# Development and Application of Novel Electron Transfer Dissociation-based Technologies for Proteomics

Christian K. Frese

dedicated to my parents
Ursula & Karl

ISBN 978-90-8891-661-8

The research in this thesis was performed in the Biomolecular Mass Spectrometry & Proteomics Group, Utrecht University, Utrecht, The Netherlands

The research was financially supported by the Netherlands Proteomics Center (NPC)

Printing of this thesis was supported by Thermo Fisher Scientific and the Utrecht Institute for Pharmaceutical Sciences (UIPS)

Cover artwork by Silvia Dierkes

Printed by proefschriftmaken.nl || Uitgeverij BOXPress

# Development and Application of Novel Electron Transfer Dissociation-based Technologies for Proteomics

Ontwikkeling en gebruik van nieuwe peptide fragmentatie technieken in proteomics (met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op vrijdag 6 september 2013 des middags te 12.45 uur

door

Christian Karl Frese

geboren op 21 december 1983, te Marsberg, Duitsland

**Promotor** Prof. dr. A.J.R. Heck

**Co-promotor** Dr. A.F.M. Altelaar

"Die meisten Menschen wollen nicht eher schwimmen als bis sie es können."

Ist das nicht witzig?
Natürlich wollen sie nicht schwimmen! Sie sind ja für den Boden geboren, nicht fürs Wasser.
Und natürlich wollen sie nicht denken; sie sind ja fürs Leben geboren, nicht fürs Denken!

Ja, und wer denkt, wer das Denken zur Hauptsache macht, der kann es darin zwar weit bringen,
aber er hat doch eben den Boden mit dem Wasser vertauscht, und einmal wird er ersaufen.

Der Steppenwolf, *Hermann Hesse*

# Table of contents

Chapter 1
**Introduction** **9**

Chapter 2
**Improved Peptide Identification by Targeted Fragmentation Using CID, HCD and ETD on a LTQ-Orbitrap Velos** **53**

Chapter 3
**Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry** **79**

Chapter 4
**Unambiguous Phosphosite Localization Using Electron-Transfer / Higher-Energy Collision Dissociation (EThcD)** **95**

Chapter 5
**Characterization of Electron-Transfer Dissociation in the Orbitrap Velos HCD Cell** **109**

Chapter 6
**Benchmarking Stable Isotope Labeling Based Quantitative Proteomics** **127**

Chapter 7
**Profiling of Diet-Induced Neuropeptide Changes in Rat Brain by Quantitative Mass Spectrometry** **155**

Chapter 8
**Summary, Samenvatting, Future Outlook, Publications, Curriculum Vitae, Acknowledgements, Abbreviations** **179**

Chapter 1

# Introduction

in parts taken from

Christian K. Frese[1,2,#], Henk van den Toorn[1,2,#],
Albert J.R. Heck[1,2] and Shabaz Mohammed[1,2]

Quantitative Proteomics for Differential Protein Expression Profiling
in
Proteomics for Biological Discovery
2nd Edition, WILEY-VCH, in press

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

[#] These authors contributed equally to this work

## 1 Mass Spectrometry-based Proteomics

Proteins are key actors in all cellular processes and pathways. Moreover, almost all diseases are linked to perturbations of proteins, their modification state, structure and interaction networks. Consequently, the fundamental study of proteins is underlying most biological questions. The term proteome was introduced in 1996 and describes all proteins, including their modifications, that are expressed by a certain cell, tissue or organism at a given time [1]. In analogy to genomics, the field of protein sciences was later termed proteomics [2]. In contrast to the genome, which is believed to consists of 20,000-30,000 genes, the proteome is highly dynamic. Not all genes are expressed at equal levels at all times. Gene expression is controlled and influenced by cell state and extracellular stimuli. Moreover, the proteome complexity is further increased by splicing variants and post-translational modifications. In the past, proteins were usually individually studied by well-established biochemical techniques such as immunoassays, two-hybrid assays and antibody or gel-based methods. However, analysis of the entire proteome is desired to understand the complex regulatory mechanisms underlying cellular processes. Mass spectrometry (MS) has become a powerful tool in life science research and is nowadays an integral analytical method in proteomics [3-5]. MS-based proteomics has matured to a high-throughput quantitative technology, aiming to provide sensitive, accurate and complete information on protein abundance, interactions and networks to a similar depth as currently possible in genome and transcriptome sequencing [6-8].

In the standard approach in MS-based proteomics, the sample is lysed and proteins are extracted followed by proteolytic cleavage into peptides. The peptides are often fractionated to reduce sample complexity or to enrich for certain subsets of peptides, e.g. phosphorylated or acetylated peptides. Peptide identification is usually performed by reversed-phase liquid chromatography coupled to mass spectrometry. This approach enables simultaneous acquisition of qualitative and quantitative information about the peptides in the sample. The first MS stage measures the accurate mass of the peptides. In the seconds stage, referred to as tandem mass spectrometry (MS/MS), peptides are fragmented and the masses of the resulting fragment ions are detected. Peptide identification is conducted by automated database search algorithms that match peptide and fragment ion masses to theoretically computed peptides. Identified peptides are mapped to the corresponding proteins followed by statistical analysis and validation.

In the following, principles of MS-based peptide sequencing are introduced. Moreover, an overview on theory and recent developments of the Orbitrap, an orbital trapping mass analyzer, is given. The second part of the introduction focuses on quantification approaches in MS-based proteomics.

## 1.1 Peptide Sequencing Technologies

Mass spectrometry allows for determination of the molecular weight of peptides and proteins. Tandem mass spectrometry (MS/MS) is applied to obtain more information about structure, composition, amino acid sequence and modifications of a selected precursor [9]. Key to any MS-based proteomics experiment is peptide sequencing by means of gas phase dissociation [10]. Several fragmentation techniques have been developed, all of which can induce cleavage of at least one of the three peptide backbone bonds: the C-N peptide bond, the Cα-C bond and the N-Cα bond (Figure 1.1). The most commonly utilized fragmentation techniques in proteomics are based on (i) collisions with neutral gas molecules and (ii) low-energy electron capture or transfer.

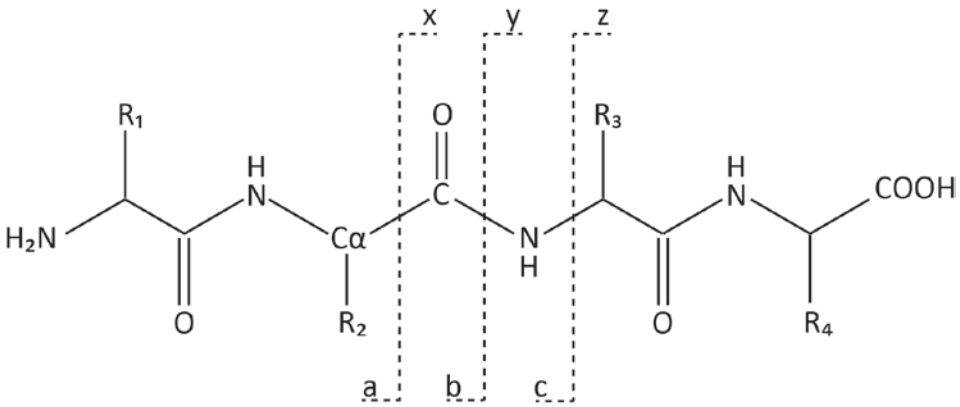



**Figure 1.1.** Activation of peptide cations induces cleavage of different bonds of the peptide backbone, yielding N-terminal a-, b- and c-ions and C-terminal x-, y- and z-ions (nomenclature according to [11])

### 1.1.1 Collision Induced Dissociation

Collision induced dissociation (CID) is the most frequently employed peptide fragmentation technique in current MS-based proteomics [12]. In CID, selected peptides are subjected to collisions with neutral gas molecules such as Helium, Nitrogen or Xenon [13]. Upon impact of the gas molecules, parts of the kinetic energy are converted into vibrational energy which ultimately leads to dissociation of certain bonds. CID, employing energies between 1-100 eV on the millisecond timescale, is usually considered a "slow-heating" method. It can be further classified into resonant-excitation and beam-type CID.

Resonant-excitation CID is utilized in ion trap mass spectrometers and, thus, is also referred to as ion trap-CID or "tandem in time". After trapping of selected precursor ions, an AC waveform, corresponding to the secular resonance of the ions, is applied to the end caps

or lenses. This increases the kinetic energy of the ions, resulting in collisions with the bath gas in the trap. The internal vibrational energy increases through multiple collisions over a few milliseconds. The internal energy is redistributed throughout all bonds of the peptide. Consequently, the dissociation is usually channeled through the lowest energy pathway, resulting in cleavage of the weakest bond. For peptides, this is usually the C-N peptide bond (Figure 1.1), yielding b- and y-type fragment ions. The fragmentation pathways in CID can be explained with the mobile proton model [14]. According to this model, charge-directed peptide bond cleavage requires a proton at the cleavage site. In non-excited states protons are sequestered by basic groups of amino acid side chains. Upon excitation, i.e. an increase of internal energy, the localized protons can move intramolecularly and reside at heteroatoms of the peptide backbone. Migration to an amide bond nitrogen atom weakens the C-N bond and generates b- and y-ions as the internal energy increases during activation [15]. The energy that is necessary to mobilize a proton depends on the amino acid composition and gas-phase basicity. It should be noted, that the mobile protons do not reside stochastically at the different amide bonds throughout the peptide backbone, i.e. not all peptide bond cleavages are observed in CID. The group of Wysocki reported that the cleavage pattern depends (i) on the peptide charge state, (ii) on the position of the most basic residue and (iii) on the presence of proline residues [16]. Depending on peptide structure, charge state and the amino acid composition, ions corresponding to the loss of ammonia or water, internal fragments or a-ions can also be observed in CID MS/MS spectra. A known limitation of ion trap-CID is the loss of ions in the low mass region [17]. Due to the waveform activation ions below 30% of the precursor m/z are rarely observed.

In beam-type CID, peptide fragmentation occurs via "tandem in space" activation. This mode of operation is typically employed in QqQ and Q-ToF instruments, but is also available on hybrid instruments through dedicated collision cells [18]. Ions are typically selected in a quadrupole mass filter and accelerated into a collision cell that is filled with the collision gas. Compared to resonant-excitation CID, the energy applied is slightly higher (usually 20-100 eV) and energy deposition occurs on a shorter timescale (< 1 ms). Together, this implicates that bond cleavage usually occurs prior energy redistribution, opening up alternative fragmentation channels. For instance, an advantage of beam-type CID is the less pronounced observation of neutral loss of the phosphate group of phosphorylated peptides, which hampers peptide identification and phosphorylation site localization in ion trap-CID [19].

### 1.1.2 Electron Transfer Dissociation

Over the last decade, the development of electron capture dissociation (ECD) [20] and electron transfer dissociation (ETD) [21] have greatly advanced the capabilities of MS-based peptide and protein sequencing. Both are electron-driven dissociation techniques and are based on the generation of odd-electron peptide cations, inducing peptide backbone cleavage of the N-Cα bond (Figure 1.2).

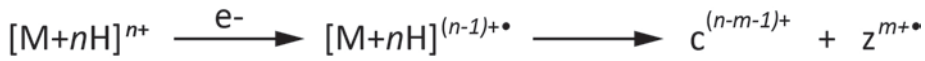



**Figure 1.2.** Generation of c/z-ions in electron driven dissociation techniques (n > 1; m < n).

In ECD, trapped peptide cations capture one or more near-thermal electrons, resulting in the formation of an odd-electron species. This radical intermediate molecule undergoes rapid and nonergodic dissociation [20]. This means that, in contrast to slow-heating collisional activation methods, the dissociation process does not involve complete intramolecular energy redistribution. In 2004, Hunt and co-workers utilized ion/ion-chemistry to implement a similar fragmentation technique termed electron transfer dissociation [21]. They used a negative chemical ionization source to generate anions (e.g. from Fluoranthene or Azulene) with low electron affinities. Electron transfer is induced upon mixing with peptide cations in a linear quadrupole trap. Both ECD and ETD generate primarily c- and z-ions and labile post-translational modifications remain intact [21].

McLafferty and co-workers initially suggested the Cornell mechanism to explain the fragmentation underlying electron capture and electron transfer dissociation (Figure 1.3A). They proposed electron capture into a Rydberg orbital at a positively charged site such as a protonated amine group of the N-terminus or any basic residue side chain (Arg, Lys, His) [20]. Relaxation into the ground state induces hydrogen atom transfer to an amide oxygen, forming an aminoketyl radical intermediate. Subsequently, this results in cleavage of the N-Cα bond c-terminal to the carbonyl group, yielding c- and z•-product ions.

In order to explain experimental observation of charge-remote fragmentation, Simons and co-workers proposed a different mechanism [22] that was later extended by Turecek and co-workers [23]. This so-called Utah-Washington mechanism suggests capture of an electron directly into the π*-antibonding orbital of an amine group (Figure 1.3B). This leads to aminoketyl radical formation. The resulting anion is neutralized by proton transfer due to the high basicity of the anion, ultimately resulting in cleavage of the N-Cα bond and generation of c- and z•- fragments. Alternatively, the Utah-Washington mechanism also acknowledges the possibility of electron capture at a positively charged site. In contrast to the Cornell mechanism, the electron is supposed to be transferred to an amide π*-antibonding orbital during relaxation into lower Rydberg states, followed by proton transfer [24].

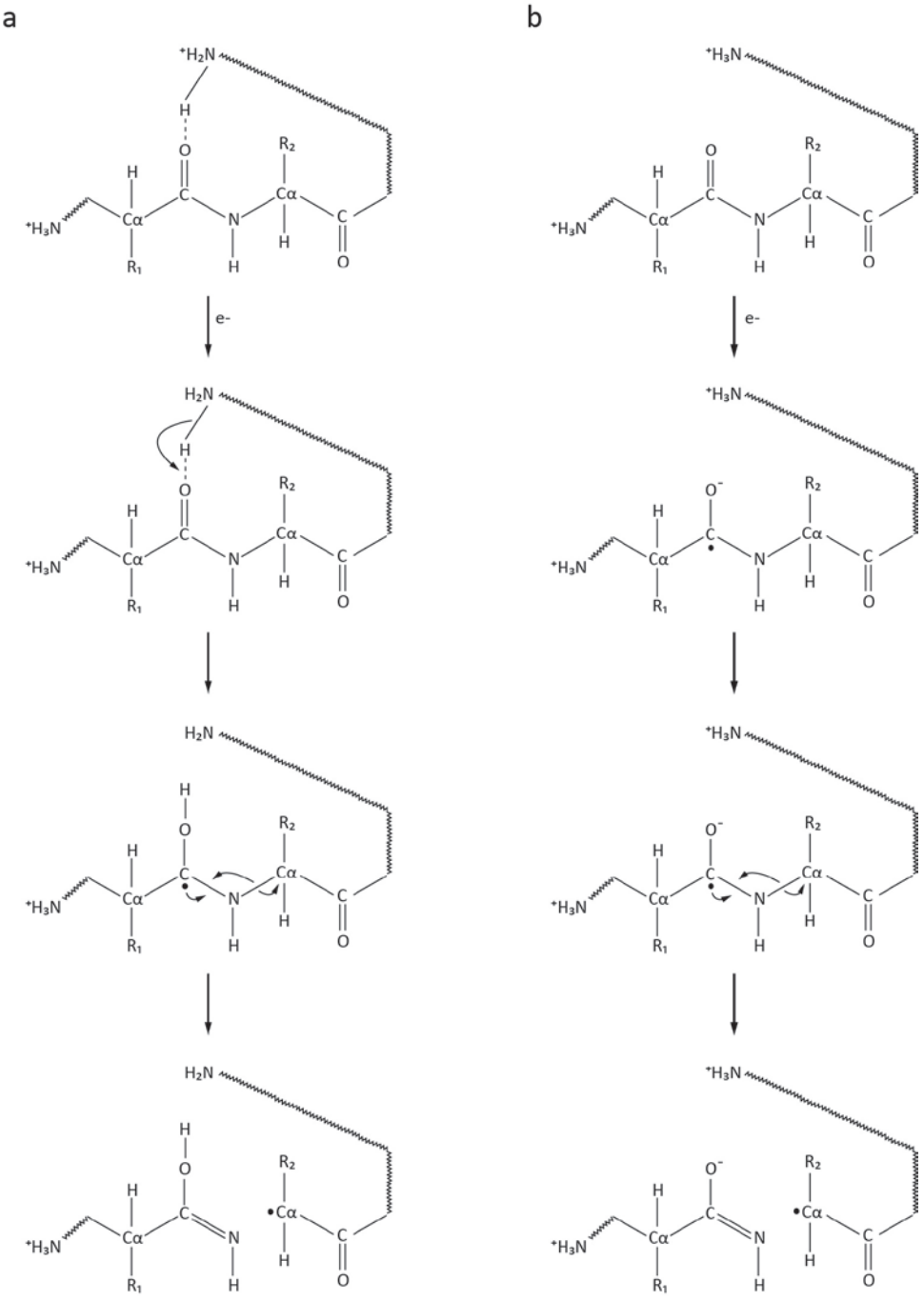



**Figure 1.3.** Proposed fragmentation mechanism for ECD/ETD. A) Cornell mechanism [20] and B) Utah-Washington mechanism [22, 23]. Figure adapted from [25].

It is believed that the fragmentation in ETD follows the same mechanisms. However, the conditions under which ECD and ETD are performed are quite different. The level of energy underlying the capture of free electrons in ECD is higher than in the transfer of electrons in ETD. Moreover, ETD reactions are performed in ion traps or quadrupole collision cells, which are operated in pressure regimes that are about six orders of magnitude higher compared to ECD which is usually performed in a very low pressure FT-ICR cell. As a consequence, ions are cooled much faster in ETD, resulting in overall lower internal energies and leading to more abundant charge-reduced radical intermediates, $[M+nH]^{(n-1)+\bullet}$ [26]. It was proposed that the $c^+$ and $z^{+\bullet}$-ions, generated upon electron transfer, are held together in close proximity via non-covalent interactions, resulting in this long-lived intermediate radical species [27]. This can result in observation of hydrogen atom rearrangement, leading to c-1 and z+1 ions [28]. The extent of the formation of these potentially undesired side products increases with longer lifetimes of the charge-reduced precursor [29] and was found to be higher for doubly charged compared to higher charged precursors [30]. The typically interrogated class of peptides in proteomics experiments are predominantly present as doubly and triply charged ions. The ETD fragmentation efficiency for these peptides is usually quite low due to the low charge-density [31]. This can be partly overcome by increasing the internal energy of the precursor before or during the electron transfer reaction [29, 32-34], or through additional collisional activation of the charge-reduced precursor. Coon and co-workers implemented this supplemental activation by applying resonant-excitation CID following an initial electron transfer step, which resulted in a substantial increase in the overall product ion yield [28].

Over the past years, several groups have reported on comparisons between electron-driven and collisional activation methods [31, 35, 36]. In summary, it is nowadays widely accepted that both methods have their strengths and weaknesses and their complementary use is highly recommended [37, 38]. In particular, CID is routinely used for sequencing of predominantly doubly charged peptides in bottom-up proteomics. On the other hand, ETD is the method of choice for the analysis of peptides carrying post-translational modifications [39, 40], multiply charged peptides [41] and also intact proteins [42].

## 1.2 Orbitrap Mass Spectrometry

The Orbitrap is the most recently developed mass analyzer that complements the array of established mass spectrometers such as sector field, time-of-flight (TOF), Fourier Transform-Ion Cyclotron Resonance and quadrupole mass analyzers. The initial description of the Orbitrap was published by Makarov in 2000 [43]. He proposed the principal of orbital trapping as a novel method for the mass-to-charge analysis of ions, based on its first de-

scription by Kingdon [44]. The Kingdon trap consist of a thin wire surrounded by a cylindrical electrode. Application of a voltage between the wire and the cylinder forces ions to the wire. Ions with sufficient kinetic energy in tangential direction can oscillate in stable trajectories around the wire. Makarov adopted this principle and designed the Orbitrap in a similar fashion. The Orbitrap is composed of an inner spindle-shaped electrode, surrounded by a barrel-shaped outer electrode (Figure 1.4) [43]. Application of a static electrostatic field to the axially symmetric electrodes creates a quadrologarithmic electrostatic potential that forces trapped ions to oscillate around the inner electrode in a harmonic orbit.

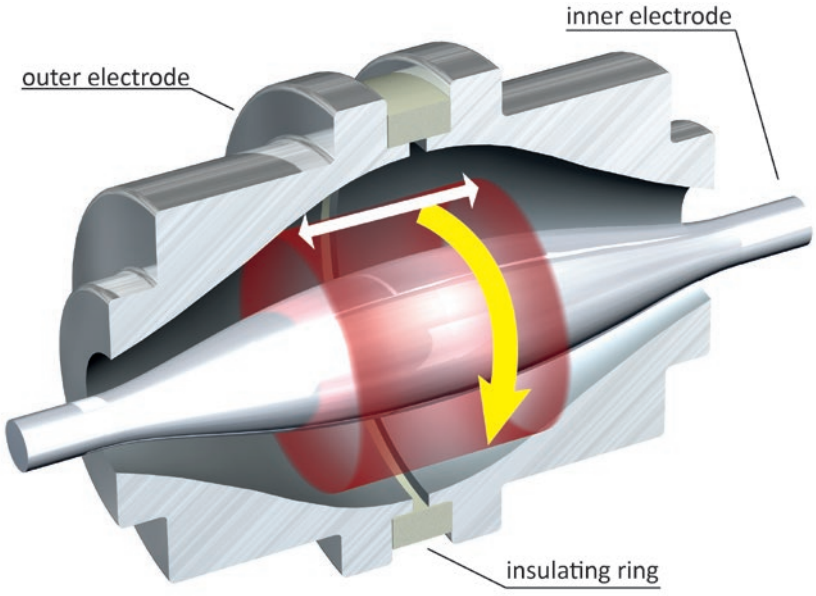



**Figure 1.4.** Schematic overview of the Orbitrap. Injected ions move in a spiral orbit around the inner spindle-shaped electrode (yellow arrow). They harmonically oscillate in axial direction due to the quadrologarithmic potential distribution (white arrow). Frequency of oscillation is only dependent on the m/z of the ions.

It follows from the potential distribution that the frequency of ion oscillation in axial direction is only dependent on the m/z-ratio of the ions:

$$\omega = \sqrt{(z/m) \times k}$$

The oscillating ions induce an image current which is detected via a differential amplifier. The recorded image current forms a transient in the time-domain, which exhibits a characteristic beat pattern (Figure 1.5). Longer transient recording times allow acquisition of more beats which increases the achievable resolving power [45]. However, frequency information can only be obtained from these beats, not from the signal in-between. The amplitude

decay of the beats mostly depends on the vacuum level in the Orbitrap as collisions with residual gas molecules result in dephasing or loss of ion packets. Consequently, transient recording is only useful as long as the amplitude of the beats is higher than the amplitude of the background noise. After acquisition, the recorded time-domain transient is digitized and converted into the m/z dimension using fast Fourier transformation [46], similar to FT-ICR signal processing [47]. Key performance metrics of the Orbitrap analyzer are the very high achievable resolving power, mass accuracy, dynamic range and its high upper mass limit [48-50].

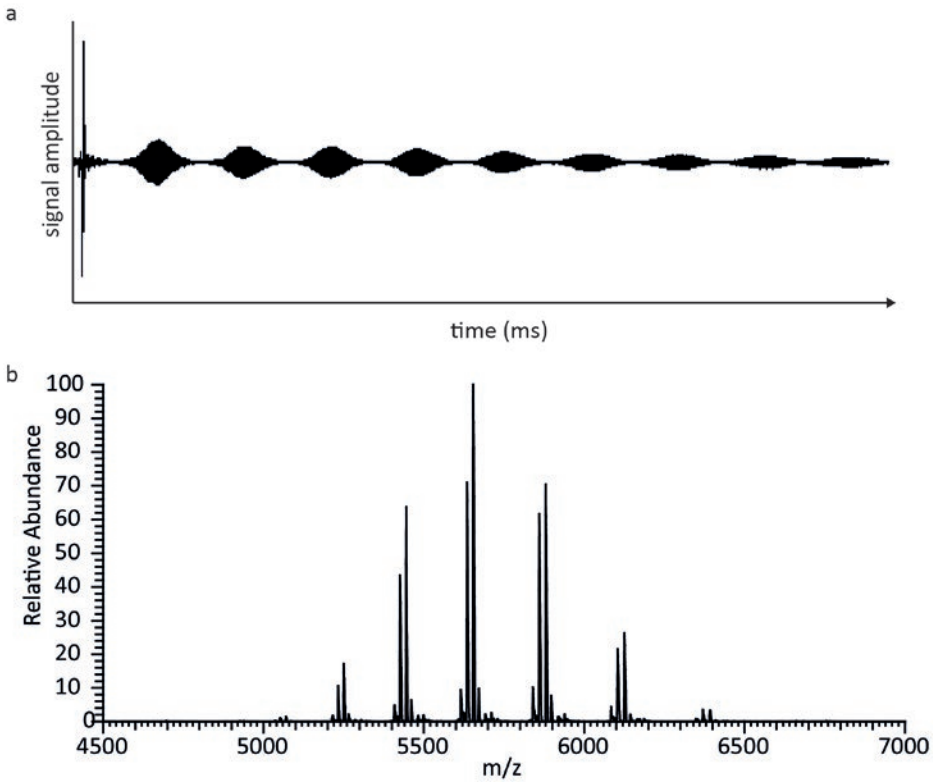



**Figure 1.5.** The transient of an antibody mixture acquired on an Orbitrap Exactive Plus exhibits the typical beat pattern (A). Fourier transformation into the m/z domain generates the spectra in panel B.

### 1.2.1 Recent Developments in Orbitrap Instrumentation

Over the last years the Orbitrap mass analyzer has been implemented into multiple configurations. Shortly after the initial description of the Orbitrap in 2000 [43] Hardman and Makarov interfaced the Orbitrap to an electrospray ionization source (ESI) [51], demonstrating high mass accuracy, high resolving power and good sensitivity for the analysis of

pulsed ion beams. The major breakthrough was achieved when the Orbitrap was coupled to a linear quadrupole ion trap (LTQ) soon after [49, 50, 52]. This class of hybrid instrumentation combined the high resolving power and mass accuracy of the Orbitrap analyzer with the sensitivity and MSn capabilities of ion trap mass spectrometers. Hyphenation of the two devices had become feasible through an additional C-trap, a curved RF-only trap, which enables storage of defined ions packages and orthogonal ejection into the Orbitrap analyzer. This first generation of the LTQ-Orbitrap featured a resolving power up to 100,000 at 400 m/z [50], dynamic range of more than 5000 [49] and very high mass accuracy <2 ppm [52]. Based on these three performance characteristics the hybrid LTQ-Orbitrap became the workhorse in many proteomics laboratories. The standard method for peptide sequencing utilizes a survey scan in the Orbitrap followed by CID fragmentation of five to ten of the most abundant peptides in the LTQ. The resulting scan cycle speed made this instrument capable of routine peptide sequencing on the LC timescale.

In 2007 Olsen et al. reported on the implementation of beam-type higher-energy collision induced dissociation (HCD) on the LTQ-Orbitrap [18]. They initially adjusted the RF voltages of the C-trap to enable triple quadrupole-like fragmentation followed by Orbitrap detection of the resulting product ions. Observation of decreased trapping capacity in the low mass range was an undesired disadvantage of this mode of operation. Therefore, an additional dedicated multipole collision was installed and directly coupled to the C-trap. This 2$^{nd}$ generation configuration was commercialized as the LTQ-Orbitrap XL and enhanced the analysis of intact proteins [45], phosphopeptides [53] and quantification of isobaric labeled peptides using iTRAQ or TMT [54, 55].

Soon after, the peptide activation toolbox of the Orbitrap platform was extended by the implementation of electron transfer dissociation (ETD) by Coon and co-workers [56]. The initial configuration utilized a pulsed, dual electrospray ion source [57], enabling front end-ETD via sequential injection of cations and anions. This setup demonstrated the potential of ETD in combination with Orbitrap detection for the analysis of larger peptides and even proteins. However, the anion reagent generated by the negative ESI source showed a relatively low ETD efficiency [56]. Moreover, the rather long cycle times limited the application for routine LC-MS/MS analyses. These drawbacks were overcome by installation of a negative chemical ionization source at the rear of the instrument [58], similar to the initial implementation on a standalone LTQ [21]. The NCI source generates radical anions that are moved via a transfer multipole through the C-trap to the linear ion trap. This configuration did not require polarity switching of the front end of the instrument and allowed for faster ETD cycle times with product ion analysis in the ion trap or Orbitrap. The implementation of ETD on the hybrid LTQ-Orbitrap greatly increased the analytical capabilities, opening up novel applications for peptide and protein sequencing. Based on the complementarity of

CID and ETD [35, 59] Swaney et al. designed a data-dependent decision tree logic that uses information on precursor charge and m/z from the MS survey scan to apply ETD or CID fragmentation on-the-fly [60]. The high resolution of the Orbitrap analyzer allowed for instant precursor charge state assignment, enabling the selection of the most suitable fragmentation technique. In general, CID was found to be the method of choice for doubly charged peptides whereas ETD provided better fragmentation for peptides with charge states ≥3+. The decision-tree method was later refined by Frese et al.[36].

In 2009 the hybrid ion trap-Orbitrap platform underwent major revision. The 3rd generation of this instrument was termed LTQ-Orbitrap Velos and featured a substantial increase in scan speed [61]. The linear ion trap was replaced by a dual-pressure ion trap that consists of two identical linear ion traps that are operated at different gas pressure regimes (Figure 1.6). The first trap is held at higher pressure for improved trapping and fragmentation whereas the second trap is kept at lower pressure for enhanced scanning capabilities [62]. Furthermore, the HCD collision cell was modified for faster extraction and more efficient transmission of all ions. This was achieved by an additional axial electric field applied to printed circuit boards in-between the rods of the HCD cell. Besides the increase in speed and sensitivity the new axial electric field also improved precision of reporter ion-based quantification approaches [63]. Together these improvements also facilitated more sensitive detection of CID and ETD product ions in the Orbitrap analyzer. A thorough comparison of all possible combinations of fragmentation technique and mass analyzer is described in chapter 2 of this thesis.

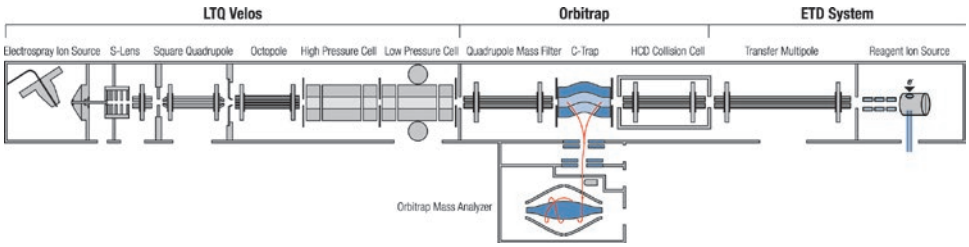



**Figure 1.6.** Schematic overview of the hybrid LTQ-Orbitrap Velos/Elite. The instrument consists of the dual pressure linear quadrupole trap which is coupled via a transfer multipole to the C-trap. From there, ions are sent in packages to the Orbitrap mass analyzer. The HCD collision cell is directly attached to the C-trap. A negative chemical ionization source, enabling ETD, can be mounted at the rear of the instrument. Figure adapted from [64].

The Orbitrap Elite is the latest hybrid ion trap-Orbitrap generation and was introduced in 2012 [65]. Key improvements compared to the Orbitrap Velos are faster scan speed and better dynamic range of the linear ion trap. The new Velos Pro ion trap allowed for rapid

CID scans with an MS/MS scan rate up to 12.5 Hz. Moreover, the Orbitrap analyzer was upgraded to a compact high-field Orbitrap. The smaller radius of the electrodes improves sensitivity and, together with a novel enhanced FT processing of the recorded transients, provides for up to fourfold higher resolving power at the same transient length. To date the Orbitrap Elite is the most versatile hybrid mass spectrometer, combining fast scanning, multiple fragmentation techniques, high resolution and mass accuracy.

Alongside the hybrid instruments, the Orbitrap has also been developed as a standalone benchtop platform, introduced as the Orbitrap Exactive [66]. This first generation benchtop Orbitrap featured an API source coupled to the Orbitrap via a C-trap and could be additionally equipped with an HCD collision cell for fragmentation [67]. The Exactive was initially designed for the MS analysis of small molecules, however, the additional HCD collision cell also enabled the use within proteomic workflows. The lack of the linear ion trap as accumulation and isolation device only allows for data-independent MS/MS acquisition [68]. This so-called all-ion fragmentation enables peptide identification via alignment of "co-eluting" precursor and product ion masses, similar to the $MS^E$ technology on Q-TOF instruments [69]. Soon after, Heck and co-workers extended the capabilities of the benchtop Orbitrap platform for the application in native mass spectrometry [70]. Instrumental modifications were made to enable the detection of ions up to 30,000 m/z, allowing for straightforward analysis of antibodies and intact protein complexes with high mass accuracy [70, 71].

Combination of the standalone Orbitrap with a quadrupole mass filter has been realized and defines a novel Orbitrap platform, termed Q Exactive [72]. The key difference to the Exactive is the transfer multipole that has been replaced with a quadrupole mass filter, enabling very fast HCD MS/MS scans. Basically, the setup of the Q Exactive corresponds to triple quadrupole (QqQ) instruments, except that the third quadrupole is represented by the Orbitrap analyzer. In contrast to hybrid ion trap-Orbitrap instruments, peptides selected for MS/MS are isolated "in space", subjected to HCD fragmentation and stored in the C-trap which allows for ion accumulation and Orbitrap mass analysis of ions from the preceding event in parallel, resulting in fast duty cycle. Moreover, this opens up capabilities for multiplexed scan modes, e.g. selected ion monitoring of multiple peptides or multiplexed MS/MS spectra [72]. Coon and co-workers recently reported on parallel reaction monitoring on a Q Exactive [73]. This method enables analysis and monitoring of all product ions of a given precursor with high resolution and mass accuracy, matching the performance of traditional SRM assays in terms of dynamic range and linearity. Together, recent developments of the Orbitrap mass analyzer have greatly expanded the toolbox for mass spectrometry-based proteomics.

## 2 Quantitative Proteomics for Differential Protein Expression Profiling

Over the past decade MS has matured from a technique that produces purely qualitative information into a versatile tool that provides accurate quantitative data on protein abundance. Quantitative information is a fundamental necessity to interrogate the highly dynamic global proteome of living organisms. Driven by recent advances in mass spectrometric instrumentation and data analysis software, today, MS-based proteomics builds the analytical framework for various challenges in biomarker discovery, systems biology and structural biology [5, 74]. Coupling liquid chromatography to MS facilitated the routine analysis of several thousands of proteins in parallel within a few hours and has mostly replaced 2D-gel electrophoresis based approaches.

Multiple quantification strategies for mass spectrometry-based proteomics have been reported. They can be distinguished into two main regimes: absolute and relative quantification. In this chapter we will focus on the latter and discuss practical aspects, applicability and problems of the currently most popular quantification strategies. For those interested in absolute quantification a detailed insight is given in a review by Brun et al. [75]. The need for reliable peptide identification is underlying all MS-based quantification strategies. Protein identification in the commonly employed bottom-up proteomics strategy involves enzymatic proteolysis followed by 1 dimensional- or 2 dimensional-peptide chromatographic separations. Trypsin and Lys-C are the routinely used enzymes where the former generates peptide lengths between 10-20 amino acids for the majority of the proteolytic peptides. These "MS-friendly" peptides are sequenced by tandem mass spectrometry (MS/MS) and identified by database search algorithms [76, 77]. Peptide quantification is performed either at the MS or MSn-level, depending on the quantification strategy.

The use of isotopes to label specific molecules has greatly expanded the toolbox in biochemistry. Radioisotopes such as $^{32}P$, $^{33}P$, $^{35}S$ or $^{125}I$ and stable isotopes such as $^{2}H$, $^{13}C$ and $^{15}N$ are widely utilized for quantitation, to track protein-protein interactions, to determine enzyme kinetics or for the dissection of metabolic pathways. Also nucleic acids, lipids and carbohydrates have been subjected to isotopic labeling. The main techniques to analyze biomolecules involving the use of non-radioactive, stable isotopes are nuclear magnetic resonance (NMR) and mass spectrometry. Today, mass spectrometry-based proteomics aims to provide large-scale quantitative information on protein abundances. The basic principle of stable-isotope based labeling for peptide and protein quantification is that the physicochemical characteristics of the differentially labeled peptides are nearly identical [78]. This includes sample preparation procedures, LC-separation performance, ionization efficiency and MS/MS fragmentation behavior. An additional assumption is that the different isotope variants do not influence any cellular process in the case of metabolic

labeling. Most labeling strategies introduce stable (non-radioactive) isotopes with distinct masses that allow each sample to be distinguishable at the mass spectrometry stage. The stable isotopes can be introduced via chemical reactions of a labeling reagent with a distinct functional group, by metabolic processes or enzymatic activity. Differentially labeled samples are combined prior to MS-analysis assuming an overall near-complete labeling efficiency. The shift in absolute mass depends on the number of heavy stable isotopes incorporated into the peptide or protein. The actual signal shift between the isoforms (in the m/z range) of a differentially labeled peptide depends on the absolute mass shift and the peptide ion charge state. For each peptide the area under the curve of the isotopic envelope is integrated over the LC elution time. Relative quantification is performed by calculating the ratio of the peak areas of the differentially labeled peptides. Intensity-based label-free quantification is based on the same principle; however, peptide abundances are retrieved from consecutive LC-MS/MS analyses since each sample is analyzed individually.

## 2.1 Metabolic Labeling

In metabolic labeling the stable isotopes are introduced into the proteins in living cells or organisms during protein synthesis. This requires that the organism or cell line has to be auxotroph for the source of the isotope used for labeling. Thus, after a few cell cycles the heavy isotopes are fully incorporated into all proteins. One of the first applications of metabolic labeling in proteomics involved the use of $^{15}N$ for labeling of bacteria [79]. The heavy nitrogen is introduced into the proteins mostly via $^{15}N$-ammonium or nitrate salts in the culture medium. This leads to all N-atoms being replaced by the heavy nitrogen isotope during cell growth and division. Complete labeling is usually achieved after 5-10 cell cycles. Consequently, all peptides can be quantified by MS independent of the amino acid composition. Thus, every enzyme can be utilized for proteolytic digestion. $^{15}N$-labeling is mostly applied in bacterial or S. cerevisiae cultures [80] but it has also been successfully applied in E. coli [81]. Multicellular organisms can be labeled by feeding them $^{15}N$ enriched yeast, bacteria or algae. This approach has extended the technique to model systems such as D. melanogaster and C. elegans [82]. Metabolic labeling of mammals has first been reported by Yates and co-workers who introduced stable isotopes into rat [83]. Moreover, metabolic labeling of chicken has also been reported [84]. However, complete incorporation in rodents is hard to achieve even after long-term feeding with $^{15}N$ enriched nutrients. In plant biology $^{15}N$ can be introduced via liquid nutrition solutions [85], axenic liquid cultures [86] or agar [87]. A number of different plants have been metabolically labeled for quantitative proteomic experiments using $^{15}N$, including the most studied plant model Arabidopsis thaliana [88]. A comprehensive review on the use of $^{15}N$ in plants has recently been published by Arsova et al. [89].

The main advantage underlying metabolic labeling strategies is that the isotopic labels are introduced during protein synthesis in living cells. After a few cell cycles the heavy isotopes are typically fully incorporated into all proteins. This allows combination of the differentially labeled samples at the earliest possible stage in proteomic workflows which thepretically minimizes variation that can potentially be introduced by systematic errors at every sample preparation step (Figure 2.1).

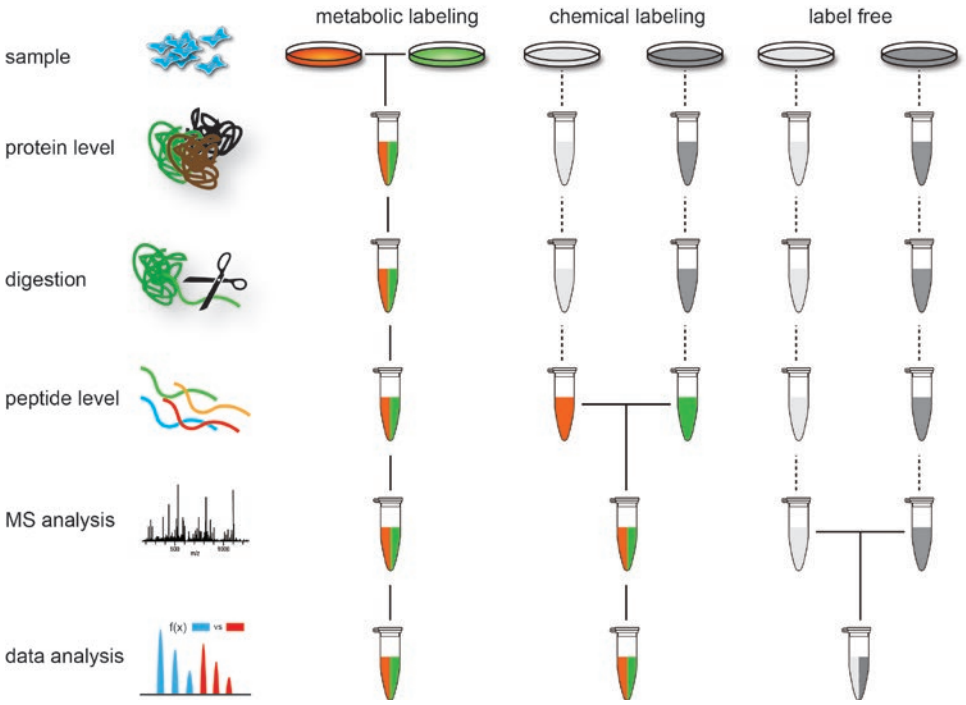



**Figure 2.1.** Workflows in quantitative proteomics. Dashed lines highlight the steps where samples are treated separately which introduces experimental variation. Horizontal lines illustrate the step where samples are combined. Upon introduction of a label samples are highlighted in orange and green, respectively. In metabolic labeling, samples are combined at the earliest possible step which, in theory, minimizes variation introduced at each sample preparation step. In chemical labeling samples are combined after labeling. Label-free quantification requires every sample to be processed individually. Here, the samples are combined at the data analysis level. Figure adapted from [91].

One of the caveats of $^{15}N$ based labeling is that the mass difference between the labeled and unlabeled version of the peptide varies with peptide length and amino acid composition, which complicates data analysis. Moreover, small $^{14}N$-contaminations can already cause incomplete peptide labeling and thus hamper peptide identification and correct assignment of the corresponding peptide pairs. One also has to consider that, despite the fact

that all isotopes have the same chemical properties, enzymatic reactions can be compromised which may lead to a bias against the $^{15}N$ isotopes in some cellular processes [90]. Besides these drawbacks, for some biological systems $^{15}N$-based labeling is too expensive or laborious.

Labeling via stable isotopes by amino acids in cell culture (SILAC) is another strategy that facilitates complete and proteome-wide labeling. For this purpose, amino acid deficient cell culture medium is supplemented with stable isotope enriched amino acids which are taken up by the cells and incorporated into all proteins. Logically, this strategy requires the cells to be auxotroph for the labeled amino acid [92]. Typically one or two specific amino acids which contain the stable isotopes are supplemented to the culture medium. Unlabeled and labeled versions of the same peptide are separated by a defined mass difference, depending on the number of labeled amino acids per peptide. Leucine [92] and methionine [93], among other amino acids, have been utilized for introducing deuterium as the stable heavy isotope. However, due to the different RP-HPLC retention behavior of deuterium-labeled peptides compared to their unlabeled counterparts most labeling approaches have switched to $^{13}C$ and $^{15}N$ as a source of stable isotopes.

Besides the need for the cell line being auxotroph for the particular amino acid one also has to consider the variation in the frequency of occurrence of the amino acids in the total proteome and the need for a corresponding protease. For instance, tryptophan and cysteine are relatively less frequent amino acids which make them less suitable for stable isotope labeling. Nowadays, arginine [94] and lysine [95] are the most commonly utilized amino acids to introduce stable isotopes into the proteome (Figure 2.2 A). Using these two amino acids in combination with Trypsin or Lys-C digestion generates peptides that contain at least one c-terminal labeled amino acid. Up to three samples (unlabeled, $^{13}C$- or $^{13}C/^{15}N$-containing amino acids) are usually differentially labeled and mixed in a single experiment. A caveat of SILAC labeling is the potential metabolic conversion of arginine to proline [96]. This conversion results in an undesired mass shift of proline-containing peptides which hampers quantification and leads to an underestimation of the original heavy peptide abundance. This issue may be addressed by either careful titration of the amount of arginine [97], the addition of proline to the culture medium [98] or by computational correction of the obtained peptide ratios [99].

Over the past years the SILAC toolbox has been extended to label also whole organisms such as Drosophila melanogaster [100] and Mus musculus [101]. The so-called SILAC mouse has been successfully used for in-depth comparison of the left and right ventricle proteomes [102] as well as for investigation of the global proteome changes during aging [103]. SILAC has also been used for pulse labeling to study global protein turnover [104].

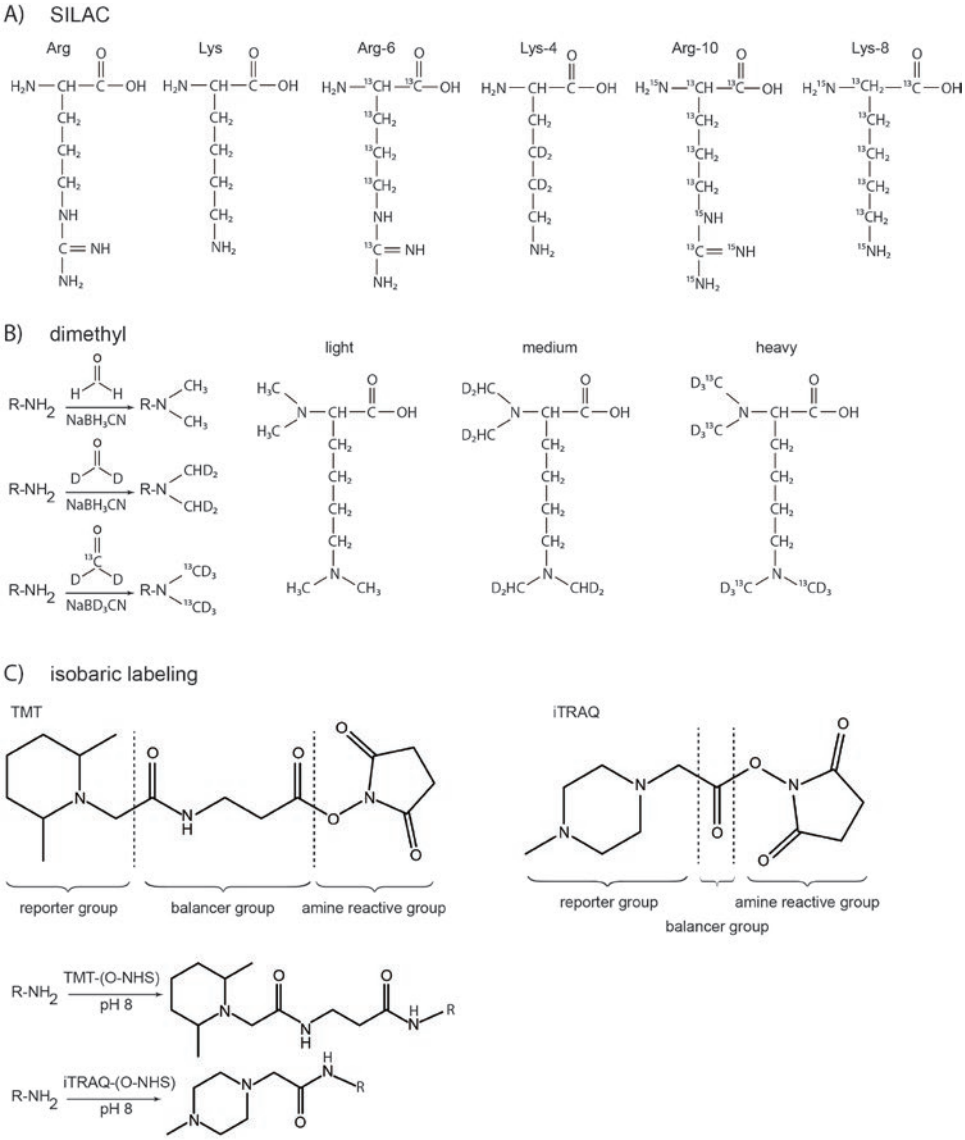



**Figure 2.2.** Chemical structures of stable isotope labeled peptides in quantitative proteomics. A) In SILAC, $^{13}C$, $^{2}H$ and $^{15}N$ isotopes are introduced into the proteins via lysine and arginine ($\Delta m$ = +4/+6 Da for intermediate and +8/+10 Da for heavy label, respectively). B) Dimethyl labeling targets free amine groups via reductive amination. Formaldehyde and cyanoborohydride with varying number of heavy isotopes ($^{13}C$ and $^{2}H$) are employed to methylate free amine groups which results in a mass increase of 28, 32 and 36 Da for the light, intermediate and heavy label, respectively. C) Isobaric labeling via TMT or iTRAQ utilizes tags that are composed of three parts which add up to the same molecular mass for all channels. The isobaric tags label peptides via NHS-chemistry through an amine-specific reactive group. The total number of $^{13}C$, and $^{15}N$ atoms is constant, however, the position of the heavy atoms varies between the different labels. The identity of the tags is revealed by the reporter group upon fragmentation of the peptide (the balancer group undergoes neutral loss).

In this so-called pSILAC experiments cells are exposed to heavy amino acids for a defined period of time prior to lysis. The incorporation of the heavy stable isotopes into the proteins, i.e. the protein turnover rate, can be determined by the ratio between the heavy and light peptide versions. Selbach and co-workers extended the experimental procedures to allow investigation of changes in global protein turnover upon an external stimulus [105]. To date SILAC labeling is not applicable to human tissue or any clinical sample. However, the Mann group recently introduced the so-called super SILAC approach for quantification of animal and human tissue proteins [106]. For this purpose they used a mixture of SILAC-labeled cells as an internal standard [107]. The tissue complexity arises from its heterogeneous composition from various cell types. To mimic the tissue of interest as close as possible they combined different cell lines at differing levels for the super SILAC mixture. The proteins of the unlabeled samples of interest are first quantified against the internal standard. Next, the differences between the tissue samples are determined by calculation of the ratios between initially determined ratios against the internal standard. The super SILAC approach expands the SILAC toolbox to theoretically all kind of samples. However, the challenge arising is the need for a defined reference sample that ideally represents exclusively the same proteome as the tissue sample of interest without adding too many unspecific proteins that would increase complexity and complicate quantification. Moreover, the super SILAC approach cannot reflect the tissue microenvironment of living cells.

## 2.2 Chemical Labeling

Chemical stable isotope labeling strategies make use of the presence of reactive groups of peptides and proteins. The main difference in contrast to metabolic labeling is that chemical labeling approaches are virtually applicable to all samples, including body fluids and any clinical samples because the stable isotope labels are introduced in the sample preparation process. In 1999, Aebersold and co-workers introduced isotope coded affinity tags (ICAT) which targeted the thiol group of cysteine residues [108]. This tag is composed of three parts: a thiol-specific reactive group, a linker that contains either zero or eight deuterium atoms and a biotin affinity tag. The latter enables affinity purification of the labeled peptides via streptavidin, which leads to a reduction in complexity of the sample. On the contrary, ICAT labeling is difficult for smaller proteins that contain only a few cysteines. ICAT labeling has also been successfully employed for organelle proteomics [109]. Since the introduction of ICAT other chemical groups such as carboxyls [110, 111] have been explored for chemical labeling. However, the majority of the techniques target the amine group at the peptide n-terminus and at the ε-position of lysines. Primary amines can be easily derivatized using various approaches, e.g. acylation or sulfonation [112].

Derivatizations via specific N-hydroxysuccinimide (NHS) chemistry or reductive amination are among the more frequently used techniques. Labeling can be performed on either peptide or protein level. The former approach has the advantage that the samples can be combined at an earlier step in the workflow. However, the fact that the most popular proteases Trypsin and Lys-C do not cleave derivatized lysine residues limits the applicability. Moreover, incomplete labeling is also more likely to occur when labeling is performed on intact protein level. This is likely the reason why chemical labeling of intact proteins is not so frequently used in proteomic experiments.

In stable isotope dimethyl labeling the amine groups are labeled by forming a Schiff base upon reaction with formaldehyde which is then rapidly reduced by cyanoborohydride [113-116]. Two cycles of these reactions will convert primary amines into tertiary amines with two methyl groups (Figure 2.2 B). Isotopomers of formaldehyde and cyanoborohydride are utilized to introduce Cα-bound Deuterium and $^{13}C$ atoms as stable isotope labels to each peptide. Typical tryptic peptides contain either one (C-terminal arginine) or two labels (C-terminal lysine) which translates to a mass increase of four or eight Da, respectively. As in SILAC, commonly three samples are differentially labeled and analyzed together in one run (Figure 2.3). Several advantages make dimethyl labeling an attractive option in quantitative proteomics. First, the required reagents are cheap and the reaction is very fast and efficient. Secondly, the labeling protocol allows great scalability from low µg to mg amounts of sample. Thirdly, using dimethyl labeling prior to immunoprecipitation (IP) on peptide level reduces variation potentially being introduced at the IP- or LC-MS/MS steps [117]. Moreover, the labeling can be performed in solution, on C18 solid phase extraction columns or on-line coupled to HPLC and MS [118, 119]. The use of deuterium leads to a small retention time shift when using reversed phase-based peptide separation coupled to MS [120]. This can influence accuracy if peptide quantification is solely based on a single MS spectrum, but has little to no effect on quantification if the area under the whole chromatographic peak is integrated [121, 122]. A comprehensive overview of applications using dimethyl labeling was recently published by Kovanich et al. [123].

Acylation of primary amines via esters such as N-hydroxysuccinimide provides another elegant way to introduce stable isotopes to peptides and proteins. Münchbach et al. used $H_4$/$D_4$-Nicotinoyloxysuccinimide esters to selectively label the N-terminus of tryptic peptides which induces a mass shift of four Da [124]. However, mixing differentially coded samples increases sample complexity which can negatively affect proteome coverage and quantification accuracy. A potential solution to this are so-called "isobaric" tandem mass tags which also employ NHS chemistry to target free amines and are currently among the most popular chemical labeling strategies (Figure 2.2 C). The most common ones allow multiplexing of up to six (TMT) [125] or eight (iTRAQ) [126] samples.

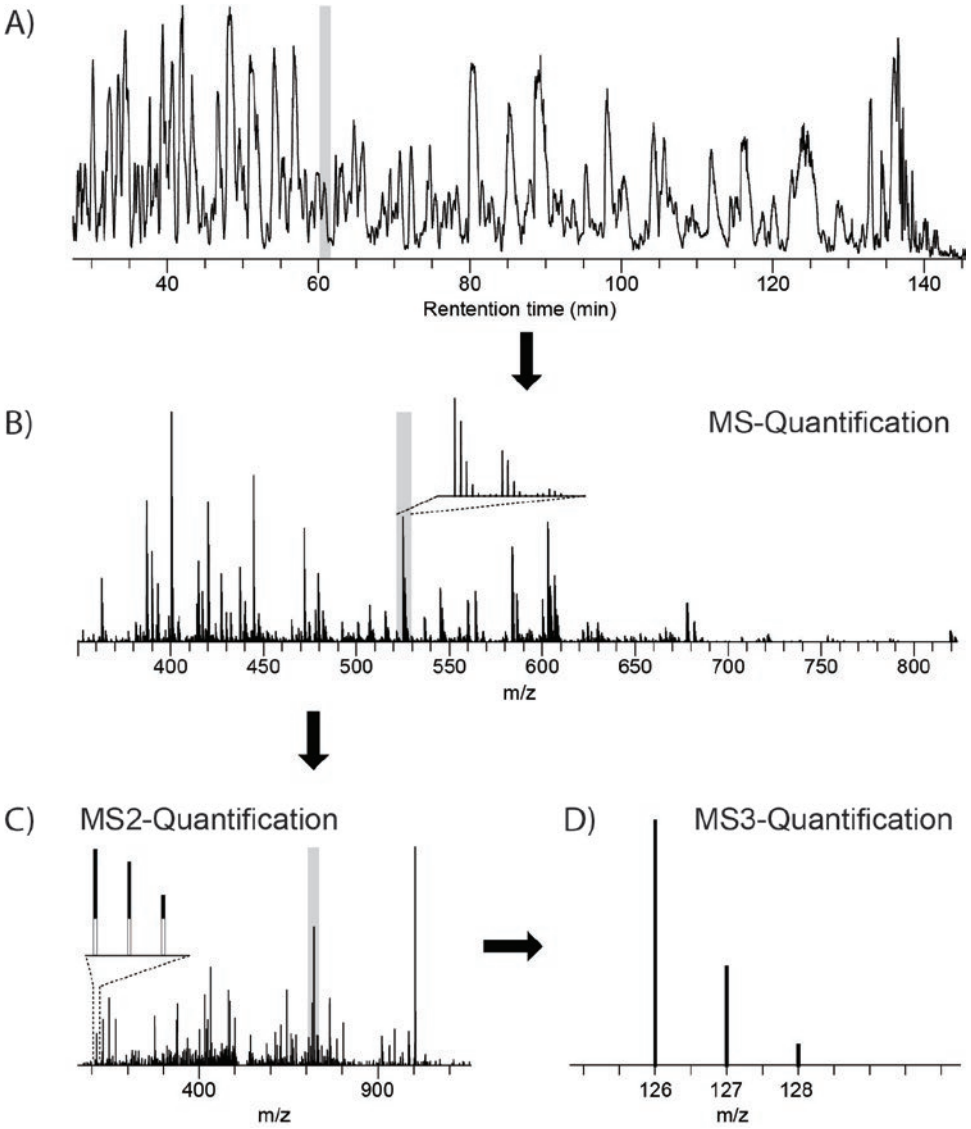



**Figure 2.3.** Peptide quantification on MS- and MSn-level. A) Typical chromatogram of a complex peptide mixture separated by reversed-phase nano-HPLC. B) Full scan MS spectrum at a given time of the analysis. Inset shows a zoom-in of a labeled peptide (three different labels). MS-level quantification is performed by comparing the MS abundance of the differentially labeled peptides. C) MS/MS spectrum of a peptide after collision induced dissociation. Inset shows the low mass region where the reporter ions of iTRAQ/TMT labeling are present. Accuracy in MS2-based reporter ion quantification is hampered by interference of co-isolated and co-fragmented background peptides (outlined rectangles) which leads to ratio distortion. D) The true intensities of the reporter ions are unveiled by MS3 methods which remove interfering background reporter ions.

Contrary to other chemical labeling strategies, isobaric labeling introduces labels that have the same total mass. The tags are composed of a reactive group, a reporter and a balancer group. The number of $^{13}C$ and $^{15}N$ atoms remains the same between the different tags; however, the position of the heavy carbon and nitrogen atoms within the reporter and balancer group changes. Thus, the same peptide originating from samples that are differentially labeled are indistinguishable in the full MS scan. Upon fragmentation they generate defined reporter ions in the low mass region of the MS/MS spectra which are used for quantification (Figure 2.3 C). The benefit of this approach is its superior multiplexing capabilities because neither LC separation nor full MS spectra suffer from increased complexity. This makes isobaric tagging a favorable technique for clinical samples or time-course experiments. Recently, Gygi and co-workers showed that a combination of metabolic and isobaric labeling further increases multiplexing and allows analysis of up to 18 samples simultaneously [127]. Another advantage of TMT- and iTRAQ quantification is the higher signal-to-noise ratio in the reporter ion m/z range of MSn spectra compared to MS level quantification.

One limitation of isobaric tagging approaches is that not all mass spectrometers are suitable for the analysis because of the reporter ion masses being in the low mass region (<150 m/z). Instrument configurations that support the analysis of isobaric labeled peptides are Orbitraps in combination with HCD [18, 63] and ETD [128, 129] fragmentation or Q-TOFs [130, 131]. The widely used ion trap instruments suffer from an inherent low mass cut-off that ultimately depends on the amplitude of the applied RF voltage during CID. Fragment ions below 1/3 to 1/4 of the precursor mass are ejected from the trap and cannot be observed. Thus, ion trap instruments can only be used in combination with pulsed-Q fragmentation [132]. Moreover, the isobaric label increases hydrophobicity and basicity of the peptides which is unfavorable for widely used methods employing collision induced dissociation [133]. A common problem in isobaric tagging-based quantification is interference from near-isobaric peptides that are within the isolation window of the peptide which is selected for fragmentation, first described by Bantscheff et al. [132]. This co-isolation issue leads to ratio distortion and hampers quantification accuracy (Figure 2.3 C) [134, 135]. Reporter ion interference by co-isolated contaminating peptides will skew the observed ratios towards unity as in most experiments the majority of the proteins remain unchanged between the samples [136]. Enhanced peptide fractionation and LC separation can reduce the ratio distortion but to a small extent [137]. Using narrow isolation windows theoretically reduces the amount of undesired interference, however, a recent report indicates that there is no correlation between the extent of co-isolation and the accuracy of the observed reporter ion ratios [138]. The reporter ion interference problem can be overcome by MS3-based methods (figure 2.3 D). Ting et al. proposed a method that initially subjects

peptides to ion trap-CID [139]. Subsequently, one of the fragment ions is isolated for HCD fragmentation to obtain the reporter ion intensities. Wenger et al. described a method that uses proton-transfer ion/ion reactions followed by isolation of the charge reduced precursor and HCD fragmentation [140]. The principle of both methods is based on increasing the mass-to-charge difference between the peptide of interest and contaminating ions prior to the ultimate fragmentation event which yields the reporter ions. However, these methods come at the cost of a substantial reduction in duty cycle speed and sensitivity.

### 2.3 Label-free Quantification

Label-free quantification strategies are based on either spectral counting or precursor ion signal intensity. In the latter one, label-free quantification is performed by extracted ion chromatograms (XICs) for all peptides over the whole LC-MS/MS run [141, 142]. For the ratio calculation the XICs are matched between runs from each sample. In spectral counting-based label-free quantification the number of peptide fragmentation events is used to estimate protein abundances [143]. It is widely used in proteomics to provide a semi-quantitative estimate of protein abundances, although concerns are often raised against this technique, vide infra.

In general, label-free quantification features some advantages compared to stable-isotope based labeling strategies: (i) Each sample is analyzed separately which allows for comparison of an unlimited amount of samples. (ii) The sample preparation is straightforward since no isotope-coded label has to be introduced to the peptides or proteins. (iii) samples from virtually all sources can be analyzed, which is especially of interest for clinical researchers. The main disadvantages of all label-free approaches are the massive increase in instrument analysis time (one run per sample) and compromised accuracy and precision [144]. Sample preparation and MS-analysis are carried out separately for each sample (Figure 2.1) which in turn increases the probability for an increase in variation between the samples. Thus, label-free quantification approaches require robust instrumentation that is capable of providing reproducible LC-separation and MS-detection.

## 3 Data Analysis in Differential Quantitative Proteomics

The measured intensity of peptides not only depends on quantity, but also on physicochemical properties of each peptide [145], and therefore establishing absolute quantities by mass spectrometry is not trivial. Consequently, most software do not attempt to do so, but calculate the ratio between different conditions, a method known as relative or differential quantification.

Different quantitative methods yield any number of measurements for a peptide in different conditions. In this text, we choose to refer to the measurements on a single condition as one channel. A SILAC experiment typically measures two channels, while iTRAQ or TMT have four to eight channels. Since there are many different methods of quantification, currently no single software package will analyze all of them. A list with a selection of software to analyze the various quantification methods is summarized in table 1.

