–MS/MS analysis, peptides dissolved in 0.1% formic acid were analyzed on a Dionex U3000, coupled to a Q-TOF Premier (Waters Corporation). To create high-quality spectra and abundant reporter peaks, an enhanced TIC threshold of the peptide precursors was set for MS/MS selection during the data-dependent acquisition.

Table 1. Identical parts of a peptide mixture were labeled with a different reporter label of the iTRAQ 4plex Kit (AB SCIEX, Framingham, MA, USA). Each sample (S1–S9) consists of a different combination of the four labels resulting in known (theoretical) ratios. After MS analysis, the average 114/115, 116/117, 114/117, and 116/115 log ratios and the associated SD are calculated from the spectra with reporter areas above the 0.3 threshold for each sample. For the linear regression analysis, both the slope and r2 values are presented for each ratio.

Sample	Label distribution				Theoretical ratio	Experimental ratio							
	114	115	116	117	log 114/115 116/117 114/117 116/115	log 114/115	SD	log 116/117	SD	log 114/117	SD	log 116/115	SD
S1	1	9	1	9	−0.954	−0.930	0.065	−0.855	0.115	−0.867	0.377	−0.867	0.106
S2	2	8	2	8	−0.602	−0.616	0.076	−0.532	0.087	−0.585	0.076	−0.562	0.087
S3	3	7	3	7	−0.368	−0.382	0.067	−0.319	0.079	−0.363	0.063	−0.337	0.085
S4	4	6	4	6	−0.176	−0.192	0.071	−0.114	0.071	−0.173	0.058	−0.132	0.086
S5	5	5	5	5	0.000	−0.025	0.079	0.067	0.070	0.011	0.060	0.031	0.091
S6	6	4	6	4	0.176	0.158	0.096	0.253	0.067	0.188	0.059	0.224	0.110
S7	7	3	7	3	0.368	0.333	0.077	0.449	0.064	0.387	0.063	0.395	0.084
S8	8	2	8	2	0.602	0.576	0.102	0.656	0.077	0.600	0.083	0.631	0.092
S9	9	1	9	1	0.954	0.888	0.126	1.096	0.093	1.045	0.080	0.935	0.138
Linear regression													
r^2						0.9997		0.9981		0.9965		0.9993	
Slope						0.9643		1.0160		0.9995		0.9621	

Using Mascot Distiller (Matrix Science), the raw data were processed into a MGF peak list according to parameters that were optimized to increase the quality of the iTRAQ peaks (A.3.4. and Supporting Information Methods 1 for experimental details). Subsequently, the areas from the 114.1, 115.1, 116.1, and 117.1 iTRAQ reporter peaks were extracted from each MS/MS spectrum by a freely available CompOmics script [6]. To compensate for the isotope carry-over, the correction factors of the manufacturer's certificate of analysis were applied. MS/MS spectra containing reporter masses with an area value below 0.3 were not taken into account for quantification, as these are more susceptible to variation [3, 4]. For data analysis, each data point represents the average of all the log ratios in one run since iTRAQ in method development requires comparison of all the spectra in a sample instead of defining quantitative changes of individual proteins. Apart from yielding accurate quantitative data on the optimization, iTRAQ coordinately circumvents the variation in precursor mass selection found in repeated runs of conventional analyses: an absent reporter indicates an absent peptide, a conclusion that cannot be drawn when two sequential runs are compared.

When plotting the theoretical against the experimental ratios of all the MS/MS spectra over a 1–10 range, a strong correlation was demonstrated. The high r^2 values (e.g. 0.9997 for 114/115) displaying the goodness of fit of the linear regression, indicate that the given iTRAQ ratios predict the actual relationship between two labeled samples on our system. Low SD displays the high precision of this approach, and the slopes of the regression lines around 1 (results in Table 1 and A.3.4. Table 1) demonstrate the capability to quantify small differences between two ratios, leading to accurate quantification. Most importantly, peptide identification is not required, as reporter intensities can be extracted separately. To target a specific set of proteins or peptides with, for example, similar physicochemical properties, protein annotation is possible. Note that the iTRAQ multiplex provides the opportunity to have duplicates in each run wherein one can coordinately discriminate the labels that are most impacted by near-isobaric masses and mixed MS/MS [2]. This is, for example, reflected in the slight difference in the slope and the 95% confidence interval between the 114/115 and 116/117 regression line as shown in Fig. 1, probably caused by label-specific effects [7]. However, all the labels are suited for optimization studies as the results of the regression analysis of all the calculated ratios indicated.

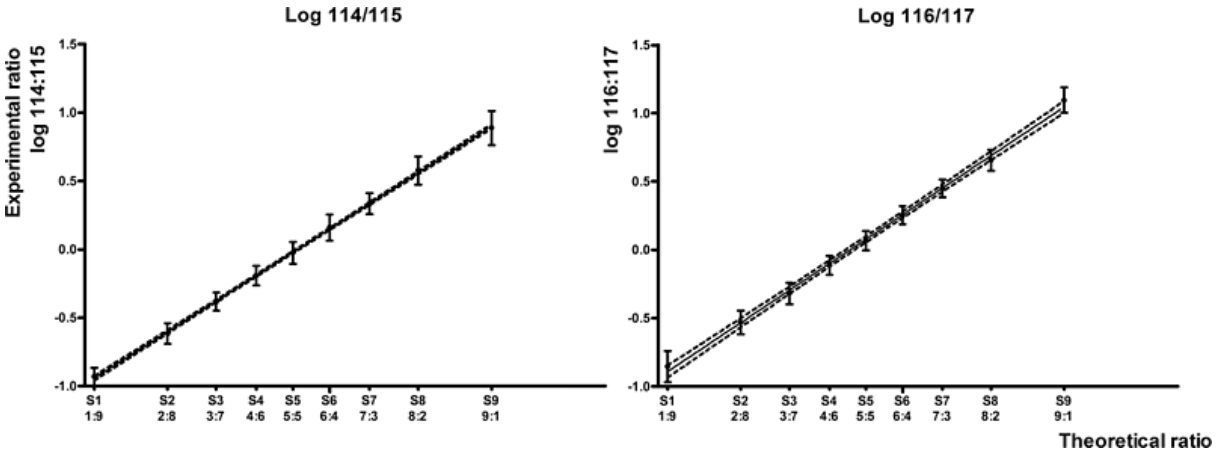


Figure 1. The plotted relationship of the theoretical and experimental ratios (data in Table 1) of two of the four ratios. Each data point represents the average of the log ratios of all MS/MS spectra. Error bars indicate the SD. The linear regression of the means results in an r2 value of 0.9997 (114/115) and 0.9981 (116/117). The dashed lines define the 95% confidence band of the regression line.

3.2. Optimization of the in-gel digestion protocol

Next, we used this iTRAQ approach to optimize our in-gel digestion protocol, the technique that bridges the gap between two keystone methodologies in proteomics: gel electrophoresis and mass spectrometry. Despite the gain in popularity of the gel-free methods, particularly due to highly sensitive

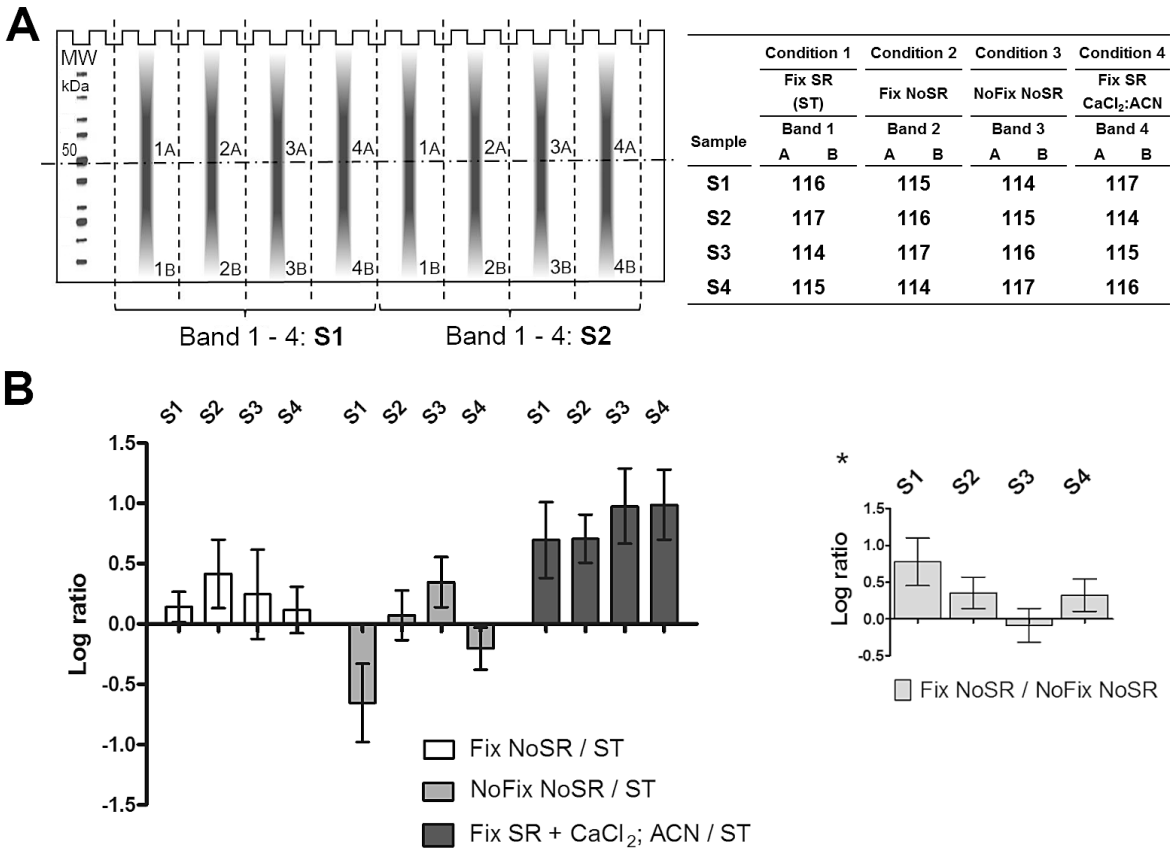
mass spectrometers, gel electrophoresis still has an added value as a molecular weight fractionator, a purification step and in the study of more hydrophobic (membrane) proteins [8]. One very interesting application lies in PTM research where isobaric labels are used to measure stoichiometry: iTRAQ after gel fractionation is capable of reducing interfering precursors and thus enhancing quantification accuracy [7, 9]. Most protocols separate the proteins by SDS-PAGE and visualize the fixed proteins by CBB, Silver stain or popular SYPRO Ruby (SR) stain [10]. After imaging, proteins in the gel bands are classically destained, reduced and alkylated before digestion [11]. Numerous in-gel digestion strategies and optimized protocols are available: some focus on the reduction of the digestion time, while others implement alternative reagents and multiple proteases to augment peptide recovery [12]. However, the consequence of several well-established steps in many protocols on the robustness of the technique is not well documented. Yet, a large (inter-run) variation in the yield of peptide extraction is a well-known downside of many in-gel digestion protocols and challenges optimization [11].

Here, we examine the impact of protein fixation and SR staining and the possible benefits of ACN and $CaCl_2$ on digest efficiency when gel electrophoresis is implemented as a sample preparation step prior to MS analysis. A HepG2 cell lysate was equally divided over sixteen wells of two gels for the electrophoresis. For each condition the extracted peptides were iTRAQ labeled and pooled immediately after the digest (Figure 2A). Every experiment was carried out four times in parallel to allow to swap iTRAQ labels to compensate for possible variation in label efficiency or label accuracy due to label specific contaminating peaks.

The gels were cut around the 50kDa marker to create two fractions and to test the possible impact of the protein molecular weight on the digest conditions. Next, the individual lanes were excised to be digested under the different conditions. For the standard condition 1, proteins were fixed within the gel with a 7% acetic acid, 10% MeOH solution twice for ten minutes. After a short wash with Milli-Q, the gel bands were incubated overnight with the SR gel stain in the dark at room temperature. For alternative conditions 2 and 3, the staining (2) or fixation and staining step (3) were skipped. The next day, the gel bands were washed, reduced, alkylated and dehydrated before modified trypsin (Promega) was added for the overnight digestion at 37°C. The digest was performed with reagents analogous to the standard iTRAQ protocol from AB Sciex (A.3.4. Experimental details). Alternatively, for condition 4, 1mM $CaCl_2$ and 5%ACN was added to the trypsin buffer on a fixed and stained gel (Figure 2A, right). After digestion, peptides were extracted with ACN in three steps, labeled and pooled according to Figure 2A. Data analysis was performed as described above.

With no obvious differences in the calculated average ratios between high and low molecular weight fractions, these data files were merged for further analysis. In Figure 2B, the conditions 2, 3 and

4 are compared against condition 1, the standard protocol where proteins are extracted from a fixed and SR stained gel. iTRAQ quantified large variations in peptide recovery between different gel digests. This coordinately emphasizes the importance of replicate analysis, especially during technical optimization. However, the main advantage of applying this strategy for method development is that the entire peptide ion yield is taken into account. Unlike most studies, we do not rely on the amount of identified proteins directly affected by unexpected modifications and inherent to different steps under investigation such as gel staining and fixation. SR staining clearly has a negative influence on peptide recovery in each replicate as shown by the increased "fixation no SR/standard" ratios where no staining was applied. Peptide loss is a known downside of fluorescent methods, yet here we show that this is not due to unexpected modifications, but rather to a loss of on average ± 40% of the extracted peptides. When implementing gel electrophoresis as a fractionation or purification step, staining should thus be avoided. Surprisingly however, fixation of a gel with acetic acid and MeOH hinted towards a positive effect on peptide recovery as suggested by three of the four significantly positive "fixation no SR/no fixation no SR" ratios (one-tailed sample t-test, $p<0.0001$) (Figure 2B, asterisk). As mentioned earlier, when looking at annotated peptides one can select certain populations of peptides within the ratio distribution (e.g. left and right of the mean) (A3.4.4. Figure 1 and Supporting Information Methods 2A for details) [6]. When these subsets of labeled peptides are compared based on certain properties such as the number of missed cleavages and GRAVY scores, one can coordinately define the gain of a certain methodology for specific characteristic sequences. Since fixation yielded a higher amount of peptides, we hypothesize that skipping fixation possibly results in spontaneous migration of proteins and peptides out of the gel. On the other hand, gel fixation resulted in an increased number of missed cleavages, which suggests that fixation could restrict the number of accessible clipping sites. Finally, despite the use of a modified trypsin, addition of $CaCl_2$ and ACN at the time of digestion was found to be very beneficial to the yield of the digestion resulting in an up to 10-fold increase in peptide recovery. To our knowledge, this is the first time that the advantages of those additives are quantified specifically for in-gel digestion. The advantage of including Ca2+ was previously explained by the beneficial effect on trypsin auto-cleavage during in-solution digestion protocols but is often overlooked since modified trypsin which is less susceptible to auto-cleavage, became a golden standard [11]. This significant difference is probably induced by the stabilizing effect of Ca2+-ions on trypsin and the improvement of the protein accessibility in the presence of ACN [13]. $CaCl_2$ and ACN supplementation does not result in recovery of more hydrophobic peptides as seen by the similar GRAVY-scores but does diminish the amount of trypsin missed cleavages from 26 to 17% (A3.4.4. Figure 2 or Supporting Information Methods 2B).



