Lys-N: A versatile enzyme for proteomics

This thesis is dedicated to my dear friend Shanna Johansen.
You were a wonderful person and friend.
I will always miss you my friend.

I'll meet you further on up the road!

ISBN: 978-90-393-5488-9

Printed by Proefschriftmaken, Oisterwijk, 2010

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

Lys-N: A versatile enzyme for proteomics

Lys-N: een voor gebruik in proteomics veelzijdig enzym

(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. J. C. Stoof, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op maandag 20 december 2010 des ochtends te 10.30 uur

door

Nadia Taouatas

Geboren op 18 mei 1978 te Vejle, Denemarken

Promoter: Prof. dr. A.J.R. Heck

Co-promoter: Dr. S. Mohammed

Table of Contents

Chapter 1

Introduction

I. Mass spectrometry based proteomics

The proteome can generally be described as the protein complement of the genome, in other terms, all the proteins expressed in a given cell type at a given time. The expression proteome was coined in 1994 to describe all of the proteins in a particular cell, tissue, or organism. Naturally, the expression "proteomics" was subsequently introduced and refers to a large-scale study of the full set of proteins encoded by the genome. Mass spectrometry is one of the most important tools in proteomic based studies.

The fundamental strategies used for protein identification and characterization in mass spectrometry based proteomics are generally divided into two sub-groups: the "bottom up" and the "top down" approach. In the "bottom up" approach purified proteins, or complex protein mixtures, are subjected to proteolytic digestion prior to mass spectrometry analysis. In the "top down" approach, intact proteins are directly introduced into the mass spectrometer for analysis. Both methods have been used for a variety of studies including: structural studies, protein-protein interactions, discovery of disease biomarkers and drug targets, identification of membrane proteins and posttranslational modifications.

The study of posttranslational modifications in particular has been favored by many research groups, as many proteins in the cell are modified following translation. In fact, posttranslational modification of proteins is one of the most important mechanisms for regulating protein function and proteolysis. There are many types of post translational modifications such as phosphorylation, sulfation, acetylation, methylation, glycosylation. Mass spectrometry analysis is, as mentioned, often used for the characterization of posttranslational modifications in proteins. The two strategies, "bottom up" and "top down", have both been applied for such studies. The advantage of the "top down" approach is that the protein remains intact and thereby retains any information about combinatorial posttranslational modifications. The method has gradually evolved and can now be applied for the analysis of a larger number of intact proteins although, up till now, it still remains a method for mainly analyzing single purified proteins. The "bottom up" proteomics approach is more widely used, both for posttranslational modifications studies, and for the identification and characterization of proteins in general and has therefore become a more established technique.

Generally, when the "bottom up" method is applied for the analysis of a complex protein mixture, it is referred to as "shotgun proteomics", which includes the four following steps: proteolytic digestion, peptide separation, peptide fragmentation in the mass spectrometer and

lastly, data analysis (database searching). This approach has been further developed by using multidimensional protein identification technology (MudPIT), which uses multidimensional liquid chromatography (LC/LC), tandem mass spectrometry and database-searching algorithms.

The four steps in "shotgun proteomics" will all be described individually in the following sections. Part I will introduce mass spectrometers and tandem mass spectrometers commonly used today in proteomics. Even though many types of mass spectrometers exist, there are certain components characteristic for all types. In brief, a mass spectrometer generally consists of three distinct parts: an ionization source that converts molecules into gaseous ions, a mass analyzer to separate the generated ions according to their mass-to-charge ratio (m/z), and a detector that converts the abundance of separated ions into a digital signal (finally a mass spectrum is generated). As ions are very reactive and short-lived, their formation and manipulation inside the instrument is performed in a vacuum. Each section of the mass spectrometer will be described below. Peptide fragmentation using tandem mass spectrometry analysis has a very important role in proteomics studies and has made it possible to study complex biological samples. The simplest form of tandem mass spectrometry combines two mass analyzers, where the first one is used to select a single precursor ion (with a particular mass to charge (m/z) value). The selected precursor ions will pass though a collision cell where the ions are activated and fragmented into product ions. The product ions are passed to a second analyzer which generates a spectrum. Such spectra consist of product ions from the selected precursor ion and the information in the spectrum can be used to perform a database search for the identification of peptides and proteins. A general description about peptide fragmentation as well as the three fragmentation techniques, collision induced/activated dissociation (CID/CAD), electron capture dissociation (ECD) and electron transfer dissociation (ETD), will be included in the discussion below.

Proteolytic enzymes are also an important factor when analyzing a complex protein mixture by mass spectrometry. Many different enzymes have been studied for the generation of proteolytic peptides for mass spectrometry analysis. The type of enzyme used, for a specific proteomic analysis, depends on several aspects: the sample type, solvents used, and fragmentation method. Part II will describe different proteolytic enzymes and their role in mass spectrometry based proteomics.

The "shotgun" proteomics approach is, as briefly described above, based on using multidimensional liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) for the analysis of complex protein mixtures. Its use for large-scale proteome profiling is now

commonly applied for the analysis of biological samples. The importance of separation techniques will be discussed in Part III.

I.I. Ionization techniques

Mass spectrometry (MS) has evolved into a mature technique for the analysis of protein and protein complexes. This is partly as a result of the development of two soft ionization methods, matrix assisted laser desorption ionization (MALDI) and electrospray ionization (ESI). These ionization methods are commonly used to transfer molecules into gas-phase as ions, from either a solution or solid state form, prior to MS analysis. However, despite the popularity of these methods in a variety of biological and analytical chemistry studies, a complete understanding of the theoretical aspects of the ionization methods has yet to be elucidated.

Matrix assisted laser desorption ionization (MALDI)

Matrix assisted laser desorption ionization (MALDI) was first introduced by Karas and Hillenkamp and, at the same time, Tanaka, in the late eighties[1-3]. In the following years MALDI developed into a widespread analytical tool for peptides, proteins, and other biomolecules such as oligonucleotides, carbohydrates and lipids.

For a typical MALDI analysis, the analyte is dissolved and mixed with an excess of a UV absorbing matrix material, usually aromatic compounds, prior to laser irradiation (Figure 1). The mixing is generally performed prior to co-crystallizing the matrix and the analyte to ensure that the analyte molecules are distributed throughout the crystallized matrix. The choice of matrix material is important as it can affect peak intensities, fragmentation of the analyte, relative abundance and the type of analyte ions detected. Extensive studies have been made over the years to define the most appropriate matrix for MALDI analysis, where one of the few agreements is that the matrix material should have a strong absorption coefficient at the wavelength of the laser used[4-7]. The absorption should occur without passing excessive internal energy to the analyte during the desorption/ionization process. This is ensured as the matrix contains chromophores that absorb the laser light and since the matrix is in excess it will, in principle, absorb the majority of the laser radiation. Several matrices and combinations of matrices have successfully been developed although only a few are used on a

regular basis[8-10]. When analyzing peptides and 'small' proteins (<5,000 Da), α-cyano-4-hydroxycinnamic acid (CHCA) is the matrix of choice[11, 12], whereas larger proteins are often analyzed using sinapinic acid (SA) [13, 14]. The use of 3-amino-4 hydroxybenzoic acid and 2,5-dihydroxybenzoic acid (DHB) are recommended for the analysis of oligosaccharides[15] and phosphopeptides[16].

Another important parameter in the MALDI process is the laser wavelength and laser type used. A range of wavelengths and alternative lasers have been studied to find the most suitable for a specific sample or matrix used in the MALDI process[17, 18]. Infrared light has, for example, been used for irradiation[19], where the matrix absorbs the laser energy via vibrational excitation. This has, however, been used to a lesser extent as commercial instruments often have nitrogen lasers implemented. Thus, the nitrogen laser is most common and generates light with a wavelength of 337nm[20, 21]. Laser irradiation excites the matrix and induces rapid heating of the sample at the surface, the combination of which causes desorption of the sample and eventually leads to the formation of gaseous protonated sample molecules. However, the exact mechanism for ionization is still being investigated[2, 22-25]. Understanding the ionization process is crucial to be able to improve ion yields, improve the control of fragmentation, to control charge state, and to identify new compounds.

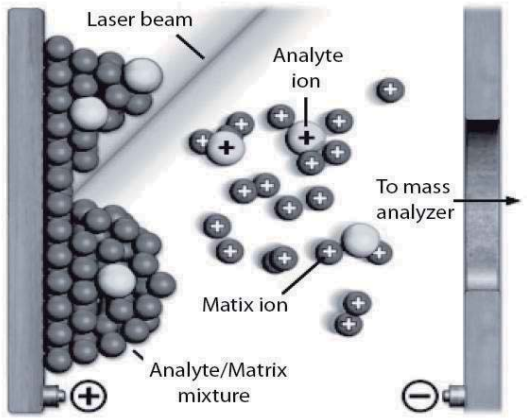

Figure 1: Illustration of the MALDI process. Sample is mixed with an excess of UV-absorbing matrix material and irradiated with a laser. Matrix molecules, upon absorption of UV light, aid the sublimation and analyte gas phase ion generation process (Figure adapted from "Tools of the Trade", Matrix Assisted Laser Desorption Ionization (MALDI), from Magnet Lab, National High Magnetic Field Laboratory).

The main advantages of using MALDI as an ionization technique are easy sample preparation and a fast analysis. Additionally, MALDI generates predominantly single-charged ions which results in spectra that are relatively easy to interpret.

Electrospray ionization (ESI)

The concept of electrospray has been known for more than 100 years[26], although the importance of the idea was not fully understood before 1937, where Chapman showed that electrified liquids can be used to spray ions and obtain mobility spectra[27]. In 1968, revolutionary experiments performed by Dole *et al* demonstrated the use of electrospray as a way to generate ions of macromolecules and thereby formed the technique electrospray ionization (ESI)[28]. Later, John Fenn demonstrated the use of ESI for the ionization of high molecular weight compounds and their subsequent analysis by mass spectrometry. The work of Fenn led the mass spectrometric community to recognize the power of electrospray, for which he later received a Nobel Prize[29]. Today, MALDI and ESI are the methods of choice in mass spectrometric analyses of biomolecules.

ESI sample preparation usually starts with mixing a given sample with an acidified volatile solvent prior to introducing the mixture into a narrow capillary. High voltage (kV) is applied to the capillary tip, creating an electric field for charge separation at the surface of the liquid. Charges are polarized at the surface of the emerging liquid close to the needle tip, by the applied electric field. When the charged liquid first exits the capillary tip, it will briefly form a cone shape (also known as a Taylor cone) before the droplets burst away from each other into a fine spray of charged droplets, due to coulombic forces (see Figure 2).

The solvent within each droplet gradually evaporates, typically with the assistance of heated nitrogen gas (known as the drying gas). The droplets will reduce in size as they drift towards the end wall of the ionization chamber towards the counter electrode, forcing ions with identical charge polarity closer together. As a result, the charge density at the surface of the droplets will increase. The droplet will, however, remain intact as long as its surface tension is stronger than the droplet's internal charge (coulomb) repulsion. The droplet diameter will continue to decrease and eventually result in the coulombic repulsion being equal to the surface tension of the solution, referred to as the Rayleigh limit. The charged droplets will undergo coulombic fission, as the instability will cause the droplets to explode and generate charged daughter droplets. The process repeats itself, each time generating smaller droplets until eventually gas phase ions are formed, free from solvents. The exact mechanism of ESI ionization is (as for MALDI) not fully understood and is still under debate. Two theories have been described to explain the final generation of gas-phase ions, the Charged Residue Model (CRM)[28] and the Ion Evaporation Model (IEM)[30].

In proteomics, electrospray is often operated at very low flow rates (μL to nL per minute)[31]. The low flow rate combined with a small needle diameter was found to generate a spray with significantly smaller droplets. Using this procedure was found to improve ionization efficiency and was first described in 1994, where the term micro-electrospray was originally used to describe the technique[32, 33]. Unlike MALDI, ESI generates multiply charges ions, which can result in more complicated mass spectra. An advantage of ESI is that that the molecules will remain intact during the ionization procedure and will not be fragmented; ESI is therefore referred to as a "soft" ionization technique. Furthermore, electrospray ionization allows easy coupling of liquid chromatography (see Section III) with mass spectrometry[34].

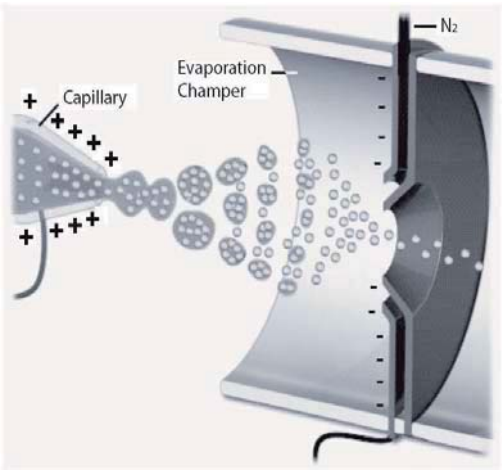

Figure 2: The electrospray process. Voltage is applied to a conductive capillary, which contains a polar liquid. The application of an electric field will cause liquid to polarize and be repulsed. The resulting Taylor cone at the capillary tip will emit highly charged droplets, creating an aerosol of charged droplets. The charged droplets will evaporate and undergo coulombic fission. The evaporation and fission process repeats itself until gas phase ions are finally formed (Figure adapted from "Tools of the Trade", Electrospray Ionization, from Magnet Lab, National High Magnetic Field Laboratory).

I.II Mass analyzers

After the ions are generated, they are introduced into a mass analyzer where they are separated according to their mass-to-charge ratio (*m*/*z*). The selection of a specific mass analyzer depends upon the resolution, mass range, scan rate, or detection limits required for a given application. Analyzers are generally divided into two subclasses of analyzers, scanning and non-scanning devices. Scanning analyzers include quadrupoles, which only transmit ions of a selected mass-to-charge ratio to the detector at a given time. When a certain mass-to-charge ratio is selected, ions at other mass-to-charge ratio are removed. Non-scanning analyzers, on the other hand, include Time-of-Flight, ion cyclotron resonance, and orbitrap

mass spectrometers. These instruments allow for the collection of an entire mass spectrum from a single pulse of ions.

Mass analyzers are often combined to create hybrid instruments and used for tandem mass spectrometry (MS/MS), which is becoming an increasingly widespread technology (Section I.III).

Quadrupole

The quadrupole (Q) mass analyzer was developed in parallel with the quadrupole ion trap by Wolfgang Paul, who received a Nobel Prize in 1989 for his work. The quadrupole analyzer is a relatively simple analyzer that consists of four parallel metal rods to which radio frequency (RF) and direct current (DC) voltages are applied (See Figure 3). Ions are guided from the ion source into the quadrupole analyzer using lenses with applied voltages. Voltages applied to the rods cause the entered ions to move in an oscillating trajectory between the rods. Only ions with a certain *m*/*z* ratio will, at a given DC and AC voltage, be able to pass through the quadrupole mass filter. The selected ions will then reach the detector while all other ions will deviate from the original path due to an unstable trajectory and collide with the rods.

Quadrupole analyzers are often used in tandem mass spectrometers for MS/MS studies. Quadrupole tandem mass spectrometers can often be found in a triple quadrupole mass spectrometer configuration[35] as well as a version where the last quadrupole is replaced with a Time-Of-Flight mass analyzer (Q-TOF)[36]. In the Q-TOF mass spectrometer, the quadrupole functions as an ion guide (MS mode) or as an ion filter (MS/MS mode), with the TOF analyzer actually being responsible for the mass measurement of the ions[37 38]. Additionally, the Q-TOF instrument is able to perform both collision induced dissociation and electron transfer dissociation (ETD) tandem mass spectrometry (MS/MS)[39, 40].

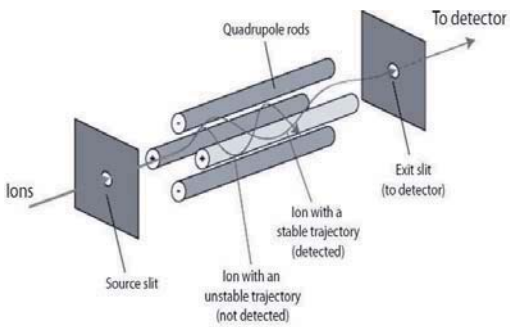

Figure 3: Schematic illustration of a quadrupole mass analyzer (Figure adapted from University of Bristol, NERC Life Sciences Mass Spectrometry Facility, Gas Chromatography Mass Spectrometry (GC/MS)).

Quadrupole ion trap

Different types of ion traps have been developed during the years, including the Penning trap[41] (used in an Fourier transform ion cyclotron resonance mass spectrometer), Kingdon trap[42] (Orbitrap mass spectrometer), linear quadrupole ion trap (two dimensional, 2D), and a Paul trap[43] (3D quadrupole ion trap). The 3D quadrupole ion trap consists of three hyberbolic electrodes, a ring electrode, and an entrance and exit endcap electrode, with the ring electrode positioned between the two endcap electrodes (see Figure 4). Generated ions are focused via electrostatic lenses into the ion trap using the inlet of the entrance endcap electrode. A gating lens at the entrance of the endcap electrode changes from positive to negative voltages to repel or attract ions. The pulsing of the ion gate, from positive to negative voltage, differs from analyzers such as the quadrupole, where ions are constantly streamed into the mass analyzer. Before the actual analysis is performed the optimal number of ions needs to be introduced. Thus an ion counting exercise is performed before each scan. Ions are loaded into the trap after which the ions are simply passed to the detector in order to determine the ion current. Using the calculated current, the required fill time for the ion trap can be determined and applied for the subsequent scan. Using such a method maximizes signal intensities and circumvents the problem of overfilling the ion trap, also referred to as "space-charge" effects. The limited trapping capacity caused by "space-charge" usually results from a high ion density, where ion-ion interactions cause distortion of the quadrupolar field. Once inside, the ions are trapped and confined to a stable oscillating motion as a RF-voltage is applied to the electrodes. The voltage that is applied is sufficient to ensure that ions over the entire m/z range of interest are trapped within the generated quadrupole field. The motion of an ion is related to its m/z ratio and the applied RF voltage. The ion trap is usually filled with helium gas (1 mTorr) to help focusing the ions toward the centre of the trap. Collision with helium dampens the kinetic energy of the ions, thereby changing their trajectories toward the center.
Trapped ions are ejected by gradually changing the electrode voltages, which results in an unstable trajectory of ions of a specific m/z value. Mass analysis can, in principle, occur in two ways, by mass selective resonance detection or by mass selective storage[44]. Using resonance detection, a voltage of frequency is applied to the endcap electrodes, on top of the existent ion trapping voltage between the endcap electrodes and ring electrode. Trapped ions will start to resonate if their m/z value results in their fundamental axial frequency being equal to the field frequency. The destabilized ions are ejected through the exit end cap to reach the detector for the generation of a mass spectrum. The m/z value of the detected ions is

defined by the resonant frequency of the excited ions and the quadrupole field voltages. In mass selective storage, only ions with a single *m*/*z* ratio will be trapped, prior to pulse-ejecting the ions from the trap. Here, the applied trapping voltage has to be selected in such a way that the range of *m*/*z* values to be trapped is as narrow as possible, finally resulting in the trapping of only ions of a particular *m*/*z* ratio of interest. The trapped and stored ions of a particular *m*/*z* ratio are ejected by applying a voltage pulse between the endcap electrodes, which result in the ions being forced out through the endcap electrode towards the detector.
Additionally, selected ions can be isolated and retained in the trap by ejecting all other ions using resonant expulsion, with the intention of fragmenting the isolated precursor ions with helium gas. The fragment ions are ejected from low to high *m*/*z* by altering amplitudes of the fundamental RF potential. This will result in sequentially unstable ion trajectories. The isolation and fragmentation step can be repeated a certain number of times, depending on fragmentation efficiency, number of ions and trapping efficiency of the specific instrument in use. The fragment ions are then ejected for scanning as described above in order to determine their *m*/*z* values.

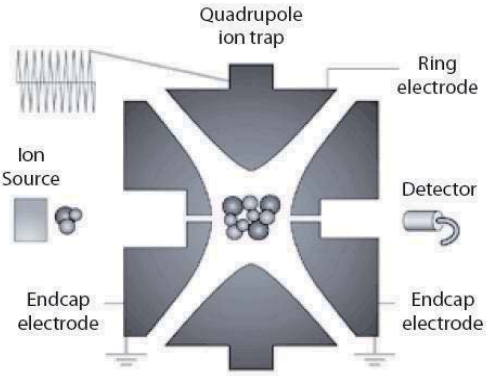

Figure 4: Schematic overview of the quadruple ion trap. Ions in the trap maintain a stable trajectory due to the application of a radio frequency voltage to the ring electrode. Mass analysis is achieved by creating unstable ion trajectories in a mass-selective manner (Figure adapted from Nat Rev Drug Discov. 2003, 2(2), 140-150).

A linear quadrupole ion trap consists of four rods (modified differently depending on the design) where ions are trapped axially, comparable to the central trapping used in 3D ion traps. Ions are trapped radially by applying an RF field and axially using static potential on the end electrodes. The difference between a linear quadrupole ion trap and a Paul quadrupole ion trap is a zero-field (no quadrupole field) node along the whole centre line instead of only a single point for the former[45, 46]. This design results in an increase in ion

storage capacity and higher injection efficiencies which again leads to a higher sensitivity and mass accuracy.

Some advantages of an ion trap instrument are the very compact size, relative low price, and the ability of the analyzer to trap and accumulate ions to improve the signal-to-noise ratio of an analysis. Another advantage is, as mentioned, the ability to perform multiple stages of MS operation (MS^n) without having multiple analyzers[47-49]. Quadrupole ion trap instruments are able to use alternative fragmentation techniques such as electron transfer dissociation (ETD),[50](Section I.III). However, ion trap mass spectrometers do have a general problem, which is referred to as low mass cutoff[51, 52]. The low mass cutoff problem occurs when the ion trap is unable to trap ions of low m/z ratios when performing MS/MS experiments. Linear ion trap analyzers are often used in hybrid instruments where they are combined with TOF, quadrupole, Orbitrap or Fourier Transform Ion Cyclotron Resonance instruments (as explained in the following sections) [53-56].

Time-of-Flight (TOF)

The principle of a linear Time-of-Flight (TOF) based mass spectrometer was first described in 1946 by Stephens[57]. The design was subsequently improved by Wiley and McLaren[58], which led to the first commercial TOF instrument. A linear TOF mass spectrometer is based on a simple mass separation principle and is often considered as one of the simplest mass analyzers as it, in its simplest form, consists only of an ion-accelerating region, a flight tube, and a detector. After the generation of ions, the produced ion package is accelerated (dependent on the number of charges) into a flight tube and their mass determined by measuring the time spent in the drift region (the flight tube) before reaching the detector. Ions in the acceleration zone will in theory experience the same potential difference and have the same kinetic energy when entering the flight tube. The velocity differences of the ions travelling in the flight tube will therefore depend upon the mass of the ions. The time the ions take to reach the detector at the end of the flight tube will then depend solely on their mass and their charge. The following equations are used to describe how the mass-to-charge ratios (m/z) of the ions relate to the flight times (t) in the flight tube:

An ion with mass *m* and total charge $q = ze$ has a kinetic energy, E_k:

$$\frac{mv^2}{2} = qV_s = zeV_s = E_k$$

The time it takes the ion to fly the distance (d) to the detector is given by: $t = d/v$

$$t^2 = \frac{m}{z}\left(\frac{d^2}{2\,V_s\,e}\right)$$

Where v is the velocity of the ion with mass *m* and charge *z*, Vs is the acceleration potential and d the length of the flight path to the detector. Even though the linear TOF analyzer is very fast in measuring ions to obtain a mass spectrum, the downsides of the instrument are poor resolution and low mass accuracy, plus there is no possibility for controlled fragmentation of ions for MS/MS analysis. To overcome the low resolution (and low mass accuracy) problem, resulting from flight time variations of ions of the same *m/z,* a number of improvements have been made to the TOF design, one of which is the "reflectron". The reflectron TOF instrument, first proposed by Mamyrin *et al*[59], is able to focus ions with the same m/z ratio but different kinetic energies. The reflectron usually consists of multiple rings and grids located at the end of the drift region, which create a decelerating/accelerating "reflecting" field. Ions entering the decelerating field will penetrate the field to a different extent, depending on their kinetic energy. As a result, ions with a high kinetic energy will penetrate the decelerating field further than ions with lower kinetic energy. This causes "faster" ions to reside longer within the reflecting field before they are sent back into the field free region of the flight tube where they will reach the detector, positioned at the end of the source, at a similar time to the "slower" low kinetic energy ions.

Both linear and reflector TOF instruments are ideally suited for pulsed ion sources and are therefore often, though not exclusively, combined with the MALDI ionization technique. The advantages of using a TOF mass spectrometer are a theoretically unlimited mass range, fast analysis, and a relatively low cost.

Even though tandem mass spectrometers usually consist of ion trap or quadrupole analyzers there are a number of instruments that employ TOF analyzers. One of these instruments is a TOF/TOF combination. The advantage of using TOF analyzers for MS/MS analysis is that

they are optimized for transmission of ions with high kinetic energy and collision of these ions results in high-energy fragments. Although a number of TOF/TOF configurations are available, they all are based on a similar design, Figure 5[60-64]. Here, a collision chamber is positioned between two TOF mass analyzers (a linear and a reflectron TOF analyzer), for fragmentation of precursor ions by collisional induced dissociation (CID). Fragmentation of ions using MALDI mass spectrometry often occurs via metastable decay where the ions' excessive internal energy or the collision with free gas cause dissociation immediately after the ions are generated. The decay of parent ions is referred to as post source decay (PSD)[65], as it occurs after the ions leave the source[66]. However, collisions in the high-energy collision cell can induce fragmentation of ions that do not decompose under normal PSD conditions. The precursor ions of interest are selected after the first TOF analyzer (linear TOF), which is only using a fraction of the normal TOF acceleration voltage (a few keV) to increase the flight time of the ions. The increase in flight time in the flight tube improves the ability of the time ion selector to isolate precursor ions. The time ion selector is a velocity selector that acts as a gate, which only allows for the transmission of precursor ions of the selected mass using its expected velocity[67, 68]. Selected precursor ions enter the high-energy collision cell where a gas is pulsed once for each cycle, to collide with the precursor ions. Prior to entering the second TOF analyzer (containing a reflectron), ions are accelerated to over 20keV in order to make the spread of kinetic energies of fragment ions insignificant relative to their total kinetic energies.

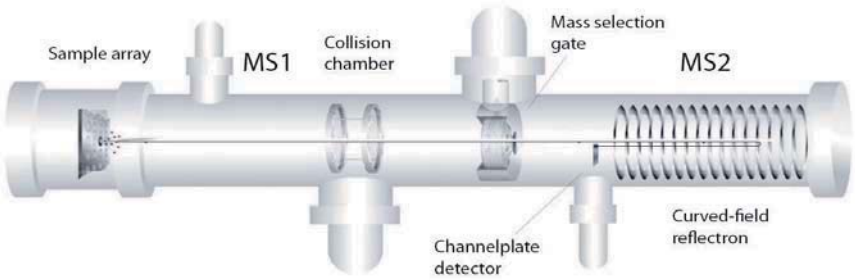

Figure 5: Schematic overview of a TOF/TOF mass spectrometer. A collision cell is positioned between two TOF analyzers, where the last is with a curved field reflectron (Figure adapted from UMass Medical School, Proteomic mass spectrometry lab, instrumentation, MALDI TOF[2]).

Fourier Transform Ion Cyclotron Resonance (FT-ICR)

Fourier Transform Ion Cyclotron Resonance mass spectrometer (FT-ICR MS or FT-MS) is a very powerful tool in mass spectrometry. The instrument was first described in 1974 by Comisarow and Marshall[69], although the basic ICR technology used has been around for more than 50 years[70]. In general, an FT-ICR MS has three main components: a superconductive magnet, an analyzer cell, and an ultrahigh vacuum system[71]. Ions are trapped by first entering an analyzer cell, which can be a cubic cell[72] or an open-ended cylindrical cell[73, 74]. The cell consists of six electrodes positioned to have the axis of the cylinder aligned with the magnetic field. Two cylinders, at the end of the cell, function as trapping electrodes. The other four (center) electrodes are used as excitation and detection plates. The cell itself is surrounded by a strong superconductive magnet with a field strength up to 25 Tesla (T)[75]. Inside the cell ions are trapped radially by a strong magnet field since they start a circular movement as soon as they enter the field, referred to as cyclotron movement. Ions are stored in the cell by using two electrodes for trapping in the axial direction. The angular frequency (ω_c) of the ions' circular motion is given by the following equation:

$$\omega_c = \frac{q\mathrm{B}}{m}$$

Where q is representing the charge of the ion, m is the mass of the ion and B is the magnetic field strength. Ions with different mass-to-charge ratios (m/q) can then be analyzed, separated and detected using their different cyclotron frequencies. The cyclotron frequency of an ion depends of the ion's charge q, its mass m and the magnetic field B, which is considered constant. A signal is obtained by first exciting the trapped ions simultaneously by applying a pulse or a radio-frequency electric field (covering the whole m/z range) using the excitation electrodes positioned parallel to the magnetic field. The excitation pulse results in ion bundles of the same m/q ratio having an increased velocity and a larger radius of motion, allowing the ions to be detected as they pass nearer the detection electrodes. When the ion bundles are passing near the detection electrodes an image current is induced in the two electrodes. The signal is then converted from a time-domain (transient alternating voltage) to a frequency-domain by Fourier transformation (FT), where the frequency spectrum can be converted into a mass spectrum (see Figure 6).

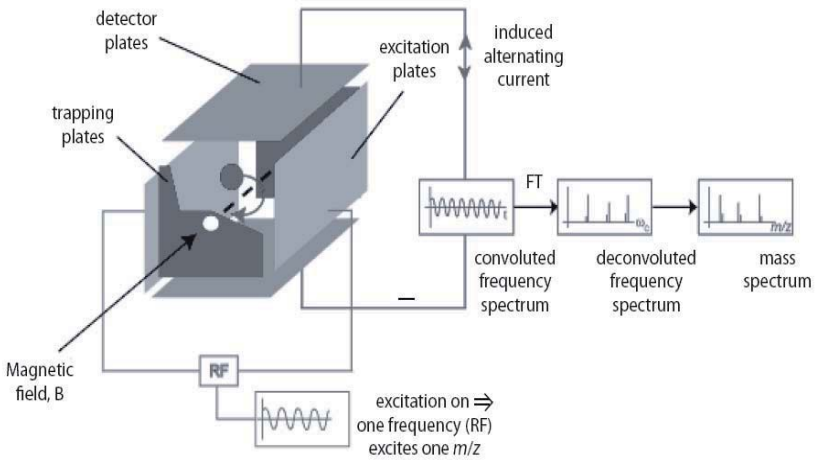

Figure 6: Schematic overview of the FT-ICR-MS showing the ion trapping, detection and signal generation (Figure adapted from University of Bristol, School of Chemistry, Mass Spectrometry Resource, Fourier-transform Ion Cyclotron Resonance (FT-ICR)).

Ions can be stored for a relatively long period of time in the cell (several seconds can be considered routine), which is possible because of a very high vacuum system. Residual gas molecules from the air would otherwise interfere with the ion bundle's circular motion and shorten the analysis and detection time.