**Table 1.** A selection of popular software packages for quantification. Licensing: OS = Open Source; $0 = free († for academic use); $ = commercial and proprietary.

| | | | Quantification methods | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | LC-MS | | | | | | MSn |
| Name | Reference | License | spectral counting | label-free | Dimethyl | SILAC | N14/N15 | ICAT | iTRAQ TMT | |
| APEX | [163] | OS | ○ | | | | | | | |
| Census | [148] | $0 | | ○ | ○ | ○ | ○ | ○ | ○ | |
| Isobar | [173] | OS | | | | | | | ○ | |
| Mascot Distiller | Matrix Science | $ | ○ | | ○ | ○ | ○ | | ○ | |
| Mascot Server | Matrix Science | $ | ○ | | | | | | ○ | |
| MaxQuant | [121] | $0 | | ○ | ○ | ○ | | | ○ | |
| MSQuant | [151] | OS | ○ | | ○ | ○ | ○ | | | |
| MzMine 2 | [176] | OS | | ○ | | | | | | |
| OpenMS/TOPP | [177] | OS | | ○ | ○ | ○ | | | ○ | |
| PEAKS | Bioinformatics Solutions | $ | ○ | ○ | ○ | ○ | | | ○ | |
| Progenesis LC-MS | Nonlinear Dynamics | $ | | ○ | | | | | | |
| Proteome Discoverer | Thermo Fisher Scientific | $ | ○ | ○ | ○ | ○ | | | ○ | |
| Skyline | [178] | OS | | ○ | | | | | | |
| TPP XPRESS | [179] | OS | | | | | ○ | ○ | | |
| TPP ASAPRatio | [180] | OS | | | | ○ | ○ | ○ | | |
| TPP SuperHirn | [181] | OS | | ○ | | | | | | |
| TPP Libra | - | OS | | | | | | | ○ | |

## 3.1 Data Analysis for Labeling Approaches

The advantage of isotope-labeled approaches is that different channels can be measured in a single LC-mass spectrometry run. Peptide peaks are 3-dimensional (m/z, time, intensity) objects on the chromatogram. Peak detection is a complex process, in which an algorithm performs a multiple fitting operation on the data. Census, for instance, predicts isotope patterns based on identifications and extracts chromatograms accordingly [147, 148]. Similarly,

MaxQuant first fits Gaussian curves to the MS data and connects the centroids of these fits to obtain an elution profile [121]. Census extracts elution profiles based on identification events, to circumvent problems caused by peptides with strong tailing behavior. MaxQuant on the other hand performs smoothing on the elution profile, and uses local minima to separate different peptides with highly similar m/z MS values. Working with centroided data, Khan et al. employ space-partitioning algorithms combined with methods from graph theory to group peptides together to substantially decrease processing time [149]. For a brief explanation on this method see [150].

The different channels can then be extracted from the chromatogram when looking at a known mass difference within an elution time window (Figure 3.1 A) for so called peak pairs (or triplets). Most software will first try to extract peak pairs and assign peptide sequencing events to the peaks afterwards. For $^{15}$N-labeling this is problematic, since the mass difference between two peaks depends on the amino acid composition of the peptides. Census [148] and MSQuant [151, 152] address this problem by using identified peptides before extracting precursor XICs. A problem that is not currently addressed in most packages is the possibility that the isotopic clusters of two labels may overlap. This effect can be corrected for by in silico deconvolution [153].

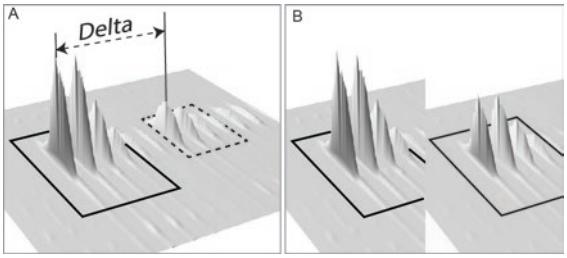



**Figure 3.1.** Stable isotope labeling and label free quantification methods. Graphical representation of extracted ion chromatograms (XICs). The height of the peaks denotes intensity, the left axis is elution time, and the horizontal axis represents the m/z dimension. Single peptides are seen as clusters of isotopic variants. A) In stable isotope labeling quantitation, differentially labeled peptides elute in a limited time window and exhibit a known m/z difference (delta). B) LC-MS label free methods compare the XICs between different runs that need to be aligned before comparison. Figure adapted from [146].

## 3.2 Label-free Approaches

### 3.2.1 MS (survey-scan) based Approaches

If a label-free experiment is conducted, the detected peptide peaks (vide supra) of two or more LC-MS based runs have to be compared (Figure 3.1 B). In order to match peaks from the same peptides, the chromatograms need to be aligned. In this alignment process, fluc-

tuations in chromatography need to be addressed [154, 155]. After the alignment, peptide intensities are extracted and compared between the two runs. For experiments involving more than two conditions or replicated measurements, multiple pair-wise alignments may be performed to make all chromatograms comparable. Similarly, the AMT (Accurate Mass and Time) tag approach normalizes peak elution times and compares the detected peaks to a pre-built database of AMT tags [156]. This has the additional advantage of avoiding additional MS/MS for the identification of the peptides.

### 3.2.2 Spectral Counting

Since the speed of modern mass spectrometers is high, a peptide is likely to be fragmented several times, depending on its abundance [157]. Therefore, protein abundance can be estimated by counting the number of peptide-to-spectrum matches (PSMs) for every protein. The main concerns raised against spectral counting are that it is prone to errors from instrument-specific parameters such as sequencing speed, ionization efficiency and dynamic exclusion. Moreover, a substantial amount of spectral counts per protein is necessary to allow for robust statistical analysis [144]. Spectral counts should be corrected to prevent overestimation of longer proteins, which produce more peptides by enzymatic digestion. The protein abundance index (PAI) is the observed number of peptides divided by the number of theoretical peptides [158]. Protein size corrected factors are for instance the spectral abundance factor (SAF), counts compensated by the number of amino acids per protein [143] and Fabb which compensates for protein mass [159]. The NSAF are SAF values that are normalized to total experimental SAF, to make different runs comparable [160]. Ishihama et al. empirically observed that the relation between protein concentration and PAI follows a logarithmic trend, which led them to propose the exponentially modified PAI (emPAI) [161]. Similarly, the iBAQ method sums all peptide XICs and divides them by the number of possible peptides [105]. Alternatively, protein abundance may be estimated by the 3 peptides with the highest intensity with T3PQ [162]. An effort to compensate for the physicochemical properties of peptides is reflected in the APEX method [163, 164], using machine learning to predict detectability of peptides and adjusting the counts accordingly. One problem of shotgun proteomics is the possibility that a peptide sequence matches several proteins [165]. More elaborate schemes are proposed to compensate for the effects of these shared peptides [166], although these are not generally applied yet. For further assessment of spectral counting-based label-free quantification the reader is referred to a review by Lundgren et al. [167]. In general, intensity-based label-free approaches are considered to provide better precision than spectral counting [168]. A comprehensive overview on label-free quantification is beyond the scope of this chapter and more detailed information can be found elsewhere [169].

### 3.3 Analysis of Fragmentation-based (MSn) Quantification

The fragments resulting from the isobaric tags in a TMT or iTRAQ experiment can simply be extracted from the fragmentation spectrum and used as a measure for quantification. It was shown that the reporter fragments of co-eluting peptides in MS/MS spectra interfere with the correct measurement of the real intensities for that peptide, leading to compression of the ratios [132]. Since the amount of compression depends on the sample it is currently not possible to compensate for these effects by software, and optimizations in LC-MS protocols are recommended, for instance the use of MS3 fragmentation [139].

### 3.4 Data Normalization

Since ratio data is inherently asymmetrical, the first step in the normalization of ratio data is the transformation into log values, typically to a base of 2. Values with a log2-ratio of zero represent equal expression, values of -1 and 1 represent twofold down- and twofold upregulation, respectively. In most experiments only a relatively low subset of proteins is expected to be differentially regulated, so the majority of the log-ratios are expected to be centered around zero. If an experimental bias occurs, e.g. by pipetting errors, the mean or median of the log-ratios may differ from zero. In that case, simply subtracting the mean or median log-ratio from all data will compensate for the bias. Moreover, non-linear trends are often observed if the log-ratios M are plotted against the log of the intensity values A (a so-called "MA plot", Figure 3.2 B). These trends may be removed by performing a moving average or locally weighted regression (LOESS) algorithm along the A axis, and subtracting the fitted line. A graphical software package implementing these methods is DanteR [170].

A problem that arises during the analysis of ratios is that some proteins will not be measured in one of the channels. A common solution is to replace the zero intensity values with very low numbers, sometimes related to the noise level. Unfortunately, the resulting ratios are usually large, disrupting normalization and significance testing. Therefore, it may be preferable to remove these measurements before further processing and treat them as special cases.

### 3.5 Significance Analysis

An important aspect after measuring and normalizing peptide and protein ratios is the determination of regulated proteins. Traditionally, a simple fold-cutoff was performed, reporting all proteins with a fold change higher than a pre-set threshold of for instance two-fold. Unfortunately, this method gives no information about the quality of the results, i.e. the

number of false positives in the reported protein list. Applying statistical analyses enables control over the quality of the results. There are different analysis tools available, which can be applied in different situations. The basic idea of statistical tests is to model variation (loosely: "noise") of some measurements, and to find out whether a single measurement deviates from some value enough to not be caused by chance. In quantification, a test would determine if the log-ratio of a protein is significantly different from zero. The variation is often modeled as a bell-shaped distribution (Figure 3.2 A) which centered around the true value and exhibits a width that depends on the variability.

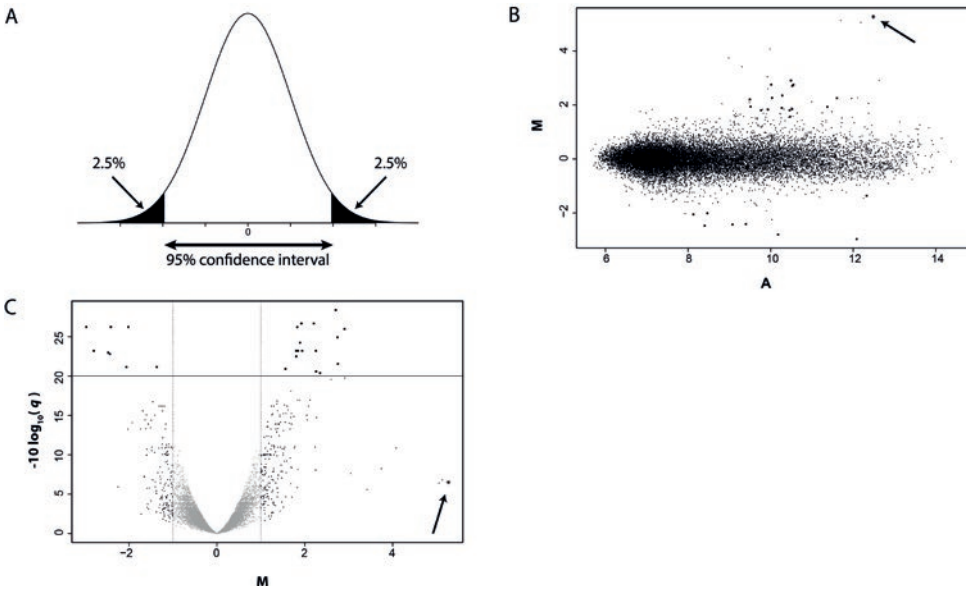



**Figure 3.2.** Statistical analysis of quantitative proteomics data. A) A bell-shaped curve shows the probability densities for a measurement. The area under the curve is comparable to a histogram, counting numbers of observations of measurements different than the expected value (0). Summing up areas under the curve gives an estimate of the probability for a certain measurement to appear within that region. Measurements that fall within 95% of the expected value will fall within the area denoted as the "95% confidence interval". The p-value of a measurement is the area under the curve that falls outside of the smallest symmetrical confidence interval that still contains that measurement. B) An MA plot of log2 ratios (M) versus the log10 of the average intensity (A). This MA plot depicts the average of several replicated experiments. The star indicates a measurement with a large average ratio that is not tested significant (see also panel C). The squares denote statistically significant values. C) A volcano plot of the data shown in panel B. The y-axis represents the -10*log10-transformed q-value (which corresponds to the p-values that are corrected for multiple testing), higher numbers denote a lower p-value. Two dotted lines indicate 2x up- and 2x downregulation, respectively. The solid horizontal line depicts the cutoff value of a false discovery rate (FDR) of 1%. Squares and star correspond to panel B. Here, the starred measurement clearly falls well below the selected FDR cutoff. Consequently, it is not statistically significant. (Dataset and analysis from The LIMMA guide, Gordon Smyth, 2012)

The height of the distribution is the "probability density", enabling calculation of the probability of a measurement within a certain area, which is chosen by the researcher. Figure 3.2 A depicts the area encompassing 95% probability, where the lower and higher bounds enclose the 95% confidence interval. A measurement outside of this interval is significantly different from the real value at a confidence level of 5%. Therefore, performing a significance test is mostly about creating the correct confidence intervals, i.e. estimating the variability.

If sufficient replicate experiments are performed, the variability of each protein can be estimated from the replicate values and used in a t-test. Increasing the number of replicates in an experiment increases the sensitivity. The type of replicate that needs to be performed depends on the experimental conditions, while the number of replicate necessary may be estimated by performing a power analysis [171, 172]. It is also important to note that proteins with a low variability can be significantly regulated, even at a very low ratio. Therefore, a second fold change threshold is often applied to account for biological significance. This is represented in a volcano plot, that shows the relation between ratio and p-value, together with the chosen thresholds (Figure 3.2 C). Time and financial constraints often limit the number of replicates that can be performed and for a single protein usually only one measurement is done. A method used to estimate the variability is to assume that most proteins are not influenced by the experimental conditions. This way the population variance is used to estimate single protein variances.

A statistical test relies on knowledge of the error of some measurements, i.e. the log-ratio. The variance for each protein can be estimated by performing multiple replicates of the same experiment. If one assumes the variance does not vary substantially between proteins, and most proteins are not regulated in the experimental circumstances, the variance can be estimated from the protein population as a whole. In Perseus, the statistical analysis tool accompanying MaxQuant, there are two ways of calculating significance based on the protein overall variation, named Significance A and Significance B [121]. Significance A calculates the variance over the whole population and performs one-sample t-tests to determine if the log-ratios are significantly different from zero. As standard deviation tends to vary with measured intensities (e.g. Figure 3.2 B), Significance B performs the analysis in based on intensity. Similarly, for reporter ion quantification-based experiments containing multiple channels the software tool Isobar can perform normalization and statistical validation [173].

### 3.5.1 Multiple Testing Correction

When considering a protein for significance based on a statistical test a so-called false positive rate (usually written as $\alpha$) is chosen: the number of times the test is expected to indicate a significant difference, when there is no real difference. Traditionally, $\alpha$ is set to 5%, indicating that if the test is performed 100 times it is expected that 5 of the tests indicate significance purely by chance. Likewise, if 5000 protein ratios are obtained in a proteomics experiment we would expect 250 significant proteins by chance. The Bonferroni correction tries to remove these false positive findings by simply dividing $\alpha$ by the number of tests, e.g. for 5000 proteins this would lead to a cut-off of 0.00001. Since this threshold is so conservative, the number of significant hits decreases dramatically. Instead, controlling the false discovery rate (FDR), i.e. allowing a fraction of the significant hits to be false, has gained popularity because of its greater sensitivity. The Benjamini-Hochberg step-up procedure is applicable for most high throughput experiments [174]. Similar to p-values resulting from statistical tests (denoting the $\alpha$ value at which a measurement would just pass as significant), the q value denotes the FDR in a multiple testing corrected context [175].

## 4 Comparison of Different Labeling Strategies

Over the past years several groups have reported on comparisons between different mass spectrometry-based quantification strategies in proteomics. Several studies have focused on head-to-head comparisons between spectral counting-based label-free quantification and metabolic labeling. These studies concordantly concluded that metabolic labeling provides better accuracy [182-184]. This holds especially for low abundant proteins for which usually only a few spectral counts are available. The occasionally reported deeper proteome coverage in label-free approaches [182, 185] is purely related to sample complexity which is two- or threefold higher in metabolic labeling. This can be overcome by using state-of-the-art MS instrumentation that provides high performance peptide separation and rapid MS/MS sequencing and is thus no systemic disadvantage. Other studies have compared isobaric tagging based approaches with label-free approaches [185-187]. Wang et al. observed that iTRAQ labeling provided more precise but less accurate data than intensity-based label-free quantification [187]. Both can be attributed to the ratio distortion caused by co-isolation of contaminating ions. Li et al. reported better accuracy and reproducibility for metabolic labeling and isobaric tagging compared to spectral counting [185].

Recently, we performed a systematic global comparison of SILAC, dimethyl labeling and isobaric tagging (TMT) [133]. We mixed differentially labeled aliquots from HeLa cells in a 1:0.5:0.1 ratio and performed a classical bottom-up proteomics experiment. This three-way

comparison allowed assessment of MS vs. MSn-based quantification as well as metabolic vs. chemical labeling and led to interesting insights into identification and quantification performance of each technique. In terms of proteome coverage all three techniques reached a similar depth level and the overlap on protein level was high. SILAC and dimethyl labeling also generated very similar results with respect to quantification precision. MS/MS-based isobaric tagging quantification was found to be heavily affected by ratio distortion through co-isolation of contaminating ions. We also employed a MS3-based method [139] to reveal the true variability in TMT quantification. Strikingly, isobaric tagging using TMT still outperformed SILAC and dimethyl labeling in terms of precision. However, this came at the cost of reduced sensitivity and duty cycle speed which led to a smaller number of quantified proteins and consequently to a less comprehensive proteome coverage. The results of this study can be found in chapter 6 of this thesis.

## 4.1 Considerations for Selecting the Most Appropriate Quantification Strategy

The ideal quantitative proteomics approach would enable reproducible, comprehensive, sensitive and unbiased analysis that provides accurate and precise quantitative data with a high dynamic range and within reasonable analysis time. However, such a single technique does not exist. All aforementioned techniques have their strengths and weaknesses, thus, choosing the most appropriate quantification strategy depends on several factors: (1) the biological research question that determines the experimental design, (2) the number and complexity of the samples as well as the total protein amount available per sample, and (3) the LC-MS instrumentation available in the lab.

The main benefit of metabolic labeling is the possibility to combine differentially labeled samples at the earliest possible stage during sample preparation. This eliminates non-systematic errors that can potentially occur at any step in the workflow. Thus, if metabolic labeling is possible it should always be considered first. $^{15}N$ or SILAC-based approaches are the methods of choice when sample preparation includes multiple steps or extensive protein fractionation such as in affinity-purification based experiments [188] or subcellular organelle enrichment [189]. Chemical labeling strategies such as dimethyl labeling [116] or mTRAQ [138] are preferred for tissue samples, body fluids, primary cell culture or cells that are not auxotroph for amino acids used in SILAC. Label-free or isobaric tagging are preferable quantification strategies for experiments that necessitate the analysis of multiple samples. Label-free quantification theoretically allows for an unlimited number of samples to be compared and is ideally employed when dealing with medium-complex samples such as in pull-down experiments [190] that can be analyzed using short LC gradients. Highly

complex samples such as whole cell or tissue lysates usually require peptide fractionation [65] prior to LC-MS/MS analysis. Label-free quantification of multiple fractions per sample is not recommended because it is time-consuming and also suffers from complicated data-analysis since peptides can elute in more than one fraction. Isobaric tagging-based labeling approaches also feature multiplexing capabilities and provide better precision than label-free approaches due to reduced technical variability and MSn-level based quantification. TMT and iTRAQ allow analysis of four to eight samples simultaneously in a single run which makes them attractive when many samples have to be screened in a short time. Furthermore, in cases where the sample amount is limited isobaric tagging based approaches are beneficial since the peptide signal in the survey MS is stacked up which improves sensitivity.

At last, one also has to take into consideration that each quantification strategy imposes special demands on instrumentation. For example, label-free quantification approaches are based on robust and highly reproducible chromatography. Isobaric tagging-based approaches require high-end mass spectrometers that are capable of detecting the reporter ions in the low mass region, such as Orbitrap [65, 72] or TripleToF [191] instruments. All of these aspects should be carefully considered to select the most appropriate quantification strategy. Many laboratories have specialized in performing one of them and in our opinion it is difficult to employ all of them at the same standard and with a similar level of expertise. Interestingly, a recent study by the Association of Biomolecular Resource Facilities (ABRF) revealed that the level of experience of the lab which performs the experiments is the most important factor that determines the quality of quantitative proteomics experiments [192].

## References

[1] M.R. Wilkins, C. Pasquali, R.D. Appel, K. Ou, O. Golaz, J.C. Sanchez, J.X. Yan, A.A. Gooley, G. Hughes, I. Humphery-Smith, K.L. Williams, D.F. Hochstrasser, From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis, Biotechnology (N Y), 14 (1996) 61-65.

[2] P. James, Protein identification in the post-genome era: the rapid rise of proteomics, Q Rev Biophys, 30 (1997) 279-331.

[3] B.T. Chait, Mass spectrometry in the postgenomic era, Annu Rev Biochem, 80 (2011) 239-246.

[4] R. Aebersold, M. Mann, Mass spectrometry-based proteomics, Nature, 422 (2003) 198-207.

[5] A. Bensimon, A.J.R. Heck, R. Aebersold, Mass Spectrometry-Based Proteomics and Network Biology, Annual Review of Biochemistry, Vol 81, 81 (2012) 379-405.

[6] A.F.M. Altelaar, J. Munoz, A.J.R. Heck, Next-generation proteomics: towards an integrative view of proteome dynamics, Nat Rev Genet, 14 (2013) 35-48.

[7] A.F.M. Altelaar, A.J.R. Heck, Trends in ultrasensitive proteomics, Curr Opin Chem Biol, 16 (2012) 206-213.

[8] M. Mann, N.A. Kulak, N. Nagaraj, J. Cox, The coming age of complete, accurate, and ubiquitous proteomes, Mol Cell, 49 (2013) 583-590.

[9] L. Sleno, D.A. Volmer, Ion activation methods for tandem mass spectrometry, Journal of Mass Spectrometry, 39 (2004) 1091-1112.

[10] A.W. Jones, H.J. Cooper, Dissociation techniques in mass spectrometry-based proteomics, Analyst, 136 (2011) 3419-3429.

[11] P. Roepstorff, J. Fohlman, Proposal for a Common Nomenclature for Sequence Ions in Mass-Spectra of Peptides, Biomed Mass Spectrom, 11 (1984) 601-601.

[12] D.F. Hunt, J.R. Yates, 3rd, J. Shabanowitz, S. Winston, C.R. Hauer, Protein sequencing by tandem mass spectrometry, Proc Natl Acad Sci U S A, 83 (1986) 6233-6237.

[13] S.A. Mcluckey, Principles of Collisional Activation in Analytical Mass-Spectrometry, J Am Soc Mass Spectr, 3 (1992) 599-614.

[14] V.H. Wysocki, G. Tsaprailis, L.L. Smith, L.A. Breci, Special feature: Commentary - Mobile and localized protons: a framework for understanding peptide dissociation, Journal of Mass Spectrometry, 35 (2000) 1399-1406.

[15] B. Paizs, S. Suhai, Fragmentation pathways of protonated peptides, Mass Spectrometry Reviews, 24 (2005) 508-548.

[16] Y.Y. Huang, J.M. Triscari, G.C. Tseng, L. Pasa-Tolic, M.S. Lipton, R.D. Smith, V.H. Wysocki, Statistical characterization of the charge state and residue dependence of low-energy CID peptide dissociation patterns, Analytical Chemistry, 77 (2005) 5800-5813.

[17] C. Cunningham, G.L. Glish, D.J. Burinsky, High amplitude short time excitation: A method to form and detect low mass product ions in a quadrupole ion trap mass spectrometer, J Am Soc Mass Spectr, 17 (2006) 81-84.

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

[19] P.J. Boersema, S. Mohammed, A.J. Heck, Phosphopeptide fragmentation and analysis by mass spectrometry, Journal of mass spectrometry : JMS, 44 (2009) 861-878.

[20] R.A. Zubarev, N.L. Kelleher, F.W. McLafferty, Electron capture dissociation of multiply charged protein cations. A nonergodic process, J Am Chem Soc, 120 (1998) 3265-3266.

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

[22] A. Sawicka, P. Skurski, R.R. Hudgins, J. Simons, Model calculations relevant to disulfide bond cleavage via electron capture influenced by positively charged groups, J Phys Chem B, 107 (2003) 13505-13511.

[23] E.A. Syrstad, F. Turecek, Toward a general mechanism of electron capture dissociation, J Am Soc Mass Spectr, 16 (2005) 208-224.

[24] I. Swierszcz, P. Skurski, J. Simons, Dipole and Coulomb Forces in Electron Capture Dissociation and Electron Transfer Dissociation Mass Spectroscopy, J Phys Chem A, 116 (2012) 1828-1837.

[25] K.O. Zhurov, L. Fornelli, M.D. Wodrich, U.A. Laskay, Y.O. Tsybin, Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis, Chem Soc Rev, (2013).

[26] H. Ben Hamidane, D. Chiappe, R. Hartmer, A. Vorobyev, M. Moniatte, Y.O. Tsybin, Electron cap-

ture and transfer dissociation: Peptide structure analysis at different ion internal energy levels, J Am Soc Mass Spectr, 20 (2009) 567-575.

[27] P.B. O'Connor, C. Lin, J.J. Cournoyer, J.L. Pittman, M. Belyayev, B.A. Budnik, Long-lived electron capture dissociation product ions experience radical migration via hydrogen abstraction, J Am Soc Mass Spectr, 17 (2006) 576-585.

[28] D.L. Swaney, G.C. McAlister, M. Wirtala, J.C. Schwartz, J.E. Syka, J.J. Coon, Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors, Anal Chem, 79 (2007) 477-485.

[29] A.R. Ledvina, G.C. McAlister, M.W. Gardner, S.I. Smith, J.A. Madsen, J.C. Schwartz, G.C. Stafford, J.E.P. Syka, J.S. Brodbelt, J.J. Coon, Infrared Photoactivation Reduces Peptide Folding and Hydrogen-Atom Migration following ETD Tandem Mass Spectrometry, Angew Chem Int Edit, 48 (2009) 8526-8528.

[30] R.X. Sun, M.Q. Dong, C.Q. Song, H. Chi, B. Yang, L.Y. Xiu, L. Tao, Z.Y. Jing, C. Liu, L.H. Wang, Y. Fu, S.M. He, Improved peptide identification for proteomic analysis based on comprehensive characterization of electron transfer dissociation spectra, J Proteome Res, 9 (2010) 6354-6367.

[31] D.M. Good, M. Wirtala, G.C. McAlister, J.J. Coon, Performance characteristics of electron transfer dissociation mass spectrometry, Mol Cell Proteomics, 6 (2007) 1942-1951.

[32] S.J. Pitteri, P.A. Chrisman, S.A. McLuckey, Electron-transfer ion/ion reactions of doubly protonated peptides: Effect of elevated bath gas temperature, Analytical Chemistry, 77 (2005) 5662-5669.

[33] D.M. Horn, Y. Ge, F.W. McLafferty, Activated ion electron capture dissociation for mass spectral sequencing of larger (42 kDa) proteins, Analytical Chemistry, 72 (2000) 4778-4784.

[34] D.M. Horn, K. Breuker, A.J. Frank, F.W. McLafferty, Kinetic intermediates in the folding of gaseous protein ions characterized by electron capture dissociation mass spectrometry, J Am Chem Soc, 123 (2001) 9792-9799.

[35] H. Molina, R. Matthiesen, K. Kandasamy, A. Pandey, Comprehensive comparison of collision induced dissociation and electron transfer dissociation, Anal Chem, 80 (2008) 4825-4835.

[36] C.K. Frese, A.F. Altelaar, M.L. Hennrich, D. Nolting, M. Zeller, J. Griep-Raming, A.J. Heck, S. Mohammed, Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos, Journal of proteome research, 10 (2011) 2377-2388.

[37] J.J. Coon, Collisions or Electrons? Protein Sequence Analysis in the 21st Century, Analytical Chemistry, 81 (2009) 3208-3215.

[38] R.A. Zubarev, A.R. Zubarev, M.M. Savitski, Electron capture/transfer versus collisionally activated/induced dissociations: Solo or duet?, J Am Soc Mass Spectr, 19 (2008) 753-761.

[39] L.M. Mikesh, B. Ueberheide, A. Chi, J.J. Coon, J.E. Syka, J. Shabanowitz, D.F. Hunt, The utility of ETD mass spectrometry in proteomic analysis, Biochim Biophys Acta, 1764 (2006) 1811-1822.

[40] J. Wiesner, T. Premsler, A. Sickmann, Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications, Proteomics, 8 (2008) 4466-4483.

[41] H.W.P. van den Toorn, S. Mohammed, J.W. Gouw, B. van Breukelen, A.J.R. Heck, Targeted SCX Based Peptide Fractionation for Optimal Sequencing by Collision Induced, and Electron

Transfer Dissociation, J Proteomics Bioinform, 1 (2008) 379-388.

[42] L. Fornelli, E. Damoc, P.M. Thomas, N.L. Kelleher, K. Aizikov, E. Denisov, A. Makarov, Y.O. Tsybin, Top-down analysis of monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS, Molecular & cellular proteomics : MCP, (2012).

[43] A. Makarov, Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis, Analytical chemistry, 72 (2000) 1156-1162.

[44] K.H. Kingdon, A method for the neutralization of electron space charge by positive ionization at very low gas pressures, Phys Rev, 21 (1923) 408-418.

[45] A. Makarov, E. Denisov, Dynamics of ions of intact proteins in the Orbitrap mass analyzer, J Am Soc Mass Spectr, 20 (2009) 1486-1495.

[46] M.W. Senko, J.D. Canterbury, S.H. Guan, A.G. Marshall, A high-performance modular data system for Fourier transform ion cyclotron resonance mass spectrometry, Rapid Commun Mass Sp, 10 (1996) 1839-1844.

[47] J.A. Marto, A.G. Marshall, L. Schweikhard, A 2-Electrode Ion-Trap for Fourier-Transform Ion-Cyclotron Resonance Mass-Spectrometry, International Journal of Mass Spectrometry and Ion Processes, 137 (1994) 9-30.

[48] Q.Z. Hu, R.J. Noll, H.Y. Li, A. Makarov, M. Hardman, R.G. Cooks, The Orbitrap: a new mass spectrometer, Journal of Mass Spectrometry, 40 (2005) 430-443.

[49] A. Makarov, E. Denisov, O. Lange, S. Horning, Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer, J Am Soc Mass Spectrom, 17 (2006) 977-982.

[50] A. Makarov, E. Denisov, A. Kholomeev, W. Balschun, O. Lange, K. Strupat, S. Horning, Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer, Anal Chem, 78 (2006) 2113-2120.

[51] M. Hardman, A.A. Makarov, Interfacing the orbitrap mass analyzer to an electrospray ion source, Analytical chemistry, 75 (2003) 1699-1705.

[52] J.V. Olsen, L.M. de Godoy, G. Li, B. Macek, P. Mortensen, R. Pesch, A. Makarov, O. Lange, S. Horning, M. Mann, Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap, Molecular & cellular proteomics : MCP, 4 (2005) 2010-2021.

[53] Y. Zhang, S.B. Ficarro, S. Li, J.A. Marto, Optimized Orbitrap HCD for quantitative analysis of phosphopeptides, J Am Soc Mass Spectrom, 20 (2009) 1425-1434.

[54] N. Mischerikow, P. van Nierop, K.W. Li, H.G. Bernstein, A.B. Smit, A.J. Heck, A.F. Altelaar, Gaining efficiency by parallel quantification and identification of iTRAQ-labeled peptides using HCD and decision tree guided CID/ETD on an LTQ Orbitrap, Analyst, 135 (2010) 2643-2652.

[55] T. Kocher, P. Pichler, M. Schutzbier, C. Stingl, A. Kaul, N. Teucher, G. Hasenfuss, J.M. Penninger, K. Mechtler, High precision quantitative proteomics using iTRAQ on an LTQ Orbitrap: a new mass spectrometric method combining the benefits of all, J Proteome Res, 8 (2009) 4743-4752.

[56] G.C. McAlister, D. Phanstiel, D.M. Good, W.T. Berggren, J.J. Coon, Implementation of electron-transfer dissociation on a hybrid linear ion trap-orbitrap mass spectrometer, Analytical Chemistry, 79 (2007) 3525-3534.

[57] D.K. Williams, Jr., G.C. McAlister, D.M. Good, J.J. Coon, D.C. Muddiman, Dual electrospray ion source for electron-transfer dissociation on a hybrid linear ion trap-orbitrap mass spectrometer, Analytical chemistry, 79 (2007) 7916-7919.

[58] G.C. McAlister, W.T. Berggren, J. Griep-Raming, S. Horning, A. Makarov, D. Phanstiel, G. Stafford, D.L. Swaney, J.E. Syka, V. Zabrouskov, J.J. Coon, A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer, J Proteome Res, 7 (2008) 3127-3136.

[59] J.M. Hogan, S.J. Pitteri, P.A. Chrisman, S.A. McLuckey, Complementary structural information from a tryptic N-linked glycopeptide via electron transfer ion/ion reactions and collision-induced

dissociation, Journal of Proteome Research, 4 (2005) 628-632.

[60] D.L. Swaney, G.C. McAlister, J.J. Coon, Decision tree-driven tandem mass spectrometry for shotgun proteomics, Nat Methods, 5 (2008) 959-964.

[61] J.V. Olsen, J.C. Schwartz, J. Griep-Raming, M.L. Nielsen, E. Damoc, E. Denisov, O. Lange, P. Remes, D. Taylor, M. Splendore, E.R. Wouters, M. Senko, A. Makarov, M. Mann, S. Horning, A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed, Mol Cell Proteomics, 8 (2009) 2759-2769.

[62] T.P. Second, J.D. Blethrow, J.C. Schwartz, G.E. Merrihew, M.J. MacCoss, D.L. Swaney, J.D. Russell, J.J. Coon, V. Zabrouskov, Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures, Anal Chem, 81 (2009) 7757-7765.

[63] P. Pichler, T. Kocher, J. Holzmann, T. Mohring, G. Ammerer, K. Mechtler, Improved Precision of iTRAQ and TMT Quantification by an Axial Extraction Field in an Orbitrap HCD Cell, Analytical Chemistry, 83 (2011) 1469-1474.

[64] C.K. Frese, M. Altelaar, H.W. van den Toorn, D. Nolting, J. Griep-Raming, A.J. Heck, S. Mohammed, Towards Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry, Analytical chemistry, (2012).

[65] A. Michalski, E. Damoc, O. Lange, E. Denisov, D. Nolting, M. Muller, R. Viner, J. Schwartz, P. Remes, M. Belford, J.J. Dunyach, J. Cox, S. Horning, M. Mann, A. Makarov, Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes, Molecular & cellular proteomics : MCP, 11 (2012) O111 013698.

[66] A. Wieghaus, A. Makarov, U. Froehlich, M. Kellmann, E. Denisov, O. Lange, Development and Application of a New Benchtop Orbitrap Mass Spectrometer, Proceedings of the 56th ASMS Conference on Mass Spectrometry and Allied Topics, Denver, CO (2008).

[67] K.P. Bateman, M. Kellmann, H. Muenster, R. Papp, L. Taylor, Quantitative-qualitative data acquisition using a benchtop Orbitrap mass spectrometer, J Am Soc Mass Spectr, 20 (2009) 1441-1450.

[68] T. Geiger, J. Cox, M. Mann, Proteomics on an Orbitrap Benchtop Mass Spectrometer Using All-ion Fragmentation, Molecular & Cellular Proteomics, 9 (2010) 2252-2261.

[69] J.C. Silva, M.V. Gorenstein, G.Z. Li, J.P. Vissers, S.J. Geromanos, Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition, Molecular & cellular proteomics : MCP, 5 (2006) 144-156.

[70] R.J. Rose, E. Damoc, E. Denisov, A. Makarov, A.J. Heck, High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies, Nature methods, 9 (2012) 1084-1086.

[71] S. Rosati, R.J. Rose, N.J. Thompson, E. van Duijn, E. Damoc, E. Denisov, A. Makarov, A.J. Heck, Exploring an orbitrap analyzer for the characterization of intact antibodies by native mass spectrometry, Angew Chem Int Ed Engl, 51 (2012) 12992-12996.

[72] A. Michalski, E. Damoc, J.P. Hauschild, O. Lange, A. Wieghaus, A. Makarov, N. Nagaraj, J. Cox, M. Mann, S. Horning, Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer, Molecular & cellular proteomics : MCP, 10 (2011) M111 011015.

[73] A.C. Peterson, J.D. Russell, D.J. Bailey, M.S. Westphall, J.J. Coon, Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics, Molecular & cellular proteomics : MCP, 11 (2012) 1475-1488.

[74] A. Altelaar, J. Munoz, A. Heck, Next-generation proteomics: towards an integrative view of proteome dynamics, Nature Reviews Genetics, In Press (in press).

[75] V. Brun, C. Masselon, J. Garin, A. Dupuis, Isotope dilution strategies for absolute quantitative proteomics, J Proteomics, 72 (2009) 740-749.

[76] J.K. Eng, A.L. Mccormack, J.R. Yates, An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database, J Am Soc Mass Spectr, 5 (1994) 976-989.

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

[78] J.W. Gouw, J. Krijgsveld, A.J.R. Heck, Quantitative Proteomics by Metabolic Labeling of Model Organisms, Molecular & Cellular Proteomics, 9 (2010) 11-24.

[79] Y. Oda, K. Huang, F.R. Cross, D. Cowburn, B.T. Chait, Accurate quantitation of protein expression and site-specific phosphorylation, P Natl Acad Sci USA, 96 (1999) 6591-6596.

[80] M.P. Washburn, R. Ulaszek, C. Deciu, D.M. Schieltz, J.R. Yates, 3rd, Analysis of quantitative proteomic data generated via multidimensional protein identification technology, Analytical chemistry, 74 (2002) 1650-1657.

[81] L. Pasa-Tolic, P.K. Jensen, G.A. Anderson, M.S. Lipton, K.K. Peden, S. Martinovic, N. Tolic, J.E. Bruce, R.D. Smith, High throughput proteome-wide precision measurements of protein expression using mass spectrometry, J Am Chem Soc, 121 (1999) 7949-7950.

[82] J. Krijgsveld, R.F. Ketting, T. Mahmoudi, J. Johansen, M. Artal-Sanz, C.P. Verrijzer, R.H.A. Plasterk, A.J.R. Heck, Metabolic labeling of C-elegans and D-melanogaster for quantitative proteomics, Nature Biotechnology, 21 (2003) 927-931.

[83] C.C. Wu, M.J. MacCoss, K.E. Howell, D.E. Matthews, J.R. Yates, Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis, Analytical Chemistry, 76 (2004) 4951-4959.

[84] M.K. Doherty, C. Whitehead, H. McCormack, S.J. Gaskell, R.J. Beynon, Proteome dynamics in complex organisms: using stable isotopes to monitor individual protein turnover rates, Proteomics, 5 (2005) 522-533.

[85] J.E. Schaff, F. Mbeunkui, K. Blackburn, D.M. Bird, M.B. Goshe, SILIP: a novel stable isotope labeling method for in planta quantitative proteomic analysis, Plant J, 56 (2008) 840-854.

[86] E.L. Huttlin, A.D. Hegeman, A.C. Harms, M.R. Sussman, Comparison of full versus partial metabolic labeling for quantitative proteomics analysis in Arabidopsis thaliana, Molecular & Cellular Proteomics, 6 (2007) 860-881.

[87] R. Hebeler, S. Oeljeklaus, K.E. Reidegeld, M. Eisenacher, C. Stephan, B. Sitek, K. Stuhler, H.E. Meyer, M.J.G. Sturre, P.P. Dijkwel, B. Warscheid, Study of early leaf senescence in Arabidopsis thaliana by quantitative proteomics using reciprocal N-14/N-15 Labeling and difference gel electrophoresis, Molecular & Cellular Proteomics, 7 (2008) 108-120.

[88] C.J. Nelson, E.L. Huttlin, A.D. Hegeman, A.C. Harms, M.R. Sussman, Implications of N-15-metabolic labeling for automated peptide identification in Arabidopsis thaliana, Proteomics, 7 (2007) 1279-1292.

[89] B. Arsova, S. Kierszniowska, W.X. Schulze, The use of heavy nitrogen in quantitative proteomics experiments in plants, Trends Plant Sci, 17 (2012) 102-112.

[90] R.D. Evans, Physiological mechanisms influencing plant nitrogen isotope composition, Trends Plant Sci, 6 (2001) 121-126.

[91] S.E. Ong, M. Mann, Mass spectrometry-based proteomics turns quantitative, Nat Chem Biol, 1

(2005) 252-262.

[92] S.E. Ong, B. Blagoev, I. Kratchmarova, D.B. Kristensen, H. Steen, A. Pandey, M. Mann, Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics, Molecular & Cellular Proteomics, 1 (2002) 376-386.

[93] T.C. Hunter, L. Yang, H. Zhu, V. Majidi, E.M. Bradbury, X. Chen, Peptide mass mapping constrained with stable isotope-tagged peptides for identification of protein mixtures, Analytical chemistry, 73 (2001) 4891-4902.

[94] B. Blagoev, S.E. Ong, I. Kratchmarova, M. Mann, Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics, Nature biotechnology, 22 (2004) 1139-1145.

[95] S.J. Berger, S.W. Lee, G.A. Anderson, L. Pasa-Tolic, N. Tolic, Y. Shen, R. Zhao, R.D. Smith, High-throughput global peptide proteomic analysis by combining stable isotope amino acid labeling and data-dependent multiplexed-MS/MS, Analytical chemistry, 74 (2002) 4994-5000.

[96] D. Van Hoof, M.W.H. Pinkse, D.W.-V. Oostwaard, C.L. Mummery, A.J.R. Heck, J. Krijgsveld, An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics, Nat. Meth., 4 (2007) 677-678.

[97] S.E. Ong, I. Kratchmarova, M. Mann, Properties of C-13-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC), Journal of Proteome Research, 2 (2003) 173-181.

[98] S.C. Bendall, C. Hughes, M.H. Stewart, B. Doble, M. Bhatia, G.A. Lajoie, Prevention of amino acid conversion in SILAC experiments with embryonic stem cells, Molecular & Cellular Proteomics, 7 (2008) 1587-1597.

[99] S.K. Park, L.J. Liao, J.Y. Kim, J.R. Yates, A computational approach to correct arginine-to-proline conversion in quantitative proteomics, Nature Methods, 6 (2009) 184-185.

[100] M.D. Sury, J.X. Chen, M. Selbach, The SILAC Fly Allows for Accurate Protein Quantification in Vivo, Molecular & Cellular Proteomics, 9 (2010) 2173-2183.

[101] M. Kruger, M. Moser, S. Ussar, I. Thievessen, C.A. Luber, F. Forner, S. Schmidt, S. Zanivan, R. Fassler, M. Mann, SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function, Cell, 134 (2008) 353-364.

[102] A. Scholten, S. Mohammed, T.Y. Low, S. Zanivan, T.A.B. van Veen, B. Delanghe, A.J.R. Heck, In-depth Quantitative Cardiac Proteomics Combining Electron Transfer Dissociation and the Metalloendopeptidase Lys-N with the SILAC Mouse, Molecular & Cellular Proteomics, 10 (2011).

[103] D.M. Walther, M. Mann, Accurate Quantification of More Than 4000 Mouse Tissue Proteins Reveals Minimal Proteome Changes During Aging, Molecular & Cellular Proteomics, 10 (2011).

[104] E. Milner, E. Barnea, I. Beer, A. Admon, The turnover kinetics of major histocompatibility complex peptides of human cancer cells, Molecular & Cellular Proteomics, 5 (2006) 357-365.

[105] B. Schwanhausser, M. Gossen, G. Dittmar, M. Selbach, Global analysis of cellular protein translation by pulsed SILAC, Proteomics, 9 (2009) 205-209.

[106] T. Geiger, J. Cox, P. Ostasiewicz, J.R. Wisniewski, M. Mann, Super-SILAC mix for quantitative proteomics of human tumor tissue, Nature Methods, 7 (2010) 383-U364.

[107] T. Geiger, J.R. Wisniewski, J. Cox, S. Zanivan, M. Kruger, Y. Ishihama, M. Mann, Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics, Nat Protoc, 6 (2011) 147-157.

[108] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, Nature Biotechnology, 17 (1999) 994-999.

[109] T.P.J. Dunkley, P. Dupree, R.B. Watson, K.S. Lilley, The use of isotope-coded affinity tags (ICAT) to study organelle proteomes in Arabidopsis thaliana, Biochem Soc T, 32 (2004) 520-523.

[110] J.E.P. Syka, J.A. Marto, D.L. Bai, S. Horning, M.W. Senko, J.C. Schwartz, B. Ueberheide, B. Garcia, S. Busby, T. Muratore, J. Shabanowitz, D.F. Hunt, Novel linear quadrupole ion trap/FT mass spectrometer: Performance characterization and use in the comparative analysis of histone H3 post-translational modifications, Journal of Proteome Research, 3 (2004) 621-626.

[111] D.R. Goodlett, A. Keller, J.D. Watts, R. Newitt, E.C. Yi, S. Purvine, J.K. Eng, P. von Haller, R. Aebersold, E. Kolker, Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation, Rapid Commun Mass Sp, 15 (2001) 1214-1221.

[112] F.E. Regnier, S. Julka, Primary amine coding as a path to comparative proteomics, Proteomics, 6 (2006) 3968-3979.

[113] P.J. Boersema, T.T. Aye, T.A. van Veen, A.J. Heck, S. Mohammed, Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates, Proteomics, 8 (2008) 4624-4632.

[114] J.L. Hsu, S.Y. Huang, N.H. Chow, S.H. Chen, Stable-isotope dimethyl labeling for quantitative proteomics, Analytical Chemistry, 75 (2003) 6843-6852.

[115] S.Y. Huang, M.L. Tsai, C.J. Wu, J.L. Hsu, S.H. Ho, S.H. Chen, Quantitation of protein phosphorylation in pregnant rat uteri using stable isotope dimethyl labeling coupled with IMAC, Proteomics, 6 (2006) 1722-1734.

[116] P.J. Boersema, R. Raijmakers, S. Lemeer, S. Mohammed, A.J.R. Heck, Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics, Nat Protoc, 4 (2009) 484-494.

[117] V.M. Ding, P.J. Boersema, L.Y. Foong, C. Preisinger, G. Koh, S. Natarajan, D.Y. Lee, J. Boekhorst, B. Snel, S. Lemeer, A.J. Heck, A. Choo, Tyrosine phosphorylation profiling in FGF-2 stimulated human embryonic stem cells, PLoS One, 6 (2011) e17538.

[118] R. Raijmakers, A.J.R. Heck, S. Mohammed, Assessing biological variation and protein processing in primary human leukocytes by automated multiplex stable isotope labeling coupled to 2 dimensional peptide separation, Mol Biosyst, 5 (2009) 992-1003.

[119] R. Raijmakers, C.R. Berkers, A. de Jong, H. Ovaa, A.J. Heck, S. Mohammed, Automated online sequential isotope labeling for protein quantitation applied to proteasome tissue-specific diversity, Molecular & cellular proteomics : MCP, 7 (2008) 1755-1762.

[120] R. Zhang, C.S. Sioma, S. Wang, F.E. Regnier, Fractionation of isotopically labeled peptides in quantitative proteomics, Analytical chemistry, 73 (2001) 5142-5149.

[121] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nature biotechnology, 26 (2008) 1367-1372.

[122] C. Ji, L. Li, Quantitative proteome analysis using differential stable isotopic labeling and microbore LC-MALDI MS and MS/MS, Journal of proteome research, 4 (2005) 734-742.

[123] D. Kovanich, S. Cappadona, R. Raijmakers, S. Mohammed, A. Scholten, A.J. Heck, Applications of stable isotope dimethyl labeling in quantitative proteomics, Anal Bioanal Chem, (2012).

[124] M. Munchbach, M. Quadroni, G. Miotto, P. James, Quantitation and facilitated de novo sequencing of proteins by isotopic N-terminal labeling of peptides with a fragmentation directing moiety, Analytical Chemistry, 72 (2000) 4047-4057.

[125] A. Thompson, J. Schafer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, C. Hamon,

Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS, Analytical Chemistry, 75 (2003) 1895-1904.

[126] P.L. Ross, Y.L.N. Huang, J.N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, D.J. Pappin, Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents, Molecular & Cellular Proteomics, 3 (2004) 1154-1169.

[127] N. Dephoure, S.P. Gygi, Hyperplexing: A Method for Higher-Order Multiplexed Quantitative Proteomics Provides a Map of the Dynamic Response to Rapamycin in Yeast, Sci Signal, 5 (2012).

[128] D. Phanstiel, Y. Zhang, J.A. Marto, J.J. Coon, Peptide and protein quantification using iTRAQ with electron transfer dissociation, J Am Soc Mass Spectr, 19 (2008) 1255-1262.

[129] R.I. Viner, T. Zhang, T. Second, V. Zabrouskov, Quantification of post-translationally modified peptides of bovine alpha-crystallin using tandem mass tags and electron transfer dissociation, J Proteomics, 72 (2009) 874-885.

[130] V. Tambor, C.L. Hunter, S.L. Seymour, M. Kacerovsky, J. Stulik, J. Lenco, CysTRAQ - A combination of iTRAQ and enrichment of cysteinyl peptides for uncovering and quantifying hidden proteomes, J Proteomics, 75 (2012) 857-867.

[131] L. Dayon, A. Hainard, V. Licker, N. Turck, K. Kuhn, D.F. Hochstrasser, P.R. Burkhard, J.C. Sanchez, Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags, Analytical Chemistry, 80 (2008) 2921-2931.

[132] M. Bantscheff, M. Boesche, D. Eberhard, T. Matthieson, G. Sweetman, B. Kuster, Robust and sensitive iTRAQ quantification on an LTQ orbitrap mass spectrometer, Molecular & Cellular Proteomics, 7 (2008) 1702-1713.

[133] A.F. Altelaar, C.K. Frese, C. Preisinger, M.L. Hennrich, A.W. Schram, H.T. Timmers, A.J. Heck, S. Mohammed, Benchmarking stable isotope labeling based quantitative proteomics, J Proteomics, (2012).

[134] S.Y. Ow, M. Salim, J. Noirel, C. Evans, I. Rehman, P.C. Wright, iTRAQ Underestimation in Simple and Complex Mixtures: "The Good, the Bad and the Ugly", Journal of Proteome Research, 8 (2009) 5347-5355.

[135] A.L. Christoforou, K.S. Lilley, Isobaric tagging approaches in quantitative proteomics: the ups and downs, Anal Bioanal Chem, 404 (2012) 1029-1037.

[136] D.W. Mahoney, T.M. Therneau, C.J. Heppelmann, L. Higgins, L.M. Benson, R.M. Zenka, P. Jagtap, G.L. Nelsestuen, H.R. Bergen, A.L. Oberg, Relative Quantification: Characterization of Bias, Variability and Fold Changes in Mass Spectrometry Data from iTRAQ-Labeled Peptides, Journal of Proteome Research, 10 (2011) 4325-4333.

[137] S.Y. Ow, M. Salim, J. Noirel, C. Evans, P.C. Wright, Minimising iTRAQ ratio compression through understanding LC-MS elution dependence and high-resolution HILIC fractionation, Proteomics, 11 (2011) 2341-2346.

[138] P. Mertins, N.D. Udeshi, K.R. Clauser, D.R. Mani, J. Patel, S.E. Ong, J.D. Jaffe, S.A. Carr, iTRAQ Labeling is Superior to mTRAQ for Quantitative Global Proteomics and Phosphoproteomics, Molecular & Cellular Proteomics, 11 (2012).

[139] L. Ting, R. Rad, S.P. Gygi, W. Haas, MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics, Nature Methods, 8 (2011) 937-940.

[140] C.D. Wenger, M.V. Lee, A.S. Hebert, G.C. McAlister, D.H. Phanstiel, M.S. Westphall, J.J. Coon, Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging,

Nature Methods, 8 (2011) 933-935.

[141] P.V. Bondarenko, D. Chelius, T.A. Shaler, Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry, Analytical Chemistry, 74 (2002) 4741-4749.

[142] D. Chelius, P.V. Bondarenko, Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry, Journal of Proteome Research, 1 (2002) 317-323.

[143] H.B. Liu, R.G. Sadygov, J.R. Yates, A model for random sampling and estimation of relative protein abundance in shotgun proteomics, Analytical Chemistry, 76 (2004) 4193-4201.

[144] W.M. Old, K. Meyer-Arendt, L. Aveline-Wolf, K.G. Pierce, A. Mendoza, J.R. Sevinsky, K.A. Resing, N.G. Ahn, Comparison of label-free methods for quantifying human proteins by shotgun proteomics, Molecular & Cellular Proteomics, 4 (2005) 1487-1502.

[145] H. Steen, A. Pandey, Proteomics goes quantitative: measuring protein abundance, Trends in Biotechnology, 20 (2002) 361-364.

[146] S. Cappadona, P.R. Baker, P.R. Cutillas, A.J. Heck, B. van Breukelen, Current challenges in software solutions for mass spectrometry-based quantitative proteomics, Amino Acids, 43 (2012) 1087-1108.

[147] M.J. MacCoss, C.C. Wu, H. Liu, R. Sadygov, J.R. Yates, 3rd, A correlation algorithm for the automated quantitative analysis of shotgun proteomics data, Anal Chem, 75 (2003) 6912-6921.

[148] S.K. Park, J.D. Venable, T. Xu, J.R. Yates, 3rd, A quantitative analysis software tool for mass spectrometry-based proteomics, Nat Methods, 5 (2008) 319-322.

[149] Z. Khan, J.S. Bloom, B.A. Garcia, M. Singh, L. Kruglyak, Protein quantification across hundreds of experimental conditions, (2009).

[150] Q. Eastman, Space-partitioning speeds up data processing - Journal of Proteome Research (ACS Publications), in, (Web): September 8, 2009.

[151] P. Mortensen, J.W. Gouw, J.V. Olsen, S.E. Ong, K.T.G. Rigbolt, J. Bunkenborg, J. Cox, L.J. Foster, A.J.R. Heck, B. Blagoev, J.S. Andersen, M. Mann, MSQuant, an Open Source Platform for Mass Spectrometry-Based Quantitative Proteomics, Journal of Proteome Research, 9 (2010) 393-403.

[152] J.W. Gouw, J. Krijgsveld, MSQuant: a platform for stable isotope-based quantitative proteomics, Methods Mol Biol, 893 (2012) 511-522.

[153] S. Cappadona, J. Munoz, W.P. Spee, T.Y. Low, S. Mohammed, B. van Breukelen, A.J. Heck, Deconvolution of overlapping isotopic clusters improves quantification of stable isotope-labeled peptides, J Proteomics, 74 (2011) 2204-2209.

[154] C. Christin, H.C. Hoefsloot, A.K. Smilde, F. Suits, R. Bischoff, P.L. Horvatovich, Time alignment algorithms based on selected mass traces for complex LC-MS data, J Proteome Res, 9 (2010) 1483-1495.

[155] G.L. Finney, A.R. Blackler, M.R. Hoopmann, J.D. Canterbury, C.C. Wu, M.J. MacCoss, Label-free comparative analysis of proteomics mixtures using chromatographic alignment of high-resolution muLC-MS data, Anal Chem, 80 (2008) 961-971.

[156] R.D. Smith, G.A. Anderson, M.S. Lipton, L. Pasa-Tolic, Y. Shen, T.P. Conrads, T.D. Veenstra, H.R. Udseth, An accurate mass tag strategy for quantitative and high-throughput proteome measurements, Proteomics, 2 (2002) 513-523.

[157] N.M. Griffin, J.Y. Yu, F. Long, P. Oh, S. Shore, Y. Li, J.A. Koziol, J.E. Schnitzer, Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis, Nature Biotechnology,

28 (2010) 83-U116.

[158] J. Rappsilber, U. Ryder, A.I. Lamond, M. Mann, Large-Scale Proteomic Analysis of the Human Spliceosome, Genome Research, 12 (2002) 1231-1245.

[159] T.T. Aye, A. Scholten, N. Taouatas, A. Varro, T.A.B. Van Veen, M.A. Vos, A.J.R. Heck, Proteome-wide protein concentrations in the human heart, Molecular Biosystems, 6 (2010) 1917-1927.

[160] B. Zybailov, A.L. Mosley, M.E. Sardiu, M.K. Coleman, L. Florens, M.P. Washburn, Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae, Journal of Proteome Research, 5 (2006) 2339-2347.

[161] Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, M. Mann, Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein, Molecular and Cellular Proteomics, 4 (2005) 1265-1272.

[162] J. Grossmann, B. Roschitzki, C. Panse, C. Fortes, S. Barkow-Oesterreicher, D. Rutishauser, R. Schlapbach, Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods, J Proteomics, 73 (2010) 1740-1746.

[163] J.C. Braisted, S. Kuntumalla, C. Vogel, E.M. Marcotte, A.R. Rodrigues, R. Wang, S.-T. Huang, E.S. Ferlanti, A.I. Saeed, R.D. Fleischmann, S.N. Peterson, R. Pieper, The APEX Quantitative Proteomics Tool: Generating protein quantitation estimates from LC-MS/MS proteomics results, in: BMC Bioinformatics, BioMed Central Ltd, 2008, pp. 529.

[164] P. Lu, C. Vogel, R. Wang, X. Yao, E.M. Marcotte, Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation, Nature Biotechnology, 25 (2006) 117-124.

[165] A.I. Nesvizhskii, R. Aebersold, Interpretation of Shotgun Proteomic Data, Molecular & Cellular Proteomics, 4 (2005) 1419-1440.

[166] Y. Zhang, Z. Wen, M.P. Washburn, F. L, Refinements to Label Free Proteome Quantitation: How to Deal with Peptides Shared by Multiple Proteins, (2010).

[167] D.H. Lundgren, S.I. Hwang, L.F. Wu, D.K. Han, Role of spectral counting in quantitative proteomics, Expert Rev Proteomic, 7 (2010) 39-53.

[168] W.X. Schulze, B. Usadel, Quantitation in Mass-Spectrometry-Based Proteomics, Annu Rev Plant Biol, 61 (2010) 491-516.

[169] K.A. Neilson, N.A. Ali, S. Muralidharan, M. Mirzaei, M. Mariani, G. Assadourian, A. Lee, S.C. van Sluyter, P.A. Haynes, Less label, more free: Approaches in label-free quantitative mass spectrometry, Proteomics, 11 (2011) 535-553.

[170] A.D. Polpitiya, W.-J. Qian, N. Jaitly, V.A. Petyuk, J.N. Adkins, D.G.C. II, G.A. Anderson, R.D. Smith, DAnTE: a statistical tool for quantitative analysis of -omics data, (2008).

[171] N. Karp, M. Spencer, H. Lindsay, K. O'Dell, L. KS, Impact of Replicate Types on Proteomic Expression Analysis, (2005).

[172] N. Karp, K. Lilley, Design and Analysis Issues in Quantitative Proteomics Studies, PROTEOMICS, 7 (2007) 42-50.

[173] F.P. Breitwieser, A. Müller, L.c. Dayon, T. Köcher, A. Hainard, P. Pichler, U. Schmidt-Erfurth, G. Superti-Furga, J.-C. Sanchez, K. Mechtler, K.L. Bennett, J. Colinge, General Statistical Modeling of Data from Protein Relative Expression Isobaric Tags, Journal of Proteome Research, 10 (2011) 2758-2766.

[174] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. Roy. Statist. Soc. Ser. B, 57 (1995) 289--300.

[175] J.D. Storey, R. Tibshirani, Statistical significance for genomewide studies, in: Proc Natl Acad Sci U S A, United States, 2003, pp. 9440-9445.

[176] T. Pluskal, S. Castillo, A. Villar-Briones, M. Oresic, MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, BMC Bioinformatics, 11 (2010) 395.

[177] M. Sturm, A. Bertsch, C. Gropl, A. Hildebrandt, R. Hussong, E. Lange, N. Pfeifer, O. Schulz-Trieglaff, A. Zerck, K. Reinert, O. Kohlbacher, OpenMS - an open-source software framework for mass spectrometry, BMC Bioinformatics, 9 (2008) 163.

[178] B. MacLean, D.M. Tomazela, N. Shulman, M. Chambers, G.L. Finney, B. Frewen, R. Kern, D.L. Tabb, D.C. Liebler, M.J. MacCoss, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments, Bioinformatics, 26 (2010) 966-968.

[179] D.K. Han, J. Eng, H. Zhou, R. Aebersold, Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry, Nature biotechnology, 19 (2001) 946-951.

[180] X.J. Li, H. Zhang, J.A. Ranish, R. Aebersold, Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry, Analytical chemistry, 75 (2003) 6648-6657.

[181] L.N. Mueller, O. Rinner, A. Schmidt, S. Letarte, B. Bodenmiller, M.Y. Brusniak, O. Vitek, R. Aebersold, M. Muller, SuperHirn - a novel tool for high resolution LC-MS-based peptide/protein profiling, Proteomics, 7 (2007) 3470-3480.

[182] T.S. Collier, P. Sarkar, W.L. Franck, B.M. Rao, R.A. Dean, D.C. Muddiman, Direct comparison of stable isotope labeling by amino acids in cell culture and spectral counting for quantitative proteomics, Analytical chemistry, 82 (2010) 8696-8702.

[183] R. Usaite, J. Wohlschlegel, J.D. Venable, S.K. Park, J. Nielsen, L. Olsson, J.R. Yates, Characterization of global yeast quantitative proteome data generated from the wild-type and glucose repression Saccharomyces cerevisiae strains: The comparison of two quantitative methods, Journal of Proteome Research, 7 (2008) 266-275.

[184] E.L. Hendrickson, Q.W. Xia, T.S. Wang, J.A. Leigh, M. Hackett, Comparison of spectral counting and metabolic stable isotope labeling for use with quantitative microbial proteomics, Analyst, 131 (2006) 1335-1341.

[185] Z. Li, R.M. Adams, K. Chourey, G.B. Hurst, R.L. Hettich, C.L. Pan, Systematic Comparison of Label-Free, Metabolic Labeling, and Isobaric Chemical Labeling for Quantitative Proteomics on LTQ Orbitrap Velos, Journal of Proteome Research, 11 (2012) 1582-1590.

[186] V.J. Patel, K. Thalassinos, S.E. Slade, J.B. Connolly, A. Crombie, J.C. Murrell, J.H. Scrivens, A Comparison of Labeling and Label-Free Mass Spectrometry-Based Proteomics Approaches, Journal of Proteome Research, 8 (2009) 3752-3759.

[187] H.X. Wang, S. Alvarez, L.M. Hicks, Comprehensive Comparison of iTRAQ and Label-free LC-Based Quantitative Proteomics Approaches Using Two Chlamydomonas reinhardtii Strains of Interest for Biofuels Engineering, Journal of Proteome Research, 11 (2012) 487-501.

[188] S.E. Ong, X. Li, M. Schenone, S.L. Schreiber, S.A. Carr, Identifying cellular targets of small-molecule probes and drugs with biochemical enrichment and SILAC, Methods in molecular biology, 803 (2012) 129-140.

[189] G.H. Borner, R. Antrobus, J. Hirst, G.S. Bhumbra, P. Kozik, L.P. Jackson, D.A. Sahlender, M.S. Robinson, Multivariate proteomic profiling identifies novel accessory proteins of coated vesicles, J

Cell Biol, 197 (2012) 141-160.

[190] R. Raijmakers, P. Dadvar, S. Pelletier, J. Gouw, K. Rumpel, A.J. Heck, Target profiling of a small library of phosphodiesterase 5 (PDE5) inhibitors using chemical proteomics, Chemmedchem, 5 (2010) 1927-1936.

[191] G.L. Andrews, B.L. Simons, J.B. Young, A.M. Hawkridge, D.C. Muddiman, Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600), Analytical chemistry, 83 (2011) 5442-5446.

[192] A.M. Falick, W.S. Lane, K.S. Lilley, M.J. MacCoss, B.S. Phinney, N.E. Sherman, S.T. Weintraub, H.E. Witkowska, N.A. Yates, ABRF-PRG07: advanced quantitative proteomics study, J Biomol Tech, 22 (2011) 21-26.