Sample	Condition 1 Fix SR (ST) Band 1 A	B	Condition 2 Fix NoSR Band 2 A	B	Condition 3 NoFix NoSR Band 3 A	B	Condition 4 Fix SR $CaCl_2$:ACN Band 4 A	B
S1	116		115		114		117	
S2	117		116		115		114	
S3	114		117		116		115	
S4	115		114		117		116	

Figure 2. Figure 2. (A) Left panel: Equal amounts of a HepG2 cell lysate are divided over 16 lanes of two gels, whereby a loaded lane is alternated with an empty one to allow easy cutting of the gel (schematic of one gel is shown). After electrophoresis, the gels are cut around the 50 kDa marker to create a high (A) and low (B) molecular weight fraction. The different lanes were excised and in-gel digestion was performed on the different gel bands according to different conditions and pooled for each replicate after labeling. Right panel: Condition 1—digestion of a fixed and SR-stained gel, marked as the standard (ST) protocol. Condition 2—digestion of a fixed, nonstained gel. Condition 3—digestion of a nonfixed, nonstained gel. Condition 4—ST supplemented with 1 mM $CaCl_2$ and 5% ACN. Extracted peptides from bands 1 to 4 were labeled and pooled according to the presented schedule, together forming four high and four low molecular weight samples. S1–S4: different replicates. (B) Each bar represents the average and SD of all the reporters of one of the four replicas. The different conditions are compared to the standard situation where the gel is fixed and SR stained. Despite a large variation, ratios indicate that SR has a negative effect on peptide recovery and fixation a positive influence. Addition of $CaCl_2$ and ACN during trypsin digestion increases the peptide yield over sevenfold in average. Asterisk: The positive effect of gel fixation is verified by three of four replicates from the “fixation no SR/no fixation no SR” ratios.

4. Conclusion

Isobaric tags are mostly applied to study the proteome expression between different biological samples. The label based bias or variation induced by the required annotation of the proteins is often overlooked or corrected for by using complex post-acquisition tools. Our standard series emphasize the accuracy of the iTRAQ technique in quantifying small loading differences between samples when analyzing the average ratios of all the data in a run. iTRAQ allows for multiplexing and internal replicates which makes the technique highly suitable for method optimization. Using all the data excludes the variation associated with protein identification and quantification, caused by unexpected modifications or fragmentation. For gel electrophoreses, particularly when used as a fractionation and purification tool for subsequent MS analysis, we recommend fixation of the gel and skipping SR staining before digestion. Finally, we corroborate the positive influence of $CaCl_2$ and ACN addition in tryptic in-gel digestion protocols to increase reproducibility, above all in automated workflows [7].

SUPPLEMENTARY FILES AVAILABLE IN ADDENDUM AND ONLINE

http://onlinelibrary.wiley.com/doi/10.1002/pmic.201300444/full

References

[1] DeSouza, L. V., Siu, K. W. M., Mass spectrometry-based quantification. *Clinical Biochemistry* 2013, *46*, 421-431.
[2] Evans, C., Noirel, J., Ow, S. Y., Salim, M.*, et al.*, An insight into iTRAQ: where do we stand now? *Anal Bioanal Chem* 2012, *404*, 1011-1027.
[3] Hultin-Rosenberg, L., Forshed, J., Branca, R. M., Lehtio, J., Johansson, H. J., Defining, comparing, and improving iTRAQ quantification in mass spectrometry proteomics data. *Molecular & cellular proteomics : MCP* 2013, *12*, 2021-2031.
[4] Ow, S. Y., Salim, M., Noirel, J., Evans, C.*, et al.*, iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". *J Proteome Res* 2009, *8*, 5347-5355.
[5] Burkhart, J. M., Vaudel, M., Zahedi, R. P., Martens, L., Sickmann, A., iTRAQ protein quantification: A quality-controlled workflow. *Proteomics* 2011, *11*, 1125-1134.
[6] Barsnes, H., Vaudel, M., Colaert, N., Helsens, K.*, et al.*, compomics-utilities: an open-source Java library for computational proteomics. *BMC bioinformatics* 2011, *12*, 70.
[7] Schmidt, C., Hesse, D., Raabe, M., Urlaub, H., Jahn, O., An automated in-gel digestion/iTRAQ-labeling workflow for robust quantification of gel-separated proteins. *Proteomics* 2013, *13*, 1417-1422.
[8] Shevchenko, A., Loboda, A., Ens, W., Schraven, B.*, et al.*, Archived polyacrylamide gels as a resource for proteome characterization by mass spectrometry. *Electrophoresis* 2001, *22*, 1194-1203.
[9] Przybylski, C., Junger, M. A., Aubertin, J., Radvanyi, F.*, et al.*, Quantitative analysis of protein complex constituents and their phosphorylation states on a LTQ-Orbitrap instrument. *J Proteome Res* 2010, *9*, 5118-5132.
[10] White, I. R., Pickford, R., Wood, J., Skehel, J. M.*, et al.*, A statistical comparison of silver and SYPRO Ruby staining for proteomic analysis. *Electrophoresis* 2004, *25*, 3048-3054.
[11] Granvogl, B., Ploscher, M., Eichacker, L. A., Sample preparation by in-gel digestion for mass spectrometry-based proteomics. *Anal Bioanal Chem* 2007, *389*, 991-1002.
[12] Switzar, L., Giera, M., Niessen, W. M., Protein digestion: an overview of the available techniques and recent developments. *J Proteome Res* 2013, *12*, 1067-1077.
[13] Sipos, T., Merkel, J. R., An effect of calcium ions on the activity, heat stability, and structure of trypsin. *Biochemistry* 1970, *9*, 2766-2775.

Chapter 4

FINAL DISCUSSION

Quantitative MS techniques allow detecting protein and PTM changes between different biological samples. The latest trends in MS instrumentation now provide "state-of-the-art" label-free quantification but these instruments and associated methods are not yet routine. Labeling strategies are nevertheless gradually losing ground compared to these label-free strategies due to the higher cost and comprehensive sample preparation but could still be of great value in the field [1]. As demonstrated in the different parts of the results presented in this work, the multiplexing potential of the labels still creates opportunities for answering specific questions. With iTRAQ quantification bias being increasingly addressed recently (cfr. introduction part 2, 2.2.1.2.), featuring the advantages of examining protein expression alterations with iTRAQ as opposed to label-free approaches are now more applicable than ever: (i) variation associated with the subsequent steps in the analysis is excluded as samples are combined at the peptide level after the labeling [2], (ii) the signal and sensitivity is increased as differentially labeled peptides from different samples accumulate in the MS1 spectrum creating one isotope envelope, and (iii) the background noise of complex samples is diminished compared to isotopic labels as quantification takes only place at the MS2 level. Compared to isotopic labels, the isobaric strategies also have the benefit of labeling up to ten samples simultaneously, which apart from saving a lot of analysis time, equally reduces technical variation over a large sample cohort [3, 4]. In this dissertation, bottom-up label-based quantitative MS was applied to study both PTM biology - H2A proteolysis and large-scale phosphorylation – and to optimize sample preparation protocols.

The most popular application of label-based quantitative proteomics still is to monitor the differences in protein expression between biological samples, the application for which label methods were originally designed [3]. In a first study, we did exactly this, by applying the isobaric iTRAQ technique for a biomarker analysis on CLL cells derived from patients with a distinct predicted outcome [5]. In our experimental setup, the effect of co-eluting peptides was reduced by an orthogonal fractionation before the LC-analysis, coordinately leading to a deeper coverage of the proteome. Further, to increase the confidence of the results, proteins were only included for quantification based on stringent conditions. Nevertheless, caution remains advised when analyzing proteomics data, even when all the technical pitfalls are taken into account. Here, our data surfaced a general up-regulation of nuclear and a down-regulation of cytoplasmic proteins, pointing towards an easily overseen difference in morphology between the diseased and healthy B-cells. Such cell-wide differences hamper any proteome-wide comparison. Thus, thorough prior knowledge of the biological background and the use of other techniques that complement MS are always essential for correct data interpretation. However, this inability to make proteome-wide conclusions inspired us to conduct a manual data analysis. This resulted in the discovery of a clipping product of histone H2A. This was not obviously exhibited by the software since conventional iTRAQ research focusses only on the protein level, in part because this

partially levels out many of the iTRAQ-related technical pitfalls seen at the single MS/MS level [5]. The clipping of the H2A C-tail after V_{114} was already described in leukemia cell lines but remained surprisingly underreported given its enormous biological potential, especially in a time with a growing buzz around histone epigenetics [6, 7]. cH2A prevalence was thus further inspected in CLL cells and healthy B-cells, initially by classical gel-based experiments based on anti-H2A immunoblotting. Gel-based methods unfortunately could not corroborate our findings due to the low mass resolution, lack of information about the exact clipping site and lack of quantitative accuracy. An alternative, more sensitive MS strategy was thus developed, based on spiked isotopically labeled peptides, wherein a second peptide compensated for the amount of total histone H2A present in each sample. This application of AQUA peptides is to our knowledge the first time a technique was defined to quantify a specific histone clipping event in a high throughput fashion. Screening a larger CLL patient population by this optimized protocol however, rejected any connection with the disease. Moreover, flow cytometry screening of these samples showed that clipping was coupled to cells of myeloid origin, a finding that was confirmed by screening cH2A in histone extracts from cells of different hematopoietic origin [5]. Thus, again these findings demonstrate that the results of an iTRAQ screening applied for a biomarker study require confirmation by a targeted, selective technique, preferably on a larger sample size. It would not be the first time that biomarkers are assigned incorrectly to a disease [8].

Subsequently, we identified the enzyme NE as being in the capacity of the “H2Asp”, the term that has been used to refer to the previously unknown enzyme responsible for the clipping of H2A V_{114} [9]. Apart from potential biological implications, this annotation also helps framing and critically revising our earlier findings as well as previous reports comprising H2A clipping, since complete *in vitro* inactivation of NE is extremely difficult [10]. Unintentional histone proteolysis during cell disruption now hampers pursuing the biological role of H2A clipping, if there is one. Equally, due the high degree of variation, our sensitive method could not confirm the previous findings of Okhawa *et al.* describing a brief cH2A formation during induced differentiation of the myeloid THP-1 cell-line [11]. Further research reflecting on histone proteolysis thus becomes even more pertinent if the biological impact of this phenomenon is ever to be unravelled. The scientific community working on histone PTMs is still surprisingly unaware of (uncontrolled) histone clipping both *in vivo* and during experimental procedures. So before classic degradomics methods are being applied on histones, accurately miming *in vitro* versus *in vivo* proteolysis is important.

Apart from technical pitfalls, biological dynamics also hinders the understanding of molecular mechanisms and biomarker studies [8]. Instead of defying a single disease-related protein or PTM, differential clusters of proteins or PTMs are by definition more robust and less prone to the variables

when trying to elucidate cellular changes. In this dissertation, the targeted screening of proteolysis as a PTM was extended and iTRAQ was applied in a very specific way for proteome-wide quantitative screening of one of the most studied and omnipresent modifications in the cell: phosphorylation. Erratic phosphorylations are coupled to disease mechanisms, so since the investigation can play a role in the development of new therapeutics, it is not surprising that multiple strategies exist for analyzing this PTM. Despite the technical challenges (introduction part 3, 1.2.2.) identification of phosphorylation sites is common practice in specialized laboratories. Quantification of phosphorylation between different samples becomes also more prominent, mainly after a parallel enrichment step. Techniques to determine the phosphorylation stoichiometry i.e. calculating the ratio of the phosphorylation site versus the non-phosphorylated counterpart, is on the other hand only described in a handful of reports.

Distinguishing high and low phosphorylation stoichiometry peptides is important to detect the changes in phosphorylation and to scrutinize biologically relevant sites [12]. One interesting methodology makes use of chemical labeling and partial dephosphorylation of peptide samples to do so [13, 14]. Surprisingly, the one report that implemented this principle for the large-scale stoichiometry study has thus far not led to the routine application of this approach, despite the many references to it. By using iTRAQ, we persisted in the optimization of an analogous technique, named Phospho-iTRAQ, for the large-scale phospho-analysis without enrichment. The multiplexing potential of the labels also allows to add internal replicates or simultaneous (phospho)proteome comparison between samples. After thoroughly optimization of the protocol, the technique was still confronted by two major pitfalls after analysis on our in-house (lower resolution) Q-TOF MS: (i) the small amount of phosphopeptides was lost in the background of the vast majority of non-phosphopeptides (ii) contaminating peaks in the iTRAQ reporter region hampered accurate quantification.

To overcome these challenges we turned to newer generation mass spectrometers with improved performance, simultaneously testing the MS systems for purchase purposes. HeLa cells were subjected to the Phospho-iTRAQ protocol and the sample was split into three parts for the acquisition on high resolution and fast acquisition instruments, thereby creating the largest dataset thus far analyzed in the proteome facility in our lab. The tested instruments rely on different software so it was crucial to standardize the data analysis workflow. Although the number of phosphopeptides filtered out of the data was elevated exponentially compared to the initial analysis on our Q-TOF Premier, drawing quantitative conclusions on the peptide level remains challenging. In large-scale proteomics, the amount of (phospho) identifications is often defined as the benchmark of success, but we prefer to elaborate more on the critical and transparent representation of the presented data. Also on these higher resolution instruments, the diluting effect of the iTRAQ ratios caused by co-eluting and co-selected

peptides was demonstrated, still hampering the accuracy of the Phospho-iTRAQ method. We did however provide a workflow to filter out either the high stoichiometry phosphopeptides or for selecting a larger, but overall less reliable list of peptides for a preliminary study. For the large-scale study, the effect of co-selected, co-fragmented peptides can potentially be diminished by extensive fractionation such as gel electrophoresis or by future technologies such as IM-separation [3, 15, 16]. Gel fractionation equally extends the proteome coverage, which is especially favourable when looking at PTMs that are only present in a low concentration.

These faster instruments with a high resolving power generate more and more data, thereby increasing the sample coverage and decreasing the amount of unidentified spectra. Concerning the workload, a shift from sample prep to more elaborate data analysis was thus experienced. MS-based proteomics with the newer generation instruments is clearly accompanied by increasing demands on the computational framework. Remarkably, even though we applied the same software and parameters for the multiple Phospho-iTRAQ DB searches, the overlap between the acquisitions on the different instruments was quite scarce owing to the inter run variation of DDA acquisition strategies both at the level of precursor selection as well as differential MS/MS quality and software analysis. Compared to the DDA approach, data-independent acquisition or DIA label-free quantification methods have the advantage of scanning the proteome sample in a relative unbiased way. Due to the lower cost, simpler sample preparation and easy-to-use bioinformatics software with built-in statistical software tools, it is expected that DIA label free techniques will become the new method of choice for standard comparative studies. Correspondingly, research projects in the lab of Pharmaceutical Biotechnology will increasingly focus on quantitative MS, more and more relying on the methodologies provided by the new instruments and building on the lessons learned in this dissertation. To guaranty the reproducibility, spiked in standards are however vital to create confident results. On the other hand, low-reproducibility steps, often mandatory in PTM research, keep the door open for label-based quantitative techniques.

For both label-free and label-based quantitative MS, it is important to realize that quantification strategies are merely an introductory stepping-stone to identify subsets of deviating protein patterns. For example, for our Phospho-iTRAQ method, additional confirmation of the identified phosphopeptides with complementary techniques remains necessary. We thus emphasize that detailed validation – even/especially of large datasets – still is a great asset in mass spectrometry, despite the tendency for automation. In light of this, label-free techniques are somewhat ambiguous since the complexity of the coupled (patented) software packages does not promote data transparency. Due to the biological complexity and overall technical variation, replicate analysis with robust protocols is designated as the

cornerstone of future quantitative proteomics research. The potential to multiplex samples makes label-based methods an important tool to fulfil this requirement as technical variation is minimized and replicates are combined [1]. The higher resolution instruments now even allow for the parallel quantification of up to ten samples simultaneously using isobaric labels.

In a final chapter, we again relied on the iTRAQ chemistry and its capacity to multiplex and implement replicates. We demonstrated that the isobaric tags are a suitable tool for method optimization and present a strategy wherein (i) proteins are not necessarily annotated and (ii) the average of all the data in a run is compared to study yield differences. This alternative analysis of labeled peptides circumvents the known bias associated with iTRAQ and simultaneously minimizes the variation in DDA peptide identification since all acquired data is analyzed. As continued optimization and validation of already existing techniques is mandatory, especially with the label-free revolution ahead which require a greater reproducibility, we applied the method to validate the steps of the in-gel digestion protocol, resulting to a ten-fold increased protein yield. Coordinately, we bridge the gap between the gel-based and MS-based methods by again emphasizing the value of gel fraction steps prior to LC-MS/MS analysis [17].