FTICR instruments are able to perform multiple-stage MS (MS^n), as with ion trap instruments, where a selected product ion (generated from a specific precursor ion) can be retained for fragmentation and generation of an additional set of product ions. Other advantages of an FTICR mass spectrometer are very high resolution (which scales with B), mass accuracy and dynamic range, where the downsides are high cost, high magnetic field, space requirements and slow acquisition times. To solve the slow acquisition rate for MS/MS in the ICR cell, the FT-ICR analyzer is, nowadays, often used in combination with a linear quadrupole ion trap analyzer[54, 76, 77]. In this hybrid instrument the analyzers can be operated in parallel (MS in the ICR cell and MS/MS in the linear ion trap) for faster speeds. The FTICR can also be used in combination with more exotic fragmentation techniques such as infrared laser activation[78], or electron capture dissociation[79, 80].

Orbitrap

The Orbitrap is a relatively new mass analyzer which employs the trapping of ions in an electrostatic field[81]. The Orbitrap mass analyzer consists of an outer barrel-like electrode and a central spindle-like electrode along the axis, as illustrated in Figure 7. Orbital trapping is not a new invention as it dates back to 1923 where the concept was first implemented by Kingdon[42]. However, the model proposed by Makarov is the first to be used for the generation of mass spectra. Ions are injected into the Orbitrap and are trapped by applying a DC voltage to the central electrode while the outer electrode is set at ground potential. The ions then orbit around the central electrode due to the generated electrostatic field and will oscillate along the axis of the central electrode (Figure 7). The angular frequency (ω_z) of the axial oscillation is dependent on the mass to charge ratio of the ion and the potential between the electrodes (which is constant) and can be described as:

$$\omega_z = \sqrt{\frac{k}{m / q}}$$

Where *k* is the axial restoring force (the value is dependent on the shape of the electrode and the potential applied), *m* and *q* are the mass and charge of the ion. Ion oscillations are recorded using image current detection and the measured signals are amplified and transformed into mass spectra as described in the FT-ICR section above.

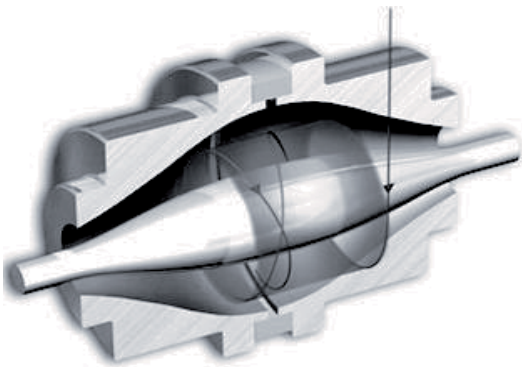

Figure 7: A cut-away model of the Orbitrap mass analyzer. Illustration of the outer barrel-like electrode of the mass analyzer and central spindle-like electrode along the axis. The arrows demonstrate the ion trajectories in the Orbitrap mass spectrometer. The ions orbit around the central electrode while oscillating back and forth along the axis (Figure adapted from Practical Proteomics 1-2/2006, Education & training, Scigelova and Makarov).

The first commercial instrument using the Orbitrap analyzer was introduced as a hybrid mass spectrometer, were the Orbitrap was coupled to a linear iontrap (LTQ). A schematic overview of the LTQ Orbitrap is shown in Figure 8. The analyte is introduced into the mass spectrometer using an ESI source. The generated ions are accumulated in the linear ion trap (LTQ part) and subsequently transferred to the Orbitrap part via a curved RF-only linear quadrupole ion trap termed the C-trap[82-84]. The shape and position of the C-trap enables fast and homogeneous injection of large populations of ions into the Orbitrap. In addition, studies performed by Olsen *et al.* showed that the C-trap could be used as a collision cell to obtain MS/MS information[85]. This hybrid instrument has the advantage of high resolution (>150.000 FWHM) and mass accuracy (<5ppm) obtained by the Orbitrap and the speed and sensitivity of the LTQ[82, 86]. Additionally, the instrument can be operated in parallel where the Orbitrap acquires MS scans while the LTQ is used for sequencing of peptide ions (MS/MS)[87-89].

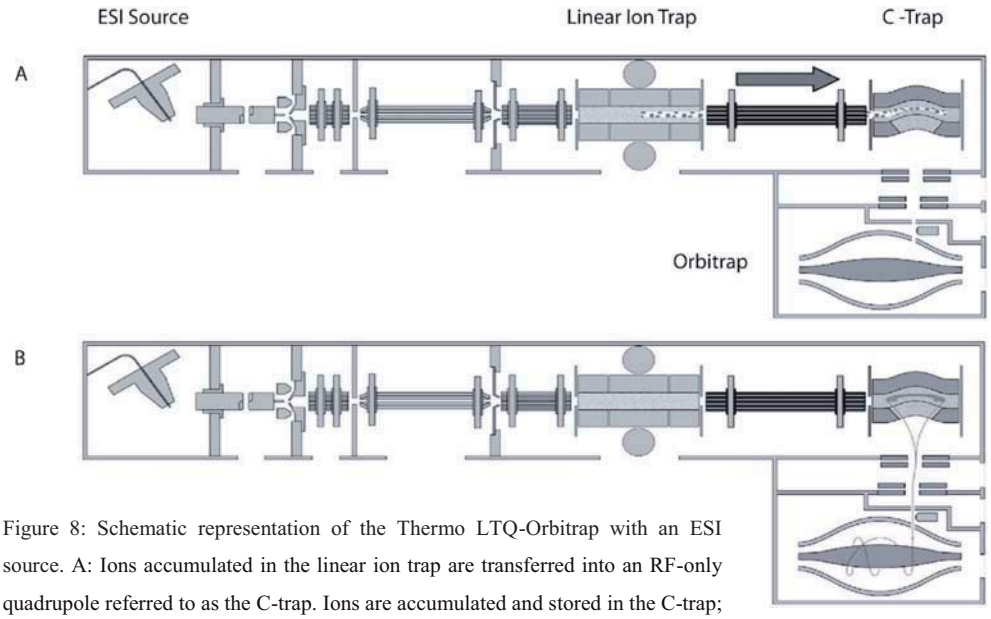

Figure 8: Schematic representation of the Thermo LTQ-Orbitrap with an ESI source. A: Ions accumulated in the linear ion trap are transferred into an RF-only quadrupole referred to as the C-trap. Ions are accumulated and stored in the C-trap; B: A pulse of injected ions from the C-trap is filled into the Orbitrap for later ion detection (Figure adapted from J Am Soc Mass Spectrom. 2006, 17(7) 977-982).

Later iterations of the instrument introduced a Higher Energy Collisional Dissociation cell (HCD cell)[85] and a dual pressure ion trap[90].

I.III Fragmentation techniques

A large range of mass analyzers and instrument designs are now available for proteomics analysis. The most popular type are hybrid instruments combining two or more mass analyzers, consecutive either in space (triple quadrupole and QTOF) or in time (ion storage devices), to allow tandem mass spectrometry (MS/MS). An MS/MS or MS^n experiment from a tandem in space instrument is usually performed by first measuring the mass to charge ratios of intact peptide ions. Subsequently, the first mass analyzer is used to select specific precursor ions, which are then transferred into a collision cell where excitation and dissociation occurs. The fragment or product ions are subsequently transferred to a second mass analyzer for the determination of their *m*/*z* values. The excitation of precursor ions is often achieved by collisions with inert gas molecules (helium, nitrogen or argon) and is referred to as collision induced dissociation (CID). However, different variations of CID exist depending on the instrument type and CID in beam type instruments can, in terms of energy, be further divided into high-energy (keV) CID and low-energy (energies less than 100-200 eV) CID[91]. Triple quadrupole and QTOF instruments use low-energy CID for parent ions fragmentation where instruments such as the TOF/TOF use energies that can reach up to several keV, defined as high-energy CID. Ion trap CID is again different as it makes use of resonance excitation or, in the case of an ICR cell, sustained off-resonance excitation (SORI)[92, 93].

The degree of peptide fragmentation depends on several factors such as the size of the precursor ion, amino acid composition and charge state. Another important parameter is the excitation method used, as there are several versions of CID in addition to alternative fragmentation methods. These include surface-induced dissociation (SID)[94], blackbody infrared radiative dissociation (BIRD)[95], electron-capture dissociation (ECD)[79] and electron transfer dissociation (ETD)[50], the latter two of which will be discussed below. Peptides fragment in a predictable fashion and generate sequence specific ions. The predictability of fragmentation has led to a nomenclature for annotating fragment ions in mass spectra, illustrated in Figure 10. The nomenclature was originally proposed by Roepstorff and Fohlman [96] and was later modified by Biemann[97]. Fragment ions that retain the positive charge on the amino terminal side are termed *a*, *b,* or *c* type ions and fragment ions retaining the positive charge on the carboxy terminal side are named *x*, *y*, or *z* type ions. If neutral losses are observed then they can be indicated by ° for the loss of water or by * for the loss of ammonia.

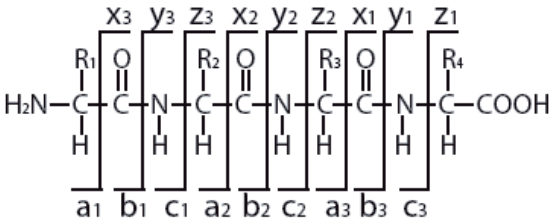

Figure 10: Illustration of the fragments formed through bond cleavages along the backbone of protonated peptides.

Collision induced/activated dissociation (CID/CAD)

Collision induced dissociation (CID) is the most common method of fragmentation in proteomics research and is also referred to as collision activated dissociation (CAD)[98, 99]. Many research groups have investigated the fragmentation of protonated peptide ions and a general model has been proposed called the mobile proton model[94, 100-108]. According to the model, the CID cleavage process of protonated peptides is charge directed. The mobile proton is transferred to various backbone amide sites and with the applied energy results in facile dissociation of the peptide backbone generating fragment ions[94, 105, 109-111]. Additionally, according to the model, more energy is needed to initiate peptide fragmentation of peptides lacking mobile protons[112]. The peptides belonging to this group are peptides with a number of basic residues that equals or exceeds the number of ionizing protons and ions that contain fixed charges[107, 113]. Examples of these types of peptides are doubly protonated peptides with two or more arginines in the peptide sequence and MALDI generated peptides with arginine at the C-terminus. It is essential to understand and be able to predict how peptides fragment in order to improve peptide sequencing and protein identification algorithms, commonly used to extract data from the MS/MS spectra.

Dissociation of a precursor ion with low-energy CID/CAD primarily results in the cleavage of the peptide-backbone at the amide bonds (C_o-N) generating *b* (N-terminal ions) and *y* type ions (C-terminal ions)[114]. A typical CID MS/MS spectrum contains both *b* and *y*-ions and ideally a full set of both ion types. The dissociation, however, does not always occur equally along the peptide backbone as cleavage near glutamic acid, glutamine and proline are often preferred. In addition, when tryptic peptides are fragmented using CID there is a tendency that *y*-ions will have a stronger signal than the equivalent *b*-ion series. The observed pattern was studied by different groups where it was found that the position of the basic residues in the peptide backbone influences the peak intensity of *b* and *y*-ions. Peptides with a basic

residue in close proximity to the N-terminus leads to the generation of prominent *b* series peaks in contrast to the intense *y* series normally observed for peptides with a basic residue close to the C-terminal (tryptic peptides)[115]. The *y*-ions are therefore the main ions used for the identification of peptides and proteins in CID MS/MS analysis[116]. Both ion series will however have additional characteristic fragments in the spectrum, which can be used for the identification of peptides. The loss of ammonia from the amino residues arginine, lysine, asparagine and glutamine is often observed, where serine, threonine, aspartic acid and glutamic acid often lose water[117].

Peptides with multiple basic residues generally have a negative effect on the CID/CAD fragmentation[118]. The random protonation along the peptide backbone is affected in such peptides and the generation of sufficient sequence information for a positive sequence analysis can therefore be hindered. An additional issue is the analysis of specific PTMs using CID-MS/MS, which often result in uninformative MS/MS spectra containing only a few fragment ions[119-122]. This is due to the preferred neutral loss of the post translational modification during the fragmentation process, instead of normal backbone sequence specific fragmentation which leads to the generation *b* and *y* fragment ions.

Electron capture/transfer dissociation (ECD/ETD)

McLafferty and co-workers introduced the fragmentation technique electron capture dissociation (ECD)[79] in 1998. ECD fragmentation is induced by a gas-phase reaction of low energy electrons (<1 eV) with multiply charged peptide and protein ions and is mainly performed in the Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer[123, 124]. Analyzed precursor ions have to be at least doubly charged, as the ECD product ions are otherwise not observed by mass spectrometry. If a singly charged positive ion would capture an electron the resulting net charge would be zero. Since it was also found that the electron capture cross section is proportional to the square of the ion charge, multiply charged ions are preferred for ECD reaction[123].

ECD cleavage pattern and ion formation is relatively unique among fragmentation techniques, as peptide backbone cleavage occurs at the N-Cα bonds[125], forming primarily c' (the ”prime” sign indicates transfer of $H^{\cdot}$ to the peptide fragment prior to bond cleavage, c + 1) and $z^{\cdot}$ (radical ions, z - 1) fragment ions[126]. ECD was found to be a complementary

dissociation technique to the conventional collision induced/activated dissociation (CID/CAD) method[127, 128].

Other less abundant product ions observed in ECD spectra are $a^{\cdot}$ – and y'-ions[126, 129] and to a slightly lesser extent *w*-ions[130, 131]. The formation of *b*-ions in ECD has likewise been observed in different studies[132, 133]. Cooper was the first to investigate the unconventional ECD fragment ion formation typically associated with CID-type processes[134]. She found that the generation of *b* type ions is depended on the type of charge carrier in the peptide sequence (lysine or arginine), as lysine containing peptides resulted in a higher abundance of *b*-ions[134]. ECD is considered "soft" fragmentation, which is an advantage when analyzing peptides with labile post translational modifications such as phosphorylation[135, 136], glycosylation[137, 138], and sulfation[80, 139] as these PTMs can then be preserved, thus revealing their position on the peptide[140]. However, the detailed mechanism of how ECD bond breakage occurs is still under debate.

An analogue to ECD, which uses electron transfer to induce peptide dissociation, was later introduced: referred to as electron transfer dissociation (ETD) [50, 141]. ETD, unlike ECD, can be performed in radio-frequency (RF) quadrupole ion trap instruments and has therefore become a more widespread fragmentation technique due to lower restrictions for the type of mass spectrometer required. In ETD an electron is transferred from a radical anion, usually fluoranthene, to the (multiply) protonated peptide, inducing fragmentation by cleavage along the peptide backbone at the amide groups (Cα-N). Fragmentation via ETD generates predominantly *c* and *z* type ions, similar to ECD, where the *c*-ion is often observed as an even-electron species (c') and the *z*-ion as an odd-electron species ($z^{\bullet}$). However, a hydrogen transfer reaction between the two ion products is possible where $c^{\cdot}$ (c – 1) and z' (z + 1) are formed[142, 143]. The dissociation mechanism of charged-reduced peptides (after electron transfer) in ETD is thought to be the same as for ECD and different theories have been published, with some arguing that the radical site on $z^{\bullet}$- ions is located on the Cα cleavage site where the electron or H-atom can migrate along the peptide backbone causing fragmentation by radical initiated rearrangement. However, the exact mechanism for ETD induced peptide dissociation is still under debate; whether it is the H-atom or electron that is transferred to a nearby bond to induce fragmentation is still under investigation[144-146].

ETD, as well as ECD, has shown to work particularly well for the fragmentation of large, highly charged peptides and small proteins[124, 147, 148]. However, earlier ETD studies have shown that doubly charged peptide precursors $[M + 2H]^{2+}$ do not readily fragment when performing ETD experiments, due to the fact that the dissociation efficiency is related to the

number of charges present in the precursor ion (where it has been suggested that the ion-ion reaction rates increase proportionally to the square of the charge).

To circumvent the difficulties of analyzing doubly protonated peptides, different techniques are used to activate precursor cations after electron transfer for more efficient fragmentation[149-151]. This was first recognized from early ECD studies where different approaches had been made to heat or activate cations prior to, or during, ECD analysis. To preheat cations, an elevated ICR cell temperature has been used, as well as infrared photons[152-155]. For activation prior to ECD analysis, collision-activation has been used, referred to as activated-ion ECD[156]. Fragmentation of peptides by ETD preserves modifications (phosphorylation, methylation, acetylation, sulfation, nitrosylation and glycosylation[50, 157-160]) on the peptide backbone, similar to ECD[161-163]. Analyzing these modified peptides by ETD has shown to generally increase the amount of ions observed in spectra compared to CID, creating more complete ion series and additionally increase sequence information as modifications are left intact for site determination.

As CID and ETD have shown to be efficient for the fragmentation of different subgroups of peptides (large peptides, highly basic peptides, peptides with PTMs, doubly charged peptides) researchers have looked into the possibility of utilizing the advantage of both fragmentation techniques[160]. That has, for instance, led to the development of “decision tree” methods where the mass spectrometer is set to choose between CID and ETD fragmentation depending on charge state and *m*/*z* value of the peptide to be fragmented[164].

II. Proteolytic enzymes for mass spectrometry based proteomics

There are, as mentioned earlier, two main approaches for the characterization of proteins, the direct MS analysis of intact proteins (also referred to as the "top-down" approach)[125, 165] and the analysis of a mixture of separated peptides generated by an enzymatic digest (referred to as the "bottom-up" approach)[166, 167]. While the "top-down" approaches has shown to be problematic when analyzing larger proteins (above 50 kDa) due to incomplete fragmentation spectra, the "bottom-up" approach also possesses limitations, limited sequence coverage and the requirement of improved separation and enrichment strategies. The necessity to improve separation techniques due to increase sample complexity will be discussed further in Section III. The aspects associated with proteolysis in bottom up proteomics workflows are very important both for sequence coverage and PTM analysis. Many proteases have therefore been tested for proteomics workflows in order to obtain optimal sequence coverage and to detect all possible PTMs on peptides and proteins[168].

Identification of amino acid sequences from MS/MS spectra are likewise generally divided into two approaches, *de novo* sequencing and database searching. Both approaches have been used extensively but have limitations. Database searching, for instance, might seem uncomplicated and straightforward; nevertheless it can be difficult to identify all proteins and their amino acid sequences as not all sequences are included in the databases and prior knowledge of the sample composition is required. *De novo* sequencing does not suffer from this constraint as it requires no preliminary information. Instead it is a requirement that all amino acid sequence peaks are generated and can be easily identified in the spectrum, which is normally difficult to achieve with the current strategies used today. Proteolytic enzymes influence the outcome of both approaches (*de novo* and database searching), as each enzyme generates a unique pool of peptides with different length distributions, solubility, and sequence characteristics. The length and charge state of generated peptides is important for successful peptide sequencing and combining data from different proteolytic digests can be useful for more comprehensive data analysis[169-171]. The importance of length, charge states and sequence characteristics of the generated peptides resulted in a number of studies to discover the most suitable 'general' enzyme for different MS/MS sequencing approaches[172, 173]. Tryptic peptides, subjected to ESI, result in mainly the generation of mainly doubly charged ions well suited for CID sequencing. Indeed, protein analysis combining tryptic proteolysis and tandem mass spectrometry using CID for peptide sequencing is undoubtedly the most common approach for protein identification. However, as CID does not offer a

complete solution for the analysis of proteins, alternative complementary fragmentation techniques, such as ETD, are becoming more established. As ETD has shown to be suited for the fragmentation of highly charged peptides, alternative proteases, as for example Lys-C, have been used to create longer peptides retaining more charges than tryptic peptides[174].

II.I. Enzymes used in proteomics workflows

Trypsin is the enzyme of choice for proteomic approaches because of its specific cleavage behavior, relatively well studied enzymatic behavior and since its proteolytic peptides are near-ideal for CID[175-178]. Even though trypsin is commonly used for proteomics analysis, new studies are still being presented, which contain improvements to current protocols. Klammer and MacCoss, for example, found that digestion of a sample using 1 hour of incubation identifies more proteins than the same sample subjected to a 24 hour digestion[179]. Pham *et al*. studied the thermostability of trypsin and found that glycation of lysine residues in trypsin greatly increases the thermostability of the enzyme[180]. This is particularly useful when studying native proteins, which can be more difficult to digest, and occasionally need to be subjected to a high temperature during digestion to remain unfolded. The use of high temperatures during the digestion could also exclude any chemical denaturants normally used for denaturing proteins prior to an enzymatic digestion. They additionally found that using high temperatures for protein digestion, possible by using glycated trypsin, could shorten incubation times or decrease the amount of enzyme normally used for a tryptic digestion. Studies performed by Petritis *et al*. showed that trypsin can be inactivated by boiling the enzyme in the absence of organic solvent. Trypsin is commonly used to catalyze stable isotope $^{16}O/^{18}O$ labeling of the C-terminal carboxyl groups of peptides, which can be reversed by the activity of residual trypsin in the sample. Boiling the peptide/trypsin solution can quench any remaining trypsin activity and thereby prevent oxygen back-exchange in ^{18}O labeled samples[181].

Even though trypsin is the most widely used protease for proteomic studies it is, at times, necessary to use alternative proteases, which have different cleavage specificity to obtain information of certain sub-groups of proteins or simply to increase sequence coverage. Table 1 highlights some of the most common enzymes used in proteomic workflows and their specificity.

Table 1. Proteolytic enzymes used in proteomic workflows.

Enzyme	Specificity
Trypsin	Arginine (R), Lysine (K)
Lys-C	Lysine (K)
Glu-C (V8)	Glutamic acid (E)
Arg-C	Arginine (R)
Asp-N	Asparagine (N)
Lys-N	Lysine (K)
Chymotrypsin	Phenylalanine (F), Leucine (L), Tyrosine (T), Tryptophan (W)
Elastase	Alanine (A), Valine (V), Leucine (L), Isoleucine (I), Serine (S), Threonine (T)
Pepsin	Phenylalanine (F), Leucine (L), Glutamic acid (E), Alanine (A)
Proteinase K	None

Where trypsin has shown to generate peptides suited for CID fragmentation, other proteases, such as Glu-C, Arg-C, Asp-N and Lys-C, have often been used to generate peptides for ETD/ECD fragmentation[173, 174]. The enzyme Asp-N has been used in multiple histone studies in combination with ETD for the identification of isoforms and PTM variants[182, 183]. Histones often contain a high amount of lysines (often modified) in the sequence, which make it difficult to determine the entire sequence or locate modification sites using trypsin. However, for the identification of phosphopeptides in ETD experiments Lys-C is more commonly used and a few groups have stated that it is more favorable than Glu-C, as more proteins can be identified[174, 184]. They argue that a notably impaired ETD fragmentation of Glu-C generated peptides is the explanation for the decrease in protein identification. The same has been noted for CID fragmentation, where Crockett *et al*. studied the identification of proteins from formalin-fixed paraffin-embedded cells[185]. The proteases Lys-C, Glu-C, Arg-N and Asp-N have all been used to improve protein identification and characterization in yeast and in general to increase sequence coverage[172].

The use of multiple enzymes has often been implemented to improve sequence coverage of specific proteins, complexes or organisms. The reason is that even though trypsin is commonly used for MS-based proteomics it does not provide a complete solution, as mentioned previously. Many tryptic peptides will have a length of 5-6 amino acids, which are too small for detection by typical proteomics MS-methods and thereby decrease sequence coverage of proteins and general protein identification. Combinations of different proteases have therefore often been applied to proteomic studies to gain more comprehensive sequence

coverage of identified proteins[169, 172]. In addition, combining data from different enzymatic digestions has shown to improve the identification of phosphorylated peptides. This has been shown by combining two specific enzymes, for example Lys-N and trypsin[186, 187] as well as combining trypsin with less specific enzymes as elastase, proteinase K and thermolysin[188]. For the analysis of membrane proteins the use of less specific enzymes (sometimes in combination with trypsin) is commonly exploited as specific enzymes often result in poor identification due to insufficient cleavage sites[189].

The relatively little explored enzyme in proteomics, Lys-N, was up to the work described in this thesis only used in a few studies. For instance, Rao *et al*. demonstrated that Lys-N is more compatible with ^{18}O labeling then previous ^{18}O labeling methods using Lys-C, trypsin or Glu-C, as these methods often generate a mixture of isotopic isoforms[190, 191]. The enzyme Lys-N will be described in more detail in Section II.II.

As mentioned briefly earlier, various other proteases, referred to as less specific, have been implemented in proteomics studies. They have shown to be helpful in the analysis of, for example, protein complexes and in phosphoproteomics and membrane proteomics, as they can access alternative cleavage sites and provide data that is otherwise not accessible using more conventional proteases.

Less specific enzymes are often used in membrane proteomics. Popular choices include pepsin, which, for example, was used for the characterization of aquaporin[192]. Chymotrypsin[193] has been used for the analysis of the membrane proteome of a number of organisms[189]. A less conventional enzyme, elastase[194], has likewise successfully been used in membrane proteomics[195]. However, others would argue that the extremely robust and non-specific enzyme Proteinase K is a better choice of enzyme for the identification of membrane proteins[196]

The mapping of phosphorylation sites has often only been made possible by the use of the more non conventional enzymes. An example is elastase, which in one example was used to provide a more complete coverage of PKA phosphorylation sites which was simply not possible using trypsin[197-199].

II.II. The Metalloendopeptidase Lys-N

The enzyme Lys-N belongs to the zinc metalloendopeptidase (MEP) superfamily, which as the name implies, require zinc for proteolytic activity. In the case of Lys-N, the enzyme was found to contain 1 atom of zinc per molecule[200]. The zinc-metalloendopeptidase Lys-N, used in the work presented in Chapter 2-5, is purified from the fruiting bodies of the edible mushroom Grifola frondosa (GF) and is therefore also referred to as GFMEP[200-203]. Nonaka *et al*. characterized the enzyme and found that it consists of a single polypeptide chain with 167 amino acids residues and found the molecular mass of the enzyme to be approximately 18 kDa[202]. Later studies performed by Hori *et al*. made it possible to obtain the first crystal structure of GFMEP (Figure 11)[203]. The enzyme activity of Lys-N/GFMEP was also studied by Nonaka *et al*., where it was found that the activity of Lys-N is inhibited by metallo chelators such as EDTA and *o*-phenanthroline. Additionally, they found that the activity can be fully restored by the addition of either Zn^{2+}, Mn^{2+}, Ca^{2+} or Co^{2+} to the zinc depleted enzyme[200, 201]. They further characterized Lys-N and found that the enzyme has a unique specificity towards peptidyl-lysine bonds (-X-Lys-) in peptides and proteins. Other interesting properties of the enzyme found include high thermostability, tolerance towards denaturing agents, binding properties to β-glycans and a maximal activity in an alkaline range of pH 9-10[201].

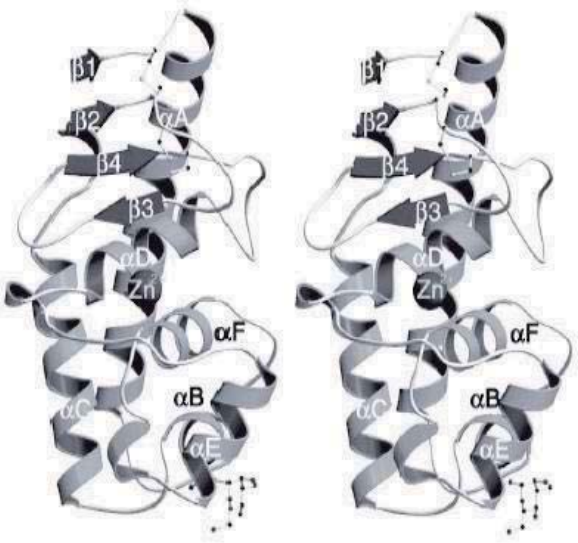

Figure 11. The structure of Lys-N/GFMEP. The figure illustrates a stereo ribbon representation of the overall structure of Lys-N/GFMEP. The structure shows 6 α-helices (light grey), 4 β-strands (dark grey), the β-turn and the loops. The catalytic zinc ion is illustrated as a dark grey ball[203] (Figure adapted from Acta Crystallogr D Biol Crystallogr 2001, 57, (Pt 3), 361-8).

II.III. Peptide sequencing

Amino acid sequencing of peptides and proteins started around 1950 with the development of Edman degradation [204]. Edman degradation was gradually replaced by MS based protein sequence determination especially after the advance of MS/MS techniques[100, 205-207]. The increased popularity and applicability brought by sensitivity and efficiency of MS/MS based sequence determination is also a result of advances in liquid chromatography (LC) techniques in combination with ESI[29]. Today, MS-based proteomics is mainly used for the identification of proteins in a complex mixture in combination with different search engines, as described below. Proteolytic digested peptides are usually analyzed by either LC-ESI-MS/MS or MALDI-MS/MS where generated data are compared to *in silico* generated comprehensive spectra using dedicated algorithms. Manual interpretation of spectra is vastly more time-consuming than algorithms, however, manual *de novo* sequencing (sequencing performed without prior knowledge of the amino acid sequence) of peptides can sometimes be a necessity[208]. It is the method of choice when analyzing samples from organisms with an unsequenced genome. Moreover, *de novo* sequencing can be helpful in analyses of protein sequence variants and their splice isoforms as well as for unexpected or unknown modified amino acids. Naturally, *de novo* sequencing can benefit if spectra can be made easier to interpret. For the generation of easily interpretable spectra alternative fragmentation techniques as well as alternative enzymes have been used[209]. The most common approach to simplify spectra is by using chemical modification at the peptide level. Two such methods are chemically assisted fragmentation (CAF)[210, 211] and 4-sulfophenyl isothiocyanate (SPITC), both simplify spectra by removing N-terminal fragment ions[210].

A number of *de novo* sequencing algorithms have been developed over the years to assist in the determination of peptide sequences from spectra such as PEAKS[212], PepNovo[213], SHERENGA[214], SeqMS[215, 216], Lutefisk[217], MSNovo[218], PFIA[219], AUDENS[220], NovoHMM[221], and pNovo[209]. It is not only simplified spectra that can facilitate *de novo* sequencing of MS/MS spectra, either manually or by using algorithms. It is just as salient that the spectra are generated from an instrument with high mass accuracy and high resolution, such as QTOFs, FTICRs and orbitraps[222-224]. The quality of the spectra is important as well, which can be affected by poor peptide fragmentation and cleavage abnormalities. To improve *de novo* sequencing alternative fragmentation techniques complementary to CID/CAD are commonly used. Furthermore, CID and ETD or ECD spectra derive from the same precursor and can therefore be paired for increased fragmentation information[225-227].

II.IV. Database searching

Database searching became a necessity due to the vast number of MS/MS spectra generated when the popularity of tandem mass spectrometers exploded. Several search algorithms have been developed (SEQUEST[116], MASCOT[228], OMSSA[229], X!TANDEM[230], VEMS[231], Prospector[232] etc.) to analyze MS/MS spectra, where SEQUEST[116] and MASCOT[228] remain the two most popular choices. The basic principle of database searching algorithms is that they match experimental spectra with theoretical spectra generated from *in silico* peptide sequences (peptide to spectrum matches or PSMs) where each algorithm uses its own unique scoring system to evaluate the peptide match. SEQUEST uses a crosscorrelation algorithm (XCorr) whereas MASCOT uses probability-based algorithms for the identification of peptides[228]. The XCorr value is dependent on the quality of the MS/MS spectrum and the quality of the match to the theoretical spectrum. The probability in the MASCOT database is calculated based on the match observed between experimental data and the sequence database. The size of the sequence database is also used to validate the significance of the score. The scores provide an aid to discriminate between correct and incorrect peptide assignments.

A common weakness for all search engines is that they generally do not perform well when peptides contain PTMs. Peptide identification by MS/MS analysis with the presence of PTMs therefore provides an important issue that requires further attention.

Several groups have compared search engines using data generated by CID fragmentation[233-236]. When using CID spectra only, the comparisons are generally based on *b* and *y*-ion formation, as they are usually observed in CID spectra, neglecting the *c* and *z*-ions produced in ECD and ETD.