Chapter 2

# Improved Peptide Identification by Targeted Fragmentation Using CID, HCD and ETD on a LTQ-Orbitrap Velos

Christian K. Frese[1,2], A.F. Maarten Altelaar[1,2], Marco L. Hennrich[1,2], Dirk Nolting[3], Martin Zeller[3], Jens Griep-Raming[3], Albert J.R. Heck[1,2] and Shabaz Mohammed[1,2]

published in
Journal of Proteome Research, 2011 May 6

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

[3] Thermo Fisher Scientific, Hanna-Kunath-Str. 11, 28199 Bremen, Germany

## ABSTRACT

Over the past decade peptide sequencing by collision induced dissociation (CID) has become the method of choice in mass spectrometry-based proteomics. The development of alternative fragmentation techniques such as electron transfer dissociation (ETD) has extended the possibilities within tandem mass spectrometry (MS/MS). Recent advances in instrumentation allow peptide fragment ions to be detected with high speed and sensitivity (e.g. in a 2D or 3D ion trap) or at high resolution and high mass accuracy (e.g. an Orbitrap or a ToF). Here, we describe a comprehensive experimental comparison of using ETD, ion trap CID and beam type CID (HCD) in combination with either linear ion trap or Orbitrap readout for the large-scale analysis of tryptic peptides. We investigate which combination of fragmentation technique and mass analyzer provides the best performance for the analysis of distinct peptide populations such as N-acetylated, phosphorylated and tryptic peptides with up to two missed cleavages. We found that HCD provides more peptide identifications than CID and ETD for doubly charged peptides. In terms of Mascot score, ETD FT outperforms the other techniques for peptides with charge states higher than 2. Our data shows that there is a trade-off between spectral quality and speed when using the Orbitrap for fragment ion detection. We conclude that a decision-tree regulated combination of higher-energy collisional dissociation (HCD) and ETD can improve the average Mascot score.

## INTRODUCTION

Peptide sequencing is the key step in commonly employed bottom-up approaches within mass spectrometry-based proteomics. Usually proteins are digested with enzymes such as trypsin followed by 1- or 2-dimensional chromatographic separation techniques prior to tandem MS analysis [1]. The current method of choice for peptide sequencing is collision induced dissociation (CID) [2]. Dependent on the instrument type three collision energy regimes can be distinguished. High energy CID (>1 keV) is commonly employed in sector field and TOF/TOF instruments. Low energy CID, which is predominantly applied within proteomics workflows, can be subdivided into resonant-excitation CID (<2 eV; used in ion traps) and beam-type CID (~100 eV; QqQ or QqTOF instruments and HCD on Orbitraps). In general, the fragmentation process starts with the isolation of the precursor ions followed by collisions with neutral gas atoms which leads to a conversion of the kinetic energy into internal vibrational energy. As the vibrational energy exceeds a certain threshold covalent peptide bonds may break. Since the dissociation rate is lower than the energy redistribution rate the energy is randomly distributed over all atom bonds. Consequently, it is predominantly the weakest bond that breaks. The resulting fragment ions are used to obtain the peptide sequence by a database search. Resonant excitation or so called ion trap CID

differs slightly from beam-type CID. The fragmentation pattern of beam-type CID is due to the higher energy applied and the shorter activation time when compared to ion trap CID. Both b- and y-ions are observed in ion trap CID whereas the higher energy levels in beam-type CID lead to a predominance of y-ions whereas b-ions can fragment further to a-ions or smaller species [3, 4].



Available fragmentation techniques besides CID include electron-based approaches such as electron capture dissociation (ECD) and electron transfer dissociation (ETD) [5, 6]. Tandem mass spectrometers enabled with ETD are becoming more popular within the mass spectrometry community due to the technique being implement on a number of configurations including Q-TOFs [7], Q-TRAPs [8], hybrid linear ion trap-Orbitrap instruments [9] or Fourier transform ion cyclotron resonance mass spectrometers [10]. Unfortunately, ECD is currently limited to a few instrument types such as FT-ICR cells or linear ion trap-QTOF instruments since they allow trapping of electrons and analyte cations simultaneously [11, 12]. In contrast to collision induced dissociation, ETD and ECD induce cleavage of the N-C$\alpha$ bond and generate c- and z-type fragment ions. Several studies demonstrated that ETD is particularly useful for peptides with charge states >2 [13-16]. Furthermore, electron driven fragmentation techniques offer advantages for the analysis of labile post-translational modifications such as phosphorylation [17-19]. Overall, it is widely accepted that the outcome of fragmentation upon ETD activation is highly dependent on precursor charge state, and that ETD can be used complementary to CID to increase the number of identifications as well as peptide sequence coverage [15, 20-22]. In 2008, Coon and co-workers introduced a data-dependent decision tree logic that chooses 'on-the-fly' the appropriate activation technique dependent on the precursor charge state and m/z. The method increases the number of successful MS/MS events and improved the level of protein IDs that one can achieve using solely a single method [23].

In recent years hybrid tandem mass spectrometers have become available that combine a linear ion trap with an Orbitrap [24]. This combination allows for a high resolution MS spectrum with an excellent mass accuracy in the Orbitrap and the ability to perform rapid MS/MS (several Hz) in the ion trap. Olsen et al. reported, in 2007, on the development on the addition of a multipole collision cell to the LTQ-Orbitrap instrument which then allowed low energy beam-type CID generating spectra similar to those of triple quadrupole and Q-TOF instruments [25]. In order to distinguish between ion trap CID and beam type CID performed in the multipole (and to not confuse it with high energy CID) the latter was referred to as higher-energy collisional dissociation (HCD). Through such an activation process, low mass ions are now observable since there is no low mass cutoff present for the multipole which makes it more suitable for the analysis of reporter ions derived from chemical labeling reagents such as iTRAQ or TMT [26-29]. In 2009, the dual-pressure linear ion trap-Orbitrap

became available which allowed the generation of faster and more efficient ion trap MS/MS. The improvements were made possible through the use of two ion traps operating at different pressure regimes. The instrument was also equipped with a more optimized geometry for HCD transmission and extraction [30, 31]. This instrument has three possible ways of sequencing a peptide; CID, HCD and ETD. Each has been shown to be very capable and each having advantages and disadvantages.

The hybrid linear ion trap-Orbitrap instrument is currently one of the most used instruments in high-throughput proteomics and offers multiple peptide fragmentation techniques. Several studies have been performed comparing CID and ETD or CID and HCD. It is widely accepted that CID and ETD can be used complementary to cover different parts of a complex peptide mixture and to increase sequence coverage [15-17, 19-21, 23, 32, 33]. However, there is still ambiguity whether CID or HCD performs best for peptide identification [26, 30, 34, 35]. To the best of our knowledge, there is no direct comparison between all these three fragmentation techniques in combination with either linear ion trap or Orbitrap detection. Choosing the best fragmentation technique for different populations of peptides is crucial to increase the number of peptides and proteins that can be identified from a complex sample. The recently updated Orbitrap instrument allows the use of all three fragmentation modes CID, HCD and ETD in combination with either ion trap or Orbitrap mass analysis. Here, we report a comprehensive comparison of all combinations of fragmentation techniques and mass analyzers for peptides derived from a tryptic digest. We analyzed distinct populations of peptides as well as whole cell lysate and answer the question which combination of fragmentation technique and mass analyzer gives the highest number of peptide IDs.

## EXPERIMENTAL SECTION

**Sample preparation.** HEK293 cells were cultured at 37°C and 5% CO2 in DMEM high glucose medium containing 10% FCS, 10mM L-Glutamine and 5% Penicillin/Streptomycin (all Lonza). Cells were harvested by trypsinization and washed three times with PBS buffer. Cells were lysed in lysis buffer (8M Urea, 1 tablet Complete Mini EDTA-free Cocktail (Roche), 1 tablet PhosSTOP Phosphatase Inhibitor Cocktail (Roche), ad 10 mL with 50 mM ammonium bicarbonate buffer) by sonication. The lysate was centrifuged for 15 min at 20.000 × g and the supernatant was transferred into new tubes. The BCA assay (Pierce) was used to determine the protein concentration. 2,750 µg protein were reduced (2mM Dithiothreitol, 25 min at 56°C) and alkylated (4 mM Iodoacetamide, 30 min at room temperature in the dark). Lys-C was added in a 1:75 (w/w) ratio and incubated for 4h at 37 degree. The sample was diluted four times with 50 mM ammonium bicarbonate buffer. Sequencing grade trypsin (Promega) was added in a 1:100 (w/w) ratio and the incubation was carried out overnight at 37 degree.

The digested lysate was desalted by C18 solid phase extraction (Sep-Pak Vac C18 cartridge 3cc/200mg, Waters), dried down and reconstituted in 10% formic acid.

**Peptide fractionation and mass spectrometry.** Fractionation by low-pH strong cation exchange chromatography was performed as described previously using a Polysulfoethyl A column (200 mm × 2.1 mm × 5 µm, 200 Å, PolyLC Inc.) [36]. Collected fractions were dried down in a vacuum concentrator and reconstituted in 100 µl 10% formic acid. The samples were analyzed on an ETD enabled Orbitrap Velos instrument (Thermo Fisher Scientific, Bremen) connected to an Agilent 1200 HPLC system. All columns were packed in-house. The trap column was made using Aqua C18 material (Phenomenex, Torrance, CA). Reprosil-pur C18 3 µm (Dr. Maisch, Ammerbuch-Entringen, Germany) was used for the 35 cm analytical column with 50 µm inner diameter. Solvent A consisted of 0.1 M acetic acid (Merck) in deionized water (MilliQ, Millipore), and solvent B consisted of 0.1 M acetic acid in 80% acetonitrile (Biosolve). The flow rate of 5 µl/min was passively split and reduced to an effective flow rate of 100 nL/min. The gradients were as follows: 60 min LC method, 10 min solvent A; 10-40% solvent B within 30 minutes; 100% solvent B for 2 minutes; 15 min solvent A. 180 min LC method: 10 min solvent A; 10-25% solvent B within 107 minutes; 25-50% solvent B within 35 minutes; 100% solvent B for 2 minutes; 15 minutes solvent A. 300 min LC method: 10 min solvent A; 10-23% solvent B within 230 minutes; 23-50% solvent B within 37 minutes; 100% solvent B for 2 minutes; 15 min solvent A.

All instrument methods for the Orbitrap Velos were set up in the data dependent acquisition mode. After the survey scan the ten most intense precursors were selected for subsequent fragmentation using optimal settings for each activation technique (see table 1 for parameter details). The normalized collision energy was set to 35% for both CID and HCD. Supplemental activation was enabled for ETD. The signal threshold for triggering an MS/MS event was set to 500 counts. For internal mass calibration the 445.120025 ion was used as lock mass with a target lock mass abundance of 0%. The low mass cutoff for HCD and ETD FT was set to 180 m/z. Charge state screening was enabled and precursors with unknown charge state or a charge state of 1 were excluded. Dynamic exclusion was enabled (exclusion size list 500, exclusion duration 25 seconds for the 60 minutes methods and 90 seconds for the 180 and 300 minutes methods). The settings for the FT based decision tree were as follows: ETD was performed instead of HCD if the charge state is 3 and m/z is less than 800, or if the charge state is 4 and the m/z is less than 900, or if the charge state is 5 and m/z is less than 950. The modified decision tree was set up to use ETD with LTQ readout instead of HCD if the charge state is 3 and m/z is less than 750. ETD was performed for all precursor ions with charge states >3. The fragment ions were analyzed in the Orbitrap instead of the LTQ if the charge state is 4 and m/z is less than 1000. The preview mode option for the survey scan was disabled for both decision tree methods.

**Data analysis.** Peak lists were generated from the raw data files using Proteome Discoverer version 1.2 (Thermo Fisher Scientific, Bremen). The non-fragment filter was used to simplify ETD spectra with the following settings: the precursor peak was removed within a 4 Da window, charged reduced precursors were removed within a 2 Da window and neutral losses from charged reduced precursors were removed within a 2 Da window (the maximum neutral loss mass was set to 120 Da). An in-house developed Perl script was used to filter CID peak lists (minimum fragment ion count was set to 100, maximum number of fragment ions was set to 100). Peak list files from Orbitrap detection (HCD, CID FT and ETD FT) were deisotoped and charge deconvoluted, as described elsewhere [37]. Briefly, spectra are simplified by removing all except the monoisotopic peak from isotopic clusters and converting the m/z value of monoisotopic peaks from higher charged species into the 1+ charged state [38]. Peak lists were searched against swissprot (version 56.2, taxonomy filter was set to homo sapiens) including a list of common contaminants using Mascot (version 2.3.02, Matrix Science, UK). Trypsin was chosen as cleavage specificity allowing a maximum of two number of allowed missed cleavages of two. Carbamidomethylation (C) was set as a fixed modification and oxidation (M), acetylation (Protein N-term) and phosphorylation (S, T, Y) were used as variable modifications. The searches were performed using a peptide tolerance of 50 ppm and a product ion tolerance of 0.6 Da (ion trap readout) and 0.05 Da (Orbitrap readout), respectively. For further filtering the decoy search option was enabled. The resulting dat files were exported and filtered for <1% false discovery rate at peptide level using in-house developed software "Rockerbox" (version 1.2.6) utilizing the Percolator algorithm [39]. In brief, Percolator extracts features from the dat files and uses a support vector machine to create parsed PSMs [40]. Widening the Mascot search space by setting the peptide tolerance to 50 ppm provides a sufficient training dataset for the Percolator algorithm to distinguish between true positive and decoy hits. Hence there is no fixed Mascot score cut-off because peptide matches are accepted until a FDR rate of 1% is reached. However, only PSMs with Mascot scores >20 were accepted to ensure that only high quality data is allowed for this study.

## RESULTS & DISCUSSION

This study is aimed at providing a fair, comprehensive evaluation of all possible combinations of peptide fragmentation techniques and mass spectrometric detection modes and to elucidate the feasibility and performance of each method for efficient, routine peptide sequencing. For the purpose of the evaluation we generated populations of the main peptide compositions generally sequenced by mass spectrometry: N-acetylated peptides, singly phosphorylated, tryptic peptides with 0, 1 and 2 missed cleavages (5 discrete peptide pools). All peptide pools were generated from a trypsinised HEK293 cell lysate that had

been subjected to fractionation by low-pH SCX [22, 36]. These peptide pools were analyzed using a 150 min gradient (180 min total analysis time) and the 10 most intense precursors from each full scan were selected for MS/MS. Each pool was analyzed five times using exclusively either CID or ETD with LTQ mass analysis (referred to as CID, ETD), CID or ETD with Orbitrap readout (CID FT, ETD FT) or HCD (only Orbitrap readout available), see Table 1.

**Table 1.** Orbitrap Velos parameter settings of the methods used for the analyses.

| | Res. | MS2 | MS2 activation | Res. | MS2 target | | anion target | |
|---|---|---|---|---|---|---|---|---|
| mode | MS | analyzer | (msec) | MS2 | AGC | max. inj. (msec) | AGC | max. inj. (msec) |
| CID | 60k | LTQ | 10 | normal | 5000 | 100 | - | - |
| HCD | 30k | FT | 0.1 | 7500 | 30000 | 500 | - | - |
| ETD | 60k | LTQ | 50 | normal | 5000 | 100 | 200000 | 100 |
| CID FT | 30k | FT | 10 | 7500 | 30000 | 500 | - | - |
| ETD FT | 30k | FT | 50 | 7500 | 200000 | 500 | 400000 | 200 |
| FT based ddDT | 30k | FT (HCD) | 0.1 | 7500 | 100000 | 500 | - | - |
| | | FT (ETD FT) | 50 | | | | 300000 | 150 |
| modified ddDT | 30k | FT (HCD) | 0.1 | 7500 | 30000 | 500 | - | - |
| | | LTQ (ETD) | 50 | normal | 5000 | 100 | 200000 | 100 |
| | | FT (ETD FT) | 50 | 7500 | 200000 | 500 | 400000 | 200 |

The instrument operates in parallel acquisition mode when the LTQ is used for the product ion analysis. This is, of course, not possible for the modes where all scans are carried out in the Orbitrap. The summation of data for each activation method led to an average of ~90,000 spectra, Table 2. We found that the CID LTQ analyses generated the most tandem MS spectra followed by HCD, ETD, CID FT and ETD FT (Table 2).

**Table 2.** Number of PSMs and average Mascot scores from the analyses of distinct peptide pools generated by low-pH SCX fractionation of a tryptic HeLa digest. All samples were analyzed consecutively five times by CID, CID FT, HCD, ETD and ETD FT.

| peptide | CID | | ETD | | HCD | | CID FT | | ETD FT | |
|---|---|---|---|---|---|---|---|---|---|---|
| population | PSM | av. score | PSM | av. score | PSM | av. score | PSM | av. score | PSM | av. score |
| 2+ | 7775 | 47 | 5445 | 43 | 10869 | 56 | 8781 | 61 | 4461 | 44 |
| 3+ | 3309 | 38 | 5163 | 55 | 6698 | 58 | 5488 | 56 | 4055 | 68 |
| 4+ | 573 | 33 | 1815 | 45 | 2470 | 52 | 2068 | 49 | 1924 | 89 |
| N- acetylated | 1018 | 60 | 583 | 51 | 1105 | 70 | 1066 | 75 | 500 | 54 |
| phosphorylated | 1558 | 40 | 1222 | 44 | 1982 | 46 | 1822 | 50 | 1051 | 47 |
| # spectra | 119986 | | 83264 | | 101360 | | 90629 | | 58943 | |
| total PSM | 14233 | | 14228 | | 23124 | | 19225 | | 11991 | |

The differences can be directly linked to the use of parallel acquisition and the average scan time duration which is comprised of the time for ion accumulation, isolation of the precursor, fragmentation and detection of the fragment ions. When analyzing the data at

the peptide identification level, the data from the HCD analyses provided, by far, the highest number of peptide to spectrum matches (PSMs) followed by CID FT (summarized in Table 2). The sequencing modes CID and ETD performed similarly whereas ETD FT provided the least number of PSMs. Considering the number of parameters that are variables for each analysis, the overall numbers are just a crude benchmark that does not provide information about each of the techniques strengths and weaknesses. Thus, we analyzed the data set of each activation method on a peptide population level, specifically at the N-acetylated, phosphorylated, doubly, triply and quadruply charged peptide levels.

### Effect of deconvolution

Fragmentation spectra generated by CID from precursors with charge states >2+ potentially contain multiply charged fragment ions. These multiply charged ions are often weighted lower than singly charged ions when using the Mascot search engine. Furthermore, spectra possessing multiple forms of the same fragment ion (with respect to charge state) complicate spectra and thus can cause a higher level of misassignment. These issues can be overcome by deconvolution of spectra which increases the intensity of fragment ion peaks by summing up the intensities from their multiple charge states. Consequently, deisotoping and charge deconvolution of higher charge state peptides when using high resolution MS/MS data is required to perform a fair comparison between CID (in all forms) and ETD. For this purpose we applied the H-Score script, published recently by Savitski et al. [37], to our data. The peak lists for HCD, CID FT and ETD FT were searched in Mascot both with and without deconvolution (Table 3).

**Table 3.** Effect of charge deconvolution. Deconvolution of MS/MS peak list files provides for an increase in number of PSMs and average Mascot score (in brackets).

| peptide | HCD | | CID FT | | ETD FT | |
|---|---|---|---|---|---|---|
| charge | raw data | deconv. | raw data | deconv. | raw data | deconv. |
| 2+ | 10037 (51) | 10869 (56) | 7768 (50) | 8781 (61) | 4320 (45) | 4461 (44) |
| 3+ | 4865 (38) | 6698 (58) | 3375 (35) | 5488 (56) | 3853 (56) | 4055 (68) |
| 4+ | 943 (30) | 2470 (52) | 365 (29) | 2068 (49) | 1676 (45) | 1924 (89) |

Not surprisingly, the number of PSMs as well as the average Mascot score increased significantly for HCD and CID FT, while the increase in PSMs for ETD FT was nominal. The limited effect of deconvolution on ETD FT data can partly be explained by the charge reduction of the precursor inherent to ETD, an initial step of the fragmentation process. Note that the

presence of c-1 and z+1 ions due to hydrogen transfer potentially complicates the deconvolution of ETD FT spectra. All subsequent data analyses were carried out using solely the deconvoluted data.



### Mascot score distribution

We plotted the Mascot score distribution for each peptide fragmentation mode in the five classified types of peptides (Figure 1). To be consistent we based the cut-off point for each population to correspond to an FDR of <1% and thus the score cutoff differs between each fragmentation mode. Using this FDR, the score cutoff was lowest for HCD and highest for ETD and ETD FT. Since the mass accuracy at the MS level was similar for all modes of analysis, differences are caused by the activation technique and the type of analyzer used for the readout of fragmentation patterns. Supplemental activation after the electron transfer step improves backbone cleavage, but also induces hydrogen-atom transfer from c- to z- ions [14, 20]. This mass shift of 1 Da increases the number of fragment ions that cannot be assigned by the Mascot search engine and thus lead to the higher score cutoff for the ETD modes [41] although other search engines can accommodate the effect [42, 43].

The lower cutoff for HCD leads to a significantly higher number of spectra (and therefore peptides) being accepted that have a score between 20 and 30 compared to the other peptide activation techniques. Overall our data indicates that there is a tradeoff between spectral quality and speed when using the Orbitrap. For CID the high resolution and mass accuracy of the Orbitrap significantly increase the number of PSMs and average Mascot ion scores compared to conventional ion trap CID. The difference increases with higher charge states which illustrates the benefit of deconvolution (Figure 1 A,B,C). It should be noted that the superior mass accuracy and resolution of the Orbitrap will not directly increase Mascot scores but does help the peptide identification process by eliminating more false positives, i.e. lower Mascot scores can be reached with the same FDR value using higher quality tandem mass spectra. Regular tryptic peptides contain two basic groups which are located at the N-terminus and at the C-terminal lysine or arginine and such peptides are often doubly charged in the gas phase. Clearly for these peptides the collisional dissociation modes HCD and CID FT provide almost equally the highest number of PSMs also in the higher Mascot ion score ranges (Figure 1A). However, this result is partially helped by more peptides with lower scores being disproportionately added to the dataset at a taken threshold of 1% FDR when compared to other activation modes. The slightly lower number of PSMs for ETD FT compared to ETD can be attributed to the slower scan cycle time.

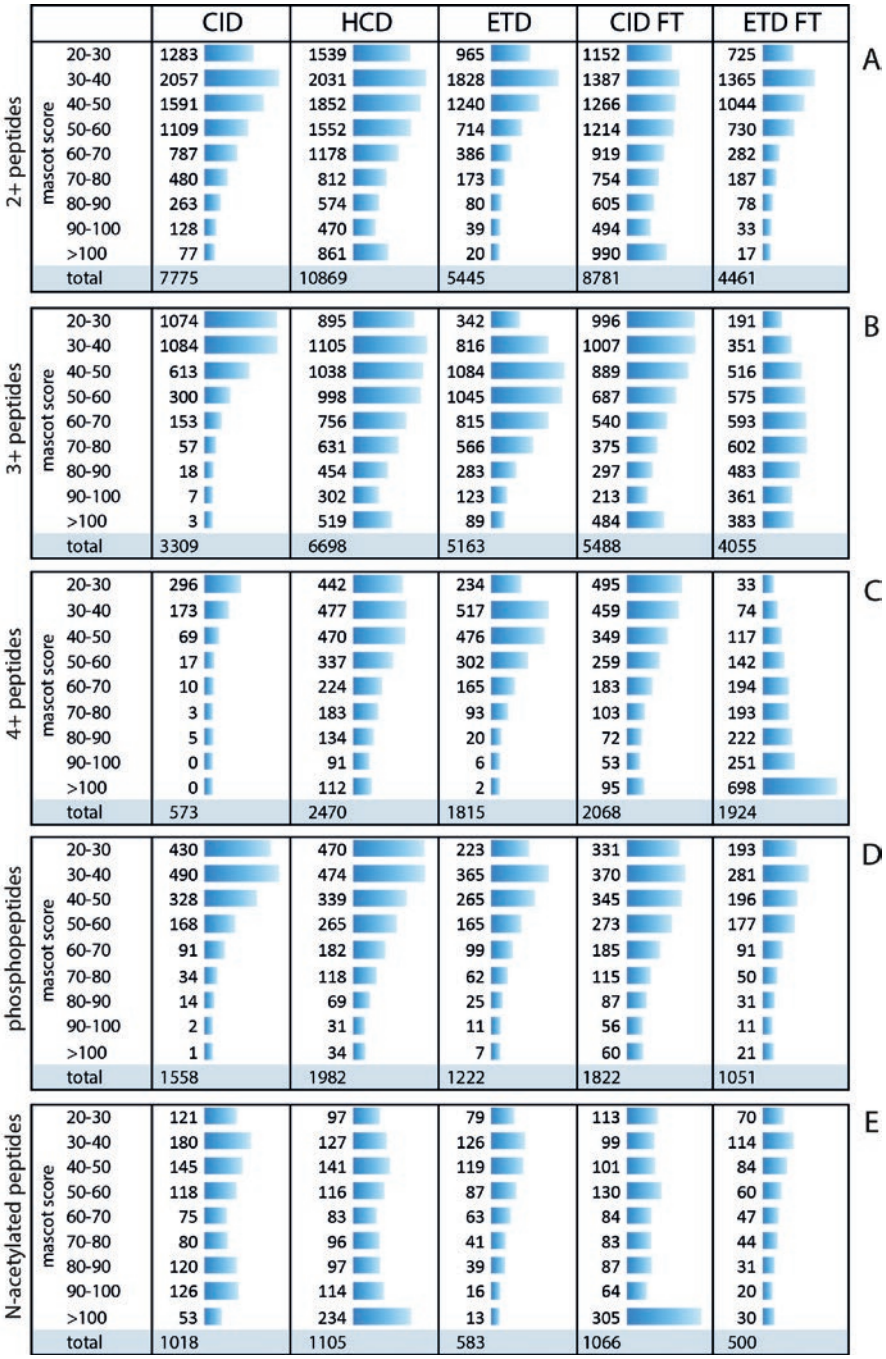



**Figure 1.** Histograms illustrating the Mascot ion score distribution of the doubly (A), triply (B), quadruply (C), phosphorylated (D), and N-acetylated (E) peptides derived from SCX-fractionated tryptic peptides. Samples were analyzed consecutively for five times employing either CID, HCD, ETD, CID FT or ETD FT. Data bars are normalized to the highest value within each population.

Previous studies have shown that ETD is more suitable than CID for the identification of peptides that carry more than two charges [14, 15]. Not surprisingly, for 3+ peptides the majority of the PSMs identified by ETD are in the Mascot score range of 40-60 whereas the CID and CID FT modes provide the highest number of PSMs between Mascot scores from 20 to 40 (Figure 1B). For both CID and ETD higher Mascot scores are obtained when the Orbitrap is chosen as mass analyzer. In particular, ETD FT shows a Mascot score distribution that is shifted towards higher ion scores which is also reflected in the highest average Mascot score among all fragmentation modes. HCD provides by far the highest number of PSMs for triply charged peptides. However, the average Mascot score and Mascot score distribution is similar to ETD and CID FT. For quadruply charged peptides the Mascot score distributions for all fragmentation modes except ETD FT resemble the 3+ peptide distributions but are shifted towards lower Mascot score ranges (Figure 1C). Here, ETD FT clearly outperforms all other fragmentation modes in terms of Mascot score, providing more than one third of all PSMs above Mascot ion scores of 100. Taking into account that the scan cycle time of ETD FT is significantly higher compared to HCD, our data suggests that ETD in combination with Orbitrap fragment ion detection provides spectra with the highest quality compared to all other fragmentation modes.

Globally, the differences in performance of the five fragmentation modes observed for the doubly charged peptides (Figure 1A) are similar in the case of phosphopeptides (Figure 1D) and N-acetylated peptides (Figure 1E). Activation with HCD provides the most PSMs followed by CID FT, CID, ETD and ETD FT being last. The Mascot score distribution of N-acetylated peptides differs from the other peptide populations. Close inspection revealed that the high scoring PSMs observed in the data in Figure 1E originated largely from a few N-acetylated peptides that had been sequenced many times. The performance of the ETD modes in terms of number of PSMs as well as Mascot scores is poor compared to HCD and the CID modes when analyzing N-acetylated peptides (Figure 1E, Table 2). This finding is in line with our expectations since N-acetylated peptides are less basic due to the blocked amine group at the N-terminus which prevents protonation at this position. In summary, HCD and CID FT perform equally well for both N-acetylated and phosphopeptides in terms of number of identifications when there is a single basic amino acid present. This is in line with a recent study published by Mann and co-workers [34].

### Influence of peptide mass on Mascot scores

Next, we investigated the influence of peptide mass on Mascot scores for doubly, triply and quadruply charged peptides. The general trend for doubly charged peptides is that the average Mascot score increases with peptide mass (Figure 2), which is particularly amplified for CID, CID FT and HCD.

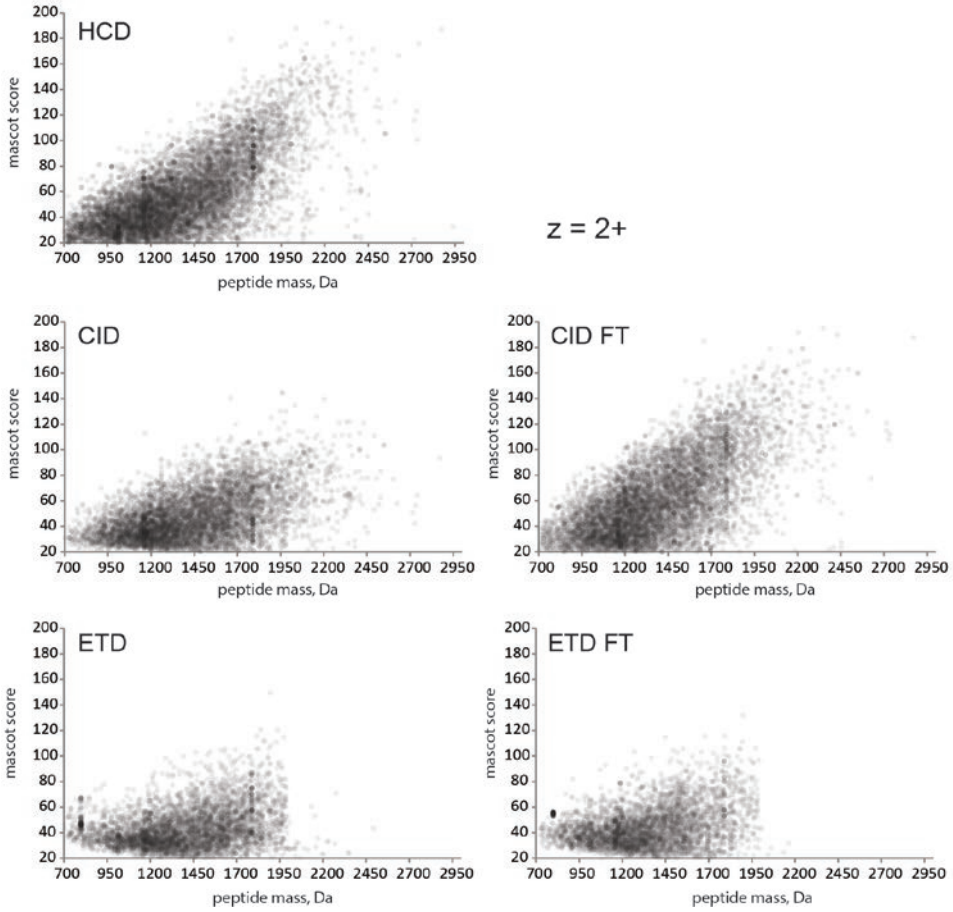



**Figure 2.** Density scatter plots display the influence of peptide mass on Mascot ion scores for doubly charged peptides identified from the analysis of tryptic peptides after SCX fractionation. Darker color indicates more data points.

The increase in score with mass can be explained simply by the increase in probable fragment ions with mass. For ETD and ETD FT the Mascot scores also increase with higher peptide mass but scores are, in general, significantly lower compared to the other fragmentation modes. The ETD modes provide few doubly charged PSMs with masses above 2000 Da. This cutoff seems artificial and may be attributed to supplemental activation being unable to operate on charge reduced species with m/z above 2000. In the case of triply charged peptides the influence of peptide mass on Mascot scores decreases for ion trap CID (Figure 3).

Comparing the result to CID FT indicates that deconvolution of multiply charged fragment ions is required for obtaining higher Mascot ion scores. Conversely, the outcome of the fragmentation by ETD seems to be less dependent on precursor mass. Here, high and low scoring PSMs are observed over the whole mass range. Again, being able to deconvolute the ETD FT data leads to the generation of higher ion scores. However, the positive effect of deconvolution in the case of triply charged peptides analyzed by ETD FT is smaller compared to CID FT. The limited effect of deconvolution of spectra from triply charged peptides being fragmented by ETD FT can be explained by the fact that such spectra consist mostly of singly charged fragments.

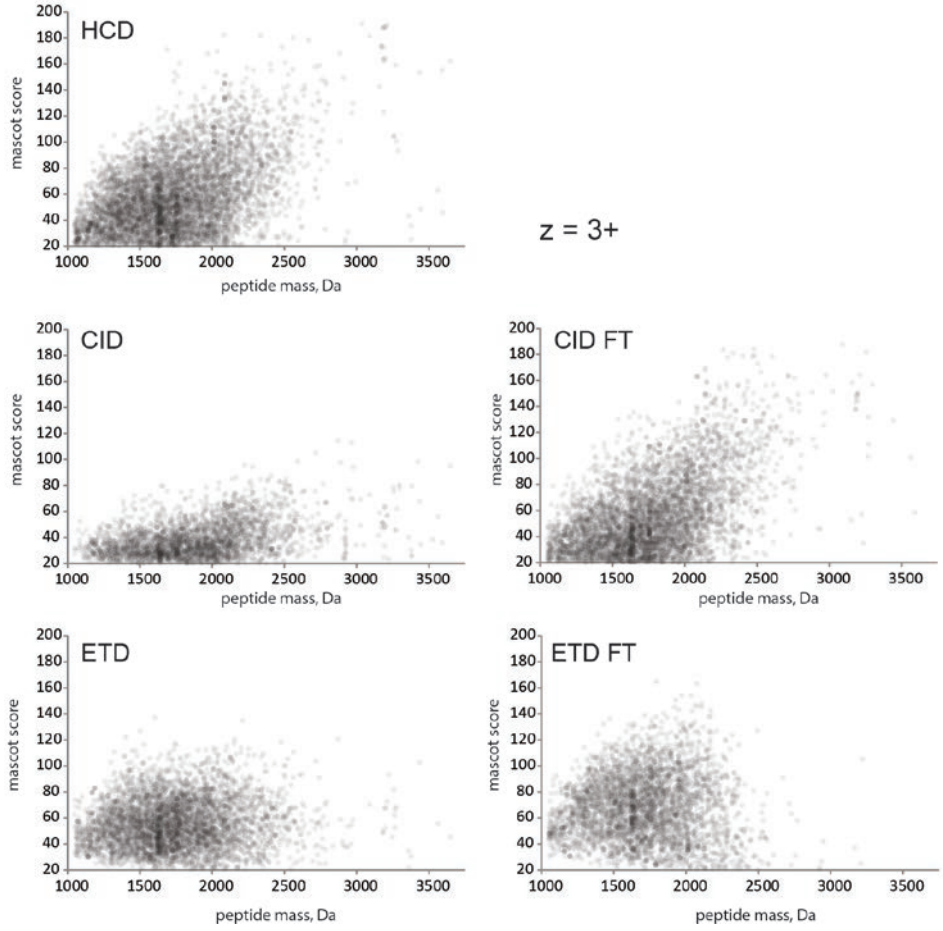



**Figure 3.** Density scatter plots display the influence of peptide mass on Mascot ion scores for triply charged peptides identified from the analysis of tryptic peptides after SCX fractionation. Darker color indicates more data points. Mascot ion scores increase substantially with peptide length for CID FT and HCD. The influence of peptide mass on Mascot scores is negligible low for CID, ETD and ETD FT.

In general, the same holds for quadruply charged peptides (Figure 4). The Mascot ion scores seem to be less influenced by peptide mass. For 4+ peptides the effect of deconvolution also boosts the ETD FT Mascot scores since there is now high probability of generating doubly and triply charged fragments.

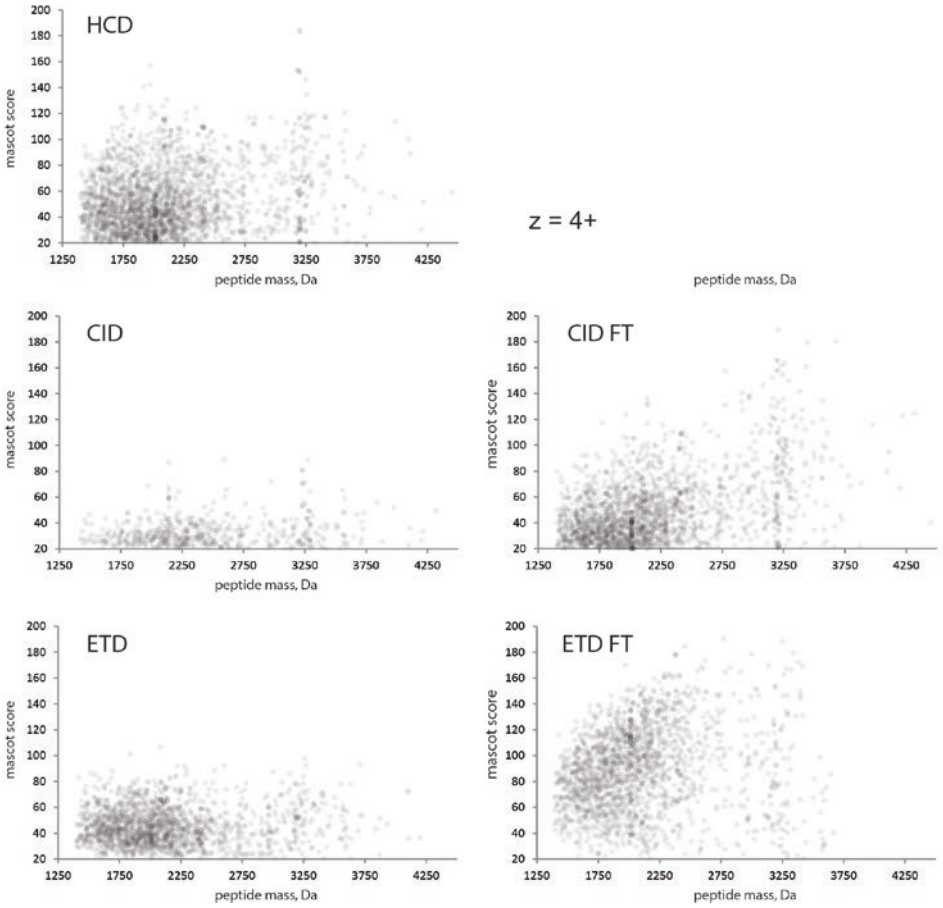



**Figure 4.** Density scatter plots display the influence of peptide mass on Mascot ion scores for quadruply charged peptides, as in Figure 3. Mascot ion scores are significantly higher for ETD FT when compared to all other fragmentation modes.

### Overlap in peptide identifications between the fragmentation modes

We investigated the overlap in identifications for each peptide population to examine whether fragmentation modes were either more alike or more complementary. We calculated the overlap in peptide identifications between CID, HCD and ETD and also between CID FT, HCD and ETD FT. We further investigated which fragmentation mode provides the highest Mascot score from the overlapping unique peptides that were identified by all three

fragmentation modes. Not surprisingly, for the doubly charged peptide population HCD provides the highest number of unique peptides (22%). Only 5% and 11% of the doubly charged peptides are exclusively identified by ETD and CID, respectively (Figure 5A). Between all three techniques 2812 unique peptides were successfully sequenced. Within the overlap, HCD provides for 72% of these overlapping peptides the highest Mascot score followed by CID (17%) and ETD (11%). A different distribution is observed when comparing HCD, CID FT and ETD FT (Figure 5B). Here, CID FT provides for 55% of the overlapping peptides the highest Mascot score. CID FT spectra exhibit b- and y-ions over the whole mass range in contrast to HCD spectra that are dominated by y-ions. The addition of b-ions lead to higher Mascot scores which might explain the prevalence of CID FT over HCD.

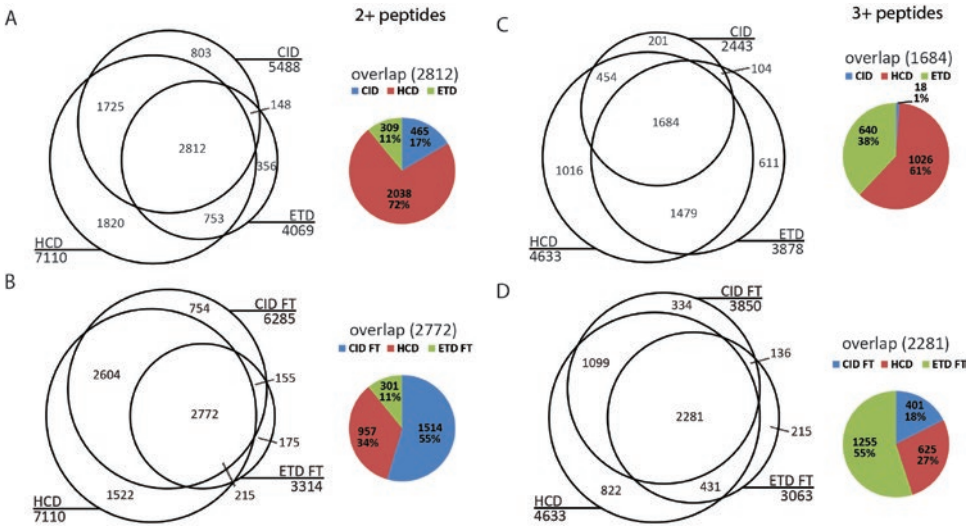



**Figure 5.** Overlap of doubly (A,B) and triply (C,D) charged peptides that were identified from the SCX-fractionated tryptic HeLa digest. Venn diagrams illustrate the overlap of the unique peptides identified between CID, HCD and ETD (A,C) and between CID FT, HCD and ETD FT (B,D). Pie charts show the number of unique peptides with the highest Mascot score per fragmentation mode within the peptides that were identified by all three fragmentation modes.

For triply charged peptides the number of unique IDs is higher for HCD compared to ETD (Figure 5C). The overlap between CID and HCD is high whereas HCD and ETD provide more complementary unique peptide. The distribution of the highest scoring peptides within the overlap suggests that ETD can be used complementary to HCD for peptides with charge states higher than 2+. The comparison between HCD, CID FT and ETD FT for triply charged peptides further indicates that ETD FT spectra can be considered to be more data-rich compared to HCD and CID FT as the highest Mascot score of 55% of all overlapping peptides was provided by ETD FT (Figure 5D).

Comparing ETD FT with CID FT and HCD for quadruply charged peptides shows that 82% of all peptides exhibit higher Mascot scores when sequenced by ETD FT which clearly demonstrates the superior ability of ETD for the analysis of multiply charged peptides (Figure 6B). In the case of N-acetylated and singly phosphorylated peptides the distribution of the highest Mascot scores from the overlapping unique peptides resembles the results of the regular tryptic 2+ peptides to a great extent. HCD provides the majority of the highest scoring peptides in comparison with ion trap CID and ETD (Figure 6C and 6E) and on the other hand CID FT dominates the distribution when the FT based fragmentation modes are compared (Figures 6D and 6F). Note, however, that a higher Mascot score by itself does not necessarily translate into improved phosphosite localization [19, 44, 45].

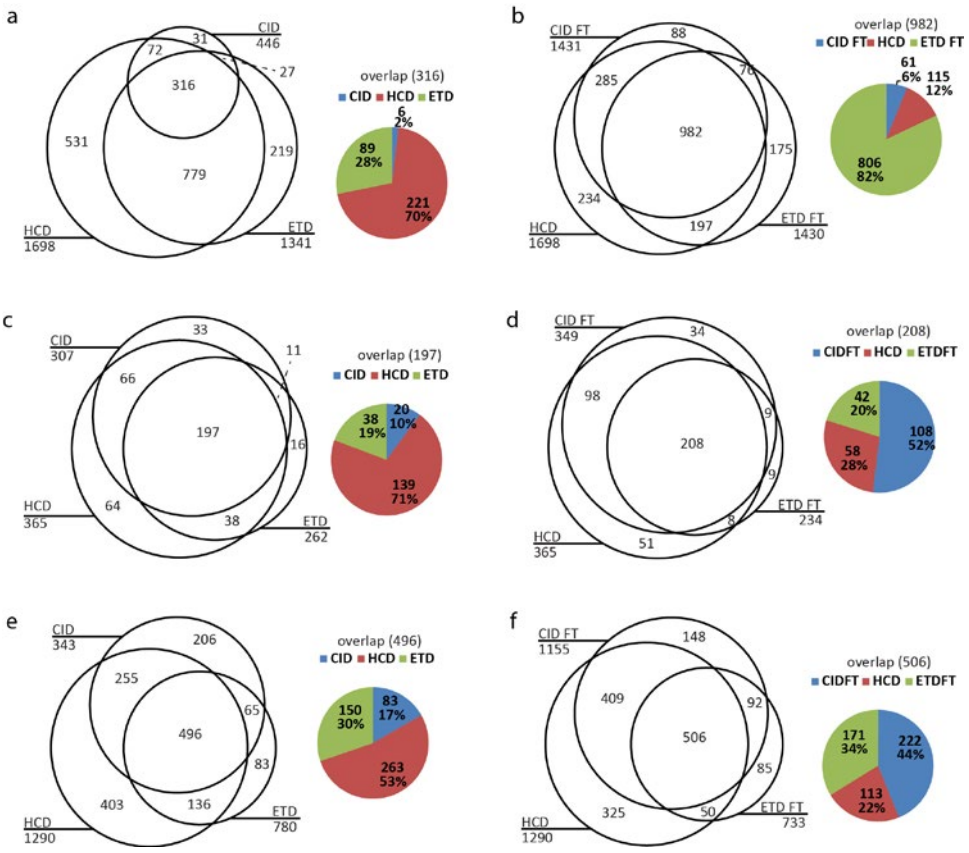



**Figure 6.** Overlap between CID, HCD and ETD (a,c,e) and between CID FT, HCD and ETD (b,d,f) for quadruply charged peptides (a,b), N-acetylated peptides (c,d) and phosphorylated peptides (e,f). Pie charts show the number of peptides with the highest Mascot score per fragmentation mode within the overlapping peptides that were identified by all three fragmentation modes.

### Analysis of a complex peptide mixture

We then investigated the performance of each fragmentation mode for the analysis of non-fractionated samples. We analyzed whole human cell lysate digests directly by RP-LC-MS/MS using a 5h LC analysis and a range of Top10 methods. Data for all acquisition modes was acquired in triplicate to investigate the repeatability of the results. The general trends for the cell lysate digests are in line with the outcome of the analyses of the SCX fractionated separate peptide pools, discussed above. In brief, HCD outperforms all other fragmentation modes followed by CID FT. The activation strategies ETD FT, ETD and CID perform approximately on the same level (Table 4, separate results from triplicate runs can be found in the online supporting information).

**Table 4.** Triplicate analysis of a complex peptide mixture derived from a tryptic HeLa cell digest. Different amounts were analyzed over a 5h LC-gradient employing CID, CID FT, HCD, ETD or ETD FT methods. Additionally the same sample was analyzed using the FT based and a modified decision tree method. The success rate was calculated by dividing the number of peptide identifications by the number of MS/MS scans.

| sample load | fragmentation mode | number of PSMs | average mascot score | protein IDs | MSMS scans | success rate |
|---|---|---|---|---|---|---|
| 1 µg | CID | 10464 ± 420 | 41 | 1971 ± 87 | 37950 ± 567 | 28% |
| | HCD | 18368 ± 71 | 50 | 2562 ± 31 | 37517 ± 318 | 49% |
| | ETD | 11007 ± 520 | 41 | 1850 ± 78 | 30182 ± 396 | 36% |
| | CID FT | 14288 ± 352 | 52 | 2213 ± 30 | 33180 ± 549 | 43% |
| | ETD FT | 11120 ± 175 | 52 | 1878 ± 37 | 25709 ± 180 | 43% |
| | modified DDDT | 18402 ± 570 | 56 | 2586 ± 112 | 33314 ± 479 | 55% |
| | HCD | 11521 ± 275 | 56 | 2304 ± 96 | 19328 ± 340 | 60% |
| | ETD | 5435 ± 264 | 43 | 1442 ± 77 | 10729 ± 209 | 51% |
| | ETD FT | 1447 ± 46 | 68 | 717 ± 17 | 3241 ± 30 | 45% |
| | FT based DDDT | 18263 ± 323 | 54 | 2587 ± 36 | 34178 ± 285 | 53% |
| | HCD | 11382 ± 100 | 56 | 2318 ± 22 | 19229 ± 414 | 59% |
| | ETD FT | 6880 ± 258 | 51 | 1576 ± 59 | 14925 ± 142 | 48% |
| 100 ng | CID | 5052 | 42 | 1071 | 25628 | 20% |
| | HCD | 12064 | 48 | 1653 | 24952 | 48% |
| | ETD | 3738 | 41 | 740 | 23006 | 16% |
| 10 ng | CID | 1191 | 45 | 420 | 7372 | 16% |
| | HCD | 2824 | 45 | 605 | 7823 | 36% |
| | ETD | 859 | 44 | 315 | 8007 | 11% |

As expected the number of MS/MS spectra acquired differs largely between the fragmentation modes because of the differences in duty cycle. Interestingly, the number of PSMs does not correlate with the number of acquired MS/MS spectra. For example, the number of PSMs identified by CID and HCD differs largely, even though approximately the same number of MS/MS scans was acquired in both modes. Consequently, the success rate, defined as the number of PSMs divided by the number of MS/MS scans acquired, varies among the fragmentation modes. The highest success rate for an individual technique was obtained by HCD (49%). Not surprisingly, the ion trap detection modes clearly exhibit a lower success rate compared to the equivalent FT detection modes which is caused both by the larger number of MS/MS spectra acquired (faster scan cycle time) and the poorer spectral quality (low resolution) of the ion trap.

Furthermore, we performed additional experiments where the amount of material analyzed was reduced to test if the apparent superior sensitivity of the linear ion trap can overcome the advantages of the high mass accuracy and resolution of the Orbitrap [30]. We performed equivalent analyses but now using 100 ng and 10 ng of human cell lysate. For this particular analysis, we restricted the types of analyses to HCD, CID or ETD. The number of MS/MS spectra acquired for each activation mode is almost identical (Table 4) when analyzing these lower amounts of material indicating that the CID mode does not benefit from the higher duty cycle of the LTQ. Strikingly, HCD results in more than twice as many PSMs as CID and three times more than ETD. This trend remains the same when analyzing 10 ng of digested human cell lysate, indicating that the improved HCD cell performs on the same level as the linear ion trap in terms of sensitivity.

We repeated the 100 ng experiment but using a shorter LC gradient (total analysis time 60 min) and Top10 methods (Table 5). Additionally, we performed a CID Top20 method. As expected, the number of MS/MS spectra acquired directly reflects the scan cycle duration of each fragmentation mode. In the CID analyses 9,021 (Top10) and 9,408 (Top20), respectively, spectra were acquired. Switching from Top10 to Top20 had a minimal effect. Due to the slower cycle time there were fewer spectra acquired in the HCD (6,462), ETD (4,096), CID FT (4,957) and ETD FT (3,549) analyses. The average number of MS/MS scans per second was highest for CID (4.7 for top10 and 5.0 for top20) followed by HCD (4.0), CID FT (2.8), ETD (2.3) and ETD FT (1.8). The short LC gradient time leads to a large number of peptides co-eluting at any time and so under-sampling would be guaranteed and, in principle, there would be constant maxima of MS/MS events per cycle. Therefore, we calculated the number of MS/MS scans during each cycle over the main elution window (between RT 23.5 and 52 minutes of the LC gradient). All fragmentation modes except for CID were triggering the maximum number of MS/MS scans during the analysis (Table 5).

**Table 5.** Triplicate analysis of a complex peptide mixture using 60 min LC-gradients and different fragmentation modes. 100 ng of a tryptic HeLa digest was analyzed consecutively by Top10 methods employing CID, CID FT, HCD, ETD and ETD FT. Additionally a Top20 method was set up and used for CID analysis.

| fragmentation mode | number of PSMs | average mascot score | protein IDs | MSMS scans | success rate | max. number of MSMS events | average MSMS/sec. |
|---|---|---|---|---|---|---|---|
| CID Top10 | 3568 ± 34 | 42 | 754 ± 23 | 9129 ± 180 | 39% | 85% | 4.7 |
| CID Top20 | 3604 ± 36 | 44 | 758 ± 16 | 9617 ± 191 | 37% | 38% | 5.0 |
| HCD | 5314 ± 170 | 54 | 863 ± 5 | 7615 ± 198 | 70% | 97% | 4.0 |
| ETD | 2359 ± 46 | 46 | 470 ± 13 | 4571 ± 115 | 52% | 99% | 2.3 |
| CID FT | 3445 ± 52 | 57 | 618 ± 23 | 5461 ± 113 | 63% | 99% | 2.8 |
| ETD FT | 2020 ± 26 | 54 | 410 ± 6 | 3546 ± 58 | 57% | 99% | 1.8 |

In the case of CID the instrument was triggering in 85% (Top10) and 38% (Top20), respectively, of the time the maximum number of MS/MS scans per MS scan which demonstrates the higher speed compared to the other fragmentation modes. The complexity of the sample is of such significance that it was surprising that the Top20 method did not completely max out. Possible reasons are poor choice of triggering thresholds, imperfect peak picking and insufficient dynamic range of the instrument. In an ion trap the total charge is split between all ions trapped; peptides and chemical noise. The 1 million 'charges' limitation (caused by the C-trap [46]) combined with the charge being distributed will lead to a compression of the dynamic range. Low abundant components will have too few charges in order to have an acceptable signal for detection. Nevertheless, the success rates of all fragmentation modes are higher for the 60 min analyses when compared to their 5h analogues, a fact that can be explained only by the high abundant 'proteotypic' precursor ions being selected for fragmentation due to the limitations in scan cycle speed. Interestingly, in terms of number of PSMs there is little difference between Top10 and Top20 CID. The Top20 method results in approximately 500 additional MS/MS spectra, however, the number of PSMs and proteins identified is on the same level. HCD still outperforms all other fragmentation modes in terms of PSMs but in the case of CID and ETD the faster scan cycle time make the ion trap modes preferable.

### Performance of a data dependent decision tree employing HCD, ETD and ETD FT

Our results show that HCD performs very well for doubly charged peptides. ETD and ETD FT provided the best results in terms of Mascot score for peptides with charge states higher than two. We next investigated if it is possible to benefit from the complementary performance of the two techniques in a data-dependent fashion. Earlier, Swaney et al. showed that a data-dependent decision tree (ddDT) algorithm utilizing CID and ETD improves the sequencing success rate and generates more identifications than CID or ETD alone [23].

For our experiment we initially had to choose a compromise for the instrument settings. So far the instrument control software did not allow the programming of a ddDT logic that uses HCD with Orbitrap readout and ETD with ion trap readout because it is not possible to switch the mass analyzer used for MS/MS readout. Furthermore, we had optimized the AGC settings for each fragmentation mode for best results. Again, the control software allowed only one AGC value for the activation mode (regardless of the type of activation) so we chose target settings that are a compromise for HCD and ETD FT (Table 1). The FT based ddDT performs HCD for all doubly charged precursors and ETD for higher charged precursors.

By addressing the most suitable fragmentation technique for each peptide we aimed to increase the number of peptide IDs, the average Mascot score and the overall success rate. We analyzed the same amount of the human cell digest (1 µg) using the Top10 FT based decision tree HCD/ETD FT method on a 5h gradient. In total 34,178 MS/MS spectra were acquired where HCD contributed with 19,229 and ETD FT with 14,925 spectra. The success rates for HCD and ETD FT within the FT based ddDT are 59% and 48%, respectively. As expected, the FT based ddDT method provides about the same number of PSMs as HCD alone (Table 4). However, the average Mascot score significantly increases. Next, we modified the instrument software to allow switching between LTQ and Orbitrap analyzer and to apply separate AGC target values for the ddDT. This enabled us to use HCD, ETD and ETD FT in one single run and address every precursor by the most suitable fragmentation mode. Not surprisingly, the modified ddDT performs on the same level as HCD in terms of identified PSMs and proteins. The majority of the PSMs is provided by HCD (11,521) followed by ETD (5,435) and ETD FT (1,447). The overall success rate of 55% illustrates that more than every second MS/MS spectrum led to statistically significant peptide identification. Compared to the regular FT based ddDT the modified one performs equally but provides slightly higher Mascot scores. Although the ddDT logic directs every precursor to either HCD, ETD or ETD FT fragmentation, there are several peptides that were identified by multiple fragmentation modes (Figure 7). This can be explained by the fact that some peptide ions can be present in different charge states. Altogether this illustrates the feasibility of a data-dependent decision tree logic combining the advantages of different fragmentation techniques (HCD, ETD) and mass analyzers (LTQ, Orbitrap) for enhanced peptide identification.

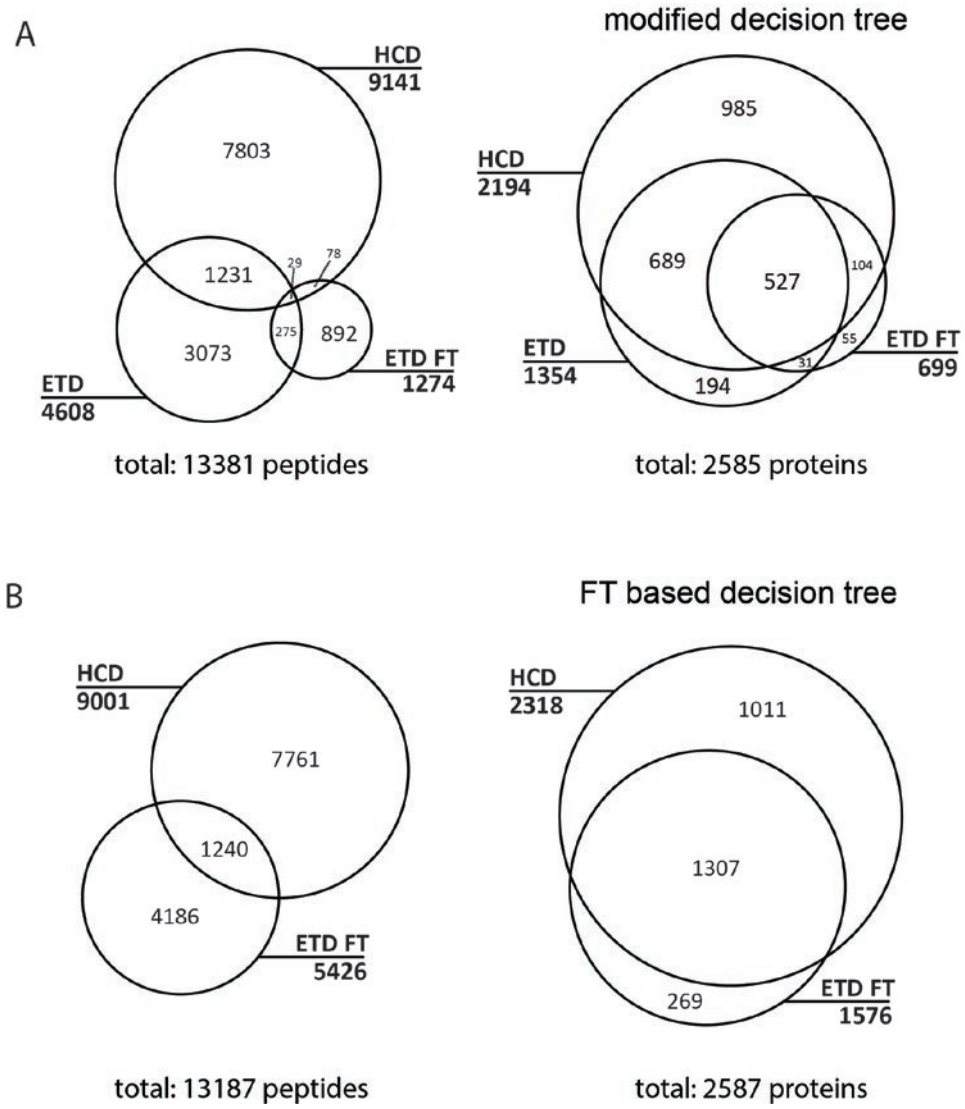



**Figure 7.** Venn diagrams illustrating the contribution of the different fragmentation modes in terms of unique peptides and proteins for the modified ddDT (A) and FT based ddDT (B).

## CONCLUSION & OUTLOOK

Our experiments were aimed to evaluate all possible combinations of fragmentation and detection modes on the current version of an Orbitrap Velos instrument. In all analyses the success rate for HCD was significantly higher than for all other modes. HCD performs best for doubly charged peptides, singly phosphorylated peptides and N-acetylated peptides. ETD FT outperforms the other techniques for peptides with charges states >2 in terms of

Mascot scores. We found that the spectral quality of Orbitrap detection data is superior to ion trap data, especially when considering charge deconvolution. However, the main discriminating factor for obtaining the highest number of PSMs is scan cycle speed, as indicated by the larger number of identified PSMs from HCD fragmentation when compared to ETD FT. We showed that despite the high sensitivity of the dual pressure linear trap HCD still generates better results than CID even for very low amounts of peptides. We expect that ETD FT will benefit dramatically from further improvements in speed and product ion transmission from the ion trap to the Orbitrap. We further expect that a dynamic AGC parameter dependent on precursor charge and m/z value could improve the success rate for multiply charged peptides. The number of possibly observable fragment ions increases with increasing peptide length and charge state. Consequently, the product ion signal-to-noise ratio gets worse. Higher AGC values for peptides with higher charges states and m/z could help compensate. Assigning the peptides to the most suitable fragmentation mode is a key challenge for obtaining the highest number of PSMs. We showed that a modified data-dependent decision tree method employing HCD, ETD and ETD FT performs similar to HCD alone for the analysis of a complex peptide mixture, however, with a higher average Mascot score.

## ACKNOWLEDGEMENTS

This research was performed within the framework of CTMM, the Center for Translational Molecular Medicine (www.ctmm.nl), project CIRCULATING CELLS (grant 01C-102), and supported by the Netherlands Heart Foundation. Additionally, the Netherlands Proteomics Centre, a program embedded in the Netherlands Genomics Initiative, is kindly acknowledged for financial support.

Supporting information available

An excel sheet with all results (including separate results from triplicate analyses). All data is available as a scaffold file on tranche at https://proteomecommons.org using the following hash:

L0u/QC4fuLOfEI43/5+Aea2ko37clzjOU4IfSa5rZxID9DmXcUvo1TFD7HYus8V3YRKX9/Hg0IlpXy5t2q5SJ2BOEbMAAAAAAAABdw==

## REFERENCES

[1] M.P. Washburn, D. Wolters, J.R. Yates, 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology, Nat Biotechnol, 19 (2001) 242-247.

[2] D.F. Hunt, J.R. Yates, 3rd, J. Shabanowitz, S. Winston, C.R. Hauer, Protein sequencing by tandem mass spectrometry, Proc Natl Acad Sci U S A, 83 (1986) 6233-6237.

[3] R.W. Vachet, K.L. Ray, G.L. Glish, Origin of product ions in the MS/MS spectra of peptides in a quadrupole ion trap, J Am Soc Mass Spectrom, 9 (1998) 341-344.

[4] K.W. Lau, S.R. Hart, J.A. Lynch, S.C. Wong, S.J. Hubbard, S.J. Gaskell, Observations on the detection of b- and y-type ions in the collisionally activated decomposition spectra of protonated peptides, Rapid Commun Mass Spectrom, 23 (2009) 1508-1514.

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

[6] R.A. Zubarev, Electron-capture dissociation tandem mass spectrometry, Curr Opin Biotechnol, 15 (2004) 12-16.

[7] Y. Xia, P.A. Chrisman, D.E. Erickson, J. Liu, X. Liang, F.A. Londry, M.J. Yang, S.A. McLuckey, Implementation of ion/ion reactions in a quadrupole/time-of-flight tandem mass spectrometer, Anal Chem, 78 (2006) 4146-4154.

[8] S.J. Pitteri, P.A. Chrisman, J.M. Hogan, S.A. McLuckey, Electron transfer ion/ion reactions in a three-dimensional quadrupole ion trap: reactions of doubly and triply protonated peptides with SO2*, Anal Chem, 77 (2005) 1831-1839.