In conclusion, MS-based proteomics is a research field in continuous progression. Depending on the research question, different complementary strategies can be applied, making a compromise between sensitivity, speed, selectivity and robustness. Quantitative approaches are now the centre of most studies, and although label-free techniques are gaining popularity, labeling strategies could very well continue to play a pivotal role as demonstrated by the different projects described in this work [1]. The different applications of iTRAQ that were presented, all exhibited the strength of isobaric methods, albeit known downsides were equally observed. The methodology is primarily suited for (PTM) discovery studies and also benefits from improved instruments although validation and replicate analysis obviously remain essential. Due to the elaborated sample prep, extensive data analysis and overall cost-effectiveness, many would probably argue that label-based quantitative proteomics strategies will primarily remain limited to early exploration studies and the academic world.

References

[1] Bantscheff, M., Lemeer, S., Savitski, M. M., Kuster, B., Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* 2012, *404*, 939-965.
[2] Russell, M. R., Lilley, K. S., Pipeline to assess the greatest source of technical variance in quantitative proteomics using metabolic labelling. *Journal of proteomics* 2012, *77*, 441-454.
[3] Evans, C., Noirel, J., Ow, S. Y., Salim, M.*, et al.*, An insight into iTRAQ: where do we stand now? *Anal Bioanal Chem* 2012, *404*, 1011-1027.
[4] Vaudel, M., Sickmann, A., Martens, L., Peptide and protein quantification: a map of the minefield. *Proteomics* 2010, *10*, 650-670.
[5] Glibert, P., Vossaert, L., Van Steendam, K., Lambrecht, S.*, et al.*, Quantitative Proteomics to Characterize Specific Histone H2A Proteolysis in Chronic Lymphocytic Leukemia and the Myeloid THP-1 Cell Line. *International journal of molecular sciences* 2014, *15*, 9407-9421.
[6] Eickbush, T. H., Watson, D. K., Moudrianakis, E. N., A chromatin-bound proteolytic activity with unique specificity for histone H2A. *Cell* 1976, *9*, 785-792.
[7] Osley, M. A., Epigenetics: how to lose a tail. *Nature* 2008, *456*, 885-886.
[8] Angel, T. E., Aryal, U. K., Hengel, S. M., Baker, E. S.*, et al.*, Mass spectrometry-based proteomics: existing capabilities and future directions. *Chemical Society reviews* 2012, *41*, 3912-3928.
[9] Dhaenens, M., Glibert, P., Lambrecht, S., Vossaert, L.*, et al.*, Neutrophil Elastase in the capacity of the "H2A-specific protease". *Int J Biochem Cell Biol* 2014, *51*, 39-44.
[10] Groth, I., Alban, S., Elastase inhibition assay with peptide substrates - an example for the limited comparability of *in vitro* results. *Planta medica* 2008, *74*, 852-858.
[11] Minami, J., Takada, K., Aoki, K., Shimada, Y.*, et al.*, Purification and characterization of C-terminal truncated forms of histone H2A in monocytic THP-1 cells. *Int J Biochem Cell Biol* 2007, *39*, 171-180.
[12] Dephoure, N., Gould, K. L., Gygi, S. P., Kellogg, D. R., Mapping and analysis of phosphorylation sites: a quick guide for cell biologists. *Mol Biol Cell* 2013, *24*, 535-542.
[13] Pflieger, D., Junger, M. A., Muller, M., Rinner, O.*, et al.*, Quantitative proteomic analysis of protein complexes: concurrent identification of interactors and their state of phosphorylation. *Molecular & cellular proteomics : MCP* 2008, *7*, 326-346.
[14] Wu, R., Haas, W., Dephoure, N., Huttlin, E. L.*, et al.*, A large-scale method to measure absolute protein phosphorylation stoichiometries. *Nature methods* 2011, *8*, 677-683.
[15] Ting, L., Rad, R., Gygi, S. P., Haas, W., MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nature methods* 2011, *8*, 937-940.
[16] Sturm, R. M., Lietz, C. B., Li, L., Improved isobaric tandem mass tag quantification by ion mobility mass spectrometry. *Rapid Commun Mass Spectrom* 2014, *28*, 1051-1060.
[17] Glibert, P., Van Steendam, K., Dhaenens, M., Deforce, D., iTRAQ as a method for optimization: enhancing peptide recovery after gel fractionation. *Proteomics* 2014, *14*, 680-684.

Chapter 5

SUMMARY

This PhD dissertation comprises the main findings of a five years' engagement at the proteomics facility at the Laboratory of Pharmaceutical Biotechnology of Ghent University. Proteomics, i.e. the analysis of the expression, localization, function, and interaction of the proteins expressed by an organism, is a discipline that has evolved from a supporting science in biochemistry into a field in its own right. The technical improvements of the mass spectrometer, today the primary instrument applied for the characterization of proteins, have contributed significantly to this progression. In a routine proteomics study, proteins are extracted from a biological sample and digested into peptides for a data-directed/depended MS analysis on a tandem mass spectrometer. This approach, known as bottom-up MS, subsequently identifies the proteins by matching masses of acquired peptides and peptide fragments to a database. Gel-based techniques, still the method of choice at the onset of my PhD research, continue to play a valuable role to address specific needs and in the vital conformation of MS results.

The introduction of this thesis gives an overview of the current status of MS-based proteomics, including future capabilities and overall difficulties. In brief, studies are challenged by the dynamic complexity and high dynamic range of the proteome. This complexity is further increased by post-translational modifications (PTM), i.e. chemical groups that are covalently bound to the proteins or proteolytic products. Since alteration of proteins by PTMs has severe biological consequences, many specialized methodologies have been developed to identify as many proteins and PTM sites as possible. Advanced MS-instruments have paved the way to detect more proteins than ever before and simultaneously allow implementing complementary strategies based on the increased resolution and speed of the machines. However, the need for PTM confirmation, quantification and functional interpretation remains omnipresent. Nowadays, quantitative MS techniques are therefore playing a pivotal role in most proteomics studies for detecting (disease specific) protein or PTM changes to address these requirements.

Labeling strategies are the basis of most quantitative strategies although modern instrumentation now equally provides promising label-free quantification approaches. These newer methods are not yet routinely applied but have the possibility to gain in popularity. Then again, labeling strategies continue to be important in quantitative proteomics, especially in terms of PTM research. The final discussion provides a knowledge base to highlight the power of labeling strategies, simultaneously describing the course of events. Multiple labeling strategies exist, each with their strengths and limitation. As a central theme throughout this work, the MS2 label method iTRAQ is used in multiple approaches, thereby leading to the discovery of an underreported proteolytic H2A product as well as contributing to the large-scale mapping of high-stoichiometry phosphorylated peptides. The results

demonstrate the immense potential of iTRAQ but equally surface the limitations and the stringent prerequisites in terms of data interpretation.

In a first part of this work, the iTRAQ method was applied for its "common" purpose: relative protein expression analysis between samples. In order to define disease related proteins, six protein samples of chronic lymphocytic leukemia (CLL) patients with a severe predicted outcome were compared with six samples of patients with a more indolent prognosis and with samples from healthy control patients. A substantial part of the observed differences between healthy and CLL cells was however attributed to the inequality in morphology between healthy and neoplastic cells, a caveat that should be taken into account in future biomarker research. Interestingly, due to the increased sensitivity introduced by the iTRAQ labels, a proteolytic product of histone H2A after V_{114} (cH2A) was rediscovered by screening the data at the peptide level. The overall literature regarding histone proteolysis is quite scattered, yet cH2A was already described in some reports and even observed in leukemia. As iTRAQ is predominantly a discovery technique, requiring extensive validation, we developed a more sensitive method based on the isotopically labeled AQUA peptides to gain more insight in the specific phenomenon of H2A V_{114} clipping. This optimized application of spiked standard peptides marks the first time a technique was defined to quantify a specific histone clipping in a high throughput fashion. Screening of 36 CLL patients refuted any link with the disease and suggested that cH2A was related to cells of myeloid origin. This observation was in alignment with the observed rise up of cH2A during induced differentiation of THP-1 promonocytes, implying a potential biological role of H2A processing in the hematopoietic development of cells from myeloid origin. However, synchronization of the identical THP-1 cell line before differentiation abrogated these results, suggesting that differences in cH2A is caused by variation in cell cycle stage.

Screening of hematopoietic cells from different origin derived from blood of healthy subjects, by both Western blot and the AQUA method, confirmed the myeloid character of H2A V_{114} clipping *in vivo*. *In vitro*-assays and *null* mice further led to the identification of Neutrophil Elastase (NE) as being in the capacity to be the "H2A-specific protease", the previously unknown enzyme described as the mediator of H2A clipping. This discovery provided the background to critically revise earlier reports on H2A clipping and confirms that drawing biological conclusions on proteolysis remains a hard task since distinguishing *in vivo* and *in vitro* protease activity is extremely difficult.

Next, another application of iTRAQ was developed to quantify protein phosphorylation stoichiometry, named Phospho-iTRAQ. Protein phosphorylation controls cellular communication and is one of the most extensively studied PTMs as irregular phosphorylation patterns are associated with disease states. Although detection is challenging due to the dynamic regulation, low stoichiometry and

heterogeneous character of phosphorylation, specialized strategies have enabled the identification of thousands of phosphosites. Quantitative approaches in phosphoproteomics exist but mainly aim to elucidate relative phosphorylation changes between samples. However, information on the unmodified counterpart of identified phosphopeptides and measuring changes in protein expression is equally important to interpret the relevancy of a phosphosite. With the Phospho-iTRAQ method, we optimized a protocol that focuses only on the non-phosphorylated counterparts of the phosphopeptides for stoichiometry measurement, avoiding specialized and relatively imprecise workflows such as phospho-enrichment. This is the first strategy, whereby a peptide sample is briefly split in two identical parts and differentially labeled preceding the phosphatase treatment of one part, which relies on isobaric tags for the large-scale analysis. The technique was eventually validated on control and EGF-stimulated HeLa cells, acquired on three different high quality LC-MS/MS instrument. The spiked internal standards demonstrated the ability of Phospho-iTRAQ to detect phosphopeptides of high stoichiometry out of the data consisting for the most part of non-phosphorylated peptides. By a standardized workflow, long lists of phosphopeptides were filtered out of the large datasets from the different instruments. Remarkably, on average 35% of the annotated phosphopeptides were shared between two independently generated protein lysates and, in general around 20% of the peptides were previously reported as phosphosites.

As multiple labels are used in our method, information on relative protein expression can be obtained or internal duplicates can be implemented, all in a single analysis. Further, the flexibility of the method was emphasized by coupling Phospho-iTRAQ to an additional step of gel-fractionation, significantly increasing the proteome coverage. However, observed variation and limited overlap between duplicate analyses exhibited that the accuracy of the methodology is still challenged by iTRAQ bias, a phenomenon that can be expected for any non-targeted quantitative approach. Confirmation of identified phosphopeptides with complementary techniques thus remains mandatory, also because the approach is not able to pinpoint the exact phosphorylation site. Depending on the goal of the study, Phospho-iTRAQ may serve as a tool to annotate high stoichiometry peptides out of a complex dataset or may focus on a larger list of peptides containing a potentially larger subset of falsely accepted phosphopeptides.

Finally, iTRAQ was adopted for optimizing and comparing methodologies since multiplexing obliterates all variation after testing various experimental conditions. When avoiding annotation of the peptides and analyzing the average ratios of all the spectra in a run, all the acquired data is taken into account, thereby minimizing the common observed problems associated with iTRAQ. Standard series of known ratios demonstrated the high accuracy and precision of the approach, also on a lower-resolution MS-instrument. Since gel-based methods prior to MS analysis are still important in sample fractionation

or specific inquiries, the steps from the in-gel protocol were validated. The negative effect of SYPRO Ruby staining and the positive effect of gel fixation were displayed and the addition of $CaCl_2$ and ACN to a tryptic digest resulted in an up to tenfold enhancement of the peptide yield.

The different projects conducted in this work exhibited the strengths of labeling methods such as iTRAQ for diverse applications. Label based quantitative MS strategies have a certain flexibility and can get around low-reproducibility steps in a protocol, making them an important tool in PTM research. Critical interpretation of the data is however essential and as for all research, validation of preliminary results remains needed.

Hoofdstuk 6

SAMENVATTING

Dit proefschrift beschrijft de belangrijkste conclusies van mijn doctoraat aan de proteomics faciliteit van het Laboratorium voor Farmaceutische Biotechnologie van de Universiteit Gent waarvoor ik mij vijf jaar lang heb geëngageerd. Proteomics wordt gedefinieerd als “de wetenschap die de expressie, lokalisatie, functies en interacties van eiwitten (proteïnen) binnenin een organisme analyseert”. Proteïnen bestaan uit lange ketens van aminozuren en spelen een essentiële rol in zo goed als elk biologisch proces. De laatste tien jaar is deze discipline geëvolueerd van een ondersteunende tak binnen de biochemische wetenschappen naar een volwaardig onderzoeksdomein. De technische verbeteringen van de massaspectrometer, tegenwoordig hét belangrijkste instrument dat wordt gehanteerd om eiwitten te karakteriseren, hebben sterk bijgedragen aan deze evolutie.

In een routine proteomics onderzoek worden de eiwitten eerst geëxtraheerd uit een biologisch staal (bijvoorbeeld kankercellen) en vervolgens verknipt in kleinere stukken, de peptiden, voor een twee-staps analyse op de massaspectrometer. Peptiden worden geïoniseerd (krijgen een lading) en in een eerste stap meet de massaspectrometer de massa (correcter is “massa op lading”) van de volledige peptiden (MS1). In een tweede stap worden de peptiden die het sterkste signaal geven geselecteerd uit het peptidemengsel en gefragmenteerd in kleinere peptide fragmenten (MS2). De informatie van de massa’s van zowel de peptiden als hun fragmenten worden softwarematig gekoppeld aan een database waardoor peptiden en eiwitten geïdentificeerd kunnen worden. Deze analysemethode, waarbij eiwitten via de peptiden worden geïdentificeerd aan de hand van een twee-staps analyse wordt ook wel “bottom-up” tandem massaspectrometrie genoemd.

Naast de massaspectrometer spelen ook andere, complementaire technieken, een belangrijke rol in proteomics. Bij de start van mijn onderzoek waren de populairste analytische methoden vooral gebaseerd op gel, waarbij complexe eiwitmengsels gescheiden worden binnen een gelmatrix. Deze technieken spelen nog steeds een belangrijke rol en helpen om specifieke onderzoeksvragen te beantwoorden en resultaten van een massaspectrometrische analyse te confirmeren.

In de inleidende hoofdstukken van dit proefschrift wordt de huidige positie van de massaspectrometer binnen de proteomics beschreven en toekomstige (technische) evoluties geschetst. Naast de technische toelichting wordt ook stilgestaan bij de uitdagingen en moeilijkheden van massaspectrometrische proteomics. Proteomics studies worden namelijk gehinderd door het dynamische karakter van de proteïnen, de complexiteit van peptidestalen en het groot concentratieverschil tussen verschillende eiwitten, zowel binnen eenzelfde mengsel, als tussen verschillende stalen. Deze complexiteit is onder meer te wijten aan de oneindig grote variatie aan post-translationele modificaties (PTM). PTM's zijn veranderingen van eiwitten die voorkomen na de translatie, het proces waarbij eiwitten worden aangemaakt uit mRNA, een afgeleide van het DNA, de

belangrijkste drager van onze erfelijke informatie. Deze veranderingen zijn meestal chemische groepen die covalent gebonden worden aan de eiwitbouwstenen (aminozuren), maar ook fragmenten van eiwitten die in de cel worden gegenereerd, worden beschouwd als PTM's. Het proces waarbij een proteïne wordt verknipt, wordt proteolyse genoemd en wordt uitgevoerd door de enzymenfamilie van proteasen.

Aangezien de aan- of afwezigheid van een PTM vergaande biologische gevolgen kan hebben en zowel de structuur en functie van eiwitten kan wijzigen, zijn er vele gespecialiseerde methoden ontwikkeld om eiwitten en hun PTM's te analyseren. Met de nieuwste generatie massaspectrometers worden er vandaag de dag in één analyse al over de duizend eiwitten geïdentificeerd. De toenemende resolutie en snelheid van deze toestellen zorgt er eveneens voor dat nieuwe strategieën die complementair zijn met de bovenvermelde “bottom-up” massaspectrometrie in populariteit toenemen. De nood om PTM's correct te identificeren, kwantificeren en interpreteren blijft weliswaar groot. Om wijzigingen van een eiwit of PTM te detecteren en de toenemende hoeveelheid massaspectrometrische data beter te begrijpen spelen kwantitatieve methoden bijgevolg een sleutelrol.