The available software for the analysis of CID spectra is, in general, far more advanced than the equivalent software for analyzing both ETD and ECD data. One of the reasons being that the CID fragmentation behavior has been more extensively characterized, which has helped in the development of software more suitable for predicting expected fragment ions. Although many groups have documented commonly observed fragment ions produced by ETD/ECD fragmentation[237-239] the relative peak intensities and frequency of ion types observed in spectra are not as well characterized. It is therefore important that search engines are able to consider all possible ion types present in an ETD spectrum.

An additional scoring imbalance was observed when analyzing ECD/ETD derived data as it has previously been reported that the Mascot search engine seems to have a higher MS/MS

protein identification score for triply charged peptides, generated from either ECD or ETD, compared to doubly charged ions[240, 241]. This was found to be a result of different factors such as higher MS/MS peptide sequence coverage, higher number of fragment pairs and lower degree of hydrogen atom rearrangement for triply charged peptides compared to their doubly charged counterpart. It is indeed a setback that Mascot does not support hydrogen atom rearrangements for ETD and ECD produced data. To circumvent this problem certain groups have worked on increasing peptide charge state using for example *m*-nitrobenzyl alcohol as charge enhancer to improve identifications using Mascot[240].

III. Liquid Chromatography (LC) - Mass spectrometry

The current advances in mass spectrometry (MS) technology have lead to the possibility of analyzing more complex biological samples and of the characterization of proteomes. This, however, increases the necessity of advance sample separation to improve identification of complex samples. A commonly used separation technique to simplify peptide or protein mixtures prior to ESI-MS analysis is liquid chromatography (LC). One of the most applied strategies for LC- mass spectrometry (LC-MS) based proteomics is the "bottom-up" approach mentioned previously. However, when proteins are cleaved using a proteolytic enzyme prior to MS analysis the complexity of the proteome samples will increase tremendously. Another important issue is the high dynamic range of peptide concentrations, as certain peptides will be highly abundant compared to other peptides within the same sample. In general, when a complex sample mixture is analyzed by LC- tandem mass spectrometry (LC-MS/MS), the possibility for peptides to have the same retention time, and thus co-elute in the chromatogram, will increase. Consequently, the amount of proteins identified in the sample mixture will decrease as the number of ions selected for fragmentation will be limited. In addition, ion suppression will occur, in which abundant peptides will suppress the signal of lesser abundant peptides in the mixture. To overcome these problems (e.g. complexity and dynamic concentration range) pre-fractionation techniques are often implemented to reduce the complexity of the sample to be analyzed and hence increase peptide identification. To improve peptide separation prior to MS analysis two dimensional (2D) or multidimensional chromatographic steps based on different separation techniques (orthogonal separation) have been developed[242].

III.I Reverse Phase Liquid Chromatography (RP-LC)

One of the most common stationary phases used for 1D-LC separation is reversed phase liquid chromatography (RP-LC)[243-245]. The stationary phases for RP-LC generally consist of alkyl chains, which commonly have one of three lengths; C_4, C_8 or C_{18}. The stationary mechanism is based on hydrophobic interactions between the analyte and the column material coated with alkyl chains. The choice of column material (length of the alkyl chains) depends on the samples to be analyzed, as C4 is generally used for proteins and C8 and C18 for peptides and small molecules. Large molecules will most likely have more hydrophobic

moieties for the interaction with the column material and a short alkyl chain is therefore sufficient for the analyte to interact. Peptides and smaller molecules, on the other hand, will need longer alkyl chains to be retained on the column, C_8 and C_{18} columns with hydrophobic longer carbon chains are therefore used.

The separation of a peptide mixture is generally performed by loading the peptides onto a RP-column packed with C18 material, often with a particle size of 5 μm. Smaller particle sizes are preferred to achieve higher separation efficiencies. A common set-up for peptide separation, in a HPLC system, often combines the use of a "trapping" RP-column and analytical RP-column. Here, peptides are loaded onto the trap column using a highly aqueous mobile phase and later eluted onto the analytical RP-column using a highly organic mobile phase. The peptides are separated on the analytical column based on their hydrophobicity and the more hydrophobic the peptides are the longer they will be retained on the stationary phase of the analytical column. The peptide mixture is often separated using a linear gradient of organic solvent, where the retained compounds are gradually eluted off the analytical column[246]. Additionally, the mobile phases are generally acidified to improve chromatographic peak shape and to increase retention of charged analytes. Two of the most commonly used acids in mobile phases are formic acid and acetic acid, which are compatible with ESI.

Protein and peptide separation in proteomics research is, as mentioned earlier, often performed using RP-LC as it has, generally, the highest peak capacity[247, 248] and is directly compatible with ESI-MS where the column eluate can be directly analyzed without additional treatment. Indeed, the direct elution of peptides from an RP column into the ESI source is one of the most used LC-MS based approaches. Two important factors should, however, be considered when combining a chromatographic device directly with an ESI source: a low flow rate between the LC-device and the ESI source, and the selection of the appropriate mobile-phase composition, as they are both crucial to obtain high sensitivity in LC-ESI-MS

However, as the complexity of proteomics samples continuously increases, the separation capability is often insufficient. To improve separation performance of complex samples prior to MS analysis, multidimensional chromatography is often used where different stationary phases are combined sequentially, as described below (section III.III).

III.II Strong cation exchange (SCX)

Another popular chromatographic technique for the separation and characterization of peptides is strong cation exchange (SCX)[249, 250]. Generally, SCX is performed at low-pH (pH 3.0) as all the peptide carboxylates will be primarily in a protonated form and the separation procedure is then based on difference in positive net charge in solution. The SCX column is based on sulfonic acid groups linked to the surface of the stationary phase[251]. The positively charged peptides are mainly retained on the stationary phase based on ionic interactions with the negatively charged sulfonic groups. Elution of peptides from the stationary phase is performed with an increasing salt concentration in the buffer or by a stepwise increase of the buffer. The elution depends on the strength of the interaction between the peptides and SCX resin. A higher salt concentration is needed to interfere with the interaction of peptides that bind strongly to the SCX resin.

A drawback of the SCX packing material is the weak retention between the stationary phase and neutral species. Different groups have, as a result, used the developed "mixed mode" stationary phases to achieve additional hydrophobic selectivity[252, 253].

III.III. Multidimensional LC

Multidimensional protein identification technology is commonly referred to as MudPIT, which uses multidimensional liquid chromatography (LC/LC) combined with MS/MS analysis. Multidimensional liquid chromatography has been developed as a method to analyze highly complex samples necessary for large-scale proteome analysis and often combines the use of proteolytic digestion, ESI, MS/MS, and database searching. The challenge, at the peptide level, is the need to separate more than hundreds of thousands of peptides. Several multidimensional LC methods have been introduced to reduce sample complexity prior to MS analysis in order to decrease under-sampling by the mass spectrometer and to increase the level of proteome characterization[166, 254, 255]. One of the important factors concerning multidimensional separations is the orthogonality of the specific separation as it is often an advantage combining methods using different separation mechanisms. Some of the advantages are a higher peak capacity, better resolved peptides as they, for instance, can be separated by charge first (SCX) and by hydrophobicity second

(RP)[166], which again reduces the degree of ion suppression and simplifies sample complexity leading to minimized under-sampling of peptides injected into the MS.

One of the earliest approaches that was introduced, notably still one of the most powerful for such a task, is MudPIT (multidimensional protein identification technology)[254, 256]. The technique couples 2D-LC to MS/MS often using a bi- or triphasic microcapillary column packed with strong cation exchange (SCX) resin and reversed phase (RP) beads[254, 257], (Figure 12). The example shown in Figure 12 is based on an on-line triphasic separation system. On-line analysis usually refers to a separation system where the analyte is transferred between separation dimensions automatically without flow interruption. Here, the peptides from a digested sample are first loaded onto a fused silica microcapillary column (approximately 100 µm i.d.). The column is packed with RP material (C_{18}) followed by SCX material and last a second RP (C_{18}) phase upstream from the SCX phase, which elute directly into a mass spectrometer. The first C_{18} phase is used for desalting samples which can also be done offline prior to loading samples onto a biphasic column[258]. The second C_{18} phase is primarily used as an analytical column separating the peptides as they are displaced from the SCX column to the hydrophobic C_{18} resin. The system relies on two buffer systems, one system is used for displacing peptides from the first C_{18} phase to the SCX phase and a second buffer system is used to displace peptides from the SCX phase to the second C_{18} phase. Peptide separation using a triphasic separation system can be performed by using elution cycles, changing between SCX solvent and RP organic solvent where the SCX salt gradient is gradually increasing. Peptides eluted from the SCX phase are displaced to the second C_{18} phase where they are finally separated according to size and hydrophobicity prior to elution for direct MS analysis, using an increasing gradient of organic solvent. The advantage of the MudPIT method is the orthogonality of the chromatographic phases as peptides are separated according to charge (SCX) by controlling the salt concentration and separated according to hydrophobicity by using an acetonitrile gradient. The salt steps and length of RP gradient can vary and depend on the complexity of the sample.

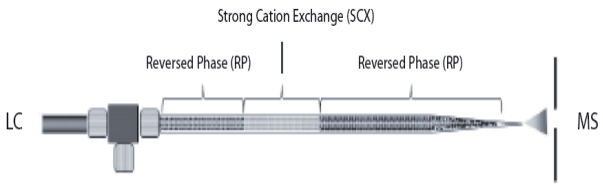

Figure 12: Example of a triphasic column packed with C_{18} material followed by strong cation exchange (SCX) resin and last a second C_{18} phase (Figure adapted from BioTechniques 2007, Vol. 43, No. 5, 563-569).

The SCX and RP separation can also be applied offline using two different chromatographic systems. The complexity of the peptide mixture is simplified in the first step (SCX) where the peptides are sequentially eluted into different fractions using a salt gradient. The obtained SCX fractions are then subjected to multiple RP-LC analyses on a second system for desalting and separation prior to MS/MS analysis.

Currently, the MudPIT separation technique is one of the most widely used approaches for the separation of a complex peptide mixture, whether it is on-line or off-line. It has a broad range of application areas ranging from membrane protein analysis[196] to phosphoproteome analysis[259-261] but also, for instance, for the targeted isolation of N-acetylated protein N-termini[262].

Although MudPIT is a highly developed technique in proteomic based mass spectrometry the approach does have certain limitations. The analysis leads to the generation of an incredibly large number of MS/MS spectra, which leaves a large demand on storage and data analysis. On the other hand, the complexity of some proteolytical digested samples, such as those belonging to human tissue, is such that only a fraction of these generated peptides will actually be selected for MS/MS fragmentation. Fragmentation of the peptides is based on semi-random selection which can result in low reproducibility.

The issue of "undersampling" has been discussed extensively in literature where it was argued that one of the reasons for the detected "undersampling" by single MudPIT experiments is due to the stochastic nature of the data-dependent sampling, where the most intense peptide peaks are selected for fragmentation. As duty cycle and dynamic range are limited, data-dependent acquisition will be biased towards the more abundant proteins in a sample, as they often have higher signal intensities. Repeated MS-measurements of the same sample therefore appears to partially solve the problem of "undersampling" and result in improved proteome coverage[263-268].

IV. Outline of the thesis

To overcome the difficulties of analyzing proteins in highly complex samples an improvement in proteomics strategies is needed. The combination of multiple proteases, peptide separation and fragmentation techniques may reduce sample complexity and improve the analysis of different sub-groups of peptides, including low abundant proteins and peptides. In this thesis, I introduce a relatively new protease in proteomics workflows and demonstrate the strength of combining its proteolytic peptides with multi-dimensional separation techniques and different fragmentation techniques, to decrease sample complexity and to improve sample identification.

In chapter 2, we evaluate the fragmentation pattern observed for peptides generated by the metalloendopeptidase Lys-N using electron transfer dissociation (ETD). The enzyme Lys-N generates peptides with a lysine residue at the N-terminal. We show that ETD sequencing of BSA generated peptides with an N-terminal lysine, and no other basic residue in the sequence, result in spectra dominated by *c*-type fragment ions. To confirm the result, fragment ion statistics were increased by analyzing Lys-N generated peptides from a cell lysate. Additionally, we show that these doubly charged Lys-N peptides containing a single lysine at the N-terminal can be selectively enriched for by using low-pH strong cation exchange (SCX).

In chapter 3, we further evaluate the SCX based fractionation of peptides generated from the metalloendopeptidase Lys-N. Here, we interestingly show that it is possible to obtain fractionation profiles where different subgroups of Lys-N generated peptides such as, acetylated N-terminal peptides, singly phosphorylated peptides, peptides with a single basic residue and peptides with multiple basic residues can be separated. We demonstrate that the combination of Lys-N digestion, low-pH SCX and reversed phase (RP) separation, with CID and ETD induced fragmentation, is a powerful approach for global proteome and phosphoproteome analysis.

In chapter 4, the metalloendopeptidase was explored for its use in MALDI-MS/MS proteomics applications. Lys-N generated peptides from a BSA digest were analyzed by MALDI-MS/MS, which resulted in simple and straightforward CID spectra, containing complete *b*-ion series. Statistical analysis was again performed to confirm the results where a cell lysate was digested to obtain a higher number of doubly charged Lys-N peptides. Last, it

was found that the simple straightforward MALDI CID spectra can be used to facilitate *de novo* sequencing.

In chapter 5, the proteolytic performance of the metalloendopeptidase Lys-N was evaluated. As a model system BSA was used to validate the performance of Lys-N when using a number of classical proteomics sample handling conditions. We demonstrate that Lys-N has many interesting and useful characteristics as it was found to be highly thermo-stable and to have a high tolerance towards certain denaturing agents, such as urea and acetonitrile. Furthermore, it was found that by increasing the digestion temperature a decrease in incubation time could be achieved. Additionally, we demonstrate that Lys-N is able to cleave adjacent to single-methylated lysines and partially adjacent to di-methylated lysines, which may be useful when analyzing naturally occurring post-translational modifications.

References

1. Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **1988,** 60, (20), 2299-301.
2. Karas, M.; Bachmann, D.; Bahr, U.; Hillenkamp, F., Matrix-assisted ultraviolet laser desorption of non-volatile compounds *International Journal of Mass Spectrometry and Ion Processes* **1987,** 78, 53-68
3. Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T., Protein and Polymer Analyses up to m/z 100 000 by Laser Ionization Time-of flight Mass Spectrometry. *Rapid. Commun. Mass Spectrom* **1988,** 2 (20), 151–153.
4. Gobom, J.; Schuerenberg, M.; Mueller, M.; Theiss, D.; Lehrach, H.; Nordhoff, E., Alpha-cyano-4-hydroxycinnamic acid affinity sample preparation. A protocol for MALDI-MS peptide analysis in proteomics. *Anal Chem* **2001,** 73, (3), 434-8.
5. Strupat, K.; Karas, M.; Hillenkamp, F., 2,5-DihydroxybenzoicAcid - a New Matrix for Laser Desorption Ionization Mass-Spectrometry. *Int. J. Mass Spectrom. Ion Processes* **1991,** 111, 89–102.
6. Schuerenberg, M.; Luebbert, C.; Eickhoff, H.; Kalkum, M.; Lehrach, H.; Nordhoff, E., Prestructured MALDI-MS sample supports. *Anal Chem* **2000,** 72, (15), 3436-42.
7. Towers, M. W.; Mckendrick, J. E.; Cramer, R., Introduction of 4-Chloro-r-cyanocinnamic Acid Liquid Matrices for High Sensitivity UV-MALDI MS. *J Proteome Res* **2010,** 9, 1931–1940.
8. Jaskolla, T. W.; Lehmann, W. D.; Karas, M., 4-Chloro-alpha-cyanocinnamic acid is an advanced, rationally designed MALDI matrix. *Proc Natl Acad Sci U S A* **2008,** 105, (34), 12200-5.
9. Mirza, S. P.; Raju, N. P.; Vairamani, M., Estimation of the proton affinity values of fifteen matrix-assisted laser desorption/ionization matrices under electrospray ionization conditions using the kinetic method. *J Am Soc Mass Spectrom* **2004,** 15, (3), 431-5.
10. Laugesen, S.; Roepstorff, P., Combination of two matrices results in improved performance of MALDI MS for peptide mass mapping and protein analysis. *J Am Soc Mass Spectrom* **2003,** 14, (9), 992-1002.
11. Beavis, R. C.; Chaudhary, T.; Chait, B. T., Alpha-Cyano-4-hydroxycinnamic acid as a matrix for matrixassisted laser desorption mass spectromtry. *Org Mass Spectrom* **1992,** 27, 156-158
12. Rehulkova, H.; Chalupova, J.; Sebela, M.; Rehulka, P., A convenient purification and preconcentration of peptides with alpha-cyano-4-hydroxycinnamic acid matrix crystals in a pipette tip for matrix-assisted laser desorption/ionization mass spectrometry. *J Mass Spectrom* **2010,** 45, (1), 104-11.
13. Andersson, M.; Groseclose, M. R.; Deutch, A. Y.; Caprioli, R. M., Imaging mass spectrometry of proteins and peptides: 3D volume reconstruction. *Nat Methods* **2008,** 5, (1), 101-8.
14. Beavis, R. C.; Chait, B. T., Matrix-assisted laser-desorption mass spectrometry using 355 nm radiation. *Rapid Commun Mass Spectrom* **1989,** 3, (12), 436-9.
15. Morelle, W.; Slomianny, M. C.; Diemer, H.; Schaeffer, C.; van Dorsselaer, A.; Michalski, J. C., Fragmentation characteristics of permethylated oligosaccharides using a matrix-assisted laser desorption/ionization two-stage time-of-flight (TOF/TOF) tandem mass spectrometer. *Rapid Commun Mass Spectrom* **2004,** 18, (22), 2637-49.
16. Kjellstrom, S.; Jensen, O. N., Phosphoric acid as a matrix additive for MALDI MS analysis of phosphopeptides and phosphoproteins. *Anal Chem* **2004,** 76, (17), 5109-17.
17. Dreisewerd, K., The desorption process in MALDI. *Chem Rev* **2003,** 103, (2), 395-426.
18. Chen, X.; Caroll, J. A.; Beavis, R. C. J., Near-ultraviolet-induced matrix-assisted laser desorption/ionization as a function of wavelength. *J. Am. Soc. Mass Spectrom.* **1998,** 9, 885.
19. Berkenkamp, S.; Kirpekar, F.; Hillenkamp, F., Infrared MALDI mass spectrometry of large nucleic acids. *Science* **1998,** 281, (5374), 260-2.
20. Durrant, E. E.; Brown, R. S., Wavelength Dependence On The Level Of Post-Source Metastable Ion Decay Observed In Infrared Matrix-Assisted Laser Desorption Ionization. *Int J Mass Spectrom* **2009,** 287, (1-3), 119-127.

21. Holle, A.; Haase, A.; Kayser, M.; Hohndorf, J., Optimizing UV laser focus profiles for improved MALDI performance. *J Mass Spectrom* **2006,** 41, (6), 705-16.
22. Knochenmuss, R., A quantitative model of ultraviolet matrix-assisted laser desorption/ionization. *J Mass Spectrom* **2002,** 37, (8), 867-77.
23. Knochenmuss, R., A quantitative model of ultraviolet matrix-assisted laser desorption/ionization including analyte ion generation. *Anal Chem* **2003,** 75, (10), 2199-207.
24. Karas, M.; Gluckmann, M.; Schafer, J., Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *J Mass Spectrom* **2000,** 35, (1), 1-12.
25. Ehring, E.; Karas, M.; Hillenkamp, F., Role of photoionization and photochemistry in ionization processes of organic molecules and relevance for matrix-assisted laser desorption ionization mass spectrometry. *Org. Mass Spectrom.* **1992,** 27, 472–480.
26. Rayleigh, L., On the Equilibrium of Liquid Conducting Masses charged with Electricity. *Philosophical Magazine* **1882** 14, 184–186.
27. Chapman, S., Carrier Mobility Spectra of Spray Electrified Liquids. *Phys. Rev.* **1937,** 52, 184-190.
28. Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B., Molecular Beams of Macroions. *J. Chem. Phys.* **1968,** 49, 2240-2249.
29. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989,** 246, (4926), 64-71.
30. Iribarne, J. V.; Thomson, B. A., On the evaporation of small ions from charged droplets. *J. Chem. Phys.* **1976,** 64, 2287-2294.
31. Wilm, M.; Mann, M., Analytical properties of the nanoelectrospray ion source. *Anal Chem* **1996,** 68, (1), 1-8.
32. Emmett, M. R.; Caprioli, R. M., Micro-electrospray mass spectrometry: ultra-high-sensitivity analysis of peptides and proteins *Journal of the American Society for Mass Spectrometry* **1994,** 5, (7), 605-613.
33. Wilm, M. S.; Mann, M., Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? . *International Journal of Mass Spectrometry and Ion Processes* **1994,** 136, (2-3), 167-180.
34. Whitehouse, C. M.; Dreyer, R. N.; Yamashita, M.; Fenn, J. B., Electrospray interface for liquid chromatographs and mass spectrometers. *Anal Chem* **1985,** 57, (3), 675-9.
35. Yost, R. A.; Enke, C. G., Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation. *Anal Chem* **1979,** 51, 1251A-1264A.
36. Glish, G. L.; Goeringer, D. E., Tandem Quadrupole/Time-of-Flight Instrument for Mass Spectrometry/Mass Spectrometry. *Anal Chem* **1984,** 56, 2291-2295.
37. Morris, H. R.; Paxton, T.; Panico, M.; McDowell, R.; Dell, A., A novel geometry mass spectrometer, the Q-TOF, for low-femtomole/attomole-range biopolymer sequencing. *J Protein Chem* **1997,** 16, (5), 469-79.
38. Morris, H. R.; Paxton, T.; Dell, A.; Langhorne, J.; Berg, M.; Bordoli, R. S.; Hoyes, J.; Bateman, R. H., High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Commun Mass Spectrom* **1996,** 10, (8), 889-96.
39. Hartmer, R. G.; Kaplan, D. A.; Stoermer, C.; Lubeck, M.; Park, M. A., Data-dependent electron transfer dissociation of large peptides and medium size proteins in a QTOF instrument on a liquid chromatography timescale. *Rapid Commun Mass Spectrom* **2009,** 23, (15), 2273-82.
40. Xia, Y.; Chrisman, P. A.; Erickson, D. E.; Liu, J.; Liang, X.; Londry, F. A.; Yang, M. J.; McLuckey, S. A., Implementation of ion/ion reactions in a quadrupole/time-of-flight tandem mass spectrometer. *Anal Chem* **2006,** 78, (12), 4146-54.
41. Dehmelt, H. G.; Walls, F. L., Bolometric Technique for the rf Spectroscopy of Stored Ions. *Phys. Rev. Lett.* **1968,** 21, 127-131.

42. Kingdon, K. H., A method for the neutralization of electron space charge by positive ionization at very low gas pressures. *Phys Rev* **1923,** 21, 408-418.
43. Paul, W.; Steinwedel, H., Ein neues Massenspektrometer ohne Magnetfeld. *Zeitschrift für Naturforschung* **1953,** 8a, 448-450.
44. Stafford, G. C. J.; Kelley, P. E.; Syka, J. E. P.; Reynolds, W. E.; Todd, J. F. J., Recent improvements in and analytical applications of advanced ion trap technology *International Journal of Mass Spectrometry and Ion Processes* **1984,** 60, 85-98
45. Hager, J. W., A new linear ion trap mass spectrometer. *Rapid Commun Mass Spectrom* **2002,** 16, 512-526.
46. Schwartz, J. C.; Senko, M. W.; Syka, J. E., A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom* **2002,** 13, (6), 659-69.
47. Johnson, J. V.; Yost, R. A.; Kelley, P. E.; Bradford, D. C., Tandem-in-space and Tandem-in-time mass-spectrometry-triple quadrupoles and quadrupole ion traps. . *Anal. Chem.* **1990,** 62, 2162–2172.
48. McLuckey, S. A.; Glish, G. L.; Van Berkel, G. J., Charge determination of product ions formed from collision-induced dissociation of multiply protonated molecules via ion/molecule reactions. *Anal Chem* **1991,** 63, (18), 1971-8.
49. Louris, J. N.; Cooks, R. G.; Syka, J. E. P.; Kelley, P. E.; Stafford, G. C.; Todd, J. F. J., Instrumentation, application and energy deposition in quadrupole ion trap MS/MS spectrometry. *Anal. Chem.* **1987,** 59, 1677-1685.
50. Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* **2004,** 101, (26), 9528-33.
51. Yang, Y. H.; Lee, K.; Jang, K. S.; Kim, Y. G.; Park, S. H.; Lee, C. S.; Kim, B. G., Low mass cutoff evasion with qz value optimization in ion trap *Analytical Biochemistry* **2009,** 387, (1), 133-135.
52. Zhang, M. Y.; Pace, N.; Kerns, E. H.; Kleintop, T.; Kagan, N.; Sakuma, T., Hybrid triple quadrupole–linear ion trap mass spectrometry in fragmentation mechanism studies: application to structure elucidation of buspirone and one of its metabolites. *J Mass Spectrom* **2005,** 40, 1017–1029.
53. Hopfgartner, G.; Varesio, E.; Tschappat, V.; Grivet, C.; Bourgogne, E.; Leuthold, L. A., Triple quadrupole linear ion trap mass spectrometer for the analysis of small molecules and macromolecules. *J Mass Spectrom* **2004,** 39, (8), 845-55.
54. Syka, J. E.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; Ueberheide, B.; Garcia, B.; Busby, S.; Muratore, T.; Shabanowitz, J.; Hunt, D. F., Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J Proteome Res* **2004,** 3, (3), 621-6.
55. Dangi, B. B.; Sassin, N. A.; Ervin, K. M., Pulsed ion extraction diagnostics in a quadrupole ion trap linear time-of-flight mass spectrometer. *Rev Sci Instrum* **2010,** 81, (6), 063302.
56. Rousu, T.; Herttuainen, J.; Tolonen, A., Comparison of triple quadrupole, hybrid linear ion trap triple quadrupole, time-of-flight and LTQ-Orbitrap mass spectrometers in drug discovery phase metabolite screening and identification in vitro--amitriptyline and verapamil as model compounds. *Rapid Commun Mass Spectrom* **2010,** 24, (7), 939-57.
57. Stephens, W. E., A Pulsed Mass Spectrometer with Time Dispersion. *Phys. Rev.* **1946,** 69, 691.
58. Wiley, W. C.; MacLaren, I. H., Time-of-Flight Mass Spectrometer with Improved Resolution. *Rev. Sci. Instrum.* **1955,** 26, 1150-1157.
59. Mamyrin, B. A.; Karataev, V. I.; Shmikk, D. V.; Zagulin, V. A., The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution. *Sov. Phys. JETP* **1973,** 37, 45.
60. Medzihradszky, K. F.; Campbell, J. M.; Baldwin, M. A.; Falick, A. M.; Juhasz, P.; Vestal, M. L.; Burlingame, A. L., The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal Chem* **2000,** 72, (3), 552-8.

61. Yergey, A. L.; Coorssen, J. R.; Backlund, P. S., Jr.; Blank, P. S.; Humphrey, G. A.; Zimmerberg, J.; Campbell, J. M.; Vestal, M. L., De novo sequencing of peptides using MALDI/TOF-TOF. *J Am Soc Mass Spectrom* **2002,** 13, (7), 784-91.
62. Cornish, T. J.; Cotter, R. J., Tandem time-of-flight mass spectrometer. *Anal Chem* **1993,** 65, (8), 1043-7.
63. Cotter, R. J.; Gardner, B. D.; Iltchenko, S.; English, R. D., Tandem time-of-flight mass spectrometry with a curved field reflectron. *Anal Chem* **2004,** 76, (7), 1976-81.
64. Kim, T. Y.; Schwartz, J. C.; Reilly, J. P., Development of a linear ion trap/orthogonal-time-of-flight mass spectrometer for time-dependent observation of product ions by ultraviolet photodissociation of peptide ions. *Anal Chem* **2009,** 81, (21), 8809-17.
65. Spengler, B.; Kirsch, D.; Kaufmann, R., Metastable decay of peptides and proteins in matrix-assisted laser-desorption mass spectrometry. *Rapid Commun Mass Spectrom* **1991,** 5, 198-202.
66. Spengler, B., Post-source Decay Analysis in Matrix-assisted Laser Desorption/Ionization Mass Spectrometry of Biomolecules. *J Mass Spectrom* **1997,** 32, 1019-1036.
67. Vestal, M. L.; Campbell, J. M., Tandem time-of-flight mass spectrometry. *Methods Enzymol* **2005,** 402, 79-108.
68. Piyadasa, C. K.; Hakansson, P.; Ariyaratne, T. R.; Barofsky, D. F., A high resolving power ion selector for post-source decay measurements in a reflecting time-of-flight mass spectrometer. *Rapid Commun Mass Spectrom* **1998,** 12, (22), 1655-64.
69. Comisarow, M. B.; Marshall, A. G., Frequency-sweep fourier transform ion cyclotron resonance spectroscopy. *Chem. Phys. Lett.* **1974,** 26, (4), 489-490
70. Sommer, H.; Thomas, H. A.; Hipple, J. A., A Precise Method of Determining the Faraday by Magnetic Resonance. *Phys. Rev.* **1949,** 76, 1877-1878.
71. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S., Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev* **1998,** 17, (1), 1-35.
72. Comisarow, M. B., Cubic trapped-ion cell for ion-cyclotron resonance. *Int. J. Mass Spectrom. Ion Proc.* **1981,** 37, 251–257.
73. Frankevich, V.; Zenobi, R., Dynamic ion trapping in a cylindrical open cell for Fourier transform ion cyclotron resonance mass spectrometry. *Int. J. Mass Spectrom* **2001,** 207, 57–67.
74. Gabrielse, G.; Haarsma, L.; Rolston, S. L., Open-endcap Penning traps for high precision experiments *Int. J. Mass Spectrom. Ion Proc.* **1989,** 88, (2-3), 319-332
75. Shi, S. D. H.; Drader, J. J.; Hendrickson, C. L.; Marshall, A. G., Fourier Transform Ion Cyclotron Resonance Mass Spectrometry in a High Homogeneity 25 Tesla Resistive Magnet. *J Am Soc Mass Spectrom* **1999,** 10, 265–268.
76. Belov, M. E.; Nikolaev, E. N.; Alving, K.; Smith, R. D., A new technique for unbiased external ion accumulation in a quadrupole two-dimensional ion trap for electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun Mass Spectrom* **2001,** 15, (14), 1172-80.
77. Niu, C. Y.; Wu, C. S.; Sheng, Y. X.; Zhang, J. L., Identification and characterization of flavonoids from semen zizyphi spinosae by high-performance liquid chromatography/linear ion trap FTICR hybrid mass spectrometry. *J Asian Nat Prod Res* **2010,** 12, (4), 300-12.
78. Laskin, J.; Futrell, J. H., Activation of large ions in FT-ICR mass spectrometry. *Mass Spectrom Rev* **2005,** 24, (2), 135-67.
79. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W., Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998,** 120, (13), 3265-3266.
80. Haselmann, K. F.; Budnik, B. A.; Olsen, J. V.; Nielsen, M. L.; Reis, C. A.; Clausen, H.; Johnsen, A. H.; Zubarev, R. A., Advantages of external accumulation for electron capture dissociation in Fourier transform mass spectrometry. *Anal Chem* **2001,** 73, (13), 2998-3005.
81. Makarov, A., Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* **2000,** 72, (6), 1156-62.