[9] G.C. McAlister, W.T. Berggren, J. Griep-Raming, S. Horning, A. Makarov, D. Phanstiel, G. Stafford, D.L. Swaney, J.E. Syka, V. Zabrouskov, J.J. Coon, A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer, J Proteome Res, 7 (2008) 3127-3136.

[10] D.A. Kaplan, R. Hartmer, J.P. Speir, C. Stoermer, D. Gumerov, M.L. Easterling, A. Brekenfeld, T. Kim, F. Laukien, M.A. Park, Electron transfer dissociation in the hexapole collision cell of a hybrid quadrupole-hexapole Fourier transform ion cyclotron resonance mass spectrometer, Rapid Commun Mass Spectrom, 22 (2008) 271-278.

[11] H. Satake, H. Hasegawa, A. Hirabayashi, Y. Hashimoto, T. Baba, K. Masuda, Fast multiple electron capture dissociation in a linear radio frequency quadrupole ion trap, Anal Chem, 79 (2007) 8755-8761.

[12] J. Axelsson, M. Palmblad, K. Hakansson, P. Hakansson, Electron capture dissociation of substance P using a commercially available Fourier transform ion cyclotron resonance mass spectrometer, Rapid Commun Mass Spectrom, 13 (1999) 474-477.

[13] A.F. Altelaar, S. Mohammed, M.A. Brans, R.A. Adan, A.J. Heck, Improved identification of endogenous peptides from murine nervous tissue by multiplexed peptide extraction methods and multiplexed mass spectrometric analysis, J Proteome Res, 8 (2009) 870-876.

[14] D.L. Swaney, G.C. McAlister, M. Wirtala, J.C. Schwartz, J.E. Syka, J.J. Coon, Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors, Anal Chem, 79 (2007) 477-485.

[15] D.M. Good, M. Wirtala, G.C. McAlister, J.J. Coon, Performance characteristics of electron transfer dissociation mass spectrometry, Mol Cell Proteomics, 6 (2007) 1942-1951.

[16] H.W.P. van den Toorn, S. Mohammed, J.W. Gouw, B. van Breukelen, A.J.R. Heck, Targeted SCX Based PeptideFractionation for Optimal Sequencing by Collision Induced, and Electron Ttransfer Dissociation, J proteomics Bioinform, 1 (2008) 379-388.

[17] L.M. Mikesh, B. Ueberheide, A. Chi, J.J. Coon, J.E. Syka, J. Shabanowitz, D.F. Hunt, The utility of ETD mass spectrometry in proteomic analysis, Biochim Biophys Acta, 1764 (2006) 1811-1822.

[18] J. Wiesner, T. Premsler, A. Sickmann, Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications, Proteomics, 8 (2008) 4466-4483.

[19] P.J. Boersema, S. Mohammed, A.J. Heck, Phosphopeptide fragmentation and analysis by mass spectrometry, J Mass Spectrom, 44 (2009) 861-878.

[20] H. Molina, R. Matthiesen, K. Kandasamy, A. Pandey, Comprehensive comparison of collision induced dissociation and electron transfer dissociation, Anal Chem, 80 (2008) 4825-4835.

[21] H.W.P. van den Toorn, S. Mohammed, J.W. Gouw, B. van Breukelen, A.J.R. Heck, Targeted SCX Based Peptide Fractionation for Optimal Sequencing by Collision Induced, and Electron

Transfer Dissociation, J Proteomics Bioinform, 1 (2008) 379-388.

[22] N. Taouatas, A.F. Altelaar, M.M. Drugan, A.O. Helbig, S. Mohammed, A.J. Heck, Strong cation exchange-based fractionation of Lys-N-generated peptides facilitates the targeted analysis of post-translational modifications, Mol Cell Proteomics, 8 (2009) 190-200.

[23] D.L. Swaney, G.C. McAlister, J.J. Coon, Decision tree-driven tandem mass spectrometry for shotgun proteomics, Nat Methods, 5 (2008) 959-964.

[24] A. Makarov, E. Denisov, A. Kholomeev, W. Balschun, O. Lange, K. Strupat, S. Horning, Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer, Anal Chem, 78 (2006) 2113-2120.

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

[26] Y. Zhang, S.B. Ficarro, S. Li, J.A. Marto, Optimized Orbitrap HCD for quantitative analysis of phosphopeptides, J Am Soc Mass Spectrom, 20 (2009) 1425-1434.

[27] T. Kocher, P. Pichler, M. Schutzbier, C. Stingl, A. Kaul, N. Teucher, G. Hasenfuss, J.M. Penninger, K. Mechtler, High precision quantitative proteomics using iTRAQ on an LTQ Orbitrap: a new mass spectrometric method combining the benefits of all, J Proteome Res, 8 (2009) 4743-4752.

[28] N. Mischerikow, P. van Nierop, K.W. Li, H.G. Bernstein, A.B. Smit, A.J. Heck, A.F. Altelaar, Gaining efficiency by parallel quantification and identification of iTRAQ-labeled peptides using HCD and decision tree guided CID/ETD on an LTQ Orbitrap, Analyst, 135 (2010) 2643-2652.

[29] J. Zhang, Y. Wang, S. Li, Deuterium isobaric amine-reactive tags for quantitative proteomics, Anal Chem, 82 (2010) 7588-7595.

[30] J.V. Olsen, J.C. Schwartz, J. Griep-Raming, M.L. Nielsen, E. Damoc, E. Denisov, O. Lange, P. Remes, D. Taylor, M. Splendore, E.R. Wouters, M. Senko, A. Makarov, M. Mann, S. Horning, A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed, Mol Cell Proteomics, 8 (2009) 2759-2769.

[31] T.P. Second, J.D. Blethrow, J.C. Schwartz, G.E. Merrihew, M.J. MacCoss, D.L. Swaney, J.D. Russell, J.J. Coon, V. Zabrouskov, Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures, Anal Chem, 81 (2009) 7757-7765.

[32] F. Sobott, S.J. Watt, J. Smith, M.J. Edelmann, H.B. Kramer, B.M. Kessler, Comparison of CID versus ETD based MS/MS fragmentation for the analysis of protein ubiquitination, J Am Soc Mass

Spectrom, 20 (2009) 1652-1659.

[33] N.E. Scott, B.L. Parker, A.M. Connolly, J. Paulech, A.V. Edwards, B. Crossett, L. Falconer, D. Kolarich, S.P. Djordjevic, P. Hojrup, N.H. Packer, M.R. Larsen, S.J. Cordwell, Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, HCD and ETD-MS applied to the N-linked glycoproteome of Campylobacter jejuni, Mol Cell Proteomics, (2010).

[34] N. Nagaraj, R.C. D'Souza, J. Cox, J.V. Olsen, M. Mann, Feasibility of large scale phosphoproteomics with HCD fragmentation, J Proteome Res, (2010).

[35] G.C. McAlister, D. Phanstiel, C.D. Wenger, M.V. Lee, J.J. Coon, Analysis of tandem mass spectra by FTMS for improved large-scale proteomics with superior protein quantification, Anal Chem, 82 (2010) 316-322.

[36] S. Gauci, A.O. Helbig, M. Slijper, J. Krijgsveld, A.J. Heck, S. Mohammed, Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach, Anal Chem, 81 (2009) 4493-4501.

[37] M.M. Savitski, T. Mathieson, I. Becher, M. Bantscheff, H-score, a mass accuracy driven rescoring approach for improved peptide identification in modification rich samples, J Proteome Res, 9 (2010) 5511-5516.

[38] B.N. Green, T. Hutton, S.N. Vinogradov, Analysis of complex protein and glycoprotein mixtures by electrospray ionization mass spectrometry with maximum entropy processing, Methods Mol Biol, 61 (1996) 279-294.

[39] H.W. van den Toorn, J. Munoz, S. Mohammed, R. Raijmakers, A.J. Heck, B. van Breukelen, RockerBox: analysis and filtering of massive proteomics search results, J Proteome Res, (2010).

[40] L. Kall, J.D. Canterbury, J. Weston, W.S. Noble, M.J. MacCoss, Semi-supervised learning for peptide identification from shotgun proteomics datasets, Nat Methods, 4 (2007) 923-925.

[41] R.X. Sun, M.Q. Dong, C.Q. Song, H. Chi, B. Yang, L.Y. Xiu, L. Tao, Z.Y. Jing, C. Liu, L.H. Wang, Y. Fu, S.M. He, Improved Peptide Identification for Proteomic Analysis Based on Comprehensive Characterization of Electron Transfer Dissociation Spectra, J Proteome Res, (2010).

[42] R.X. Sun, M.Q. Dong, C.Q. Song, H. Chi, B. Yang, L.Y. Xiu, L. Tao, Z.Y. Jing, C. Liu, L.H. Wang, Y. Fu, S.M. He, Improved peptide identification for proteomic analysis based on comprehensive characterization of electron transfer dissociation spectra, J Proteome Res, 9 (2010) 6354-6367.

[43] S. Kim, N. Mischerikow, N. Bandeira, J.D. Navarro, L. Wich, S. Mohammed, A.J. Heck, P.A. Pevzner, The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search, Mol Cell Proteomics, 9 (2010) 2840-2852.

[44] N. Mischerikow, A.F. Altelaar, J.D. Navarro, S. Mohammed, A.J. Heck, Comparative assessment of site assignments in CID and electron transfer dissociation spectra of phosphopeptides discloses limited relocation of phosphate groups, Mol Cell Proteomics, 9 (2010) 2140-2148.

[45] A.M. Palumbo, S.A. Smith, C.L. Kalcic, M. Dantus, P.M. Stemmer, G.E. Reid, Tandem mass spectrometry strategies for phosphoproteome analysis, Mass Spectrom Rev, (2011).

[46] A. Makarov, E. Denisov, O. Lange, S. Horning, Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer, J Am Soc Mass Spectrom, 17 (2006) 977-982.

Chapter 3

# Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry



Christian K. Frese[1,2], A.F. Maarten Altelaar[1,2], Henk van den Toorn[1,2], Dirk Nolting[3], Jens Griep-Raming[3], Albert J.R. Heck[1,2] and Shabaz Mohammed[1,2]

published in
Analytical Chemistry, 2012 Nov 20

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

[3] Thermo Fisher Scientific, Hanna-Kunath-Str. 11, 28199 Bremen, Germany

## ABSTRACT

Increasing peptide sequence coverage by tandem mass spectrometry improves confidence in database search-based peptide identification and facilitates mapping of post-translational modifications and de novo sequencing. Inducing two-fold fragmentation by combining electron-transfer and higher-energy collision dissociation (EThcD) generates dual fragment ion series and facilitates extensive peptide backbone fragmentation. After an initial electron-transfer dissociation step all ions including the unreacted precursor ions are subjected to collision induced dissociation which yields b/y- and c/z-type fragment ions in a single spectrum. This new fragmentation scheme provides richer spectra and substantially increases the peptide sequence coverage and confidence in peptide identification.

## INTRODUCTION

Peptide sequencing by tandem mass spectrometry is a central step in any proteomics experiment [1]. The most popular method for sequencing peptides involves collision induced dissociation (CID) [2]. Peptide identification is commonly achieved by database search algorithms that compare experimentally acquired fragmentation spectra against theoretically constructed spectra based on genomics derived protein databases [3]. Alternatively, database-independent de novo sequencing allows identification of proteins through direct interpretation of the MS/MS spectra. Both methods benefit from the level of completeness in sequencing. Often not all cleavages that are possible for the peptide are represented by fragment ions in CID. The resulting sequence gaps will increase ambiguity in peptide identification and hamper de novo sequencing. Conversely, the observation of C- and N-terminal fragment ion pairs strengthens the confidence in the presence of a certain backbone cleavage [4]. Over the past decade, electron capture dissociation (ECD) [5] and electron transfer dissociation (ETD) [6] have emerged as valuable alternative fragmentation techniques for peptide sequencing including the analysis of post-translational modifications (PTMs) [7].

The principle of ECD/ETD is based on generating odd-electron peptide cations. These cations then undergo free-radical driven N-C$\alpha$ backbone cleavage, which results primarily in c- and z$^{\bullet}$-type product ions. Both ECD and ETD facilitate random peptide backbone cleavages. The mechanism underlying electron-driven fragmentation techniques is fundamentally different from CID [8]. Peptide cations with high charge density are found to be favorable for ETD-based peptide sequencing [9] and it is generally accepted that ETD is most suitable for longer, more basic peptides with charge states $\geq$3 [10]. Both ECD and ETD provide spectral information that can often be used complementary to CID.[8] The first step in the ETD process is the formation of an intermediate radical cation, $[M+nH]^{(n-1)+\bullet}$ [11]. This non-dis-

sociative, charge reduced precursor and the unreacted precursor ion are often found to be dominant peaks in tandem mass spectra. Several techniques that aim to improve the dissociation of the charge reduced precursor ions, enhancing the total product ion yield, have been reported. These methods increase the internal energy of the peptide by using either elevated bath gas temperatures [12], photoactivation [13] or collisional activation [14]. For ETD, supplemental activation is usually performed in an ion trap by collisional activation of the charge reduced precursor ions. Coon and co-workers introduced a method that utilizes low-energy resonant-excitation CID (pseudo-MS3) of the charge reduced precursor subsequent to the electron transfer process (ETcaD) [15]. This method enhances the fragmentation efficiency (precursor-to-product-conversion) of ETD and results in increased formation of c- and z-ions, especially of doubly charged peptides. Other groups have reported on the use of collisional activation of the unreacted and charge-reduced precursor, respectively. Liu et al. applied "high-energy" resonant-excitation CID to activate the charge-reduced precursor (ET/CID-MS3) [16]. However, isolating the charge reduced precursor discards valuable sequence information that is generated by the MS2 ETD step. Campbell et al. applied resonant-excitation CID to the unreacted precursor ions simultaneously to the ETD reaction [17]. This fragmentation technique provided some additional b- and y-ions but did not increase the yield of c- and z-type ETD product ions from the charge reduced precursor.



The fact that the unreacted precursor and charge reduced precursor remain highly abundant after ETD is not desirable but common to all of the techniques described above. We reasoned, that both ion populations can be utilized in parallel to obtain b- and y-type ions alongside an increase in c- and z-type ions by means of beam-type collision induced dissociation. Here, we report on a two-fold fragmentation strategy implementing higher-energy collision dissociation (HCD) following an electron transfer reaction on a hybrid linear ion trap-Orbitrap mass spectrometer, referred to as EThcD.

## EXPERIMENTAL SECTION

**Materials.** Iodoacetamide (IAA), urea, sodium chloride, triethyl ammonium bicarbonate (TEAB), bovine serum albumin (BSA), alcohol dehydrogenase (ADH) and Glu-Fibrinopeptide B (Glufib) were obtained from Sigma-Aldrich (Steinheim, Germany). Formic acid, acetic acid, methanol and dimethylsulfoxide (DMSO) were obtained from Merck (Darmstadt, Germany). Acetonitrile was purchased from Biosolve (Valkenswaard, The Netherlands). Dithiothreitol (DTT) was supplied from Fluka (Buchs, Switzerland). Protease inhibitor cocktail and PhosSTOP phosphatase inhibitor were obtained from Roche (Mannheim, Germany). Sep-Pak Vac C18 1cc cartridges were obtained from Waters (Milford, MA). Sequencing grade trypsin was obtained from Promega (Madison, WI), and Lys-C was supplied by Wako Chemicals (Rich-

mond, VA). Metalloendopeptidase Lys-N was obtained from U-protein Express BV (Utrecht, The Netherlands). DMEM high glucose medium was purchased from PAA (Pasching, Austria). Fetal calf serum, penicillin/streptomycin, L-glutamine and phosphate buffered saline was obtained from Lonza (Breda, The Netherlands).

**Cell culture.** HeLa cells were cultured at 37°C and 5% CO2 in DMEM high glucose medium containing 10% FCS, 10mM L-Glutamine and 5% Penicillin/Streptomycin. Cells were harvested by trypsinization and washed three times with PBS buffer. HeLa cells were lysed in lysis buffer (8M Urea, 1 tablet Complete Mini EDTA-free protease inhibitor, 1 tablet PhosSTOP Phosphatase Inhibitor in 10 mL 50 mM TEAB) by sonication. The lysate was centrifuged for 15 min at 20.000 x g and the supernatant was transferred into new tubes. Protein concentration was determined by the BCA assay (Thermo Fisher Scientific, Rockford, IL).



**Protein digestion.** Proteins were reduced (2mM Dithiothreitol, 25 min at 56 degree) and alkylated (4 mM Iodoacetamide, 30 min at room temperature in the dark). Lys-C was added in a 1:75 (w/w) ratio and incubated for 4h at 37°C. The sample was diluted four times with 50 mM ammonium bicarbonate buffer. Trypsin was added in a 1:100 (w/w) ratio and the incubation was carried out overnight at 37°C. For the Lys-N digestion Lys-N was added in a 1:75 (w/w) ratio and incubated for 4h at 37°C. The sample was diluted four times with 50 mM ammonium bicarbonate buffer. Lys-N was added in a 1:100 (w/w) ratio and the incubation was carried out overnight at 37°C. Peptide mixtures were desalted by C18 solid phase extraction (Sep-Pak Vac C18 cartridge), lyophilized and reconstituted in 10% formic acid.

**Strong cation exchange.** Peptides were fractionated by strong-cation exchange (SCX) as previously reported [18]. Briefly, SCX was performed using a Zorbax BioSCX-Series II column (0.8 mm × 50 mm, 3.5 µm). Solvent A consisted of 0.05% formic acid in 20% acetonitrile, solvent B consisted of 0.05% formic acid, 0.5 M NaCl in 20% acetonitrile. The following gradient was used: 0–0.01 min (0–2% B); 0.01–8.01 min (2–3% B); 8.01–14.01 min (3–8% B); 14.01–28 min (8–20% B); 28–38 min (20–40% B); 38–48 min (40–90% B); 48–54 min (90% B); 54–60 min (0% B). 50 fractions were collected per sample, dried in vacuo, reconstituted in 10% formic acid and 5% DMSO and stored at -20°C.

**Mass Spectrometry.** All data was acquired on an ETD enabled Thermo Scientific LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). For direct infusion Glufib was dissolved in methanol/water/formic acid in a 50:49:1 ratio at a final concentration of 1 µM. A 100 µL Hamilton syringe was used for direct infusion and the flow rate was set to 1 µL/min. For LC-MS/MS experiments a Thermo Scientific EASY-nLC 1000 (Thermo Fisher Scientific, Odense, Denmark) was connected to the LTQ Orbitrap Velos mass spectrometer. Both the trap (200 mm × 100 µm I.D.) and the analytical column (40 cm × 75 µm I.D.) were packed in-house using Zorbax SB-C18, 1.8 µm (Agilent Technologies). The flow

rate was set to 150 nL/min. Mobile phase solvent A consisted of 0.1 M acetic acid in water, and mobile-phase solvent B consisted of 0.1 M acetic acid in 80% acetonitrile. The 110 min gradient was as follows: 10-40% solvent B within 91 min; 40-100% solvent B within 3 min; 100% solvent B for 2 min; 13 min solvent A. MS/MS methods were programmed in the data dependent acquisition mode. All data were acquired in the Orbitrap mass analyzer. The FT survey scan was set to a scan range from 350 to 1500 m/z at a resolution of 30,000. The 10 most intense peaks were subjected to either ETcaD, HCD or EThcD fragmentation. For all methods the MS/MS resolution was set to 7,500; precursor AGC target was set to 2e5; precursor isolation width was set to1.5 Da and the maximum injection time was set to 500 ms. For ETcaD the ETD reaction time was set to 50 ms. Charge state dependent ETD reaction time was enabled. Supplemental activation was enabled. For HCD normalized collision energy was set to 40%. For EThcD we modified the instrument firmware to allow all-ion HCD fragmentation after an ETD step (Figure 1). The ETD reaction time was set to 50 ms and HCD normalized collision energy to 30% (calculation based on precursor m/z and charge). The anion AGC target was set to 4e5 for both ETcaD and EThcD. The effective average, minimum and maximum anion injection times were 12, 3 and 20 ms, respectively.

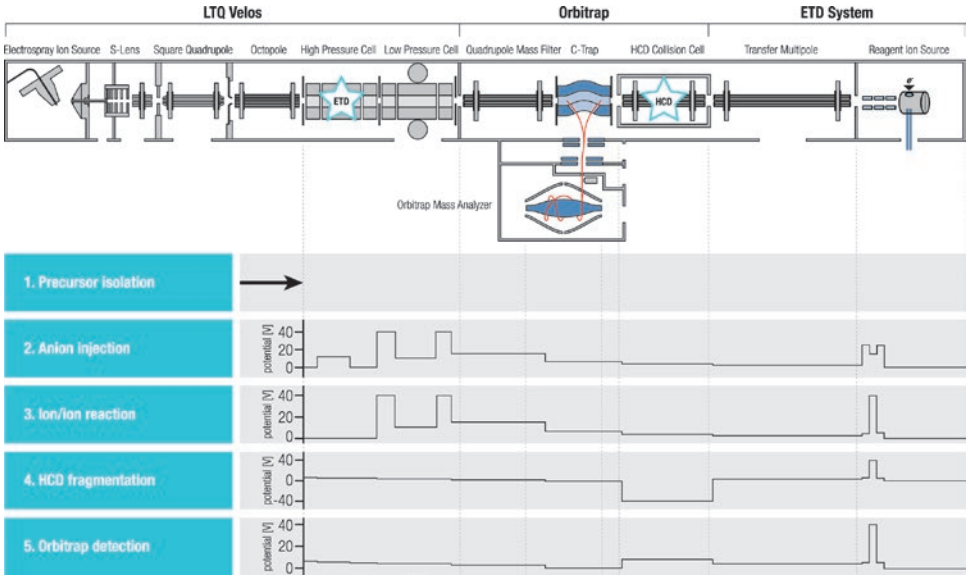



**Figure 1.** EThcD scan sequence setup. (1) Peptide precursor ions are accumulated in the linear ion trap (LTQ). (2) Fluoranthene anions are generated in the chemical ionization source at the rear of the instrument and transferred to the LTQ. (3) Cations and anions are mixed in the LTQ (reaction time 50 ms). (4) All ions are transmitted to the HCD collision cell and subjected to beam-type collision induced dissociation. Collision energy calculation is based on the precursor ion charge state and mass-to-charge ratio. (5) Fragment ions are detected in the Orbitrap mass analyzer.

**Data analysis.** Peak lists were generated using Thermo Scientific Proteome Discoverer 1.3 software (Thermo Fisher Scientific, Bremen, Germany). The non-fragment filter was used to simplify ETD spectra with the following settings: the precursor peak was removed within a 4 Da window, charged reduced precursors were removed within a 2 Da window, and neutral losses from charge reduced precursors were removed within a 2 Da window (the maximum neutral loss mass was set to 120 Da). MS/MS spectra were searched against the human Uniprot database (version v2010-12) including a list of common contaminants using SEQUEST [3]. The precursor mass tolerance was set to 10 ppm, the fragment ion mass tolerance was set to 0.02 Da. Enzyme specificity was set to Trypsin or Lys-N with 2 missed cleavages allowed. The following ion series were considered: c- and z- ions (ETcaD); b- and y- ions (HCD); and c-, z-, b- and y- ions (EThcD). Oxidation of methionine was used as variable modification and carbamidomethylation of cysteines was set as fixed modification. Percolator [19] was used to filter the PSMs for <1% false-discovery-rate. Additionally, only PSMs that pass the following filters were accepted: search engine rank 1, peptide length between 6 and 35 residues. The peptide hydrophobicity was calculated based on the Hopp-Wood scale [20].

## RESULTS & DISCUSSION

Initial proof of concept was obtained using the Glu-Fibrinopeptide B standard. Tandem mass spectra of the doubly charged precursor were acquired with ETD, ETcaD, HCD and EThcD (Figure 2). ETD fragmentation generated only three z-type fragment ions. Supplemental activation in ETcaD led to seven additional z-type ions and one additional c-type ion. HCD fragmentation generated twelve y-ions and one b-ion. For EThcD we subjected the same number of peptide cations to ETD followed by HCD all-ion fragmentation. Compared to ETcaD and HCD we detected all observed c-, z-, b- and y- ions plus two additional z-ions. These z-ions were not observed in ETcaD likely due to the low-mass cut-off in the linear ion trap. These data show that one can use a level of energy that is appropriate to fragment both the unreacted and charge reduced precursor whilst preserving all c- and z-ions generated by the preceding ETD reaction.

The most important parameters that drive the generation of dual ion series in EThcD are the ETD ion/ion reaction time and the HCD collision energy. For optimization we analyzed a simple peptide mixture (derived from tryptic digestion of bovine serum albumin and alcohol dehydrogenase) by LC-MS/MS using combinations of these two parameters. We opted to use the SEQUEST algorithm for data analysis as it is currently one of the few that allows for searching b-, y-, c- and z-ion series within the same spectrum.

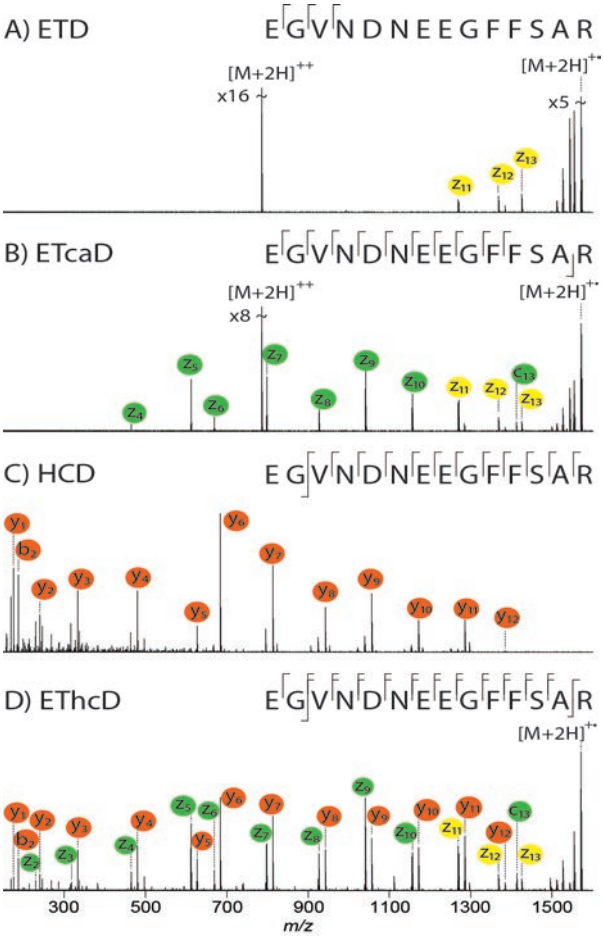



**Figure 2.** Tandem mass spectra of the 2+ peptide EGVNDNEEGFFSAR (i.e. Glufib) after fragmentation by either A) ETD, B) ETD with supplemental activation (ETcaD), C) HCD and D) EThcD. Colors represent the origin of the fragment ions: yellow, c/z-ions generated by ETD; green, c/z-ions derived from CID of the charge reduced precursor; orange, b/y ions derived from HCD of the precursor.

We found that an ion-ion reaction time of 50 ms and a HCD normalized collision energy of 30% provided the highest number of peptide-spectrum-matches (PSMs) and on average the highest Xcorr values.

We inferred from our initial observations that EThcD could improve confidence in peptide identification and increase the overall peptide sequence coverage. To further test our hypothesis we performed a large scale shotgun proteomics experiment. Aliquots of lysed HeLa cells were digested with either trypsin or metallopeptidase Lys-N [21]. The resulting peptide mixtures were fractionated by strong cation exchange (SCX), as previously

described [22]. The SCX fractions that contain peptides with up to one missed cleavage were analyzed three times by LC-MS/MS using either ETcaD, HCD or EThcD. We used identical instrument settings for all three methods with the only exception being the peptide ion activation method. The data was searched, initially, with SEQUEST and filtered for <1% false-discovery-rate (FDR) using percolator [19], a semi-supervised machine learning-based algorithm. Compared with a standard target-decoy approach [23] we found that percolator performed much better for all fragmentation techniques including EThcD, even though it was not designed for EThcD spectra.

**Table 1.** Results from large-scale analysis of SCX fractionated peptides.

| | Trypsin | | | Lys-N | | |
|---|---|---|---|---|---|---|
| | **ETcaD** | **HCD** | **EThcD** | **ETcaD** | **HCD** | **EThcD** |
| # MS/MS spectra | 100,204 | 118,047 | 105,266 | 118,877 | 124,352 | 114,267 |
| # PSM | 19,958 | 79,226 | 67,530 | 43,320 | 57,717 | 53,313 |
| ID % | 20% | 67% | 64% | 36% | 46% | 47% |
| # peptides | 10,848 | 34,859 | 30,760 | 15,814 | 21,752 | 20,277 |
| average SEQUEST Xcorr | 1.6 | 2.8 | 3.2 | 1.5 | 2.8 | 4.0 |
| average peptide sequence coverage | 85% | 85% | 94% | 69% | 84% | 93% |
| average number of b/c-ion pairs | 0.4 | | 1.9 | 0.4 | | 7.0 |
| average number of y/z-ion pairs | 2.1 | | 4.8 | 0.6 | | 4.3 |

We acquired for each dataset about 110,000 MS/MS spectra. In total we obtained 19,958 (ETcaD), 79,226 (HCD) and 67,530 (EThcD) PSMs at 1% FDR level from the trypsin dataset. ETcaD performance, as expected, was not optimal due to the use of doubly charged precursors and the need to keep consistent AGC values for FT detection. Strikingly, EThcD and HCD produced a very similar number of PSMs and performed equally well in terms of peptide identification success rate. Both techniques identified about two-third of all MS/MS spectra for the trypsin dataset (Table 1). We found that the overlap at the peptide level was very high between the three fragmentation techniques (Figure 3A). The distributions of peptide precursor area and peptide hydrophobicity were similar across all experiments (Figure 3 B,C). Compared to ETcaD we found that EThcD is not biased towards shorter peptides (Figure 3D). Similar trends were observed for the Lys-N dataset (Table 1, Figure 3 E-H).

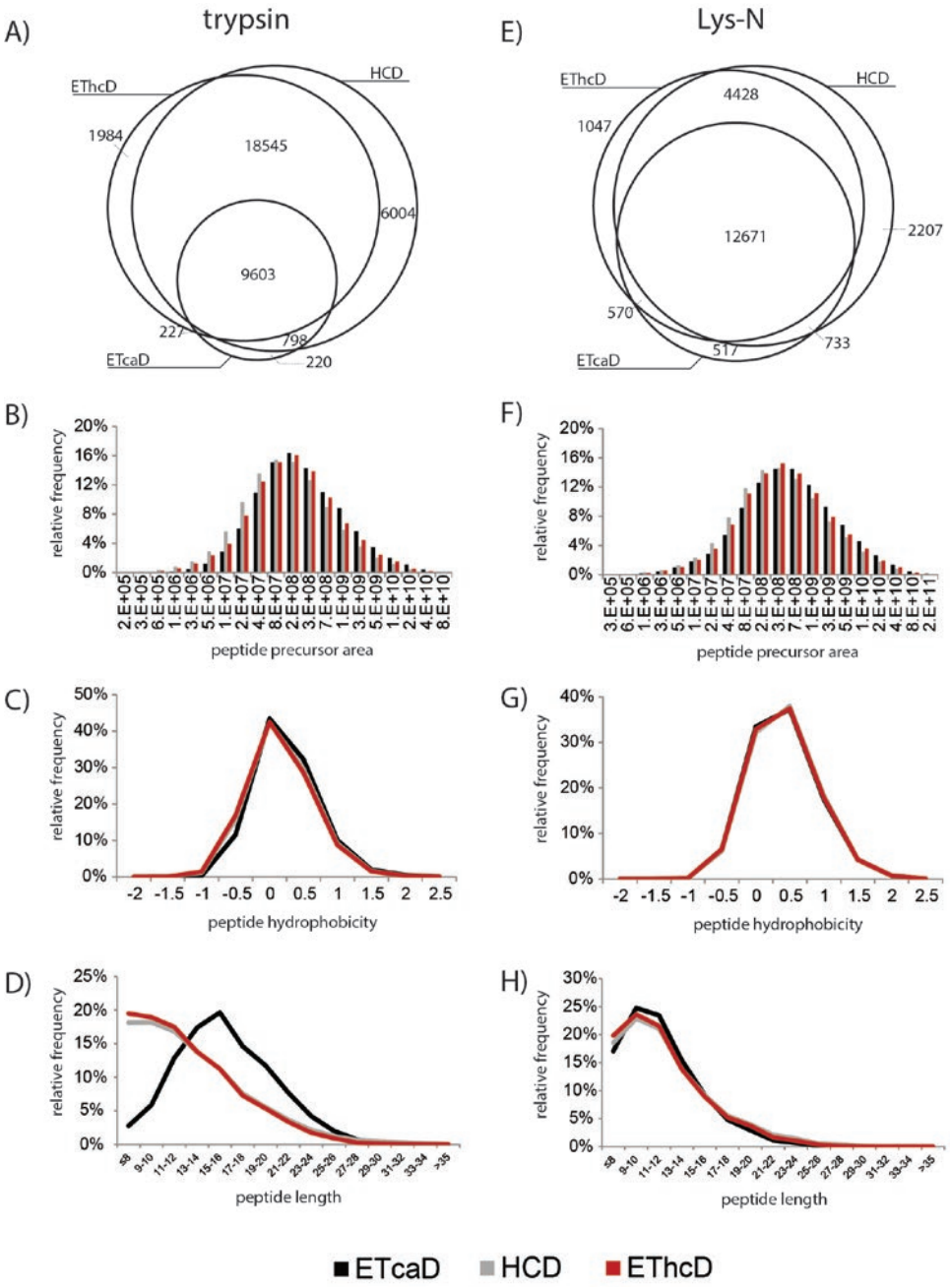



**Figure 3.** Peptide characteristics for the tryptic (A-D) and Lys-N dataset (E-H). Overlap of unique peptide identifications (A,E). Precursor area distribution (B,F). Distribution of peptide hydrophobicity (C,G). Distribution of peptide length (D,H).

Together these findings indicate that the performance of EThcD is independent of the amino acid composition. Across both datasets we found the average SEQUEST Xcorr to be substantially higher for EThcD than HCD, followed by ETcaD (Figure 4A,B). We calculated the ΔXcorrs for 28,059 unique tryptic peptides that are in common between EThcD and HCD (Figure 4C). Strikingly, we found that EThcD outperforms HCD for 72% of all peptides. This finding was more apparent in the Lys-N dataset where EThcD outperforms HCD for 91% of all PSMs (Figure 4D). This strengthens our assumption that EThcD generates more fragment ions, which yield more sequence information, consequently leading to higher Xcorrs.

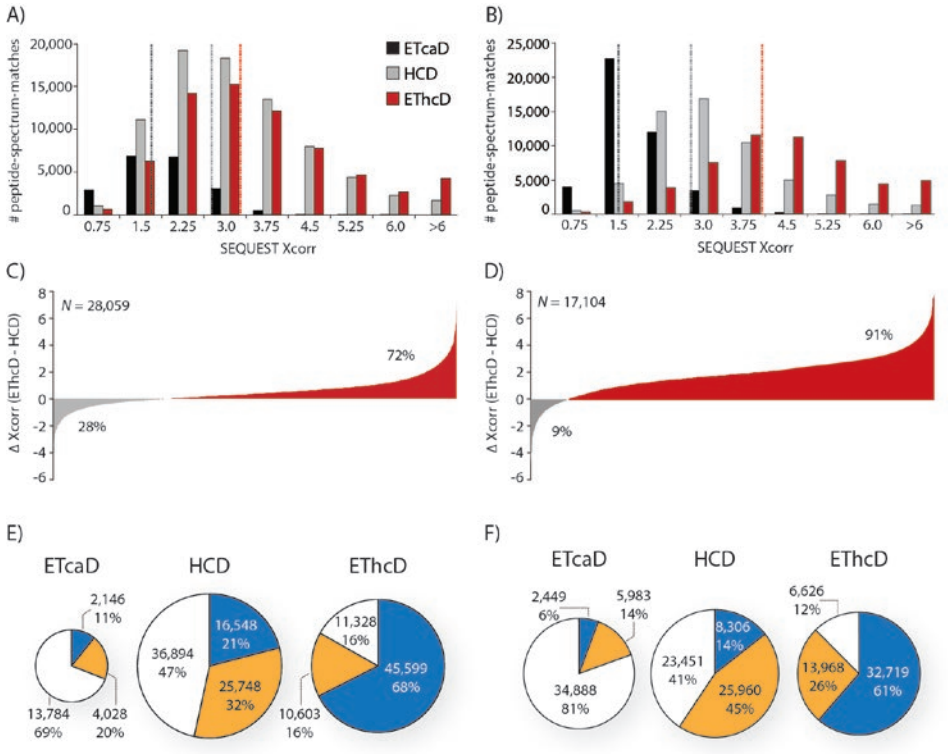



**Figure 4.** Performance statistics of ETcaD, HCD and EThcD for trypsin (A,C,E) and Lys-N datasets (B,D,F). Histograms illustrate the SEQUEST Xcorr distribution (A,B), whereby the dashed lines indicate the average Xcorr values. ΔXcorrs for peptides that were identified by both EThcD and HCD (C,D), highlighting that EThcD outscores HCD 3- and 10-fold for trypsin and Lys-N, respectively. Number of peptide sequence gaps (E,F). Please note all data was filtered to an FDR of 1% at the PSM level.

Maximizing peptide sequence coverage increases confidence in database search-based peptide identification (especially for site assignments of PTMs) but is also a fundamental necessity for de novo sequencing. We extracted the number of fragment ions from the raw data and calculated the number of gaps in the peptide sequence for each PSM. Remarkably, we found that EThcD provided zero gaps for 68% of all PSMs in the tryptic dataset (Figure 4E). This translates in a 3.2- and 6.1-fold increase compared to HCD and ETcaD, respectively. Moreover, we observed for only 16% of all PSMs more than one gap for EThcD whereas these numbers were 47% and 69% for HCD and ETcaD. The overall peptide sequence coverage was 85% for ETcaD, 85% for HCD and, strikingly, 94% for EThcD. This finding is in agreement with observations by Molina et al. who merged separately acquired CID and ETD spectra in silico [24]. From the Lys-N dataset we obtained similar results (Figure 4F). Here, zero gaps were observed for 61% of all PSMs and the average sequence coverage was 93%.



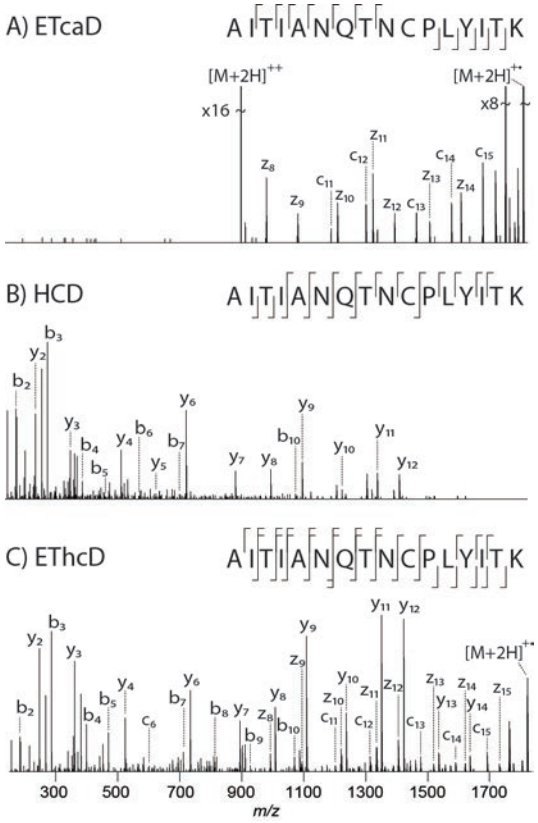



**Figure 5.** A) ETcaD, B) HCD and C) EThcD spectra of the doubly charged tryptic peptide AITIANQTNCPLYITK.

Figure 5 shows illustrative MS/MS spectra of a typical doubly charged tryptic peptide. Both ETcaD and HCD yielded sufficient fragment ions for identification. ETcaD generated 12 peptide backbone cleavages (Xcorr 1.03), whereas we observed 13 backbone cleavages in the HCD spectrum (Xcorr 4.35). However, only EThcD provided all 15 backbone cleavages and thus full peptide sequence coverage (Xcorr 7.22).

Next, we determined the number of b/c- and y/z-ion pairs in EThcD. These so-called "golden pairs" add another level of confidence for peptide identification [25]. Taking singly and doubly charged fragment ions into account we found on average 1.9 b/c-ion pairs and 4.8 y/z-ion pairs per tryptic peptide at an average peptide length of 12.5 amino acids. For Lys-N peptides we observed 7.0 b/c- and 4.3 y/z-ion pairs (average peptide length 11.8 residues). This observation provides further evidence that EThcD generates exceptionally data-rich spectra that facilitate complete peptide sequencing.

## CONCLUSION & OUTLOOK

Here, we introduce EThcD and highlight its great potential for peptide sequence analysis. This novel fragmentation scheme substantially improves the level of peptide backbone fragmentation observed which translates into a substantial increase in peptide sequence coverage. Furthermore, we show that EThcD provides more reliable peptide identification than HCD at a similar identification success rate. To open up future prospects in de novo sequencing and PTM mapping specialized bioinformatics tools need to be further developed to fully benefit from these very data-rich spectra.

## ASSOCIATED CONTENT

Supporting Information

Additional information as noted in the text. This material is available free of charge via the internet at http://pubs.acs.org

## ACKNOWLEDGMENTS

This work was supported by The Netherlands Proteomics Center, by the PRIME-XS project, grant number 262067, funded by the European Union 7th Framework Program and the Netherlands Organization for Scientific Research (NWO) with the VIDI grant (700.10.429).

## REFERENCES


[1] R. Aebersold, M. Mann, Mass spectrometry-based proteomics, Nature, 422 (2003) 198-207.

[2] D.F. Hunt, J.R. Yates, 3rd, J. Shabanowitz, S. Winston, C.R. Hauer, Protein sequencing by tandem mass spectrometry, Proc Natl Acad Sci U S A, 83 (1986) 6233-6237.

[3] J.K. Eng, A.L. Mccormack, J.R. Yates, An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database, J Am Soc Mass Spectr, 5 (1994) 976-989.

[4] D.M. Horn, R.A. Zubarev, F.W. McLafferty, Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry, P Natl Acad Sci USA, 97 (2000) 10313-10317.

[5] R.A. Zubarev, N.L. Kelleher, F.W. McLafferty, Electron capture dissociation of multiply charged protein cations. A nonergodic process, J Am Chem Soc, 120 (1998) 3265-3266.

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

[7] L.M. Mikesh, B. Ueberheide, A. Chi, J.J. Coon, J.E. Syka, J. Shabanowitz, D.F. Hunt, The utility of ETD mass spectrometry in proteomic analysis, Biochim Biophys Acta, 1764 (2006) 1811-1822.

[8] R.A. Zubarev, A.R. Zubarev, M.M. Savitski, Electron capture/transfer versus collisionally activated/induced dissociations: Solo or duet?, J Am Soc Mass Spectr, 19 (2008) 753-761.

[9] D.M. Good, M. Wirtala, G.C. McAlister, J.J. Coon, Performance characteristics of electron transfer dissociation mass spectrometry, Mol Cell Proteomics, 6 (2007) 1942-1951.

[10] C.K. Frese, A.F. Altelaar, M.L. Hennrich, D. Nolting, M. Zeller, J. Griep-Raming, A.J. Heck, S. Mohammed, Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos, Journal of proteome research, 10 (2011) 2377-2388.

[11] K. Breuker, H.B. Oh, C. Lin, B.K. Carpenter, F.W. McLafferty, Nonergodic and conformational control of the electron capture dissociation of protein cations, P Natl Acad Sci USA, 101 (2004) 14011-14016.

[12] S.J. Pitteri, P.A. Chrisman, S.A. McLuckey, Electron-transfer ion/ion reactions of doubly protonated peptides: Effect of elevated bath gas temperature, Analytical Chemistry, 77 (2005) 5662-5669.

[13] A.R. Ledvina, G.C. McAlister, M.W. Gardner, S.I. Smith, J.A. Madsen, J.C. Schwartz, G.C. Stafford, J.E.P. Syka, J.S. Brodbelt, J.J. Coon, Infrared Photoactivation Reduces Peptide Folding and Hydrogen-Atom Migration following ETD Tandem Mass Spectrometry, Angew Chem Int Edit, 48 (2009) 8526-8528.

[14] D.M. Horn, Y. Ge, F.W. McLafferty, Activated ion electron capture dissociation for mass spectral sequencing of larger (42 kDa) proteins, Analytical Chemistry, 72 (2000) 4778-4784.

[15] D.L. Swaney, G.C. McAlister, M. Wirtala, J.C. Schwartz, J.E. Syka, J.J. Coon, Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors, Anal Chem, 79 (2007) 477-485.

[16] C.W. Liu, C.C. Lai, Effects of Electron-Transfer Coupled with Collision-Induced Dissociation (ET/CID) on Doubly Charged Peptides and Phosphopeptides, J Am Soc Mass Spectr, 22 (2011) 57-66.

[17] J.L. Campbell, J.W. Hager, J.C. Le Blanc, On performing simultaneous electron transfer dissociation and collision-induced dissociation on multiply protonated peptides in a linear ion trap, J Am Soc Mass Spectr, 20 (2009) 1672-1683.

[18] E.L. de Graaf, A.F. Altelaar, B. van Breukelen, S. Mohammed, A.J. Heck, Improving SRM Assay Development: A Global Comparison between Triple Quadrupole, Ion Trap, and Higher Energy CID Peptide Fragmentation Spectra, Journal of proteome research, 10 (2011) 4334-4341.

[19] L. Kall, J.D. Canterbury, J. Weston, W.S. Noble, M.J. MacCoss, Semi-supervised learning for peptide identification from shotgun proteomics datasets, Nat Methods, 4 (2007) 923-925.

[20] T.P. Hopp, K.R. Woods, A computer program for predicting protein antigenic determinants, Mol Immunol, 20 (1983) 483-489.

[21] N. Taouatas, M.M. Drugan, A.J. Heck, S. Mohammed, Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase, Nature methods, 5 (2008) 405-407.

[22] J. Munoz, T.Y. Low, Y.J. Kok, A. Chin, C.K. Frese, V. Ding, A. Choo, A.J. Heck, The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells, Mol Syst Biol, 7 (2011) 550.

[23] J.E. Elias, S.P. Gygi, Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry, Nature methods, 4 (2007) 207-214.

[24] H. Molina, R. Matthiesen, K. Kandasamy, A. Pandey, Comprehensive comparison of collision induced dissociation and electron transfer dissociation, Anal Chem, 80 (2008) 4825-4835.

[25] M.L. Nielsen, M.M. Savitski, R.A. Zubarev, Improving protein identification using complementary fragmentation techniques in fourier transform mass spectrometry, Molecular & cellular proteomics : MCP, 4 (2005) 835-845.

Chapter 4

# Unambiguous Phosphosite Localization Using Electron-Transfer / Higher-Energy Collision Dissociation (EThcD)

Christian K. Frese[1,2], Houjiang Zhou[1,2],
Thomas Taus[3], A.F. Maarten Altelaar[1,2], Karl Mechtler[3,4],
Albert J.R. Heck[1,2] and Shabaz Mohammed[1,2]

published in
Journal of Proteome Research, 2013 Mar 1

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

[3] Research Institute of Molecular Pathology (IMP), Dr. Bohrgasse 7, A-1030 Vienna, Austria

[4] Institute of Molecular Biotechnology (IMBA), Vienna, Austria

## ABSTRACT

We recently introduced a novel fragmentation scheme combining electron-transfer and higher-energy collision dissociation (coined EThcD) for improved peptide fragmentation and identification. We reasoned that phosphosite localization, one of the major hurdles in high-throughput phosphoproteomics, could also highly benefit from the generation of such EThcD spectra. Here, we systematically assessed the impact on phosphosite localization utilizing EThcD in comparison to methods employing either ETD or HCD, respectively, using a defined synthetic phosphopeptide mixture and also using a larger dataset of $Ti^{4+}$-IMAC enriched phosphopeptides from a tryptic human cell line digest. In combination with a modified version of phosphoRS we observed that in the majority of cases EThcD generated richer and more confidently identified spectra, resulting in superior phosphosite localization scores. Our data demonstrates the distinctive potential of EThcD for PTM localization, also beyond protein phosphorylation.



## INTRODUCTION

Reversible phosphorylation of proteins is a key regulatory mechanism in living cells [1]. Protein phosphorylation can modulate protein activity, turnover, sub-cellular localization, complex formation, folding and degradation. Dynamic phosphorylation plays a pivotal role in almost all biological processes including cell division, differentiation, polarization and apoptosis [2]. Moreover, it is an important switch in cellular signal transduction [3]. The importance of this post-translational modification (PTM) for cell biology has driven the development of novel mass spectrometric tools for sensitive and global detection of phosphorylation [4, 5]. However, the analysis of phosphorylated peptides by mass spectrometry is still not as straightforward as for "regular", unmodified peptides. One of the major challenges in phosphoproteomics is to improve MS level representation since phosphopeptides are usually present at sub-stoichiometric levels. Hence, an enrichment step is necessary to enable deeper penetration of the phosphoproteome. Enrichment is typically performed by chromatography [6], antibodies [7] or metal-ion/metal oxide affinity-based [8, 9] techniques. Two other main challenges are the identification of phosphopeptides and confident localization of the corresponding phosphosite [10]. The challenge is caused by the higher lability of the phosphate group when compared to the amide bond. A number of strategies have been proposed to circumvent poor fragmentation and improve sequence and site diagnostic fragmentation, including the use of neutral loss-triggered MS/MS/MS [11] and multistage activation (MSA) [12] in ion traps, the use of beam-type CID fragmentation [13], and electron capture/transfer dissociation [14] or a combination of some of these approaches [9, 15].

Once phosphopeptide identification is feasible through sufficient peptide backbone fragments it can still be challenging to pinpoint the true phosphosite. This becomes more difficult as the number of potential phosphorylation sites within the peptide sequence increases. In principle, unambiguous phosphosite localization requires site-determining fragment ions [16]. Direct validation is feasible through detection of a fragment ion that carries the phosphate group. Neutral loss fragment ions can be used as well; however, since they exhibit the same mass as a water loss from an unmodified residue they do not directly confirm the correct site [17]. Diagnostic phosphosite-specific fragments facilitate pinpointing the correct phosphosite [18-20]. Several algorithms and programs have been developed to enable automatic phosphosite localization [3, 16, 21-26]. These software tools are based on distinct but similar approaches and they all aim to provide a metric that allows for assessment of the confidence in phosphosite localization. Recently, Taus et al. have reported on a new algorithm, termed phosphoRS [27], which presently is uniquely compatible with CID, HCD and ETD fragmentation and was optimized for both low- and high- resolution MS/MS spectra. PhosphoRS provides individual localization probabilities for all potential phosphosites in a given peptide.



Generally, all scoring tools depend on the quality of the MS/MS spectra. The more site-determining ions are detected the higher the confidence in phosphosite localization. We have recently introduced a novel fragmentation scheme combining electron-transfer and higher-energy collision dissociation, termed EThcD [28]. This method employs dual fragmentation to generate both b/y- and c/z- ions which leads to very fragment ion and thus data-rich MS/MS spectra. Compared to HCD and ETD, we found a substantial increase in peptide backbone fragmentation, which translated into a remarkable average peptide sequence coverage of ~94% for tryptic peptides. We reasoned that localization of post-translational modifications could also highly benefit from EThcD spectra. Here, we systematically assessed the impact on phosphosite localization using EThcD. In this work we evaluate the performance of EThcD in comparison to ETD and HCD using a defined synthetic phosphopeptide mixture and also on a larger dataset of $Ti^{4+}$-IMAC enriched phosphopeptides, all in combination with a modified version of phosphoRS.

## EXPERIMENTAL SECTION

**Materials**. All chemicals were purchased from Sigma-Aldrich (Steinheim, Germany) unless otherwise stated. Formic acid and ammonia were obtained from Merck (Darmstadt, Germany). Acetonitrile was purchased from Biosolve (Valkenswaard, The Netherlands).

**Sample preparation**. Protein from HeLa cells was harvested and digested with trypsin, as previously described [29]. $Ti^{4+}$-IMAC beads were prepared as reported elsewhere [30, 31]. Phosphopeptides were enriched as previously described.[32] Briefly, GeI-loader tips that were plugged with C8 material (3M, Zoeterwoude, The Netherlands) were filled up to 1 cm with $Ti^{4+}$-IMAC beads. Columns were equilibrated with loading buffer (80% ACN, 6% TFA). Peptides were reconstituted in loading buffer, loaded onto the columns and washed with washing buffer 1 (50% ACN, 0.5% TFA, 200 mM NaCl) and subsequently washing buffer 2 (50% ACN, 0.1% TFA). Phosphopeptides were eluted with elution buffer 1 (10% $NH_3$ in $H_20$) followed by elution buffer 2 (80% ACN, 2% FA). Eluate was acidified and diluted with formic acid to a final acetonitrile concentration of <5%, split into three equal amounts and directly analyzed by single run LC-MS/MS utilizing ETD, HCD and EThcD, respectively.

**Mass spectrometry.** All data was acquired on an ETD enabled Thermo Scientific LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). A Thermo Scientific EASY-nLC 1000 (Thermo Fisher Scientific, Odense, Denmark) was connected to the LTQ Orbitrap Velos mass spectrometer. ETD, HCD and EThcD methods were set up as previously described [28]. Briefly, all spectra were acquired in the Orbitrap at a resolution of 7,500. For HCD the normalized collision energy was set to 40%. The ETD reaction time was set to 50 ms for ETD and EThcD. Supplemental activation was enabled for ETD. HCD normalized collision energy was set to 30% for EThcD (calculation based on precursor m/z and charge state). The anion AGC target was set to 4e5 for both ETD and EThcD.

**Data analysis.** Peak lists were generated using Thermo Scientific Proteome Discoverer 1.3 software (Thermo Fisher Scientific, Bremen, Germany). The non-fragment filter was used to simplify ETD spectra with the following settings: the precursor peak was removed within a 4 Da window, charged reduced precursors were removed within a 2 Da window, and neutral losses from charge reduced precursors were removed within a 2 Da window (the maximum neutral loss mass was set to 120 Da). MS/MS spectra were searched against a database containing the synthetic phosphopeptide sequences and the human Uniprot database (version v2010-12), respectively, including a list of common contaminants using SEQUEST or Mascot (Matrix Science, UK). The precursor mass tolerance was set to 10 ppm, the fragment ion mass tolerance was set to 0.02 Da. Enzyme specificity was set to Trypsin with two missed cleavages allowed. Data from the synthetic phosphopeptide mixture were searched with no enzyme specificity. Oxidation of methionine and phosphorylation (S,T,Y) were used as variable modification and carbamidomethylation of cysteines was set as fixed modification. Percolator [33] was used to filter the PSMs for <1% false-discovery-rate. Phosphorylation sites were localized by applying a custom version of phosphoRS [27] (v3.0 – EThcD enabled) that has been expanded to allow analysis of EThcD data [28]. Briefly, the algorithm considers both HCD- and ETD-type fragment ions at the same time. While singly and doubly

charged b- and y-type fragment ions including neutral loss of phosphoric acid ($H_3PO_4$) are considered for site localization, only singly charged c-, z-radical and z-prime ions are scored.

## RESULTS & DISCUSSION

Increasing the confidence in phosphosite localization is a key challenge in phosphoproteomics. Site-determining fragment ions are required to unambiguously pinpoint the correct phosphosite. Observing all possible peptide backbone cleavages in a single MS/MS spectrum substantially simplifies phosphosite localization. Recently, we showed that EThcD enables complete peptide sequencing through dual fragmentation [28]. In EThcD, the peptide precursor is initially subjected to an ion/ion reaction with fluoranthene anions in a linear ion trap, which generates c- and z-ions. However, the unreacted precursor and the charge-reduced precursor remain highly abundant after ETD. In the second step HCD all-ion fragmentation is applied to all ETD derived ions. This generates b- and y-ions from the unreacted precursor and simultaneously increases the yield of c- and z-ions by fragmentation of the charge reduced precursor. Since the remaining unreacted precursor population is higher charged than the ETD-derived fragment ions one can apply a level of energy that fragments the precursor but does not induce secondary fragmentation of c- and z-ions. Here, we continue to explore the benefits of this novel fragmentation mode for the analysis of phosphopeptides.



### Evaluation of phosphosite localization by EThcD using a defined phosphopeptide mixture

To evaluate the potential added value of phosphopeptide analysis by EThcD we initially used a defined mixture of well-characterized synthetic phosphopeptides. This mixture consists of 30 phosphopeptides of varying length with up to four phosphorylated residues (see Table 1 for a complete list). We analyzed this mixture by LC-MS/MS employing ETD, HCD and EThcD fragmentation, respectively. We used identical instrument settings with the only exception being the parameters for peptide dissociation, which were set to the for each method optimized values. The data was searched with SEQUEST and the PSMs were manually validated and filtered (7 ppm peptide mass tolerance, search engine rank 1, absolute Xcorr threshold 0.4). Additionally, we considered only PSMs for which the injection time did not max out (<500 ms), i.e. the target number of ions was reached. Note that this precaution was taken to exclude the number of ions as a variable that might impair the quality of fragmentation. We calculated the average precursor ion purity (PIP) [34] for each dataset and found similar values, which were approximately 95% for all three techniques.

**Table 1.** Overview of synthetic phosphopeptide sequences used in this study. a) Peptide identifier. b) Peptide sequence. Red letters indicate phosphorylated residues. c) Position of phosphorylation. d) Number of PSMs (Phosphosites correctly annotated by SEQUEST/identified). e) Average SEQUEST Xcorrs.

| | | | number of PSM[d] | | | average SEQUEST Xcorrs[e] | | |
|---|---|---|---|---|---|---|---|---|
| primeXS #[a] | phosphopeptide sequence[b] | phosphosite[c] | ETD | HCD | EThcD | ETD | HCD | EThcD |
| 1 | NSVEQGRRL | S2 | | | 16/16 | | | 1.9 |
| 2 | APPDNLPSPGGSR | S12 | 18/22 | 18/26 | 24/30 | 0.7 | 1.2 | 2.0 |
| 3 | LIEDNEYTAR | Y7 | | 20/22 | 26/26 | | 2.3 | 2.4 |
| 4 | RSDGGHTVLHR | S2 | 1/1 | 1/1 | 2/2 | 2.3 | 2.3 | 2.6 |
| 5 | ENIMRSENSESQLTSK | S6 | 11/11 | 11/11 | 6/9 | 1.7 | 3.1 | 3.8 |
| 6 | QLGEPEKSQDSSPVLSELK | S8, S16 | 1/4 | 7/8 | 3/3 | 1.5 | 2.1 | 4.4 |
| 7 | QLGEPEKSQDSSPVLSELK | S8, S11, S16 | 1/4 | 2/3 | 2/2 | 1.3 | 2.1 | 4.5 |
| 8 | KFLSLASNPELLNLPSSVIK | S4, S17 | | | | | | |
| 9 | THILLFLPKSVSDYEGK | S10 | | | | | | |
| 10 | WWGSGPSGSGGSGGGK | S12 | 6/8 | 4/9 | 9/9 | 0.9 | 1.2 | 2.5 |
| 11 | WWGSGPSGSGGSGGGK | S9, S12 | 1/5 | 0/7 | 8/9 | 0.8 | 1.0 | 2.6 |
| 12 | WWGSGPSGSGGSGGGK | S7, S9, S12 | 1/2 | 2/4 | 4/4 | 0.8 | 0.7 | 1.9 |
| 13 | WWGSGPSGSGGSGGGK | S4, S7, S9, S12 | | | | | | |
| 14 | GSRAAKTLGSLGK | S2, T7, S10 | 41/41 | 32/32 | 35/35 | 2.1 | 0.9 | 2.2 |
| 15 | RYRVAKLGSSLGK | Y2, S9, S10 | 31/31 | 18/18 | 3/3 | 1.5 | 1.4 | 2.1 |
| 16 | RSRVVKLGSSSLGK | S2, S9, S10, S11 | 22/22 | 4/4 | | 1.4 | 1.5 | |
| 18 | SAGSRLTLSGR | S4, T7, S9 | 6/6 | | 10/12 | 1.9 | | 2.0 |
| 19 | LQLQGLDLSSR | S9, S10 | 3/3 | 6/6 | 5/5 | 0.6 | 0.8 | 1.3 |
| 20 | LEKTIDDLEDK | T4 | 10/10 | 14/14 | 16/16 | 1.3 | 2.1 | 2.9 |
| 21 | RGDDSFGDK | S5 | | | 1/1 | | | 2.2 |
| 22 | GSAHGSGYSGASGHSGAGSGGK | S2 | 3/3 | 1/1 | 2/2 | 2.1 | 4.1 | 2.9 |
| 23 | GSAHGSGYSGASGHSGAGSGGK | S2, S6 | 6/7 | 7/7 | 5/5 | 2.2 | 3.6 | 3.0 |
| 24 | GSAHGSGYSGASGHSGAGSGGK | S2, S6, S9 | 5/6 | 1/2 | | 2.1 | 2.5 | |
| 25 | GSAHGSGYSGASGHSGAGSGGK | S2, S6, S9, S12 | 1/3 | | | 1.6 | | |
| 26 | LQLQGLDLSSR | S9 | | 4/7 | 8/8 | | 0.9 | 2.0 |
| 27 | TYETTLEK | T1, Y2, T4 | | | 3/3 | | | 1.9 |
| 28 | VPQVSTPTLVEVSR | S5, T8, T8 | | 5/5 | 7/7 | | 1.5 | 1.4 |
| 29 | TASDTDSSYAIPTAGMSPSR | Y9 | 5/6 | 12/14 | 10/13 | 1.4 | 3.1 | 3.1 |
| 30 | SVENLPEAGITHEQR | T11 | 17/17 | 22/22 | 11/11 | 1.3 | 2.6 | 2.6 |
| 31 | FGSSNTDSAGALGTLR | S3, S4, T14 | 4/4 | 3/14 | 15/17 | 1.6 | 1.6 | 2.8 |

Together, these stringent criteria ensure that the activation technique is the only variable that controls the fragmentation behavior. A summary of the data from this direct comparison is given in table 2. Similar numbers of PSMs were identified for all three fragmentation techniques. We found that EThcD provided 248 PSMs while these numbers were 237 and 216 for HCD and ETD, respectively. Out of the 30 unique synthetic phosphopeptides injected ETD, HCD and EThcD identified 21, 22 and 24, respectively. We found the average SEQUEST Xcorr being highest for EThcD (2.5) followed by HCD (1.9) and ETD (1.5), which is in line with our previous results for non-modified peptides [28]. The SEQUEST algorithm correctly annotated the known phosphosites in 79% of ETD and 78% of HCD data. Significantly, for EThcD this was over 95% (of all PSMs), which directly reflects the higher spectral quality, due to the generation of both b/y and c/z ions. This initial data suggests that EThcD provides even more extensive backbone fragmentation of phosphorylated peptides than ETD or HCD alone, facilitating sensitive phosphosite localization with very high confidence. It should be noted that the application of a site localization algorithm would be prudent for real-life samples since the true phosphorylation sites are unknown.

**Table 2.** Analysis of 30 synthetic phosphopeptides.

| | ETD | HCD | EThcD |
|---|---|---|---|
| # PSM | 216 | 237 | 248 |
| # unique peptides | 21/30 | 22/30 | 24/30 |
| average Xcorr | 1.5 | 1.9 | 2.5 |
| % PSM with correctly localized phosphosite (SEQUEST) | 79% | 78% | 95% |
| # phosphosites with phosphoRS site probability >99% | 478 | 410 | 423 |
| % phosphosites with phosphoRS site probability >99% | 96% | 95% | 97% |

Recently, Taus et al. described phosphoRS, a novel tool to improve confident localization of phosphosites [27]. The software is based on validated peptide identifications provided by database search engines and calculates site probabilities for each potential phosphosite in the peptide sequence. For this study we used a modified version of phosphoRS that also enables assessment of individual phosphosite probabilities for EThcD fragmentation. We analyzed each dataset using phosphoRS and found that it performs equally well for all three fragmentation techniques. Of all true phosphosites 96% (ETD), 95% (HCD) and 97% (EThcD), respectively, were assigned a site probability >99%, which corresponds to a very high confidence in site localization (Table 2). Together, these findings suggest that EThcD generates MS/MS spectra that contain sufficient fragment ions for the unambiguous and sensitive phosphorylation site localization.