Het leeuwendeel van de kwantitatieve methoden zijn gebaseerd op chemische labels. Met de nieuwste massaspectrometers is kwantificatie ook mogelijk met strategieën die niet afhankelijk zijn van chemische labels, maar deze technieken worden voorlopig nog niet routinematig toegepast. Label-methoden blijven evenwel van vitaal belang in kwantitatieve proteomics, vooral om PTM's in kaart te brengen. In de finale discussie van de thesis worden de voordelen van label-kwantificatie benadrukt en gekoppeld aan de verschillende onderdelen van de resultaten. Er bestaan verschillende label-strategieën om proteïnen te kwantificeren, ieder met zijn sterktes en zwaktes. In deze thesis werd bij elk onderzoeksproject gebruik gemaakt van de iTRAQ-label- methode.

Met iTRAQ worden peptiden afkomstig van verschillende stalen gelabeld met een verschillend iTRAQ-label en onmiddellijk daarna samengevoegd. Omdat iTRAQ- labels isobaar zijn (wat wil zeggen dat de verschillende labels een andere samenstelling maar dezelfde chemische eigenschappen en massa hebben) zullen ze bij de massaspectrometrische analyse niet gescheiden worden tot na de fragmentatie waardoor kwantificatie pas op het MS2-niveau zal plaatsvinden. Dit heeft als groot voordeel dat het signaal van een peptide afkomstig van verschillende stalen wordt versterkt en dat de technische variatie die steeds wordt geïntroduceerd bij de staalvoorbereiding wordt beperkt. Het gebruik van iTRAQ leidde tijdens mijn doctoraatsonderzoek o.a. tot de ontdekking van een proteolytisch product van het histoneiwit H2A en werd ook gehanteerd om een techniek te ontwikkelen die de verschillen in fosforylatie tussen stalen kan meten. De verschillende resultaten demonstreren het immense potentieel van iTRAQ maar etaleren ook de gebreken eigen aan de techniek, waardoor een kritische benadering

van de gegenereerde resultaten vereist is. De belangrijkste onderzoeksprojecten uitgevoerd gedurende mijn doctoraatsthesis staan beschreven in de resultatensectie van het proefschrift en worden hier kort besproken.

In een eerste project werd de iTRAQ-methode gehanteerd voor zijn gangbare toepassing: het meten van de relatieve eiwitverschillen tussen stalen. De eiwitmonsters, afkomstig van patiënten met chronische lymfatische leukemie (CLL), de meest voorkomende vorm van leukemie, werden onderzocht om meer inzicht te krijgen in deze pathologie. Het ziekteverloop van CLL kan enorm gevarieerd zijn: sommige patiënten hebben een vrij agressief ziektebeeld met korte levensverwachting, anderen hebben een veel gunstigere prognose met een levensverwachting van tientallen jaren. Voor een verkennende studie werden de eiwitextracten van zes patiënten met de agressieve vorm van CLL en zes extracten van patiënten met de mildere vorm a.d.h.v. iTRAQ met elkaar vergeleken. Omdat CLL een kanker is van de B-lymfocyten werden in dezelfde analyse ook B-cellen van gezonde personen getest om simultaan potentiële ziekte-specifieke eiwitten in kaart te brengen. Een groot deel van de gemeten proteïneverschillen tussen gezonde en kankercellen was echter toe te schrijven aan de morfologische verschillen tussen beide celpopulaties: kankercellen hebben over het algemeen een grotere kern dan gezonde cellen, hetgeen eenduidige besluitvorming bemoeilijkt, een gegeven waarmee te weinig rekening gehouden wordt bij kankeronderzoek. Verdere manuele analyse van de dataset leidde tot de ontdekking van een proteolytisch product van het eiwit histon H2A na aminozuurnummer 114 van de sequentie valine 114. Histon H2A is één van de “kernhistonen”, een beruchte klasse van proteïnen waarrond ons DNA gebonden is. De PTM's van deze histonen bepalen de activiteit van het DNA: bepaalde modificaties hebben een stimulerend effect op de DNA- transcriptie, andere inhiberen deze. De proteolytische H2A-modificatie na valine 114, eerder al gedefinieerd als cH2A, werd midden jaren zeventig voor het eerst beschreven in de literatuur en werd in sommige rapporten gekoppeld aan bepaalde vormen van leukemie. Om de bevindingen van de verkennende iTRAQ-analyse te confirmeren en na te gaan of cH2A een kenmerk/biomerker is van CLL, werd een tweede, meer specifieke en gevoelige methode opgesteld om cH2A te kwantificeren. Deze geoptimaliseerde methode, gebaseerd op een interne standaard mix van isotopisch gelabelde AQUA-peptiden, is de eerste techniek die werd ontwikkeld om specifieke histon proteolyse high-throughput te kwantificeren. De cH2A-screening van eiwitstalen, geëxtraheerd uit de witte bloedcellen van 36 CLL-patiënten met verschillende prognose, weerlegde echter duidelijke de link met de pathologie. Uit de resultaten werd besloten dat cH2A eerder gerelateerd is aan cellen van myeloïde in plaats van lymfatische oorsprong (de myeloïde lijn staat in voor de vorming van een bepaalde categorie witte bloedcellen in het beenmerg). Onze observatie stemde overeen met een in de literatuur beschreven experiment dat de opkomst van cH2A aantoont tijdens geïnduceerde differentiatie van THP-1-promonocyten, een kankercellijn afgeleid van acute

monocytische leukemie. Omdat deze referentie een potentiële biologische rol van cH2A in de hematopoietische ontwikkeling van cellen van myeloïde oorsprong suggereert, werd dit experiment herhaald met behulp van onze gevoelige AQUA-peptiden. De synchronisatie van de identieke THP-1-cellijn voor differentiatie weerlegde de vooropgestelde hypothese en impliceerde dat de waargenomen verschillen in cH2A vooral worden veroorzaakt door de variatie in celcyclus bij staalname.

De verdere screening van verschillende types hematopoietische cellen uit het bloed van gezonde proefpersonen met Western blot (een techniek waarbij op gel gescheiden eiwitten selectief kunnen worden gedetecteerd m.b.v. specifieke antilichamen) en de AQUA-methode, bevestigde het myeloïde karakter van cH2A. Laboratorium assays en knock-out muizen leiden in een parallel onderzoeksproject tot de identificatie van Neutrofiel Elastase (NE) als het enzym verantwoordelijk voor het knippen van histon H2A. Dit protease werd in de literatuur gedefinieerd als het “H2A-specifiek protease” maar werd nooit eerder geïdentificeerd. Dankzij deze annotatie kunnen verschillende voorgaande analyses en rapporteringen van cH2A verklaard worden. Anderzijds tonen de experimenten ook aan dat het trekken van biologische conclusies over proteolyse een bijzonder moeilijke opgave is aangezien de in vivo en in vitro activiteit van het NE protease uiterst moeilijk van elkaar te onderscheiden is.

In een tweede project werd iTRAQ aangewend om de fosforylatie stoichiometrie (de hoeveelheid van een specifiek fosfopeptide dat gefosforyleerd is) te meten door een techniek te optimaliseren die de naam “Phospho-iTRAQ” kreeg. De proteïnefosforylatie is een PTM die de communicatie tussen en binnenin de cellen controleert. Aangezien onregelmatige fosforylatiepatronen geassocieerd worden met verschillende pathologieën is de fosforylatie één van de meest bestudeerde modificaties. Hoewel de detectie van fosforylaties bemoeilijkt wordt door de dynamisch regulatie, de lage stoichiometrie en het heterogene karakter van deze PTM, hebben gespecialiseerde strategieën in totaal al duizenden fosfosites geïdentificeerd. De kwantitatieve massaspectrometrisch methoden die reeds beschreven werden focussen echter vooral op de relatieve veranderingen van de fosforylatie tussen verschillende stalen. Om de relevantie van geïdentificeerde fosfosites in te schatten is het echter belangrijk om ook informatie te winnen over de niet-gemodificeerde tegenhanger van fosfopeptiden binnen elk staal.

Met de “Phospho-iTRAQ”-methode werd een protocol geoptimaliseerd die zich focust op de niet-gefosforyleerde tegenhanger van de fosfopeptiden om de stoichiometrie te analyseren. Dankzij dit principe wordt de variatie-inducerende aanrijkingsstap die eigen is aan klassieke fosfo-bepalende-technieken vermeden. Voor deze strategie wordt een peptidestaal in twee identieke delen gesplitst en differentieel gelabeld, om vervolgens één van deze delen te behandelen met een fosfatase. Het is de eerste maal dat dit principe wordt toegepast met iTRAQ-labels voor de analyse van een volledig

proteoom. De techniek werd na een uitgebreide optimalisatie in een finaal experiment op een staal van controle en EGF-behandelde Hela-cellen (een populair onderzoeksmodel in de fosfoproteomics), gevalideerd op drie nieuwste generatie massaspectrometers. De toegevoegde interne standaarden toonden aan dat de “Phospho-iTRAQ”-techniek in staat is om peptiden van een hoge stoichiometrie uit een dataset te filteren die voornamelijk bestaat uit niet-gefosforyleerde peptiden. Dankzij een gestandaardiseerde workflow werden eveneens lange lijsten fosfopeptiden uit de grote dataset geselecteerd. Het was opmerkelijk dat gemiddeld 35% van alle geannoteerde fosfopeptiden voorkwamen in beide onafhankelijk gegenereerde eiwitextracten en dat 20% van de peptiden reeds gerapporteerd werden als fosfosites.

Omdat iTRAQ-kits bestaan uit meerdere labels (kits van 4 en 8 labels bestaan), kan in hetzelfde experiment de relatieve proteïne-expressie tussen stalen worden vergeleken of kan er gekozen worden om een experiment met een interne replica uit te voeren. Deze flexibiliteit werd aangetoond door een extra stap van gel-fractionering toe te voegen aan het “Phospho-iTRAQ”-protocol, wat leidde tot een uitgebreidere analyse van het proteoom. Doch, in de dataset werd ook een grote variatie en een beperkte overlap tussen duplicaatanalyses waargenomen. Dit fenomeen is vooral te wijten aan reeds eerder gerapporteerde onnauwkeurigheden die eigen zijn aan iTRAQ en andere “niet-gerichte” kwantitatieve technieken die op grote schaal worden toegepast. Confirmatie van de geïdentificeerde fosfopeptiden met complementaire technieken blijft dus noodzakelijk, temeer omdat onze methode wel de stoichiometrie van fosfopeptiden kan kwantificeren maar niet exact het aminozuur kan aanduiden dat is gemodificeerd. Afhankelijk van het doel van het experiment kan de ”Phospho-iTRAQ”-methode worden gebruikt om fosfopeptiden uit een complexe dataset in kaart te brengen of kan de onderzoeker beslissen om zich te concentreren op een langere lijst van fosfopeptiden voor een verkennende studie, wetende dat er in dat geval een groter aandeel vals positieven wordt geselecteerd.

In een laatste project werd iTRAQ specifiek voorgesteld als een tool om nieuwe onderzoeksmethoden te optimaliseren of om verschillende protocols met elkaar te vergelijken. iTRAQ is uiterst geschikt voor dit doeleinde aangezien de peptidestalen afkomstig van verschillende protocols onmiddellijk na het labelen opnieuw kunnen worden samengevoegd waardoor verdere variatie in de staalvoorbereiding vermeden wordt. Voor de data analyse werden peptiden niet geannoteerd zoals gebruikelijk, maar werd de gemiddelde intensiteit van alle spectra in een run berekend om stalen met elkaar te vergelijken. Deze benadering heeft als voordeel dat alle data van een analyse wordt bekeken en dat de gevolgen van de waargenomen onnauwkeurigheden die gepaard gaan met iTRAQ worden geminimaliseerd. Standaardreeksen van gekende ratios demonstreerden de hoge nauwkeurigheid en precisie van deze werkwijze, ook op massaspectrometers met een relatief lage resolutie. Omdat

gelelektroforese als voorbereidende stap op een massaspectrometrische analyse nog steeds een belangrijke toegevoegde waarde heeft, o.a. als fractionatiestap, werden de verschillende stappen van het in-gel digest protocol geoptimaliseerd. Tijdens dit protocol worden op gel-gescheiden eiwitten in een gel verknipt tot peptiden en geëxtraheerd. Het negatieve effect van SYPRO Ruby kleuring en het positieve effect van de gelfixatie werden duidelijk aangetoond en de toevoeging van $CaCl_2$ en acetonitrile aan het trypsine digest resulteerde zelf in een vertienvoudiging van de peptide opbrengst.

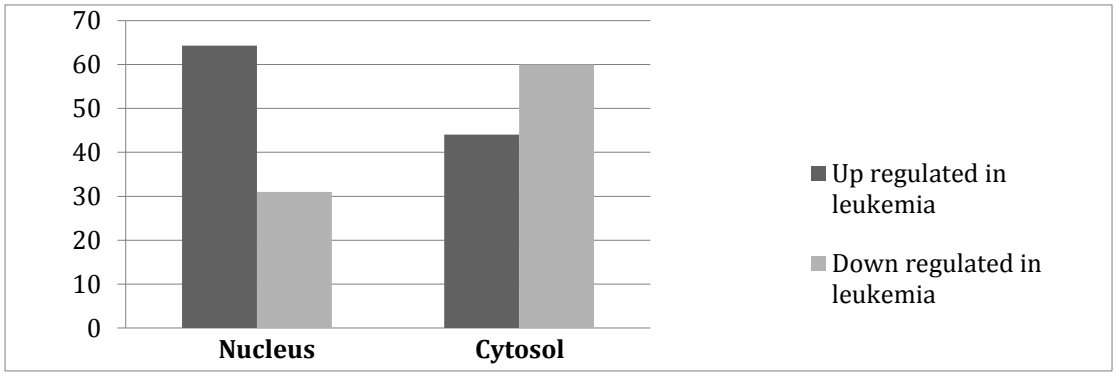
Samengevat: De verschillende onderzoeksprojecten uitgevoerd in dit doctoraat demonstreerden de sterkte van label-gebaseerde kwantificatie methoden zoals iTRAQ, en dit voor verschillende toepassingen. Omdat deze massaspectrometrische label-strategieën zeer flexibel kunnen worden aangewend en omdat de gevolgen van experimentele stappen van een lage reproduceerbaarheid kunnen worden omzeild, blijven deze methoden zeer belangrijk voor het onderzoek naar PTM's. Zoals bij alle wetenschappelijk onderzoek blijft een kritische interpretatie van de data echter noodzakelijk en is verdere validatie van de resultaten onontbeerlijk.

Chapter 7

ADDENDUM

SUPPORTING INFORMATION CHAPTER 3 – PART 1

A.3.1. Figure 1. Cellular component analysis of CLL (UM+ and M-) VS Healthy

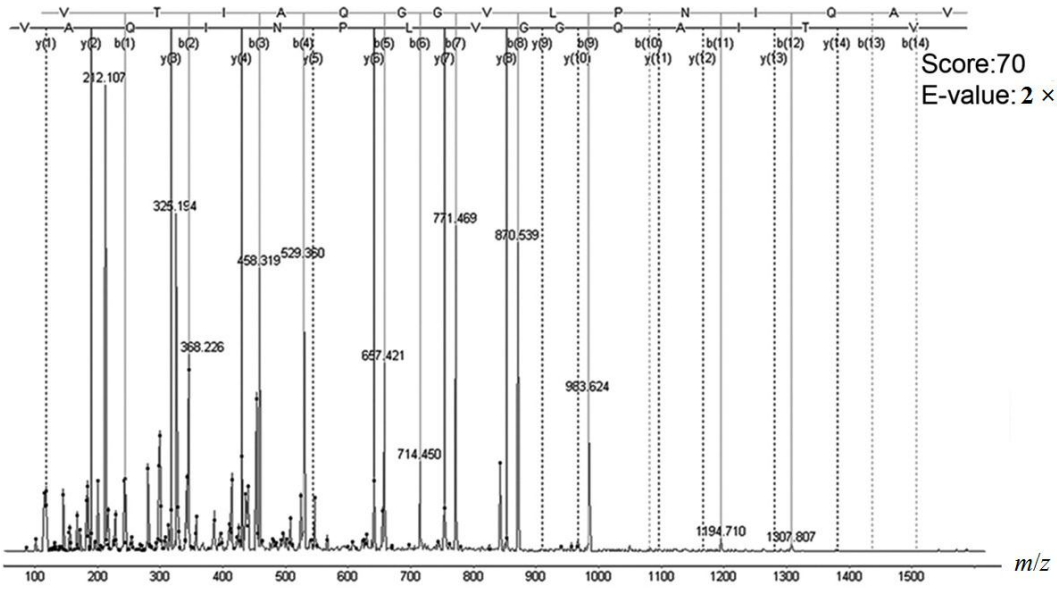


Main results from the GO analysis of the proteins upregulated and downregulated in [CLL (UM+ and M-) / Healthy] and downregulated in [CLL (UM+ and M-) / Healthy]. The cellular component analysis from WebGestalt [22] indicated that 64% of the proteins up-regulated in CLL are categorized as nuclear whereas down-regulated proteins are primarily located in the cytosol (60%) [24]. These results suggest morphological differences between the healthy and cancerous cells rather than true molecular aberrancies.