82. Makarov, A.; Denisov, E.; Lange, O.; Horning, S., Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. *J Am Soc Mass Spectrom* **2006,** 17, (7), 977-82.
83. Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S., Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem* **2006,** 78, (7), 2113-20.
84. Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M., Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* **2005,** 4, (12), 2010-21.
85. Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M., Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* **2007,** 4, (9), 709-12.
86. Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R., The Orbitrap: a new mass spectrometer. *J Mass Spectrom* **2005,** 40, (4), 430-43.
87. Yates, J. R.; Cociorva, D.; Liao, L.; Zabrouskov, V., Performance of a linear ion trap-Orbitrap hybrid for peptide analysis. *Anal Chem* **2006,** 78, (2), 493-500.
88. Han, X.; Aslanian, A.; Yates, J. R., 3rd, Mass spectrometry for proteomics. *Curr Opin Chem Biol* **2008,** 12, (5), 483-90.
89. Kim, M. S.; Kandasamy, K.; Chaerkady, R.; Pandey, A., Assessment of Resolution Parameters for CID-Based Shotgun Proteomic Experiments on the LTQ-Orbitrap Mass Spectrometer. *J Am Soc Mass Spectrom* **2010**.
90. Olsen, J. V.; Schwartz, J. C.; Griep-Raming, J.; Nielsen, M. L.; Damoc, E.; Denisov, E.; Lange, O.; Remes, P.; Taylor, D.; Splendore, M.; Wouters, E. R.; Senko, M.; Makarov, A.; Mann, M.; Horning, S., A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol Cell Proteomics* **2009,** 8, (12), 2759-69.
91. Xia, Y.; Liang, X.; McLuckey, S. A., Ion trap versus low-energy beam-type collision-induced dissociation of protonated ubiquitin ions. *Anal Chem* **2006,** 78, (4), 1218-27.
92. Goeringer, D. E.; McLuckey, S. A., Evolution of Ion Internal Energy During Collisional Excitation in the Paul Ion Trap: A Stochastic Approach. *J. Chem. Phys.* **1996,** 104, 2214-2221.
93. Heck, A. J. R.; Derrick, P. J., Ultrahigh Mass Accuracy in Isotope-Selective Collision-Induced Dissociation Using Correlated Sweep Excitation and Sustained Off-Resonance Irradiation—A Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Case Study on the [M + 2H](2+) Ion of Bradykinin. *Anal. Chem.* **1997,** 69, 3603–3607

94. Dongre, A. R.; Somogyi, A.; Wysocki, V. H., Surface-induced dissociation: an effective tool to probe structure, energetics and fragmentation mechanisms of protonated peptides. *J Mass Spectrom* **1996,** 31, (4), 339-50.
95. Schnier, P. D.; Price, W. D.; Jockusch, R. A.; Williams, E. R., Blackbody infrared radiative dissociation of bradykinin and its analogues: energetics, dynamics, and evidence for salt-bridge structures in the gas phase. *J Am Chem Soc* **1996,** 118, (30), 7178-89.
96. Roepstorff, P.; Fohlman, J., Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* **1984,** 11, (11), 601.
97. Biemann, K., Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol* **1990,** 193, 886-7.
98. McLuckey, S. A.; Goeringer, D. E.; Glish, G. L., Collisional activation with random noise in ion trap mass spectrometry. *Anal Chem* **1992,** 64, (13), 1455-60.
99. Hayes, R. N.; Gross, M. L., Collision-induced dissociation. *Methods Enzymol* **1990,** 193, 237-63.
100. Biemann, K.; Martin, S. A., Mass-spectrometric determination of the amino-acid-sequence of peptides and proteins. *Mass Spectrom. Rev.* **1987,** 6, 1-75.
101. McCormack, A. L.; Somogyi, A.; Dongre, A. R.; Wysocki, V. H., Fragmentation of protonated peptides: surface-induced dissociation in conjunction with a quantum mechanical approach. *Anal Chem* **1993,** 65, (20), 2859-72.

102. Summerfield, S. G.; Whiting, A.; Gaskell, S. J., Intra-ionic interactions in electrosprayed peptide ions. *Int. J. Mass Spec. and Ion Proc.* **1997,** 162, 149-161.
103. Burlet, O.; Orkiszewski, R. S.; Ballard, K. D.; Gaskell, S. J., Charge promotion of low-energy fragmentations of peptide ions. *Rapid Commun Mass Spectrom* **1992,** 6, (11), 658-62.
104. Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A., Mobile and localized protons: a framework for understanding peptide dissociation. *J Mass Spectrom* **2000,** 35, (12), 1399-406.
105. Dongre, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H., Influence of Peptide Composition, Gas-Phase Basicity and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model *J. Am. Chem. Soc.* **1996,** 118, (35), 8365-8374
106. Cox, K. A.; Gaskell, S. J.; Morris, M.; Whiting, A., Role of the Site of Protonation in the Low-Energy Decompositions of Gas-Phase Peptide Ions *J. Am. Soc. Mass Spectrom.* **1996,** 7, 522-531.
107. Gu, C.; Tsaprailis, G.; Breci, L.; Wysocki, V. H., Selective gas-phase cleavage at the peptide bond C-terminal to aspartic acid in fixed-charge derivatives of Asp-containing peptides. *Anal Chem* **2000,** 72, (23), 5804-13.
108. Shukla, A. K.; Futrell, J. H., Tandem mass spectrometry: dissociation of ions by collisional activation. *J Mass Spectrom* **2000,** 35, (9), 1069-90.
109. Zhang, Z. Q., Prediction of low-energy collision-induced dissociation spectra of peptides. *Anal Chem* **2004,** 76, 3908-3922.
110. Madine, J.; Jack, E.; Stockley, P. G.; Radford, S. E.; Serpell, L. C.; Middleton, D. A., Structural insights into the polymorphism of amyloid-like fibrils formed by region 20-29 of amylin revealed by solid-state NMR and X-ray fiber diffraction. *J Am Chem Soc* **2008,** 130, (45), 14990-5001.
111. Sun, F.; Liu, R.; Zong, W.; Tian, Y.; Wang, M.; Zhang, P., A Unique Approach to the Mobile Proton Model: Influence of Charge Distribution on Peptide Fragmentation. *J. Phys. Chem. B* **2010,** 114, 6350-6353.
112. Bythell, B. J.; Suhai, S.; Somogyi, A.; Paizs, B., Proton-driven amide bond-cleavage pathways of gas-phase peptide ions lacking mobile protons. *J Am Chem Soc* **2009,** 131, (39), 14057-65.
113. Tsaprailis, G.; Nair, H.; Somogyi, Á.; Wysocki, V. H.; Zhong, W.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J., Influence of secondary structure on the fragmentation of protonated peptides. *J Am Chem Soc* **1999,** 121, (22), 5142-5154.
114. Steen, H.; Mann, M., The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* **2004,** 5, (9), 699-711.
115. Tabb, D. L.; Huang, Y.; Wysocki, V. H.; Yates, J. R., 3rd, Influence of basic residue content on fragment ion peak intensities in low-energy collision-induced dissociation spectra of peptides. *Anal Chem* **2004,** 76, (5), 1243-8.
116. Eng, J. K.; McCormack, A. L.; Yates, J. R. I., An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **1994,** 5, 976-989.
117. Savitski, M. M.; Kjeldsen, F.; Nielsen, M. L.; Zubarev, R. A., Relative specificities of water and ammonia losses from backbone fragments in collision-activated dissociation. *J Proteome Res* **2007,** 6, (7), 2669-73.
118. Bridgewater, J. D.; Srikanth, R.; Lim, J.; Vachet, R. W., The effect of histidine oxidation on the dissociation patterns of peptide ions. *J Am Soc Mass Spectrom* **2007,** 18, (3), 553-62.
119. Leitner, A.; Foettinger, A.; Lindner, W., Improving fragmentation of poorly fragmenting peptides and phosphopeptides during collision-induced dissociation by malondialdehyde modification of arginine residues. *J Mass Spectrom* **2007,** 42, (7), 950-9.
120. DeGnore, J. P.; Qin, J., Fragmentation of phosphopeptides in an ion trap mass spectrometer. *J Am Soc Mass Spectrom* **1998,** 9, (11), 1175-88.
121. Tholey, A.; Reed, J.; Lehmann, W. D., Electrospray tandem mass spectrometric studies of phosphopeptides and phosphopeptide analogues. *J Mass Spectrom* **1999,** 34, (2), 117-23.
122. Nemeth-Cawley, J. F.; Karnik, S.; Rouse, J. C., Analysis of sulfated peptides using positive electrospray ionization tandem mass spectrometry. *J Mass Spectrom* **2001,** 36, (12), 1301-11.

123. Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W., Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem* **2000,** 72, (3), 563-73.
124. McLafferty, F. W.; Horn, D. M.; Breuker, K.; Ge, Y.; Lewis, M. A.; Cerda, B.; Zubarev, R. A.; Carpenter, B. K., Electron capture dissociation of gaseous multiply charged ions by Fourier-transform ion cyclotron resonance. *J Am Soc Mass Spectrom* **2001,** 12, (3), 245-9.
125. McLafferty, F. W.; Fridriksson, E. K.; Horn, D. M.; Lewis, M. A.; Zubarev, R. A., Techview: biochemistry. Biomolecule mass spectrometry. *Science* **1999,** 284, (5418), 1289-90.
126. Kruger, N. A.; Zubarev, R. A.; Horn, D. M.; McLafferty, F. W., Electron capture dissociation of multiply charged peptide cations. *International Journal of Mass Spectrometry* **1999,** 185-187, 787-793
127. Leymarie, N.; Berg, E. A.; McComb, M. E.; O'Connor, P. B.; Grogan, J.; Oppenheim, F. G.; Costello, C. E., Tandem mass spectrometry for structural characterization of proline-rich proteins: application to salivary PRP-3. *Anal Chem* **2002,** 74, (16), 4124-32.
128. Nielsen, M. L.; Savitski, M. M.; Zubarev, R. A., Improving protein identification using complementary fragmentation techniques in fourier transform mass spectrometry. *Mol Cell Proteomics* **2005,** 4, (6), 835-45.
129. Zubarev, R. A.; Kruger, N. A.; Fridriksson, E. K.; Lewis, M. A.; Horn, D. M.; Carpenter, B. K.; McLafferty, F. W., Electron Capture Dissociation of Gaseus Multiply-Charged Proteins Is Favored at Disulfid Bonds and Other High Hydrogen Atom affinity *J. Am. Chem. Soc.* **1999,** 121, (12), 2857-2862.
130. Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A., Side-chain losses in electron capture dissociation to improve peptide identification. *Anal Chem* **2007,** 79, (6), 2296-302.
131. Li, X.; Lin, C.; Han, L.; Costello, C. E.; O'Connor, P. B., Charge remote fragmentation in electron capture and electron transfer dissociation. *J Am Soc Mass Spectrom* **2010,** 21, (4), 646-56.
132. Liu, H.; Hakansson, K., Abundant b-type ions produced in electron capture dissociation of peptides without basic amino acid residues. *J Am Soc Mass Spectrom* **2007,** 18, (11), 2007-13.
133. Haselmann, K. F.; Schmidt, M., Do b-ions occur from vibrational excitation upon H-desorption in electron capture dissociation? *Rapid Commun Mass Spectrom* **2007,** 21, (6), 1003-8.
134. Cooper, H. J., Investigation of the presence of b ions in electron capture dissociation mass spectra. *J Am Soc Mass Spectrom* **2005,** 16, (12), 1932-40.
135. Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A., Electron capture dissociation of singly and multiply phosphorylated peptides. *Rapid Commun Mass Spectrom* **2000,** 14, (19), 1793-800.
136. Shi, S. D.; Hemling, M. E.; Carr, S. A.; Horn, D. M.; Lindh, I.; McLafferty, F. W., Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry. *Anal Chem* **2001,** 73, (1), 19-22.
137. Mirgorodskaya, E.; Roepstorff, P.; Zubarev, R. A., Localization of O-glycosylation sites in peptides by electron capture dissociation in a Fourier transform mass spectrometer. *Anal Chem* **1999,** 71, (20), 4431-6.
138. Hakansson, K.; Cooper, H. J.; Emmett, M. R.; Costello, C. E.; Marshall, A. G.; Nilsson, C. L., Electron capture dissociation and infrared multiphoton dissociation MS/MS of an N-glycosylated tryptic peptic to yield complementary sequence information. *Anal Chem* **2001,** 73, (18), 4530-6.
139. Liu, H.; Hakansson, K., Electron capture dissociation of tyrosine O-sulfated peptides complexed with divalent metal cations. *Anal Chem* **2006,** 78, (21), 7570-6.
140. Bakhtiar, R.; Guan, Z., Electron capture dissociation mass spectrometry in characterization of post-translational modifications. *Biochem Biophys Res Commun* **2005,** 334, (1), 1-8.
141. McLuckey, S. A.; Stephenson, J. L., Jr., Ion/ion chemistry of high-mass multiply charged ions. *Mass Spectrom Rev* **1998,** 17, (6), 369-407.
142. Chalkley, R. J.; Medzihradszky, K. F.; Lynn, A. J.; Baker, P. R.; Burlingame, A. L., Statistical analysis of Peptide electron transfer dissociation fragmentation mass spectrometry. *Anal Chem* **2010,** 82, (2), 579-84.

143. Ledvina, A. R.; McAlister, G. C.; Gardner, M. W.; Smith, S. I.; Madsen, J. A.; Schwartz, J. C.; Stafford, G. C., Jr.; Syka, J. E.; Brodbelt, J. S.; Coon, J. J., Infrared photoactivation reduces peptide folding and hydrogen-atom migration following ETD tandem mass spectrometry. *Angew Chem Int Ed Engl* **2009,** 48, (45), 8526-8.
144. Anusiewicz, I.; Berdys-Kochanska, J.; Simons, J., Electron attachment step in electron capture dissociation (ECD) and electron transfer dissociation (ETD). *J Phys Chem A* **2005,** 109, (26), 5801-13.
145. Leymarie, N.; Costello, C. E.; O'Connor, P. B., Electron capture dissociation initiates a free radical reaction cascade. *J Am Chem Soc* **2003,** 125, (29), 8949-58.
146. Han, H.; Xia, Y.; McLuckey, S. A., Ion trap collisional activation of c and z* ions formed via gas-phase ion/ion electron-transfer dissociation. *J Proteome Res* **2007,** 6, (8), 3062-9.
147. Kjeldsen, F.; Haselmann, K. F.; Budnik, B. A.; Sorensen, E. S.; Zubarev, R. A., Complete characterization of posttranslational modification sites in the bovine milk protein PP3 by tandem mass spectrometry with electron capture dissociation as the last stage. *Anal Chem* **2003,** 75, (10), 2355-61.
148. Coon, J. J.; Ueberheide, B.; Syka, J. E.; Dryhurst, D. D.; Ausio, J.; Shabanowitz, J.; Hunt, D. F., Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc Natl Acad Sci U S A* **2005,** 102, (27), 9463-8.
149. Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E.; Coon, J. J., Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal Chem* **2007,** 79, (2), 477-85.
150. Pitteri, S. J.; Chrisman, P. A.; Hogan, J. M.; McLuckey, S. A., Electron transfer ion/ion reactions in a three-dimensional quadrupole ion trap: reactions of doubly and triply protonated peptides with SO2*. *Anal Chem* **2005,** 77, (6), 1831-9.
151. Pitteri, S. J.; Chrisman, P. A.; McLuckey, S. A., Electron-transfer ion/ion reactions of doubly protonated peptides: effect of elevated bath gas temperature. *Anal Chem* **2005,** 77, (17), 5662-9.
152. Breuker, K.; Oh, H.; Horn, D. M.; Cerda, B. A.; McLafferty, F. W., Detailed unfolding and folding of gaseous ubiquitin ions characterized by electron capture dissociation. *J Am Chem Soc* **2002,** 124, (22), 6407-20.
153. Horn, D. M.; Breuker, K.; Frank, A. J.; McLafferty, F. W., Kinetic intermediates in the folding of gaseous protein ions characterized by electron capture dissociation mass spectrometry. *J Am Chem Soc* **2001,** 123, (40), 9792-9.
154. Tsybin, Y. O.; Witt, M.; Baykut, G.; Kjeldsen, F.; Hakansson, P., Combined infrared multiphoton dissociation and electron capture dissociation with a hollow electron beam in Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun Mass Spectrom* **2003,** 17, (15), 1759-68.
155. Mihalca, R.; van der Burgt, Y. E.; McDonnell, L. A.; Duursma, M.; Cerjak, I.; Heck, A. J.; Heeren, R. M., Combined infrared multiphoton dissociation and electron-capture dissociation using co-linear and overlapping beams in Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun Mass Spectrom* **2006,** 20, (12), 1838-44.
156. Horn, D. M.; Ge, Y.; McLafferty, F. W., Activated ion electron capture dissociation for mass spectral sequencing of larger (42 kDa) proteins. *Anal Chem* **2000,** 72, (20), 4778-84.
157. Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E.; Shabanowitz, J.; Hunt, D. F., The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta* **2006,** 1764, (12), 1811-22.
158. Udeshi, N. D.; Shabanowitz, J.; Hunt, D. F.; Rose, K. L., Analysis of proteins and peptides on a chromatographic timescale by electron-transfer dissociation MS. *FEBS J* **2007,** 274, (24), 6269-76.
159. Udeshi, N. D.; Compton, P. D.; Shabanowitz, J.; Hunt, D. F.; Rose, K. L., Methods for analyzing peptides and proteins on a chromatographic timescale by electron-transfer dissociation mass spectrometry. *Nat Protoc* **2008,** 3, (11), 1709-17.
160. Wu, S. L.; Huhmer, A. F.; Hao, Z.; Karger, B. L., On-line LC-MS approach combining collision-induced dissociation (CID), electron-transfer dissociation (ETD), and CID of an isolated charge-

reduced species for the trace-level characterization of proteins with post-translational modifications. *J Proteome Res* **2007,** 6, (11), 4230-44.
161. Kelleher, N. L.; Zubarev, R. A.; Bush, K.; Furie, B.; Furie, B. C.; McLafferty, F. W.; Walsh, C. T., Localization of labile posttranslational modifications by electron capture dissociation: the case of gamma-carboxyglutamic acid. *Anal Chem* **1999,** 71, (19), 4250-3.
162. Vosseller, K.; Trinidad, J. C.; Chalkley, R. J.; Specht, C. G.; Thalhammer, A.; Lynn, A. J.; Snedecor, J. O.; Guan, S.; Medzihradszky, K. F.; Maltby, D. A.; Schoepfer, R.; Burlingame, A. L., O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. *Mol Cell Proteomics* **2006,** 5, (5), 923-34.
163. Chalkley, R. J.; Thalhammer, A.; Schoepfer, R.; Burlingame, A. L., Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. *Proc Natl Acad Sci U S A* **2009,** 106, (22), 8894-9.
164. Swaney, D. L.; McAlister, G. C.; Coon, J. J., Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat Methods* **2008,** 5, (11), 959-64.
165. Meng, F.; Forbes, A. J.; Miller, L. M.; Kelleher, N. L., Detection and localization of protein modifications by high resolution tandem mass spectrometry. *Mass Spectrom Rev* **2005,** 24, (2), 126-34.
166. Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., 3rd, Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* **1999,** 17, (7), 676-82.
167. Delahunty, C.; Yates, J. R., 3rd, Protein identification using 2D-LC-MS/MS. *Methods* **2005,** 35, (3), 248-55.
168. Meyer, B.; Papasotiriou, D. G.; Karas, M., 100% protein sequence coverage: a modern form of surrealism in proteomics. *Amino Acids* **2010**.
169. Mohammed, S.; Lorenzen, K.; Kerkhoven, R.; van Breukelen, B.; Vannini, A.; Cramer, P.; Heck, A. J., Multiplexed proteomics mapping of yeast RNA polymerase II and III allows near-complete sequence coverage and reveals several novel phosphorylation sites. *Anal Chem* **2008,** 80, (10), 3584-92.
170. Aye, T. T.; Scholten, A.; Taouatas, N.; Varro, A.; Van Veen, T. A.; Vos, M. A.; Heck, A. J., Proteome-wide protein concentrations in the human heart. *Mol Biosyst* **2010**.
171. Helbig, A. O.; Gauci, S.; Raijmakers, R.; van Breukelen, B.; Slijper, M.; Mohammed, S.; Heck, A. J., Profiling of N-acetylated protein termini provides in-depth insights into the N-terminal nature of the proteome. *Mol Cell Proteomics* **2010,** 9, (5), 928-39.
172. Swaney, D. L.; Wenger, C. D.; Coon, J. J., Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J Proteome Res* **2010,** 9, (3), 1323-9.
173. Ni, W.; Dai, S.; Karger, B. L.; Zhou, Z. S., Analysis of Isoaspartic Acid by Selective Proteolysis with Asp-N and Electron Transfer Dissociation Mass Spectrometry. *Anal Chem* **2010**.
174. Molina, H.; Horn, D. M.; Tang, N.; Mathivanan, S.; Pandey, A., Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A* **2007,** 104, (7), 2199-204.
175. Olsen, J. V.; Ong, S. E.; Mann, M., Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* **2004,** 3, (6), 608-14.
176. Viswanatha, T.; Liener, I. E., The inhibition of trypsin. III. Influence of urea. *J Biol Chem* **1955,** 215, (2), 777-85.
177. Viswanatha, T.; Pallansch, M. J.; Liener, I. E., The inhibition of trypsin. II. The effect of synthetic anionic detergents. *J Biol Chem* **1955,** 212, (1), 301-9.
178. Chen, E. I.; Cociorva, D.; Norris, J. L.; Yates, J. R., 3rd, Optimization of mass spectrometry-compatible surfactants for shotgun proteomics. *J Proteome Res* **2007,** 6, (7), 2529-38.
179. Klammer, A. A.; MacCoss, M. J., Effects of modified digestion schemes on the identification of proteins from complex mixtures. *J Proteome Res* **2006,** 5, (3), 695-700.

180. Pham, V. T.; Ewing, E.; Kaplan, H.; Choma, C.; Hefford, M. A., Glycation improves the thermostability of trypsin and chymotrypsin. *Biotechnol Bioeng* **2008,** 101, (3), 452-9.
181. Petritis, B. O.; Qian, W. J.; Camp, D. G., 2nd; Smith, R. D., A simple procedure for effective quenching of trypsin activity and prevention of 18O-labeling back-exchange. *J Proteome Res* **2009,** 8, (5), 2157-63.
182. Eliuk, S. M.; Maltby, D.; Panning, B.; Burlingame, A. L., High resolution electron transfer dissociation studies of unfractionated intact histones from murine embryonic stem cells using on-line capillary LC separation: determination of abundant histone isoforms and post-translational modifications. *Mol Cell Proteomics* **2010,** 9, (5), 824-37.
183. Phanstiel, D.; Brumbaugh, J.; Berggren, W. T.; Conard, K.; Feng, X.; Levenstein, M. E.; McAlister, G. C.; Thomson, J. A.; Coon, J. J., Mass spectrometry identifies and quantifies 74 unique histone H4 isoforms in differentiating human embryonic stem cells. *Proc Natl Acad Sci U S A* **2008,** 105, (11), 4093-8.
184. Chi, A.; Huttenhower, C.; Geer, L. Y.; Coon, J. J.; Syka, J. E.; Bai, D. L.; Shabanowitz, J.; Burke, D. J.; Troyanskaya, O. G.; Hunt, D. F., Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. *Proc Natl Acad Sci U S A* **2007,** 104, (7), 2193-8.
185. Crockett, D. K.; Lin, Z.; Vaughn, C. P.; Lim, M. S.; Elenitoba-Johnson, K. S., Identification of proteins from formalin-fixed paraffin-embedded cells by LC-MS/MS. *Lab Invest* **2005,** 85, (11), 1405-15.
186. Mischerikow, N.; Altelaar, A. F.; Navarro, J. D.; Mohammed, S.; Heck, A., Comparative assessment of site assignments in CID and ETD spectra of phosphopeptides discloses limited relocation of phosphate groups. *Mol Cell Proteomics* **2010**.
187. Gauci, S.; Helbig, A. O.; Slijper, M.; Krijgsveld, J.; Heck, A. J.; Mohammed, S., Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. *Anal Chem* **2009,** 81, (11), 4493-501.
188. Schlosser, A.; Vanselow, J. T.; Kramer, A., Mapping of phosphorylation sites by a multi-protease approach with specific phosphopeptide enrichment and NanoLC-MS/MS analysis. *Anal Chem* **2005,** 77, (16), 5243-50.
189. Fischer, F.; Poetsch, A., Protein cleavage strategies for an improved analysis of the membrane proteome. *Proteome Sci* **2006,** 4, 2.
190. Rao, K. C.; Palamalai, V.; Dunlevy, J. R.; Miyagi, M., Peptidyl-Lys metalloendopeptidase-catalyzed 18O labeling for comparative proteomics: application to cytokine/lipolysaccharide-treated human retinal pigment epithelium cell line. *Mol Cell Proteomics* **2005,** 4, (10), 1550-7.
191. Rao, K. C.; Carruth, R. T.; Miyagi, M., Proteolytic 18O labeling by peptidyl-Lys metalloendopeptidase for comparative proteomics. *J Proteome Res* **2005,** 4, (2), 507-14.
192. Han, J.; Schley, K. L., Proteolysis and mass spectrometric analysis of an integral membrane: aquaporin 0. *J. Proteome Res.* **2004,** 3, 807–812.
193. Lucas, F.; Barber, M.; Wolstenholme, W. A.; Geddes, A. J.; Graham, G. N.; Morris, H. R., Mass-spectrometric determination of the amino acid sequences in peptides isolated from the protein silk fibroin of Bombyx mori. *Biochem J* **1969,** 114, (4), 695-702.
194. Morris, H. R.; Batley, K. E.; Harding, N. C.; Bjur, R. A.; Dann, J. G.; King, R. W., Dihydrofolate reductase: low-resolution mass-spectrometric analysis of an elastase digest as a sequencing tool. *Biochem J* **1974,** 137, (2), 409-11.
195. Rietschel, B.; Arrey, T. N.; Meyer, B.; Bornemann, S.; Schuerken, M.; Karas, M.; Poetsch, A., Elastase digests: new ammunition for shotgun membrane proteomics. *Mol Cell Proteomics* **2009,** 8, (5), 1029-43.
196. Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Yates, J. R., 3rd, A method for the comprehensive proteomic analysis of membrane proteins. *Nat Biotechnol* **2003,** 21, (5), 532-8.

197. Schlosser, A.; Pipkorn, R.; Bossemeyer, D.; Lehmann, W. D., Analysis of protein phosphorylation by a combination of elastase digestion and neutral loss tandem mass spectrometry. *Anal Chem* **2001,** 73, (2), 170-6.
198. Schlosser, A.; Bodem, J.; Bossemeyer, D.; Grummt, I.; Lehmann, W. D., Identification of protein phosphorylation sites by combination of elastase digestion, immobilized metal affinity chromatography, and quadrupole-time of flight tandem mass spectrometry. *Proteomics* **2002,** 2, (7), 911-8.
199. MacCoss, M. J.; McDonald, W. H.; Saraf, A.; Sadygov, R.; Clark, J. M.; Tasto, J. J.; Gould, K. L.; Wolters, D.; Washburn, M.; Weiss, A.; Clark, J. I.; Yates, J. R., 3rd, Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc Natl Acad Sci U S A* **2002,** 99, (12), 7900-5.
200. Nonaka, T.; Hashimoto, Y.; Takio, K., Kinetic characterization of lysine-specific metalloendopeptidases from Grifola frondosa and Pleurotus ostreatus fruiting bodies. *J Biochem* **1998,** 124, (1), 157-62.
201. Nonaka, T.; Ishikawa, H.; Tsumuraya, Y.; Hashimoto, Y.; Dohmae, N., Characterization of a thermostable lysine-specific metalloendopeptidase from the fruiting bodies of a basidiomycete, Grifola frondosa. *J Biochem* **1995,** 118, (5), 1014-20.
202. Nonaka, T.; Dohmae, N.; Hashimoto, Y.; Takio, K., Amino acid sequences of metalloendopeptidases specific for acyl-lysine bonds from Grifola frondosa and Pleurotus ostreatus fruiting bodies. *J Biol Chem* **1997,** 272, (48), 30032-9.
203. Hori, T.; Kumasaka, T.; Yamamoto, M.; Nonaka, N.; Tanaka, N.; Hashimoto, Y.; Ueki, U.; Takio, K., Structure of a new 'aspzincin' metalloendopeptidase from Grifola frondosa: implications for the catalytic mechanism and substrate specificity based on several different crystal forms. *Acta Crystallogr D Biol Crystallogr* **2001,** 57, (Pt 3), 361-8.
204. Edman, P., Method for the determination of the amino acid sequence of peptides. *Acta Chem. Scand* **1950,** 4, 283-293.
205. Biemann, K., Sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation. *Methods Enzymol* **1990,** 193, 455-79.
206. Hunt, D. F.; Buko, A. M.; Ballard, J. M.; Shabanowitz, J.; Giordani, A. B., Sequence analysis of polypeptides by collision activated dissociation on a triple quadrupole mass spectrometer. *Biomed Mass Spectrom* **1981,** 8, (9), 397-408.
207. Hunt, D. F.; Yates, J. R., 3rd; Shabanowitz, J.; Winston, S.; Hauer, C. R., Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci U S A* **1986,** 83, (17), 6233-7.
208. Standing, K. G., Peptide and protein de novo sequencing by mass spectrometry. *Curr Opin Struct Biol* **2003,** 13, (5), 595-601.
209. Chi, H.; Sun, R. X.; Yang, B.; Song, C. Q.; Wang, L. H.; Liu, C.; Fu, Y.; Yuan, Z. F.; Wang, H. P.; He, S. M.; Dong, M. Q., pNovo: de novo peptide sequencing and identification using HCD spectra. *J Proteome Res* **2010,** 9, (5), 2713-24.
210. Keough, T.; Lacey, M. P.; Youngquist, R. S., Derivatization procedures to facilitate de novo sequencing of lysine-terminated tryptic peptides using postsource decay matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom* **2000,** 14, (24), 2348-56.
211. Keough, T.; Youngquist, R. S.; Lacey, M. P., A method for high-sensitivity peptide sequencing using postsource decay matrix-assisted laser desorption ionization mass spectrometry. *Proc Natl Acad Sci U S A* **1999,** 96, (13), 7131-6.
212. Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G., PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* **2003,** 17, (20), 2337-42.
213. Frank, A.; Pevzner, P., PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal Chem* **2005,** 77, (4), 964-73.
214. Dancik, V.; Addona, T. A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A., De novo peptide sequencing via tandem mass spectrometry. *J Comput Biol* **1999,** 6, (3-4), 327-42.