### Phosphosite localization of $Ti^{4+}$-IMAC enriched phosphopeptides by EThcD

Next, we assessed the performance of EThcD for phosphosite localization on a larger dataset. We used $Ti^{4+}$-IMAC material for the enrichment of phosphopeptides from a tryptic digest of HeLa cells and analyzed equal amounts (corresponding to enriched phosphopeptides from 100 µg of protein) by LC-MS/MS with ETD, HCD and EThcD, respectively. All three methods generated a similar number of MS/MS spectra. All spectra were searched with SEQUEST. The ETD data was also searched with Mascot because we found SEQUEST to perform poorly for doubly charged phosphopeptides. Note that other search engines such as OMSSA or SpectrumMill might provide larger number of identifications for ETD data [35]. However, these algorithms are currently not compatible with EThcD data and phosphoRS analysis within the Proteome Discoverer software environment. All identified PSMs were then filtered for <1% FDR using percolator to ensure consistency.

In total we identified 2217 (ETD), 4179 (HCD) and 3594 (EThcD) phospho-PSMs (Table 3). Our initial analysis of a defined synthetic phosphopeptide mixture demonstrated that EThcD performs at least on the same level as HCD in terms of peptide identification. However, the overall identification success rate in the $Ti^{4+}$-IMAC dataset was slightly lower for EThcD compared to HCD. This can be attributed to the rigid automatic FDR filtering. The MS/MS

spectra from the synthetic phosphopeptide mixture were manually validated whereas the $Ti^{4+}$-IMAC dataset was computationally filtered to <1% FDR.

**Table 3.** LC-MS/MS analysis of $Ti^{4+}$-IMAC enriched tryptic phosphopeptides originating from a cellular lysate using ETD, HCD and EThcD, respectively.

| | ETD | HCD | EThcD |
|---|---|---|---|
| # PSM | 2266 | 4282 | 3679 |
| ID success rate | 25% | 51% | 44% |
| average Xcorr | 1.9 | 2.5 | 3.2 |
| % average peptide sequence coverage | 83% | 81% | 92% |
| # phospho-PSM | 2217 | 4179 | 3594 |
| # phospho-sites >99% pRS probability | 2002 | 4291 | 3942 |
| % phospho-sites >99% pRS probability | 81% | 89% | 95% |



The application of EThcD, in comparison to ETD or HCD alone, significantly increases the number of fragment ions observed in the MS/MS scans. On the one hand EThcD spectra contain more sequence information, which is beneficial for inferring the peptide sequence and PTM localization. On the other hand, these additional fragment ions may also match to random peptide sequences, increasing their score and hampering the differentiation between correct and incorrect matches. Consequently, the chance for a high scoring random match will be elevated. Similar to the increased average score of decoy hits also the true hits are likely to provide on average higher scores. Depending on whether the distance between the two score distributions decreases or increases, the identification success rate will be higher or lower. Since the ID success rate is slightly lower for EThcD compared to HCD alone, the negative effect of higher-scoring random matches might be more pronounced. Thus, higher score cut-offs need to be applied in order to reach the desired FDR. A standard target-decoy approach [36] against a reversed concatenated database revealed the FDR for EThcD (2.6%) being almost twice as high compared to HCD (1.4%), which provides further evidence for this hypothesis.

Next, we calculated the average peptide sequence coverage for all PSM. As expected, EThcD provided a substantial increase in sequence coverage (92%) compared to HCD (81%) and ETD (83%). Obtaining near-complete peptide sequence coverage tremendously simplifies phosphosite localization. We used the extended phosphoRS algorithm to validate our assumption. Remarkably, EThcD provided for 95% of all phosphosites a confident site localization probability of >99%. In the HCD dataset we found that 89% of all phosphosites were assigned with a confident site localization probability >99%, while this was only 81% for ETD dataset. We recalculated these number for all peptides that contain >2 residues that can be phosphorylated because singly phosphorylated peptides with only one potential

phosphorylation site could bias the results towards HCD. Of all phosphosites from this subset of peptides 97% (ETcaD), 93% (EThcD) and 87% (HCD), respectively, were assigned a localization probability >99%.

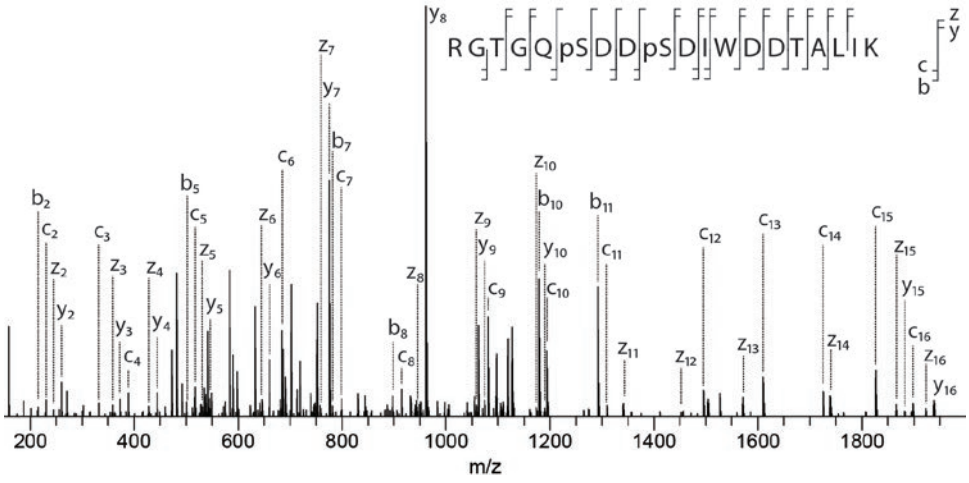



**Figure 1.** EThcD MS/MS spectrum of the doubly phosphorylated peptide RGTGQSDDSDIWDDTALIK, containing in total four potential phosphorylation sites. EThcD generates dual ion series that enable phosphorylation site localization with very high confidence (phosphoRS site probabilities: T(3): 0.0%; S(6): 100.0%; S(9): 100.0%; T(15): 0.0%). SEQUEST Xcorr 7.79.



For multiply phosphorylated peptides site localization becomes more challenging. Figure 1 shows an MS/MS spectrum of a doubly phosphorylated peptide upon EThcD fragmentation. The overall sequence coverage is 89% taking b/y- and c/z-ions into account. Six out of 18 amino acid bond cleavages are represented by c- and b-ions (referred to as "golden pairs" [37]). Additionally, we observed 11 z/y-ion pairs, which strengthens the argument that EThcD provides extensive sequence information that facilitates pinpointing the correct phosphorylation site. More than 95% of the phosphosites from all doubly phosphorylated peptides were assigned with a site localization probability >99%, highlighting that EThcD performs equally well with singly and doubly phosphorylated peptides. A known limitation of ETD is its inability to cleave the N-Cα bond N-terminal to proline [38, 39]. This can hamper phosphosite localization for proline-rich peptides. Generation of dual ion series in EThcD can overcome this issue. Figure 2 shows the EThcD spectrum of a singly phosphorylated peptide that contains four serine residues. The c- and z-ions derived from the ETD step cover only the N-terminal part of the peptide and the site probability is only 50%. The additional y-ions derived from the subsequent HCD activation provide supporting sequence information and cover also the two serine residues next to the prolines which enables unambiguous phosphosite localization.

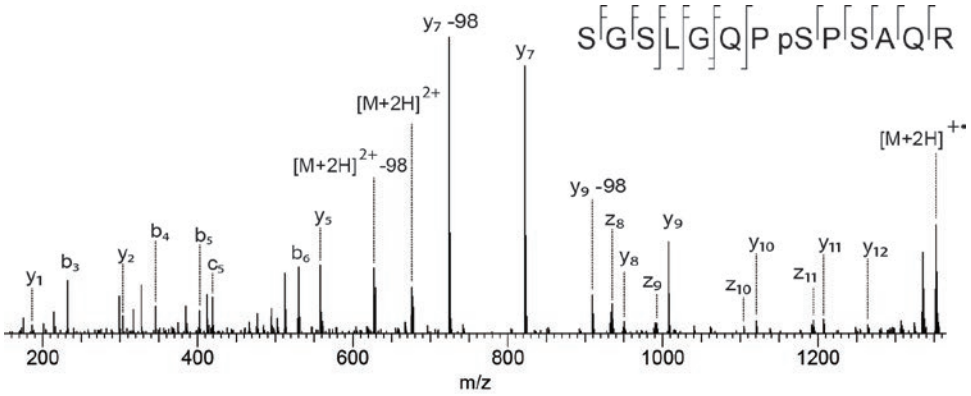



**Figure 2.** EThcD spectrum of a proline-containing phosphopeptide that contains four serine residues, one of which is phosphorylated. ETD does not cleave the N-Cα bond N-terminal to proline and the phosphorylation site probability is only 50% based on c- and z-ions alone. Dual fragmentation by EThcD generates complementary sequence information from c/z- and b/y-ions. Here, the exact phosphosite is revealed by y-ions that cover the corresponding phosphosite (phosphoRS site probabilities: S(1): 0.0; S(3): 0.0; S(8): 99.5; S(10): 0.5). SEQUEST Xcorr 4.10.



## CONCLUSION & OUTLOOK

Here we have evaluated the potential of EThcD in improving the analysis of phosphopeptides. Our data highlights the benefit of dual ion series as generated by EThcD fragmentation. We observed for a defined phosphopeptide mixture average higher SEQUEST Xcorr values, higher peptide sequence coverage and more confident phosphosite localization in EThcD compared to ETD and HCD. This finding was confirmed when we analyzed a complex phosphopeptide sample resulting from a $Ti^{4+}$-IMAC enrichment of peptides from a cellular lysate. This is in line with recent reports that showed that confidence in phosphorylation site localization increases when multiple separately acquired MS/MS spectra (e.g. ETD/CID or MSA/ETD) are combined for scoring [25, 26]. For this larger dataset we observed that the identification success rate was slightly lower for EThcD compared to HCD. This can be attributed to the use of conventional database search engines that are not optimized for spectra that contain dual ion series [40]. However, the fact that both peptide sequence coverage and the percentage of localized phosphosites are higher for EThcD than for HCD suggests that once a peptide was identified, further analyses such as site localization benefit from the more data-rich EThcD spectra.

In EThcD often c/b- and z/y-ion pairs are observed that increase the confidence in a particular peptide backbone cleavage [41]. We speculate that the identification success rate of EThcD for phosphopeptides can be improved by novel or optimized data analysis tools.

Finally, we reason that EThcD can also be beneficial and used to improve the localization of other post-translational modifications such as ubiquitination, glycosylation or acetylation.

## ASSOCIATED CONTENT

Supporting Information

Additional information as noted in the text. This material is available free of charge via the internet at http://pubs.acs.org

## ACKNOWLEDGMENTS

We thank Mathias Madalinski for peptide synthesis of the primeXS phosphopeptide mixture. This research was performed within the framework of the PRIME-XS project, grant number 262067, funded by the European Union 7th Framework Program. Additionally, the Netherlands Proteomics Centre, a program embedded in the Netherlands Genomics Initiative, is kindly acknowledged for financial support as well as the Netherlands Organization for Scientific Research (NWO) with the VIDI grant (700.10.429). Work in the Mechtler lab was supported by the European Commission via the FP7 projects MeioSys and PRIME-XS, the Austrian Science Fund via the Special Research Program Chromosome Dynamics (SFB-F3402).



## REFERENCES

[1] P. Cohen, The regulation of protein function by multisite phosphorylation--a 25 year update, Trends Biochem Sci, 25 (2000) 596-601.

[2] T. Hunter, Signaling--2000 and beyond, Cell, 100 (2000) 113-127.

[3] J.V. Olsen, B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, M. Mann, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks, Cell, 127 (2006) 635-648.

[4] B. Eyrich, A. Sickmann, R.P. Zahedi, Catch me if you can: mass spectrometry-based phosphoproteomics and quantification strategies, Proteomics, 11 (2011) 554-570.

[5] M. Mann, O.N. Jensen, Proteomic analysis of post-translational modifications, Nature biotechnology, 21 (2003) 255-261.

[6] S. Di Palma, M.L. Hennrich, A.J. Heck, S. Mohammed, Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis, J Proteomics, 75 (2012) 3791-3813.

[7] J. Rush, A. Moritz, K.A. Lee, A. Guo, V.L. Goss, E.J. Spek, H. Zhang, X.M. Zha, R.D. Polakiewicz, M.J. Comb, Immunoaffinity profiling of tyrosine phosphorylation in cancer cells, Nature biotechnology, 23 (2005) 94-101.

[8] S.B. Ficarro, M.L. McCleland, P.T. Stukenberg, D.J. Burke, M.M. Ross, J. Shabanowitz, D.F. Hunt, F.M. White, Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces

cerevisiae, Nature biotechnology, 20 (2002) 301-305.

[9] H. Zhou, T.Y. Low, M.L. Hennrich, H. van der Toorn, T. Schwend, H. Zou, S. Mohammed, A.J. Heck, Enhancing the identification of phosphopeptides from putative basophilic kinase substrates using Ti (IV) based IMAC enrichment, Molecular & cellular proteomics : MCP, 10 (2011) M110 006452.

[10] P.J. Boersema, S. Mohammed, A.J. Heck, Phosphopeptide fragmentation and analysis by mass spectrometry, J Mass Spectrom, 44 (2009) 861-878.

[11] S.A. Beausoleil, M. Jedrychowski, D. Schwartz, J.E. Elias, J. Villen, J. Li, M.A. Cohn, L.C. Cantley, S.P. Gygi, Large-scale characterization of HeLa cell nuclear phosphoproteins, P Natl Acad Sci USA, 101 (2004) 12130-12135.

[12] M.J. Schroeder, J. Shabanowitz, J.C. Schwartz, D.F. Hunt, J.J. Coon, A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry, Analytical chemistry, 76 (2004) 3590-3598.

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

[14] A. Chi, C. Huttenhower, L.Y. Geer, J.J. Coon, J.E. Syka, D.L. Bai, J. Shabanowitz, D.J. Burke, O.G. Troyanskaya, D.F. Hunt, Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry, P Natl Acad Sci USA, 104 (2007) 2193-2198.

[15] H. Molina, R. Matthiesen, K. Kandasamy, A. Pandey, Comprehensive comparison of collision induced dissociation and electron transfer dissociation, Anal Chem, 80 (2008) 4825-4835.

[16] S.A. Beausoleil, J. Villen, S.A. Gerber, J. Rush, S.P. Gygi, A probability-based approach for high-throughput protein phosphorylation analysis and site localization, Nature biotechnology, 24 (2006) 1285-1292.

[17] A. Stensballe, O.N. Jensen, J.V. Olsen, K.F. Haselmann, R.A. Zubarev, Electron capture dissociation of singly and multiply phosphorylated peptides, Rapid Commun Mass Sp, 14 (2000) 1793-1800.

[18] C.D. Kelstrup, O. Hekmat, C. Francavilla, J.V. Olsen, Pinpointing phosphorylation sites: Quantitative filtering and a novel site-specific x-ion fragment, Journal of proteome research, 10 (2011)

[19] Y.S. Shin, J.H. Moon, M.S. Kim, Observation of phosphorylation site-specific dissociation of singly protonated phosphopeptides, J Am Soc Mass Spectr, 21 (2010) 53-59.

[20] P.M. Gehrig, B. Roschitzki, D. Rutishauser, S. Reiland, R. Schlapbach, Phosphorylated serine and threonine residues promote site-specific fragmentation of singly charged, arginine-containing peptide ions, Rapid communications in mass spectrometry : RCM, 23 (2009) 1435-1445.

[21] B. Lu, C. Ruse, T. Xu, S.K. Park, J. Yates, 3rd, Automatic validation of phosphopeptide identifications from tandem mass spectra, Analytical chemistry, 79 (2007) 1301-1310.

[22] C.M. Bailey, S.M. Sweet, D.L. Cunningham, M. Zeller, J.K. Heath, H.J. Cooper, SLoMo: automated site localization of modifications from ETD/ECD mass spectra, Journal of proteome research, 8 (2009) 1965-1971.

[23] M.M. Savitski, T. Mathieson, I. Becher, M. Bantscheff, H-score, a mass accuracy driven rescoring approach for improved peptide identification in modification rich samples, J Proteome Res, 9 (2010)

[24] B.E. Ruttenberg, T. Pisitkun, M.A. Knepper, J.D. Hoffert, PhosphoScore: an open-source phosphorylation site assignment tool for MSn data, Journal of proteome research, 7 (2008) 3054-3059.

[25] T.A. Hansen, M. Sylvester, O.N. Jensen, F. Kjeldsen, Automated and High Confidence Protein Phosphorylation Site Localization Using Complementary Collision-Activated Dissociation and Elec-

tron Transfer Dissociation Tandem Mass Spectrometry, Analytical chemistry, (2012).

[26] M. Vandenbogaert, V. Hourdel, O. Jardin-Mathe, J. Bigeard, L. Bonhomme, V. Legros, H. Hirt, B. Schwikowski, D. Pflieger, Automated Phosphopeptide Identification Using Multiple MS/MS Fragmentation Modes, Journal of proteome research, (2012).

[27] T. Taus, T. Kocher, P. Pichler, C. Paschke, A. Schmidt, C. Henrich, K. Mechtler, Universal and confident phosphorylation site localization using phosphoRS, Journal of proteome research, 10 (2011) 5354-5362.

[28] C.K. Frese, M. Altelaar, H.W. van den Toorn, D. Nolting, J. Griep-Raming, A.J. Heck, S. Mohammed, Towards Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry, Analytical chemistry, (2012).

[29] A.F. Altelaar, C.K. Frese, C. Preisinger, M.L. Hennrich, A.W. Schram, H.T. Timmers, A.J. Heck, S. Mohammed, Benchmarking stable isotope labeling based quantitative proteomics, J Proteomics, (2012).

[30] H. Zhou, M. Ye, J. Dong, G. Han, X. Jiang, R. Wu, H. Zou, Specific phosphopeptide enrichment with immobilized titanium ion affinity chromatography adsorbent for phosphoproteome analysis, Journal of proteome research, 7 (2008) 3957-3967.

[31] Z. Yu, G. Han, S. Sun, X. Jiang, R. Chen, F. Wang, R. Wu, M. Ye, H. Zou, Preparation of monodisperse immobilized Ti(4+) affinity chromatography microspheres for specific enrichment of phosphopeptides, Anal Chim Acta, 636 (2009) 34-41.

[32] H. Zhou, M. Ye, J. Dong, E. Corradini, A. Cristobal, A.J.R. Heck, H. Zou, S. Mohammed, Robust phosphoproteome enrichment using monodisperse microspheres-based immobilized titanium (IV) ion affinity chromatography, Nat Protoc, accepted (2012).

[33] L. Kall, J.D. Canterbury, J. Weston, W.S. Noble, M.J. MacCoss, Semi-supervised learning for peptide identification from shotgun proteomics datasets, Nat Methods, 4 (2007) 923-925.

[34] P. Mertins, N.D. Udeshi, K.R. Clauser, D.R. Mani, J. Patel, S.E. Ong, J.D. Jaffe, S.A. Carr, iTRAQ labeling is superior to mTRAQ for quantitative global proteomics and phosphoproteomics, Molecular & cellular proteomics : MCP, 11 (2012) M111 014423.

[35] K. Kandasamy, A. Pandey, H. Molina, Evaluation of Several MS/MS Search Algorithms for Analysis of Spectra Derived from Electron Transfer Dissociation Experiments, Analytical Chemistry, 81 (2009) 7170-7180.

[36] J.E. Elias, S.P. Gygi, Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry, Nature methods, 4 (2007) 207-214.

[37] D.M. Horn, R.A. Zubarev, F.W. McLafferty, Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry, P Natl Acad Sci USA, 97 (2000) 10313-10317.

[38] H.J. Cooper, R.R. Hudgins, K. Hakansson, A.G. Marshall, Secondary fragmentation of linear peptides in electron capture dissociation, Int J Mass Spectrom, 228 (2003) 723-728.

[39] W. Li, C. Song, D.J. Bailey, G.C. Tseng, J.J. Coon, V.H. Wysocki, Statistical analysis of electron transfer dissociation pairwise fragmentation patterns, Analytical chemistry, 83 (2011) 9540-9545.

[40] M.S. Kim, J. Zhong, K. Kandasamy, B. Delanghe, A. Pandey, Systematic evaluation of alternating CID and ETD fragmentation for phosphorylated peptides, Proteomics, 11 (2011) 2568-2572.

[41] M.L. Nielsen, M.M. Savitski, R.A. Zubarev, Improving protein identification using complementary fragmentation techniques in fourier transform mass spectrometry, Molecular & cellular proteomics : MCP, 4 (2005) 835-845.

Chapter 5

# Characterization of Electron-Transfer Dissociation in the Orbitrap Velos HCD Cell

Christian K. Frese[1,2], Dirk Nolting[3],
A.F. Maarten Altelaar[1,2], Jens Griep-Raming[3],
Shabaz Mohammed[1,2] and Albert J.R. Heck[1,2]

published in
Journal of the American Society for Mass Spectrometry, 2013 Apr 20



[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

[3] Thermo Fisher Scientific, Hanna-Kunath-Str. 11, 28199 Bremen, Germany

## ABSTRACT

Electron transfer dissociation (ETD) is commonly employed in ion traps utilizing RF fields that facilitate efficient electron transfer reactions. Here, we explore performing ETD in the HCD collision cell on an Orbitrap Velos instrument by applying a static DC gradient axially to the rods. This gradient enables simultaneous three dimensional, charge sign independent, trapping of cations and anions, initiating electron transfer reactions in the center of the HCD cell where oppositely charged ions clouds overlap. Here, we evaluate this mode of operation for a number of tryptic peptide populations and the top-down sequence analysis of ubiquitin. Our preliminary data shows that performing ETD in the HCD cell provides similar fragmentation as ion trap-ETD but requires further optimization to match performance of ion trap-ETD.

## INTRODUCTION

In 1998, McLafferty and co-workers introduced electron capture dissociation (ECD) as a novel dissociation technique for protein sequence analysis [1, 2]. The dissociation process involves the capture of one or more near-thermal electrons by protein cations that are kept in an electro-magnetic trap of a FT-ICR instrument. The strong magnetic fields enable efficient trapping and storage of electrons on a millisecond timescale. Not soon after, the implementation of ECD on ion trap instruments using RF fields has also been reported [3, 4]. However, simultaneous trapping of near-thermal electrons and peptide cations for several milliseconds is technically difficult on commercial instrument. Nevertheless, it has been shown that it is possible to obtain similar performance to an FT-ICR [5]. Moreover, maintaining trapped electrons in an RF field can increase their kinetic energy which might induce "hot-ECD"-like secondary fragmentation [4, 6]. To date, ECD is available on FT-ICR, digital ion traps and QTOF instruments [1, 3, 5, 7-9].

A similar dissociation technique that is equally based on the transfer of an electron from an anion to a cation has been developed by Hunt and co-workers in 2004 [10-12]. They utilized superimposed RF voltages on a quadrupole linear ion trap (QLT) to enable electron transfer dissociation (ETD) which provides ECD-like fragmentation. The principle of ETD involves the generation of peptide cations and reagent anions, where electron transfer is achieved by mixing of the two ion populations, followed by straightforward mass analysis of resulting product ions. They accomplished this by applying secondary RF fields to the end lenses of a linear ion trap, which enables three dimensional charge sign independent trapping [10]. In recent years, ETD has evolved into a widely utilized fragmentation technique for peptide and protein sequence analysis [13-15] and is often used complementary to collision induced

dissociation (CID) [16-18]. Furthermore, ETD preserves labile post-translational modifications [11, 19, 20] and is also a very versatile tool for top-down protein analysis [15, 21-23].

Over the past years the research into the implementation of ETD on a plethora of different configurations has gained momentum [10, 24-27]. The increasing popularity is likely based on the prevalence of ion trap instruments, which are also easily adaptable to allow ion/ion reactions. Isolation and storage of cations is the first step of commonly employed ion trap-CID [28] and is also essential for ETD. The additional steps to enable ETD reactions include the injection of anions and the actual electron transfer by ion/ion reactions through combined or overlapping storage of cations and anions in a confined space. Reagent anions can be introduced either by axial injection via the front-end [29, 30], through the rear of the ion-trap [10, 26] or in-between the rods [31]. There are in principle two ways to initiate the electron transfer process: passing ion/ion reactions [31, 32] or mutual storage [10, 33, 34] of both cations and anions in the trap. The former approach requires only one of the ions to be trapped while the oppositely charged ions are forced to pass through the trapped ion cloud. On the other hand, mutual storage requires charge sign independent trapping [10]. This is usually achieved by additional RF voltages applied to the front and end lens of the QLT, which increases complexity of the instrumentation and requires additional RF generators.

Here, we explore the capabilities of performing electron transfer dissociation in the HCD collision cell on an Orbitrap Velos utilizing a static DC gradient along the axis. The DC gradient enables three-dimensional charge state independent trapping in combination with the radial trapping capabilities of the HCD collision cell. The main advantages of performing ETD in the HCD cell over ion trap-ETD is a higher space charge capacity, potentially allowing more ions in a single scan event which increases the probability of detecting low abundant fragment ions. Coon and co-workers have recently reported on a dedicated segmented reaction cell for the same purpose [35, 36]. Our preliminary data reveal that ETD in the HCD cell is feasible but does not outperform ETD in the Orbitrap's linear ion trap. Nevertheless, using ETD in the HCD cell we demonstrate near full sequence coverage on ubiquitin in a top-down proteomics approach.



## EXPERIMENTAL SECTION

**Sample preparation.** Total protein from HeLa cells was lysed, digested with trypsin and resulting peptides were fractionated by strong cation exchange chromatography as previously described [37]. Bovine ubiquitin (Sigma-Aldrich, Steinheim, Germany) was dissolved in methanol/water/formic acid in a 50:49:1 ratio to a final concentration of 2 µM.

**Mass spectrometry.** Ubiquitin data was acquired using a 100 µL Hamilton syringe in direct infusion mode. LC-MS/MS data was acquired on an Orbitrap Velos (Thermo Fisher Scientific, Bremen, Germany) coupled to an Agilent 1200 HPLC system previously described [16]. The peptide precursor automatic gain control (AGC) target value was varied as indicated. The precursor isolation window was set to 4.0 m/z to keep the injection times low for high AGC target values. All MS/MS data was acquired in the Orbitrap (resolution LC-MS/MS 7,500; direct infusion 100,000). After the survey scan (m/z 350-1500) the four most abundant peptides were subjected to fragmentation. For normal ETD the reaction time was set to 50 ms (anion AGC 4e5 for LC-MS/MS and 1e7 for direct infusion of ubiquitin). We used modified instrument firmware to enable ETD reactions in the HCD cell. Briefly, an additional DC gradient voltage was applied to the four printed circuit boards (PCBs) located in-between the rods of the HCD collision cell (see also Figure 1) [38]. The DC gradient voltage applied over the length of the auxiliary electrodes was set to 100 V. Note, that the actual electrical field strength depends on the cell geometry. The simulated trapping field in the center of the HCD cell is approximately 100 mV/cm. For ETD in the HCD cell the anion injection time equals the effective ETD reaction time and was varied as mentioned in the text. The scan sequence is as follows: 1) all voltages of the HCD cell including the gradient DC are set. 2) Precursor cations are injected into the HCD cell. 3) Anions are injected and accumulated for the specified injection time. The ETD reaction takes place in the center of the HCD cell where ion clouds overlap. 4) The voltages are changed to transfer the resulting product ions to the Orbitrap mass analyzer. Note that ETD in the HCD cell requires larger numbers of cations and anions to achieve sufficient overlap of the oppositely charged ion clouds.

**Data analysis.** LC-MS/MS data was processed as previously described [16]. Briefly, MS/MS spectra were converted into peak lists, charge deconvoluted and searched using Mascot 2.3.02 (Matrix Science, UK). Ubiquitin data was charge deconvoluted using the Xtract function within Xcalibur 2.2 (Thermo Fisher Scientific, San Jose, CA). Data were analyzed using ProteinProspector (http://prospector.ucsf.edu) and ProSight PTM (http://prosightptm.northwestern.edu), as described elsewhere [23, 39].

## RESULTS & DISCUSSION

In this work we use an additional DC gradient to enable ETD reactions via three-dimensional charge sign independent trapping in the HCD collision cell of an Orbitrap Velos instrument. Initial experiments have demonstrated the general feasibility of this mode of operation [38]. The static DC gradient voltage is applied to PCBs that are installed in-between the rods of the HCD collision cell (Figure 1A). This gradient generates an additional axial field, similar to what can be obtained by using conical rods [40] or by the installation of additional

electrodes along the rods of a linear collision cell [41]. The Z-shaped potential generates two potential minima for different charge signs (Figure 1B). Consequently, cations and anions are trapped at opposing sides of the HCD cell. Increasing the number of ions leads to higher space charge, which, as a consequence, force the oppositely charged ion clouds to partially overlap in the center of the HCD cell (Figure 1C).

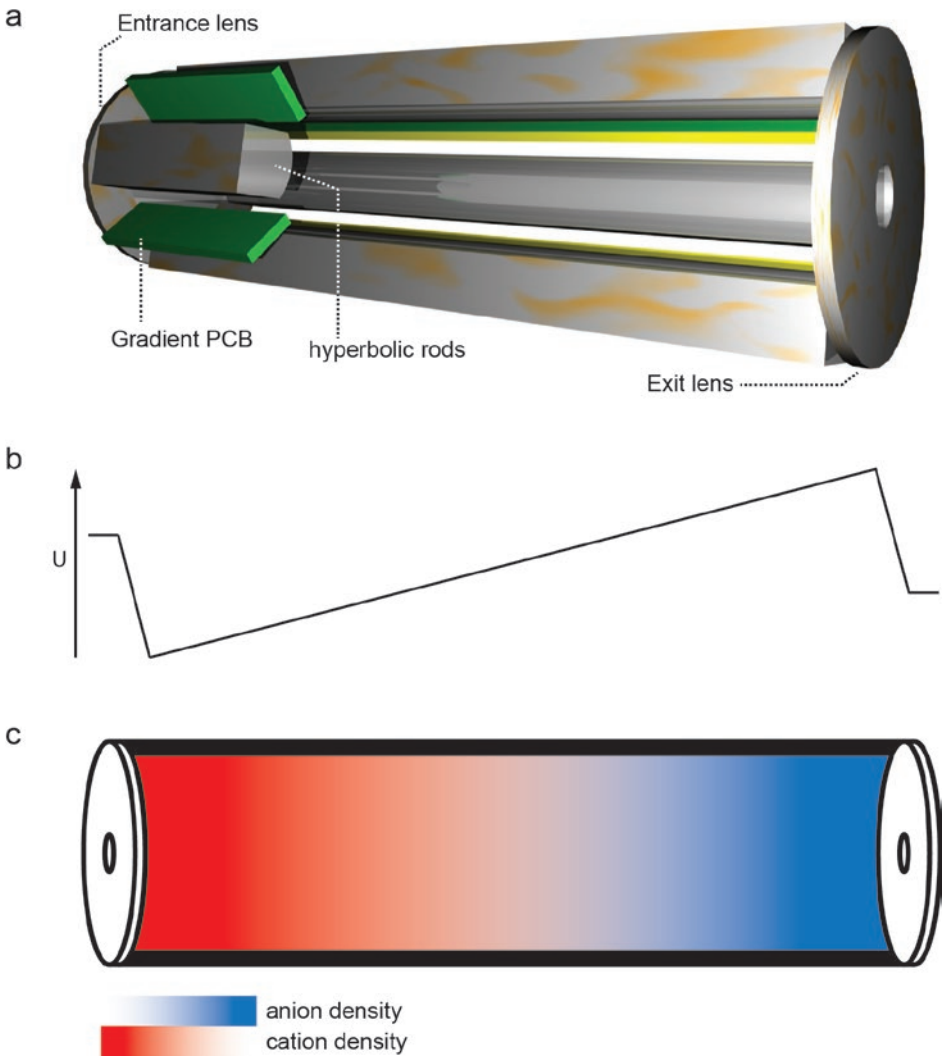



**Figure 1.** A static DC gradient voltage enables electron transfer reactions in the HCD collision cell. A) An additional static DC voltage of 100 V is applied over the length of four printed circuit boards that are placed in-between the rods. The effective DC potential in the center of the HCD cell is approximately 100 mV/cm. B) The resulting Z-shaped potential in axial dimension enables three-dimensional charge sign independent trapping of oppositely charged ions. C) High ion numbers increase space charge which forces the ion cation and anion clouds to overlap in the center of the collision cell.

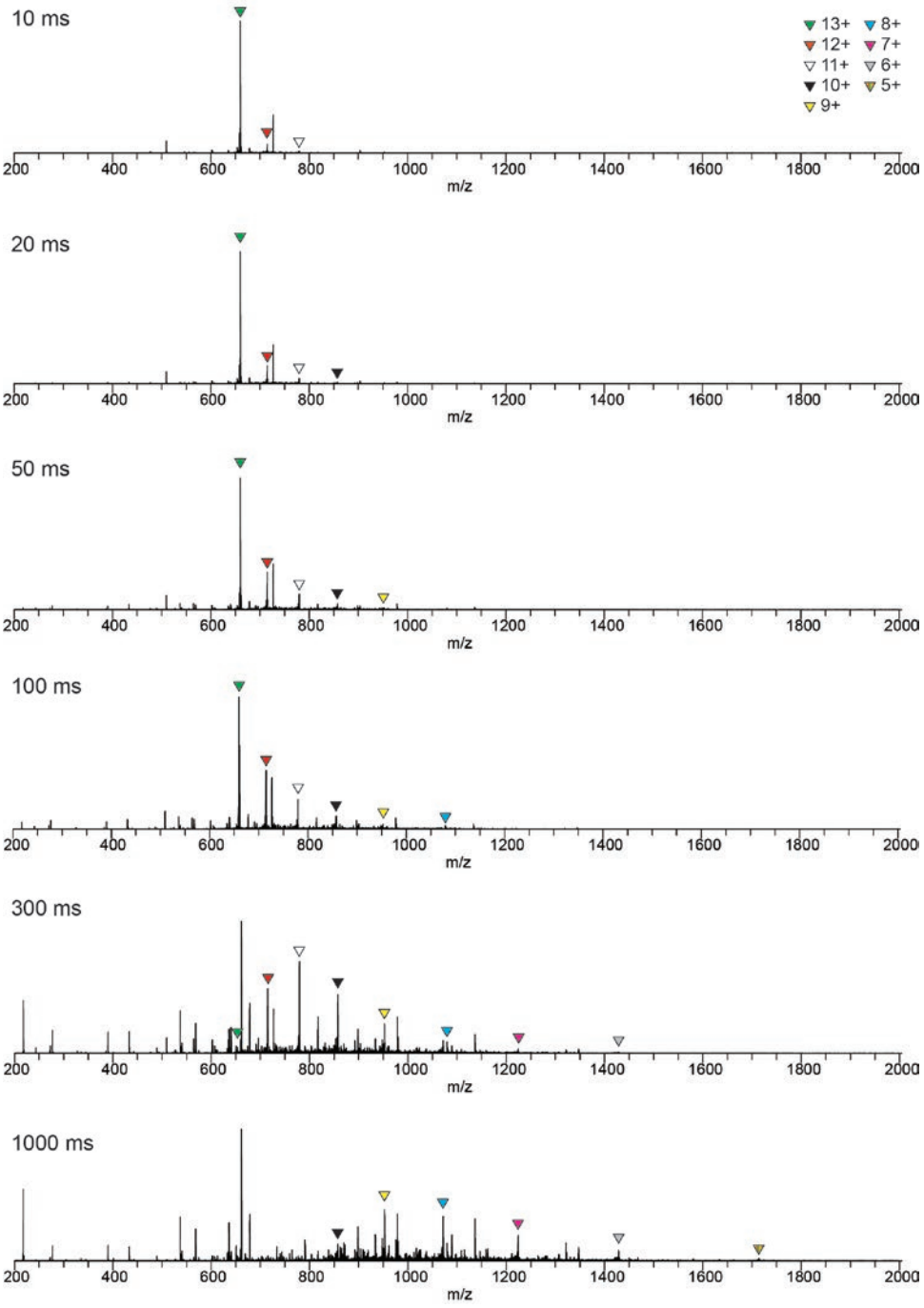



**Figure 2.** Fragmentation of 13+ ubiquitin ions by ETD in the HCD cell using different ETD reaction times. Color coded triangles highlight charge-reduced species of the ubiquitin precursor.

To assess the performance of this mode of operation we analyzed ubiquitin via direct infusion nano-ESI. Ubiquitin is a typical protein for ECD/ETD performance evaluation [1, 15, 27, 42-49]. We subjected the 13+ charge state (m/z 659.8) to ETD fragmentation with Orbitrap detection. At this charge state all basic residues (histidine, lysine and arginine) and the N-terminus are protonated [45, 50]. After a short 10 ms ETD reaction the unreacted precursor is the most abundant species (Figure 2). Increasing the number of anions in the HCD cell through longer injection times results in multiple electrons being transferred to the precursor ions. After a 100 ms ETD reaction the unreacted precursor is still the most abundant ion; however, the charge-reduced precursor is present at multiple charge states (12+ to 8+). We noted that the charge-reduced precursor is present even after very long reaction times. Subjecting the same number of ubiquitin ions to a 1 s ETD reaction we observed the charge-reduced precursor at charge states ranging from 10+ to 5+ (Figure 2). Moreover, the majority of the product ions is detected between 800-1400 m/z. This shift towards higher m/z values indicates that longer ETD reaction times induce secondary (or higher order) electron transfer to product ions as well.

The resolution of the Orbitrap analyzers allows for charge deconvolution of the MS/MS spectra. To assess the differences in amino acid bond cleavages we deconvoluted the MS/MS spectra and analyzed the number of singly charged product ions with increasing ETD reaction time. The number of c/z-ions after a 10 ms ETD reaction in the HCD cell was low; only 21 c- and 33 z-ions were detected (Figure 3A). The number of detected product ions increases with longer ETD reaction times and peaks at 100 ms. This can be explained by the fact that longer reaction times increase the total number of anions in the HCD cell. In turn, this increases the overlap of cations and anions and thus leads to a more efficient electron transfer process. Surprisingly, the number of c/z-ions remains almost constant up to a reaction time of 1 s (Figure 3A). This is likely due to consecutive electron transfer reactions that lead to overall lower charged fragment ions but do not affect the total number of fragment ions [38]. On the contrary, the number of c/z-ions from ion trap-ETD was highest after a 10 ms reaction and quickly decreases with longer reaction times (Figure 3B). This is likely because of an overreaction, inducing undesired consecutive electron transfer to c/z-ions and ultimately to their neutralization. Please note that here we do not take into account the anion injection time for ion trap-ETD. The reported reaction times equal the anion injection for ETD in the HCD cell while anion injection and the actual electron transfer reaction are two separate events in ion trap-ETD mode.

Figure 4A shows the MS/MS spectrum after reaction of 1e6 ubiquitin cations with approximately 1e6 fluoranthene anions (300 ms injection/reaction time) in the HCD cell at a DC gradient of 100 V (Note that the number of anions is estimated based on the anion flux in ion trap-ETD mode).

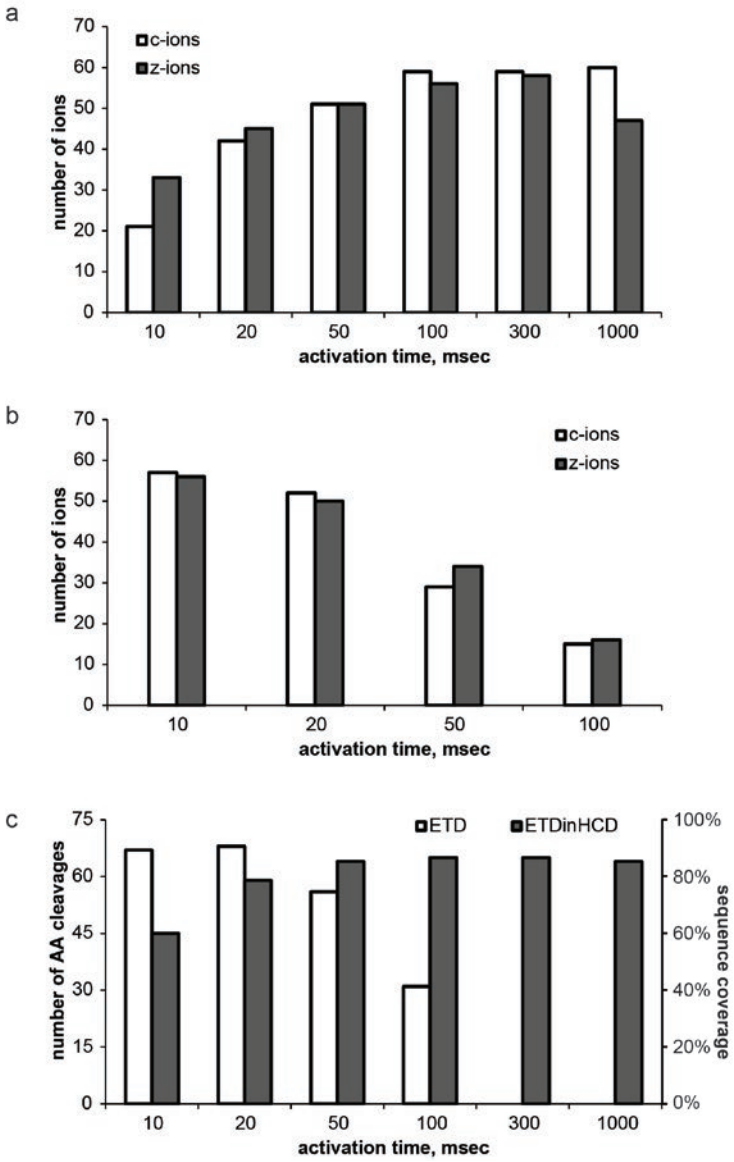



**Figure 3.** Comparison of ion trap-ETD and ETD in the HCD cell with respect to the fragment ions generated from the 13+ ubiquitin precursor. Number of c/z-ions observed at different ETD reaction times for ETD in the HCD cell (A) and ion trap-ETD (B). C) Number of amino acids bond cleavages observed. 100% sequence coverage correlates to the observation of all 72 potentially observable amino acid bond cleavages.

Only a few intense peaks are detected below 600 m/z which correspond to some N-terminal fragments. The main product ion population is observed between 600-1200 m/z. Fragment ion charge states vary from 1+ up to 12+. Among the most intense peaks are the nondissociative, charge-reduced precursor ions. The predominance of charge-reduced precursor ions

is an indicator for low precursor-to-product-conversion efficiency [29]. For comparison we subjected the same number of ubiquitin ions to ion trap-ETD. Figure 4B shows the MS/MS spectrum after a 10 ms ETD reaction. The absence of unreacted or charge-reduced precursor species indicates a high electron transfer and fragmentation efficiency.

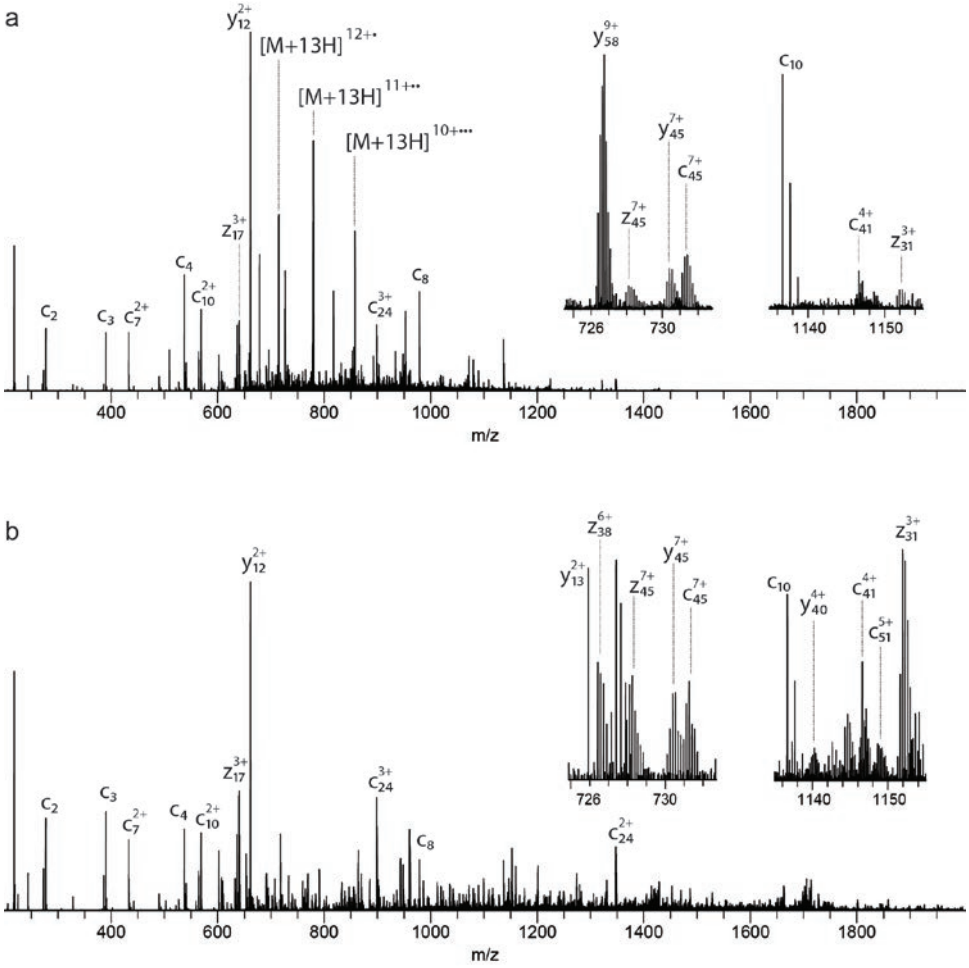



**Figure 4.** MS/MS spectra upon ETD of the 13+ ubiquitin precursor. A) 300 ms ETD reaction in the HCD cell. Many high abundant ions are charge-reduced precursor species. The majority of the product ions is observed between 600-1200 m/z. B) 10 ms ETD in the ion trap (anion AGC 1e7). The intact precursor ion is not observed. Insets are zoom-ins of identical m/z ranges.

Charge-reduced precursor ions were only observed for very short reaction times (<2ms). Note that for proteins larger than ubiquitin absence of charge-reduced precursor ions might also be addressed to insufficient ion transmission. The majority of the product ions is observed between 600-2000 m/z. The insets in figure 4 highlight that the extent of electron

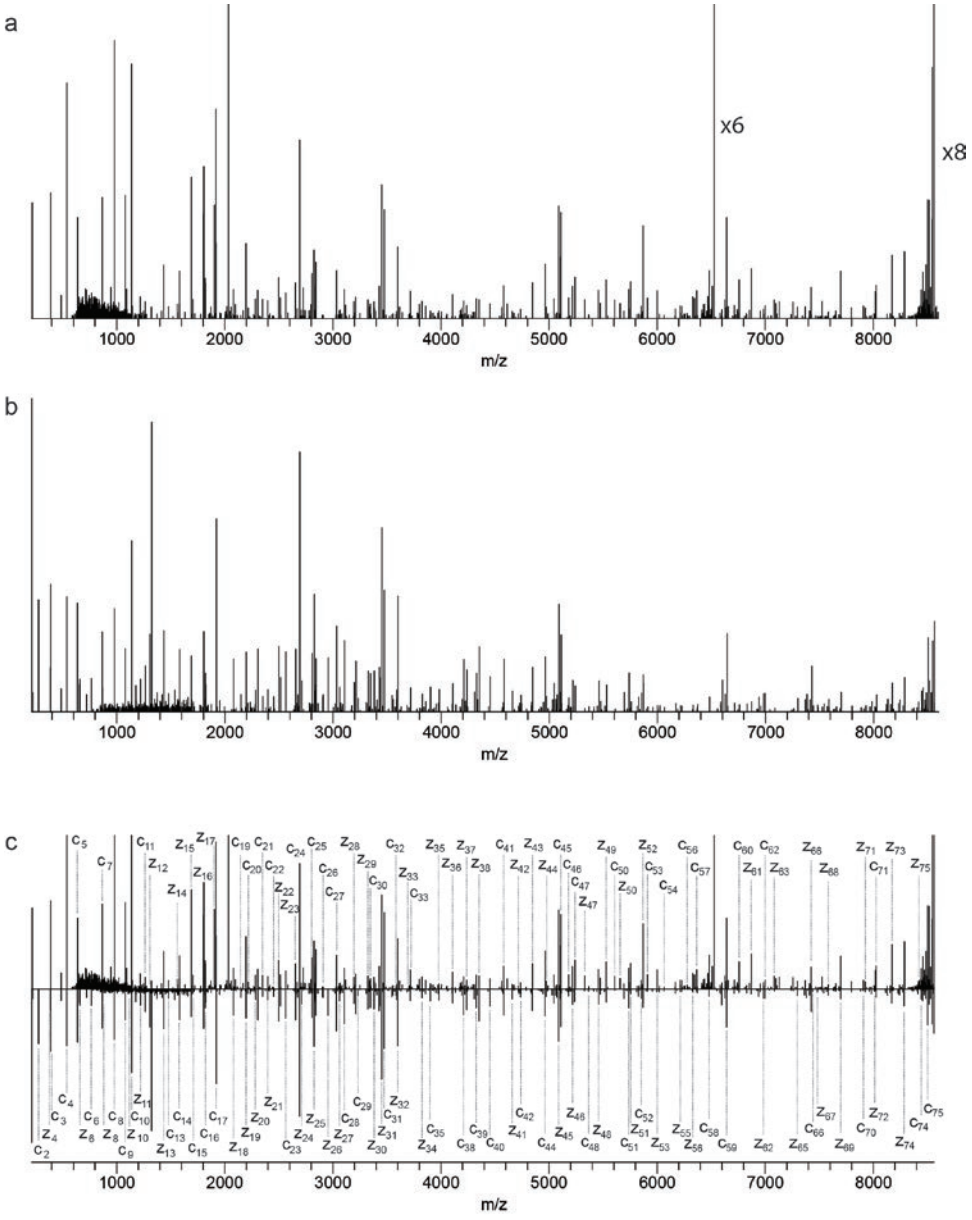



**Figure 5.** Charge deconvoluted ETD-MS/MS spectra of Ubiquitin. A) MS/MS spectrum of ubiquitin after a 300 ms ETD in the HCD cell (see also Figure 4A). The highly abundant precursor ion shows that the ETD reaction efficiency was low. B) Deconvoluted MS/MS spectrum of ubiquitin after a 10 ms ETD reaction in the ion trap (see also Figure 4B). C) Mirror image of ETD in the HCD cell (top panel) and ion trap-ETD (bottom panel) highlights that similar amino acid bond cleavages are obtained using ETD in the ion-trap and the HCD cell.

transfer and fragmentation is different between the two modes of operation. Product ions derived are similar, but not identical between ion trap-ETD and ETD in the HCD cell. Ion trap-ETD produces product ions that are widely spread over the whole detection range. On the contrary, ETD in the HCD cell generates product ions that feature a broader charge state distribution. We calculated the ubiquitin sequence coverage based on the obtained c- and z-ions. Similar to electron capture dissociation cleavage N-terminal to proline is not observed in ETD [51], thus 100% sequence coverage corresponds to observation of 72 out of 75 amino acid bond cleavages due to three proline residues in the sequence. Overall, the sequence coverage follows the trend of the total number of c/z-ions, as expected (Figure 3C). Using ETD in the HCD cell we observed that the sequence coverage reaches an appreciable 89% at a reaction time of 50 ms and remains at this level up to a reaction time of 1 s. For regular ion trap-ETD we found that the ubiquitin sequence coverage is highest at a reaction time of 10-20 ms (94%) but decreases quickly with longer reaction times (Figure 3C). With respect to protein sequence coverage in  a top-down approach ETD in the HCD cell can thus compete with traditional ETD in the ion trap.

Comparing the spectra that generated the highest number of c/z-ions revealed that the precursor-to-product conversion is substantially different between the two modes of operation (Figure 4). However, we found the total number of product ions and the overall sequence coverage at the same level (Figure 3). This indicates that the same level of sequence information is available through both techniques. Figure 5 shows the deconvoluted versions of the MS/MS spectra from figure 4. The main difference is the intensity of the precursor ion, which is the most abundant peak in the spectrum of a 300 ms ETD reaction in the HCD cell (Figure 5A). On the contrary, the precursor in the ion trap-ETD spectrum (Figure 5B, 10 ms reaction time) is present at the same intensity level as the other product ions, illustrating a higher ETD efficiency, mainly because of the effective mixing of anions and cations. However, direct comparison of the two different approaches reveals that the extent of observable c/z-ions is almost identical (Figure 5C).

Next, we evaluated the performance of ETD in the HCD cell in a proteomics LC-MS/MS experiment, targeting smaller peptides rather than proteins, to evaluate the new set-up for applications in bottom-up approaches. We subjected tryptic peptides with charge states ≥3+ to LC-MS/MS analysis performing ETD either in the ion trap or in the HCD collision cell. We varied the AGC target value starting from 5e5 up to 5e6 and assessed the impact on number of identified peptides and the obtained Mascot ion score. Example MS/MS spectra of a quadruply charged peptide composed of 26 amino acids are displayed in figure 6.

Both modes yield similar fragments; however, ion trap-ETD shows a higher extent of electron-transfer. Overall, the data shows that ion trap-ETD outperforms ETD in the HCD cell for tryptic peptides regardless of the precursor AGC target value (Figure 7).

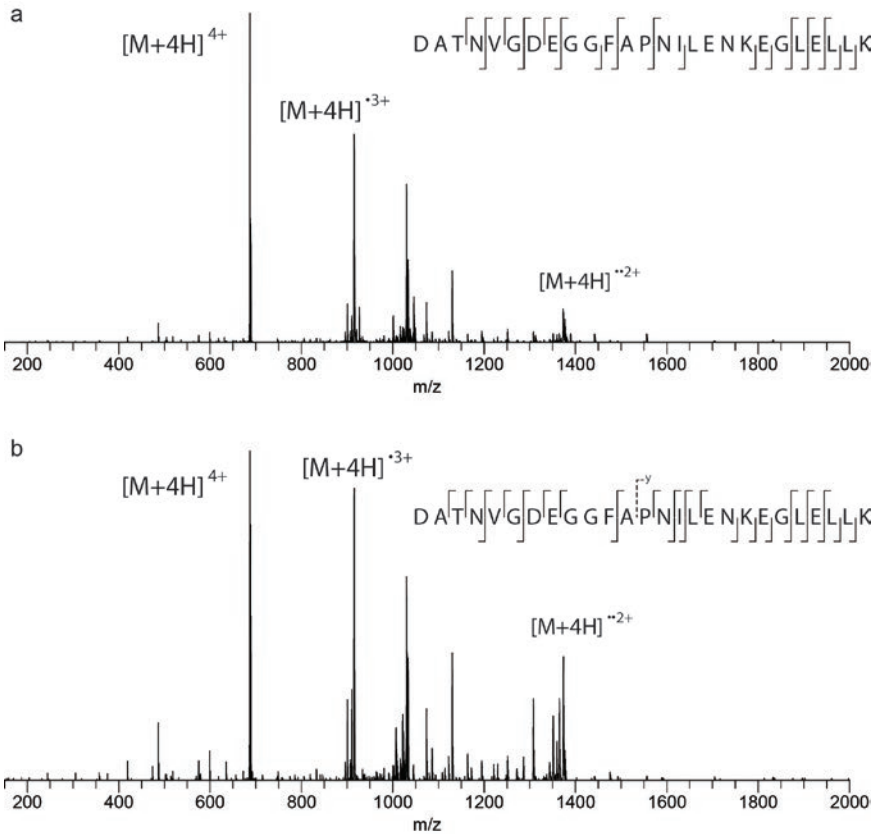



**Figure 6.** ETD MS/MS spectra of the quadruply charged peptide DATNVGDEGGFAPNILENKEGLELLK. A) ETD in HCD cell (150 ms). B) conventional ion trap ETD (25 ms).

The absolute number of peptide-spectrum-matches (PSMs) decreases with higher precursor AGC target values (Figure 7A). This is caused by the precursor injection times that rise for higher AGC values. This decreases the duty cycle and reduces the total number of spectra acquired. Thus, the absolute number of PSMs between the two different methods level up with increasing precursor AGC values. However, the peptide identification success rate, defined as the number of PSMs divided by the number of MS/MS spectra acquired, increases for both approaches ion trap-ETD and ETD in the HCD cell (Figure 7B) but remains superior for the former approach regardless of the precursor AGC settings. The identification rate increases likely because of the reduced duty cycle speed that prevents the instrument from targeting the low abundant peptides for fragmentation. The average Mascot ion score, as a measure for spectral quality, slightly increases with higher precursor AGC values (Figure 7C). This can be explained by a better signal-to-noise on product ions. Overall, the average Mascot ion score was lower using ETD in the HCD cell compared to ion trap-ETD.

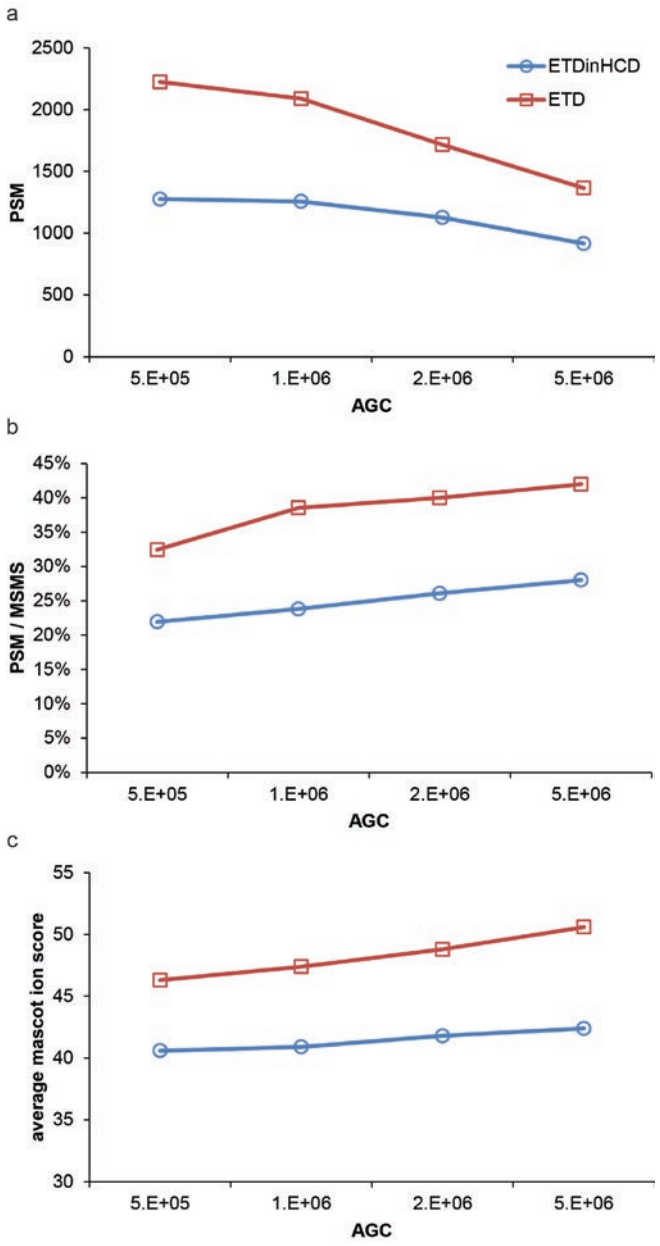



**Figure 7.** Analysis of tryptic peptides with charge states ≥3 following SCX fractionation. Sample was analyzed consecutively at different precursor AGC target values using ion trap ETD and the new approach utilizing ETD in the HCD cell. Spectra were charge deconvoluted and analyzed with Mascot. A) Number of peptide-spectrum-matches. B) Peptide identification success rate. C) Average Mascot ion score.

## CONCLUSION & OUTLOOK

Here we demonstrate that ETD can be performed in the HCD collision cell of an Orbitrap using an additional static DC gradient that enables simultaneous charge sign independent trapping of peptide cations and reagent anions. Ubiquitin was used for evaluating the fragmentation by ETD in the HCD cell, as this protein was used for ETD and ECD on various instruments. Here, we show that the extent of peptide backbone fragmentation obtained and the resulting sequence coverage for ubiquitin is around 90% using ETD in the HCD cell, reaching the same level as when using ETD in a ion trap-ETD experiment on the same Orbitrap instrument. However, ETD in the HCD cell comes at the cost of longer duty cycle times because this mode of operation requires longer ETD reaction (i.e. anion injection) times to achieve sufficient fragmentation. For smaller tryptic peptides we observed that performing ETD in the HCD cell does not reach the level attainable using the ion trap-ETD. Future developments improving the overlap between cations and anions could enhance the fragmentation efficiency. Our preliminary data reveal a relatively low ETD reaction efficiency in the HCD cell, which however could be increased by improved mixing of cations and anions. For example, future experiments could aim on varying the DC voltage that is used to create the axial gradient as the steepness of the gradient determines the overlap between the oppositely charged ions.



## ACKNOWLEDGEMENTS

The Netherlands Proteomics Centre, a program embedded in the Netherlands Genomics Initiative, is kindly acknowledged for financial support. Part of this research was performed within the framework of CTMM, the Center for Translational Molecular Medicine (www.ctmm.nl), project CIRCULATING CELLS (grant 01C-102).

## REFERENCES

[1] R.A. Zubarev, N.L. Kelleher, F.W. McLafferty, Electron capture dissociation of multiply charged protein cations. A nonergodic process, J Am Chem Soc, 120 (1998) 3265-3266.

[2] F.W. McLafferty, D.M. Horn, K. Breuker, Y. Ge, M.A. Lewis, B. Cerda, R.A. Zubarev, B.K. Carpenter, Electron capture dissociation of gaseous multiply charged ions by Fourier-transform ion cyclotron resonance, J Am Soc Mass Spectr, 12 (2001) 245-249.

[3] O.A. Silivra, F. Kjeldsen, I.A. Ivonin, R.A. Zubarev, Electron capture dissociation of polypeptides in a three-dimensional quadrupole ion trap: Implementation and first results, J Am Soc Mass Spectr, 16 (2005) 22-27.

[4] T. Baba, Y. Hashimoto, H. Hasegawa, A. Hirabayashi, I. Waki, Electron capture dissociation in a radio frequency ion trap, Analytical chemistry, 76 (2004) 4263-4266.

[5] H. Satake, H. Hasegawa, A. Hirabayashi, Y. Hashimoto, T. Baba, K. Masuda, Fast multiple electron capture dissociation in a linear radio frequency quadrupole ion trap, Anal Chem, 79 (2007) 8755-8761.

[6] F. Kjeldsen, K.F. Haselmann, B.A. Budnik, F. Jensen, R.A. Zubarev, Dissociative capture of hot (3-13 eV) electrons by polypeptide polycations: an efficient process accompanied by secondary fragmentation, Chem Phys Lett, 356 (2002) 201-206.

[7] J. Axelsson, M. Palmblad, K. Hakansson, P. Hakansson, Electron capture dissociation of substance P using a commercially available Fourier transform ion cyclotron resonance mass spectrometer, Rapid Commun Mass Spectrom, 13 (1999) 474-477.

[8] V.G. Voinov, M.L. Deinzer, J.S. Beckman, D.F. Barofsky, Electron capture, collision-induced, and electron capture-collision induced dissociation in Q-TOF, J Am Soc Mass Spectr, 22 (2011) 607-611.

[9] L. Ding, F.L. Brancia, Electron capture dissociation in a digital ion trap mass spectrometer, Analytical Chemistry, 78 (2006) 1995-2000.

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

[11] A. Chi, C. Huttenhower, L.Y. Geer, J.J. Coon, J.E. Syka, D.L. Bai, J. Shabanowitz, D.J. Burke, O.G. Troyanskaya, D.F. Hunt, Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry, P Natl Acad Sci USA, 104 (2007) 2193-2198.