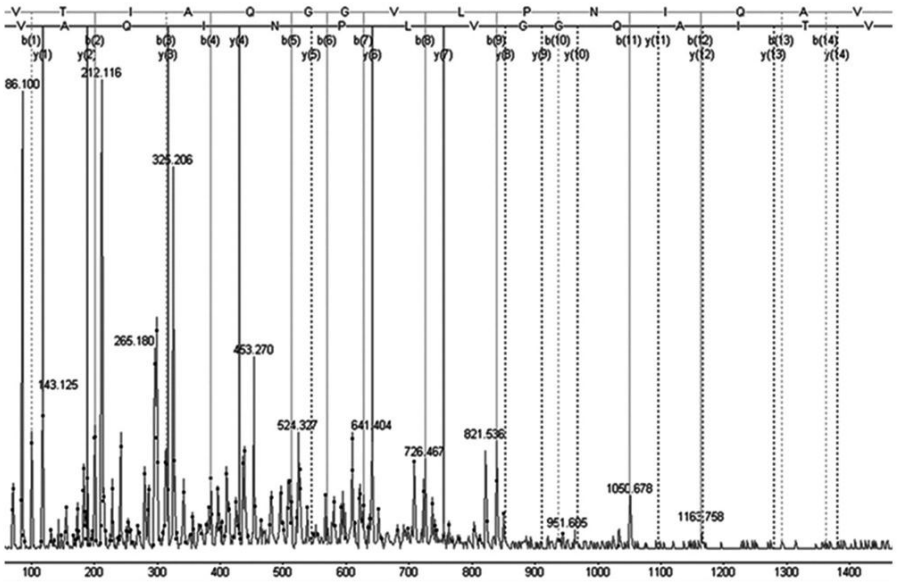
A.3.1. Figure 2. Annotation of the semi-tryptic cH2A$_{114}$ peptide

Validation of the annotation of the semi-tryptic peptide VTIAQGGVLPNIQAV on the raw data of the MSMS spectrum in Mascot Distiller (Matrix Science). *Upper panel:* The iTRAQ-labeled H2A peptide. The Mascot score of the labeled peptide is 70 and the expectation value is $2e^{-6}$. *Lower panel:* De novo sequencing of the non-labeled H2A peptide. The Mascot score is only 40 with a expectancy value of $3e^{-2}$.

This difference could be explained by the contribution of the iTRAQ labels which are covalently bound to the peptide N-termini, generating more intense b-series and consequently resulting in an increased peptide score. For tryptic peptides, this effect is of minor importance, but when the peptide does not have the conventional K or R at y1, this significantly changes the fragmentation pattern and facilitates the automated identification.

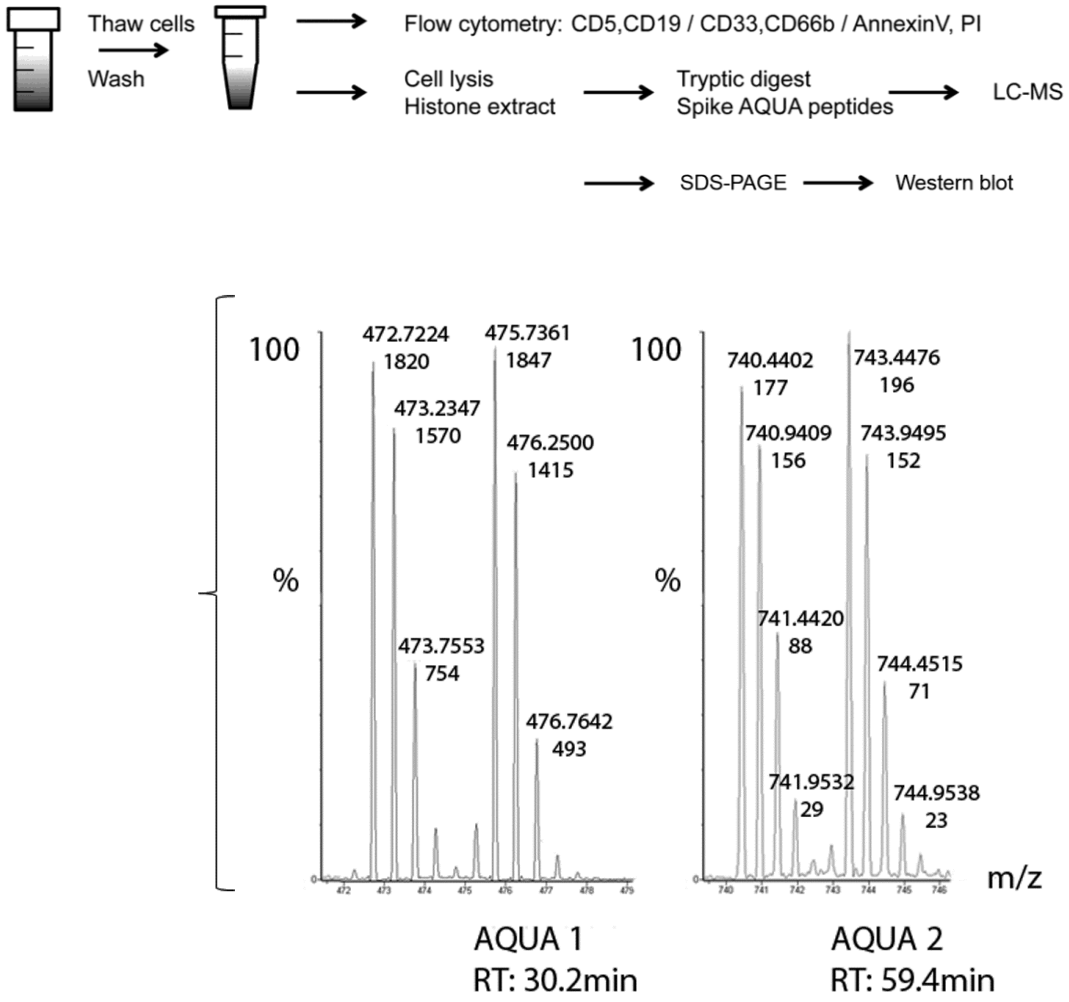


Score:70
E-value: 2×10^{-6}



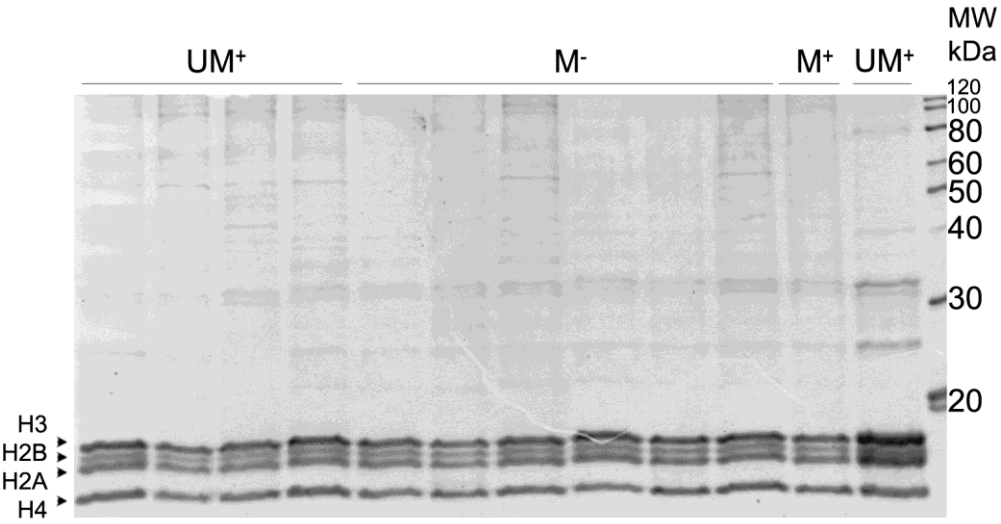
Score:40
E-value: 3×10^{-2}

A.3.1. Figure 3. Workflow of the CLL screening.



Scheme of the workflow of the CLL screening on 36 different patient cells with a variable prognosis. Frozen cells are thawed and washed with PBS. Flow cytometry was performed for CD5:CD19, CD33:CD66b and AnnexinV: PI analysis on $5x10^5$ cells. The rest of the cells were lysed and histones were extracted with acid. After protein quantitation, 3µg histones was applied for gel electrophoresis and subsequent Western blot. 2 µg was digested and the AQUA peptides were added before LC-MSMS. *Right panel:* At the top of the elution peak as seen on the extracted ion chromatogram of m/z 474,8 2+, the amount of total H2A in the sample was quantified with the AQUA 1 peptide. The AQUA 2 peptide (m/z 743,4 2+) was spiked to quantify cH2A in order to subsequently calculate clipped H2A. %cH2A = $\{(740.4/743.4)_{TIC}\}$ / $\{10(470.7/473.7)_{TIC}\}$ x 100.

A.3.1. Figure 4. Sypro staining of CLL-cells



The histone extracts from 12 samples (*From left to right:* patient samples UM^+ 8, 9, 10 and 11, M^- 9,10, 11, 12, 13 and 14, M^+ 5 and UM^- 6). Sypro staining of all the proteins demonstrate the purity of the HE. Clean bands of all the core histones were visualized together with the molecular weight marker in the exterior lane.

A.3.1. Table 1. Results CLL patient screening

Quantified parameters for the 36 CLL patient cohort:

The most recurrent diagnostic parameters in CLL are the mutational status of the B-cell receptor and the expression of the ZAP70 protein. Mutational status and ZAP70 expression were measured during clinical diagnosis. Mutationonal status is expressed as % DNA sequence homology compared to germline sequence. ZAP70 expression was quantified seperately by flow cytometry during clinical diagnosis. Unmutated (UM) B-cell recepetor and high ZAP70 expression predict poor clinical outcome.

Legend:

UM: unmutated CLL

M: Mutated CLL

ZAP: ZAP70 expression level **+/-:** ZAP positive or ZAP negative

%VH: mutational status of the B-cell receptor.

%CD5CD19: % CLL B-cells in the sample (combined CD5CD19 expression is a CLL B-cell marker)

%CD33: myeloid cell fraction

%CD66b: granulocyte cell fraction

PI: Propidium Iodide staining for quantitation of dead cells

AnV: Annexin V staining for quantitation of apoptotic cells

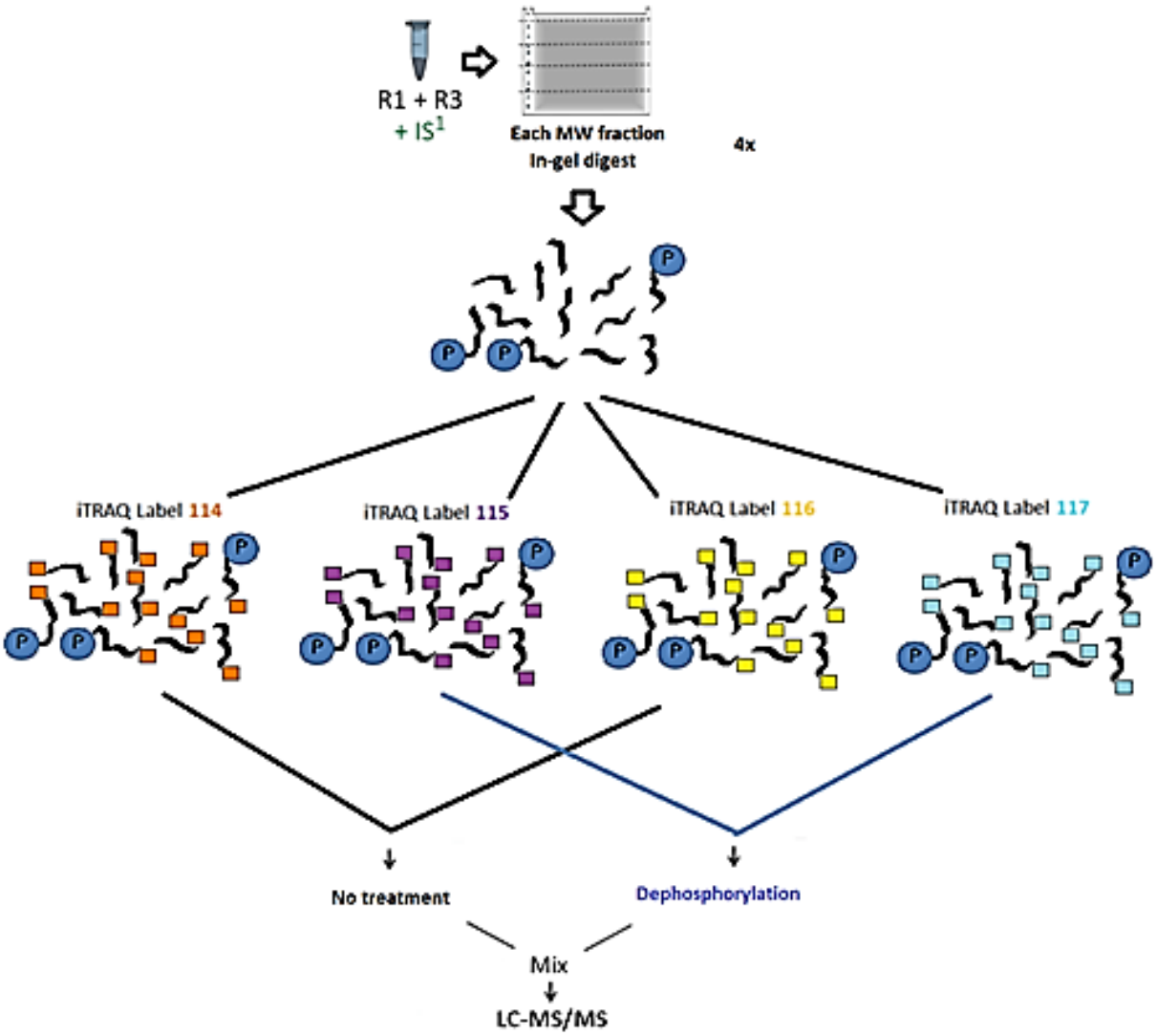
%cH2A: percentage of histone H2A that is cleaved at V114 as seen by our AQUA screening tool (Figure 1 and supplementary figure 3)