215. Fernandez-de-Cossio, J.; Gonzalez, J.; Betancourt, L.; Besada, V.; Padron, G.; Shimonishi, Y.; Takao, T., Automated interpretation of high-energy collision-induced dissociation spectra of singly protonated peptides by 'SeqMS', a software aid for de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* **1998,** 12, (23), 1867-78.
216. Fernandez-de-Cossio, J.; Gonzalez, J.; Satomi, Y.; Shima, T.; Okumura, N.; Besada, V.; Betancourt, L.; Padron, G.; Shimonishi, Y.; Takao, T., Automated interpretation of low-energy collision-induced dissociation spectra by SeqMS, a software aid for de novo sequencing by tandem mass spectrometry. *Electrophoresis* **2000,** 21, (9), 1694-9.
217. Taylor, J. A.; Johnson, R. S., Sequence database searches via de novo peptide sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* **1997,** 11, (9), 1067-75.
218. Mo, L.; Dutta, D.; Wan, Y.; Chen, T., MSNovo: a dynamic programming algorithm for de novo peptide sequencing via tandem mass spectrometry. *Anal Chem* **2007,** 79, (13), 4870-8.
219. Jagannath, S.; Sabareesh, V., Peptide Fragment Ion Analyser (PFIA): a simple and versatile tool for the interpretation of tandem mass spectrometric data and de novo sequencing of peptides. *Rapid Commun Mass Spectrom* **2007,** 21, (18), 3033-8.
220. Grossmann, J.; Roos, F. F.; Cieliebak, M.; Liptak, Z.; Mathis, L. K.; Muller, M.; Gruissem, W.; Baginsky, S., AUDENS: a tool for automated peptide de novo sequencing. *J Proteome Res* **2005,** 4, (5), 1768-74.
221. Fischer, B.; Roth, V.; Roos, F.; Grossmann, J.; Baginsky, S.; Widmayer, P.; Gruissem, W.; Buhmann, J. M., NovoHMM: a hidden Markov model for de novo peptide sequencing. *Anal Chem* **2005,** 77, (22), 7265-73.
222. Fu, Q.; Li, L., De novo sequencing of neuropeptides using reductive isotopic methylation and investigation of ESI QTOF MS/MS fragmentation pattern of neuropeptides with N-terminal dimethylation. *Anal Chem* **2005,** 77, (23), 7783-95.
223. Savitski, M. M.; Nielsen, M. L.; Kjeldsen, F.; Zubarev, R. A., Proteomics-grade de novo sequencing approach. *J Proteome Res* **2005,** 4, (6), 2348-54.
224. Spengler, B., De novo sequencing, peptide composition analysis, and composition-based sequencing: a new strategy employing accurate mass determination by fourier transform ion cyclotron resonance mass spectrometry. *J Am Soc Mass Spectrom* **2004,** 15, (5), 703-14.
225. Zubarev, R. A.; Zubarev, A. R.; Savitski, M. M., Electron capture/transfer versus collisionally activated/induced dissociations: solo or duet? *J Am Soc Mass Spectrom* **2008,** 19, (6), 753-61.
226. Horn, D. M.; Zubarev, R. A.; McLafferty, F. W., Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry. *Proc Natl Acad Sci U S A* **2000,** 97, (19), 10313-7.
227. Datta, R.; Bern, M., Spectrum fusion: using multiple mass spectra for de novo Peptide sequencing. *J Comput Biol* **2009,** 16, (8), 1169-82.
228. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999,** 20, (18), 3551-67.
229. Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H., Open mass spectrometry search algorithm. *J Proteome Res* **2004,** 3, (5), 958-64.
230. Craig, R.; Beavis, R. C., TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **2004,** 20, (9), 1466-7.
231. Matthiesen, R.; Lundsgaard, M.; Welinder, K. G.; Bauw, G., Interpreting peptide mass spectra by VEMS. *Bioinformatics* **2003,** 19, (6), 792-3.
232. Clauser, K. R.; Baker, P.; Burlingame, A. L., Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal Chem* **1999,** 71, (14), 2871-82.
233. Sadygov, R. G.; Cociorva, D.; Yates, J. R., 3rd, Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. *Nat Methods* **2004,** 1, (3), 195-202.

234. Chamrad, D. C.; Korting, G.; Stuhler, K.; Meyer, H. E.; Klose, J.; Bluggel, M., Evaluation of algorithms for protein identification from sequence databases using mass spectrometry data. *Proteomics* **2004,** 4, (3), 619-28.
235. Kapp, E. A.; Schutz, F.; Connolly, L. M.; Chakel, J. A.; Meza, J. E.; Miller, C. A.; Fenyo, D.; Eng, J. K.; Adkins, J. N.; Omenn, G. S.; Simpson, R. J., An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis. *Proteomics* **2005,** 5, (13), 3475-90.
236. Balgley, B. M.; Laudeman, T.; Yang, L.; Song, T.; Lee, C. S., Comparative evaluation of tandem MS search algorithms using a target-decoy search strategy. *Mol Cell Proteomics* **2007,** 6, (9), 1599-608.
237. Zubarev, R. A., Electron-capture dissociation tandem mass spectrometry. *Curr Opin Biotechnol* **2004,** 15, (1), 12-6.
238. Cooper, H. J.; Hudgins, R. R.; Hakansson, K.; Marshall, A. G., Characterization of amino acid side chain losses in electron capture dissociation. *J Am Soc Mass Spectrom* **2002,** 13, (3), 241-9.
239. Fung, Y. M.; Chan, T. W., Experimental and theoretical investigations of the loss of amino acid side chains in electron capture dissociation of model peptides. *J Am Soc Mass Spectrom* **2005,** 16, (9), 1523-35.
240. Kjeldsen, F.; Giessing, A. M.; Ingrell, C. R.; Jensen, O. N., Peptide sequencing and characterization of post-translational modifications by enhanced ion-charging and liquid chromatography electron-transfer dissociation tandem mass spectrometry. *Anal Chem* **2007,** 79, (24), 9243-52.
241. Kalli, A.; Hakansson, K., Comparison of the electron capture dissociation fragmentation behavior of doubly and triply protonated peptides from trypsin, Glu-C, and chymotrypsin digestion. *J Proteome Res* **2008,** 7, (7), 2834-44.
242. Cooper, J. W.; Wang, Y.; Lee, C. S., Recent advances in capillary separations for proteomics. *Electrophoresis* **2004,** 25, (23-24), 3913-26.
243. Ogiso, H.; Taguchi, R., Reversed-phase LC/MS method for polyphosphoinositide analyses: changes in molecular species levels during epidermal growth factor activation in A431 cells. *Anal Chem* **2008,** 80, (23), 9226-32.
244. Dillon, T. M.; Bondarenko, P. V.; Rehder, D. S.; Pipes, G. D.; Kleemann, G. R.; Ricci, M. S., Optimization of a reversed-phase high-performance liquid chromatography/mass spectrometry method for characterizing recombinant antibody heterogeneity and stability. *J Chromatogr A* **2006,** 1120, (1-2), 112-20.
245. Yang, F. Q.; Li, D. Q.; Feng, K.; Hu, D. J.; Li, S. P., Determination of nucleotides, nucleosides and their transformation products in Cordyceps by ion-pairing reversed-phase liquid chromatography-mass spectrometry. *J Chromatogr A* **2010,** 1217, (34), 5501-10.
246. Keppel, T. R.; Jacques, M. E.; Weis, D. D., The use of acetone as a substitute for acetonitrile in analysis of peptides by liquid chromatography/electrospray ionization mass spectrometry. *Rapid Commun Mass Spectrom* **2010,** 24, (1), 6-10.
247. Shen, Y.; Zhao, R.; Berger, S. J.; Anderson, G. A.; Rodriguez, N.; Smith, R. D., High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics. *Anal Chem* **2002,** 74, (16), 4235-49.
248. Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C., Orthogonality of separation in two-dimensional liquid chromatography. *Anal Chem* **2005,** 77, (19), 6426-34.
249. Alpert, A. J.; Andrews, P. C., Cation-exchange chromatography of peptides on poly(2-sulfoethyl aspartamide)-silica. *J Chromatogr* **1988,** 443, 85-96.
250. Burke, T. W.; Mant, C. T.; Black, J. A.; Hodges, R. S., Strong cation-exchange high-performance liquid chromatography of peptides. Effect of non-specific hydrophobic interactions and linearization of peptide retention behaviour. *J Chromatogr* **1989,** 476, 377-89.
251. Li, Y.; Xiang, R.; Wilkins, J. A.; Horvath, C., Capillary electrochromatography of peptides and proteins. *Electrophoresis* **2004,** 25, (14), 2242-56.

252. Scherer, B.; Steiner, F., Application of hydrophobic anion-exchange phases in capillary electrochromatography. *J Chromatogr A* **2001,** 924, (1-2), 197-209.
253. Klampfl, C. W.; Haddad, P. R., Behavior of basic compounds in ion-exchange capillary electrochromatography with low-pH carrier electrolytes. *J Chromatogr A* **2000,** 884, (1-2), 277-85.
254. Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **2001,** 19, (3), 242-7.
255. Boersema, P. J.; Divecha, N.; Heck, A. J.; Mohammed, S., Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. *J Proteome Res* **2007,** 6, (3), 937-46.
256. Delahunty, C. M.; Yates, J. R., 3rd, MudPIT: multidimensional protein identification technology. *Biotechniques* **2007,** 43, (5), 563, 565, 567 passim.
257. Motoyama, A.; Xu, T.; Ruse, C. I.; Wohlschlegel, J. A.; Yates, J. R., 3rd, Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. *Anal Chem* **2007,** 79, (10), 3623-34.
258. McDonald, W. H.; Ohi, R.; Miyamoto, D. T.; Mitchison, T. J.; Yates, J. R., Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. *International Journal of Mass Spectrometry* **2002,** 219, (1), 245-251
259. Villen, J.; Gygi, S. P., The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat Protoc* **2008,** 3, (10), 1630-8.
260. Gruhler, A.; Olsen, J. V.; Mohammed, S.; Mortensen, P.; Faergeman, N. J.; Mann, M.; Jensen, O. N., Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics* **2005,** 4, (3), 310-27.
261. Gnad, F.; de Godoy, L. M.; Cox, J.; Neuhauser, N.; Ren, S.; Olsen, J. V.; Mann, M., High-accuracy identification and bioinformatic analysis of in vivo protein phosphorylation sites in yeast. *Proteomics* **2009,** 9, (20), 4642-52.
262. Dormeyer, W.; Mohammed, S.; Breukelen, B.; Krijgsveld, J.; Heck, A. J., Targeted analysis of protein termini. *J Proteome Res* **2007,** 6, (12), 4634-45.
263. Xia, Y. Q.; Miller, J. D.; Bakhtiar, R.; Franklin, R. B.; Liu, D. Q., Use of a quadrupole linear ion trap mass spectrometer in metabolite identification and bioanalysis. *Rapid Commun Mass Spectrom* **2003,** 17, (11), 1137-45.
264. Cagney, G.; Park, S.; Chung, C.; Tong, B.; O'Dushlaine, C.; Shields, D. C.; Emili, A., Human tissue profiling with multidimensional protein identification technology. *J Proteome Res* **2005,** 4, (5), 1757-67.
265. Kislinger, T.; Gramolini, A. O.; MacLennan, D. H.; Emili, A., Multidimensional protein identification technology (MudPIT): technical overview of a profiling method optimized for the comprehensive proteomic investigation of normal and diseased heart tissue. *J Am Soc Mass Spectrom* **2005,** 16, (8), 1207-20.
266. Chakraborty, A. B.; Berger, S. J.; Gebler, J. C., Use of an integrated MS--multiplexed MS/MS data acquisition strategy for high-coverage peptide mapping studies. *Rapid Commun Mass Spectrom* **2007,** 21, (5), 730-44.
267. de Godoy, L. M.; Olsen, J. V.; de Souza, G. A.; Li, G.; Mortensen, P.; Mann, M., Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol* **2006,** 7, (6), R50.
268. Liu, H. B.; Sadygov, R. G.; Yates, J. R., A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **2004,** 76, 4193-4201.

Chapter 2

Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase

Nadia Taouatas, Madalina M Drugan, Albert J R Heck & Shabaz Mohammed

Biomolecular Mass Spectrometry and Proteomics group, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, the Netherlands.

Based on Nature Methods. 2008, (5), 405-407.

Abstract

We introduce a method for sequencing peptides by mass spectrometry using a metalloendopeptidase that cleaves proteins at the amino side of lysine (Lys-N). When analyzed by electron transfer dissociation (ETD) – based mass spectrometric sequencing, Lys-N – digested peptides that contain a single lysine residue produce spectra dominated by *c*-type fragment ions, providing simple ladders for sequence determination. This method should be a valuable strategy for *de novo* sequencing and the analysis of post-translational modifications.

Abbreviations used:

CID - collision induced dissociation; ESI – electrospray ionization; ETD - electron transfer dissociation; ETcaD - electron transfer collisional activated dissociation; HEK - human embryonic kidney; HPLC - high pressure liquid chromatography; LC-MS/MS - liquid chromatography tandem mass spectrometry; MS - mass spectrometry; nanoLC - nanoflow liquid chromatography; SCX - strong cation exchange.

Introduction

Mass spectrometry based protein identification and characterization is now a mature and routinely used technology[1]. At present the preferred analytical method for protein identification proceeds via trypsin proteolysis followed by nanoliter flow reversed phase liquid chromatography (nanoLC) peptide separation. Identification is performed by collision-induced dissociation (CID) tandem mass spectrometry of the peptide fragments[2]. Although CID spectra of tryptic peptides generate many fragment ions that can be used for identification in automated database search protocols, the magnitude and thus the complexity of fragment ion signals is not always useful when peptides need to be identified by *de novo* sequencing[3], such as when the parent protein for the proteolytic peptide is not present in a database - for example, for a previously unreported protein isoform or when the protein originates from a species with an unknown genome. The recently introduced peptide fragmentation methods electron capture dissociation and electron transfer–induced dissociation (ETD) have proven to be complementary to CID as they are better suited for sequencing larger and more basic peptides (such as histones) and are now becoming more established technologies[4,5]. Several groups have explored alternative proteases for mass spectrometry analysis of proteolytic peptides[6,7]. In particular, for ETD applications, Lys-C, an endoprotease that cleaves proteins at the carboxy side of lysine, might offer a useful alternative to trypsin as it creates large peptides with tryptic peptide-like features[6,7]. Lys-C generated peptides attain on average higher charges in the electrospray ionization process owing to the possible presence of multiple arginine residues, which is welcomed as it improves electron capture efficiency thus providing richer peptide sequence data. Here we explored the application of Lys-N in combination with ETD, compared to the use of Lys-C and trypsin.

Materials and methods

Materials.

Protease inhibitor cocktail and Lys-C were obtained from Roche Diagnostics (Mannheim. Germany). Metalloendopeptidase from Grifola Frondosa (Lys-N) was obtained from Seikagaku Corporation (Tokyo, Japan). Bovine Serum Albumin (BSA) and Iodoacetamide were obtained from Sigma-Aldrich (Steinheim, Germany). DLDithiothreitol was obtained

from Fluka biochemical (Steinheim, Germany). HEK293 cells were a gift from the ABC Protein Expression Center (Utrecht University, The Netherlands). HPLC-S gradient grade acetonitrile was purchased from Biosolve (Valkenwaard, The Netherlands). Acetic acid was obtained by MERCK KGaA (Damstadt, Germany) and high purity water obtained from Milli-Q system (Millipore, Bedford, MA).

Human Embryonic kidney (HEK) 293T cells.

HEK293 cells were harvested at a density of approx 1.5 x 106 cells/mL and stored at -30 °C. Cells were thawed and resuspended in ice-cold lysisbuffer (15 mL PBS, 150 µl Tween 20 and protease inhibitor cocktail). After daunce homogenizing on ice, the lysate was stored at 0°C for 10 min. Subsequently centrifugation at 20000 x g in a tabletop centrifuge (Eppendorf, Hamburg, Germany) at 4°C yielded separation of soluble and insoluble protein fractions. The soluble fraction was collected and the concentration determined by a Bradford assay. The lysate was dissolved in 50 mM ammonium bicarbonate to a concentration of 4mg/mL.

Lysate and BSA in-solution digestion.

100 mg digested protein/lysate was reduced with 45 mM dithiothreitol (50 °C, 15 min) followed by alkylation using 110 mM iodoacetamide (dark, RT, 15min). Buffer exchange was performed with 50 mM ammonium bicarbonate using 5 kD spin columns. The resulting solutions were dried in a vacuum centrifuge and resuspended in 50 mM ammonium bicarbonate. The purified digests were aliquoted. One part was digested with Lys-C and an equal amount with Lys-N. Lys-C was added to the samples at a 1:50 (w/w) ratio and incubated at 25 °C over night and Lys-N was added at a ratio of 1:85 (w/w) and also incubated over night at 25 °C.

BSA in-gel digestion.

Gel bands containing 5 µg of BSA were cut out of the gel and washed with water. After shrinking the gel pieces with acetonitrile they were reduced with 10 mM of dithiothreitol (60 °C, 1 hour) followed by alkylation using 55 mM iodoacetamide (dark, RT, 30min). After shrinking the gel pieces with acetonitrile they were incubated with trypsin or Lys-N (10 ng / µl) and left on ice for 30 min. Excess trypsin or Lys-N was then removed and enough 50 mM ammonium bicarbonate was added to cover the gel pieces. The gel pieces were incubated over night at 37 °C. Supernatant was transferred to new eppendorf tubes. Peptides were

extracted by adding 5 % formic acid to the gel pieces following heating (65 °C, 2 min) and shaking (RT, 20 min). The supernatant was added to the previous supernatant.

Strong Cation Exchange.

Strong cation exchange was performed using a zorbax BioSCX-Series II column (0.8 mm i.d. X 50 mm length, 3.5µm), a FAMOS autosampler (LCpacking, Amsterdam, The Netherlands), a Shimadzu LC-9A binary pump and a SPD-6A UV detector (Shimadzu, Tokyo, Japan). Prior to SCX chromatography, protein digests were desalted using a small plug C18 material (3 M Empore C18 extraction disk) packed into a GELoader tip (Eppendorf) onto which 10 µL of Aqua C18 (5 µm, 200 Å) material was placed. The eluate was dried completely and subsequently reconstituted in 20 % acetonitrile and 0.05 % formic acid. After injection, a linear gradient of 1% min –1 solvent B (500 mM KCL in 20 % acetonitrile and 0.05% formic acid, pH 3.0) was performed. A total of 30 SCX fractions (1 min each, i.e., 50 µL elution volume) were manually collected and dried in a vacuum centrifuge.

ETD experiments.

Digested BSA and HEK293 lysate were subjected to nanoscale liquid chromatography tandem mass spectrometry (nanoLC-MS/MS) analysis, performed on an Agilent 1100 HPLC system (Agilent Technologies) connected to a LTQ XL Linear Ion Trap Mass Spectrometer with an ETD source at the back from Thermo Fisher Scientific, Inc. (Waltham, MA). The samples (0.5µg digested HEK293 lysate and 50 fmol of BSA) were diluted in 5 % formic acid and injected on the trap column (Aqua C18 (phenomenex, Torrance, CA)), 20 mm x 100 µm I.D.) at a flow rate of 5 µL / min. The peptides were transferred with a split-reduced flow rate of 100 nL / min solvent A (0.1 M acetic acid) onto an analytic column (Reprosil C18 RP (Dr Maisch, Germany), 20 cm x 50 µm I.D.). Elution of peptides from digested lysate was achieved with a linear gradient from 0 to 60% B (acetonitrile/water (v/v) containing 0.1 M acetic acid) in 95 min and digested BSA was eluted with a linear gradient from 0 to 40% B in 75min. The column effluent was directly introduced into the ESI source of the MS. The mass spectrometer was operated in positive ion mode, from 350 to 1500 m/z in MS mode and with an AGC value of 1.00e+05. Parent ions were isolated for a more accurate mass measurement by performing a SIM scan and fragmented by CID or ETD in data-dependent mode with an AGC value of 1.00e+04. Ions were fragmented using CID with normalized collision energy of 35 and 30 ms activation time. ETD fragmentation was performed with supplemental activation, fluoranthene was used as reagent anion and ion/ion reaction in the ion trap was

taking place for 100 ms. After MS measurements data was analyzed with the MASCOT software version 2.2.0 (www.matrixscience.com). The database search was made with parameters set to consider a peptide tolerance of ± 0.5 Da, a fragment tolerance of ± 0.9 Da, a static modification of + 57 Da on cystein residues (Carbamidomethylation) and a differential modification of + 16 Da on methionine (oxidation). CID and ETD spectra of digested lysate were searched in NCBInr 20070713 (5269953 sequences; 1825351362 residues) database and BSA in a BSA database.

Lys-N and Lys-C fragment ion statistics.

Peptides with a minimum MASCOT score of 30 were considered. The Mascot identification allowed automated peak counting and intensity summing where all possible c and z ions (plus known common neutral losses) were considered. Exceptions occurred when c- or z-ions (or related ions) were assigned to the same isotope cluster in which case the most appropriate assignment was chosen (e.g. based on mass accuracy trend, mono isotopic peak etc) in an automated fashion.

Results and discussion

We first compared Lys-N and Lys-C using the standard protein bovine serum albumin (BSA). We subjected the generated proteolytic peptides of BSA to CID and ETD (Figure 1). In all ETD experiments we applied supplemental collisional activation, often referred to as ETcaD[9]. Lys-C proteolytic peptides generally contain a basic entity at both termini (Figure 1a) that usually becomes protonated in the electrospray ionization process, leading to doubly charged peptide ions[8]. A typical ETcaD fragmentation spectrum of a doubly charged Lys-C proteolytic peptide ion (Figure 1c) reveals the formation of multiple *c* and *z* fragment ions with the *z* ions being dominant[9,10]. When performing ETcaD on doubly charged peptide ions, a single charge remains after the induced charge reduction, and thus the position of this remaining charge dictates whether *c* or *z* fragment ions will be formed (Figure 1a).

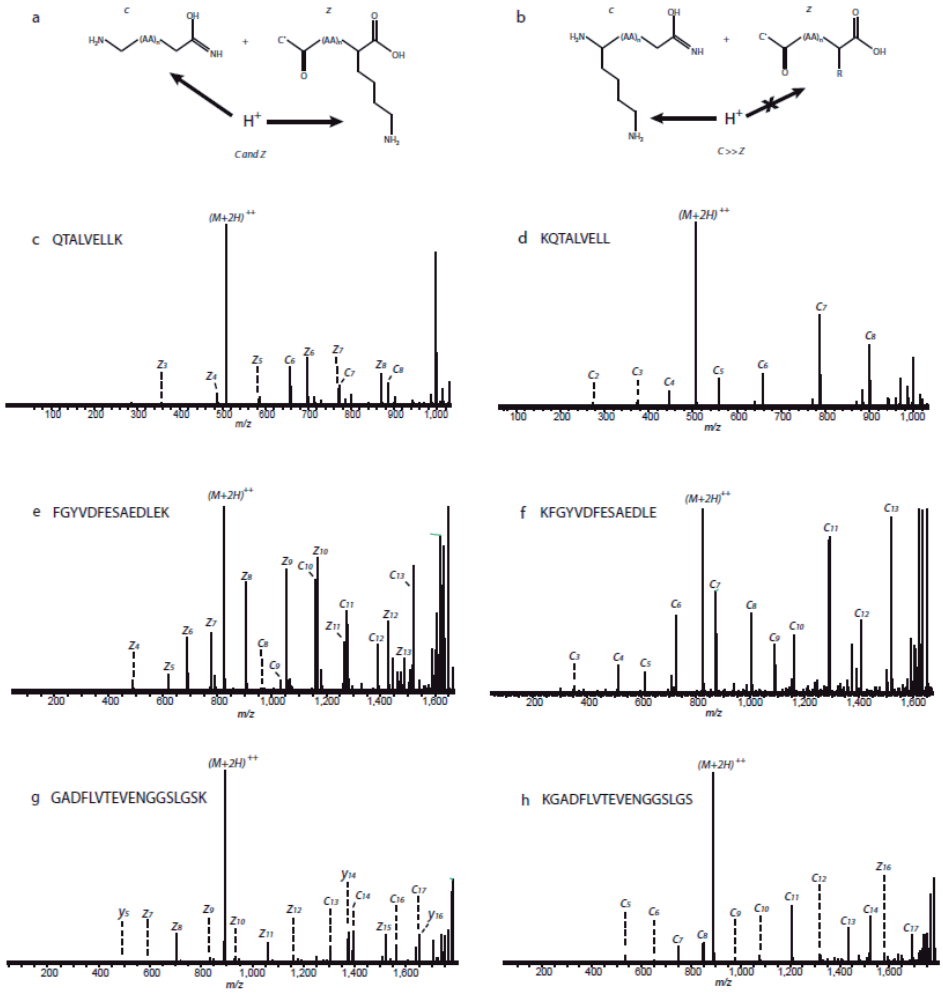

Figure 1: ETcaD peptide fragmentation spectra of doubly charged ions originating from Lys-C and Lys-N digested peptides. (a,b) Simplified *c* and *z* structures for proteolytic peptide fragment ions resulting from Lys-C (a) and Lys-N (b) digest. (c–h) ETcaD peptide fragmentation mass spectra for Lys-C-digested (c,e,g) and Lys-N-digested (d,f,h) doubly charged peptide ions from BSA (c,d), human nucleolin (e,f), human pyruvate kinase (g,h).

Like Lys-C, Lys-N has an enzymatic selectivity for lysine residues except that the cleavage occurs on the N-terminal side[11,12]. Lys-N proteolytic peptides that contain a single lysine will have two basic entities at the N-terminal site (Figure 1b). When we subjected these Lys-N proteolytic peptides to ETcaD, the vast majority of fragment ions we observed were N-

terminal *c* ions (Figure 1d). We rationalize that the strong basic nature of the N terminus creates a strong attraction for the remaining proton (Figure 1b). This behavior is somewhat similar to that observed for singly charged peptides formed in matrix-assisted laser desorption/ionization where, if there is a highly basic entity present at the C terminus, CID fragmentation produces preferentially *y* ions[13]. Using Lys-N in combination with ETcaD, we observed cleavage at almost every amide bond in the peptide backbone, yielding easy-to-interpret sequence ladders. The signal intensity of the fragment ions seemed to diminish with decreasing *m/z*, as observed previously[9]. We provide a rough comparison of the performance of Lys-N and Lys-C in Table 1.

Table 1: Summary of Lys-C and Lys-N digests using BSA as the analyte.

	Suitable for in-gel digestion	Obtained sequence coverage	Number of peptides identified	Number of peptides with no arginine or histidine (dominated by c-ion series)	Protease / protein ratio used
Lys-C	No	59%	28	19 (0)	1:50
Lys-N	Yes	56%	27	14 (14)	1:85

Many protein analyses are performed on analytes that originate from SDS-PAGE separations. Therefore, we tested the applicability of the Lys-N for in-gel (one-dimensional SDS-PAGE) digestion. As Lys-C is not easily adaptable to in-gel digestion, we compared the performance of Lys-N with trypsin. Lys-N proved to be at least as efficient as trypsin for in-gel proteolysis of BSA. To test the general applicability of using Lys-N in "shotgun proteomics" and its feasibility of using it to digest complex biological samples, we digested a HEK293 cell lysate with Lys-N and Lys-C in parallel. We separated the resulting complex peptide mixtures in single nanoLC - mass spectrometry (nanoLC-MS) runs and analyzed them by both ETcaD and CID-based peptide sequencing. Counting the frequency of the various types of peptides (Figure 2 and Supplementary Tables 1 and 2) indicated that the number of missed cleavages observed for Lys-N was lower than that for Lys-C in our data, although the total number of peptides observed remained comparable. In contrast to typical spectra originating from the Lys-C digested peptides (Figure 1e,g), the spectra originating from Lys-N digested peptides were heavily dominated by *c* ions (Figure 1f,h). The overall features of the spectra were quite similar and resembled those observed for BSA (Figure 1c,d). The observed behavior was

general based on data in Supplementary Figure 1 and the full dataset (Proteomics Identifications Database (PRIDE: http://www.ebi.ac.uk/pride/) accession numbers 3268–3278).

We performed a fragment-ion count for doubly charged and triply charged ions on the dataset obtained from the Lys-C and Lys-N digested cell lysates (Figure 2 and Supplementary Tables 1 and 2). Fragment-ion peak counting revealed that in a set of nearly 100 Lys-N digested peptides that contained a single basic residue (lysine), just over 90% of all fragment ions were *c* ions. In the analogous Lys-C dataset, the dominant *z* ions represented only 62% of the fragment-ion population. However, this specificity was largely lost when the Lys-N–digested peptides contained an additional lysine or other basic residues, that is, arginine or histidine (Figure 2, Supplementary Figure 2 and Supplementary Tables 1 and 2). Also, triply charged Lys-C– and Lys-N–digested peptides possessed a more equal abundance of *c* and *z* ions (Figure 2, Supplementary Figure 2 and Supplementary Table 2).

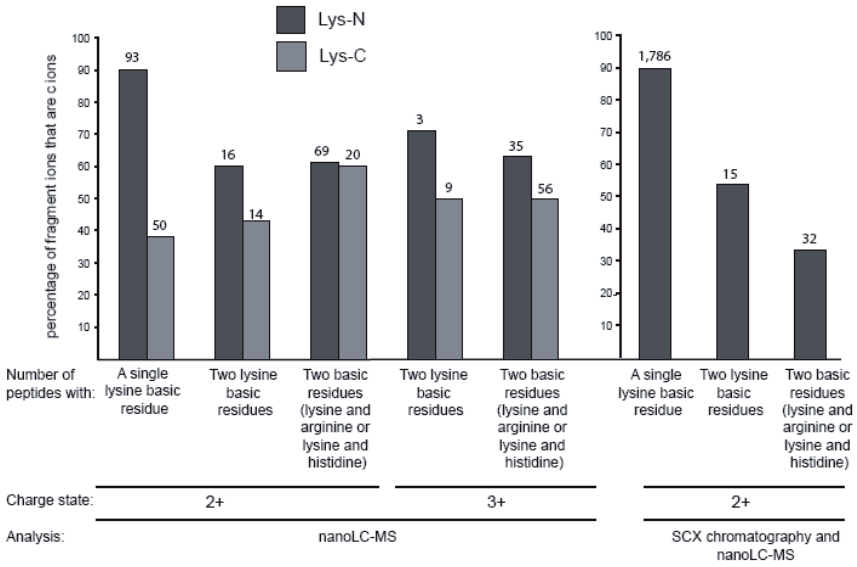

Figure 2: Propensity of *c*-ion formation by ETcaD fragmentation for peptides originating from Lys-N and Lys-C digests. Each bar represents a specific peptide population based on charge state and number of basic residues from single nanoLC-MS analyses of HEK 293 Lys-N and Lys-C cell lysate digests (left) and from nanoLC-MS analyses of a HEK 293 Lys-N cell lysate digest enriched for the population of peptides that contain a single basic residue (a lysine at the N terminus), in an initial low-pH SCX chromatography fractionation (right). On top of each bar is the number of peptides identified.

To evaluate the feasibility of applying the Lys-N digest approach for global proteome analysis strategies, we performed a large-scale targeted multidimensional analysis. We fractionated the Lys-N–digested HEK293 cell lysate by low-pH strong cation exchange (SCX) chromatography[14], a separation that is primarily dictated by the number of basic moieties present in the peptides and thus facilitates the targeted enrichment of peptides that contain a single basic residue. We subjected six fractions, rich in single lysine–containing peptides, to nanoLC-MS where each peptide chosen by the mass spectrometer for sequencing was subjected to both ETcaD and CID peptide ion fragmentation. We identified nearly 1,900 peptides, of which ~1,800 contained a single lysine residue (Figure 2 and Supplementary Table 3). We confirmed, via fragment ion statistics, the domination of *c* ions for these peptides as this ion series represented over 90% of all fragment ions. Taking signal intensity into account, we found that the *c*-ion population represented over 98% of the total fragment ion signal (data not shown). Furthermore, on average 75% of the sequence of these peptides was directly represented by *c* ions, and this coverage increased to 85% if we allowed a single *c* ion to be missed in a sequence, underscoring the high potential of the combination of Lys-N proteolysis and ETD for facile spectral interpretation and *de novo* sequencing.