[12] A. Chi, D.L. Bai, L.Y. Geer, J. Shabanowitz, D.F. Hunt, Analysis of intact proteins on a chromatographic time scale by electron transfer dissociation tandem mass spectrometry, Int J Mass Spectrom, 259 (2007) 197-203.

[13] M.S. Kim, A. Pandey, Electron transfer dissociation mass spectrometry in proteomics, Proteomics, 12 (2012) 530-542.

[14] L.M. Mikesh, B. Ueberheide, A. Chi, J.J. Coon, J.E. Syka, J. Shabanowitz, D.F. Hunt, The utility of ETD mass spectrometry in proteomic analysis, Biochim Biophys Acta, 1764 (2006) 1811-1822.

[15] J.J. Coon, B. Ueberheide, J.E.P. Syka, D.D. Dryhurst, J. Ausio, J. Shabanowitz, D.F. Hunt, Protein identification using sequential ion/ion reactions and tandem mass spectrometry, P Natl Acad Sci USA, 102 (2005) 9463-9468.

[16] C.K. Frese, A.F. Altelaar, M.L. Hennrich, D. Nolting, M. Zeller, J. Griep-Raming, A.J. Heck, S. Mohammed, Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos, Journal of proteome research, 10 (2011) 2377-2388.

[17] R.A. Zubarev, A.R. Zubarev, M.M. Savitski, Electron capture/transfer versus collisionally activated/induced dissociations: Solo or duet?, J Am Soc Mass Spectr, 19 (2008) 753-761.

[18] J.J. Coon, Collisions or Electrons? Protein Sequence Analysis in the 21st Century, Analytical Chemistry, 81 (2009) 3208-3215.

[19] Q.B. Zhang, A. Frolov, N. Tang, R. Hoffmann, T. van de Goor, T.O. Metz, R.D. Smith, Application of electron transfer dissociation mass spectrometry in analyses of non-enzymatically glycated peptides, Rapid Commun Mass Sp, 21 (2007) 661-666.

[20] H. Molina, D.M. Horn, N. Tang, S. Mathivanan, A. Pandey, Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry, P Natl Acad Sci USA, 104 (2007) 2199-2204.

[21] M.K. Bunger, B.J. Cargile, A. Ngunjiri, J.L. Bundy, J.L. Stephenson, Automated proteomics of

E-coli via top-down electron-transfer dissociation mass spectrometry, Analytical Chemistry, 80 (2008) 1459-1467.

[22] N. Siuti, N.L. Kelleher, Decoding protein modifications using top-down mass spectrometry, Nature Methods, 4 (2007) 817-821.

[23] L. Fornelli, E. Damoc, P.M. Thomas, N.L. Kelleher, K. Aizikov, E. Denisov, A. Makarov, Y.O. Tsybin, Top-down analysis of monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS, Molecular & cellular proteomics : MCP, (2012).

[24] Y. Xia, P.A. Chrisman, D.E. Erickson, J. Liu, X. Liang, F.A. Londry, M.J. Yang, S.A. McLuckey, Implementation of ion/ion reactions in a quadrupole/time-of-flight tandem mass spectrometer, Anal Chem, 78 (2006) 4146-4154.

[25] S.J. Pitteri, P.A. Chrisman, J.M. Hogan, S.A. McLuckey, Electron transfer ion/ion reactions in a three-dimensional quadrupole ion trap: reactions of doubly and triply protonated peptides with SO2*, Anal Chem, 77 (2005) 1831-1839.

[26] G.C. McAlister, W.T. Berggren, J. Griep-Raming, S. Horning, A. Makarov, D. Phanstiel, G. Stafford, D.L. Swaney, J.E. Syka, V. Zabrouskov, J.J. Coon, A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer, J Proteome Res, 7 (2008) 3127-3136.

[27] D.A. Kaplan, R. Hartmer, J.P. Speir, C. Stoermer, D. Gumerov, M.L. Easterling, A. Brekenfeld, T. Kim, F. Laukien, M.A. Park, Electron transfer dissociation in the hexapole collision cell of a hybrid quadrupole-hexapole Fourier transform ion cyclotron resonance mass spectrometer, Rapid Commun Mass Spectrom, 22 (2008) 271-278.

[28] D.F. Hunt, J.R. Yates, 3rd, J. Shabanowitz, S. Winston, C.R. Hauer, Protein sequencing by tandem mass spectrometry, Proc Natl Acad Sci U S A, 83 (1986) 6233-6237.

[29] G.C. McAlister, D. Phanstiel, D.M. Good, W.T. Berggren, J.J. Coon, Implementation of electron-transfer dissociation on a hybrid linear ion trap-orbitrap mass spectrometer, Analytical Chemistry, 79 (2007) 3525-3534.

[30] T.Y. Huang, J.F. Emory, R.A.J. O'Hair, S.A. McLuckey, Electron-transfer reagent anion formation via electrospray ionization and collision-induced dissociation, Analytical Chemistry, 78 (2006) 7387-7391.

[31] J. Wu, J.W. Hager, Y. Xia, F.A. Londry, S.A. McLuckey, Positive ion transmission mode ion/ion reactions in a hybrid linear ion trap, Analytical chemistry, 76 (2004) 5006-5015.

[32] X. Liang, J.W. Hager, S.A. McLuckey, Transmission mode ion/ion electron-transfer dissociation in a linear ion trap, Analytical chemistry, 79 (2007) 3363-3370.

[33] Y. Xia, J. Wu, S.A. McLuckey, F.A. Londry, J.W. Hager, Mutual storage mode ion/ion reactions in a hybrid linear ion trap, J Am Soc Mass Spectr, 16 (2005) 71-81.

[34] J.J. Coon, J.E.P. Syka, J.C. Schwartz, J. Shabanowitz, D.F. Hunt, Anion dependence in the partitioning between proton and electron transfer in ion/ion reactions, Int J Mass Spectrom, 236 (2004) 33-42.

[35] J.D. Russell, A.R. Ledvina, G.C. McAlister, M.S. Westphall, J.E. Syka, J. Griep-Raming, J.J. Coon, Single-scan, top-down intact protein analysis on a Velos Orbitrap modified with a dedicated high-capacity ion/ion reaction cell, Proc. 59th ASMS Conf. Mass Spectrometry and Allied Topics, Denver, CO, (2011).

[36] C. Yang, J.D. Russell, A.R. Ledvina, M.S. Westphall, J. Brumbaugh, J.J. Coon, Profiling Histone H3 Isoforms in Human Embryonic Stem Cells using a Velos Orbitrap Modified with a Segmented Reaction Cell, Proc. 58th ASMS Conf. Mass Spectrometry and Allied Topics, Denver, CO, (2011).

[37] C.K. Frese, M. Altelaar, H.W. van den Toorn, D. Nolting, J. Griep-Raming, A.J. Heck, S. Mohammed, Towards Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry, Analytical chemistry, (2012).

[38] D. Nolting, J. Griep-Raming, R. Malek, A. Makarov, Implementation of ETD reactions in a linear quadrupole collision cell utilizing a static DC gradient for dual ion trapping, Proc. 58th ASMS Conf. Mass Spectrometry and Allied Topics, Salt Lake City, UT, 2010, (2010).

[39] R.J. Chalkley, K.F. Medzihradszky, A.J. Lynn, P.R. Baker, A.L. Burlingame, Statistical Analysis of Peptide Electron Transfer Dissociation Fragmentation Mass Spectrometry, Analytical Chemistry, 82 (2010) 579-584.

[40] B.A. Mansoori, E.W. Dyer, C.M. Lock, K. Bateman, R.K. Boyd, B.A. Thomson, Analytical performance of a high-pressure radio frequency-only quadrupole collision cell with an axial field applied by using conical rods, J Am Soc Mass Spectr, 9 (1998) 775-788.

[41] A. Loboda, A. Krutchinsky, O. Loboda, J. McNabb, V. Spicer, W. Ens, K. Standing, Novel Linac II electrode geometry for creating an axial field in a multipole ion guide, Eur J Mass Spectrom, 6 (2000) 531-536.

[42] K. Hakansson, M.J. Chalmers, J.P. Quinn, M.A. McFarland, C.L. Hendrickson, A.G. Marshall, Combined electron capture and infrared multiphoton dissociation for multistage MS/MS in a Fourier transform ion cyclotron resonance mass spectrometer, Analytical chemistry, 75 (2003) 3256-3262.

[43] S.K. Sze, Y. Ge, H. Oh, F.W. McLafferty, Plasma electron capture dissociation for the characterization of large proteins by top down mass spectrometry, Analytical chemistry, 75 (2003) 1599-1603.

[44] R.A. Zubarev, D.M. Horn, E.K. Fridriksson, N.L. Kelleher, N.A. Kruger, M.A. Lewis, B.K. Carpenter, F.W. McLafferty, Electron capture dissociation for structural characterization of multiply charged protein cations, Analytical Chemistry, 72 (2000) 563-573.

[45] K. Breuker, H.B. Oh, D.M. Horn, B.A. Cerda, F.W. McLafferty, Detailed unfolding and folding of gaseous ubiquitin ions characterized by electron capture dissociation, J Am Chem Soc, 124 (2002) 6407-6420.

[46] F.W. McLafferty, E.K. Fridriksson, D.M. Horn, M.A. Lewis, R.A. Zubarev, Biochemistry - Biomolecule mass spectrometry, Science, 284 (1999) 1289-1290.

[47] D.M. Horn, R.A. Zubarev, F.W. McLafferty, Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry, P Natl Acad Sci USA, 97 (2000) 10313-10317.

[48] E.W. Robinson, R.D. Leib, E.R. Williams, The role of conformation on electron capture dissociation of ubiquitin, J Am Soc Mass Spectr, 17 (2006) 1469-1479.

[49] V.A. Mikhailov, H.J. Cooper, Activated Ion Electron Capture Dissociation (AI ECD) of proteins: synchronization of infrared and electron irradiation with ion magnetron motion, J Am Soc Mass Spectr, 20 (2009) 763-771.

[50] P.D. Schnier, D.S. Gross, E.R. Williams, On the Maximum Charge-State and Proton-Transfer Reactivity of Peptide and Protein Ions Formed by Electrospray-Ionization, J Am Soc Mass Spectr, 6 (1995) 1086-1097.

[51] W. Li, C. Song, D.J. Bailey, G.C. Tseng, J.J. Coon, V.H. Wysocki, Statistical analysis of electron transfer dissociation pairwise fragmentation patterns, Analytical chemistry, 83 (2011) 9540-9545.

Chapter 6

# Benchmarking Stable Isotope Labeling Based Quantitative Proteomics

A.F. Maarten Altelaar[1,2,#], Christian K. Frese[1,2,#], Christian Preisinger[1,2], Marco L. Hennrich[1,2], Andree W. Schram[3], H. Th. Marc Timmers[3], Albert J.R. Heck[1,2] and Shabaz Mohammed[1,2]

published in
Journal of Proteomics, 2012 Oct 22



[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

[3] Molecular Cancer Research, UMC Utrecht, 3584 CG Utrecht, The Netherlands

[#] These authors contributed equally to this work

## ABSTRACT

Several quantitative mass spectrometry based technologies have recently evolved to interrogate the complexity, interconnectivity and dynamic nature of proteomes. Currently, the most popular methods use either metabolic or chemical isotope labeling with MS based quantification or chemical labeling using isobaric tags with MS/MS based quantification. Here, we assess the performance of three of the most popular approaches through systematic independent large scale quantitative proteomics experiments, comparing SILAC, dimethyl and TMT labeling strategies. Although all three methods have their strengths and weaknesses, our data indicate that all three can reach a similar depth in number of identified proteins using a classical (MS2 based) shotgun approach. TMT quantification using only MS2 is heavily affected by co-isolation leading to compromised precision and accuracy. This issue may be partly resolved by using an MS3 based acquisition; however, at the cost of a significant reduction in number of proteins quantified. Interestingly, SILAC and chemical labeling with MS based quantification produce almost indistinguishable results, independent of which database search algorithm used.

## INTRODUCTION

Robust and accurate quantification of differential protein expression levels is essential for deciphering the dynamics of proteomes. Mass spectrometry provides excellent means for quantitative proteomics [1] whereby the most common and accurate quantitative approaches utilize stable isotopes [2-4]. In recent years, several strategies have been developed in parallel where isotope incorporation can occur either metabolically in cell culture [3, 5] or whole organisms [6-8] or via chemical labeling at the protein [2] or - more frequently - peptide level [9-13].

As holds true for any analytical method, all labeling approaches have their strengths and weaknesses. It is widely accepted that the error in measurement may be lower if the isotope can be introduced as early as possible into the sample of interest. The earliest time point is at the cellular level that is via metabolic means and currently the most popular labeling approach is through 'stable isotope labeling with amino acids in cell culture' (SILAC) [5]. The most common variant of SILAC involves the use of labeled lysine and arginine amino acids in the growth media [14-16]. This metabolic incorporation of heavy amino acids is very effective resulting in high labeling efficiency where exact levels mostly depend on the purity of the amino acids [17]. However, the technique is more or less restricted to cell cultures and selected model systems such as the mouse [7, 18], as other mammalian/human tissue samples are not easily SILAC compatible. Although recently reports appeared trying

to circumvent this limitation by building a heavy SILAC reference sample as internal standard [19, 20], issues remain with increased experiment time and dilution of low abundant proteins or even absence of tissue specific proteins.

Isotopic labeling after sample lysis offers a valuable alternative as it is applicable to virtually any sample. Several amino acids are amenable for chemical labeling through which stable isotopes can be incorporated [21]. The most successful approaches are highly selective and target the vast majority of the proteolytic peptides. Many chemical labeling strategies target primary amines which are present at the N-terminus of every peptide [22-24], the two most common being based on NHS chemistry [25] and reductive amination [12, 13] (dimethyl labeling). In both SILAC and dimethyl labeling, quantification occurs at the MS level, comparing the ion intensities of peptide pairs or triplets. An alternative peptide labeling approach whereby quantification occurs at the MS/MS level utilizes a set of isobaric chemical labels. Currently, two commercial variants of isobaric tags allow either 2-plex or 6-plex (TMT) [9] and 4-plex or 8-plex (iTRAQ) [10] quantification. The multiplexing capabilities of these isobaric techniques offer great potential in the analysis of larger sample sets such as the analysis of multiple time points, treatments or clinical samples. [26, 27]

Here, we set out to assess and compare the most common quantification strategies using a controlled large-scale shotgun proteomics approach. The quantification labeling methods used are triplex SILAC, triplex dimethyl and TMT (6-plex). These three cover the majority of approaches available in proteomics (metabolic versus chemical labeling and MS versus MS/MS based quantification). Our data indicate that all these three approaches are capable of equivalent depths of analysis; however, quantification by TMT is severely compromised due to the unwanted co-isolation of precursor ions. If the co-isolation issue is resolved by use of an MS3 based approach then the comprehensiveness of the experiment becomes severely compromised. We believe our unbiased, detailed analysis provides an excellent resource for researchers designing quantitative proteomics experiments. All identified peptides and proteins in all the quantitative experiments are reported in the supplemental Tables S2 to S25 and the raw files have been made available through Tranche (see supporting information online).



## EXPERIMENTAL SECTION

**Cell Culture.** HeLa Kyoto cells were cultured at 37°C and 5% CO2 in DMEM (PAA) medium containing 10% dialyzed fetal bovine serum (Invitrogen), 1% L-Glutamine (Lonza) and 1% Penicillin/Streptomycin (Lonza). For SILAC labeling the cells were grown in three conditions using isotopically labeled L-Lysine (500 µM, Sigma) and L-Arginine (170 µM, Sigma) that were added to the medium. "Light label" medium contained "normal" lysine and arginine,

"medium label" medium contained 4,4,5,5-$D_4$ lysine and $^{13}C_6$ arginine, and "heavy label" medium contained $^{13}C_6{}^{15}N_2$ lysine and $^{13}C_6{}^{15}N_4$ arginine. Cells were cultured in SILAC medium for >6 population doublings to reach >98% labeling. For SILAC experiments the three conditions were mixed (1) in a 1:0.5:0.1 and (2) 1:1:1 ratio based on the cell count. Additionally, SILAC light labeled cells were used separately for dimethyl-, TMT- and iTRAQ-labeling. Cells were harvested by trypsinization and washed three times with PBS-buffer.

**Sample Preparation.** Cells were lysed in lysis buffer (8 M urea, 1 tablet Complete Mini EDTA-free protease inhibitor Cocktail (Roche), 1 tablet PhosSTOP Phosphatase Inhibitor Cocktail (Roche), in 10 mL 50 mM triethyl ammonium bicarbonate (Sigma) by sonication. Cell debris was spun down by centrifugation at 20,000 × g for 15 min and the supernatant was transferred into new tubes. The protein concentration was determined by the BCA assay (Pierce). Samples were reduced for 25 min at 56°C (2 mM dithiothreitol, Fluka) and alkylated for 30 min at room temperature in the dark (4 mM iodoacetamide, Sigma). LysC (Wako Chemicals) was added in a 1:75 (w/w) ratio and incubated for 4 h at 37°C. The sample was diluted four times with 50 mM triethyl ammonium bicarbonate and sequencing grade trypsin (Promega) or LysC was added in a 1:100 (w/w) ratio. Incubation was carried out overnight at 37°C.

**Dimethyl-, TMT- and iTRAQ-labeling.** Dimethyl triplex labeling was performed according to the standard protocol [28]. Equal amounts of protein were labeled "on-column" using "light", "medium" and "heavy" dimethyl labeling reagents and the three channels were mixed in a 1:0.5:0.1 ratio. TMT-labeling was carried out using the TMT 6plex labeling kit (Pierce) according to the manufacturer's protocol. In brief, 100 µg of protein per channel were put in new Eppendorf tubes. The TMT labeling reagents were dissolved in 41 µL acetonitrile (Biosolve) per vial and added to the samples. The reaction was incubated for 1h at room temperature and quenched by adding 8 µL of 5% hydroxylamine. iTRAQ labeling was performed according to the manufacturers protocol. Briefly, 70 µL Ethanol were added to each reagent vial, mixed with 100 µg of digested protein and incubated for 1h at room temperature. For the triplex experiment three channels were mixed in a 1:0.5:0.1 ratio. For the sixplex experiment the channels were mixed in a 1:0.5:0.1:1:0.5:0.1 ratio. All SILAC, dimethyl and TMT samples were desalted using C18 solid phase extraction (Sep-Pak Vac C18 cartridge 3 cc/200 mg, Waters), dried in vacuo and reconstituted in 10% formic acid (Merck).

**Strong cation exchange.** Peptides were fractionated using strong-cation exchange (SCX) as described previously [29]. In short, SCX was performed using a Zorbax BioSCX-Series II column (0.8 mm × 50 mm, 3.5 µm). Solvent A consisted of 0.05% formic acid in 20% acetonitrile, solvent B consisted of 0.05% formic acid, 0.5 M NaCl in 20% acetonitrile. The following gradient was used: 0–0.01 min (0–2% B); 0.01–8.01 min (2–3% B); 8.01–14.01

min (3–8% B); 14.01–28 min (8–20% B); 28–38 min (20–40% B); 38–48 min (40–90% B); 48–54 min (90% B); 54–60 min (0% B). A total of 50 fractions were collected per sample, dried in vacuo, reconstituted in 10% formic acid and stored at -20°C.

**Mass spectrometry.** All samples were analyzed on an Orbitrap Velos (Thermo Fisher Scientific, Bremen) that was connected to an Agilent 1200 HPLC system. All columns were packed in-house. Both the trap (20 mm × 100 µm I.D.) and the analytical column (35 cm × 50 µm I.D.) were packed using Reprosil-pur C18 3µm (Dr. Maisch). The sample was delivered to the trap column at 5 µL/min in 100% solvent A (0.1 M acetic acid (Merck). Solvent B consisted of 0.1 M acetic acid in 80% acetonitrile and the gradient flow rate of 0.6 mL/min was passively split and reduced to an effective flow rate of 100 nL/min. The gradient was as follows: 10 min 100% solvent A; 10-25% solvent B within 107 min; 25-50% solvent B within 35 min; 100% solvent B for 2 min; 15 min solvent A. The mass spectrometer was programmed in the data dependent acquisition mode. Two instrument methods were set up. For the MS2 method a FT survey scan from 350 to 1500 m/z (resolution 30, 000; AGC target 5e5) was performed. The 10 most intense peaks were subjected to HCD fragmentation (resolution 7,500; AGC target 3e4, NCE 40% (SILAC, dimethyl) and 45% (TMT), max. injection time 500 ms, dynamic exclusion 60 sec). Predictive AGC was enabled. The MS3 method was programmed as described by Ting et al. [30]. Briefly, the ten most intense ions were subjected to CID fragmentation (CE 35%, AGC 5e3, max. injection time 150 ms). Wideband activation was enabled. The most intense product ion from each MS2 step was selected for HCD-MS3 fragmentation (NCE 60%, AGC 2e5, max. injection time 300 ms). Charge state screening was disabled for the MS3 events.

**Data analysis.** Raw data was analyzed by Proteome Discoverer (version 1.3, Thermo). MS/MS peak lists were created without any pre-processing. Mascot (version 2.3.02, Matrix science) was used to search the MS/MS data against the human Uniprot database (version v2010-12) including a list of common contaminants and concatenated with the reversed versions of all sequences (41,008 sequences). Trypsin or LysC were chosen as cleavage specificity allowing two missed cleavages. Carbamidomethylation (C) was set as a fixed modification. For SILAC labeling oxidation (M), Arginine-6 (R), Arginine-10 (R), Lysine-4 (K), Lysine-8 (K), were used as variable modifications. Additionally, for the SILAC 1:1:1 experiment we used Proline-5 (P) and Proline-6 (P) as variable modifications. For dimethyl labeling oxidation (M), dimethyl (K), dimethyl (N-term), dimethyl-4 (K), dimethyl-4 (N-term), dimethyl-8 (K) and dimethyl-8 (N-term) were set as variable modifications. For TMT labeling oxidation (M), TMT6plex (K) and TMT6plex (N-term) were used as variable modifications. The database searches were performed using a peptide tolerance of 50 ppm and a fragment mass tolerance of 0.05 Da (HCD) and 0.5 Da (CID), respectively. The peptide validator node was set to 'false'. SILAC and dimethyl datasets were quantified using peak area with

the 'precursor ions quantifier' node. Peaks were quantified from MS scans within 2 ppm mass precision between consecutive scans. The 'RT tolerance of isotope pattern multiplets' parameter was set to 0.2 min (SILAC) and 0.5 min (dimethyl), respectively. The 'single-peak quan channels' option was enabled. The TMT datasets were quantified using centroid peak intensity with the 'reporter ions quantifier' node. For all experiments only unique peptides were considered for protein quantification. To achieve a false discovery rate <1% the following filter were applied for all datasets: peptide length 6-35; Mascot ion score >20; maximum search engine rank 1; peptide mass deviation 7 ppm. Leucine and Isoleucine were considered as equal for protein grouping.

MaxQuant (software version 1.1.1.36) data analysis was carried out using the following parameters: Peptide FDR was set to 1%; max peptide PEP was set to 1; protein FDR was set to 1%; minimum peptide length was set to 6; minimum score was set to 60; minimum peptides and minimum razor peptides was set to 1; peptides used for protein quantification was set to razor and unique peptides; the re-quantify option was enabled. The first search was performed within 20 ppm mass tolerance and HCD fragment ion tolerance was set to 20 ppm.

The peptide hydrophobicity was calculated based on the Hopp-Wood scale [31].

## RESULTS & DISCUSSION

The aim of this study was to perform an unbiased global comparison in terms of both identification and quantification of three commonly employed quantitative proteomics strategies. Several crucial parameters had to be considered to perform each of the different labeling strategies under optimal non-biased conditions. One of the great advantages of SILAC is that the heavy amino acid has no discernible differences regarding chemistry (including mass spectrometry) thus little optimization is needed for a quantitative SILAC-based experiment. However, a potential issue that needs to be solved is the arginine to proline conversion, which can be rather significant depending on the cell line used [32]. When this arginine to proline conversion occurs, additional versions of the heavy peptide, containing heavy proline residues, appear which leads to underestimation of the 'original' heavy peptide intensity. Although several means have been introduced to correct for this artifact, experimentally or computationally [32-34], it might still hamper accurate quantification. In our experiments we used a HeLa cell line and culture conditions that have been previously used extensively for SILAC experiments [16], wherein this conversion rate is rather low. We confirmed these previous findings by performing a 1:1:1 SILAC experiment (Table 1) which, indeed, showed less than 1% conversion.

**Table 1.** Summary of identification and quantification at the peptide and protein level for each of the three quantitative proteomics strategies at different mixing ratios. a) Software used for data analysis, PD = Proteome Discoverer, MQ = MaxQuant. b) Number of MS/MS events acquired in each dataset. c) Number of peptide-to-spectrum matches. d) Percentage of the number of PSMs and the number of MS/MS spectra. e) Number of unique identified peptides. f) Number of identified proteins. g) Number of proteins that were quantified in all channels. h) Ratio between the number of quantified and identified proteins. [1)] Automatic Gain Control, AGC = 3e4. [2)] AGC = 6e4.

| labeling | ratio | software[a] | MS/MS[b] | PSM[c] | % identified[d] | peptide ID[e] | protein ID[f] | protein quant[g] | quant/ID[h] |
|---|---|---|---|---|---|---|---|---|---|
| SILAC | 1:1:1 | PD | 215816 | 88915 | 41% | 29483 | 5347 | 4807 | 90% |
| SILAC | 1:0.5:0.1 | MQ | 210964 | 80355 | 38% | 33016 | 5972 | 4507 | 75% |
| SILAC | 1:0.5:0.1 | PD | 210962 | 78496 | 37% | 31940 | 5651 | 3849 | 68% |
| dimethyl | 1:0.5:0.1 | PD | 235366 | 84955 | 36% | 31572 | 5344 | 3894 | 73% |
| TMT 3x | 1:0.5:0.1 | PD | 229269 | 60178 | 26% | 31041 | 5046 | 4858 | 96% |
| TMT 6x[1] | 1:0.5:0.1:1:0.5:0.1 | PD | 231907 | 61911 | 27% | 31142 | 5214 | 4661 | 89% |
| TMT 6x[2] | 1:0.5:0.1:1:0.5:0.1 | PD | 220853 | 61721 | 28% | 31600 | 5172 | 4884 | 94% |

In the case of the two chemical labeling strategies the peptide N-termini and lysine residues are modified (see also chapter 1, Figure 2.2). We first evaluated the tandem MS settings for each labeling strategy since peptide properties such as basicity can be altered by the chemical tags. Specifically, we evaluated the optimal collision energy for peptide fragmentation since there is a well-established correlation between optimal collision energy and basicity [35]. As expected, we found the optimal value for TMT labeled peptides (normalized collision energy 45%) to be slightly higher than that for SILAC and dimethyl (normalized collision energy 40%). We optimized the precursor isolation window with respect to specificity and sensitivity and found optimal values of 1.2 Da (TMT) and 1.5 Da (SILAC, dimethyl), respectively. Finally, the signal to noise ratio of all fragment ions in the tandem mass spectra depend on the lowest mass to charge ratio chosen for detection. This is particularly important in the case of TMT labeling since the quantification is performed using ions at a lower than normal mass to charge ratio. We found that a minimum m/z value of 100 was optimal. Subsequently, these conditions were applied appropriately to all three quantitative approaches



### Effect of labeling and definition of experimental design

The quality of quantification is directly linked to the quality and comprehensiveness of the identification process. Parameters such as the amount of material, the level of separation power applied and the speed and type of sequencing performed were considered. A schematic overview of the chosen workflow is shown in supplemental figure 1. Importantly, to be able to show the effect of the labeling procedure by itself the variability in enzymatic digestion was removed from the initial experiment.

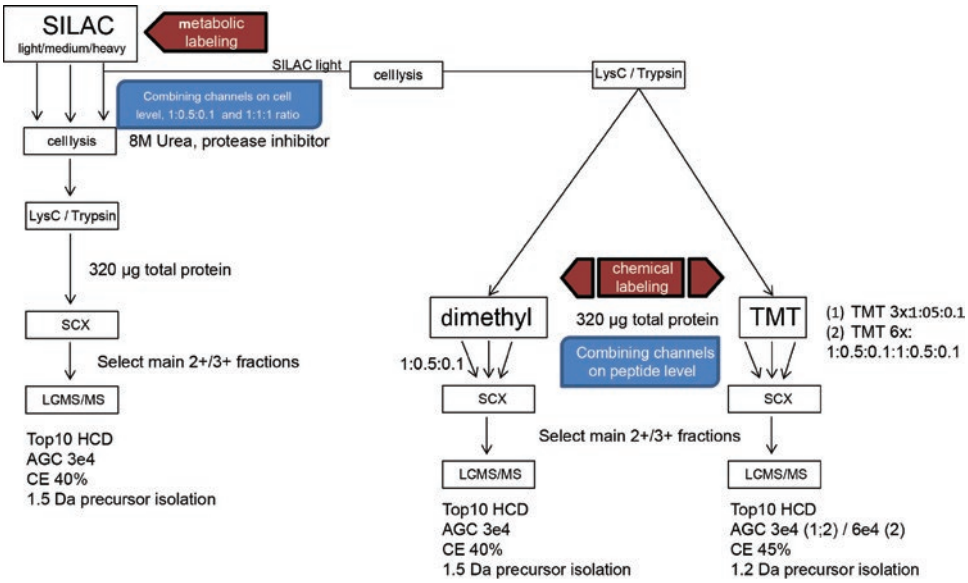



**Figure 1.** Extended workflow of the experiment. HeLa Kyoto cells were cultured in SILAC medium. The three channels were mixed in a 1:1:1 and 1:0.5:0.1 ratio based on cell count. Cells were lysed and digested using LysC and Trypsin. 320 µg of total protein were subjected to strong cation exchange chromatography. The fractions which contain the main 2+/3+ peptides were selected for mass spectrometric analysis on an Orbitrap Velos. For TMT and dimethyl labeling the SILAC light cell culture was used. The main difference is that in SILAC the different labeling channels were combined before sample preparation whereas in TMT/dimethyl the channels were combined after chemical labeling.

For all three quantitative approaches (SILAC, dimethyl and TMT) we prepared triple labeled samples and mixed them in a 1:0.5:0.1 ratio corresponding to a total of 320 µg of material (Figure 1). Furthermore, we performed a SILAC 1:1:1 experiment as described above and an additional 6-plex TMT experiment, using all 6 channels, which we mixed at 1:0.5:0.1:1:0.5:0.1, to also assess the performance under increased complexity. These five different samples were fractionated by strong cation exchange (SCX) followed by the analysis of the equivalent fractions (across all SCX separations) on an Orbitrap Velos with HCD peptide fragmentation. Additionally, we injected (from every SCX fraction) a certain percentage, so as to obtain comparable MS signal. These precautions were taken to minimize experimental deviation between quantification methods caused by MS signal intensity. Finally, the 6-plex TMT experiment was repeated in which we doubled the number of precursor ions used for fragmentation in order to test if quantification is affected by the size of the ion populations. For the identification and quantification of all data we used the same bioinformatics pipeline, based on Mascot for identification and Proteome Discoverer for quantification. We also compared for the SILAC data Proteome Discoverer and MaxQuant.

### Effect of labeling: protein identification

In our experiment we found that the number of sequencing events (spectra) was similar across all employed quantitative proteomics strategies (Table 1). Furthermore, the distribution of peptide signal was also similar across all experiments (Figure 2A). Interestingly, a similar success rate between the number of spectra and peptide to spectrum matches (PSMs) was achieved for SILAC and dimethyl. The 1:1:1 experiment produced a similar success rate for identification when compared to the 1:0.5:0.1 experiment.

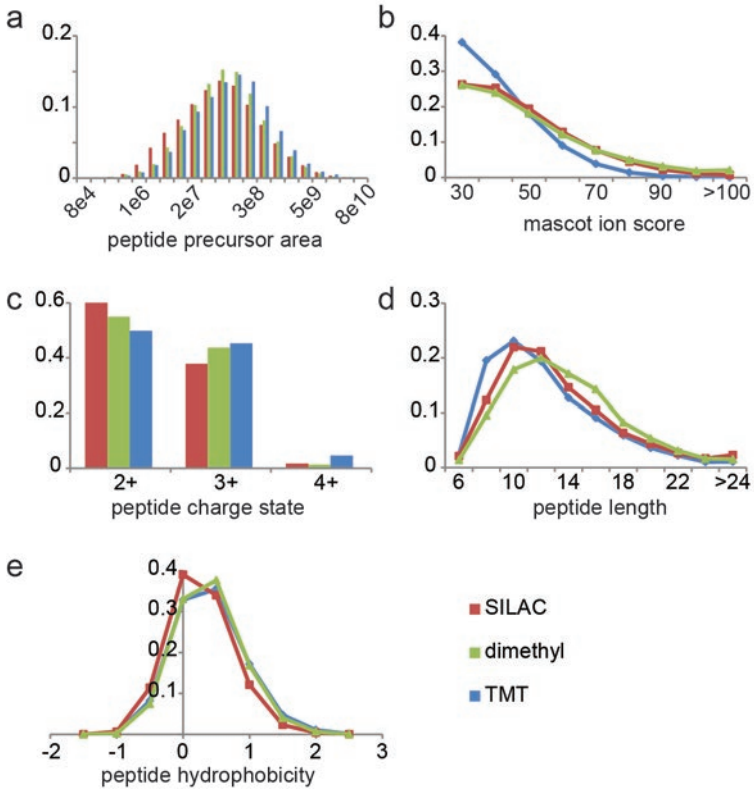



**Figure 2.** Global characteristics of peptide properties identified in the three quantitative proteomics strategies. (a) Histogram illustrating the distribution of peptide ion abundances as reflected by the precursor area for all PSMs. Similar peptide precursor areas were observed in all quantitative proteomics strategies. (b) Normalized Mascot ion score distributions, revealing that TMT provides on average PSMs with somewhat lower confidence whereas dimethylation and SILAC perform equally well. (c) Normalized peptide charge state distribution. Chemical labeling increases the basicity of the peptides and can elevate the charge state of the peptides, especially for TMT. (d) Distribution of the length of the identified peptides. TMT provides on average the shortest and dimethyl the longest peptides. (e) Distribution of peptide hydrophobicity. TMT and dimethyl labeled peptides show a slight increase in the observed hydrophobicity compared to SILAC.

The percentage of spectra leading to identification in the TMT experiments is significantly lower, where only approximately a quarter of all queries are translated into PSMs. A lower success rate is not unexpected since in TMT often optimal conditions for quantification of the low mass reporter ions comes at the expense of sufficient ion intensity of fragment ions needed for identification. The relatively decreased percentage of spectra leading to identification in the TMT experiments has also been noted by Pichler et al. [36]. Comparing Mascot ion score distributions for the three quantitative techniques reveals that TMT performs poorest (Figure 2B). The basicity of the labeling reagents also affected the observed charge state with TMT being the most basic, producing on average a slightly higher charge state envelope for its peptides (Figure 2C). The peptide length distribution showed some small differences between the three techniques, with the TMT peptide distribution being somewhat skewed to lower and dimethyl peptides to higher values (Figure 2D) while the hydrophobicity distribution looked very similar, with, on average, slightly higher values for TMT and dimethyl peptides (Figure 2E).

Switching from total number of PSMs to the number of unique peptides identified produces a different picture. We found that the number of unique peptides identified was quite similar in all three quantitative proteomics strategies. The reduction in complexity of the TMT experiment seems to compensate for the poorer success rate of spectra to PSMs. In the end all experiments identified over 5000 proteins and about 30,000 unique peptides. All three methods showed similar levels of comprehensiveness with the number of proteins commonly identified between all three methods being around 50% and between any of two methods 75-80%  (Figure 3). The observed overlap is in line with that observed between technical replicates, and thus shows that the differently labeled samples are very comparable in number and quality of identifications.



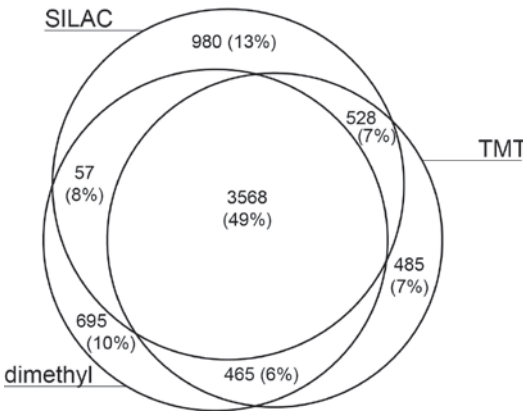



**Figure 3.** Protein overlap. Venn diagrams illustrate the overlap on protein level between the three labeling strategies for the 1:0.5:0.1 samples. All three methods showed similar levels of comprehensiveness with the number of proteins commonly identified between all three methods being 3568, around 50%, and between two methods 75-80% .

**Effect of labeling: protein quantification**

Issues affecting the quality of the studied three quantitative proteomics approaches are the quality/completeness of the labeling and its performance/co-elution in the (multidimensional) LC-MS analysis. The accuracy of the quantification is also related to the quality of mixing. Here, a fundamental difference between SILAC and chemical labeling is the stage in the experiment where the mixing of the samples occurs. Often, in SILAC experiments, mixing is performed at the cellular level, preventing errors being introduced during sample preparation but potentially skewing the 1:1 ratio as it is less accurate to estimate total protein content at the cellular level. The introduced bias is typically corrected by assuming the majority of the protein ratios in any biological experiment is 1:1, allowing normalization. For the SILAC 1:1:1 experiment 90% of all identified proteins were quantified (Table 1). In the case of the 1:0.5:0.1 SILAC mixture, success rate of identification to quantification drops to 68%. The performance of the dimethyl labeling experiment was found to be in the same range, with 73% of the identified proteins quantified. Noticeably, we found that 96% of all proteins identified in the TMT (3-plex) experiment were quantified in all three channels, clearly outperforming the two other quantitative proteomics approaches. As expected, for all three labeling strategies we found differences in the number of quantified proteins, albeit small, for the different ratios, with the number of quantified proteins for the 1:0.5 channel being higher than for the 1:0.1 channel (Figure 4A), caused by the poorer signal-to-noise ratio for the latter (least intense) channel.

The boxplots in Figure 4b, normalized to the expected ratios, illustrate precision of the three quantitative proteomics approaches. The precision for SILAC and dimethyl are very comparable. Surprisingly, TMT produced by far the best precision. For example, for the 1:0.5 ratio the standard deviation was 0.08 for TMT compared to 0.35 and 0.34 for SILAC and dimethyl, respectively (Figure 4b). In the case of the 1:0.1 ratio the spread of the protein ratios increases for all three labeling strategies, reflected in the standard deviation being 0.17 for TMT compared to 0.68 and 0.75 for SILAC and dimethyl. The precision performance of TMT was unexpected as TMT MS/MS based quantification will be more often based on fewer spectra when compared to MS level quantification strategies. The better precision performance of the TMT approach can largely be explained by the near absence of noise in the Orbitrap MS/MS readout. This in combination with the low-mass scanning efficiency of the Orbitrap and the accumulation of low abundant ions results in higher signal to noise ratios and therefore increased quantification efficiency (Figure 5).

The examples in Figure 5 (d to f) correspond to approximately 0.5 ms of accumulation of the desired peptide for the MS level while for the MS/MS level, as used by TMT, the accumulation was close to 30 ms. The TMT experiment accumulated 60-fold more of the

desired peptide while eliminating the chemical noise. Although both SILAC and dimethyl use multiple scans for quantification the signal gains are not comparable to those obtained by TMT quantification.

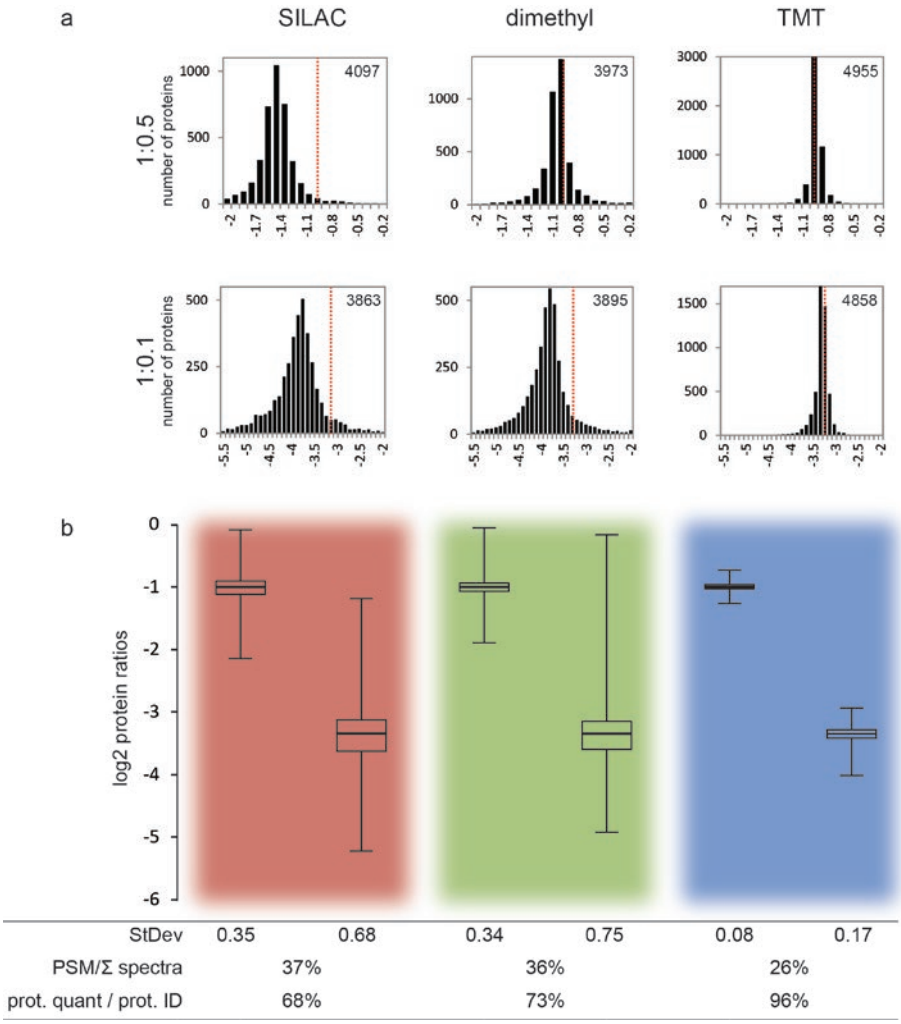



**Figure 4.** Protein quantification data summarized for the 1:0.5:0.1 samples. (a) Histograms illustrate the distribution of the protein log2 ratios. The dashed red lines indicate the exact log2 values for the 1:0.5 and 1:0.1 ratios. (b) Precision of protein quantification. Boxplots indicate the median (middle line), the interquartile range (box), and the 1st and 99th percentile for each population (whiskers). Table underneath the boxplots represents the standard deviations, percentage of identified spectra and percentage of proteins that were quantified.

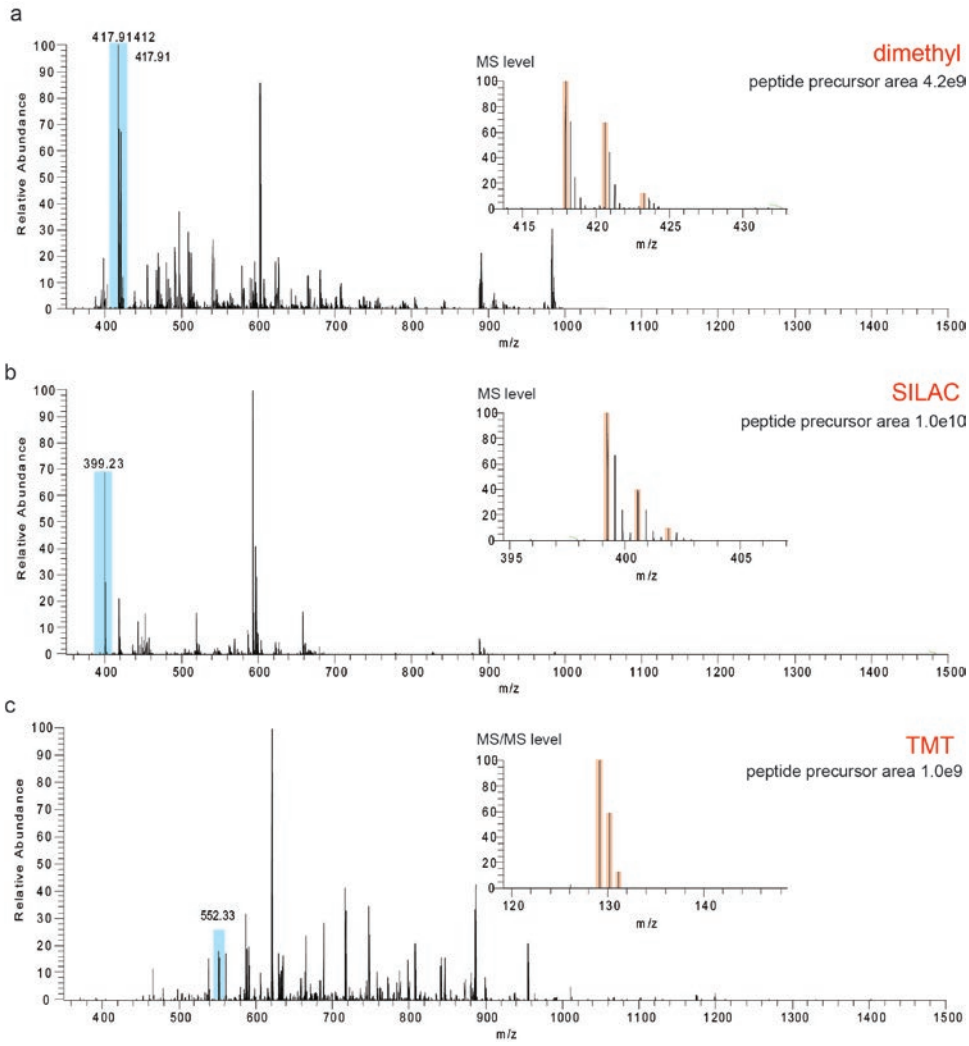
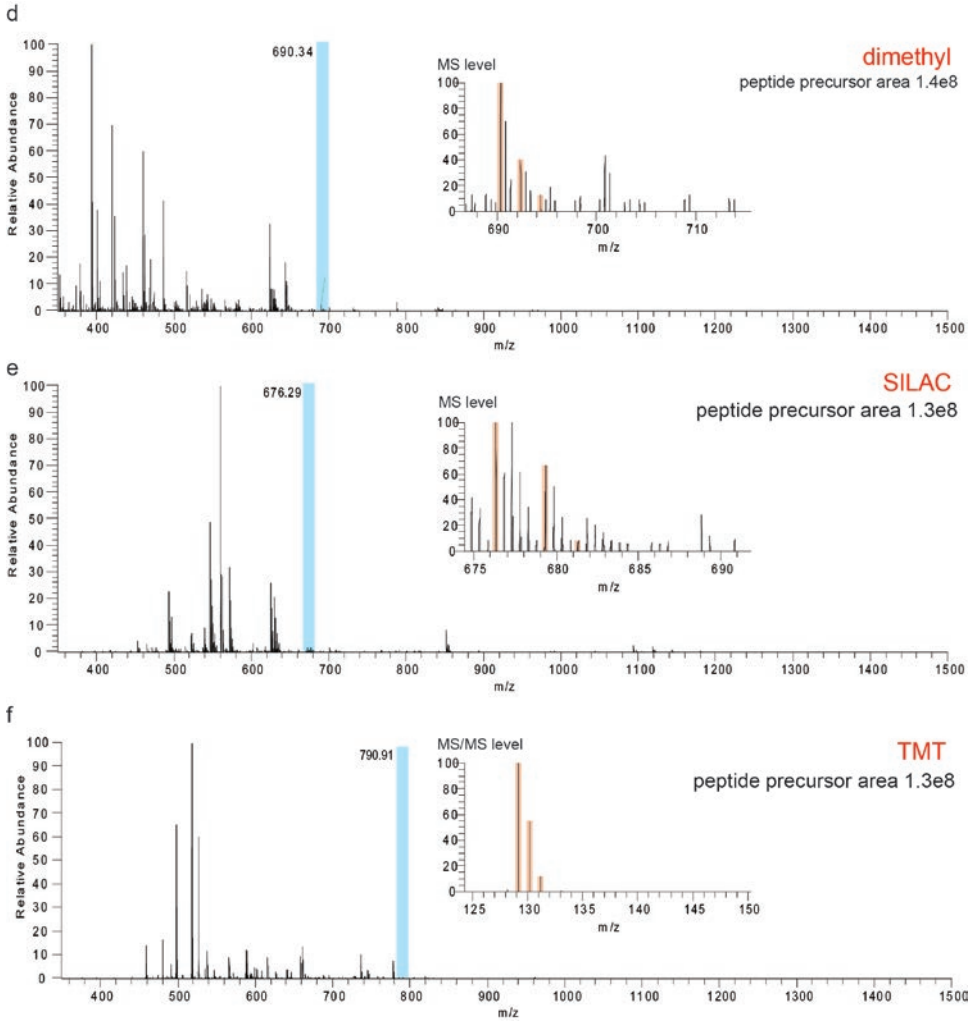



**Figure 5.** Example spectra illustrating the influence of signal-to-noise on peptide quantification. The insets are close-ups of the blue highlighted regions, showing the peptide (SILAC, dimethyl) or reporter ion (TMT) triplets for each labeling strategy. Dimethyl and SILAC quantification is performed on the MS level, whereas TMT quantification is based on reporter ions in the MS/MS spectra. (a-c) The high abundant peptide LTISSPLEAHK shows a good signal-to-noise ratio of the ion triplets for all three labeling strategies. (d, e) The low abundant peptide LCTSVTESEVAR shows a poor signal-to-noise on the MS level (dimethyl, SILAC) which hampers quantification. (e) In TMT the same peptide still shows a good signal-to-noise ratio of the reporter ions on the MS/MS level. (Figure continues on next page)

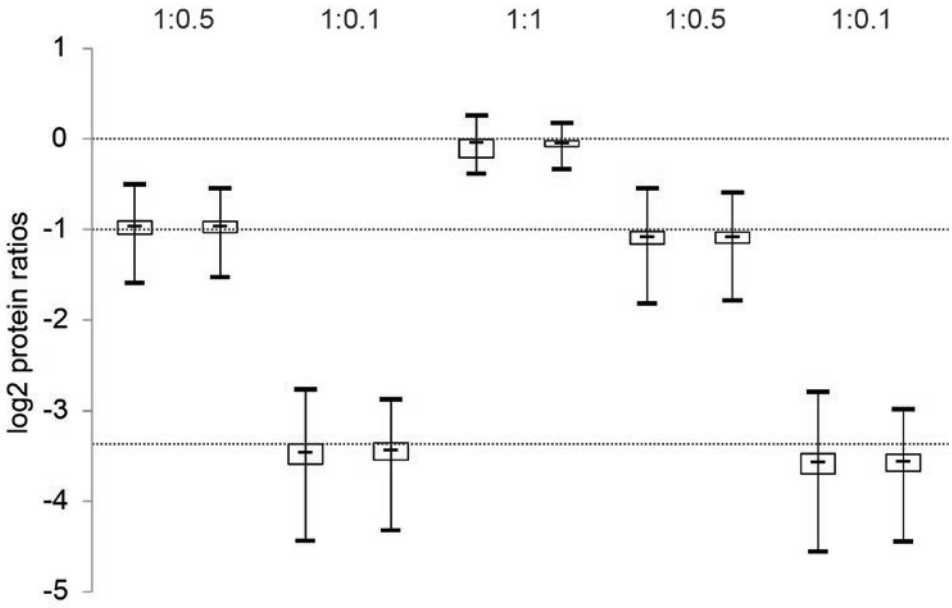





Not surprisingly, for both SILAC and dimethyl we found that the variability of the protein ratios decreases with higher numbers of peptide counts per protein. When we analyzed the 1:0.5:0.1 SILAC dataset by MaxQuant [37] we found that the number of quantified proteins is higher (Table 1). Moreover, while the same level of precision and accuracy in quantification was observed between Proteome Discoverer and MaxQuant for the 1:0.5 ratio, the 1:0.1 ratio showed better accuracy for MaxQuant at the cost of slightly decreased precision.

When designing the TMT experiment we suspected that the total number of ions used for MS/MS would be important. To test this hypothesis we performed a 6-plex experiment alongside the 3-plex experiment, through which the signal per channel would be in theory halved. Moreover, we repeated the 6-plex experiment trapping twice as many ions for the

tandem MS level, creating a 6-plex experiment with the same ion statistics as the 3-plex experiment. As we increased the complexity for TMT from 3plex to 6plex the quantification rate decreases from 96% to 89% when keeping the AGC settings unchanged (Table 1). When we doubled the AGC value the quantification rate increased to 94%. The increased AGC target setting also resulted in a better precision as reflected by the lower standard deviations for all reporter ion channels (Figure 6).




| | 126/131 | | 127/131 | | 128/131 | | 129/131 | | 130/131 | |
|---|---|---|---|---|---|---|---|---|---|---|
| AGC | 3.E+04 | 6.E+04 | 3.E+04 | 6.E+04 | 3.E+04 | 6.E+04 | 3.E+04 | 6.E+04 | 3.E+04 | 6.E+04 |
| proteins | 5094 | 5080 | 4773 | 4942 | 5112 | 5090 | 5082 | 5081 | 4752 | 4928 |
| StDev | 0.18 | 0.16 | 0.28 | 0.22 | 0.10 | 0.08 | 0.20 | 0.16 | 0.32 | 0.24 |

**Figure 6.** Protein quantification of the TMT 6-plex sample for two different AGC settings. Boxplots indicate the median (middle line), the interquartile range (box), and the 1st and 99th percentile (whiskers) for each population. The dashed lines indicate the exact log2 values for the 1:0.5 and 1:0.1 ratios. Doubling the automatic gain control (AGC) leads to a higher quantification rate and better precision, as reflected in the standard deviation for each channel.



### Effect of co-isolation in TMT experiments

For TMT quantification it is a prerequisite that the desired precursor peptide is isolated exclusively, which experimentally cannot always be achieved [30, 38-41]. Co-isolation of more than one precursor peptide can cause the observed ratio to become skewed towards unity. Under optimal conditions, as described above, such an effect would not be observed. To quantitatively assess the amount of precursor ion interference we designed and performed an additional isobaric labeling experiment (Figure 7), in which we took aliquots of the same

HeLa digest and added three quite distinct isobaric tags. One aliquot was labeled with TMT (mass tag is 224.15 Da, reporter channel 131), one with TMT$^{zero}$ containing no heavy isotopes, which possess a different mass (5 Da) to the classic TMT (mass tag is 229.16 Da, reporter channel 126) and a third aliquot with iTRAQ (mass tag is 144.1 Da, reporter channel 114), which differs 80 and 85 Da in mass to the two TMT labels, respectively. These three samples were mixed 1:1:1, allowing random co-isolation of the different isobaric labels to be observed and the amount of interference extracted. Figure 8A shows spectra for three identified iTRAQ labeled peptides, depicting respectively no interference, interference with just one of the TMT labels and interference with both TMT labels.

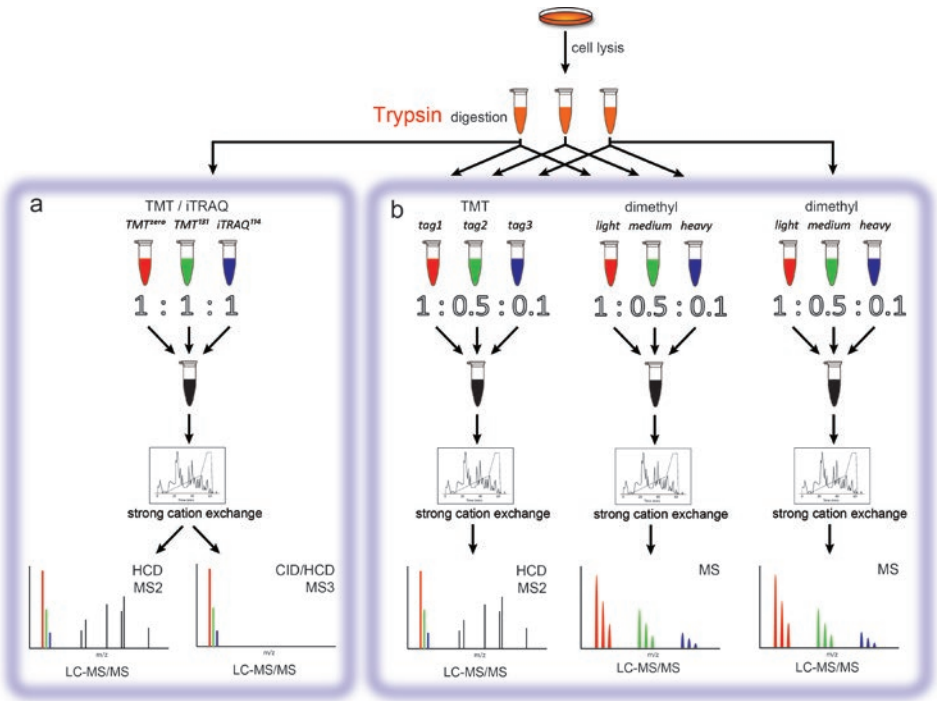



**Figure 7.** Flowchart illustrating the workflow used to evaluate the level of variation caused by the digestion process and to observe the level of co-isolation in reporter ion-based quantification. HeLa cells were lysed and aliquots were digested using either Trypsin or LysC. (A) A single tryptic digest was split into three aliquots and labeled with TMT$^{zero}$, TMT$^{131}$ and iTRAQ$^{114}$, respectively. Samples were combined in a 1:1:1 ratio and separated by strong cation exchange. SCX fractions were analyzed with either MS2-HCD or MS3-CID-HCD. (B) Three separate tryptic digests were split and labeled with either TMT- or dimethyl-labeling, respectively. Additionally we used a single tryptic digest for dimethyl labeling to assess error in precision that is introduced by the separate digests. Channels were mixed in a 1:0.5:0.1 ratio and fractionated by SCX. The main fractions were analyzed by MS2-HCD.

Upon quantifying the level of co-isolation (Figure 8B) it can be clearly seen that the amount of interference is a considerable problem in the standard MS2 approach and a narrow isolation window (1.2 Da) does not resolve this issue. The percentage of completely 'clean' reporter ion spectra is just 28% for the MS2 method while producing a worrisome 52% of spectra showing both interfering reporter ions. The ratio of interfering to expected reporter ions (Figure 8C) shows that in 18% of the cases the interfering reporter ion signal intensity is over 50%, emphasizing the observed compression of isobaric ratios to unity in previously reported studies.

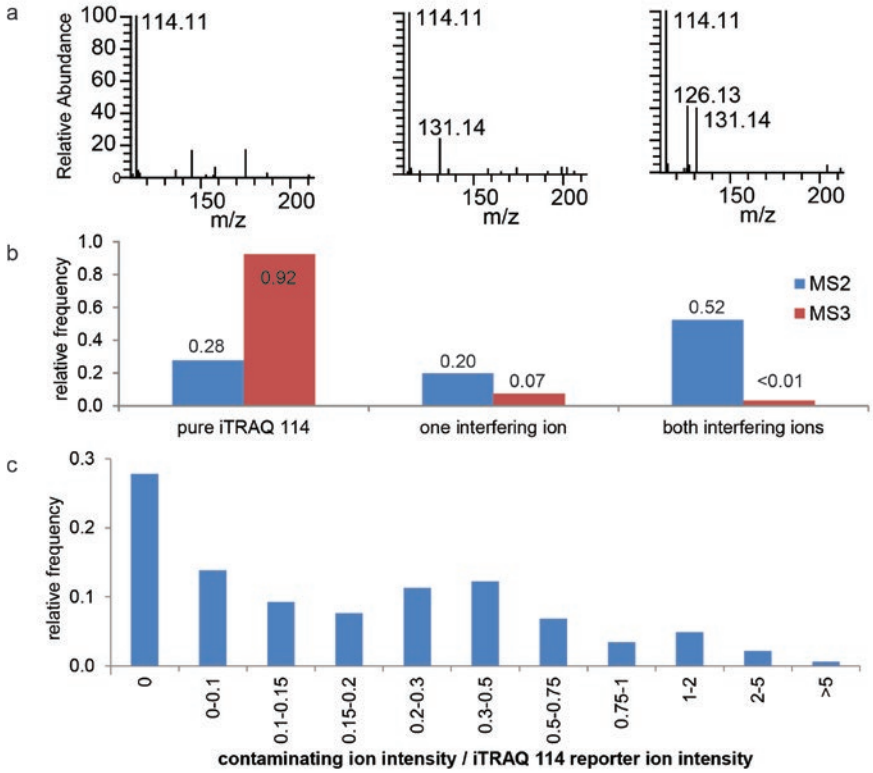



**Figure 8.** Direct evidence of reporter ion interference. Tryptic peptides were prepared and analyzed as described in figure 7. (A) Example spectra of peptides identified with the iTRAQ modification. The spectra from left to right focus on the reporter region illustrating zero, one or two interfering reporter ions next to the iTRAQ 114 reporter ion. (B) Analysis of relative frequency of iTRAQ-labeled peptides with respect to the presence of interfering TMT reporter ions. More than 90% of the peptides that were identified by the MS3 method show no reporter ion. The peptides that were identified by the MS2 method show extensive interfering reporter ion signals. Only 28% of the MS2 method spectra contained no interference while both interfering ions are present in 52% of all MS2 spectra. (C) For each peptide the ratio between the summed intensities of the interfering TMT reporter ions and the iTRAQ reporter ion was calculated. The histogram shows the distribution of the ratios for the MS2 method.

When comparing the level of isolation interference against interfering reporter ions intensity (Figure 9A) little correlation is observed, as reported previously [39]. In contrast, isolation interference had a negative correlation with Mascot ion score (Figure 9B). This co-isolation issue may however be minimized by applying a double isolation event, either of a charge reduced species by using proton transfer reactions [41] or of a dominant fragment ion by applying MS3 [30]. The double isolation leads to removal of the 1:1 background interference and resolves the true TMT ratio, although at the expense of acquisition speed, identification rate and quantification efficiency. Ting et al.[30] first fragmented the precursor by ion trap CID followed by fragmentation of the dominant fragment ion (MS3) by HCD.

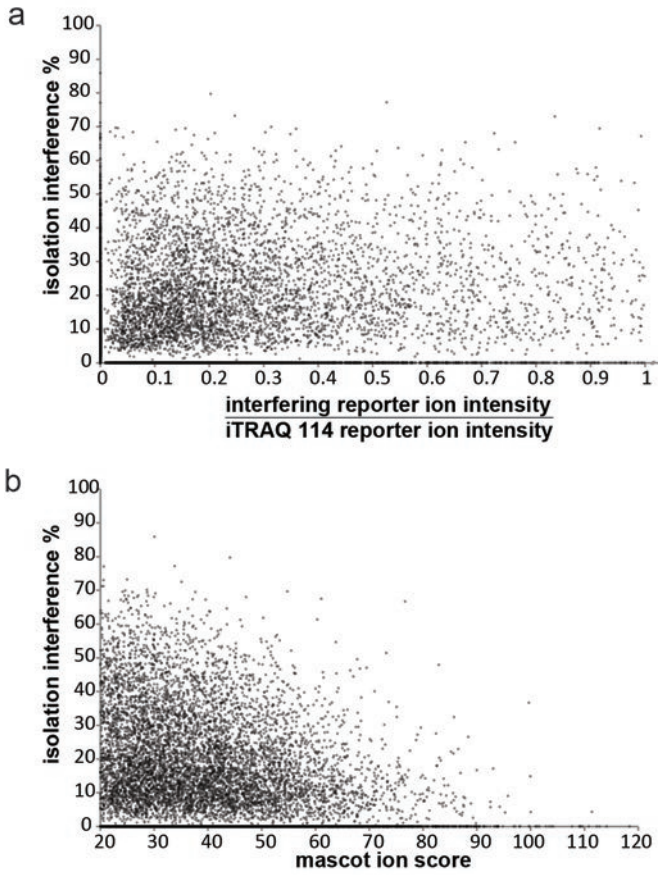



**Figure 9.** Influence of precursor isolation interference on peptide identification and the presence of interfering reporter ions in MS2-HCD. (A) Ratio between the summed intensities of the interfering TMT reporter ions and the iTRAQ reporter ion as a function of precursor isolation interference. No clear correlation was observed. (B) Mascot ion scores as a function of precursor isolation interference. The probability of obtaining higher Mascot ion scores decreases with increasing precursor isolation interference.