			Measured by Flow cytometry					Measured by AQUA
Diagnosed:	**ZAP**	**% VH**	**%CD5CD19**	**%CD33**	**%CD66b**	**%PI**	**% AnV**	**%cH2A**
UM^+1	95,2	100	75,4	0,5	2,1	54,9	86,3	**19,7**
UM^+2	94,3	100	93	0,5	1,8	79,5	41,4	**6,7**
UM^+3	94,1	98,9	84,8	1,5	1,6	54,4	53,9	**3,2**
UM^+4	93,8	100	75,5	1	2	29,9	57,4	**13,7**
UM^+5	91,8	100	71,9	0,3	1,7	13,4	82,8	**9,1**
UM^+6	86,1	100	64,5	0,4	3,7	46,1	88,4	**5,3**
UM^+7	84,9	99,3	74,9	0,9	0,8	5,9	94,6	**3,3**
UM^+8	83,7	100	61,3	0,9	1,3	7,6	77,5	**13,2**
UM^+9	81,7	100	38,7	4,3	6,4	51,9	94	**14,3**
UM^+10	54,2	100	76,4	1,1	5,7	84,7	99,6	**11,3**
UM^+11	53	100	77,63	0,6	0,9	22,5	71,85	**4,6**
M^-1	8,9	91,6	72,4	0,7	1,6	71,5	87,9	**12,1**
M^-2	9,8	95,1	91,5	0,1	2	73,8	75,2	**38,5**
M^-3	7,93	87,85	71,8	2,2	3,3	42	90,2	**14,2**
M^-4	8,3	92,2	94,8	0,7	1,8	11,6	69,9	**17,5**
M^-5	7,5	92	70,3	0,8	8,6	31,9	86,5	**33,2**
M^-6	7,9	96,2	74,8	1,5	4,6	20,3	93,6	**2,9**
M^-7	7,5	95,1	82	2	3,6	30	93,9	**3,2**
M^-8	8,91	92,98	65,6	5,7	2,6	24,5	83,8	**13,2**
M^-9	8,06	93,06	63	1,4	1,1	6,5	87,2	**17,6**
M^-10	7,74	93,75	80,6	1,8	14,1	6,5	86,2	**67**
M^-11	8,7	93,7	88,5	0,9	0,1	23,2	68,7	**0,7**
M^-12	7,9	93,47	66	1,9	1,9	6,2	97,2	**3,5**
M^-13	8,9	93,3	55,2	2	5	13	77,6	**21,1**
M^-14	7,3	94	61,1	4,7	1,9	46	78,8	**6,4**
M^+1	92,7	94	78,8	0,5	6,4	67,2	64,6	**43,3**
M^+2	97,3	95,4	91,4	1,2	3,3	21,5	54,6	**10,7**
M^+3	80,3	92,98	70,5	0,7	6,1	38,8	85,4	**19,4**
M^+4	77,9	94,21	66,3	3,6	2,5	37,9	96	**9,7**
M^+5	99,8	96,2	81,5	2,5	1,1	31,9	86,1	**2,9**
UM^-1	9,8	98,6	82,2	1	0,8	49,4	61	**6,9**
UM^-2	9,8	100	82,7	1,9	3,5	20,7	36,4	**3,4**
UM^-3	7,6	100	72,9	0,4	1,6	19,3		**9,5**
UM^-4	7,5	99,6	75,4	3,1	2,9	16,3	86,2	**8,8**
UM^-5	7,5	100	69,7	0,4	2,9	24,5	90,8	**6,6**
UM^-6	9,8	100	57,7	0,4	2,1	13,2	93,3	**13,9**

SUPPORTING INFORMATION CHAPTER 3 – PART 3

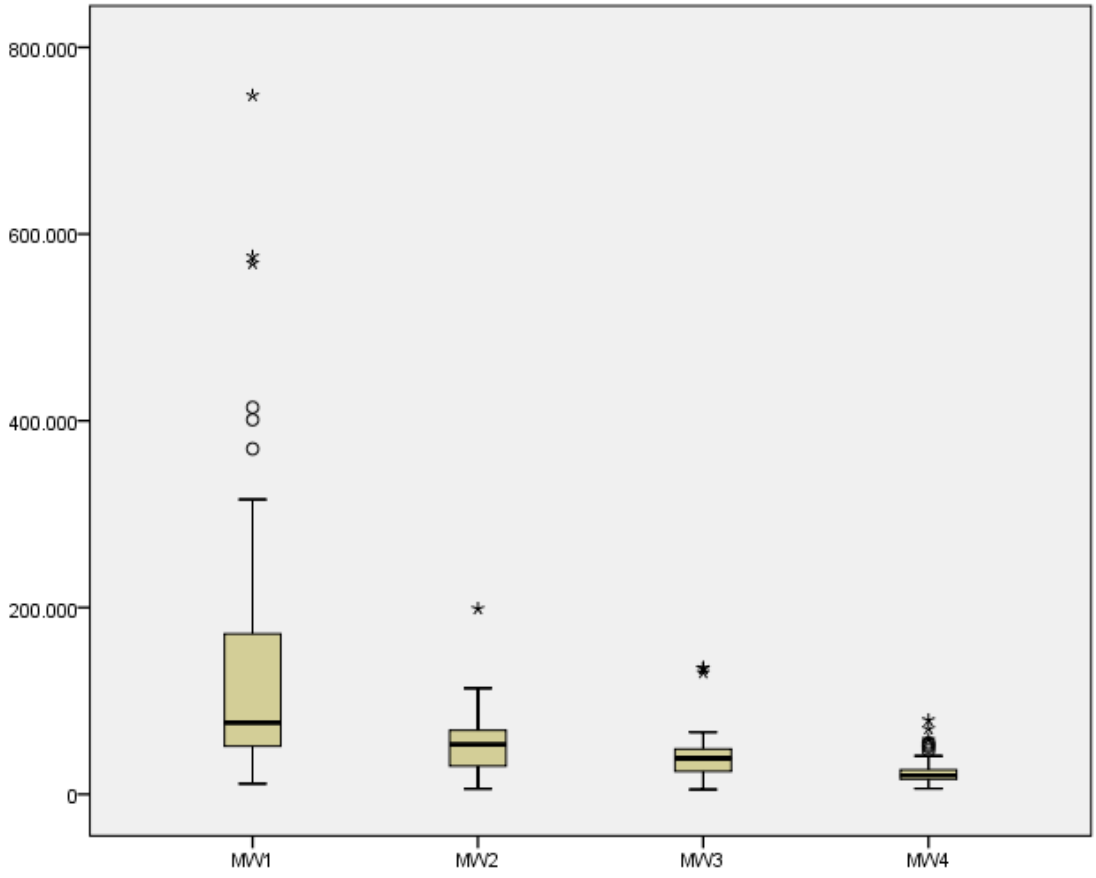
A.3.3. Figure 1. Workflow of R1+R3 analysis

The R1+R3 extract of the EGF-stimulated cells were first gel separated in four fractions, which were each split into four equal parts. The labeled peptides 114-116 and 115-117 were combined. To the 115-117 samples dephosphorylating enzyme cocktail was added and after the inhibition of the phosphatases, all parts were pooled.



A.3.3. Figure 2. Validation of the molecular weight separation

A small fraction of the peptide pool, obtained after the in-gel digestion, was analyzed by a Waters™ Premier mass spectrometer in order to validate the molecular weight gel fractionation. Data was processed and searched against Swissprot (peptide tolerance: ± 0.25 Da) with the Mascot Daemon software. The masses (Da) of the identified proteins from the four molecular weight fractions (MW 1, 2,3 and 4) were exported to SPSS Statistics 20™ (IBM, Armonk, NY, USA) for the creation the Boxblot. Asterisks and circles display outliers.



A.3.3. Table 1A & B: Composition Internal standards

A) IS[1]: Peppermint Stick internal protein standard

Protein	Molecular Weight	Phosphoprotein?
Bèta-Galactosidase*	116.250	no
Bovine serum albumin †	66.200	no
Ovalbumin ‡	45.000	yes
Bèta-Casein §	23.600	yes
Avidin ‡	18.000	no
Lysozyme ‡	14.400	no

* From E.coli. † From bovine serum. ‡ From chicken egg. § From bovine milk.

Source: http://tools.lifetechnologies.com/content/sfs/manuals/mp33350.pdf
Molecular Weight in Dalton.

B) IS[2]: Internal phosphopeptide standard

Initial phosphopeptide *	Unmodified counterpart	Relative signal intensity **
> Phosphopeptide mix 1 (MSP1H)		
VLHSGpS[R]	VLHSGSR	Weak
RSpYpSRS[R]	RSYSRSR	Weak
RDSLGpTYSS[R]	RDSLGTYSSR	Medium
pTKLIpTQLRDA[K]	TKLITQLRDAK	Strong
EVQAEQPSSpSSP[R]	EVQAEQPSSSSPR	Medium
ADEPpSSEESDLEID[K]	ADEPSSEESDLEIDK	Strong
ADEPSpSEEpSDLEID[K]	ADEPSSEESDLEIDK	Medium
FEDEGAGFEESpSETGDYEE[K]	FEDEGAGFEESSETGDYEEK	Strong
ELSNpSPLRENSFGpSPLEF[R]	ELSNSPLRENSFGSPLEFR	Medium
SPTEYHEPVpYANPFYRPTpTPQ[R]	SPTEYHEPVYANPFYRPTTPQR	Strong
> Phosphopeptide mix 2 (MSP2H)		
LPQEpTA[R]	LPQETAR	Weak
RYpSpSRS[R]	RYSSRSR	Weak
EpTQSPEQV[K]	ETQSPEQVK	Weak
VIEDNEpYTA[R]	VIEDNEYTAR	Medium
pSRSPpSSPELNN[K]	SRSPSSPELNNK	Medium
ADEPSSEEpSDLEID[K]	ADEPSSEESDLEIDK	Strong
HQYSDYDpYHSSpSE[K]	HQYSDYDYHSSSEK	Medium
NTPpSQHSHpSIQHSPE[R]	NTPSQHSHSIQHSPER	Medium
ELpSNpSPLRENSFGSPLEF[R]	ELSNSPLRENSFGSPLEFR	Medium
LGPGRPLPTFPpTSE(CAM)TSDVEPDT[R]	LGPGRPLPTFPTSECTSDVEPDTR	Strong
> Phosphopeptide mix 3 (MSP3H)		
SLpSpYpSP[V]ER	SLSYSPVER	Weak
LQGpSGVpS[L]ApSK	LQGSGVSLASK	Medium
PPpYpSRV[I]pTQR	PPYSRVITQR	Strong
pSRS[R]pSYpTPEpYR	SRSRSYTPEYR	Weak
ADEPpSpSEEpSDLE[I]DK	ADEPSSEESDLEIDK	Medium

* Amino acid in [brackets] denotes site of label incorporation for heavy mixes as follows:
[K], $^{13}C_6{}^{15}N_2$ [R], $^{13}C_6{}^{15}N_4$ [V], $^{13}C_5{}^{15}N_1$ [L], $^{13}C_6{}^{15}N_1$ [I], $^{13}C_6{}^{15}N_1$
(CAM) denotes carbamidomethyl cysteine
**As determined using electrospray ionization following standard reverse phase chromatography

Source: http://www.sigmaaldrich.com/content/dam/sigma-aldrich/life-science/mass-spectrometry/phosphomix-composition.jpg

A.3.3. Table 2. Database search parameters

Database search parameters (Mascot Daemon 2.4.0.)

MGF-files from all the samples were searched against the SwissProt Human database.

Constant parameters:

Type of search	:	MS/MS Ion Search
Enzyme	:	Trypsin
Mass values	:	Monoisotopic
Protein mass	:	Unrestricted
Max missed cleavages	:	2

Parameters changed between searches:

For the error tolerance searches

Error tolerance	:	Error tol. search of all significant protein families
Peptide mass tolerance	:	± 10 ppm (# 13C = 1)
Fragment mass tolerance	:	± 0.4 Da

For the quantitation searches

Quantitation	:	iTRAQ 4-plex quantitation reporter tolerance 0.01
Peptide mass tolerance	:	± 15 ppm (# 13C = 1)
Fragment mass tolerance	:	± 0.1 Da

Fixed modifications:

Error tolerance	Quantitation search R1 data	Quantitation search R1+R3 data (gel)	IS^2 phosphomix search 1	IS2 phosphomix search 2
iTRAQ4plex (K)	iTRAQ4plex (K)	iTRAQ4plex (K)	iTRAQ4plex (K)	iTRAQ4plex (K)
iTRAQ4plex (N-term)	iTRAQ4plex (N-term)	iTRAQ4plex (N-term)	iTRAQ4plex (N-term)	iTRAQ4plex (N-term)
Methylthio (C)	Methylthio (C)	Methylthio (C)	Methylthio (C)	

Variable modifications:

Error tolerance	Quantitation search R1 data	Quantitation search R1+R3 data (gel)	IS^2 phosphomix search 1	IS2 phosphomix search 2
iTRAQ4plex (Y)	iTRAQ4plex (Y)	iTRAQ4plex (Y)	iTRAQ4plex (Y)	iTRAQ4plex (Y)
	Deamidated (NQ)	Deamidated (NQ)	Label:13C(5)15N(1) (V)	Label:13C(6)15N(2) (K)
	Oxidation (M)	Oxidation (M)	Label:13C(6)15N(1) (I)	Label:13C(6)15N(4) (R)
	Gln->pyro-Glu (N-term Q)	Gln->pyro-Glu (N-term Q)	Label:13C(6)15N(1) (L)	Carbamidomethyl (C)
	Glu->pyro-Glu (N-term E)	Glu->pyro-Glu (N-term E)	Label:13C(6)15N(4) (R)	Phospho (ST)
	Guanidinyl (N-term)	Carbamyl (N-term)	Label:13C(6)15N(2) (K)	Phospho (Y)
	+100.02 Da (N-term)	Propionamide (C)		

SUPPORTING INFORMATION CHAPTER 3 – PART 4

A.3.4. Experimental details

Validating iTRAQ for optimization: Experimental details :

- The protein mixture digest (Dionex Nr. 161088) was suspended in 80µl of 50mM tri-ethylamonium buffer, pH 7,8 (TEABC) and equally split into four.
- Each sample was labeled with a different label according to the manufacturer's protocol (ABSciex, Framingham, MA, USA):

1. Add 70 µL of ethanol to each iTRAQ™ Reagent vial.
2. Vortex each vial for 1 minute to dissolve the iTRAQ™ Reagent,
then spin.
3. Transfer the contents of one iTRAQ™ Reagent vial to one
sample tube.
4. Vortex each tube to mix, then spin.
5. Incubate the tubes at room temperature for 1 hour.

- Free labels were hydrolyzed with an excess of Milli-Q for 1h at room temperature and dried out in the Speedvac.
- The four differentially labeled peptide mixtures were mixed resolved in 0,1% formic acid and mixed in different ratios according to Table 1.

Samples were purified by C18 zip tips (Sigma) and again dried in the Speedvac.

- For LCMSMS analysis, dried peptides were analyzed on a Dionex U3000,coupled to a Q-TOF Premier (Waters) .
- Dried peptides were dissolved in 0,1% FA, concentrated on a C18 Pepmap (Dionex) with 0,1% formic acid (FA) as loading buffer and subsequently eluted according to a linear gradient wherein the eluting buffer (0,1% FA 80% ACN) increases from 4 to 100% in 70 min.
- LC and MS run details are described in worksheet 4 and 5.

Optimization of the in-gel digestion protocol: Experimental details:

- 16µl of a cell lysate derived from 5 million HepG2 cells was mixed with 5µl Laemmli 4x (Biorad) and reduced with 2µl of BME for 10 min at 95°C before loading onto the gel.
- According to the presented schedule (Figure 2A), the sixteen identical samples were split over two 18 well 10% precast Tris-Glycine polyacrylamide gels together with a Precision Plus Protein Marker (Biorad) at the exterior well.
- Electrophoreses of the two gels was peformed simultaniously for 30 min at 150V and 60 min at 200V in the same Criterion Cell (Biorad).
- The gels were cut around the 50kDa marker to create two molecular weight fractions.
- Individual bands were excised to be digested under the different conditions.
- Proteins were fixed within the gel with a 7% acetic acid, 10% MeOH solution twice for ten minutes.
- After a short wash with Milli-Q, the gel bands were incubated overnight with the SR gel stain in the dark at room temperature.
- For alternative conditions 2 and 3, the fixation (2) or fixation and staining step (3) were skipped.
- All gel pieces were washed 3 x 10 min with a 25mM tri-ethylamonium buffer (TEABC), 50% ACN solution with intermediate shaking.
- he gel pieces were reduced with freshly prepared 50mM tris(2-carboxyethyl)phosphine (TCEP) 25mM TEABC solution for 2h at 56°C.
- After discarding the TCEP, a 200mM S-Methyl thiomethanesulfonate (MMTS) solution was added to the gel pieces for alkylating the proteins in the gel for 1h at room temperature.
- Gel bands were dehydrated with 100% ACN twice for 5 minutes and brought to complete dryness in the Speedvac.
- Modified trypsin (Promega) dissolved in 50mM TEABC, was added in a concentration of 10ng/µl until

all dried gel-pieces were rehydrated.