The Lys-C-digested doubly charged peptides fragment under ETcaD conditions into a complex mixture of both *c* and *z* ions and are therefore more difficult to interpret *de novo*, whereas equivalent Lys-N generated peptides are dominated by *c* ion sequence ladders. For the BSA peptide fragmentation spectra shown in Figure 1c,d, the spectra interpretation software Mascot generated similar scores and levels of confidence (information present in the PRIDE dataset). This can be attributed to the fact that the number of peaks matched in both cases is rather similar. However, one could envision that a database search strategy specifically targeted at the unique fragmentation behavior observed under ETcaD conditions for Lys-N generated peptides could lead to a marked improvement in the confident assignment of these spectra.

We also analyzed the CID spectra of all these peptides and observed that peptides generated by Lys-C proteolysis fragmented into the well-known *b* and *y* ions series, *y* being prominent, whereas peptides produced by Lys-N digest yielded preferentially *b* ions (Supplementary Figure 3). However, the tendency of Lys-N–digested peptides to uniquely form one type of ion is much more pronounced in ETD compared to CID (Supplementary Figure 4), most likely because of the single proton/charge present for peptide fragments under ETD conditions.

Conclusion

In terms of sensitivity, selectivity and applicability Lys-N performs as well as currently used proteases and is also adaptable for in-gel digestion. Notably, we found that Lys-N, used in combination with ETcaD, provides a clear advantage over other proteases for spectral interpretation because fragmentation spectra of doubly charged peptides consist almost exclusively of *c*-type fragment ions, generating simple sequence ladders of the peptides of interest. We show that these peptides can be enriched for by SCX chromatography, and therefore propose that a combination of Lys-N proteolysis, SCX chromatography and ETcaD, will provide a valuable targeted strategy for the analysis of post-translational modifications and for *de novo* sequencing.

Acknowledgement

We thank A.F.M Altelaar and B. van Breukelen for fruitful discussions and support. This work was supported by the Netherlands Proteomics Centre.

Supplementary data:

Supplementary Figure 1: Further illustrative examples of MS/MS spectra from ETcaD analysis of Lys-N digested doubly charged peptides derived from the HEK293 cell lysate.

Supplementary Figure 2: ETcaD peptide ion fragmentation spectra of doubly and triply charged ions originating from Lys-N generated peptides from HEK293 cells.

Supplementary Figure 3: Typical CID peptide ion fragmentation spectra of doubly charged ions originating from Lys-C and Lys-N generated peptides from the HEK293 cell lysate.

Supplementary Figure 4: Typical CID and ETcaD peptide ion fragmentation spectra of a Lys-N generated doubly charged peptide ion.

Supplementary Table 1: A count of *c* and *z* ion (and related neutral losses) abundance ratios for doubly charged peptides originating from Lys-C and Lys-N HEK 293 cell lysate digests.

Supplementary Table 2: A count of *c* and *z* ion (and related neutral losses) abundance ratios for triply charged peptides originating from Lys-C and Lys-N HEK 293 cell lysate digests.

Supplementary Table 3: A count of *c* and *z* ion (and related neutral losses) abundance ratios for doubly charged peptides originating from Lys-C and Lys-N HEK 293 cell lysate digests initially fractionated by low pH SCX.

References

1. Aebersold, R. & Mann, M. *Nature* **422**, 198–207 (2003).
2. Witze, E.S., Old, W.M., Resing, K.A. & Ahn, N.G. *Nat. Methods* **4**, 798–806 (2007).
3. Standing, K.G. *Curr. Opin. Struct. Biol.* **13**, 595–601 (2003).
4. Zubarev, R.A., Kelleher, N.L. & McLafferty, F.W. *J. Am. Chem. Soc.* **120**, 3265–3266 (1998).
5. Syka, J.E.P., Coon, J.J., Schroeder, M.J., Shabanowitz, J. & Hunt, D.F. *Proc. Natl. Acad. Sci. USA* **101**, 9528–9533 (2004).
6. Molina, H., Horn, D.M., Tang, N., Mathivanan, S. & Pandey, A. *Proc. Natl. Acad. Sci. USA* **104**, 2199–2204 (2007).
7. Good, D.M., Wirtala, M., McAlister, G.C. & Coon, J.J. *Mol. Cell. Proteomics* **6**, 1942–1951 (2007).
8. Wysocki, V.H., Tsaprailis, G., Smith, L.L. & Breci, L.A. *J. Mass Spectrom.* **35**, 1399–1406 (2000).
9. Swaney, D.L. *et al. Anal. Chem.* **79**, 477–485 (2007).
10. Han, H., Xia, Y. & McLuckey, S.A. *J. Proteome Res.* **6**, 3062–3069 (2007).
11. Nonaka, T., Hashimoto, Y. & Takio, K. *J. Biochem.* **124**, 157–162 (1998).
12. Rao, K.C.S., Palamalai, V., Dunlevy, J.R. & Miyagi, M. *Mol. Cell. Proteomics* **4**, 1550–1557 (2005).
13. Peters, E.C., Horn, D.M., Tully, D.C. & Brock, A. *Rapid Commun. Mass Spectrom.* **15**, 2387–2392 (2001).
14. Beausoleil, S.A. *et al. Proc. Natl. Acad. Sci. USA* **101**, 12130–12135 (2004).

Chapter 3

Strong cation exchange-based fractionation of Lys-N-generated peptides facilitates the targeted analysis of post-translational modifications.

Nadia Taouatas[1], A.F. Maarten Altelaar[1], Madalina M. Drugan, Andreas O. Helbig, Shabaz Mohammed and Albert J.R. Heck.

[1] These authors contributed equally to this work and should be considered joint first authors

Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, The Netherlands

Based on Mol Cell Proteomics. 2009, (1), 190-200.

Abstract

In proteomics multi-dimensional fractionation techniques are widely used to reduce the complexity of peptide mixtures subjected to mass spectrometric analysis. Here, we describe the sequential use of strong cation exchange and reversed phase chromatography in the separation of peptides generated by a relatively little explored metalloendopeptidase with Lys-N cleavage specificity. When such proteolytic peptides are subjected to low-pH SCX we obtain fractionation profiles in which, peptides from different functional categories are well separated. The four categories we distinguish and are able to separate to near completion are I) acetylated N-terminal peptides, II) singly phosphorylated peptides containing a single basic (lysine) residue, III) peptides containing a single basic (lysine) residue and IV) peptides containing more than one basic residue.

Analysing these peptides by LC-MS/MS using an ion trap with both collision as well as electron transfer induced dissociation provides unique optimal targeted strategies for proteome analysis. The acetylated peptides in category I can be identified confidently by both CID and ETcaD, where the ETcaD spectra are dominated by sequence informative *z*-ion series. For the phosphorylated peptides in II and the "normal" single lysine containing peptides in III ETcaD provides unique straightforward sequence ladders of *c*'-ions, from which the exact location of possible phosphorylation sites can be easily determined. The later fractions, category IV, require analysis by both ETcaD and CID, where it is shown that ETD performs relatively well for these multiple basic residues containing peptides, as is expected. We argue that the well-resolved separation of functional categories of peptides observed is characteristic for Lys-N generated peptides. Overall, the combination of Lys-N proteolysis, low-pH SCX and RP separation, with CID and ETD induced fragmentation, adds a new very powerful method to the toolbox of proteomic analyses.

Abbreviations used:

CID - collision induced dissociation; ESI – electrospray ionization; ETD - electron transfer dissociation; ETcaD - electron transfer collisional activated dissociation; HEK - human embryonic kidney; HPLC - high pressure liquid chromatography; LC-MS/MS - liquid chromatography tandem mass spectrometry; MS - mass spectrometry; nanoLC - nanoflow liquid chromatography; PTMs – post translational modifications; RP - reversed phase; SCX - strong cation exchange.

Introduction

The enormous complexity of the proteome poses a considerable analytical challenge for global protein identification. An additional order of complexity is caused by protein post-translational modifications (PTMs) of which many different variants are known[1]. Mass spectrometry (MS) is, nowadays, routinely used for such complex protein identification and characterization studies[2, 3]. Common practice, in high throughput global proteome analyses, is to start with proteolytic cleavage of all proteins using, most often, trypsin. The generated peptides are then separated by reversed phase (RP) nanoflow liquid chromatography (nanoLC) and subjected to tandem MS sequencing with CID as the preferred fragmentation method[3]. Protein identification is then accomplished by searching the acquired peptide fragmentation data against large protein sequence databases, which is greatly aided by the availability of sequenced genomes[4]. The above described general procedure for proteomics analysis provides a far from comprehensive view of the proteome, and leaves room for improvement in areas such as proteolytic cleavage[5, 6], peptide enrichment and fractionation[7-18] and peptide activation and fragmentation in the mass spectrometer[19-22]. More specifically, smart combinations of proteases, peptide separations and peptide fragmentation techniques may allow the targeted analysis of specific functional groups of peptides, enabling the analysis of low abundant proteins and peptides.

Recently, electron transfer dissociation (ETD) of peptides was introduced as an alternative peptide fragmentation method. ETD cleaves peptides at the N-C_α bond producing *c'*-and *z*-type ions[23-25]. From the early work it is clear that ETD can be complementary to CID since it prefers larger and more basic peptides, which attain multiple charges during electrospray ionization (ESI)[26, 27]. Reduced fragmentation efficiency of ETD for doubly charged peptides is compensated by the use of supplemental collisional activation (ETcaD)[28]. Interestingly, ETD (and ETcaD) leaves PTMs largely intact on the peptide backbone during fragmentation thus providing, potentially, simpler spectra in which the site of modification can be easily annotated[29]. Recently, we explored and introduced the use of a relatively little known metalloendopeptidase for digestion of proteins in combination with ETD[30]. This metalloendopeptidase, termed Lys-N, has enzymatic cleavage specificity for lysine residues with cleavage occurring at the N-terminal side[31, 32]. Lys-N was shown to be as sensitive and selective as currently used proteases, and can be used for both in-gel and in-solution digestion experiments[30]. Our preliminary data indicated that the resulting proteolytic peptides are favourable for ETD sequencing, with respect to peptide size and the number of charges after

electrospray ionization (ESI), with the added advantage of the lysine residue being situated at the peptide N-terminus. A large proportion of these proteolytic Lys-N peptides do not contain any other basic residue leaving only two basic entities each residing at the N-terminus. We argued and demonstrated that the strong basic nature of the N-terminal side of Lys-N proteolytic peptides attracts the proton providing the final fragment ion charge before/after electron transfer, causing the observed fragments to be almost exclusively *c'*-type ions. Utilizing this unique fragmentation behaviour of Lys-N peptides under ETcaD conditions could markedly reduce the dependence on sequenced genomes and open up a complete new window for *de novo* sequencing and potentially the analysis of post-translational modifications in a facile manner[30]. In the accompanying paper we show that this unique feature is not solely a characteristic of ETD fragmentation, as the same category of peptides provides straightforward peptide sequence ladders in MALDI-CID-MS/MS fragmentation[41].
Here, we further explore the potential of this protease in proteomics, focusing on the analysis of post-translational modifications. We developed a method for global protein analysis of whole cell lysates using a combination of Lys-N proteolytic cleavage followed by low-pH SCX fractionation and RP-nanoLC-ETcaD-MS analysis and RP-nanoLC-CID-MS analysis. Low-pH SCX chromatography has shown to be a valuable tool for phosphopeptide enrichment after tryptic digestion, although they often co-elute with (more abundant) acidic and N-acetylated peptides[8, 34, 35]. We show that the combination of Lys-N and low-pH SCX is an ideal combination for global proteome and phosphoproteome analysis, as well as for the selective enrichment and analysis of protein N-terminal peptides[36, 37]. We argue that the benefits of this approach are multiple since we take full advantage of the proteolytic peptide properties after Lys-N digestion in both SCX enrichment and ETcaD analysis.

Materials and methods

Materials.

Protease inhibitor cocktail was obtained from Roche Diagnostics (Mannheim, Germany). Metalloendopeptidase from Grifola Frondosa (Lys-N) was obtained from Seikagaku Corporation (Tokyo, Japan). Iodoacetamide and sodium orthovanadate were obtained from Sigma-Aldrich (Steinheim, Germany). DL-Dithiothreitol was obtained from Fluka Biochemical (Steinheim, Germany). HEK293 cells were provided by Dr. Pantelis Hatzis and Dr. Tokameh Mahmoudi from the Netherlands Institute for Developmental Biology,

Hubrecht Institute, The Netherlands. HPLC-S gradient grade acetonitrile was purchased from Biosolve (Valkenwaard, The Netherlands). Acetic acid was obtained by MERCK KGaA (Damstadt, Germany) and high purity water obtained from Milli-Q system (Millipore, Bedford, MA).

Human Embryonic Kidney (HEK) 293T cells.

HEK 293 cells were harvested at a density of approx 1.5 x 10^6 cells/mL by centrifugation for 5 min at 1200 rpm and after removal of the medium, cells were resuspended in 50 mM ammonium bicarbonate containing 25 mM sodium phosphate, 1 mM potassium fluoride and 1 mM sodium orthovanadate. After centrifugation at 1200 rpm and removal of the supernatant ice-cold lysisbuffer (50 mM ammonium bicarbonate containing 8 M urea, protease inhibitor, 5 mM sodium phosphate, and 1 M potassium fluoride and 1 M sodium orthovanadate) was added to the cell pellet and the cells were lysed on ice for 30 min. Subsequently, centrifugation at 20000 x g in a tabletop centrifuge (Eppendorf, Hamburg, Germany) at 4 °C separated the soluble and insoluble protein fractions. The soluble fraction was collected and the protein concentration determined by a Bradford assay.

Lysate in-solution digestion.

One mg of lysate was reduced with 45 mM dithiothreitol (50 °C, 15 min) followed by alkylation using 100 mM iodoacetamide (dark, RT, 15 min) and digested with Lys-N. Lys-N was added at a ratio of 1:85 (w/w) and the sample was incubated over night for 37 °C. The digest was dried in a vacuum centrifuge and resuspended in 0.05 % formic acid.

Strong Cation Exchange (SCX) Chromatography.

SCX was performed using an Agilent 1100 HPLC system (Agilent Technologies) with two C18 Opti-Lynx (Optimized Technologies, Oregon OR) guard columns and a Polysulfoethyl A SCX column (PolyLC, Colombia, MD; 200 mm x 2.1 mm i.d., 5 µm, 200-Å). The digested cell lysate was dissolved in 0.05 % formic acid and 750 µg was loaded onto the guard column at 100 µl / min and subsequently eluted onto the SCX column with 80 % acetonitrile and 0.05 % formic acid. SCX buffer A was made of 5 mM KH_2PO_4, 30 % acetonitrile and 0.05% formic acid, pH 2.7; SCX buffer B consist of 350 mM KCL, 5 mM KH_2PO_4, 30 % acetonitrile and 0.05% formic acid, pH 2.7. The gradient was performed in the following way: 0% B for 10 min, 0-85 % B in 35 min, 85-100 % B in 6 min and 100 % B for 4 min. A total of 49 fractions were collected and dried in a vacuum centrifuge.

ETD experiments.

The dried fractions were diluted in 10 % formic acid and aliquots of all the SCX fractions were subjected to nanoscale liquid chromatography tandem mass spectrometry (nanoLC-MS/MS) analysis, performed on an Agilent 1100 HPLC system (Agilent technologies) connected to a LTQ XL Linear Ion Trap Mass Spectrometer with an ETD source at the back from Thermo Fisher Scientific, Inc. (Waltham, MA).

The instrument was equipped with a 20 mm x 100 µm i.d. Aqua C18 trap column (Phenomenex, Torrance, CA) and a 200 mm x 50 µm i.d. Reprosil C18 RP analytical column (Dr Maisch, Ammerbuch-Entringen, Germany). Trapping was performed at a flow of 5 µL/min for 10 min and the fractions were eluted using a 75 min linear gradient from 0 to 40 % solvent B (0.1 M acetic acid in 80 % acetonitrile (v/v), in which solvent A was 0.1 M acetic acid), 40 to 100 % solvent B in 2 min and 100 % B for 2.5 min. The flow rate from the LC was passively split from 0.360 ml/min to 100 nl/min. The column effluent was directly introduced into the ESI source of the MS using a standard coated fused silica emitter (New Objective, Woburn, MA, USA) (o.d. 360 µm, tip i.d. 10 µm) biased to 1.7 kV. The mass spectrometer was operated in positive ion mode, from 350 to 1500 *m/z* in MS mode and with an AGC value of 1.00e+05 and a max injection time of 50 ms. Parent ions were isolated for a more accurate measurement by performing a SIM scan and fragmented by CID and ETD in data dependent mode with an AGC value of 3.00e+04 and a max injection time of 500 ms. Ions were fragmented using CID with normalized collision energy of 35 and 30 ms activation time. ETD fragmentation was performed with supplemental activation, fluoranthene was used as reagent anion and ion/ion reaction in the ion trap was taking place for 100 ms.

Protein identification.

Raw MS data were converted to peak lists using Bioworks Browser software, version 3.1.1. Spectra were searched against the IPI (International Protein Index) Human database version 3.36 (69012 sequences; 29002682 residues) using Mascot software version 2.2.0 (www.matrixscience.com), with Lys-N cleavage specificity. The database search was made with the following parameters set to consider a peptide tolerance of ± 0.5 Da, a fragment tolerance of ± 0.6 Da, allowing 2 missed cleavages, Carbamidomethyl (C) as fixed modification, Oxidation (M), phosphorylation (ST), Phosphorylation (Y) and Acetylation (N-term) as variable modifications. Tandem mass spectra assigned with a Mascot score ≥ 30 (p-value ≤ 0.05) were accepted providing a false discovery rate for the CID data of 4.68 % and 2.53 % for ETD, determined using a decoy database. Mascot interpretation was accepted for

N-terminal acetylation and phosphorylation site assignment. All data are stored in the public depository PRIDE under accession numbers: 3746-3843 and the project description "*A simple strategy for straightforward proteome analysis by a combination of Lys-N, strong cation exchange and electron transfer dissociation*".

Lys-N fragment ion statistics.

For the calculation of the frequency of occurrence of *c*- and *z*-type ions ETD MS/MS fragmentation spectra of unique peptides with a minimum Mascot score of 30 and with a Lysine at the N-terminus (this does not apply for N-acetylated peptides) were considered. The Mascot identification of the spectra was used for automated peak fragment ion counting. For ETD spectra, exceptions occurred when *c*- or *z*-type ions (or related ions with ammonia/water losses) were assigned to the same isotope cluster in which case the most appropriate assignment (e.g. based on mass accuracy trend, mono isotopic peak etc) was chosen in an automated fashion.

Results

We explored a method for global protein analysis of whole cell lysates using a combination of Lys-N proteolytic cleavage followed by low-pH SCX fractionation and RP-nanoLC-MS/MS analysis using both CID and ETcaD. After in-solution digestion of a whole HEK293 cell lysate with Lys-N, the resulting peptides were separated into 49 fractions using a low-pH SCX separation. SCX chromatography separates peptides primarily based on their charge state in-solution, which is governed by the protonation and deprotonation of the basic and acidic residues, and thus the pH of the solution. We hypothesized that low-pH SCX chromatography of Lys-N proteolytic peptides would be an ideal tool for the separation and enrichment of acetylated protein N-termini, phosphorylated and unphosphorylated single lysine containing peptides, as illustrated schematically in Figure 1. Many of the Lys-N generated peptides will contain a single basic lysine residue at the N-terminus of the peptides and therefore carry two positive charges in solution[30]. Similarly, a significant number of the Lys-N generated peptides, with acetylated N-termini (i.e. mostly protein N-termini), will not contain any basic group and therefore be uncharged. Singly phosphorylated Lys-N peptides, with no additional basic residues, will carry a single charge in solution and should potentially be separated from the unphosphorylated single lysine containing peptides and, more

importantly, from the acetylated protein N-terminal peptides described above. Naturally, in the final SCX fractions peptides will be found that contain multiple basic residues, for instance due to the presence of arginine and histidine or more than one lysine residue due to miss-cleavages. This clear separation, as represented in Figure 1, will allow prior knowledge of each fraction composition (i.e. 'peptide must be N-terminally acetylated' or 'peptide must contain a single lysine and one phosphorylated residue') that can be used in the database analysis. This can lead to removal of certain peptides that are false positives leaving behind a dataset with higher overall confidence.

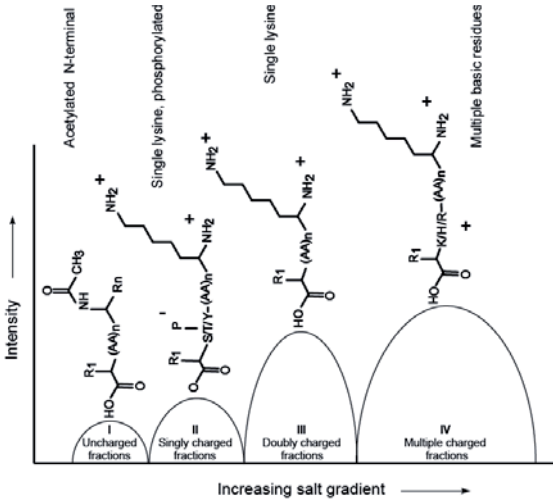

Figure 1: Expected SCX separation scheme for peptides from a Lys-N digest. The horizontal axis shows an increase in salt concentration in the eluted sample and the vertical axis shows an increase in intensity of the eluted peptide fractions. The uncharged Lys-N peptides are by theory expected to elute first from an SCX column and thereafter the singly charged Lys-N peptides and so forth. Using Lys-N, acetylated and phosphorylated peptides should be separated on the SCX column as they possess different charge states in-solution.

We hypothesized that ionization of peptides from the early eluting SCX fractions by electrospray ionization (ESI) could be less efficient due to the lack of potential basic protonation sites. If these peptides would attain just a single charge in the ESI process, analysis with ETD would not be possible since ion-neutralization will occur by the invoked charge reduction. In order to probe this effect we decided to conduct alternating ETcaD and CID fragmentation experiments on the same precursors. Each of the 49 SCX fractions were analysed by RP-nanoLC-MS/MS using, in each single run, alternating ETcaD and CID activation and fragmentation. The resulting spectra were searched against the IPI human database using the Mascot search engine, taking a Mascot score ≥ 30 as threshold. We classified, for each of the SCX fractions, all identified peptides into 8 different categories, based on the number of basic residues (from 1 to 6) present in the peptide. Additionally,

categories were defined for acetylated N-terminal peptides and phosphorylated peptides containing only a single lysine residue.

As can be seen in Figure 2, low-pH SCX separation of the Lys-N digested peptides results in a clear fractionation of all these above mentioned categories of peptides. Specifically, the enrichment of N-acetylated and singly phosphorylated peptides is outstanding. Nearly all of the acetylated N-terminal peptides are distinctly separated from the singly phosphorylated peptides, both for the peptides identified by ETD (Figure 2A) and CID (Figure 2B). Statistical analysis of the data (Figure 2) show that the fractions 7 to 18 contain predominantly acetylated protein N-terminal peptides (92% in ETD and 80% in CID) and only a very small number of phosphopeptides (4% in both ETD and CID). The following 8 SCX fractions (19 to 27) contain only ~3% acetylated protein N-terminal peptides compared to more than 75% singly phosphorylated peptides, illustrating nicely the separation power of the combination of Lys-N digestion and low-pH SCX separation. However, in the later SCX fractions (24 to 27) also several non-phosphorylated peptides were identified, which predominantly contained just a single lysine basic residue. Closer inspection of the fractions containing the N-acetylated protein termini resulted in the observation, that most of these peptides contain a single basic residue. The acetylated N-terminal peptides without basic residues are present in the very early fractions (i.e. 4 to 6) while the following fractions contain almost exclusively peptides with one basic residue. This observation explains why we identify so many of the N-terminal peptides in the ETD experiment, as without this basic residue, ionization and ETD fragmentation would be hampered. It does however raise the question why these peptides are separated from the single lysine containing phosphorylated peptides, as both these groups of peptides should have a single charge in solution. We believe that the unique combination of N-terminal lysine peptides and low-pH SCX conditions might play an important role. Although, phosporylated peptides with a single lysine have an overall charge of 1+ in solution, the N-terminal end is still doubly charged and thus retention in the SCX column might be stronger than the retention of single charged N-terminal acetylated peptides. The number of N-acetylated and phosphorylated peptides in the remaining later SCX fractions were also calculated, but were of low frequency (see Figure 2).

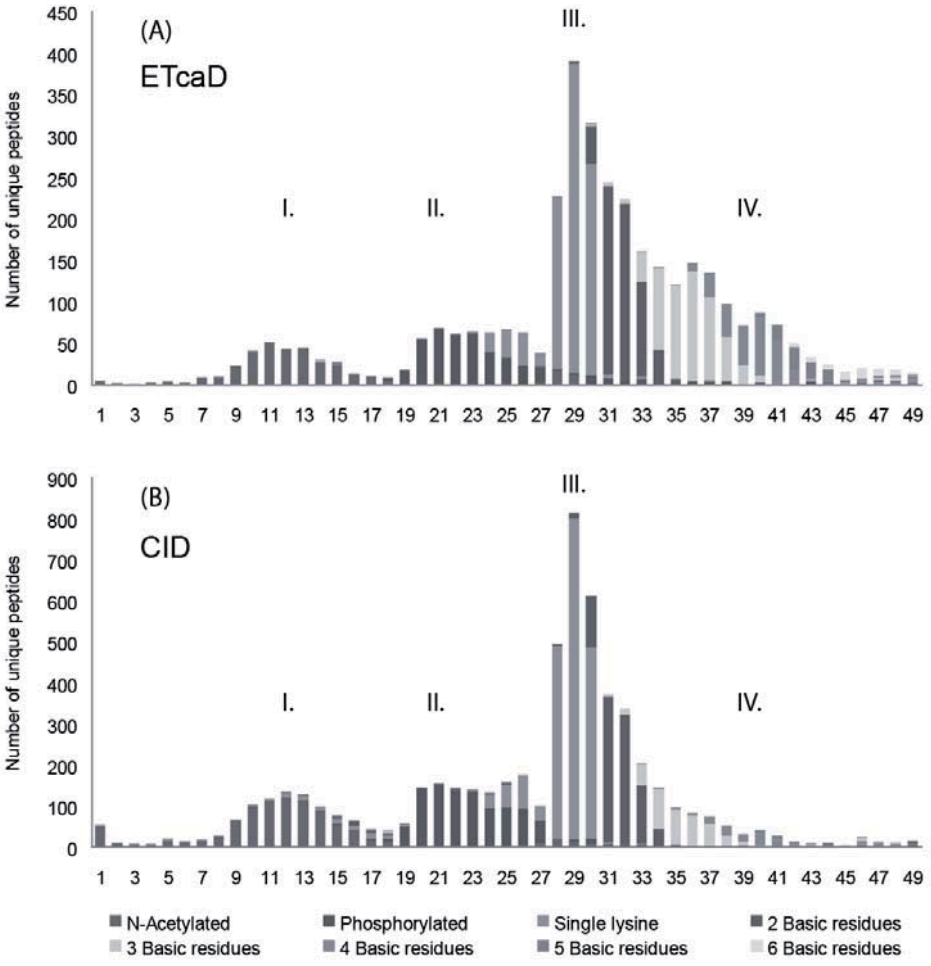

Figure 2: Total number of unique peptides in all the SCX fractions analyzed by ETcaD (A) and CID (B). A clear separation of the different peptide subgroups is observed for both ETD and CID (group 1 to 4). The different peptide subgroups are described at the bottom and illustrated in different shades of grey.

The general trends observed for ETD and CID were similar but there were also some significant differences observed, as seen in Figure 2A and 2B. Significantly more acetylated peptides are identified by CID, when comparing the same SCX fractions to ETD identified peptides (Figure 2B); over 2 fold more for the fractions 7 to 18. However, the vast majority of the identified N-acetylated peptides, in the CID experiments, had precursor peptide ions that were doubly charged making them, potentially, amenable to ETD analysis. Therefore, the

relative increase in identification efficiency, observed in the CID experiments, compared to ETD, likely points towards the increased efficiency and more confident identification of the data by Mascot for doubly charged peptides, in agreement with published data[19, 26, 27].

When inspecting the ETcaD tandem mass spectra, we observed, that the majority of the doubly charged acetylated protein N-terminal precursor ions lead to clear MS/MS spectra mainly consisting of $z^{\cdot}$- fragment ions (as illustrated in Figure 3A and C).

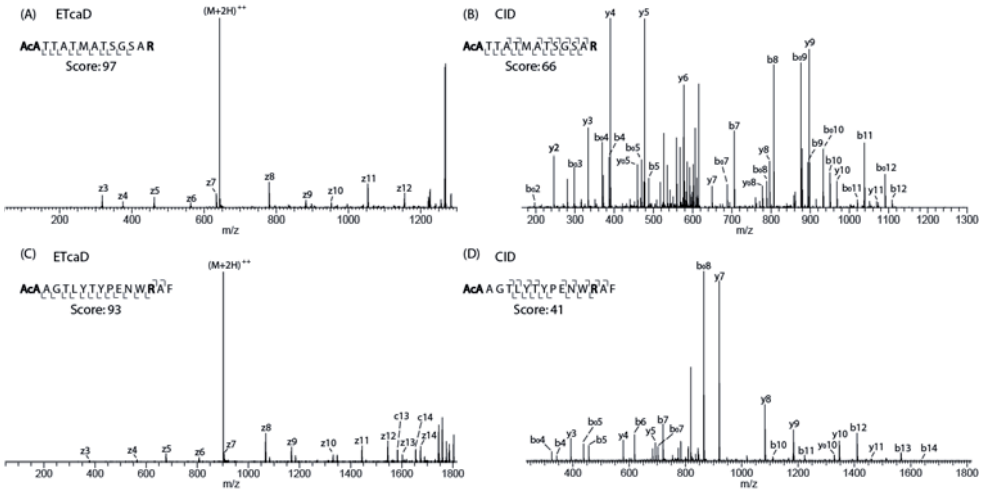

Figure 3: ETcaD and CID MS/MS spectra of N-terminally acetylated doubly charged ions originated from a Lys-N digest of a HEK293 cell lysate. (A): ETcaD spectrum of the N-terminal acetylated peptide AcATTATMATSGSAR (*m*/*z*: 642.23, 2+) from EIF4A3 Eukaryotic initiation factor 4A-III. A clear sequence of almost all the *z*-ions are being generated due to the basic residue (R) at the C-terminal. (B): CID spectrum of the N-terminal acetylated peptide AcATTATMATSGSAR (*m*/*z*: 642.16, 2+). Both *b*- and *y*-type ions are generated in the CID fragmentation process. (C): ETcaD spectrum of the N-terminal acetylated peptide AcAAGTLYTYPENWRAF (*m*/*z*: 901.26, 2+) from EEF1G Elongation factor 1-gamma. Almost a whole series of *z*-ions is being generated due to the basic residue (R) close positioned to the C-terminal. Two N-terminally *c'*-ions (*c*13 and *c*14) are generated from the position of the arginine residue. (D): CID spectrum of the N-terminal acetylated peptide AcAAGTLYTYPENWRAF (*m*/*z*: 901.26, 2+). An almost equal amount of *b*- and *y*-ions is formed.

We believe that this preference for the formation of $z^{\cdot}$-ions is caused by the acetylation of the N-terminus and the presence of a basic residue close to the C-terminus, which directs the remaining proton to the C-terminus. In agreement with this hypothesis, N-acetylated peptides

with no basic residues lead to the abundant formation of both *c'*- and *z*·- fragment ions (see supplemental dataset present in PRIDE – accession numbers: 3746-3843). As demonstrated in Figure 4, the dominant presence of *z*·-ions is consistent over the whole range of SCX fractions from 9-17, containing primarily the N-acetylated protein N-termini with a single basic residue. In Figure 3B and D the associated CID spectra of the same peptides are shown. Although the CID spectra are more fragment ion rich, they do provide a less clear picture of the peptide sequence, when compared to the ETD spectra given in Figure 3A and C.