The MS2 spectrum is then used for identification while the MS3 precursor is used for quantification. We performed such an experiment on the mixture of three isobaric label sample set. We found that now 92% of the MS3 spectra used for quantification were free of interference (Figure 8B). Thus, one should be very careful with the interpretation of ratios reported for an MS2/TMT based method.

### Effect of digestion

Another important parameter to take into account when benchmarking stable isotope based quantitative labeling is the potential effect of digestion variability on the overall performance, especially when one combines samples at peptide level.

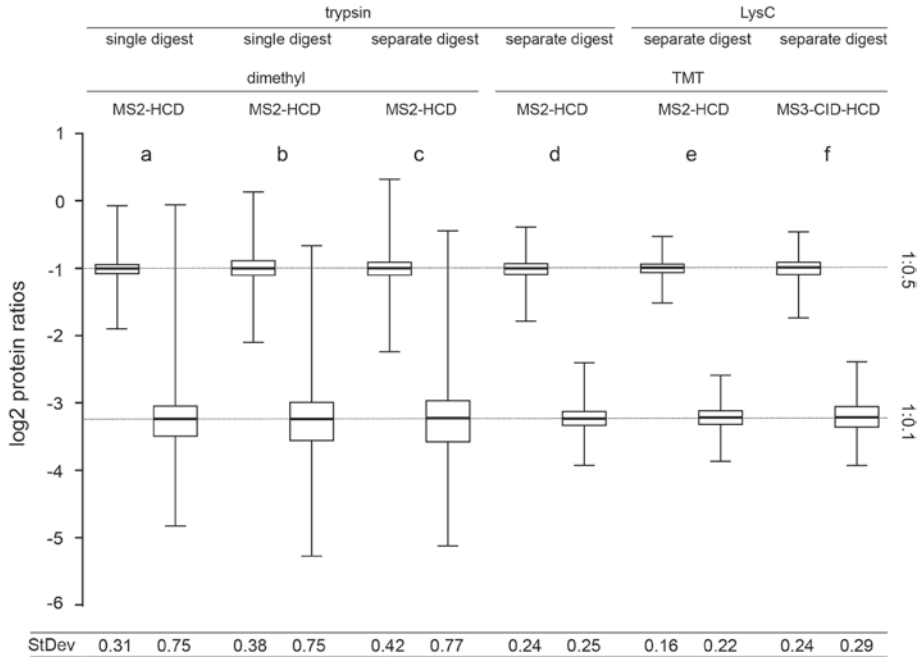



**Figure 10.** Protein quantification to assess level of variation caused by the digestion. Channels were combined in a 1:0.5:0.1 ratio. Boxplots indicate the median (middle line), the interquartile range (box), and the 1st and 99th percentile for each population (whiskers). (A, B) The single digest dimethyl labeling experiment was performed in duplicate to demonstrate repeatability and inter-experiment variability. (C) The dimethyl labeling experiment that was based on three separate digestions shows no distinct change in the standard deviation of the log2 protein ratios compared to the single digest experiment. (D) The TMT experiment based on three separate tryptic digestions shows marked increase in standard deviation of the log2 protein ratios when compared to the single digested experiment. (E, F) For LysC, precision of MS2-HCD is slightly better compared to MS3-CID-HCD which is likely because of a better signal-to-noise ratio and the reporter ion intensity that is derived from two TMT-tags (solely one TMT tag in MS3).

To assess the variability introduced by digestion we performed three experiments, one using dimethyl labeling as described above (to demonstrate repeatability) and two where three separate digests where labeled with either TMT or dimethyl (Figure 7B). Little effect of digestion variability on precision for the dimethyl quantification experiment was observed (Figure 10 A-C), possibly hidden underneath the relatively broad standard deviation. For the more precise TMT experiment we did observe a slight increase in standard deviation caused by variability in digestion performance (Figure 10D). Since TMT quantification based on MS3 spectra requires a LysC digest (to assure the presence of a TMT label on every potential peptide fragment) we performed three independent LysC digests on a HeLa cell lysate and labeled each digest with one of three TMT labels. This LysC TMT sample was then analyzed both by a conventional MS2 HCD experiment and an MS3 CID-HCD experiment as described by Ting et al.[30]. As can be seen in Figure 10E-F the effect of LysC digestion on the precision of quantification is the same as that for trypsin.

### Benchmarking protein quantification considering all effects

Finally, we compared all quantification approaches under their optimal 'realistic' working conditions (Figure 11). To do so we took the original SILAC dataset, as here digestion variation is not an issue, and the dimethyl and MS3 TMT (3X) datasets obtained from three independent digestions. All mixtures possessed a ratio of 1:0.5:0.1. As can be seen from figure 12, MS1 based quantification by either SILAC or dimethyl is still very comparable, with a slightly better precision performance for SILAC. The precision of MS3 based TMT is better than MS1 based quantification. Notably, the MS3 approach is restricted to LysC digestion with CID based identification and requires longer overall acquisition times. These restrictions lead to a severe decrease in number of peptide identifications as well as efficiency of quantification. In fact, the number of quantifiable proteins was reduced to approximately 2/3 of what is possible with either dimethyl or SILAC based quantification. One possible solution to improve the overall success rate of the MS3 method would be to select multiple fragment ions in the second isolation event to increase the total signal intensity available for quantification [42].

## CONCLUSION & OUTLOOK

The increased need for quantitative proteomics over the last decade has resulted in the parallel development of multiple isotopic labeling strategies, largely divided into three main categories; 1) metabolic labeling, 2) chemical labeling with MS based quantification and 3) chemical labeling with MS/MS based quantification. These technologies all have their mer-

its and caveats [5, 43-46]. Chemical labeling strategies have a priori often been assumed to produce the poorest precision and accuracy as it potentially introduces significant error into the experiment through incomplete labeling, poor sensitivity potential side reactions and digestion performance variation. Here, we show for the first time how well dimethyl holds up against SILAC when dealing with the same sample. TMT performance at the MS2 level was similar in terms of peptide identifications as dimethyl and SILAC; however, its ability to quantify was severely compromised by issues with co-isolation.

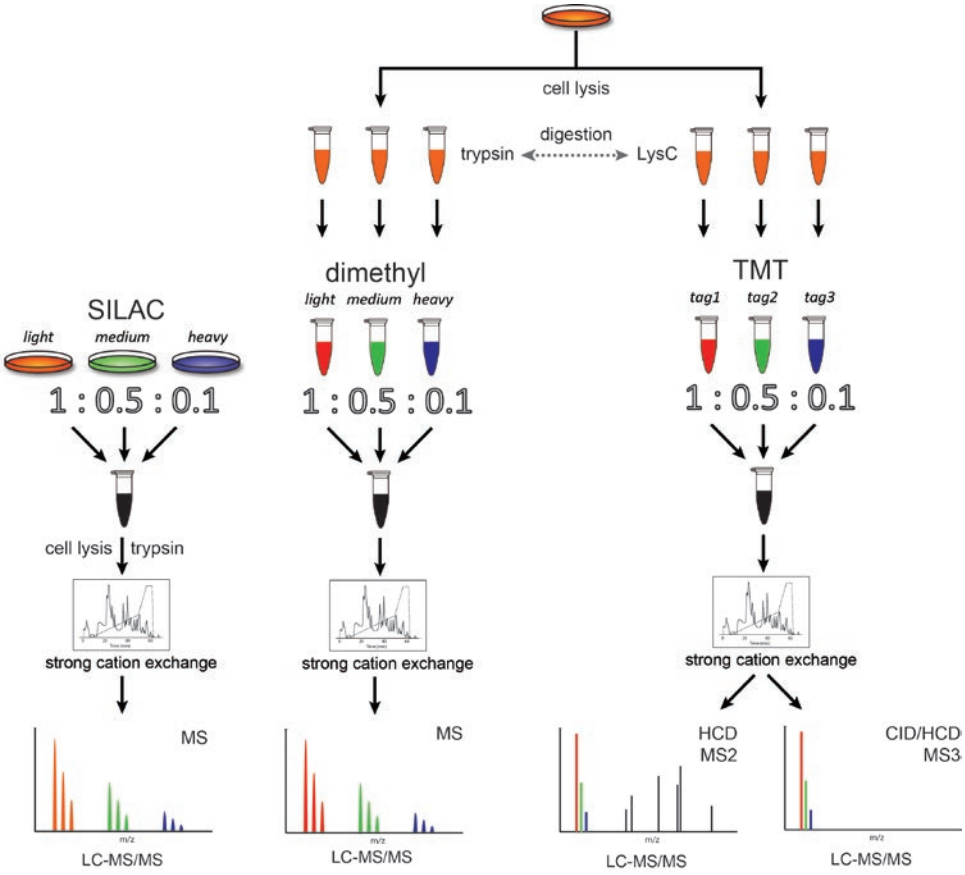



**Figure 11.** Flowchart illustrating the workflow used to benchmark the three quantitative proteomics strategies. HeLa cells were grown in SILAC medium supplemented with light, intermediate or heavy arginine and lysine amino acids. Channels were mixed based on cell count prior cell lysis. For the chemical labeling experiments lysate aliquots were individually digested with either or trypsin or LysC. After labeling and mixing the samples were fractionated by strong cation exchange chromatography (SCX) and analyzed by LC-MS/MS. In the case of LysC digested sample which was labeled with TMT, two acquisition methods were utilized: the standard MS2 HCD method and a MS3-CID-HCD method.

The use of an MS3 strategy resolved the issue but at the expense of a lower MS level duty cycle, poorer efficiency of identification and quantification which led to a less comprehensive experiment than either SILAC or dimethyl. However, MS2/MS3 based TMT did show best precision. Our multiple isobaric labeling strategy also confirms that co-isolation of peptides in a TMT/iTRAQ experiment will severely hinder accuracy of quantification and a double isolation method is required to address this issue [30, 41]. It should be noted that the MS3 method solves the co-isolation issue but then leads to significantly fewer proteins being quantified than either SILAC or dimethyl, both of which do not suffer from co-isolation issues. Nevertheless, Carr and co-workers have shown that the precision of isobaric reagent based quantification can allow compressed peptide ratios to be statistically significantly resolved [39], although this requires an additional MS1 based quantification experiment to determine the compression factor.

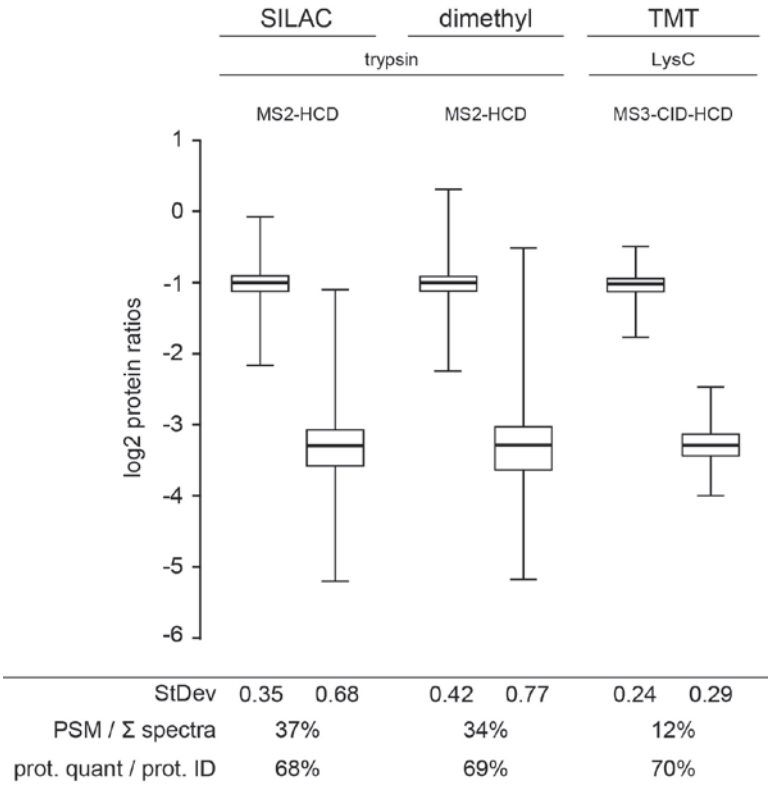



**Figure 12.** Error in Protein quantification for the SILAC dimethyl and TMT experiments under realistic conditions. Table underneath the boxplots represents the standard deviations, percentage of identified spectra and percentage of proteins that were quantified. Data was acquired by MS2-HCD except for the LysC sample, which was acquired by MS3-CID-HCD.

It was surprising that the dimethyl and SILAC strategies perform qualitatively and quantitatively alike under real life conditions. This means that either of the methods can be chosen based on the question at hand. For example, SILAC would be the method of choice for any affinity purification (AP-MS) based quantification experiment, since it allows isotope incorporation at the earliest sample preparation step. Conversely, dimethyl labeling would be preferred when dealing with primary cell cultures of tissue samples in order to prevent the hassle with approaches such as super SILAC where additional LC-MS experiments and multiple SILAC cell lines are needed to mimic the tissue heterogeneity, while still omitting the tissue microenvironment and associated cell-cell communication. Overall our comparative analysis of these three popular quantitative proteomics approaches shows that chemical labeling at the peptide level can be as robust and precise as metabolic labeling at the cellular level when performed under highly controlled conditions. Moreover, although the isobaric quantification strategy leads to a severe improvement in quantification precision, the co-isolation issue has a dramatic impact on observed ratios and is currently not adequately resolved. Therefore, currently the MS1 based methods are preferred for accurate quantification.

## ACKNOWLEDGEMENTS

We kindly acknowledge Michiel Vermeulen and Pascal Jansen for providing the SILAC HeLa cells. This work was in parts supported by the PRIME-XS project, grant agreement number 262067, funded by the European Union 7th Framework Programme; the Netherlands Proteomics Centre, embedded in the Netherlands Genomics Initiative, the Centre for Biomedical Genetics; the Netherlands Organization for Scientific Research (NWO) with a VIDI grant (700.10.429) and the Netherlands Bioinformatics Centre.

## REFERENCES

[1] B.F. Cravatt, G.M. Simon, J.R. Yates Iii, The biological impact of mass-spectrometry-based proteomics, Nature, 450 (2007) 991-1000.

[2] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, Nat. Biotech., 17 (1999) 994-999.

[3] Y. Oda, K. Huang, F.R. Cross, D. Cowburn, B.T. Chait, Accurate quantitation of protein expression and site-specific phosphorylation, Proc. Nat. Acad. Sciences USA, 96 (1999) 6591 -6596.

[4] L. Paša-Toli□, P.K. Jensen, G.A. Anderson, M.S. Lipton, K.K. Peden, S. Martinovi□, N. Toli□, J.E. Bruce, R.D. Smith, High Throughput Proteome-Wide Precision Measurements of Protein Expression Using Mass Spectrometry, J. Am. Chem. Soc., 121 (1999) 7949-7950.

[5] S.-E. Ong, B. Blagoev, I. Kratchmarova, D.B. Kristensen, H. Steen, A. Pandey, M. Mann, Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression

proteomics, Mol. Cell Proteomics, 1 (2002) 376-386.

[6] J. Krijgsveld, R.F. Ketting, T. Mahmoudi, J. Johansen, M. Artal-Sanz, C.P. Verrijzer, R.H.A. Plasterk, A.J.R. Heck, Metabolic labeling of C. elegans and D. melanogaster for quantitative proteomics, Nat. Biotech., 21 (2003) 927-931.

[7] M. Krüger, M. Moser, S. Ussar, I. Thievessen, C.A. Luber, F. Forner, S. Schmidt, S. Zanivan, R. Fässler, M. Mann, SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Function, Cell, 134 (2008) 353-364.

[8] C.C. Wu, M.J. MacCoss, K.E. Howell, D.E. Matthews, J.R. Yates, Metabolic Labeling of Mammalian Organisms with Stable Isotopes for Quantitative Proteomic Analysis, Anal. Chem., 76 (2004) 4951-4959.

[9] A. Thompson, J. Schäfer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, R. Johnstone, A.K.A. Mohammed, C. Hamon, Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS, Anal. Chem., 75 (2003) 1895-1904.

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

[11] X. Yao, A. Freas, J. Ramirez, P.A. Demirev, C. Fenselau, Proteolytic 18O Labeling for Comparative Proteomics: Model Studies with Two Serotypes of Adenovirus, Anal. Chem., 73 (2001) 2836-2842.

[12] J.-L. Hsu, S.-Y. Huang, N.-H. Chow, S.-H. Chen, Stable-isotope dimethyl labeling for quantitative proteomics, Anal. Chem., 75 (2003) 6843-6852.

[13] P.J. Boersema, R. Raijmakers, S. Lemeer, S. Mohammed, A.J.R. Heck, Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics, Nat. Protocols, 4 (2009) 484-494.

[14] N. Dephoure, C. Zhou, J. Villén, S.A. Beausoleil, C.E. Bakalarski, S.J. Elledge, S.P. Gygi, A quantitative atlas of mitotic phosphorylation, Proc. Nat. Acad. Sciences USA, 105 (2008) 10762-10767.

[15] J.V. Olsen, M. Vermeulen, A. Santamaria, C. Kumar, M.L. Miller, L.J. Jensen, F. Gnad, J. Cox, T.S. Jensen, E.A. Nigg, S. Brunak, M. Mann, Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy During Mitosis, Sci. Signal., 3 (2010) ra3.

[16] M. Vermeulen, H.C. Eberl, F. Matarese, H. Marks, S. Denissov, F. Butter, K.K. Lee, J.V. Olsen, A.A. Hyman, H.G. Stunnenberg, M. Mann, Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers, Cell, 142 (2010) 967-980.

[17] S.-E. Ong, M. Mann, Mass spectrometry-based proteomics turns quantitative, Nat. Chem. Biol., 1 (2005) 252-262.

[18] A. Scholten, S. Mohammed, T.Y. Low, S. Zanivan, T.A.B. van Veen, B. Delanghe, A.J.R. Heck, In-depth Quantitative Cardiac Proteomics Combining Electron Transfer Dissociation and the Metalloendopeptidase Lys-N with the SILAC Mouse, Mol. Cell Proteomics, 10 (2011) O111.008474.

[19] T. Geiger, J. Cox, P. Ostasiewicz, J.R. Wisniewski, M. Mann, Super-SILAC mix for quantitative proteomics of human tumor tissue, Nat. Meth., 7 (2010) 383-385.

[20] M. Monetti, N. Nagaraj, K. Sharma, M. Mann, Large-scale phosphosite quantification in tissues by a spike-in SILAC method, Nat. Methods, 8 (2011) 655-658.

[21] S. Julka, F. Regnier, Quantification in Proteomics through Stable Isotope Coding: A Review, J. Prot. Res., 3 (2004) 350-363.

[22] C.L. Nilsson, R. Dillon, A. Devakumar, S.D.H. Shi, M. Greig, J.C. Rogers, B. Krastins, M. Rosen-

blatt, G. Kilmer, M. Major, B.J. Kaboord, D. Sarracino, T. Rezai, A. Prakash, M. Lopez, Y. Ji, W. Priebe, F.F. Lang, H. Colman, C.A. Conrad, Quantitative phosphoproteomic analysis of the STAT3/IL-6/HIF1alpha signaling network: an initial study in GSC11 glioblastoma stem cells, J. Prot. Res., 9 (2010) 430-443.

[23] D. Van Hoof, J. Muñoz, S.R. Braam, M.W.H. Pinkse, R. Linding, A.J.R. Heck, C.L. Mummery, J. Krijgsveld, Phosphorylation dynamics during early differentiation of human embryonic stem cells, Cell Stem Cell, 5 (2009) 214-226.

[24] D. Wijte, B.L.M. van Baar, A.J.R. Heck, A.F.M. Altelaar, Probing the proteome response to toluene exposure in the solvent tolerant Pseudomonas putida S12, J. Prot. Res., 10 (2011) 394-403.

[25] M. Münchbach, M. Quadroni, G. Miotto, P. James, Quantitation and Facilitated de Novo Sequencing of Proteins by Isotopic N-Terminal Labeling of Peptides with a Fragmentation-Directing Moiety, Anal. Chem., 72 (2011) 4047-4057.

[26] M. Bantscheff, D. Eberhard, Y. Abraham, S. Bastuck, M. Boesche, S. Hobson, T. Mathieson, J. Perrin, M. Raida, C. Rau, V. Reader, G. Sweetman, A. Bauer, T. Bouwmeester, C. Hopf, U. Kruse, G. Neubauer, N. Ramsden, J. Rick, B. Kuster, G. Drewes, Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors, Nat. Biotech., 25 (2007) 1035-1044.

[27] D.S. Counotte, N.A. Goriounova, K.W. Li, M. Loos, R.C. van der Schors, D. Schetters, A.N.M. Schoffelmeer, A.B. Smit, H.D. Mansvelder, T. Pattij, S. Spijker, Lasting synaptic changes underlie attention deficits caused by nicotine exposure during adolescence, Nat. Neurosci., 14 (2011) 417-419.

[28] P.J. Boersema, R. Raijmakers, S. Lemeer, S. Mohammed, A.J. Heck, Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics, Nat Protoc, 4 (2009) 484-494.

[29] E.L. de Graaf, A.F.M. Altelaar, B. van Breukelen, S. Mohammed, A.J.R. Heck, Improving SRM assay development: a global comparison between triple quadrupole, ion trap, and higher energy CID peptide fragmentation spectra, J. Prot. Res., 10 (2011) 4334-4341.

[30] L. Ting, R. Rad, S.P. Gygi, W. Haas, MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics, Nature Methods, 8 (2011) 937-940.

[31] T.P. Hopp, K.R. Woods, A computer program for predicting protein antigenic determinants, Mol Immunol, 20 (1983) 483-489.

[32] D. Van Hoof, M.W.H. Pinkse, D.W.-V. Oostwaard, C.L. Mummery, A.J.R. Heck, J. Krijgsveld, An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics, Nat. Meth., 4 (2007) 677-678.

[33] S.C. Bendall, C. Hughes, M.H. Stewart, B. Doble, M. Bhatia, G.A. Lajoie, Prevention of amino acid conversion in SILAC experiments with embryonic stem cells, Mol. Cell Proteomics, 7 (2008) 1587-1597.

[34] S.K. Park, L. Liao, J.Y. Kim, J.R. Yates, 3rd, A computational approach to correct arginine-to-proline conversion in quantitative proteomics, Nat. Meth., 6 (2009) 184-185.

[35] V.H. Wysocki, G. Tsaprailis, L.L. Smith, L.A. Breci, Mobile and localized protons: a framework for understanding peptide dissociation, J. Mass. Spectrom., 35 (2000) 1399-1406.

[36] P. Pichler, T. Köcher, J. Holzmann, M. Mazanek, T. Taus, G. Ammerer, K. Mechtler, Peptide labeling with isobaric tags yields higher identification rates using iTRAQ 4-plex compared to TMT 6-plex and iTRAQ 8-plex on LTQ Orbitrap, Anal. Chem., 82 (2010) 6549-6558.

[37] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nat. Biotech., 26 (2008) 1367-1372.

[38] N.A. Karp, W. Huber, P.G. Sadowski, P.D. Charles, S.V. Hester, K.S. Lilley, Addressing accuracy and precision issues in iTRAQ quantitation, Mol. Cell Proteomics, 9 (2010) 1885-1897.

[39] P. Mertins, N.D. Udeshi, K.R. Clauser, D.R. Mani, J. Patel, S.E. Ong, J.D. Jaffe, S.A. Carr, iTRAQ labeling is superior to mTRAQ for quantitative global proteomics and phosphoproteomics, Mol. Cell Proteomics, (2011) In Press.

[40] S.Y. Ow, M. Salim, J. Noirel, C. Evans, I. Rehman, P.C. Wright, iTRAQ Underestimation in Simple and Complex Mixtures: "The Good, the Bad and the Ugly", J. Prot. Res., 8 (2009) 5347-5355.

[41] C.D. Wenger, M.V. Lee, A.S. Hebert, G.C. McAlister, D.H. Phanstiel, M.S. Westphall, J.J. Coon, Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging, Nature Methods, 8 (2011) 933-935.

[42] G. McAlister, M. Jedrychowski, Y. Yonghao, L. Ting, E.L. Huttlin, R. Rad, W. Haas, S. Gygi, Isolating multiple MS2 fragments using waveforms with multiple frequency notches improves MS3 sensitivity ~8 fold over standard MS3-based TMT methods, in: 60th ASMS Conference on Mass Spectrometry and Allied Topics, Vancouver, Canada, 2012.

[43] M. Bantscheff, M. Schirle, G. Sweetman, J. Rick, B. Kuster, Quantitative mass spectrometry in proteomics: a critical review, Anal. Bioanal. Chem., 389 (2007) 1017-1031

[44] F. Xie, T. Liu, W.-J. Qian, V.A. Petyuk, R.D. Smith, Liquid Chromatography-Mass Spectrometry-based Quantitative Proteomics, J. Biol. Chem., 286 (2011) 25443 -25449.

[45] E.I. Chen, J.R. Yates, Cancer Proteomics by Quantitative Shotgun Proteomics, Mol. Oncol., 1 (2007) 144-159.

[46] J. Cox, M. Mann, Quantitative, high-resolution proteomics for data-driven systems biology, Annu. Rev. Biochem., 80 (2011) 273-299.

Chapter 7

# Profiling of Diet-Induced Neuropeptide Changes in Rat Brain by Quantitative Mass Spectrometry

Christian K. Frese[1,2], Arjen J. Boender[3], Shabaz Mohammed[1,2], Roger A.H. Adan[3], Albert J.R. Heck[1,2] and A.F. Maarten Altelaar[1,2]

published in
Analytical Chemistry, 2013 Apr 26

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

[3] Department of Neuroscience and Pharmacology, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands

## ABSTRACT

Neuropeptides are intercellular signal transmitters that play key roles in modulation of many behavioral and physiological processes. Neuropeptide signaling in several nuclei in the hypothalamus contributes to the control of food intake. Additionally, food intake regulation involves neuropeptide signaling in the reward circuitry in the striatum. Here, we analyze neuropeptides extracted from hypothalamus and striatum from rats in four differentially treated dietary groups including a high-fat/high-sucrose diet, mimicking diet-induced obesity. We employ high-resolution tandem mass spectrometry using HCD and ETD fragmentation for sensitive identification of more than 1700 unique endogenous peptides including virtually all key neuropeptides known to be involved in food intake regulation. Label-free quantification of differential neuropeptide expression revealed comparable up-regulation of orexigenic and anorexigenic neuropeptides in rats that were fed on a high-fat/high-sucrose diet.

## INTRODUCTION

Neuropeptides are a class of naturally occurring endogenous peptides that function as intercellular neuronal signal transmitters mostly through binding of G-protein coupled receptors. They play important roles as key regulators in various physiological processes such as pain, drug addiction, food intake, memory, behavior, circadian rhythm and energy metabolism [1-7]. The alarming increase in feeding- and energy homeostasis-related disorders, such as obesity and diabetes, over the past decades has drawn interest from researchers and the pharmaceutical industry towards a better understanding of the underlying control mechanisms. The basic regulation of food intake is tightly controlled by an interplay of multiple circuits in several nuclei in the hypothalamus [8-10]. Homeostatic regulation of feeding involves a variety of well-studied endocrine peptide hormones [11], such as leptin [12] and ghrelin [13], as well as neuropeptides, such as neuropeptide Y [14, 15], galanin [16, 17], and melanocortin peptides [18, 19]. Moreover, it has been shown that the reward circuitry in the striatum and other brain areas contributes to the regulation of food intake [20, 21]. The rewarding effects of food intake have similarity to those of drug addiction [22]. This is strengthened by the fact that some neuropeptides that cause changes in food intake also play a role in cocaine reward [3].

Neuropeptides are derived from larger precursor proteins by extensive enzymatic processing [23]. The first step in the precursor processing involves the cleavage of the N-terminal signal peptide by a peptidase in the endoplasmic reticulum. The precursor proteins undergo first post-translational modifications (PTMs) throughout transportation into the Golgi. Next,

proteins are stored in large dense core vesicles and transported via the secretory pathway towards the cell membrane. Further processing occurs in the secretory vesicles by prohormone convertases which cleave primarily at dibasic sites in the sequence [24]. Removal of C-terminal basic residues by carboxypeptidases [25] yields the final neuropeptide sequence, which can then still be modified by amidation, further modulating bioactivity and stability. Various neuropeptides can be obtained from a single precursor protein and intermediate cleavage product can function as bioactive neuropeptides as well [18]. The bioactivity is mostly controlled by PTMs such as C-terminal amidation, phosphorylation, sulfation, acetylation and glycosylation. This extensive processing, makes it very difficult to predict the sequence and PTM pattern of bioactive neuropeptides, which raises the need for mass spectrometry (MS) analysis. Knowledge about the exact modification site and sequence of the neuropeptides is crucial for comprehensive understanding of their biological functionality. Over the past years, the MS toolbox has expanded into the field of peptidomics, aiming to characterize bioactive endogenous peptides [26-28]. MS-based analysis of neuropeptides is commonly performed using whole tissue extracts [28-30], extracts of distinct nuclei from tissue sections [31, 32] or via direct profiling from neuronal tissue slices [33, 34] or single cells [35, 36]. The latter approaches give additional spatial information whereas whole brain extracts provide solely tissue-averaged data and makes the interpretation on a functional level difficult. Obtaining spatial information is crucial because the brain consists of many sub-organs with distinct functions.

Challenges in the analysis of neuropeptides remain, especially in sample preparation, MS-based sequencing and data analysis. Firstly, preserving neuropeptides in their native state during sample preparation is essential. Post-mortem degradation or other enzymatic processing altering the peptide sequence has to be prevented [37, 38]. Moreover, unspecific protein degradation products can interfere with neuropeptide assignment. Rapid heating in boiling solvents or by microwave irradiation is usually employed prior to peptide extraction to denature and inactivate enzymes that would otherwise introduce artifacts [38].

Secondly, physicochemical characteristics of neuropeptides differ largely from tryptic peptides, which are routinely analyzed in MS-based proteomics. Variation in peptide length and amino acid composition as well as the occurrence of internal basic residues hamper peptide sequencing by traditionally used MS/MS methods such as resonant excitation-collision induced dissociation. Alternative fragmentation techniques such as electron capture dissociation (ECD) [39] and electron transfer dissociation (ETD) [40] are less influenced by the peptide sequence or PTMs and are nowadays increasingly used for the analysis of neuropeptides [29, 41, 42]. Furthermore, recent instrument developments improving mass accuracy and resolving power facilitate reliable, fast and sensitive neuropeptide identification [32].

Thirdly, peptidomics data analysis is not straightforward. Even if the genome is completely sequenced prediction of the sequence of bioactive neuropeptides remains challenging due to the extensive precursor processing. In traditional proteomics experiments the use of enzymes such as trypsin provides information about one or both of the terminal amino acids. Although neuropeptides are usually derived by cleavage at dibasic residues, removal of the C-terminal amino acids by carboxypeptidases abolishes this useful information. The complex processing procedure is not modeled by conventional database search engines that were initially developed for shotgun proteomics. Thus, MS/MS data needs to be analyzed without enzyme specificity which is more time-consuming and, more importantly, increases the number of potential false positive identifications within the defined peptide mass tolerance. Moreover, it remains challenging to distinguish bioactive neuropeptides from non-bioactive endogenous peptides, purely based on MS-generated sequence information [27].

Fourthly, the unique properties of neuropeptides regarding amino acid composition have to be considered when selecting a quantification strategy. In vivo metabolic labeling via stable isotopes has been demonstrated in rodents [43, 44] but is costly and it remains challenging to obtain complete labeling. Therefore, quantification of endogenous peptides is mostly accomplished by means of chemical labeling following peptide extraction [45, 46]. However, not all endogenous peptides are amenable to chemical labeling. For instance, the peptide hormone Gonadoliberin-1 contains no lysine residue and carries in its active form a N-terminal pyro-Glu modification. The lack of reactive amine groups make this class of endogenous peptides inaccessible to commonly employed chemical labeling approaches such as iTRAQ, TMT, dimethyl labeling or other anhydride reagents. These issues can be overcome by label-free approaches that allow quantification of all peptides irrespective of their amino acid composition. Label-free quantification potentially provides deeper proteome coverage because a peptide needs to be sequenced only in one of the samples in order to be quantifiable across all samples. However, individual sample preparation can increase variability at every step in the workflow [47].

Here, we combine high-resolution tandem mass spectrometry employing HCD and ETD fragmentation together with XIC-based label-free quantification for the analysis of diet-induced neuropeptide expression in rat brain. We analyzed extracts from dissected hypothalamus and striatum and assessed the endogenous peptide levels in four differentially treated dietary groups. We show that this analytical strategy allows sensitive and comprehensive global profiling of endogenous peptides. We identified and quantified virtually all key neuropeptides that are known to be involved in food intake regulation. Furthermore, we sequenced 24 previously unknown potentially novel neuropeptides.

## EXPERIMENTAL SECTION

**Materials.** All chemicals were obtained from Sigma-Aldrich (Steinheim, Germany) unless otherwise stated.

**Animals.** The experiments were approved by the animal ethics committee of the University Medical Center Utrecht, following Dutch legislation. Male Wistar rats (Charles-River, Erkrath, Germany) were used (N = 12, 3 animals per group), ranging in initial weight from 200 to 250 g. The animals were housed individually (378 × 217 × 180 cm) in a controlled environment under a 12:12 light/dark cycle with lights on at 0700 h. Rats were allowed to adapt to their environment for 7 days. Animals were equally divided into four experimental groups that were matched for body weight. Control ('CTRL' group) animals, high-fat high-sucrose ('HFHS' group) fed animals and palatable feeding schedule ('CHOC' group) animals had ad libitum access to water and chow (Special diet services, Essex, UK), while restricted feeding schedule ('RESTR' group) animals had access to chow from 1300 to 1500 h. In addition, HFHS animals had ad libitum access to saturated fat (Vandemoortele, Eeklo, Belgium) and a 30% sucrose solution (Suiker Unie, Hoogkerk, The Netherlands), while CHOC animals received 5 g of chocolate (Droste, Vaassen, the Netherlands) at 1300 h each day. After 8 days of exposure to the diet, all HFHS and two CTRL animals were decapitated, while the CHOC animals, RESTR animals and the remaining CTRL animal were decapitated after 15 days of dietary exposure.

**Sample preparation.** All animals were decapitated between 12:15 pm and 12:45 pm within 10 s after they had been taken from their cages. Immediately after decapitation, the heads were heated using a 800 W microwave (Sharp Co., Bangkok, Thailand) for 5 s after which the brains were removed. Skulls were opened and striatal and hypothalamic brain areas were dissected, snap-frozen in liquid nitrogen and stored at -80°C until further use. Trunk blood was collected in tubes containing EDTA and aprotinin and subsequently centrifuged, after which the blood plasma was stored at -20 °C. Plasma leptin and ghrelin levels were determined in triplicate using a radioimmunoassay (Linco Research, St. Charles, MO, USA).

**Peptide extraction.** Each brain sample was treated individually. Peptides were extracted by a double extraction protocol based on recent reports by Lee et al. [48] and Altelaar et al. [49] 150 µL (striatum) and 300 µL (hypothalamus), respectively, of ice-cold acidified acetone (acetone/water/HCl 40:6:1) were added to the tissue. Samples were lysed by microtip sonication and incubated on ice for 1h followed by centrifugation for 25 min at 14,000 × g. The supernatant was transferred to new tubes and neutralized by adding 1 M NaOH. Acetone was removed in vacuo. The remaining pellet was washed with 150 µL (striatum) and 300 µL (hypothalamus), respectively, of 0.25% acetic acid, incubated on ice for 1h and collected by

centrifugation for 25 min at 14,000 × g. Protein concentration of the pellet was determined using the BCA assay (Thermo Scientific, Rockford, IL). Both extracts were combined and filtered through 10 kDa cutoff membranes (Amicon ultra YM-10, Millipore) for 45 min at 14,000 × g. The flowthrough was desalted by C18 solid phase extraction (Sep-Pak Vac C18 cartridge 1 cc, Waters), lyophilized and reconstituted in 10% formic acid.

**Mass spectrometry.** Peptide extracts from each brain sample were analyzed separately. The running order of the samples was randomized in order to avoid introduction of any bias. Three technical replicates were analyzed from each sample. Peptides were analyzed on an ETD enabled Orbitrap Velos (Thermo Fisher Scientific, Bremen) that was connected to an Agilent 1200 HPLC system, as previously described [47]. Both trap (2 cm × 100 µm inner diameter) and analytical column (40 cm × 50 µm inner diameter) were packed in-house using Reprosil-pur C18 3 µm (Dr. Maisch, Ammerbuch-Entringen, Germany). Solvent A consisted of 0.1M acetic acid (Merck) in deionized water (Milli-Q, Millipore), and solvent B consisted of 0.1M acetic acid in 80% acetonitrile (Biosolve). The flow rate of 5 µL/min was passively split and reduced to an effective flow rate of 100 nL/min. The LC gradient was as follows: 10 min solvent A; 10-25% solvent B within 65 min; 25-50% solvent B within 25 min; 100% solvent B for 3 min; 15 min 100% solvent A. The autosampler was programmed to inject 20 fmol of a tryptic digest of bovine serum albumin prior to sample injection. The mass spectrometer was programmed in the data dependent acquisition mode. Survey MS scans were acquired at a resolution of 30,000 at m/z 400. The four most intense precursor ions were selected within an isolation window of 1.5 Da for consecutive fragmentation by HCD and ETD. The precursor AGC target was set to 1e5 which is a compromise between the optimal settings for HCD and ETD [50]. The normalized collision energy for HCD was set to 40%. The activation time for ETD was set to 50 ms for doubly charged peptides. Supplemental activation and charge state dependent activation time were enabled. The AGC target setting for fluoranthene anions was set to 3e5. All MS/MS spectra were acquired in the Orbitrap at a resolution of 7,500 at m/z 400.

**Peptide identification.** MS/MS spectra were processed as described previously [50]. Briefly, ETD spectra were simplified by removing the precursor, charged reduced precursor and neutral losses from the latter species. MS/MS spectra were deisotoped and charge deconvoluted. The IPI_rat database (version 3.68) was filtered and solely proteins that contain a signal peptide (based on signalP 4.0 [51]) were retained. MS/MS spectra were searched without enzyme specificity against this modified database concatenated with a list of common contaminants using Mascot (version 2.3, Matrix Science, UK). The mass tolerances for the precursor and fragment ions were set to 10 ppm and 0.02 Da, respectively. Acetylation (N-term), amidation (C-term), deamidation (NQ), glutamine to pyroglutamyl (N-term Q), oxidation (M) and phosphorylation (ST) were allowed as variable modifications. The decoy

search option was enabled. Mascot search results were filtered for <1% false discovery rate at peptide level using the percolator algorithm within Rockerbox [52]. Additionally, only peptide-spectrum-matches (PSMs) with a mascot ion score >20 were retained. The Neuropred algorithm [53] was used to identify potentially novel neuropeptides using default settings and the mammalian, rat and known-motif models, respectively. The Swepep database [54] was downloaded from http://www.swepep.org. The neuropedia database (version 063011) was obtained from http://proteomics.ucsd.edu/Software/NeuroPedia/index.html.

**Label-free quantification.** Label-free quantification was performed using Progenesis LC-MS (version 4.0, Nonlinear Dynamics, UK). Data from hypothalamus and striatum were analyzed separately. All raw files from one brain area were subjected to automatic LC retention time alignment. Alignment was manually validated using a set of 35 tryptic BSA peptides that had been spiked into each sample prior LC-MS/MS analysis as references. Features that were not subjected to MS/MS fragmentation in at least one of the raw files were removed from the analysis. Peptide abundance based on precursor area of the isotopic envelope was calculated by Progenesis LC-MS. Tryptic BSA peptides were used as internal standard to correct the intensities for technical variability introduced at the LC-MS level. Peptide abundances from technical replicates were averaged. The abundances of peptides which were identified from multiple charge states were summarized. All quantitative data are given as mean ± SEM.

## RESULTS & DISCUSSION

In this work we analyzed neuropeptides from two different brain compartments from rats in four differentially treated dietary groups. Rats were fed on a regular diet (CTRL), high-fat/high-sucrose diet (HFHS) and restricted chow diet (RESTR), respectively. Another group of rats was also fed on the regular diet but was additionally given a piece of chocolate every day (CHOC). Similar diets have previously been used to study food intake-related disorders [55, 56]. Animals fed on a restricted food diet show a substantial loss in body weight after two weeks. No significant difference in relative body weight changes is observed between the other groups, which is expected because of short timescale of the experiment. Hypothalamus and striatum were dissected immediately after decapitation, followed by rapid heating of the brain to prevent post-mortem degradation. The hypothalamus was chosen because of its well-known role in food intake regulation. On the other hand, reward-related neuropeptide signaling mainly occurs in the striatum. Endogenous peptides from the two brain areas were separately extracted by a two-step extraction procedure using acidified acetone and acetic acid. Equal amounts of the peptide extracts were analyzed in triplicate by LC-MS/MS-based label-free quantification.

### Qualitative analysis of neuropeptides from rat hypothalamus and striatum

Neuropeptides can be extensively modified by PTMs. We have previously shown that the use of the two fundamentally different but complementary fragmentation techniques CID and ETD substantially aids in neuropeptide identification [29]. Here, we employed ETD and HCD fragmentation in combination with Orbitrap detection to maximize the peptide identification rate. Figure 1 shows MS/MS spectra of the 7+ ion of the 36 amino acid-long mature neuropeptide Y upon activation with ETD and HCD, respectively. Dissociation using ETD generated extensive peptide backbone fragmentation that led to identification with very high confidence (Figure 1A, sequence coverage 74%; Mascot ion score 95). On the contrary, HCD fragmentation of the same precursor yielded fewer sequence-informative fragment ions (Figure 1B, sequence coverage 46%; Mascot ion score 28). Our peptide identification strategy involved charge deconvolution and deisotoping of the MS/MS spectra, enabling identification of highly charged endogenous peptides such as neuropeptide Y. All data was searched with Mascot without enzyme specificity against the IPI_rat database that was filtered to retain only proteins that contain a signal peptide sequence. All PSMs were filtered at a FDR-level <1%. In total we identified 53859 and 11972 PSMs in rat hypothalamus and striatum, respectively (Table 1). Across both datasets 56.7% of all peptides were successfully identified by both ETD and HCD (Figure 2A). Of all peptides 24.3% and 19.0%, respectively, were only identified by either ETD or HCD. We observed that the majority of all PSMs identified by ETD were derived from precursors with charge states ≥3+. On the contrary, 53% of all PSMs identified by HCD were doubly charged. Our data highlights the advantage of employing multiple fragmentation techniques for increasing peptide identification rates. Subjecting each precursor to both ETD and HCD fragmentation increases the probability to obtain a high-quality MS/MS spectrum. This is in line with previous studies that have shown that ETD is beneficial for higher charged peptides [50, 57].

**Table 1.** Endogenous peptide identification from rat brain.

| | hypothalamus | striatum |
|---|---|---|
| peptide-spectrum-matches | 53859 | 11972 |
| unique peptides (excluding PTMs) | 1513 (1256) | 451 (378) |
| peptides from known NP precursor proteins | 78% | 73% |
| | hypothalamus + striatum | |
| unique peptides (excluding PTMs) | 1743 (1372) | |
| - swepep | 63 (31) | |
| - neuropedia | 54 (31) | |
| - neuropred | 167 (102) | |

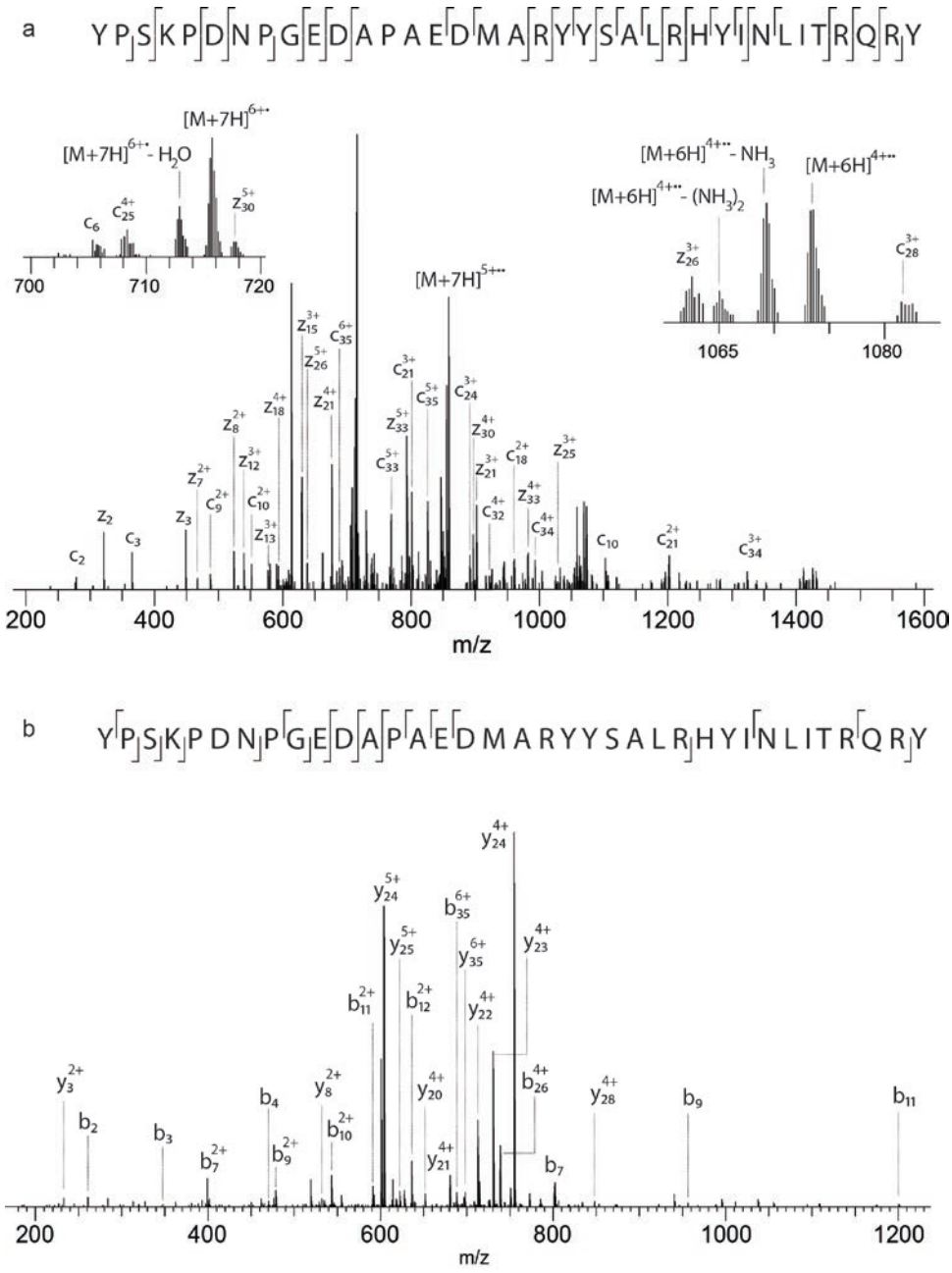



**Figure 1.** ETD and HCD MS/MS spectra of neuropeptide Y. A) ETD spectrum of the 7+ precursor of NPY. The resulting Mascot ion score was 95. Extensive peptide backbone fragmentation yields a total sequence coverage of 74%. Insets highlight zoom-ins of the corresponding m/z regions. B) Substantially less fragment ions are observed upon HCD fragmentation of the same precursor ion (Mascot score 28, sequence coverage 46%).

The total number of PSMs translates into 1513 (hypothalamus) and 451 (striatum) unique endogenous peptides when allowing for PTMs to be included. We attribute the higher number of peptides in the hypothalamus dataset to the fact that the hypothalamic brain area is larger and more heterogeneous than the striatal brain area. Consequently, the total peptide yield in the extracts is higher in the hypothalamus. Combining the hypothalamus and striatum datasets we identified in total 1743 unique endogenous peptides at 1% FDR level (Table 1). This corresponds to 1369 unique peptide sequences, omitting PTMs. A complete list of all identified peptides is provided in the supporting information (Supporting Table 1, available online). To our knowledge, this is the largest dataset of endogenous peptides from rat brain to date. Figures 2b and 2c show the distribution of the peptides per protein for both brain areas. The majority of all identified peptides in both hypothalamus and striatum are derived from well-characterized neuropeptide precursors such as Proenkephalin-A (PENK), Proenkephalin-B (PDYN), Pro-opiomelanocortin (POMC), Cholecystokinin (CCK), Pro-MCH (PMCH), Somatostatin (SST), Protachykinin-1 (TAC1), Chromogranin-A (CHGA), ProSAAS and proteins from the Secretogranin-family (SCG1, SCG2, SCG3, SCG5); figure 2b,c. Among the identified endogenous peptides are several neuropeptides that are known to modulate food intake, such as neuropeptide Y, orexin-B, galanin, PACAP-38, neuromedin-C, neuromedin-U, neurotensin, PrRP-20, α-neoendorphin, β-neoendorphin, dynorphin A, α-MSH, β-endorphin and nociceptin. Of all identified peptides 78% (hypothalamus) and 73% (striatum), respectively, are derived from known neuropeptide precursor proteins. The remaining peptides are mostly derived from secreted extracellular proteins or plasma membrane proteins, based on gene ontology classification. Several of the identified peptides are truncated versions of longer versions of the same peptide. These truncated forms are likely to contain intermediate peptide products generated by cleavage through peptidases. This process occurs during neuropeptide maturation in the secretory pathway by processing enzymes such as carboxypeptidases (e.g. CPE) or by extracellular peptidases that target secreted proteins and peptides. Note that truncated peptides can also reflect artifacts due to post-mortem protease activity [38]. However, we estimate only a minor impact of post-mortem degradation in our experiment as we observe little protein degradation products and most of the abundant peptides in this dataset are indeed well-known neuropeptides (see below). We observed many of the neuropeptides in multiple modification states. For example, one of the major final products of the POMC precursor is desacetyl-α-melanocyte stimulating hormone (desacetyl-α-MSH). N-terminal acetylation of this peptide gives α-MSH which plays a role in food intake regulation [18]. Both peptides were identified in hypothalamus whereas in the striatum only α-MSH was detected. Both desacetyl-α-MSH and α-MSH were also identified with additional methionine oxidation. Note that N-terminal acetylation and C-terminal amidation are known to modulate neuropeptide activity and stability [58]. Oxidation of methionine is likely an artifact, that can be introduced during sample preparation.

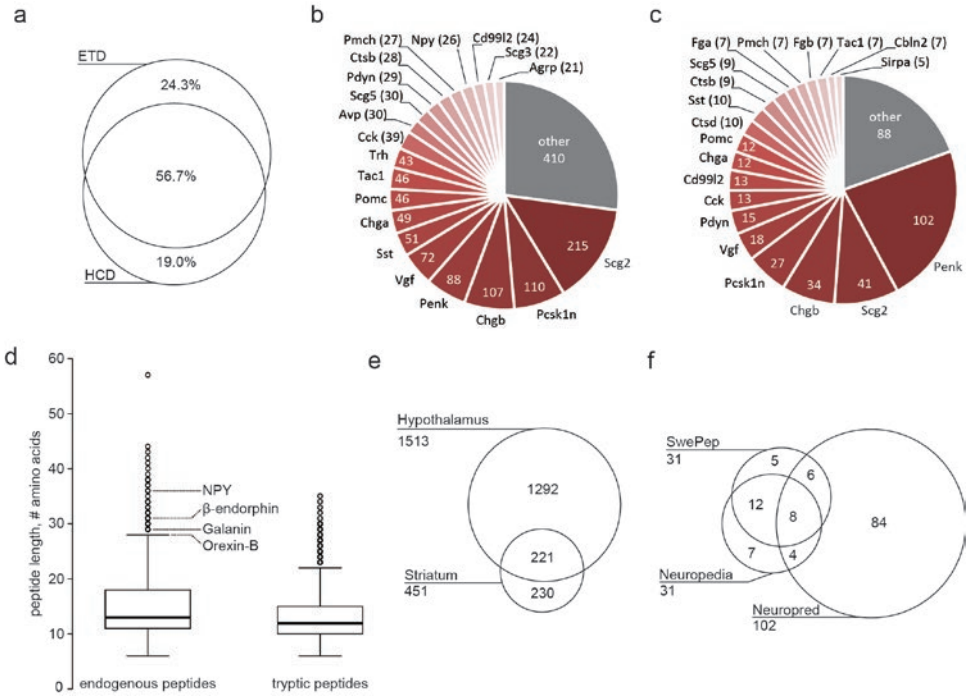



**Figure 2.** Identification of endogenous peptides from rat brain. A) Contribution of the two fragmentation techniques ETD and HCD for the total of all peptide identifications. (B, C) Distribution of the number of peptides per protein for the top 20 most abundant precursors from hypothalamus (B) and striatum (C). D) Distribution of endogenous peptide in terms of peptide length in comparison to regular tryptic peptides (reference dataset taken from [59]). Boxplots indicate the median (middle line) and the interquartile range (box). Values that exceed 1.5 × the interquartile range are marked as outliers (circles). E) Overlap of unique peptides between hypothalamus and striatum. F) Overlap of known neuropeptides (matched against two databases; SwePep and Neuropedia, respectively) and endogenous peptides that match predicted neuropeptide sequences (Neuropred).

Neuropeptides can vary heavily in length and amino acid composition, which is also evident in our data. Compared to peptides derived from tryptic digestion of proteins from a mammalian cell line we found that the endogenous peptides are more diverse in terms of peptide length (Figure 2D). The average length of the endogenous peptides in our data was 14.9 amino acids while this is 12.8 for tryptic peptides. The complex endogenous peptide processing causes the higher variability in length. Peptide outliers in the tryptic dataset are typically due to miss cleavages. Remarkably, several of the longer endogenous peptides in our dataset are mature neuropeptides that are involved in food intake regulation, such as neuropeptide Y (36 amino acids), β-endorphin (31 amino acids), galanin (29 amino acids) and orexin-B (28 amino acids). The longest peptide identified is a 57-amino acid fragment of the substance P precursor Protachykinin-1. We observed that the peptide length distribution

is slightly skewed towards higher values in the case of the endogenous peptides. Note that this can be an artifact due to the rigid 1% FDR filtering. Short endogenous peptides are less likely to be statistically significant because of the high number of theoretically expected matches by chance in database searches without enzyme specificity [60].

We determined the overlap of endogenous peptides between the two brain areas and found that 49% of the peptides from striatum were also identified in the hypothalamus (Figure 2E). Notably, endogenous peptides derived from neuropeptide Y (NPY), galanin (GAL) and agouti-related protein (AgRP) precursors were only identified in the hypothalamus, as expected. AgRP is exclusively co-expressed in NPY neurons in the ARC whereas galanin is more widely expressed in the ARC, PVN and SON in the hypothalamus [9]. We compared the unique peptide sequences (disregarding any PTMs) with the Swepep and Neuropedia databases, which contain collections of known and well-characterized neuropeptides. In total 31 sequences were matched to entries in Swepep and neuropedia, respectively (table 1), but only 20 peptides were commonly found in both databases (Figure 2F). Next, we used Neuropred [53] to identify potentially novel neuropeptides in silico. The Neuropred algorithm predicts neuropeptide sequences based on the occurrence of expected basic cleavage sites in the protein sequence.

The precursor protein sequences of all identified endogenous peptides were submitted to Neuropred and the obtained peptides were matches against the endogenous peptides in our dataset. In total 102 unique endogenous peptide sequences are predicted by Neuropred (Figure 2F). Many of the predicted neuropeptide sequences have been previously reported in other peptidomic studies (Supporting Table 2, available online). However, 24 of the predicted sequences that are derived from known neuropeptide precursors have not been reported before (Table 2). Note that six out of these contain C-terminal Arg or Lys and are thus likely intermediate processing products prior trimming of basic residues by carboxypeptidases.

**Label-free quantification of diet-induced neuropeptide level changes**

For this study samples were randomized and equal amounts based on total protein content were analyzed in triplicate. To reduce technical variability at the LC-MS-level we spiked into each sample small amounts of peptides derived from a tryptic digest of BSA. These peptides were used as internal standard to facilitate chromatographic alignment and for intensity normalization. Coefficients of variation (cv) for the retention time of the BSA peptides were between 2.1% – 3.9%, which demonstrates the robustness and reproducibility of the LC setup.

**Table 2.** Sequences of identified, previously unreported and potentially novel endogenous peptides.

| Gene | Protein name | peptide sequence |
|---|---|---|
| Adcyap1 | Pituitary adenylate cyclase-activating polypeptide | KR.YKQRVKNKGRRIAYL.- |
| Cck | Cholecystokinin | R.YIQQVRKAPSG.R |
| Chga | Chromogranin-A | KR.LEGEDDPDRSMK.L |
| Chgb | Secretogranin-1 | R.EPGAYPALDSRQE.KR |
| Npvf | Isoform 1 of FMRFamide-related peptides | R.SVTFQELKDWGA.KK |
| Npy | Pro-neuropeptide Y | KR.SSPETLISDLLMRESTENAPR.T |
| Nts | Neurotensin/neuromedin N | KR.QLYENKP.RR |
| Pcsk1n | ProSAAS | KR.LENSSPQAPARR.L |
| Pcsk2 | Neuroendocrine convertase 2 | KR.QLERDPRIKMALQQEGFD.RK |
| Pdyn | Proenkephalin-B preproprotein | KR.SSEMAGDEDRGQDGDQVGHEDLY.KR |
| Pdyn | Proenkephalin-B preproprotein | KR.YGGFLRRQFK.V |
| Pmch | Pro-MCH | K.AFQKEDTAE.R |
| Pnoc | Prenociceptin | RR.TLHQNGIQVSPR.T |
| Pomc | Pro-opiomelanocortin | RR.AEEETAGGDGRPEPSPREG.KR |
| Scg2 | Secretogranin-2 | K.ENTEKNEQINEEM.KR |
| Scg2 | Secretogranin-2 | KR.SGHLGLPDEGNRKESKDQLSEDASKVITYLRR.L |
| Scg5 | Neuroendocrine protein 7B2 | RR.SVNPYLQGKRLDNVVAKKSVPHFSEEEKEPE.- |
| Stc1 | Stanniocalcin-1 | R.GEGDSPSHI.KR |
| Tac1 | Protachykinin-1, isoform delta | KR.DAGHGQISHKMAYERSAMQNYE.RR |
| Tac2 | Tachykinin-3 | KR.DMHDFFVGLMG.KR |
| Trh | Prothyroliberin | KR.QHPGRRFIDPELQRSWEEKEGEGVLMPE.KR |
| Trh | Prothyroliberin | KR.QHPGRRSFPWMESDVT.KR |
| Trh | Prothyroliberin | R.SWEEKEGEGVLMPE.KR |
| Ucn3 | Urocortin 3 | KK.NKLEDVPVLS.KK |

Figure 3 shows the distribution of the relative abundance of all endogenous peptides that were identified, spanning four orders of magnitude. The data highlights that peptides, although derived from the same precursor protein, can greatly vary in abundance. For instance, in both hypothalamus (Figure 3A) and striatum (Figure 3B) β-endorphin peptides are present in the lower quartile range whereas α-MSH is found in the upper quartile. In the hypothalamus, many of the most abundant peptides are derived from proSAAS (PCSK1N) and proenkephalin-B (PDYN). In the striatum, several of the high abundant peptides are derived from proenkephalin-A (PENK).

Previous studies revealed that proSAAS-derived peptides co-localize with NPY in the hypothalamus and are potentially involved in orexigenic food intake signaling [61]. Enzymatic processing of the precursor protein proSAAS generates multiple, and partially overlapping neuropeptides of varying length (Figure 4A. Initial cleavage by prohormone convertases produces the terminal peptides big SAAS and PEN-LEN (not identified). Further consecutive processing yields the shorter peptides little SAAS, PEN, big LEN and little LEN. Notably, in our dataset the relative levels of these peptides do not follow the same trend between the different groups. Big SAAS and PEN are present at elevated levels in the HFHS group. On the contrary, the peptides little SAAS, big LEN and little LEN show no significant difference between the groups.

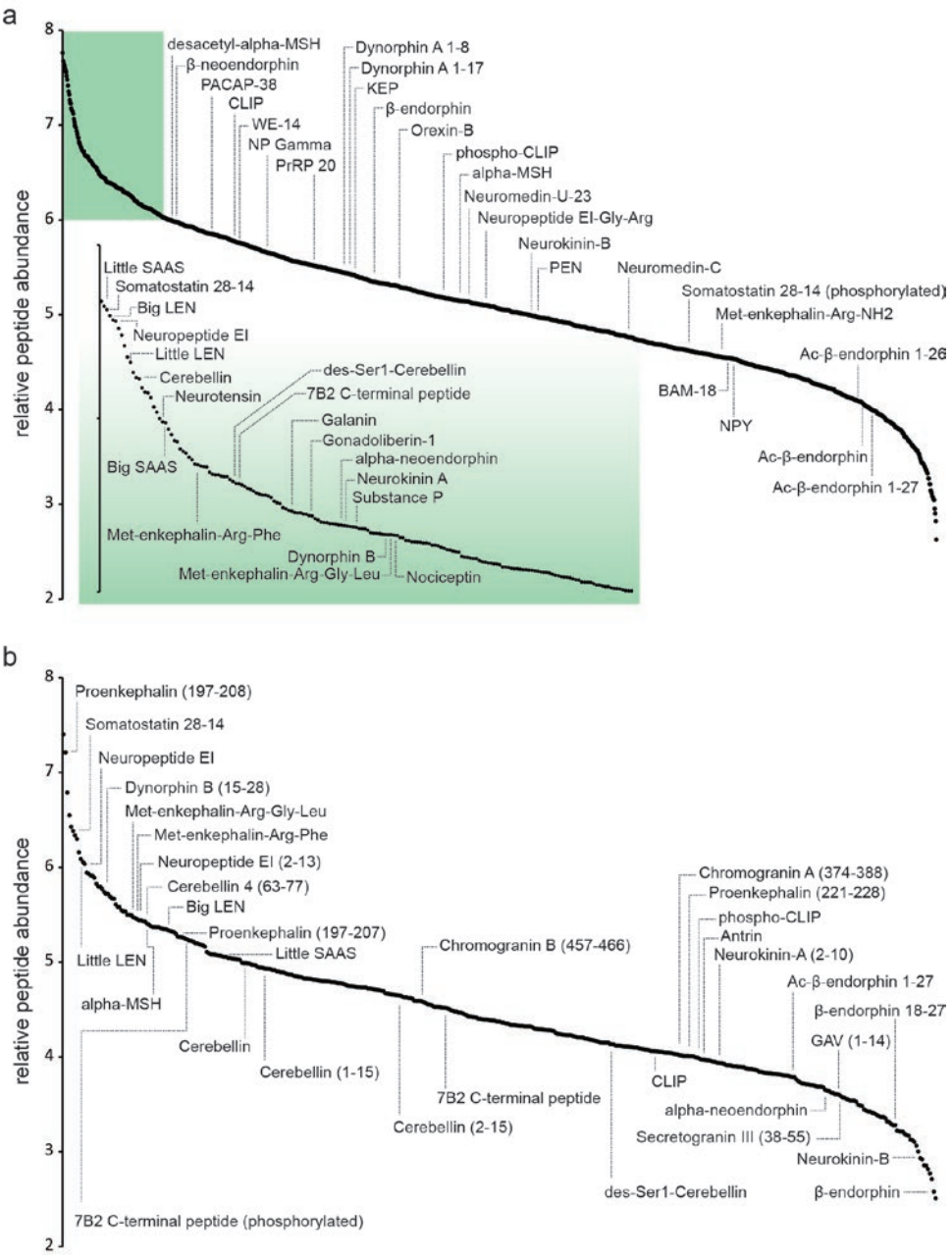



**Figure 3.** Relative abundance of endogenous peptides from hypothalamus (A) and striatum (B). For each peptide the peak area was calculated and log10 transformed. Data from all animals were averaged. The inset in panel A shows a zoom-in of the region highlighted in green.