- For condition 4, 1mM $CaCl_2$ and 5%ACN was added to the trypsin buffer on a fixed and stained gel.
- The digest was performed overnight at 37°C.
- Peptides were extracted in three steps, each time for 30 min, with an increasing amount of 50, 75 and 100% ACN.
- Extracted peptides were dried in the Speedvac and labeling was performed according to the manufactures protocol (details in Experimental Details 1).
- Samples were pooled as presented in the table in Figure 2A, purified by C18 zip tips (Sigma) and again dried in the Speedvac.
- For LCMSMS analysis, dried peptides were analyzed on a Dionex U3000,coupled to a Q-TOF Premier (Waters) .
- Dried peptides were dissolved in 0,1% FA, concentrated on a C18 Pepmap (Dionex) with 0,1% formic acid (FA) as loading buffer and subsequently eluted according to a linear gradient wherein the eluting buffer (0,1% FA 80% ACN) increases from 4 to 100% in 70 min.
- LC and MS run details are described in worksheet 4 and 5.

A.3.4. Table 1

	114/115	116/117	114/117	116/115
Best-fit values				
Slope	0.9643 ± 0.006578	1.016 ± 0.01671	0.9995 ± 0.02254	0.9621 ± 0.009822
Y-intercept when X=0.0	0.02115 ± 0.00372	0.07791 ± 0.009451	0.0270 ± 0.01275	0.03533 ± 0.005555
X-intercept when Y=0.0	0,0219	-0,0767	-0,0270	-0,0367
1/slope	1,0370	0,9845	1,0000	1,0390
95% Confidence Intervals				
Slope	0.9487 to 0.9798	0.9762 to 1.055	0.9462 to 1.053	0.9389 to 0.9853
Y-intercept when X=0.0	0.02995 to -0.0123	0.05556 to 0.1003	0.003153 to 0.0571	0.02220 to 0.04847
X-intercept when Y=0.0	0.01281 to 0.03107	0.09904 to -0.0546	0.05734 to 0.00315	0.05043 to -0.0230
Goodness of Fit				
r^2	0,9997	0,9981	0,9965	0,9993
Sy.x	0,0112	0,0284	0,0383	0,0167
Is slope significantly non-zero?				
F	21490	3695	1966	9594
DFn, DFd	1.000, 7.000	1.000, 7.000	1.000, 7.000	1.000, 7.000
P value	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Deviation from zero?	Significant	Significant	Significant	Significant
Data				
Number of X values	9	9	9	9
Maximum number of Y replicate	1	1	1	1
Total number of values	9	9	9	9
Number of missing values	0	0	0	0

Graph PRISM/prism for Windows/ Version 5.00
Analysis: Linear Regression

A.3.4. Figure 1: Detailed analysis on annotated peptides from Fixed VS non-fixed gel

Detailed analysis of physicochemical properties of a subset of the ratio distribution from annotated peptides. Difference in peptide recovery between a fixed and non-fixed gel.

Objective

The difference in recovery between two digestion conditions is defined by the average ratio of the complete sample. Following peptide or protein annotation, iTRAQ for method optimization also allows for detailed analysis of a subset of the distribution. Here we load the data into the Rover software to determine the distribution of the peptides, based on a calculated peptide Z-score. The Rover software uses robust statistics to compensate for outliers in order to visualize and validate the quantitative iTRAQ data. Applying alternative software tools is possible.

In this template we focus only on the ratios that represent the comparison between the Fix NoSR condition and the condition where NoFix was added. In three of the four samples, the average ratios of the complete sample clearly show that gel fixation significantly increases recovery, as also seen in Figure 2B. Here we validate if a difference could be observed in the physicochemical properties of a subset of the distribution. Peptide subsets (left/right from median/1xSD) were selected based on the Z-scores of the data exported from Rover. We applied an in-house excel based tool to compare the average GRAVY-score, peptide length and miss cleavages.

Methods

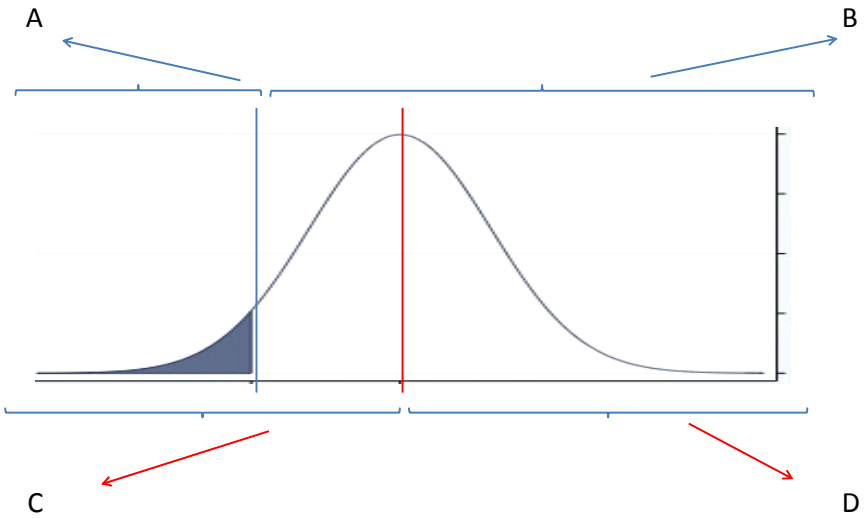
LS-MS analysis, data processing and database search was performed as described in Supporting information Methods 1. Data was validated with the open-source Rover software: http://compomics-rover.googlecode.com/ version 2.3.12. Dat- files obtained from the Mascot DB-search were loaded into Rover according to default settings, also with a SD of 0,14 for the log 1/1 ratio.

Workflow The Rover filter "unique and razor peptides only" was applied and peptides were exported as a csv-file. CSV files of each sample (S1-S4) were imported. Data was sorted on:

S1	Fix NoSR/ NoFix No SR	115/114
S2	NoFix No SR/ Fix NoSR	115/116
S3	NoFix No SR/ Fix NoSR	116/117
S4	NoFix No SR/ Fix NoSR	117/114

Peptide info, based on peptide sequence. Theoretically.	
Information of GRAVY score (J Mol Biol. 1982, PMID: 7108955), miss cleavages and peptide lenght.	
Peptides enriched in FixNoSR supplemened samples.	Threshold 1XSD (Z-score <-1,64 (or >1,64 for S1)
Other peptides in distribution	Threshold 1XSD (Z-score ≥-1,64 (or ≤1,64 for S1)
Peptides enriched in FixNo SR supplemened samples.	One side of mean (Z-score <0 (or > for S1)).
Other peptides	Other side of mean (Z-score ≥0 (or ≤ for S1)).

Final results:



Fix NoSR enriched

A

Threshold: Z-score for each sample < -1.64 (or > 1,64 for S1)

	CONCLUSION	*Result in %*
Total # peptides	*39,00*	
1 miss cleavage	16,00	41,03
2 miss cleavages	0,00	0,00
3 or more miss cleavages	0,00	0,00
Average peptide lenght	13,67	
Average Gravy index	-0,05	

B

Threshold: Z-score for each sample ≥ -1.64 (or ≤ 1,64 for S1)

	CONCLUSION	*Result in %*
Total # peptides	*986,00*	
1 miss cleavage	205,00	20,79
2 miss cleavages	1,00	0,10
3 or more miss cleavages	0,00	0,00
Average peptide lenght	11,73	
Average Gravy index	0,05	

C

Threshold: Z-score for each sample < 0 (or > 0 for S1)

	CONCLUSION	Result in %
Total # peptides	552,00	
1 miss cleavage	132,00	23,91
2 miss cleavages	1,00	0,18
3 or more miss cleavages	0,00	0,00
Average peptide lenght	11,94	
Average Gravy index	0,04	

D

Threshold: Z-score for each sample ≥ 0 (or ≤ 0 for S1)

	CONCLUSION	*Result in %*
Total # peptides	*473,00*	
1 miss cleavage	88,00	18,60
2 miss cleavages	0,00	0,00
3 or more miss cleavages	0,00	0,00
Average peptide lenght	11,65	
Average Gravy index	0,06	

Conclusion:

When looking at annotated peptides one can select certain populations of peptides within the ratio distribution (e.g. left and right of the mean).

When these subsets are compared based on certain properties as the number of miss cleavages and GRAVY scores, one can coordinately define the gain of a certain methodology for specific characteristic sequences. Although more peptides were obtained from the digest, the analysis demonstrated that gel fixation results in an increased number of miss cleavages. We thus hypothesize that skipping fixation could result in spontaneous migration of proteins and peptides out of the gel.

A.3.4. Figure 2: Detailed analysis of annotated peptides from the digestion with and without the addition of CaCl2 and ACN

Detailed analysis of physicochemical properties of a subset of the ratio distribution from annotated peptides. Difference in peptide recovery between the digestion protocols with and without the addition of CaCl2 and ACN.

Objective

The difference in recovery between two digestion conditions is defined by the average ratio of the complete sample. Following peptide or protein annotation, iTRAQ for method optimization also allows for detailed analysis of a subset of the distribution. Here we load the data into the Rover software to determine the distribution of the peptides, based on a calculated peptide Z-score. The Rover software uses robust statistics to compensate for outliers in order to visualize and validate the quantitative iTRAQ data. Applying alternative software tools is possible.

In this example we focus only on the ratios that represent the comparison between the standard condition and the condition where CaCl2 and ACN are added. In all four samples, the average ratios of the complete sample clearly show that supplementing the trypsin buffer with CaCl2 and ACN significantly increases recovery, as also seen in Figure 2B. Here we validate if a difference could be observed in the physicochemical properties of a subset of the distribution. Peptide subsets (left/right from median/1xSD) were selected based on the Z-scores of the data exported from Rover. We applied an in-house excel based tool to compare the average GRAVY-score, peptide length and miss cleavages.

Methods

LS-MS analysis, data processing and database search was performed as described in Supporting information Methods 1.Data was validated with the open-source Rover software: http://compomics-rover.googlecode.com/ version 2.3.12. Dat- files obtained from the Mascot DB-search were loaded into Rover according to default settings, also with a SD of 0,14 for the log 1/1 ratio.

Workflow

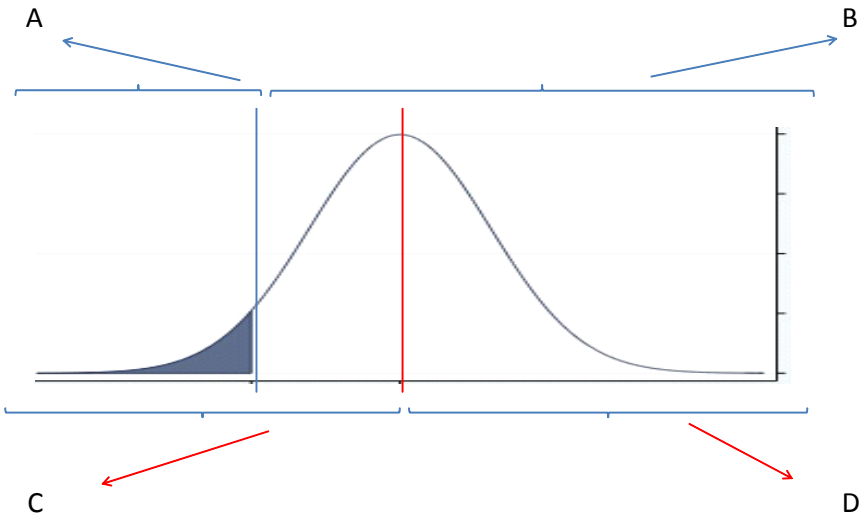
The Rover filter "unique and razor peptides only" was applied and peptides were exported as a csv-file. CSV files of each sample (S1-S4) were imported. Data was sorted on:

S1	Fix SR/ Fix SR CaCl2 ACN
S2	Fix SR/ Fix SR CaCl2 ACN
S3	Fix SR CaCl2 ACN/ Fix SR
S4	Fix SR/ Fix SR CaCl2 ACN

The Z-scores of the ratios were ranked according to enrichment in the samples supplemented with CaCl2 ACN. S1: low-high/ S2: low-high/ S3: high-low/ S4: low-high

Peptide info, based on peptide sequence. Theoretically.	
Information of GRAVY score (J Mol Biol. 1982, PMID: 7108955), miss cleavages and peptide lenght.	
Peptides enriched in $CaCl_2$ACN supplemened samples.	Threshold 1XSD (Z-score <-1,64 (or >1,64 for S3)).
Other peptides in distribution	Threshold 1XSD (Z-score ≥-1,64 (or ≤1,64 for S3)).
Peptides enriched in $CaCl_2$ACN supplemened samples.	One side of mean (Z-score <0 (or > for S3)).
Other peptides	Other side of mean (Z-score ≥0 (or ≤ for S3)).

Final results:



CaCl2 ACN enriched

A

Threshold: Z-score for each sample < -1.64 (or > 1,64 for S3)

	CONCLUSION	*Result in %*
Total # peptides	*68*	
1 miss cleavage	12	17,65
2 miss cleavages	0	0,00
3 or more miss cleavages	0	0,00
Average peptide lenght	11,24	
Average Gravy index	-0,11	

B

Threshold: Z-score for each sample ≥ -1.64 (or ≤ 1,64 for S3)

	CONCLUSION	*Result in %*
Total # peptides	*957*	
1 miss cleavage	209	21,84
2 miss cleavages	1	0,10
3 or more miss cleavages	0	0,00
Average peptide lenght	11,85	
Average Gravy index	0,06	

C

Threshold: Z-score for each sample < 0 (or > 0 for S3)

	CONCLUSION	
Total # peptides	514	
1 miss cleavage	87	16,92607
2 miss cleavages	0	0
3 or more miss cleavages	0	0
Average peptide lenght	11,45	
Average Gravy index	0,04	

D

Threshold: Z-score for each sample ≥ 0 (or ≤ 0 for S3)

	CONCLUSION	*Result in %*
Total # peptides	*511*	
1 miss cleavage	134	26,22
2 miss cleavages	1	0,20
3 or more miss cleavages	0	0,00
Average peptide lenght	12,17	
Average Gravy index	0,06	

Conclusion:

CaCl2 and ACN supplementation does not result in recovery of more hydrophobic peptides as seen by the similar GRAVY-scores but does diminish the amount of trypsin miss cleavages.

Curriculum vitae

General information

Name:	Pieter Glibert
Address:	Maquisstraat 89
	Ronse, Belgium
Date of birth:	08/07/1985
Place of birth:	Ronse
Civil status:	Married

Education

2009-present:	PhD candidate for a PhD in Pharmaceutical Sciences; Ghent University.
2008-2009:	Master in Industrial Pharmacy; Ghent, Leuven, Brussels and Antwerp University. Diploma obtained with distinction.
2003-2008:	Master in Pharmaceutical Sciences (Pharmacist); Ghent University. Diploma obtained with distinction.
1997-2003:	Mathematics-Sciences; KSO Glorieux Ronse.

Scientific Curriculum

1. Publications *(A1)*

- Phospho-IiTRAQ, "Assessing Isobaric Labels For The Large-Scale Study Of Phosphopeptide Stoichiometry".
  Glibert P, Meert P, Van Steendam K, Van Nieuwerburgh F, Martens L, Dhaenens M and Deforce D.
  Submitted to Journal Of Proteome Research (Under review; 2014 Aug).
  IF: 5.001. JCR rank: 9/78 in Biochemical Research Methods category.

- Histone Proteolysis: A proposal for categorization into 'Clipping' and 'Degradation'.
  Dhaenens M, **Glibert P**, Meert P, Vossaert L, Deforce D.
  BioEssays. 2014 Oct 28. doi: 10.1002/bies.201400118.
  IF: 4.927. JCR rank: 11/85 in Biology category.

- Identification of histone H3 clipping activity in human embryonic stem cells.
  Vossaert L, Meert P, Scheerlinck E, **Glibert P**, Van Roy N, Heindryckx B, De Sutter P, Dhaenens M, Deforce D.
  Stem Cell Res. 2014 Jul;13(1):123-34.
  IF: 3.912. JCR rank: 75/185 in Cell Biology category.