Elution of phosphorylated peptides in SCX overlaps marginally with peptides containing a single lysine residue and no PTMs, as can be seen in figure 2. The phosphorylated peptides were mainly observed as doubly charged ions, after ESI, and contained exclusively a single N-terminal lysine, as the only basic residue. ETcaD fragmentation of these doubly charged phosphorylated peptides results in 'clean' straightforward interpretable spectra consisting predominantly of *c'*-type ions, as observed and reported previously for their non-phosphorylated counterparts[30]. Utilizing Mascot annotation, we calculated the frequency of the various fragment ions observed and found that more than 90% of the ions were N-terminal in nature (i.e. *c*-type ions), as illustrated in Figure 4.

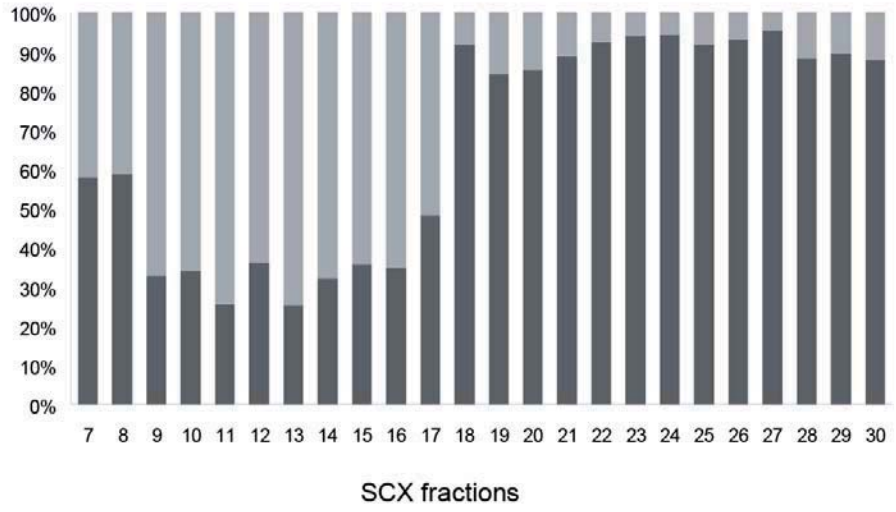

Figure 4. Occurrence (%) of *c'*- and *z*- fragment ions in the ETcaD spectra of the doubly charged peptides per SCX fraction. The dark grey color represent the *c'*-ions and the light grey the *z*-ions. Fraction 7 to 17 contain a lot of N-terminal acetylated peptides which results in mainly the generation of *z*-ions (light grey) in ETD, whereas phosphorylated peptides and non-phosphorylated peptides with a single lysine at the N-terminus (fraction 18-30) results in the generation of almost only *c'*-ions (dark grey).

Interestingly, these ETcaD spectra require very little effort to interpret since clear sequence ladders are formed with the phosphor group left intact on the peptide backbone (Figure 5), facilitating site specific phosphopeptide identification. As can be seen in Figure 5D, also tyrosine phosphorylation could be easily determined from the sequence ladders.

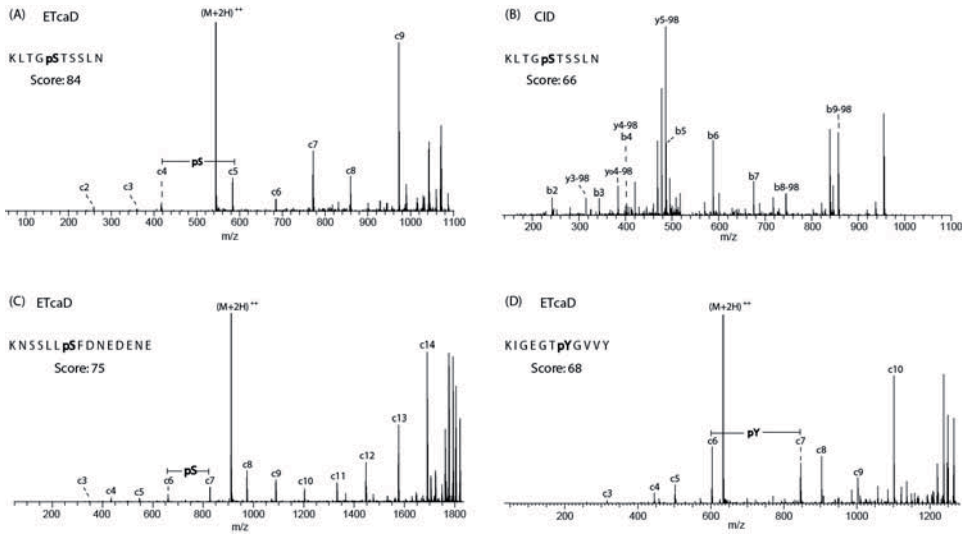

Figure 5: ETcaD and CID MS/MS spectra of phosphorylated doubly charged ions originating from Lys-N generated peptides from HEK293 cells. (A): ETcaD MS/MS spectrum of serine phosphorylated KLTGpSTSSLN (*m/z*: 544.11, 2+) from EXOC1 Isoform 1 of Exocyst complex component 1. From the sequence of *c'*-ions it is easy to determine the phosphorylation site. (B): CID MS/MS spectrum of serine phosphorylated KLTGpSTSSLN (*m/z*: 544.08, 2+). Compared to the ETcaD spectrum a clear sequence for interpretation of the phosphorylation site is not observed. (C): ETcaD MS/MS spectrum of serine phosphorylated KNSSLLpSFDNEDENE (*m/z*: 910.68, 2+) from the Uncharacterized protein ENSP00000307425. Nearly the entire sequence is fragmented into *c'*-ions by the ETcaD process. The phosphate group is not lost in the fragmentation process. (D): ETcaD MS/MS spectrum of tyrosine phosphorylated KIGEGTpYGVVY (*m/z*: 633.13, 2+) CDK3 Cell division protein kinase 3. By interpretation of the peptide fragment ions in the spectrum the sequence and site of phosphorylation can be determined, as only *c'*-ions are generated and the phosphate group is not lost in the fragmentation process.

The single lysine containing peptides are strongly enriched in mainly three SCX fractions (28 till 30), which contain almost 85% of the single lysine peptide population (Figure 2), calculated from both the ETcaD and CID analyses. In agreement with our earlier reported data[30], ETD analysis of these fractions resulted in MS/MS spectra almost exclusively

consisting of *c*-type ions in which *c'*-ions dominate, providing very simple to interpret sequence ladders (Figure 6A). For comparison, in Figure 6B, is shown the complementary CID spectrum of the same precursor peptide, which reveals again a rich, but less clear fragmentation spectrum.

In the ETcaD experiments the number of generated *c*-type ions decreases as the charge state increases, however, still more than 65% of the ions observed in the fractions containing peptides with three or more basic residues are *c*-type ions (see supplemental dataset present in PRIDE accession numbers: 3746-3843). Here, the relative contribution of *c*-type ions is determined by the position of the additional basic residue, i.e. an additional lysine, arginine or histidine. If the extra basic residue is positioned close to the C-terminus both *c'*- and $z^{\cdot}$-ions will be formed. However, if the basic residue is closer to the N-terminal site *c'*-ions are dominant, as illustrated by some examples in Figure 6. Obviously, these higher charged peptides are still very useful in the analysis of the whole cell lysate since ETD efficiency increases with increasing charge state. When comparing the global SCX profiles of the ETcaD and the CID experiments, CID performs relatively well in the SCX fractions that contain peptides with up to maximally 2 basic residues. For peptides containing more than 2 basic residues, which attain more charges in the ESI process, a rapid decrease in the absolute and relative number of peptides identified by CID is observed. In general, ETcaD performs better for these peptides containing multiple basic residues, largely due to the increased efficiency of the electron transfer process. We believe that in the current experimental set-up the number of peptides identified by ETcaD, in the lower charged fractions, are underrepresented due to the suboptimal weighting in the search algorithm for the "unique" *c'*-fragment ion ladder sequences observed for the single lysine peptides. Improving the ion scoring algorithms for these peptides might improve the number of identifications.

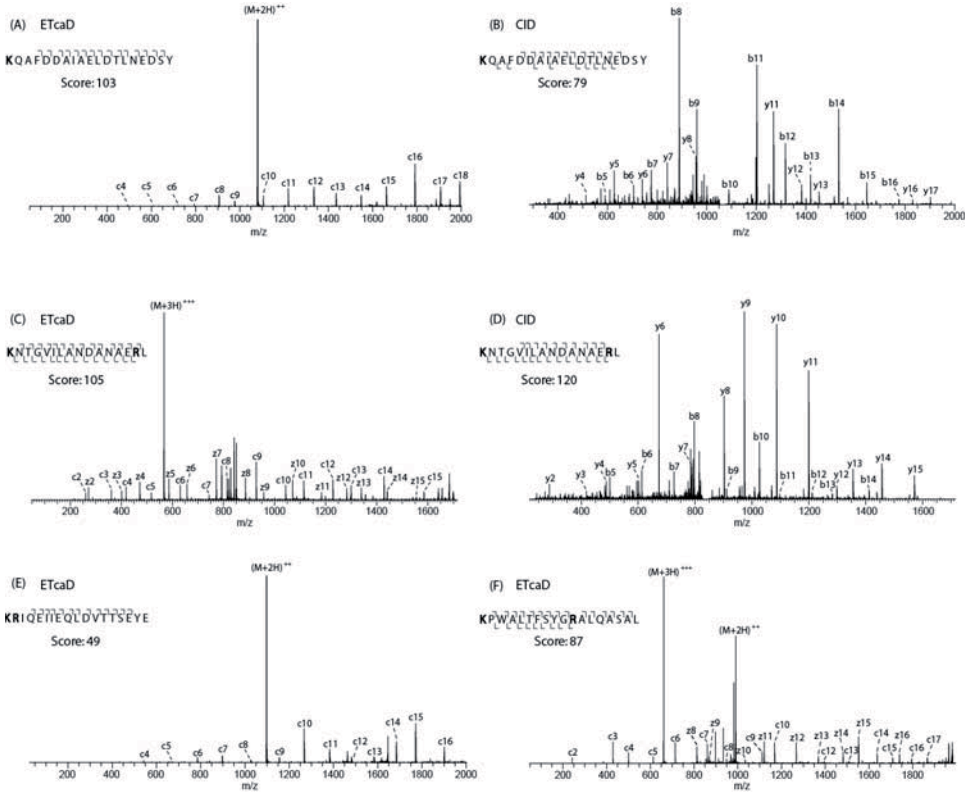

Figure 6: ETcaD and CID MS/MS spectra of doubly and triple charged ions originating from Lys-N generated peptides from HEK293 cells. (A): ETcaD spectrum of doubly charged peptide KQAFDDAIAELDTLNEDSY (*m*/*z*: 1079.34, 2+) from YWHAH 14-3-3 protein eta. Protons/charges in Lys-N generated peptides, with a single basic residue, will be preferentially located at the N-terminus due to the presence of two free amine groups, which results in the exclusive generation of *c'*-type fragment ions. (B): CID spectrum of doubly charged peptide KQAFDDAIAELDTLNEDSY (*m*/*z*: 1079.36, 2+) from YWHAH 14-3-3 protein eta. CID fragmentation does not result in the exclusive generation of N-terminal fragment ions. (C): ETcaD spectrum of triple charged peptide KNTGVILANDANAERL (*m*/*z*: 566.82, 3+) from NOL1 94 kDa protein. Lys-N generated peptides, with a basic residue contiguous to the *c*-terminal, result in the formation of almost equally numbers of *c'*- and *z*-ions. (D): CID spectrum of doubly charged KNTGVILANDANAERL (*m*/*z*: 849.86, 2+) from NOL1 94 kDa protein. In CID the same trend is not observed as mainly *y*-ions are generated. (E): ETcaD spectrum of doubly charged peptide KRIQEIIEQLDVTTSEYE (*m*/*z*: 1097.45, 2+) from HSPD1 60 kDa heat shock protein, mitochondrial precursor. Lys-N generated peptides, with a second basic residue adjacent to the N-teminal lysine, will result in the exclusive generation of *c'*-type fragment ions (F): ETcaD spectrum of triple charged peptide KPWALTFSYGRALQASAL (*m*/*z*: 660.53, 3+) from ALDOA Fructose-bisphosphate aldolase A. Lys-N generated peptides, with a second basic residue in the centre of the peptide sequence, will result in the generation of *c'*- and *z*-type fragment ions.

Discussion

Generally, the first step in a typical global proteomics analysis workflow is fractionation of a tryptic digest in order to reduce the complexity of the peptide mixtures subjected to RP-nanoLC-MS. The reduction of complexity is vital to the outcome of the experiment, since existing mass spectrometers lack the speed to be able to sequence every peptide of a complex sample eluting from the LC column. Fractionation is often accomplished by SCX chromatography of peptides generated by in-solution digestion. An additional benefit of SCX separation is the enrichment into distinct fractions of similar peptide properties before MS analysis. This advantage can be further exploited by performing the SCX separation at low pH, where, after tryptic digestion, most of the phosphopeptides are found in the early fractions of the SCX separation. This is due to their reduced charge states in solution since glutamates and aspartates will be neutralised and the phosphate group will possess a negative charge[8-34]. Here, we show the added value of using the metalloendopeptidase Lys-N to digest the proteins into peptides in combination with low-pH SCX fractionation and ETcaD-MS and CID-MS analysis. Digestion with Lys-N yields proteolytic peptides with the lysine at their N-terminus, resulting in an increased basic entity caused by the lysine and N-terminal amines. In the subsequent low-pH SCX separation distinct fractionation profiles could be observed, in which, peptides from different functional categories were extremely well separated. The four categories we were able to separate well are I) acetylated N-terminal peptides, II) singly phosphorylated peptides containing a single basic (lysine) residue, III) peptides containing a single basic (lysine) residue and IV) peptides containing more than one basic residue. Since the SCX gradient was optimized for the enrichment of acetylated protein N-termini and phosphorylated peptides, the separation of the later eluting multiple charged peptides (i.e. containing more than one basic residue) was somewhat compromised. Still, with the current SCX set-up the single lysine containing peptides elute in a very clean pool, spread over mainly three fractions. The enrichment of post-translational modified peptides is relatively high, in our current experiments, when taking into account that no additional enrichment method has been applied, and that solely in-solution charge state separation has been used.

Most interestingly, the proposed set-up results in a fractionation of peptides where there is almost no overlap between the N-acetylated and phosphorylated peptides. N-terminal acetylation, and less so propionylation, of proteins is a common and important process in cellular biology, which is for instance linked to protein stability, protecting the proteins from attack by aminopeptidases[35, 38-40]. Since approximately 90% of cellular proteins in eukaryotic

cells contain blocked N-termini, they make up a significant portion of the peptide pool and are of great interest with respect to the examination of N-terminal protein processing or alternatively terminated protein isoforms[33, 35]. Our data show, that Lys-N generated peptides with an acetylated protein N-termini, in category I, form primarily $z^{\cdot}$-ions in ETcaD fragmentation (~70%, Figure 3 and Figure 4), caused by the blockage of the N-terminus and the presence of a basic residue.

An additional advantage of the almost complete separation is that the fractions containing acetylated protein N-termini are removed as 'contamination' from the phosphorylated fractions, allowing enrichment and targeted analysis of the phosphoproteome. ETcaD analysis of the purified phosphopeptides (category II) results in clear *c'*-ion ladder sequences of the peptides, and allows a simple read out of the location of the phosphorylation. Peptides containing a single lysine and no other basic residues (category III) can also be purified via low pH SCX. These peptide ions can be preferentially analyzed by ETcaD, as with this activation method simple *c'*-ion sequence ladders are generated (as illustrated in Figure 6A). These sequence ladders can potentially facilitate *de novo* sequencing of peptides from species with unsequenced genomes or unknown isoforms, circumventing the need for the availability of databases containing known protein molecular sequences.

A general overview of the fragmentation patterns observed in ETcaD analysis of the peptides observed in the four different categories is shown in Figure 7. Figure 7A shows the preferred ETcaD fragmentation of the acetylated protein N-terminus peptides (category I). Since the N-terminus is blocked the remaining proton after the ETcaD analysis will preferentially go to the C-terminus and in the resulting spectra $z^{\cdot}$-ions will dominate. Both the phosphorylated and unphosphorylated single lysine peptides (Figure 7 B and C, category II and III, respectively) will produce spectra dominated by *c'*-ions, after ETcaD analysis, due to the favoured protonation of the N-terminus. Finally, peptides containing multiple basic residues, category IV, form both *c'*- and $z^{\cdot}$-ions (Figure 7D). Here, a preference for *c*-type ions is only observed when the basic residues are close to the N-terminus (see also Figure 6E). An equal occurrence of *c*- and *z*-type ions are observed when the basic residue is close to the C-terminus (see also figure 6C) as has also been shown for tryptic peptides using electron capture dissociation.

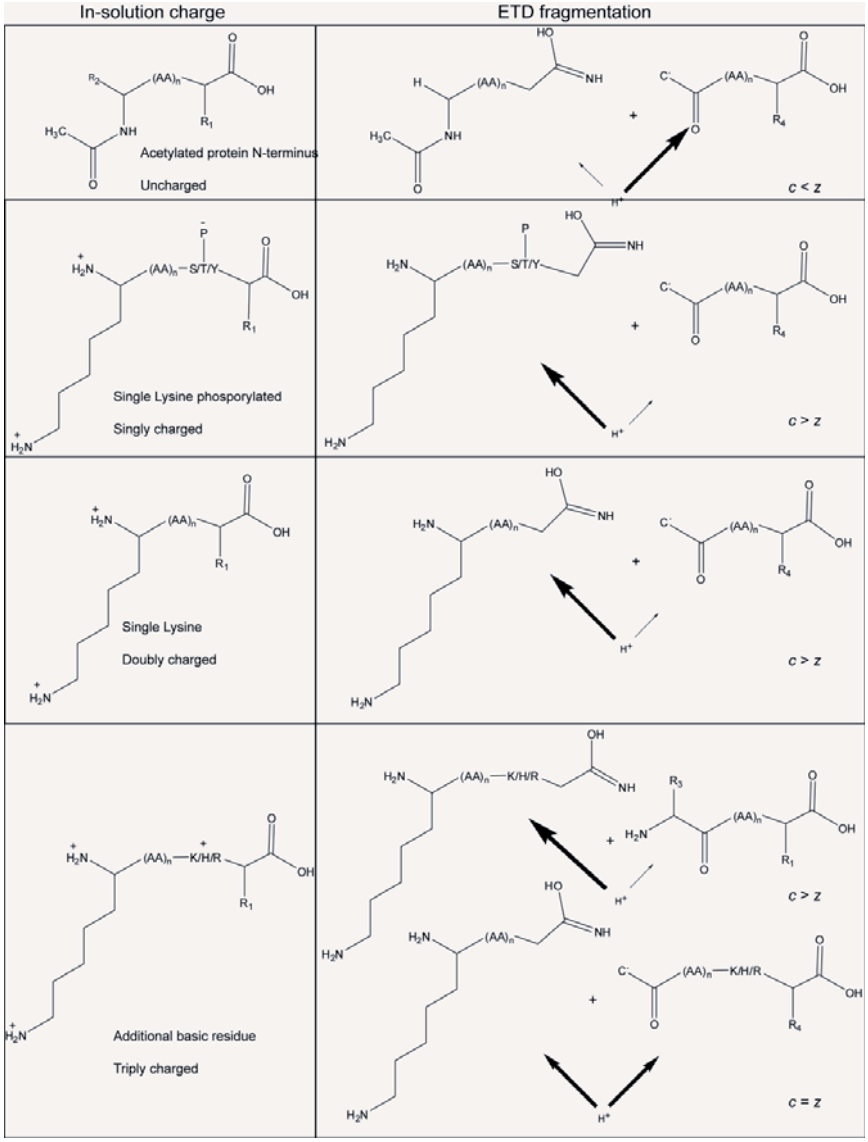

Figure 7: General overview describing the dominance of particular fragmentation patterns observed in ETcaD analysis of (A) peptides with an acetylated N-terminus containing a single basic residue (group I); (B) phosphorylated peptides containing a single N-terminal lysine (group II); (C) peptides containing a single N-terminal lysine (group III); (D) peptides with at least two basic residues in their sequence (group IV).

Conclusion

In summary, the use of the metalloendopeptidase Lys-N in combination with SCX, described here in detail, is a powerful proteomics method. It allows the separation to near completion of I) acetylated N-terminal peptides (with and without a single basic residue), II) singly phosphorylated peptides containing a single basic (lysine) residue, III) peptides containing a single basic (lysine) residue and IV) peptides containing more than one basic residue. Analysing these peptides by LC-MS/MS using both CID and ETcaD provides unique optimal targeted strategies for proteome analysis of these classes of peptides. Strikingly, ETcaD provides a facile method for site localisation of phosphorylated peptides in category II and facilitates a database independent method for sequencing of "normal" single lysine containing peptides. Overall, the combination of the use of the Lys-N protease, with SCX and RP separation, and CID and ETD induced fragmentation, adds a new very powerful method to the toolbox of proteomic analyses enabling also facile analysis of peptides and their post-translational modifications.

Acknowledgements

This work was supported by the Netherlands Proteomics Centre (http://www.netherlandsproteomicscentre.nl/).

References

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

2 Pandey, A., Andersen, J. S., and Mann, M. (2000) Use of mass spectrometry to study signaling pathways. Sci. STKE, PL1

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

4 Sadygov, R. G., Cociorva, D., and Yates, J. R., 3rd (2004) Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. Nat. Methods 1**,** 195– 202

5 Mohammed, S., Lorenzen, K., Kerkhoven, R., van Breukelen, B., Vannini, A., Cramer, P., and Heck, A. J. (2008) Multiplexed proteomics mapping of yeast RNA polymerase II and III allows near-complete sequence coverage and reveals several novel phosphorylation sites. Anal. Chem. 80**,** 3584– 3592

6 Klammer, A. A., and MacCoss, M. J. (2006) Effects of modified digestion schemes on the identification of proteins from complex mixtures. J. Proteome Res. 5**,** 695– 700

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

8 Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villen, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc. Natl. Acad. Sci. U. S. A. 101**,** 12130– 12135

9 Pinkse, M. W., Uitto, P. M., Hilhorst, M. J., Ooms, B., and Heck, A. J. (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. Anal. Chem. 76**,** 3935– 3943

10 Boersema, P. J., Divecha, N., Heck, A. J., and Mohammed, S. (2007) Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. J. Proteome Res. 6**,** 937– 946

11 Bodenmiller, B., Mueller, L. N., Mueller, M., Domon, B., and Aebersold, R. (2007) Reproducible isolation of distinct, overlapping segments of the phosphoproteome. Nat. Methods 4**,** 231– 237

12 Kweon, H. K., and Hakansson, K. (2006) Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. Anal. Chem. 78**,** 1743– 1749

13 Pinkse, M. W., Mohammed, S., Gouw, J. W., van Breukelen, B., Vos, H. R., and Heck, A. J. (2008) Highly robust, automated, and sensitive online TiO_2-based phosphoproteomics applied to study endogenous phosphorylation in Drosophila melanogaster. J. Proteome Res. 7**,** 687– 697

14 Ficarro, S. B., Parikh, J. R., Blank, N. C., and Marto, J. A. (2008) Niobium(V) Oxide (Nb2O5): application to phosphoproteomics. Anal. Chem. 80**,** 4606– 4613

15 Ruse, C. I., McClatchy, D. B., Lu, B., Cociorva, D., Motoyama, A., Park, S. K., and Yates, J. R. (2008) Motif-specific sampling of phosphoproteomes. J. Proteome Res. 7**,** 2140– 2150

16 Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., Faergeman, N. J., Mann, M., and Jensen, O. N. (2005) Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. Mol. Cell. Proteomics 4**,** 310– 327

17 Dephoure, N., Zhou, C., Villen, J., Beausoleil, S. A., Bakalarski, C. E., Elledge, S. J., and Gygi, S. P. (2008) A quantitative atlas of mitotic phosphorylation. Proc. Natl. Acad. Sci. U. S. A. 105**,** 10762– 10767
18 Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127**,** 635– 648
19 Good, D. M., Wirtala, M., McAlister, G. C., and Coon, J. J. (2007) Performance characteristics of electron transfer dissociation mass spectrometry. Mol. Cell. Proteomics 6**,** 1942– 1951
20 Molina, H., Horn, D. M., Tang, N., Mathivanan, S., and Pandey, A. (2007) Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. Proc. Natl. Acad. Sci. U. S. A. 104**,** 2199– 2204
21 Zubarev, R. A., Zubarev, A. R., and Savitski, M. M. (2008) Electron capture/transfer versus collisionally activated/induced dissociations: solo or duet? J. Am. Soc. Mass Spectrom. 19**,** 753– 761
22 Liu, H., Hakansson, K., Lee, J. Y., and Sherman, D. H. (2007) Collision-activated dissociation, infrared multiphoton dissociation, and electron capture dissociation of the Bacillus anthracis siderophore petrobactin and its metal ion complexes. J. Am. Soc. Mass Spectrom. 18**,** 842– 849
23 Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc. Natl. Acad. Sci. U. S. A. 101**,** 9528– 9533
24 Coon, J. J., Ueberheide, B., Syka, J. E., Dryhurst, D. D., Ausio, J., Shabanowitz, J., and Hunt, D. F. (2005) Protein identification using sequential ion/ion reactions and tandem mass spectrometry. Proc. Natl. Acad. Sci. U. S. A. 102**,** 9463– 9468
25 Pitteri, S. J., Chrisman, P. A., Hogan, J. M., and McLuckey, S. A. (2005) Electron transfer ion/ion reactions in a three-dimensional quadrupole ion trap: reactions of doubly and triply protonated peptides with $SO_2^{\cdot-}$. Anal. Chem. 77**,** 1831– 1839
26 Toorn van de, H. W. P., Mohammed, S., Gouw, J. W., Breukelen van, B., and Heck, A. J. R. (2008) Targeted SCX-based peptide fractionation for optimal sequencing by collision- and electron transfer-induced dissociation. J. Proteomics Bioinform. 1**,** 374– 386
27 Molina, H., Matthiesen, R., Kandasamy, K., and Pandey, A. (2008) Comprehensive comparison of collision-induced dissociation and electron transfer dissociation. Anal. Chem. 80**,** 4825– 4835
28 Swaney, D. L., McAlister, G. C., Wirtala, M., Schwartz, J. C., Syka, J. E., and Coon, J. J. (2007) Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. Anal. Chem. 79**,** 477– 485
29 Mikesh, L. M., Ueberheide, B., Chi, A., Coon, J. J., Syka, J. E., Shabanowitz, J., and Hunt, D. F. (2006) The utility of ETD mass spectrometry in proteomic analysis. Biochim. Biophys. Acta 1764**,** 1811– 1822
30 Taouatas, N., Drugan, M. M., Heck, A. J., and Mohammed, S. (2008) Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase. Nat. Methods 5**,** 405– 407
31 Nonaka, T., Hashimoto, Y., and Takio, K. (1998) Kinetic characterization of lysine-specific metalloendopeptidases from Grifola frondosa and Pleurotus ostreatus fruiting bodies. J. Biochem. 124**,** 157– 162
32 Rao, K. C., Palamalai, V., Dunlevy, J. R., and Miyagi, M. (2005) Peptidyl-Lys metalloendopeptidase-catalyzed 18O labeling for comparative proteomics: application to cytokine/lipolysaccharide-treated human retinal pigment epithelium cell line. Mol. Cell. Proteomics 4**,** 1550– 1557

33 Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G. R., and Vandekerckhove, J. (2003) Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. Nat. Biotechnol. 21**,** 566–569

34 Ballif, B. A., Villen, J., Beausoleil, S. A., Schwartz, D., and Gygi, S. P. (2004) Phosphoproteomic analysis of the developing mouse brain. Mol. Cell. Proteomics 3**,** 1093–1101

35 Dormeyer, W., Mohammed, S., Breukelen, B., Krijgsveld, J., and Heck, A. J. (2007) Targeted analysis of protein termini. J. Proteome Res. 6**,** 4634– 4645

36 Coussot, G., Hawke, D. H., Mularz, A., Koomen, J. M., and Kobayashi, R. (2007) A method for the isolation of blocked N-terminal peptides. Anal. Biochem. 361**,** 302–304

37 Staes, A., Van Damme, P., Helsens, K., Demol, H., Vandekerckhove, J., and Gevaert, K. (2008) Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). Proteomics 8**,** 1362– 1370

38 Brown, J. L., and Roberts, W. K. (1976) Evidence that approximately eighty per cent of the soluble proteins from Ehrlich ascites cells are Nα-acetylated. J. Biol. Chem. 251**,** 1009– 1014

39 Polevoda, B., and Sherman, F. (2003) N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. J. Mol. Biol. 325**,** 595– 622

40 Chen, Y., Sprung, R., Tang, Y., Ball, H., Sangras, B., Kim, S. C., Falck, J. R., Peng, J., Gu, W., and Zhao, Y. (2007) Lysine propionylation and butyrylation are novel post-translational modifications in histones. Mol. Cell. Proteomics 6**,** 812– 819

41 Boersema PJ, Taouatas N, Altelaar AF, Gouw JW, Ross PL, Pappin DJ, Heck AJ, Mohammed S. (2009) Straightforward and de novo peptide sequencing by MALDI-MS/MS using a Lys-N metalloendopeptidase. Mol. Cell. Proteomics, 4, 650-660.

Chapter 4

Straightforward and *de novo* peptide sequencing by MALDI-MS/MS using a Lys-N metalloendopeptidase

Paul J. Boersema[1,†], Nadia Taouatas[1,†], A. F. Maarten Altelaar[1], Joost W. Gouw[1], Philip L. Ross[2], Darryl J. Pappin[3], Albert J.R. Heck[1], Shabaz Mohammed[1]

[†] These authors contributed equally to this work and should be considered joint first authors

[1] Biomolecular Mass Spectrometry and Proteomics Group, Utrecht Institute for Pharmaceutical Sciences and Bijvoet Center for Biomolecular Research, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, The Netherlands. [2] Applied Biosystems, 500 Old Connecticut Path, Framingham MA 01701, USA. [3] Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA

Based on Mol Cell Proteomics. 2009, (4), 650-660.

Abstract

In this work, we explore the potential of the metalloendopeptidase Lys-N for MALDI-MS/MS proteomics applications. Initially, we digested a HEK293 cellular lysate with Lys-N and, for comparison, in parallel with the protease Lys-C. The resulting peptides were separated by SCX to enrich and isolate peptides containing a single N-terminal lysine. MALDI-MS/MS analysis of these peptides yielded CID spectra with clear and often complete sequence ladders of *b*-ions. To test the applicability for *de novo* sequencing we next separated an ostrich muscle tissue protein lysate by 1D SDS-PAGE. A protein band at 42 kDa was in-gel digested with Lys-N. Relatively straightforward sequencing resulted in the *de novo* identification of the two ostrich proteins creatine kinase and actin. We therefore conclude that this method that combines Lys-N, SCX enrichment and MALDI-MS/MS analysis provides a valuable alternative proteomics strategy.

Abbreviations used:

CAF - chemically assisted fragmentation; ETD - electron transfer dissociation; ETcaD - ETD with supplemental collisional activation; FDR - false discovery rate; MS/MS - tandem mass spectrometry; RP - reversed phase; SCX - strong cation exchange; ZIC-HILIC - zwitterionic hydrophilic interaction liquid chromatography.