Observation of differential levels of endogenous peptides that are derived from the same precursor might be related to distinct specificity and activity of the prohormone convertases involved [62]. Big SAAS and PEN are derived by N-terminal cleavage at a signal peptide cleavage site (Big SAAS) and a Arg-Arg dibasic cleavage site (PEN), respectively. The other proSAAS peptides that show no clear upregulation in one of the groups are derived from N-terminal processing at a single Arg (little SAAS) and Lys-Arg dibasic cleavage site (little LEN; big LEN), respectively [63]. Together, the data indicates that prediction of neuropeptide abundance based on detection of other peptides from the same precursor protein or the precursor protein itself is not possible and bioactive end products need to be monitored individually.

We compared the relative neuropeptide levels between hypothalamus and striatum (Figure 4B). Some neuropeptides, for example α-neoendorphin, show a very similar abundance pattern across the groups in hypothalamus and in striatum. Others, such as β-endorphin are present at different relative levels between the two brain areas. In the hypothalamus β-endorphin is highly abundant in the CHOC and HFHS groups, and very low abundant in the RESTR group. On the contrary, the same neuropeptide is present at elevated level in the control group in the striatum. Moreover, our data suggest differential neuropeptide processing in the two brain areas. The post-translational modified and truncated form Ac-β-endorphin 1-27 is equally abundant across all groups in the striatum while in the hypothalamus we found it to be relatively increased in the HFHS group (Figure 4B). Overall, the variance between the individual animals and between the differentially treated dietary groups was high. The median intra- and inter-group coefficient of variation for the peptide abundance was 78% and 124%, respectively. The high variation in this experiment is probably due to the low number of animals per group (n=3) and puts a strain on biological interpretation of the data. However, several neuropeptides show an interesting and substantial increase in the HFHS group. Orexigenic neuropeptides involved in food intake regulation, such as neuropeptide Y, orexin-B and galanin were substantially overrepresented in the HFHS group compared to the other groups in the hypothalamus (Figure 4C). This suggests that the animals on this high-energy diet develop increased appetite. On the other hand, we also found elevated levels of anorexigenic neuropeptides, such as α-MSH, neurotensin and neuromedin-U (Figure 4D), all of which mediate decreased food intake [19, 64, 65]. La Fleur et al. have previously reported on elevated NPY mRNA levels in rats that were subjected to the same HFHS diet [56]. They also observed decreased POMC mRNA in the same experiment. However, this is in contrast to our data that revealed α-MSH to be equally elevated in the HFHS group (Figure 4D).

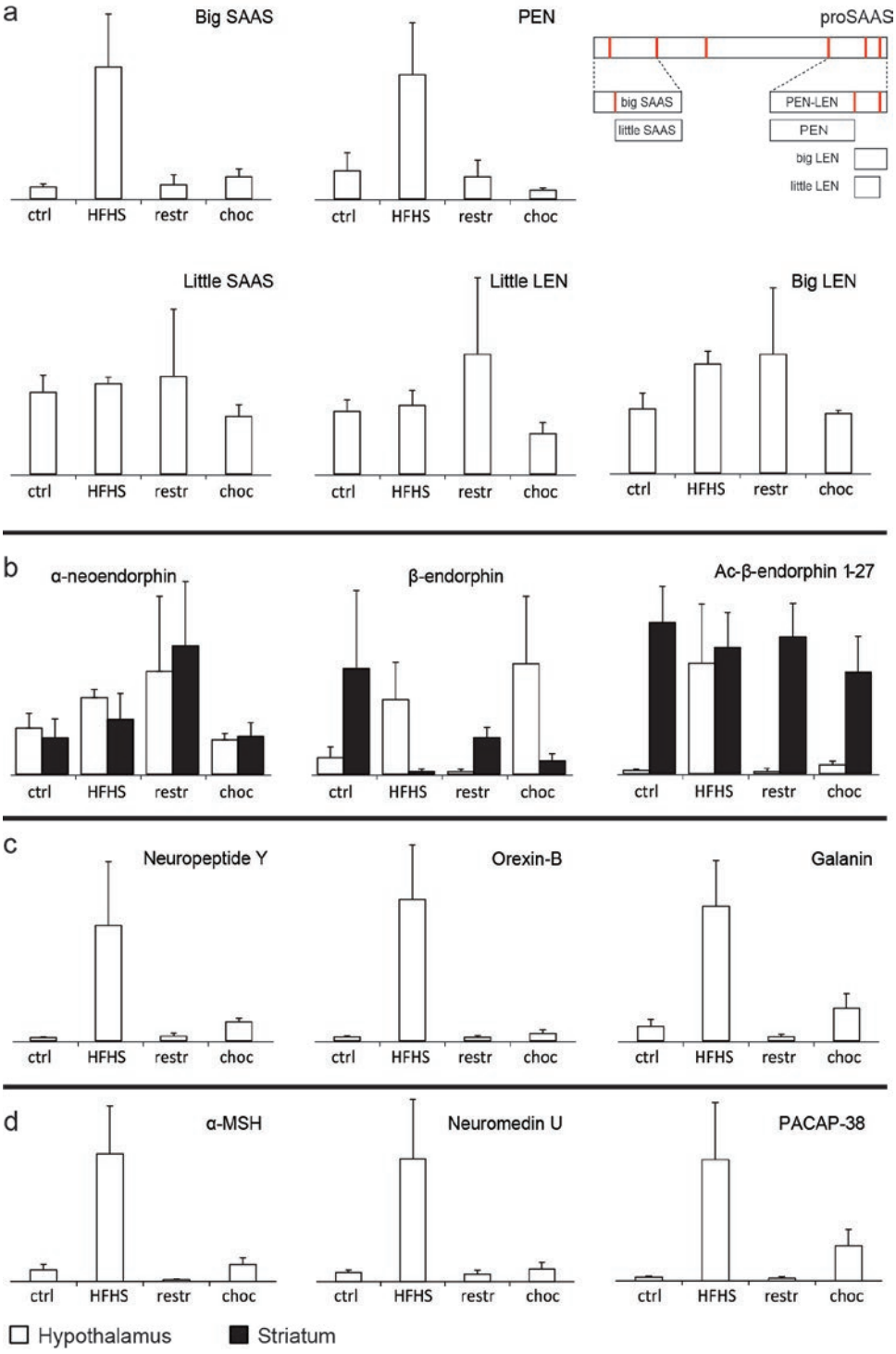

a
Big SAAS
PEN
proSAAS
big SAAS
little SAAS
PEN-LEN
PEN
big LEN
little LEN
Little SAAS
Little LEN
Big LEN
ctrl
HFHS
restr
choc
b
α-neoendorphin
β-endorphin
Ac-β-endorphin 1-27
c
Neuropeptide Y
Orexin-B
Galanin
d
α-MSH
Neuromedin U
PACAP-38
Hypothalamus
Striatum

**Figure 4.** Relative label-free quantification of neuropeptides in hypothalamus and striatum. A) Levels of five neuropeptides derived from the same precursor protein proSAAS. Vertical red lines illustrate prohormone convertase cleavage sites. B) Comparison of relative neuropeptide levels in hypothalamus and striatum. (C, D) Quantitative data of neuropeptides that play key roles in food intake regulation. Elevated levels of orexigenic (panel C, neuropeptide Y, orexin-B and galanin) and anorexigenic (panel D, alpha-MSH, neuromedin U, PACAP-38) neuropeptides were observed in HFHS rats. Abbreviations: ctrl, control; HFHS, high-fat/high-sugar; restr, restricted food; choc, chocolate anticipating. All charts display normalized neuropeptide abundances.

## CONCLUSION & OUTLOOK

In this study we have identified over 1700 unique endogenous peptides in rat hypothalamus and striatum. The high number of endogenous peptides identified highlights the efficient peptide extraction procedure using a combination of acidified acetone [32] and acetic acid [29]. Employing two complementary fragmentation techniques (HCD and ETD) provides for high peptide identification rates, because it increases the probability of obtaining sufficient peptide sequence information by at least one of the two techniques. Future instrument developments that improve sensitivity will facilitate the analysis of neuropeptide extracts from tissue slices or punches to a similar depth as from whole tissue extracts. Extracting endogenous peptides from spatially defined areas would decrease the overall variability and potentially provides additional spatial resolution, which is desired to investigate neuropeptide populations in defined nuclei (or even single cells) apart from adjacent brain areas.

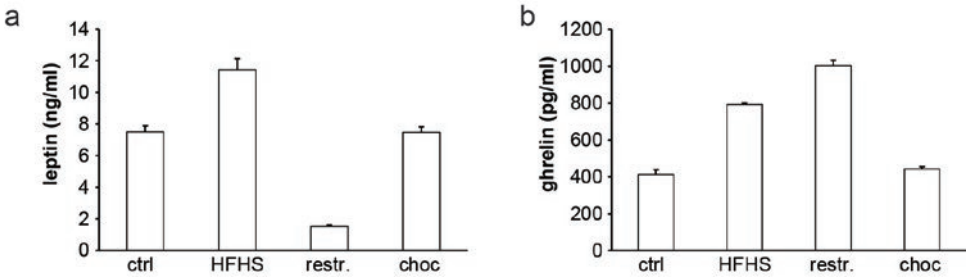



**Figure 5.** Blood plasma levels of leptin (A) and ghrelin (B). All data are given as mean ± SEM.

The quantitative data highlights the advantage of direct neuropeptide profiling by mass spectrometry that can reveal relative abundance of individual neuropeptides rather than just quantitative information on the precursor proteins. Despite the high overall variation in our dataset, which is likely caused by the small number of animals per group, we found several known neuropeptides elevated in the HFHS group. Comparable upregulation of orexigenic and anorexigenic neuropeptides might reflect conflicting signaling in HFHS diet

exposed rats as we have observed increased motivation for palatable food in these rats while they are also more obese [66]. Leptin (anorexigenic) and ghrelin (orexigenic) are endocrine peptide hormones that have opposing effects on food intake and energy balance, mediated by binding to receptors of hypothalamic neurons [13]. Leptin plasma levels were found to be decreased in RESTR but elevated in HFHS relative to the control group (Figure 5). Surprisingly, ghrelin was also elevated in HFHS compared to control, which might explain why we observe elevated levels of oppositely acting neuropeptides in the HFHS group.

## ACKNOWLEDGEMENTS

This research was performed within the framework of the PRIME-XS project, grant number 262067, funded by the European Union 7th Framework Program. Additionally, the Netherlands Proteomics Centre, a program embedded in the Netherlands Genomics Initiative, is kindly acknowledged for financial support as well as the Netherlands Organization for Scientific Research (NWO) with the VIDI grant for SM (700.10.429).

## REFERENCES

[1] T. Hokfelt, C. Broberger, Z.Q. Xu, V. Sergeyev, R. Ubink, M. Diez, Neuropeptides--an overview, Neuropharmacology, 39 (2000) 1337-1356.

[2] A.B. Hummon, T.A. Richmond, P. Verleyen, G. Baggerman, J. Huybrechts, M.A. Ewing, E. Vierstraete, S.L. Rodriguez-Zas, L. Schoofs, G.E. Robinson, J.V. Sweedler, From the genome to the proteome: uncovering peptides in the Apis brain, Science, 314 (2006) 647-649.

[3] R.J. Di Leone, J.R. Taylor, M.R. Picciotto, The drive to eat: comparisons and distinctions between mechanisms of food reward and drug addiction, Nat Neurosci, 15 (2012) 1330-1335.

[4] G.A. Tejwani, A.K. Rattan, Met-enkephalin alteration in the rat during chronic injection of morphine and/or midazolam, Brain Res, 775 (1997) 119-126.

[5] R.J. Bodnar, G.E. Klein, Endogenous opiates and behavior: 2003, Peptides, 25 (2004) 2205-2256.

[6] M.J. Kreek, S.D. Schlussman, B. Reed, Y. Zhang, D.A. Nielsen, O. Levran, Y. Zhou, E.R. Butelman, Bidirectional translational research: Progress in understanding addictive diseases, Neuropharmacology, 56 (2009) 32-43.

[7] V. Hook, S. Bark, N. Gupta, M. Lortie, W.D. Lu, N. Bandeira, L. Funkelstein, J. Wegrzyn, D.T. O'Connor, P. Pevzner, Neuropeptidomic components generated by proteomic functions in secretory vesicles for cell-cell communication, Aaps J, 12 (2010) 635-645.

[8] G.J. Morton, D.E. Cummings, D.G. Baskin, G.S. Barsh, M.W. Schwartz, Central nervous system control of food intake and body weight, Nature, 443 (2006) 289-295.

[9] J.A. Parker, S.R. Bloom, Hypothalamic neuropeptides and the regulation of appetite, Neuropharmacology, 63 (2012) 18-30.

[10] J.J. Hillebrand, D. de Wied, R.A. Adan, Neuropeptides, food intake and body weight regulation: a hypothalamic focus, Peptides, 23 (2002) 2283-2306.

[11] K.G. Murphy, S.R. Bloom, Gut hormones and the regulation of energy homeostasis, Nature, 444 (2006) 854-859.

[12] H. Munzberg, J.S. Flier, C. Bjorbaek, Region-specific leptin resistance within the hypothalamus of diet-induced obese mice, Endocrinology, 145 (2004) 4880-4889.

[13] X. Shan, G.S. Yeo, Central leptin and ghrelin signalling: comparing and contrasting their mechanisms of action in the brain, Rev Endocr Metab Disord, 12 (2011) 197-209.

[14] R.E. Mercer, M.J. Chee, W.F. Colmers, The role of NPY in hypothalamic mediated food intake, Front Neuroendocrinol, 32 (2011) 398-415.

[15] A.P. Chambers, S.C. Woods, The role of neuropeptide Y in energy homeostasis, Handb Exp Pharmacol, (2012) 23-45.

[16] G.E. Bergonzelli, F.P. Pralong, M. Glauser, C. Cavadas, E. Grouzmann, R.C. Gaillard, Interplay between galanin and leptin in the hypothalamic control of feeding via corticotropin-releasing hormone and neuropeptide Y, Diabetes, 50 (2001) 2666-2672.

[17] I. Mechenthaler, Galanin and the neuroendocrine axes, Cellular and molecular life sciences : CMLS, 65 (2008) 1826-1835.

[18] K.G. Mountjoy, Functions for pro-opiomelanocortin-derived peptides in obesity and diabetes, Biochem J, 428 (2010) 305-324.

[19] R. Pandit, J.W. de Jong, L.J. Vanderschuren, R.A. Adan, Neurobiology of overeating and obesity: the role of melanocortins and beyond, Eur J Pharmacol, 660 (2011) 28-42.

[20] A.E. Kelley, B.A. Baldo, W.E. Pratt, M.J. Will, Corticostriatal-hypothalamic circuitry and food motivation: integration of energy, action and reward, Physiol Behav, 86 (2005) 773-795.

[21] A.E. Kelley, K.C. Berridge, The neuroscience of natural rewards: relevance to addictive drugs, J Neurosci, 22 (2002) 3306-3311.

[22] N.D. Volkow, G.J. Wang, R.D. Baler, Reward, dopamine and the control of food intake: implications for obesity, Trends Cogn Sci, 15 (2011) 37-46.

[23] R. von Eggelkraut-Gottanka, A.G. Beck-Sickinger, Biosynthesis of peptide hormones derived from precursor sequences, Curr Med Chem, 11 (2004) 2651-2665.

[24] M. Rholam, C. Fahy, Processing of peptide and hormone precursors at the dibasic cleavage sites, Cell Mol Life Sci, 66 (2009) 2075-2091.

[25] J.L. Arolas, J. Vendrell, F.X. Aviles, L.D. Fricker, Metallocarboxypeptidases: Emerging drug targets in biomedicine, Curr Pharm Design, 13 (2007) 349-366.

[26] L.D. Fricker, J.Y. Lim, H. Pan, F.Y. Che, Peptidomics: identification and quantification of endogenous peptides in neuroendocrine tissues, Mass Spectrometry Reviews, 25 (2006) 327-344.

[27] L.J. Li, J.V. Sweedler, Peptides in the Brain: Mass Spectrometry-Based Measurement Approaches and Challenges, Annu Rev Anal Chem, 1 (2008) 451-483.

[28] M. Svensson, K. Skold, P. Svenningsson, P.E. Andren, Peptidomics-based discovery of novel neuropeptides, Journal of proteome research, 2 (2003) 213-219.

[29] A.F. Altelaar, S. Mohammed, M.A. Brans, R.A. Adan, A.J. Heck, Improved identification of endogenous peptides from murine nervous tissue by multiplexed peptide extraction methods and multiplexed mass spectrometric analysis, J Proteome Res, 8 (2009) 870-876.

[30] F.Y. Che, J. Lim, H. Pan, R. Biswas, L.D. Fricker, Quantitative neuropeptidomics of microwave-irradiated mouse brain and pituitary, Molecular & cellular proteomics : MCP, 4 (2005) 1391-1405.

[31] J.E. Lee, L. Zamdborg, B. Southey, N. Atkins, J.W. Mitchell, M. Li, M.U. Gillette, N.L. Kelleher, J.V. Sweedler, Quantitative Peptidomics for Discovery of Circadian-Related Peptides from the Rat Suprachiasmatic Nucleus, Journal of proteome research, (2012).

[32] J.E. Lee, N. Atkins, N.G. Hatcher, L. Zamdborg, M.U. Gillette, J.V. Sweedler, N.L. Kelleher, Endogenous Peptide Discovery of the Rat Circadian Clock A FOCUSED STUDY OF THE SUPRACHIASMATIC NUCLEUS BY ULTRAHIGH PERFORMANCE TANDEM MASS SPECTROMETRY, Molecular & Cellular Proteomics, 9 (2010) 285-297.

[33] A. Bora, S.P. Annangudi, L.J. Millet, S.S. Rubakhin, A.J. Forbes, N.L. Kelleher, M.U. Gillette, J.V. Sweedler, Neuropeptidomics of the supraoptic rat nucleus, Journal of proteome research, 7 (2008) 4992-5003.

[34] A.F. Altelaar, J. van Minnen, C.R. Jimenez, R.M. Heeren, S.R. Piersma, Direct molecular imaging of Lymnaea stagnalis nervous tissue at subcellular spatial resolution by mass spectrometry, Analytical chemistry, 77 (2005) 735-741.

[35] S. Neupert, S.S. Rubakhin, J.V. Sweedler, Targeted single-cell microchemical analysis: MS-based peptidomics of individual paraformaldehyde-fixed and immunolabeled neurons, Chem Biol, 19 (2012) 1010-1019.

[36] E.J. Lanni, S.S. Rubakhin, J.V. Sweedler, Mass spectrometry imaging and profiling of single cells, J Proteomics, 75 (2012) 5036-5051.

[37] M. Svensson, K. Skold, A. Nilsson, M. Falth, K. Nydahl, P. Svenningsson, P.E. Andren, Neuropeptidomics: MS applied to the discovery of novel peptides from the brain, Analytical Chemistry, 79 (2007) 14-21.

[38] K. Skold, M. Svensson, M. Norrman, B. Sjogren, P. Svenningsson, P.E. Andren, The significance of biochemical and molecular sample integrity in brain proteomics and peptidomics: Stathmin 2-20 and peptides as sample quality indicators, Proteomics, 7 (2007) 4445-4456.

[39] R.A. Zubarev, N.L. Kelleher, F.W. McLafferty, Electron capture dissociation of multiply charged protein cations. A nonergodic process, J Am Chem Soc, 120 (1998) 3265-3266.

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

[41] L. Hui, R. Cunningham, Z. Zhang, W. Cao, C. Jia, L. Li, Discovery and characterization of the Crustacean hyperglycemic hormone precursor related peptides (CPRP) and orcokinin neuropeptides in the sinus glands of the blue crab Callinectes sapidus using multiple tandem mass spectrometry techniques, Journal of proteome research, 10 (2011) 4219-4229.

[42] L. Fornelli, A.W. Schmid, L. Grasso, H. Vogel, Y.O. Tsybin, Deamidation and transamidation of substance P by tissue transglutaminase revealed by electron-capture dissociation fourier transform mass spectrometry, Chemistry, 17 (2011) 486-497.

[43] C.C. Wu, M.J. MacCoss, K.E. Howell, D.E. Matthews, J.R. Yates, Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis, Analytical Chemistry, 76 (2004) 4951-4959.

[44] M. Kruger, M. Moser, S. Ussar, I. Thievessen, C.A. Luber, F. Forner, S. Schmidt, S. Zanivan, R. Fassler, M. Mann, SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function, Cell, 134 (2008) 353-364.

[45] F.Y. Che, L.D. Fricker, Quantitative peptidomics of mouse pituitary: comparison of different stable isotopic tags, Journal of mass spectrometry : JMS, 40 (2005) 238-249.

[46] X. Hou, F. Xie, J.V. Sweedler, Relative Quantitation of Neuropeptides over a Thousand-fold Concentration Range, J Am Soc Mass Spectr, 23 (2012) 2083-2093.

[47] A.F. Altelaar, C.K. Frese, C. Preisinger, M.L. Hennrich, A.W. Schram, H.T. Timmers, A.J. Heck, S. Mohammed, Benchmarking stable isotope labeling based quantitative proteomics, J Proteomics, (2012).

[48] J.E. Lee, N. Atkins, Jr., N.G. Hatcher, L. Zamdborg, M.U. Gillette, J.V. Sweedler, N.L. Kelleher, Endogenous peptide discovery of the rat circadian clock: a focused study of the suprachiasmatic nucleus by ultrahigh performance tandem mass spectrometry, Mol Cell Proteomics, 9 (2010) 285-297.

[49] A.F. Altelaar, S. Mohammed, M.A. Brans, R.A. Adan, A.J. Heck, Improved identification of endogenous peptides from murine nervous tissue by multiplexed peptide extraction methods and multiplexed mass spectrometric analysis, J Proteome Res, 8 (2009) 870-876.

[50] C.K. Frese, A.F. Altelaar, M.L. Hennrich, D. Nolting, M. Zeller, J. Griep-Raming, A.J. Heck, S. Mohammed, Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos, Journal of proteome research, 10 (2011) 2377-2388.

[51] T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions, Nat Methods, 8 (2011) 785-786.

[52] H.W. van den Toorn, J. Munoz, S. Mohammed, R. Raijmakers, A.J. Heck, B. van Breukelen, RockerBox: analysis and filtering of massive proteomics search results, J Proteome Res, 10 (2011) 1420-1424.

[53] B.R. Southey, A. Amare, T.A. Zimmerman, S.L. Rodriguez-Zas, J.V. Sweedler, NeuroPred: a tool to predict cleavage sites in neuropeptide precursors and provide the masses of the resulting peptides, Nucleic Acids Res, 34 (2006) W267-272.

[54] M. Falth, K. Skold, M. Norrman, M. Svensson, D. Fenyo, P.E. Andren, SwePep, a database designed for endogenous peptides and mass spectrometry, Mol Cell Proteomics, 5 (2006) 998-1005.

[55] S.E. la Fleur, M.C. Luijendijk, A.J. van Rozen, A. Kalsbeek, R.A. Adan, A free-choice high-fat high-sugar diet induces glucose intolerance and insulin unresponsiveness to a glucose load not explained by obesity, Int J Obes (Lond), 35 (2011) 595-604.

[56] S.E. la Fleur, A.J. van Rozen, M.C. Luijendijk, F. Groeneweg, R.A. Adan, A free-choice high-fat high-sugar diet induces changes in arcuate neuropeptide expression that support hyperphagia, Int J Obes (Lond), 34 (2010) 537-546.

[57] K. Sasaki, T. Osaki, N. Minamino, Large-scale Identification of Endogenous Secretory Peptides Using Electron Transfer Dissociation Mass Spectrometry, Molecular & cellular proteomics : MCP, 12 (2013) 700-709.

[58] F. Cuttitta, Peptide amidation: signature of bioactivity, Anat Rec, 236 (1993) 87-93, 172-173; discussion 193-175.

[59] C.K. Frese, M. Altelaar, H.W. van den Toorn, D. Nolting, J. Griep-Raming, A.J. Heck, S. Mohammed, Towards Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry, Analytical chemistry, (2012).

[60] M.N. Akhtar, B.R. Southey, P.E. Andren, J.V. Sweedler, S.L. Rodriguez-Zas, Evaluation of Database Search Programs for Accurate Detection of Neuropeptides in Tandem Mass Spectrometry Experiments, Journal of proteome research, (2012).

[61] J.H. Wardman, I. Berezniuk, S. Di, J.G. Tasker, L.D. Fricker, ProSAAS-derived peptides are colocalized with neuropeptide Y and function as neuropeptides in the regulation of food intake, PLoS One, 6 (2011) e28152.

[62] J.H. Wardman, X. Zhang, S. Gagnon, L.M. Castro, X. Zhu, D.F. Steiner, R. Day, L.D. Fricker, Analysis of peptides in prohormone convertase 1/3 null mouse brain using quantitative peptidomics, J Neurochem, 114 (2010) 215-225.

[63] L.D. Fricker, A.A. McKinzie, J. Sun, E. Curran, Y. Qian, L. Yan, S.D. Patterson, P.L. Courchesne, B. Richards, N. Levin, N. Mzhavia, L.A. Devi, J. Douglass, Identification and characterization of proSAAS, a granin-like neuroendocrine peptide precursor that inhibits prohormone processing, J Neurosci, 20 (2000) 639-648.

[64] L. Mounien, J.C. Do Rego, P. Bizet, I. Boutelet, G. Gourcerol, A. Fournier, P. Brabet, J. Costentin, H. Vaudry, S. Jegou, Pituitary adenylate cyclase-activating polypeptide inhibits food intake in mice through activation of the hypothalamic melanocortin system, Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology, 34 (2009) 424-435.

[65] D.A. Bechtold, T.R. Ivanov, S.M. Luckman, Appetite-modifying actions of pro-neuromedin U-derived peptides, Am J Physiol Endocrinol Metab, 297 (2009) E545-551.

[66] S.E. la Fleur, L.J. Vanderschuren, M.C. Luijendijk, B.M. Kloeze, B. Tiesjema, R.A. Adan, A reciprocal interaction between food-motivated behavior and diet-induced obesity, Int J Obes (Lond), 31 (2007) 1286-1294.

Chapter 8

# Summary
# Samenvatting
# Future Outlook
# Publications
# Curriculum Vitae
# Acknowledgements
# Abbreviations

## SUMMARY

**Chapter 1** gives a general introduction to mass spectrometry-based proteomics. In particular, the principles of different peptide fragmentation techniques are discussed. Peptide sequencing is key to any MS-based proteomic experiment and provides the foundation for structural analysis. Collision induced dissociation is the most established dissociation technique and is nowadays routinely used for peptide sequencing. Over the last years, electron-driven approaches such as electron transfer dissociation have evolved as valuable alternative fragmentation techniques. Moreover, the Orbitrap, the most recently developed mass analyzer, is described. A brief introduction of the principles of orbital trapping is given followed by an overview of recent advances in Orbitrap technology. The second part of chapter 1 provides an overview of current approaches and challenges in quantitative proteomics. Fundamental aspects of the most commonly utilized stable isotope-based labeling strategies are discussed. Furthermore, approaches and challenges in quantitative and statistical data analysis are described.

In **chapter 2**, a systematic comparison of all possible combinations of fragmentation techniques and mass analyzers on an hybrid ion trap-Orbitrap is reported. The Orbitrap Velos features three main fragmentation techniques: resonant-excitation CID, beam-type CID and electron transfer dissociation. MS/MS scans can be performed in the linear ion trap (fast, high sensitivity) or the Orbitrap mass analyzer (high resolving power and mass accuracy). We investigate which combination of fragmentation technique and mass analyzer provides the best performance for the analysis of distinct peptide populations such as N-acetylated, phosphorylated, and tryptic peptides. Our data highlights that that there is a trade-off between spectral quality and speed when using the Orbitrap for fragment ion detection. We conclude that a enhanced decision-tree regulated combination of higher-energy collisional dissociation (HCD) and ETD can improve confidence in peptide sequencing.

In **chapter 3,** we introduce a new peptide fragmentation method combining electron transfer and higher-energy collision dissociation, coined EThcD. The unreacted and charge-reduced precursor ions are usually the dominating ions in ETD spectra due to the low efficiency for tryptic peptides. Here, we utilize beam-type collisional activation to increase the overall product ion yield. We demonstrate that fragmentation of both the unreacted and charge reduced precursor while preserving all c- and z-ions generated by the preceding ETD reaction is feasible. We observe that EThcD provides more peptide backbone fragments, translating into near-complete peptide sequence coverage. Moreover, we show that EThcD spectra contain multiple b/c- and y/z-ion pairs that facilitate de novo sequencing.

In **chapter 4,** we applied EThcD fragmentation for the analysis of phosphopeptides. Localization of post-translational modifications such as phosphorylation and glycosylation is challenging because it requires observation of site-determining fragment ions. As shown in chapter 3, EThcD generates dual ion series that increase the overall peptide sequence coverage. Here, we show that EThcD is particularly for the analysis of phosphopeptides. We analyzed (i) synthetic phosphopeptides and (ii) Ti-IMAC-enriched phosphopeptides to demonstrate the potential of EThcD for phosphosite localization. Our data shows that EThcD facilitates unambiguous phosphosite localization for 95% of all phosphorylation sites.

In **chapter 5,** we describe a novel instrumental setup to perform electron transfer reactions on a hybrid LTQ-Orbitrap. ETD is commonly performed in ion traps utilizing RF fields that facilitate efficient electron transfer reactions. In this chapter, we explore performing ETD in the HCD collision cell on an Orbitrap Velos instrument by applying a static DC gradient axially to the rods. This gradient enables simultaneous three dimensional, charge sign independent, trapping of cations and anions, initiating electron transfer reactions in the center of the HCD cell where oppositely charged ions clouds overlap. We evaluate this mode of operation for a number of tryptic peptide populations and the top-down sequence analysis of ubiquitin. Our preliminary data shows that performing ETD in the HCD cell provides similar fragmentation as ion trap-ETD but requires further optimization to match performance of ion trap-ETD.

In **chapter 6**, we benchmark three of the currently most popular approaches for MS-based protein quantification: metabolic labeling, chemical labeling utilizing MS-level quantitation and isobaric labeling. We assess the performance in large scale quantitative proteomics experiments. Our data indicate that all three can reach a similar depth in number of identified proteins using a classical (MS2 based) shotgun approach. SILAC and chemical labeling with MS based quantification performed equally well regarding precision in quantitation. We emphasize that isobaric labeling using MS2-level quantification is heavily affected by co-isolation leading to compromised precision and accuracy. This issue may be partly resolved by using an MS3 based acquisition; however, at the cost of a significant reduction in number of proteins quantified. We conclude that currently MS-level based quantification approaches are preferred for accurate protein quantification.

In **chapter 7**, we analyze endogenous neuropeptides extracted from hypothalamus and striatum in a rat model for diet-induced obesity. Neuropeptides are key players in food in-

take regulation. Here, we analyze brain extract from two different brain areas. Based on the results from chapter 2, we utilized HCD and ETD fragmentation to increase the overall peptide identification rate. Overall we identified more than 1700 unique endogenous peptides, including virtually all neuropeptides that are known to be involved in food intake regulation. The quantitative data highlights differential processing of endogenous peptides that are derived from the same precursor protein. Moreover, we found comparable upregulation of orexigenic and anorexigenic potentially indicates conflicting signaling in rats that were fed on a high fat/high sucrose diet.

## SAMENVATTING

**Hoofdstuk 1** geeft een algemene inleiding over, op massaspectrometrie gebaseerde, proteomics. Met name worden de principes van de peptide fragmentatietechnieken besproken. De bepaling van de peptide sequentie is een essentieel onderdeel van een MS-gebaseerd proteomics experiment en vormt de basis voor eiwit structuur analyse. 'Collision induced dissociation' (CID) is de meest gebruikte dissociatie techniek en wordt tegenwoordig routinematig ingezet voor het ophelderen van peptide sequenties. In de afgelopen jaren hebben elektron-gestuurde benaderingen zoals 'electron transfer dissociation' (ETD) zich ontwikkeld tot waardevol alternatief. Hiernaast wordt de Orbitrap, de laatst ontwikkelde 'mass-analyzer', beschreven. Een korte introductie over de principes van de orbitale trapping wordt gegeven, gevolgd door een overzicht van recente ontwikkelingen in Orbitrap technologie. Het tweede deel van hoofdstuk 1 geeft een overzicht van de huidige aanpak en uitdagingen in kwantitatieve proteomics. Fundamentele aspecten van de meest gebruikte strategieën om peptiden met zware isotopen te labelen worden besproken. Verder worden de aanpak van en uitdagingen in kwantitatieve en statistische data-analyse beschreven.

**Hoofdstuk 2** beschrijft een systematische vergelijking van alle mogelijke combinaties van fragmentatie technieken en 'mass-analyzers' gebruik makend van een hybride iontrap-Orbitrap. De Orbitrap Velos biedt drie belangrijke fragmentatie technieken: resonantie-excitatie CID, 'bundel-type' CID en ETD. MS/MS-scans kunnen in de lineaire iontrap worden uitgevoerd (snel, hoge gevoeligheid) of in de Orbitrap (hoog oplossend vermogen en massa nauwkeurigheid). We onderzoeken welke combinatie van fragmentatie techniek en mass-analyzer de beste prestatie levert voor de analyse van afzonderlijke peptide populaties zoals N-geacetyleerde, gefosforyleerde of tryptische peptiden. Onze gegevens wijzen op een balans tussen spectrale kwaliteit en snelheid bij gebruik van de Orbitrap voor detectie van de fragment ionen. We concluderen dat een combinatie van 'higher-energy collisional dissociation' (HCD) en ETD, aangedreven door een automatische besluitvormingscascade, het vertrouwen in de bepaling van de peptidesequentie verbetert.

In **hoofdstuk 3** introduceren we een nieuwe peptide fragmentatie methode, genoemd EThcD, waarbij we ETD en HCD combineren. De meest dominante ionen in ETD spectra van tryptische peptiden zijn meestal zowel de niet gereageerde als de lading-gereduceerde precursor ionen, dit vanwege de lage ETD efficiëntie voor peptiden met een lading $\leq 3$. In dit hoofdstuk beschrijven we het gebruik van HCD activering van al de ionen na ETD, om de totale product-ion opbrengst te verhogen. We laten zien dat fragmentatie van zowel de niet gereageerde als de lading gereduceerde precursor, met behoud alle c- en z-ionen

gegenereerd door de voorgaande ETD reactie, haalbaar is. We zien dat EThcD meer peptide fragmenten oplevert, wat leidt tot bijna volledige opheldering van de peptidesequentie. Bovendien tonen wij aan dat EThcD spectra meerdere b/c- en y / z-ion paren bevatten die de novo opheldering van peptide sequenties vergemakkelijken.

In **hoofdstuk 4** hebben we EThcD fragmentatie toegepast op de analyse van fosfopeptiden. Lokalisatie van post-translationele modificaties zoals glycosylering en fosforylering is uitdagend omdat het de aanwezigheid van plaats-bepalende fragment ionen vereist. Zoals beschreven in hoofdstuk 3 genereert EThcD een dubbele serie ionen met als gevolg toename van de dekking van de totale peptide sequentie. Hier laten we zien dat EThcD bijzonder geschikt is voor de analyse van fosfopeptiden. We analyseerden (i) synthetische fosfopeptiden en (ii) Ti-IMAC-verrijkte fosfopeptiden en tonen het potentieel van EThcD voor fosforylering site lokalisatie. Onze gegevens tonen dat EThcD ondubbelzinnige lokalisatie van de fosforylering site aanzienlijk vergemakkelijkt voor 95% van alle fosforyleringen.

In **hoofdstuk 5** beschrijven we een nieuwe instrumentele setup om ETD reacties uit te voeren op een hybride LTQ-Orbitrap. ETD wordt vaak uitgevoerd in iontraps door gebruik te maken van RF velden die efficiënte elektronenoverdracht reacties mogelijk maken. In dit hoofdstuk onderzoeken we de toepassing van ETD in de HCD botsing cel van een Orbitrap Velos instrument, door een statische DC gradiënt axiaal aan te brengen. Dit gradiënt maakt het gelijktijdig driedimensionale vangen van kationen en anionen mogelijk, onafhankelijk van lading, wat leidt tot initiatie van elektronoverdracht reacties in het midden van de HCD cel waar tegengesteld geladen ionen overlappen. We evalueren deze toepassing voor een aantal tryptische peptiden alsmede voor de top-down sequentie-analyse van ubiquitin. Uit onze gegevens blijkt dat het uitvoeren van ETD in de HCD cel tot soortgelijke fragmentatie leidt als iontrap-ETD maar verdere optimalisatie vereist om de prestaties van iontrap-ETD te evenaren.

In **hoofdstuk 6** vergelijken we drie van de momenteel meest populaire methoden voor kwantificering van eiwitexpressie m.b.v. MS: metabolisch labelen, chemische labelen in combinatie met MS-niveau kwantificatie en isobarisch labelen. Wij vergelijken de prestaties in grootschalige kwantitatieve proteomics experimenten. Onze gegevens tonen aan dat alle drie de methoden een vergelijkbare diepte kunnen bereiken in aantal geïdentificeerde eiwitten met behulp van een klassieke (MS2 gebaseerde) shotgun benadering. SILAC en chemisch labelen met MS kwantificering testen vergelijkbaar op precisie van de kwantificatie. We laten zien dat isobarisch labelen met kwantificering op MS2 niveau zwaar wordt beïnvloedt door co-isolatie, wat precisie en nauwkeurigheid compromitteert.

Dit probleem kan gedeeltelijk worden opgevangen door peptide kwantificatie gebaseerd op MS3 spectra, maar dit gaat ten koste van een aanzienlijke daling in aantal gekwantificeerde eiwitten. We concluderen dat momenteel kwantificering op MS-niveau de voorkeur heeft voor nauwkeurige eiwit kwantificering.

In **hoofdstuk 7** analyseren we endogene neuropeptiden expressie in hypothalamus en striatum, in een rat model voor dieet-geïnduceerde zwaarlijvigheid. Neuropeptiden zijn belangrijke spelers in de regeling van voedselinname. Hier analyseren we extracten uit twee verschillende hersengebieden. Op basis van de resultaten uit hoofdstuk 2 hebben we HCD en ETD fragmentatie gebruikt om de algehele peptiden identificatie te verhogen. In totaal hebben we meer dan 1700 unieke endogene peptiden geïdentificeerd, waaronder vrijwel alle neuropeptiden waarvan bekend is dat zij betrokken zijn bij de regulatie van voedselinname. De kwantitatieve gegevens tonen differentiële processing van endogene peptiden die zijn afgeleid van het zelfde precursor eiwit. Bovendien vonden we vergelijkbare toename van orexigene en anorexigenic peptiden wat mogelijk wijst op tegenstrijdige signalering bij ratten gevoed met een hoog vet/hoog sucrose dieet.

## FUTURE OUTLOOK

Over the past years, bottom-up mass spectrometry-based proteomic technologies have evolved to facilitate routine quantitative analysis of proteomes. All stages of the most commonly employed proteomic workflow, based on the enzymatic digestion of proteins and subsequent identification of the generated peptides by LC-MS/MS, have been extensively optimized and refined. Recent advances in pre-lysis sample fractionation [1, 2] sample preparation [3, 4], peptide fractionation [5, 6], LC separation [7, 8], and MS instrumentation [9-11] enable researchers to interrogate cellular proteomes to a depth up to 10,000 proteins from mammalian systems [12, 13].

Interestingly, the fundamental principle of the bottom-up approach has largely remained unchanged. Information on protein identity and abundance is inferred from the identification and quantification of peptides which are typically derived from a tryptic digest. MS/MS-based peptide sequencing usually involves collision-induced dissociation in the gas-phase. Tryptic peptides are short and typically sequester two protons, which makes them an ideal target class for sequencing by means of collisional activation. The development of electron transfer dissociation as an alternative fragmentation technique, despite its numerous advantages, has not changed this dogma yet. Strict adherence to the conventional bottom-up approach can largely be attributed to its efficiency and robustness, that have facilitated the rapid development of MS-based proteomics toward establishment as a high-throughput technology. Another reason could be that the proteomics community in the last ten years largely focused on increasing numbers; for instance, increasing the number of peptides identified, the number of proteins quantified, the number of phosphorylation sites localized. Striving for unbiased comprehensiveness, sensitivity and coverage, most technology developments in MS-based proteomics were made with the ultimate goal to get on par with high-throughput next-generation sequencing technologies [14] in terms of speed, sensitivity and level of completeness.

Despite the fact that most advances in bottom-up MS-based proteomics were largely achieved without a major contribution of ETD, I envision that ETD and other alternative fragmentation techniques will play an increasingly important role in the near future, arising from the inherent limitations of the current standard workflow in bottom-up proteomics. The advantages of ETD for the analysis of PTMs [15-17] and its complementarity to CID [18, 19] have been extensively reviewed elsewhere. Here, I will focus on describing a perspective of ETD toward increased proteome coverage and for advanced protein isoform characterization.

**Proteome coverage**

The claim of being able to analyze "complete proteomes" by bottom-up MS-based proteomics has become popular in recent years [20-22] and led to the assumption that accurate and complete characterization of all expressed proteins at a given time point is feasible [23]. However, this statement should be critically scrutinized for the following reason: identification of peptides that cover parts of all gene products of a given sample does not necessarily imply that the whole proteome is covered. More specifically, de Godoy et al. reported on unbiased proteome coverage across all protein classes and subcellular compartments including membrane proteins in yeast [22]. However, the average protein sequence coverage was 38%, which is a typical value in large-scale proteomics experiments. "Complete" proteome analysis should include observation and characterization of every single amino acid of each protein. Obtaining full proteome coverage demands by definition 100% protein sequence coverage and is crucial for PTM mapping and the detection of SNPs and other RNA editing events [24]. Achieving this goal is challenging, not just because of the large dynamic range of peptide mixtures [25]. The use of enzymes limits the theoretically observable part of the proteome. Trypsin is the preferred protease in bottom-up proteomics because of its cleavage specificity, ease of use and sequencing-friendly behaviour in the mass spectrometer. Yet, 56% of all tryptic peptides consist of six or less amino acids. In contrast, 97% of all identified peptides in large-scale proteomics studies are between 7-35 amino acids long [26]. Shorter peptides can be sequenced but they rarely pass rigid filtering to reach an acceptable level of the estimated number of false-positive identifications. The fact that the majority of the generated protein fragments are discarded also limits attempts to observe and characterize "complete" sub-proteomes such as the phospho-proteome [27, 28], simply because many potential phosphosites are not detectable using trypsin.

Increasing the proteome coverage in bottom-up approaches can be achieved by combined use of multiple proteases. Enzymes such as GluC, LysC and chymotrypsin, respectively, generate peptides that cover different parts of the proteome. Using multiple enzymes can also prevent introduction of biases [29]. Moreover, peptides derived from enzymes other than trypsin are potentially very suitable for ETD-based sequencing. For instance, Heck and co-workers used four different proteases in combination with ETD and CID for near-complete mapping of yeast polymerase II and III [30]. Coon and co-workers utilized five different proteases for characterization of the yeast proteome [26]. Interestingly, they found that the majority of the peptide identifications were obtained from ETD fragmentation. Novel enzymes such as LysN [31] can generate peptides that exhibit ETD-favorable amino acid composition. Moreover, future engineering of proteases that generate longer peptides will further elevate the need for ETD-based sequencing methods. For example, Kelleher and co-workers introduced a protease, OmpT, that cleaves at dibasic sites and generates pep-

tides with an average mass of 6.3 kDa [32]. Again, the majority of these longer peptides were successfully sequenced by ETD. Moreover, generating longer peptides also reduces sample complexity, in turn allowing deeper proteome coverage [25]. Additionally, longer peptides increase the probability to observe multiple PTMs at once, further facilitating investigation of PTM crosstalk [33]. In essence, the current standard bottom-up proteomics workflow will soon be stretched to its limits in terms of total proteome coverage. Maximizing proteome coverage will probably be achieved using multiple enzymes and multiple fragmentation techniques, with a significant contribution of ETD.

**Protein isoforms**

Another field where ETD has a distinct potential for a major contribution is in the analysis of protein isoforms. Many current quantitative proteomic studies focus on protein expression profiling. This involves, for instance, comparisons between different treatments or over several time points. One should take in mind, that not all perturbations of the proteome are exclusively reflected by changes in initial gene expression, but can also trigger alterations in post-translational modifications, splicing patterns or other RNA editing events that lead to variations in the different protein isoforms. Thus, observed relative expression changes of certain protein groups is not necessarily translated into a different biological phenotype. Conversely, protein isoform variations can influence biological processes, which might not be observed in protein expression profiling if all isoforms are reported as one protein group. Almost all current data analysis approaches involve protein grouping, acknowledging the fact that most peptides cannot be used to distinguish protein isoforms. This limitation is inherently linked to the above described basic principle of bottom-up proteomics [34]. Observation of a few proteotypic peptides per protein easily provides information on the identity of the most abundant protein isoform of any given gene; however, the direct link to protein isoforms, splice variants and their distinct PTM pattern is lost [35]. Consequently, many genes are often represented by only one protein group in proteomic datasets, neglecting any splicing and PTM variants. For example, over 170 species of histone H3.2 were identified [36], highlighting the challenge arising from protein isoforms.

Interestingly, in the earlier days of proteomics many large-scale studies were conducted utilizing two-dimensional gel electrophoresis (2D-PAGE) for protein separation prior digestion, providing additional information on size and isoelectric point, and enabling differentiation of protein isoforms [37]. Starting with the rise of two-dimensional peptide separation approaches using liquid chromatography [38, 39], the 2D-PAGE approaches fell behind. The increasing popularity of the 2D-LC approaches can mainly be attributed to their robustness, the higher dynamic range, better reproducibility, higher throughput and advanced multiplex-

ing capabilities. Nevertheless, these advances come at the cost of information on protein isoforms.

Using multiple proteases can improve the proteome coverage and facilitate differentiation of protein isoforms, as described above. Another strategy to overcome this issue is the analysis of intact proteins [40, 41]. Assuming sufficient resolving power of the mass analyzer, different protein isoforms can be distinguished based on their absolute mass. Protein sequence analysis by means of tandem mass spectrometry was termed top-down proteomics. In top-down approaches the enzymatic cleavage of proteins is omitted and intact proteins are directly analyzed [42]. SDS-PAGE-based protein separation approaches are not suitable for subsequent intact protein analysis. Consequently, a variety of separation tools based on isoelectric focusing [43], HILIC [44], ion-exchange [45] and reversed-phase chromatography [46] have been developed. Over the last years, top-down proteomics gained momentum through the development of high-resolution FT-ICR and Orbitrap analyzers and electron capture and electron transfer dissociation, respectively. Electron-driven dissociation methods are advantageous for top-down proteomics because they do not cause loss of labile PTMs and basically induce peptide backbone fragmentation indifferent of length or amino acid composition. Moreover, ETD also facilitates analysis of proteins containing disulfide bonds. For example, Tsybin and co-workers employed ETD together with high-resolution Orbitrap detection for sequence analysis of an intact monoclonal antibody [47]. In contrast to bottom-up approaches, top-down proteomics is still in the early stages of development but ETD will certainly further advance the field [48]. Recent advances in MS instrumentation provide researchers nowadays with sophisticated tools for top-down analysis, enabling to tackle the issue of protein isoforms [42, 46]. This also led to the introduction of the term "proteoform" to 'designate all of the different molecular forms in which the protein product of a single gene can be found, including changes due to genetic variations, alternatively spliced RNA transcripts and post-translational modifications' [49].

In conclusion, the field of application of ETD as complementary fragmentation method for sequencing of larger peptides and intact proteins will further expand in the future. For example, within this thesis ETD was successfully employed for the analysis of endogenous peptides, and contributed to an improved data-dependent decision tree, aiming to subject peptides to the most suitable fragmentation technique. Future combination of bottom-up and top-down approaches using multiple enzymes and multiple fragmentation techniques will further maximize proteome coverage and provide valuable insight into biology [24, 50].

## REFERENCES


[1] R. Chakraborty, T.C. Hazen, D.C. Joyner, K. Kusel, M.E. Singer, J. Sitte, T. Torok, Use of immunomagnetic separation for the detection of Desulfovibrio vulgaris from environmental samples, J Microbiol Methods, 86 (2011) 204-209.

[2] S.M. Khan, B. Franke-Fayard, G.R. Mair, E. Lasonder, C.J. Janse, M. Mann, A.P. Waters, Proteome analysis of separated male and female gametocytes reveals novel sex-specific Plasmodium biology, Cell, 121 (2005) 675-687.

[3] J.R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis, Nature methods, 6 (2009) 359-362.

[4] J.R. Wisniewski, P. Ostasiewicz, M. Mann, High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers, Journal of proteome research, 10 (2011) 3040-3049.

[5] S. Di Palma, M.L. Hennrich, A.J. Heck, S. Mohammed, Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis, J Proteomics, 75 (2012) 3791-3813.

[6] S. Di Palma, S. Mohammed, A.J. Heck, ZIC-cHILIC as a fractionation method for sensitive and powerful shotgun proteomics, Nat Protoc, 7 (2012) 2041-2055.

[7] T. Kocher, R. Swart, K. Mechtler, Ultra-high-pressure RPLC hyphenated to an LTQ-Orbitrap Velos reveals a linear relation between peak capacity and number of identified peptides, Analytical chemistry, 83 (2011) 2699-2704.

[8] A. Cristobal, M.L. Hennrich, P. Giansanti, S.S. Goerdayal, A.J. Heck, S. Mohammed, In-house construction of a UHPLC system enabling the identification of over 4000 protein groups in a single analysis, The Analyst, 137 (2012) 3541-3548.

[9] A. Michalski, E. Damoc, O. Lange, E. Denisov, D. Nolting, M. Muller, R. Viner, J. Schwartz, P. Remes, M. Belford, J.J. Dunyach, J. Cox, S. Horning, M. Mann, A. Makarov, Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes, Molecular & Cellular Proteomics, 11 (2012).

[10] A. Michalski, E. Damoc, J.P. Hauschild, O. Lange, A. Wieghaus, A. Makarov, N. Nagaraj, J. Cox, M. Mann, S. Horning, Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer, Molecular & cellular proteomics : MCP, 10 (2011) M111 011015.

[11] G.L. Andrews, B.L. Simons, J.B. Young, A.M. Hawkridge, D.C. Muddiman, Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600), Analytical chemistry, 83 (2011) 5442-5446.

[12] J. Munoz, T.Y. Low, Y.J. Kok, A. Chin, C.K. Frese, V. Ding, A. Choo, A.J. Heck, The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells, Mol Syst Biol, 7 (2011) 550.

[13] T. Geiger, A. Wehner, C. Schaab, J. Cox, M. Mann, Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins, Molecular & cellular proteomics : MCP, 11 (2012) M111 014050.

[14] C. Howald, A. Tanzer, J. Chrast, F. Kokocinski, T. Derrien, N. Walters, J.M. Gonzalez, A. Frankish, B.L. Aken, T. Hourlier, J.H. Vogel, S. White, S. Searle, J. Harrow, T.J. Hubbard, R. Guigo, A. Reymond, Combining RT-PCR-seq and RNA-seq to catalog all genic elements encoded in the human genome, Genome Res, 22 (2012) 1698-1710.

[15] J. Wiesner, T. Premsler, A. Sickmann, Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications, Proteomics, 8 (2008) 4466-4483.

[16] G. Zauner, R.P. Kozak, R.A. Gardner, D.L. Fernandes, A.M. Deelder, M. Wuhrer, Protein O-glycosylation analysis, Biol Chem, 393 (2012) 687-708.

[17] M.S. Kim, A. Pandey, Electron transfer dissociation mass spectrometry in proteomics, Proteomics, 12 (2012) 530-542.

[18] R.A. Zubarev, A.R. Zubarev, M.M. Savitski, Electron capture/transfer versus collisionally activated/induced dissociations: Solo or duet?, J Am Soc Mass Spectr, 19 (2008) 753-761.

[19] J.J. Coon, Collisions or Electrons? Protein Sequence Analysis in the 21st Century, Analytical Chemistry, 81 (2009) 3208-3215.

[20] N. Nagaraj, J.R. Wisniewski, T. Geiger, J. Cox, M. Kircher, J. Kelso, S. Paabo, M. Mann, Deep proteome and transcriptome mapping of a human cancer cell line, Mol Syst Biol, 7 (2011).

[21] M. Beck, A. Schmidt, J. Malmstroem, M. Claassen, A. Ori, A. Szymborska, F. Herzog, O. Rinner, J. Ellenberg, R. Aebersold, The quantitative proteome of a human cell line, Mol Syst Biol, 7 (2011) 549.

[22] L.M. de Godoy, J.V. Olsen, J. Cox, M.L. Nielsen, N.C. Hubner, F. Frohlich, T.C. Walther, M. Mann, Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast, Nature, 455 (2008) 1251-1254.

[23] M. Mann, N.A. Kulak, N. Nagaraj, J. Cox, The coming age of complete, accurate, and ubiquitous proteomes, Mol Cell, 49 (2013) 583-590.

[24] B. Meyer, D.G. Papasotiriou, M. Karas, 100% protein sequence coverage: a modern form of surrealism in proteomics, Amino Acids, 41 (2011) 291-310.

[25] R.A. Zubarev, The challenge of the proteome dynamic range and its implications for in-depth proteomics, Proteomics, 13 (2013) 723-726.

[26] D.L. Swaney, C.D. Wenger, J.J. Coon, Value of using multiple proteases for large-scale mass spectrometry-based proteomics, Journal of proteome research, 9 (2010) 1323-1329.

[27] H. Zhou, S. Di Palma, C. Preisinger, M. Peng, A.N. Polat, A.J. Heck, S. Mohammed, Towards a comprehensive characterization of a human cancer cell phosphoproteome, Journal of proteome research, (2012).

[28] A. Lundby, A. Secher, K. Lage, N.B. Nordsborg, A. Dmytriyev, C. Lundby, J.V. Olsen, Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues, Nat Commun, 3 (2012) 876.

[29] M. Peng, N. Taouatas, S. Cappadona, B. van Breukelen, S. Mohammed, A. Scholten, A.J. Heck, Protease bias in absolute protein quantitation, Nature methods, 9 (2012) 524-525.

[30] S. Mohammed, K. Lorenzen, R. Kerkhoven, B. van Breukelen, A. Vannini, P. Cramer, A.J. Heck, Multiplexed proteomics mapping of yeast RNA polymerase II and III allows near-complete sequence coverage and reveals several novel phosphorylation sites, Analytical chemistry, 80 (2008) 3584-3592.

[31] N. Taouatas, M.M. Drugan, A.J. Heck, S. Mohammed, Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase, Nature methods, 5 (2008) 405-407.

[32] C. Wu, J.C. Tran, L. Zamdborg, K.R. Durbin, M. Li, D.R. Ahlf, B.P. Early, P.M. Thomas, J.V. Sweedler, N.L. Kelleher, A protease for 'middle-down' proteomics, Nature methods, 9 (2012) 822-824.

[33] V. van Noort, J. Seebacher, S. Bader, S. Mohammed, I. Vonkova, M.J. Betts, S. Kuhner, R. Kumar, T. Maier, M. O'Flaherty, V. Rybin, A. Schmeisky, E. Yus, J. Stulke, L. Serrano, R.B. Russell, A.J. Heck, P. Bork, A.C. Gavin, Cross-talk between phosphorylation and lysine acetylation in a genome-reduced

bacterium, Mol Syst Biol, 8 (2012) 571.

[34] M.W. Duncan, R. Aebersold, R.M. Caprioli, The pros and cons of peptide-centric proteomics, Nature biotechnology, 28 (2010) 659-664.

[35] A.I. Nesvizhskii, R. Aebersold, Interpretation of shotgun proteomic data: the protein inference problem, Molecular & cellular proteomics : MCP, 4 (2005) 1419-1440.

[36] B.A. Garcia, J.J. Pesavento, C.A. Mizzen, N.L. Kelleher, Pervasive combinatorial modification of histone H3 in human cells, Nature methods, 4 (2007) 487-489.

[37] T. Rabilloud, M. Chevallet, S. Luche, C. Lelong, Two-dimensional gel electrophoresis in proteomics: Past, present and future, J Proteomics, 73 (2010) 2064-2077.

[38] M.P. Washburn, D. Wolters, J.R. Yates, 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology, Nat Biotechnol, 19 (2001) 242-247.

[39] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates, 3rd, Direct analysis of protein complexes using mass spectrometry, Nature biotechnology, 17 (1999) 676-682.

[40] J.D. Tipton, J.C. Tran, A.D. Catherman, D.R. Ahlf, K.R. Durbin, N.L. Kelleher, Analysis of Intact Protein Isoforms by Mass Spectrometry, J Biol Chem, 286 (2011) 25451-25458.

[41] A.J. Heck, Native mass spectrometry: a bridge between interactomics and structural biology, Nature methods, 5 (2008) 927-933.

[42] W. Cui, H.W. Rohrs, M.L. Gross, Top-down mass spectrometry: recent developments, applications and perspectives, The Analyst, 136 (2011) 3854-3864.

[43] G. Weber, P. Bocek, Recent developments in preparative free flow isoelectric focusing, Electrophoresis, 19 (1998) 1649-1653.

[44] P.J. Boersema, S. Mohammed, A.J. Heck, Hydrophilic interaction liquid chromatography (HILIC) in proteomics, Anal Bioanal Chem, 391 (2008) 151-159.

[45] S. Sharma, D.C. Simpson, N. Tolic, N. Jaitly, A.M. Mayampurath, R.D. Smith, L. Pasa-Tolic, Proteomic profiling of intact proteins using WAX-RPLC 2-D separations and FTICR mass spectrometry, Journal of proteome research, 6 (2007) 602-610.

[46] J.C. Tran, L. Zamdborg, D.R. Ahlf, J.E. Lee, A.D. Catherman, K.R. Durbin, J.D. Tipton, A. Vellaichamy, J.F. Kellie, M.X. Li, C. Wu, S.M.M. Sweet, B.P. Early, N. Siuti, R.D. LeDuc, P.D. Compton, P.M. Thomas, N.L. Kelleher, Mapping intact protein isoforms in discovery mode using top-down proteomics, Nature, 480 (2011) 254-U141.

[47] L. Fornelli, E. Damoc, P.M. Thomas, N.L. Kelleher, K. Aizikov, E. Denisov, A. Makarov, Y.O. Tsybin, Top-down analysis of monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS, Molecular & cellular proteomics : MCP, (2012).

[48] B.A. Garcia, What does the future hold for Top Down mass spectrometry?, J Am Soc Mass Spectr, 21 (2010) 193-202.

[49] L.M. Smith, N.L. Kelleher, Proteoform: a single term describing protein complexity, Nature methods, 10 (2013) 186-187.

[50] B.T. Chait, Chemistry. Mass spectrometry: bottom-up or top-down?, Science, 314 (2006) 65-66.

## PUBLICATIONS

**Frese CK***, van den Toorn H*, Heck AJ and Mohammed S. Quantitative Proteomics for Differential Protein Expression Profiling. *Proteomics for Biological Discovery*, 2nd Edition, WILEY-VCH, in press

**Frese CK**, Nolting D, Altelaar AF, Griep-Raming J, Mohammed S and Heck AJ. Characterization of Electron Transfer Dissociation in the Orbitrap Velos HCD Cell. *Journal of the American Society for Mass Spectrometry,* 2013, Apr 20

**Frese CK**, Boender AJ, Mohammed S, Heck AJ, Roger AH and Altelaar AF. Profiling of Diet-Induced Neuropeptide Changes in Rat Brain by Quantitative Mass Spectrometry. *Analytical Chemistry*, 2013, Apr 13

Jones EA, Schmitz N, Waajer CJF, **Frese CK**, van Remoortere A, van Zeijl RJM, Heck AJ, Hogendoorn PCW, Deelder AM, Altelaar AFM, Boveé JVMG and McDonnell LA. Imaging Mass Spectrometry Based Molecular Histology Differentiates Microscopically Identical and Heterogeneous Tumors. *Journal of Proteome Research*, 2013, Mar 25

**Frese CK**, Zhou H, Taus T, Altelaar AF, Mechtler K, Heck AJ and Mohammed S. Unambiguous Phosphosite Localization using Electron-Transfer / Higher-Energy Collision Dissociation (EThcD). *Journal of Proteome Research*, 2013, Mar 1, 12(3):1520-5

**Frese CK**, Altelaar AF, van den Toorn H, Nolting D, Griep-Raming J, Heck AJ and Mohammed S. Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Analytical Chemistry*, 2012, Nov 20, 84(22):9668-73

Altelaar AF*, **Frese CK***, Preisinger C, Hennrich ML, Schram AW, Timmers M, Heck AJ and Mohammed S. Benchmarking stable isotope labeling based quantitative proteomics. *Journal of Proteomics*, 2012, Oct 22

Fränzel B, Penkova M, **Frese C,** Metzler-Nolte N and Wolters D. Escherichia coli exhibits a membrane-related response to a small arginine- and tryptophan-rich antimicrobial peptide. *Proteomics*, 2012, Aug, 12(14):2319-30

Munoz J, Low TY, Kok Y, Chin A, **Frese CK**, Ding V, Choo A and Heck AJ. The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells. *Molecular Systems Biology*, 2011, Nov 22, 7:550

**Frese CK**, Altelaar AF, Hennrich ML, Nolting D, Zeller M, Griep-Raming J, Heck AJ and Mohammed S. Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *Journal of Proteome Research*, 2011, May 6, 10(5):2377-88

Fränzel B, **Frese C**, Penkova M, Metzler-Nolte N, Bandow JE and Wolters DA. Corynebacterium glutamicum exhibits a membrane-related response to a small ferrocene-conjugated antimicrobial peptide. *Journal of Biological Inorganic Chemistry*, 2010, Nov, 15(8):1293-303

*These authors contributed equally.

## CURRICULUM VITAE

Christian was born on 21 December 1983 in Marsberg, North Rhine-Westphalia, Germany. In 2004 he took up studies in Biochemistry at the Ruhr-Universität in Bochum. He obtained his B.Sc. degree in 2007. During his thesis work at the Medical Proteome Center he worked in the field of Alzheimer's disease under supervision of Prof. Dr. Katrin Marcus on the establishment of immunofluorescence in cells with silenced APP/APLP2 expression. In 2009 he graduated with a M.Sc. in Biochemistry with a major in biochemistry of the nervous system. In the course of his Master thesis project at the department of analytical chemistry Christian worked under supervision of Dr. Dirk Wolters on investigating changes of the bacterial membrane proteome in response to a novel antibiotic compound. During this final year at the Ruhr-Universität in Bochum he was introduced to mass spectrometry-based proteomics and he decided to pursue his career in this direction. After graduation in 2009 Christian started working as doctoral candidate in the Biomolecular Mass Spectrometry & Proteomics Group under supervision of Prof. Dr. Albert Heck. The research mainly focused on MS technology development in close collaboration with Thermo Fisher Scientific, Bremen. The results of his work are presented in this thesis.

## 2011

## 2013

## ABBREVIATIONS

2D-PAGE, Two-dimensional polyacrylamide gel electrophoresis
α-MSH, Alpha-melanocyte stimulating hormone
AC, Alternating current
AGC, Automatic gain control
AgRP, Agouti-related protein
AMT, Accurate mass and time
AP-MS, Affinity purification based mass spectrometry
ARC, Arcuate nucleus
BCA, Bicinchoninic acid
CID, Collision induced dissociation
ECD, Electron capture dissociation
ETD, Electron transfer dissociation
FDR, False discovery rate
FT, Orbitrap readout
FT-ICR, Fourier-transform ion cyclotron resonance
Ghrelin, Growth hormone release inducing
HCD, Higher energy collision dissociation
ICAT, Isotope coded affitnity tags
iTRAQ, Isobaric tag for relative and absolute quantification
LC, Liquid chromatography
MS, Mass spectrometry
MS/MS, Tandem mass spectrometry
NCE, Normalized collision energy
NMR, Nuclear magnetic resonance
NPY, Neuropeptide Y
PSM, Peptide to spectrum match
PTM, Post-translational modification
PVN, Paraventricular nucleus
RF, Radio frequency
SILAC, Stable isotope labeling by amino acids in cell culture
QqQ, Triple quadrupole
QqTOF, Quadrupole time-of-flight
Q-TRAP, Quadrupole linear ion trap
SCX, Strong cation exchange
SILAC, Stable isotope labeling by amino acids in cell culture
SON, Supraoptic nucleus
TMT, Tandem mass tag
TOF, Time-of-flight