- Quantitative proteomics to characterize specific histone H2A proteolysis in chronic lymphocytic leukemia and the myeloid THP-1 cell line.
  Glibert P, Vossaert L, Van Steendam K, Lambrecht S, Van Nieuwerburgh F, Offner F, Kipps T, Dhaenens M, Deforce D.
  Int J Mol Sci. 2014 May 27;15(6):9407-21.
  IF: 2.339. JCR rank: 52/148 in Chemistry, Multidisciplinary category.

- Neutrophil Elastase in the capacity of the "H2A-specific protease".
  Dhaenens M*****, **Glibert P***, Lambrecht S, Vossaert L, Van Steendam K, Elewaut D, Deforce D.
  (*****Equal contribution) Int J Biochem Cell Biol. 2014 Jun;51:39-44.
  IF: 4.240. JCR rank: 75/291 in Biochemistry and Molecular Biology category.

- iTRAQ as a method for optimization: enhancing peptide recovery after gel fractionation.
  Glibert P, Van Steendam K, Dhaenens M, Deforce D.
  Proteomics. 2014 Mar;14(6):680-4.
  IF: 3.973. JCR rank: 14/78 in Biochemical Research Methods category.

- Immunization with the immunodominant Helicobacter suis urease subunit B induces partial protection against H. suis infection in a mouse model.
  Vermoote M, Van Steendam K, Flahou B, Smet A, Pasmans F, **Glibert P**, Ducatelle R, Deforce D, Haesebrouck F.
  Vet Res. 2012 Oct 26;43:72.
  IF: 3.383. JCR rank: 1/132 in Veterinary Sciences category.

2. Participations in international conferences

- Poster Presentations at the Biochemical Society Conference “Quantitative Proteomics and Data Analysis”, 24 – 25/03/2014, Chester, UK (not attended).
  - PiTRAQ, extending a differential labeling and dephosphorylation approach to the readily detectable proteome.
    Glibert P, Martens L, Deforce D, Dhaenens M*. Poster award winner. (*Presenting author)
  - Sensitive quantification of proteolytic products by use of aqua peptides.
    Glibert P, Vossaert L*, Dhaenens M, Deforce D. (*Presenting author)
  - iTRAQ as a method for optimization: enhancing peptide recovery after gel fractionation. **Glibert P**, Meert P*, Van Steendam K, Dhaenens M, Deforce D. (*Presenting author)

- Poster Presentation at the 61st ASMS (American Society for Mass Spectrometry) “Conference on Mass Spectrometry and Allied Topics”, 09 - 13/06/2013, Minneapolis, USA. PiTRAQ: a strategy to simultaneously correlate protein expression and phosphorylation stoichiometry between different samples: evaluation on different mass spectrometers.
  Glibert P, Dhaenens M, Meert P, Colaert N, Van Nieuwerburgh F, Martens L, Deforce D.

- Poster Presentation at the Seventh General Meeting of the International Proteolysis Society, 16 - 20/10/2011, San Diego, USA.
  Neutrophil Elastase is the “histone h2a-specific protease”.
  Glibert P, Dhaenens D, Lambrecht S, Deforce D.

3. Participations in national conferences

- Oral presentation at the AB SCIEX, PEAK Scientific LC/MS/MS Laboratory, 14/05/2014, Technologiepark Zwijnaarde. Adopting the isobaric iTRAQ labels for the specific quantitation of the phosphorylation stoichiometry and the large-scale phosphorylation mapping.
  Glibert P.

- Poster presentation at “Knowledge for Growth”, FlandersBio, 08/05/2014, Gent. Mass spectrometry based proteomics: Status - Emerging technologies - Challenges of making the jump

from bench side to bedside.
Glibert P, Deforce D.

- Poster presentation at “Knowledge for Growth”, FlandersBio, 30/05/2013, Gent. Simultaneously Quantify Protein Expression And Post-Translational Modifications.
  Glibert P, Dhaenens M, Meert P, Colaert N, Van Nieuwerburgh F, Martens L, Deforce D.

- Poster presentation at “Belgian Proteomics Association”, 29 & 30/11/2012, Gent. Evaluation of a Strategy to Identify Protein Phosphorylation in an unbiased way.
  Glibert P, Dhaenens M, Meert P, Colaert N, Martens L, Deforce D.

4. Attendance of international meetings, conferences and workshops

- ASMS Short Course: Bioinformatics for Protein Identification, 08-09/06/2013, Minneapolis, USA.
- Thermo Scientific Q Exactive Demo, 07 & 08/05/2013, Basel, Swiss.
- AB Sciex TripleTof 5600 Demo, 03 & 04/04/2013, Darmstadt, Germany.
- ASMS Fall Workshop: Phosphorylation Analysis and Phosphoproteomics, 08 & 09/11/2012, Boston, USA.
- Waters Cooperation Synapt G2Si Demo, 28-30/08/2012, Manchester, UK.
- Mascot Training Course, 26 & 27/04/2010, Hannover, Germany.

5. Attendance of national meetings, conferences and workshops

- The Epigenetic Revolution, 1,2 &4/07/2013, Leuven.
- BITS - Mass spectrometry data processing, 19/11/2012, Ghent.
- Waters "MS Technology Days”, 11/10/2012, Brussels.
- BITS - Introduction to Bioinformatics, 2010, Leuven.
- KVCV Flanders Proteomics, 16 & 17/12/2010, Antwerp.
- Introduction course Flowcytometry, 30/03/2010, Erembodegem.

Educational experience

- Supervisor of practical courses in Pharmacognosy and Phytochemistry
- Supervisor of practical courses in Biotechnology
- Supervisor of student theses:
  - 2011-2012: Leen Feys (1st master Pharmaceutical Sciences; UGent)
    Optimalisatie van een kwantitatieve massaspectrometrische techniek in de fosfoproteomics ter bepaling van de activiteit van het US3-kinase
  - 2011-2012: Elien Vanderschaeve (Master of Science in Industrial Pharmacy; UGent). GlyTRAQ: ontwikkeling en optimalisatie van een massaspectrometrische techniek ter kwantificatie van glycosylaties
  - 2011-2012: Benny Vlerick (Bachelor Biochemistry; HoGent)
    Validatie van een nieuwe massaspectrometrisch-gebaseerde techniek voor de kwantificatie van gefosforyleerde eiwitten

- 2012-2013: Erika D'haenens (Master Industrial Sciences, Biochemistry; HoGent)
- 2012-2013: Lieselotte Paridaens (Master of Science in Industrial Pharmacy; UGent)
- 2012-2013: Delphine Jonckheere (1st master Pharmaceutical Sciences; UGent). Phospho-iTRAQ: optimalisatie van een kwantitatieve massaspectrometrische techniek in de fosfoproteomics

Dankwoord

Als oudste telg uit een geslacht van apothekers was Farmacie een evidente studiekeuze. Een paar goede scores, jeugdige ambitie, een drang naar een buitenlands avontuur en de nieuwsgierigheid naar de wereld buiten de vier muren van een apotheek deed me een jaartje verder studeren. Dankzij de eerste opdracht, het schrijven van een tweede masterthesis, werd ik geïntroduceerd in het laboratorium farmaceutische biotechnologie alwaar ik voor het eerst “wetenschappelijke onderzoek” aan den lijve mocht ervaren. Doctoreren stond niet op mijn to-do list maar het onderzoeksproject was ambitieus en het polsen naar een mogelijke carrièreswitch van zowel promotor en begeleider was flaterend. Waarom zou ik mij niet eens vijf jaar toeleggen op het onderzoek? “In het leven moet je immers kansen grijpen wanneer die zich voordoen.” YOLO! En zo ging ik na een maand vakantie opnieuw naar het FFW. Deze maal als doctorandus. Vandaag, vijf jaar later, na hard werken met bijhorende ontspanning, na momenten van frustratie en momenten van succes, eindigt dit academisch avontuur. Tijd om afscheid te nemen en een paar mensen te bedanken.

Prof. Dr. Apr. Deforce, u gaf mij de mogelijkheid om te doctoreren en stelde de middelen ter beschikking om dit onderzoek te financieren. Bedankt Dieter voor het vertrouwen en de gegunde vrijheid. Bedankt voor het verbeteren van de manuscripten en ons, dankzij jouw analytisch vermogen, de juiste richting uit te sturen. Dankzij de kansen die jij mij gaf heb ik mij de afgelopen vijf jaar in een inspirerende omgeving verder kunnen ontwikkelen, ook op niet-wetenschappelijk gebied.

Maarten, thesis begeleider, collega, copromotor. Tijdens mijn wetenschappelijk rijpingsproces ben jij een belangrijke katalysator geweest en ik kan niet anders dan u mijn mentor noemen. Bedankt om je kennis geduldig door te geven en mij te besmetten met de passie voor de massa. In een buitenland, na middernacht een bar afsluiten en nog steeds discussiëren over de data om daarna vroeg op te staan en de strever uit te hangen…het zijn momenten die onze samenwerking hebben versterkt en de onderzoeksresultaten steeds een boost gaven. De synergie tussen ons beider enthousiasme, jouw "disruptive thinking" in combinatie met mijn pragmatisme, het harde werk… elementen die noodzakelijk waren om te volharden in de H2A-queeste. It was a cool ride.

Verder wil ik ook mijn dank betuigen aan de overige leden van de lees- en examencommissie. Hartelijk dank om mijn doctoraat te evalueren en een constructieve bijdrage te leveren.

Als doctorandus werk je in een team met jonge, gemotiveerde en verstandige collega’s. Deze begrijpen, beter dan wie, de frustraties en momenten van vreugde die eigen zijn aan het academisch

onderzoek. Ik had dan ook graag de ondersteunende oude garde en “mes compagnons de route” bedankt.

Kelly, wij mochten maar enkele maanden onder jouw vleugels vertoeven maar jij was die sterke, rechtlijnige postdoc die ons, nog groen achter de oren, gemeend ondersteunde. Jip, jij gaf, aan het kleinste bureautje van het FFW, een nieuwe definitie aan het begrip werkethiek en stond steeds klaar om mij te assisteren met het praktisch werk. Filip, jij gaf als ancien geregeld jouw ongezouten mening over het reilen en zeilen van het labo. Reality checks die soms eens nuttig waren, en steeds, zoals jouw vele anekdotes na en tussen de werkuren, zeer entertainend. Mado, met jou kon ik het al van toen ik nog thesisstudent was goed vinden. Ik ben steeds gefascineerd gebleven door jouw beredeneerde manier van werken.

Katleen, wij waren elkaar al meermaals tegengekomen dankzij onze Ronsische roots en ik ben blij dat ik jou beter heb leren kennen op het FBT waar jij nu schittert als postdoc. Bedankt voor jouw goede raad, luisterend oor en assistentie bij het opstellen van de artikels. Marlies, je bent slecht één jaartje ouder maar was (zeker de eerste jaren), zoveel matuurder. Bedankt om mij te introduceren in de wonderlijke wereld van het proteoom! Sandra, mijn overbuur, jij stond steeds klaar om voor ons te duimen. Proficiat met je zoon. Trees, bedankt om altijd voor iedereen klaar te staan en af en toe een verslagje over te nemen zodat ik aan de massa kon prutsen.

En zo beland ik automatisch bij mijn ware “compagnons de route”: Liesbeth en Veerle. Wij startten op dezelfde dag aan ons doctoraat, wij doorstonden de koude in de practicumzaal, wij delen een IWT-trauma maar vooral heel veel momenten van plezier. Liesbeth, het H3-complement van de histone handshake en gedurende vijf jaar mijn buur, bedankt voor al de fijne gesprekken en het vele lachen. Veerle, jij was steeds een baken van rust en een vertrouwd klankbord. Bedankt. Liesbeth haar lach, Veerle haar niezen en het geluid van een vacuümpomp,… het zijn de drie geluiden die ik het meeste zal missen.

Na een jaar op het labo mochten wij het “nieuwkes” brandmerk doorgeven. Bedankt Ellen, om gemotiveerd de TripleTof optimalisatie verder voor jouw rekening te nemen. Shaid, glad you're doing well these days. Paulien, bedankt voor de leuke gesprekken en jouw nuchtere kijk. Christophe, mijn gewezen looppartner, op jou gezondheid hef ik mijn volgende streekbiertje. Bart, als jij niet op “de Diergeneeskunde” was kon ik de extra man in ons bureau wel appreciëren zodat aangebrande moppen weer mogelijk waren. Elizabeth, het was steeds leuk om jou al pendelend tegen te komen. Eindelijk iemand die histon extracties serieus neemt. Lieselot, respect om het meer dan een jaar vol te houden daar beneden. Bert, Yannick en Dieter, dank voor jullie praktische hulp bij de technische IT-problemen.

Laura en Senne, het nieuwe talent, succes met jullie doctoraat. Bedankt iedereen voor de leuke momenten!

Sofie, bedankt voor al die keren dat jij de naald of tubing hebt vervangen die door onze stalen verstopten. Bedankt om mij te begeleiden bij de mass spec en LC problemsolving.

Ik had uiteraard ook graag de collega's van de routine bedankt. Met velen had ik een uitstekend contact en regelmatig leuke gesprekjes in de wandelgangen. Een speciaal woord van dank gaat naar Astrid, Nadine en Inge. Jullie zijn de motor van het labo en hebben tot de laatste dag een waaier aan praktische zaken voor mij geregeld.

Verder had ik ook graag de wetenschappers uit andere onderzoeksgroepen bedankt (Miet, Stijn, Céline, Ut, Bonk,…) waarmee ik mocht samenwerken en zo mijn horizon kon verbreden. Thesisstudenten, jullie hebben niet alleen het praktische werk verlicht maar mij ook scherp gehouden. Bedankt: Jelena (although not really my thesis student), Leen, Benny, Elien, Lieselotte, Erika en Delphine.

De boog kan niet altijd gespannen staan. Bedankt vrienden uit Ronse voor de talrijke momenten van afleiding. Ik werd wel eens vergeleken met "Sheldon" als ik mij teveel liet gaan maar jullie hebben mij steeds doen relativeren en me er op gewezen dat het in "de reguliere business" ook niet alles is. Bedankt vrienden van de FOB (+vrouwen). Het is geweldig dat wij, na vijf jaar Farmacie, nog op regelmatige basis afspreken om de teugels andermaal los te laten. Bedankt aan die andere vrienden (de architecten, de aanhangsels van Lien haar vriendinnen, de duikers,…) om mijn gedachten van tijd tot tijd te verzetten.

Familie! Bedankt grootouders. Opa, jij was altijd de grootste voorstander van "mijn doctoreren". Mijn nieuwsgierigheid vindt ongetwijfeld zijn oorsprong in de plakboeken met fauna- en floraprentjes die jij voor mij samenstelde en de uitzendingen van "le jardin extraordinaire" die we samen keken. Bedankt mama en papa voor alles. Wij hebben veel aan jullie te danken. Studeren werd bij ons gestimuleerd (eufemisme), waarvoor dank. Na tien jaar op deze universiteit zoek ik nu wel met plezier andere oorden op! Nele en Maarten, ik ben trots dat jullie ook verder specialiseren. Slim dat jullie voor deze specialisatie wél voor een echt beroep hebben gekozen. Jaantje, je bent de jongste en moet je nog bewijzen. Wat je ook doet, komt wel goed! Ik had dit forum voor de verandering eens voor een praktisch doeleinde willen benutten. Nele, met drie broers is de concurrentie groot en omdat ik elk concurrentieel voordeel moet aangrijpen had ik via deze weg willen vragen of ik jouw getuige mag zijn? Dank! Schoonfamilie (Mireille, Phillip, Liesbeth, Ruben en Odette), jullie empathie en steun is fantastisch.

Om te eindigen wil ik nog één persoon bedanken. De belangrijkste. Een mislukt experiment, een onstabiele massaspectrometer, onvrede over de lab politics, enthousiasme over een nieuwe hypothese,… Liefje, jij bent degene bij wie ik steeds terecht kon. Even uitrazen was vaak voldoende om onze avond en de volgende dag zonder stress verder te zetten. Bedankt om mij en mijn keuzes steeds met een lach aan te moedigen. Ik ben trots dat ik mij jouw man mag noemen!


UNIVERSITEIT
GENT