Introduction

In proteomics, peptide sequencing is mainly performed by collision induced dissociation (CID) based tandem mass spectrometry[1, 2]. Generated peptide fragmentation spectra are matched against in silico derived spectra from amino acid sequences in proteomic and genomic databases. Trypsin is the most frequently utilized protease as it generates peptides in the preferred mass range for effective fragmentation by CID[3]. Trypsin has high cleavage specificity and is stable under a wide variety of conditions generating peptides with a C-terminal arginine or lysine. This C-terminal positioning of the basic residue has consequences for fragment ion formation in CID. According to the 'mobile proton' model, dissociation upon excitation is initiated by a proton that weakens an amide bond in the peptide backbone[4-7]. The proton affinity/gas phase basicity of the two conjugate fragments will then dictate which fragment will inherit the amide breaking proton, leading to the formation of respectively *b*- or *y*-ions[8]. In MALDI-MS/MS of singly charged tryptic peptides, fragmentation results in complex spectra containing not only *b*- and *y*-ions, but also some *a*- and immonium ions, internal fragments and ions resulting from neutral loss of ammonia or water[9, 10]. Although all these fragment ions are used in typical database search strategies they often complicate and hamper *de novo* sequencing, *e.g.* sequencing of peptides from species of which no genome sequence is available[11, 12]. Therefore, several attempts have been made to simplify MALDI-CID spectra[11, 13-19]. For example, the peptide N-terminus can be derivatized with sulfonic acid as in chemically assisted fragmentation (CAF)[16, 17] or with 4-sulfophenyl isothiocyanate (SPITC)[20, 21] to establish a fixed negative charge. After this reaction, primarily protonated C-terminal fragment ions are detected, giving rise to *y*-ion ladder series. The loss of sensitivity caused by adding a negative charge while performing positive ion mode analysis can, though partly, be compensated for by increasing the basicity of lysine residues [22]. Another approach to simplify MALDI spectra is to add a fixed positively charged tag to the peptide N-terminus[13] while modifying internal arginine residues[14] or removing the C-terminal lysine or arginine[11]. These modified peptides fragment to generate spectra with mainly *a*- and *b*-ions. One can also modify the basicity of the peptide to promote formation of a single series of ions. Lysine can be made more basic by guanidation[18] or treatment with 2-methoxy-4,5-dihydro-1H-imidazole[19]. In this way, the C-terminal fragment of a tryptic peptide is more likely to be protonated after fragmentation yielding spectra with more intense *y*-ions. Such techniques, however, require additional sample handling. Moreover, chemical

derivatization of minute amounts of sample is more difficult and is often hampered by the formation of unwanted side products[23].

Recently, we explored a new method for mass spectrometry based sequencing of peptides using a little explored metalloendopeptidase with Lys-N cleavage specificity[24]. We showed that the combination of this protease with ESI-MS using electron transfer induced dissociation (ETD) for peptide fragmentation produced spectra that were completely dominated by *c*-type fragment ions, providing simple sequence ladders of the peptides of interest[24]. In ETD with supplemental collisional activation (commonly referred to as ETcaD), doubly charged peptide ions generated by ESI are charge reduced during the electron transfer process, resulting in the remainder of a single free proton[25]. As the N-terminus of Lys-N peptides accommodates two basic entities, primarily N-terminal fragments are protonated, which therefore leads to the detection of mainly *c*-ions. However, ETD fragmentation requires multiply charged ions (i.e. ESI) and specific instrumentation that is not readily available. Therefore, in this paper, we comprehensively explored the use of the Lys-N metalloendopeptidase and CID fragmentation using MALDI-MS/MS. A HEK293 cellular lysate was digested by Lys-N and for a direct comparison also with Lys-C, which produces tryptic-like peptides with the basic lysine at the C-terminus. The resulting peptides were separated and enriched for peptides containing a single lysine residue by SCX and analyzed by MALDI-TOF/TOF. The combination of Lys-N and MALDI-MS/MS resulted in spectra with clear and straightforward sequence ladders, consisting of almost exclusively *b*-ions. We also performed a direct comparison of the following four combinations I) Lys-C, SCX enrichment and MALDI-MS/MS analysis, II) Lys-N, SCX enrichment and MALDI-MS/MS analysis, III) Lys-N, SCX enrichment and ESI-CID MS/MS analysis and IV) Lys-N, SCX enrichment and ESI-ETD MS/MS analysis, which clearly demonstrated that only the combinations of Lys-N generated peptides with MALDI-CID and ESI-ETD MS/MS provided very clear sequence ladders. Furthermore, the potential of this method for facilitating *de novo* sequencing was illustrated by the successful identification of proteins from an SDS-PAGE band of an ostrich tissue lysate, where ostrich represents a species with an unsequenced genome.

Materials and methods

Materials.

Protease inhibitor cocktail and Lys-C were purchased from Roche Diagnostics (Mannheim. Germany). Metalloendopeptidase Lys-N was obtained from Seikagaku Corporation (Tokyo, Japan). Iodoacetamide, trifluoroacetic acid (TFA) and α-cyano-4-hydroxycinnamic acid were purchased from Sigma-Aldrich (Steinheim, Germany). Dithiothreitol (DTT) was obtained from Fluka biochemical (Buchs, Switzerland). HEK293 cells were a gift from the ABC Protein Expression Center (Utrecht University, The Netherlands). Ostrich steak was purchased at the local butcher. Water that was used in these experiments was obtained from a Milli-Q purification system (Millipore, Bedford, MA). All other chemicals were purchased from commercial sources and were of analysis grade.

Sample preparation.

HEK293 cells were harvested at a density of ~1.5 x 10^6 cells/mL and stored at -20 °C. Cells were thawed and resuspended in ice-cold lysisbuffer (15 mL PBS, 150 µL Tween 20 and protease inhibitor cocktail). After dounce homogenizing on ice, the lysate was stored at 0 °C for 10 min. Subsequently, centrifugation at 20000 x g at 4 °C yielded separation of soluble and insoluble protein fractions. The soluble fraction was collected and the concentration determined by a Bradford assay. The lysate was dissolved in 50 mM ammonium bicarbonate to a concentration of 4 mg/mL.

Approximately 200 µg of ostrich muscle tissue was frozen in liquid nitrogen, pulverized with a mortar and pestle after which 8 M urea was added and the sample homogenized by microtip sonication. An amount of 30 µg of lysate was then separated by 1D SDS-PAGE.

In-solution digestion.

HEK293 lysate was reduced with 45 mM dithiothreitol (50 °C, 15 min) followed by alkylation using 110 mM iodoacetamide (in dark, RT, 15min). Buffer exchange was performed using 5 kDa spin columns. The resulting solutions were dried in a vacuum centrifuge and resuspended in 50 mM ammonium bicarbonate. One part was digested with Lys-C and an equal amount with Lys-N. Lys-C was added to the samples at a 1:50 (w/w) ratio while Lys-N was added at a ratio of 1:85 (w/w). Both solutions were incubated overnight at 25 °C.

In-gel digest.

A gel band at approximately 42 kDa was cut out of the gel and washed with water. After shrinking the gel piece with acetonitrile the contents were reduced with 10 mM of dithiothreitol (60 °C, 1 hour) followed by alkylation using 55 mM iodoacetamide (in dark, RT, 30min). After shrinking the gel pieces with acetonitrile the gel was incubated with Lys-N (10 ng/µL) overnight at 37 °C. Supernatant was transferred to new eppendorf tubes. Peptides were extracted by adding 50% acetonitrile, 5% formic acid to the gel pieces. The supernatant was added to the previous supernatant.

Strong cation exchange (SCX).

SCX was performed using an Agilent 1100 series LC-system with a C18 Opti-Lynx (Optimize Technologies, Oregon OR) guard column and Polysulfoethyl A SCX column (PolyLC, Columbia, MD; 200 mm x 2.1 mm i.d.). Sample was dissolved in 0.05% formic acid and loaded onto the guard column at 100 µL/min and consecutively eluted onto the SCX column with 80% acetonitrile, 0.05% formic acid. SCX buffer A was 5 mM KH_2PO_4, 30% acetonitrile, pH 2.7; SCX buffer B was 350 mM KCl, 5 mM KH_2PO_4, 30% acetonitrile, pH 2.7. Gradient elution was performed as follows: 0-85% B in 45 min, 85%-100% B in 6 min, 100% B for 4 min. A total of 53 one-minute fractions were collected and dried in a vacuum centrifuge.

Offline NanoRP-LC and MALDI Preparation.

NanoRP-LC separation of SCX fractions 31 to 33 and of the Lys-N in-gel digest of ostrich muscle tissue was performed on a Famos/Ultimate LC instrument (LC Packings, Naarden, the Netherlands), using a vented column set-up [26]. The trapping column was Aqua C18 (Phenomenex, Torrance, CA), 0.1 x 20 mm; analytical column was Aqua C18, 0.075x 230 mm. All columns were packed in-house. Trapping was performed at 5 µL/min for 10 min and analytical separation at 0.2 µL/min, passively split from 200 µL/min. Buffer A was 95% H_2O, 5% acetonitrile, 0.05% trifluoroacetic acid; buffer B was 5% H_2O, 95% acetonitrile, 0.05% trifluoroacetic acid. Gradient was 0-32% B in 35 min, 32-100% B in 2 min, 100% B for 5 min. 20-second fractions were automatically mixed with 0.5 µL MALDI matrix (3 mg/mL α-cyano-4-hydroxycinnamic acid, 80% acetonitrile, 0.1% trifluoroacetic acid) and spotted onto a MALDI target using a Probot Microfraction collector (LC Packings).

MALDI-TOF/TOF.

MALDI-TOF/TOF analysis was performed with a 4700 Proteomics analyzer (Applied Biosystems, Darmstadt, Germany). Spectra were acquired in positive and reflectron ion modes in the *m/z* range 900–4000. Maximally 1500 shots were averaged for each spectrum. Data were acquired at a laser repetition rate of 200 Hz, an acceleration voltage of 20 kV, a grid voltage of 70% and a digitizer bin size of 0.5 ns. The calibration of the spectra was done using a standard peptide calibration mixture (Applied Biosystems). CID spectra were obtained with a collisional energy of 1 keV and averaging maximally 15000 shots. Maximally 5 MS/MS precursors were selected per MS run and were excluded from further selection once sequenced.

NanoLC-ESI-CID-MS/MS and nanoLC-ESI-ETD-MS/MS.

An aliqout of SCX fraction 31 to 33 was also analyzed by nanoLC-CID/ETD-MS/MS . An Agilent 1100 HPLC system was connected to an LTQ XL Linear Ion Trap Mass Spectrometer with an ETD source at the back (Thermo Fisher Scientific Inc., Waltham, MA). The instrument was equipped with a 20 mm x 100 µm i.d. Aqua C18 trap column (Phenomenex, Torrance, CA), and a 200 mm x 50 µm i.d. Reprosil C18 RP analytical column (Dr. Maisch, Ammerbuch-Entringen, Germany). The fractions were separated by using a 95-minute 100 nL/min linear gradient from 0 to 60% solvent B (0.1 M acetic acid in 80% acetonitrile (v/v)), in which solvent A was 0.1 M acetic acid. The MS was operated in positive ion mode, and parent ions were isolated for fragmentation by CID or ETD in data-dependent mode. ETD fragmentation was performed with supplemental activation, fluoranthene was used as reagent anion and ion/ion reaction in the ion trap was allowed to take place for 100 ms.

Peptide identification.

MALDI data analysis and peak list generation was performed with the Data explorer™ software version 4.5 (Applied Biosystems). Raw ESI-CID and ESI–ETD MS data were converted to peak lists using Bioworks Browser software, version 3.3.1. For the work on the HEK293 lysate, spectra were searched against IPI Human (v3.37, 69164 entries searched) using Mascot (version 2.1.0), with Lys-C or Lys-N cleavage specificity allowing 1 missed cleavage, carbamidomethyl (C) as fixed modification, oxidation (M) as variable modification. Peptide tolerance was set to 100 ppm for 1+ peptide charge (MALDI) or 0.5 Da for 2+ and 3+ peptide charge (ESI) and MS/MS tolerance was 0.2 Da (MALDI) or 0.9 Da (ESI).

Peptides were identified with a minimum Mascot score of 30 and at these settings the false discovery rate (FDR) was less than 0.75% as estimated by using the Mascot decoy database function. For further data analysis, Mascot data were imported into Scaffold 1.7.
For the *de novo* sequencing of ostrich proteins the most abundant peptides in the MALDI MS spectra were fragmented. Manual annotation of CID spectra was performed using the mass differences between adjacent fragment ions. The obtained sequences were BLAST searched against a human and chicken UniProt database. Several assigned peptides were identical to peptides from chicken and/or human creatine kinase and actin. Other peptides were largely homologous, but revealed ostrich specific single or double amino acid differences.

Results

In a typical proteomics experiment, digestion is performed by trypsin, generating peptides with a C-terminal arginine or lysine. A mixture of N- and C-terminal fragment ions can be detected after CID of these peptides, but as the arginine and lysine residues are more basic than the α-amino group, generally, *y*-ions are more abundant[10, 19]. This paper evaluates the unique situation that arises when proteolysis is performed by Lys-N, as peptides are yielded with an N-terminus accommodating both α- and ε-amino basic entities. Fragmentation of Lys-N peptides is thus likely skewed towards the production of N-terminal ions (see Figure 1), as has been implied previously when looking at a few individual peptides[27, 28].

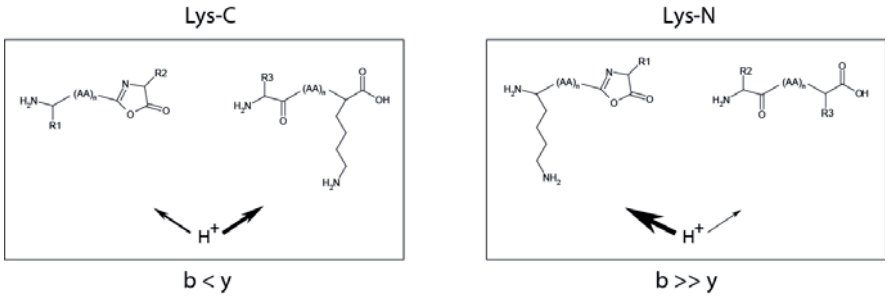

Figure 1: Schematic representation of CID fragmentation of Lys-C and Lys-N derived singly charged peptide ions. Lys-C peptides have a basic N- and C-terminus, therefore, both termini will be protonated, leading to a mixture of *b*- and *y*-ions in CID. Lys-N peptides concentrate the basicity at the N-terminus, leading to predominantly *b*-ions in CID.

To more comprehensively assess the fragmentation behavior of Lys-N produced peptides, a whole cellular lysate was digested in parallel with Lys-N and Lys-C. Lys-C, like trypsin, generates peptides with a C-terminal basic amino acid residue. However, Lys-C has no specificity for arginine and so comparing Lys-N with Lys-C is more appropriate than comparing it with trypsin. The digested cellular lysates were first subjected to low-pH SCX chromatography to enrich for and isolate peptides with a single basic lysine residue[24, 29, 30]. A few consecutive fractions of the SCX run will provide a set of peptides with a single C-terminal lysine residue for the Lys-C digested sample, and with a single N-terminal lysine from the Lys-N digest[24, 30-32]. Offline nanoRP-LC separation was performed on these selected SCX fractions. The eluent was subsequently mixed post-column with α-cyano-4-hydroxycinnamic acid and automatically fractionated and spotted onto a MALDI target plate, an experimental set-up adopted from our previously described offline ZIC-HILIC set-up[33]. The fractionated peptides were then subjected to analysis by MALDI-TOF/TOF based tandem mass spectrometry. As an illustrative result, fragmentation of the peptide KCQEVISWLDANTLAE, as depicted in Figure 2a, resulted in a CID spectrum that was typical for Lys-N proteolytic peptides (Figure 2b) and displayed a complete sequence ladder consisting of *b*-ions. As shown in Figure 2b, Lys-N peptide sequences can be easily read, as there are no significant 'interfering' ion-series present. As indicated in Figure 2, the b_1-ion is quite abundant and additionally, the b_1 related ions a_1 and a_1-NH_3 were found to be often the base peak in the spectra. C-terminal fragment ions (*y*-ions) were detected at a very low frequency and/or intensity.

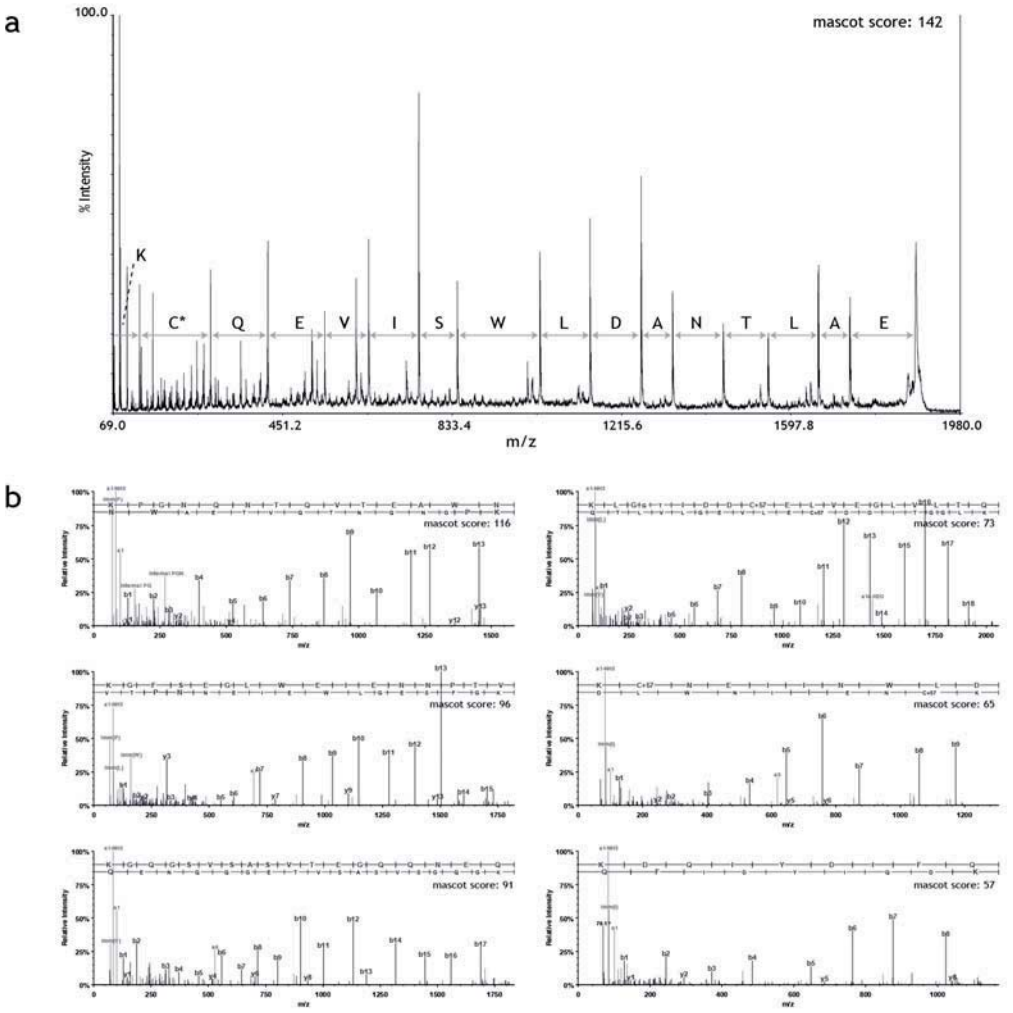

Figure 2: Representative MALDI-CID spectra of peptides identified from a Lys-N digested HEK293 cellular lysate. (a) A clean *b*-series sequence ladder is detected for peptide KC*QEVISWLDANTLAE (*m/z* 1876.73, 1+; C*, carbamidomethylated cysteine) (b) Six typical CID spectra of Lys-N peptides annotated by Scaffold, dominated by a nearly full series of *b*-ions, KPGNQNTQVTEAWN (*m/z* 1586.61, 1+); KGFSEGLWEIENNPTV (*m/z* 1819.85, 1+); KGQGSVSASVTEGQQNEQ (*m/z* 1833.79, 1+); KLGGTIDDC*ELVEGLVLTQ (*m/z* 2059.88, 1+); KC*NEIINWLD (*m/z* 1304.52, 1+); KDQIYDIFQ (*m/z* 1169.54, 1+).

We further examined and compared the effect of the N- or C-terminal position of the lysine on the fragmentation of the peptides in CID. Typical peptide fragmentation spectra are shown in Figure 3, which incorporates MALDI-CID spectra of the same peptide with a lysine either on the C-terminus or the N-terminus (respectively generated by Lys-C and Lys-N digestion) as well as spectra of the doubly charged ion of the same Lys-N peptide analyzed by ESI-CID MS/MS and ESI-ETcaD MS/MS. The MALDI tandem mass spectrum obtained for the peptide with a C-terminal lysine contains a mixture of *b*- and *y*-ions (and no complete series) and is clearly more complex than the spectrum from the analogous Lys-N peptide, which provides a nearly complete *b*-ion series. We found that fewer immonium ions were detected and the number of non-informative background peaks seemed to be lower in the spectra of the Lys-N peptide, possibly related to the reduced number of fragment ion pathways available. The tandem mass spectrum of the doubly charged ion of the Lys-N peptide obtained by ESI-CID also shows a complex spectrum with both *b*- and *y*-ions which can be attributed to the availability of two protons for the fragment ions[24]. Finally, the ETcaD spectrum of the doubly charged peptide ion is comprised of only *c*-ions, in agreement with our recent findings and in appearance similar to the MALDI-CID spectrum[24].

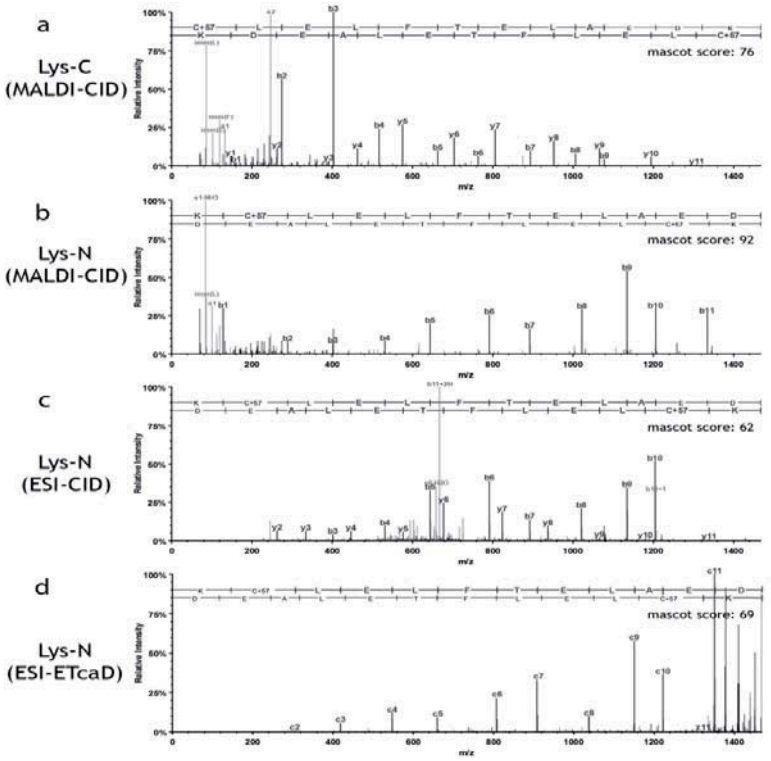

Figure 3: Representative mass spectra annotated by Scaffold of the same peptide with lysine respectively on (a) the C- or (b, c and d) the N-terminus of the peptide. In (a) (*m/z* 1467.60, 1+) and (b) (*m/z* 1467.58, 1+) are shown the MALDI-CID spectra, in (c) (*m/z* 734.20, 2+) the ESI-CID spectrum obtained by an ion trap and in (d) (*m/z* 734.26, 2+) the ESI-ETcaD-mass spectrum. MALDI-MS/MS of peptides with a C-terminal lysine provides a spectrum with both *b*- and *y*-ions, while spectra of peptides with an N-terminal lysine are less complex and dominated by *b*-ions. However, ESI-CID spectra of doubly charged peptides with an N-terminal lysine show a mixture of *b*- and *y*-ions due to the double charge, while the MALDI-TOF/TOF spectrum of the singly charged peptide results in a full *b* ion series. Also the ETD experiments result in mainly *c*-ion generation[24].

As stated, MALDI-CID of peptides with a single N-terminal lysine yielded mass spectra dominated by *b*-ions and with only minor *y*-ions. However, when a proline is present in the sequence this straightforward fragmentation pattern becomes disrupted. Peptide cleavage on the N-terminal side of a proline leads to preferential formation of a *y*-ion (Figure 4a). An illustrative example of a CID spectrum of a peptide containing two prolines is shown in

Figure 4b. The two most intense peaks correspond to *y*-ions with an N-terminal proline residue. Also, two intense internal fragments are detected with an N-terminal proline residue. Nevertheless, an almost complete and clear *b*-ion series is still detected.

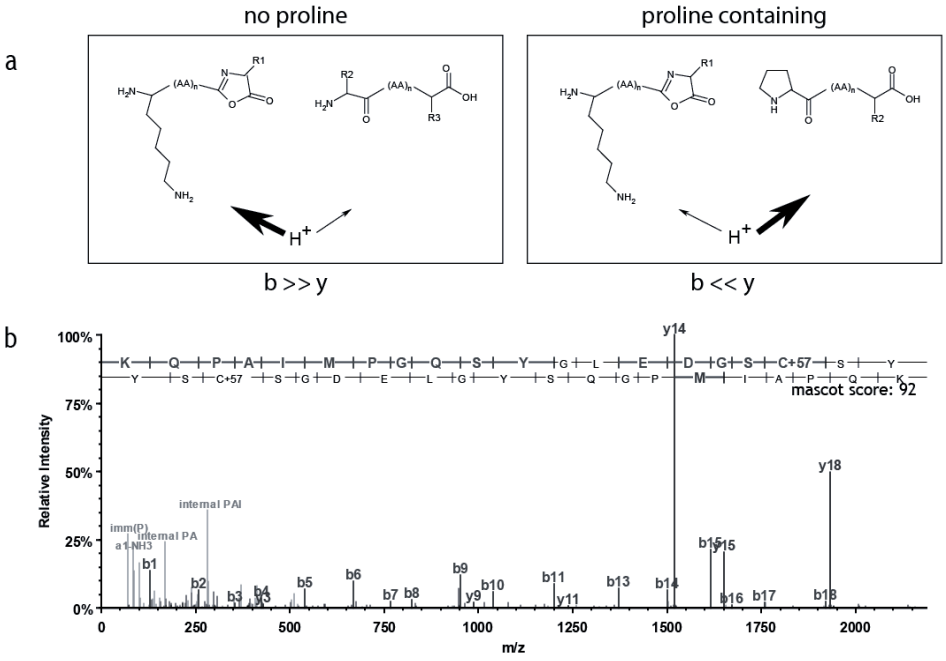

Figure 4: CID fragmentation of proline containing Lys-N peptides. (a) Schematic representation of CID fragmentation of Lys-N peptides with or without proline in the sequence. Fragmentation of a proline free peptide results in the detection of mainly *b*-ions, while cleavage N-terminal of a proline residue of a Lys-N peptide results in the detection of dominant *y*-ions. (b) Representative MALDI-CID spectrum of a proline containing Lys-N peptide (KQPAIMPGQSYGLEDGSC*SY, *m/z* 2187.78, 1+) annotated by Scaffold. The two intense *y*-ions correspond to ions with an N-terminal proline. Also two intense internal fragments with an N-terminal proline are detected. Note, that still a nearly full *b* ion series can be detected.

MALDI-CID analysis of 3 SCX fractions, enriched for peptides containing a single, N-terminal lysine, lead to the identification of a total of 247 peptides (Mascot score ≥30, FDR 0.75%) with an N-terminal lysine residue (tandem mass spectra are available as a PRIDE database, accession number 3380). Of these, 36 contained an extra basic residue and were initially removed from further analysis. Of the remaining 211 peptides, 119 contained one or more proline residues. To account for the proline effect on ion formation, proline containing

peptides were analyzed separately from proline free peptides. N-terminal *b*-ions represented in number 84% of the detected backbone fragment ion types in the 92 spectra of proline free Lys-N peptides, which corresponded to 94% of the total signal intensity of *b*- and *y*-ions, confirming its dominance (see Table 1). Furthermore, on average, two thirds of all theoretically obtainable *b*-ions were detected. A slightly lower percentage (75%) of *b*-ions was observed for proline containing peptides. If *y*-ions with an N-terminal proline were removed, the relative *b*-ion intensity percentage increased (81%), though not to the level of proline free peptides. Nevertheless, the dominant nature of *b*-ions in such spectra still allows straightforward interpretation (Figure 4). The relative intensity of *b*-ions in fragmentation spectra of Lys-C peptides is in sharp contrast with those of Lys-N peptides. In tandem mass spectra of Lys-C peptides *b*- and *y*-type ions are approximately equally abundant (see Table 1).

Table 1: Analysis of the frequency and normalized overall intensity of *b*-ions compared to *y*-ions in MALDI-CID spectra of Lys-N and Lys-C derived peptides

	Lys-N (211 peptides)		**Lys-C (220 peptides)**	
***b*-ion vs. *y*-ion**	**no proline**	**proline containing**	**no proline**	**proline containing**
Frequency (%)	84	75	46	48
Intensity (%)	94	71 (81*)	50	43

The percentage intensity is the intensity of all *b*-ions divided by the sum of intensity of all *b*- and *y*-ions. *After subtracting the intensity of *y*-ions with an N-terminal proline.

To evaluate the applicability of Lys-N digestion for genuine *de novo* sequencing, muscle tissue from an ostrich, of which the genome has not (yet) been sequenced, was lysed and separated by 1D SDS-PAGE. It should be noted that in our earlier work no significant difference in yield was found between in-gel digestion with trypsin or Lys-N[24] (Supplemental Figure 1 indicates the sequence coverage and peptide signal achieved by LC-MS for an in-gel digestion of equivalent amounts of BSA by Lys-N and trypsin). A band at approximately 42 kDa was excised from the ostrich gel and its content was digested with Lys-N. These peptides were then separated by 'offline' nano-LC with the eluent being mixed post-column with α-cyano-4-hydroxycinnamic acid and automatically spotted onto a MALDI target plate.

MALDI-TOF/TOF analysis was subsequently performed. CID spectra of approximately 20 of the most intense peptides were manually interpreted. Identified peptide sequences were subjected to homology searches, where the identified sequences were searched against UniProt human and chicken databases using BLAST. Four unique peptides were found to align with creatine kinase (Figure 5a), while five unique peptides were found to align with actin (Figure 5b). Strikingly, one of the actin peptides was found in four different forms. Alongside an unmodified form we observed sequences containing; methionine oxidation, methionine oxidation plus tryptophan hydroxylation and a decomposed carboxymethylated methionine[34]. Such modifications would be missed by a database search strategy if one does not, in advance, add these possibilities in the submission criteria. Three peptide sequences of creatine kinase and four sequences of actin were 100% homologous to human and chicken protein sequences allowing the identification of the protein. We found, for both proteins, peptides with sequences that differed slightly from the human and chicken sequences. In Figure 5c, the CID spectrum of such a peptide of creatine kinase is depicted. The clear *b*-ion series facilitated its *de novo* sequencing. The sequence deviates at the fourth (alanine) residue compared to chicken and human, while the 11th (glutamine) residue is similar to chicken, but different from human. In Figure 5d a CID spectrum is depicted of actin peptide KYPIEHAIITNWDDME. This peptide aligns with human and chicken actin except for residue seven (alanine). Additional annotated spectra can be found in Supplemental Figure 2.

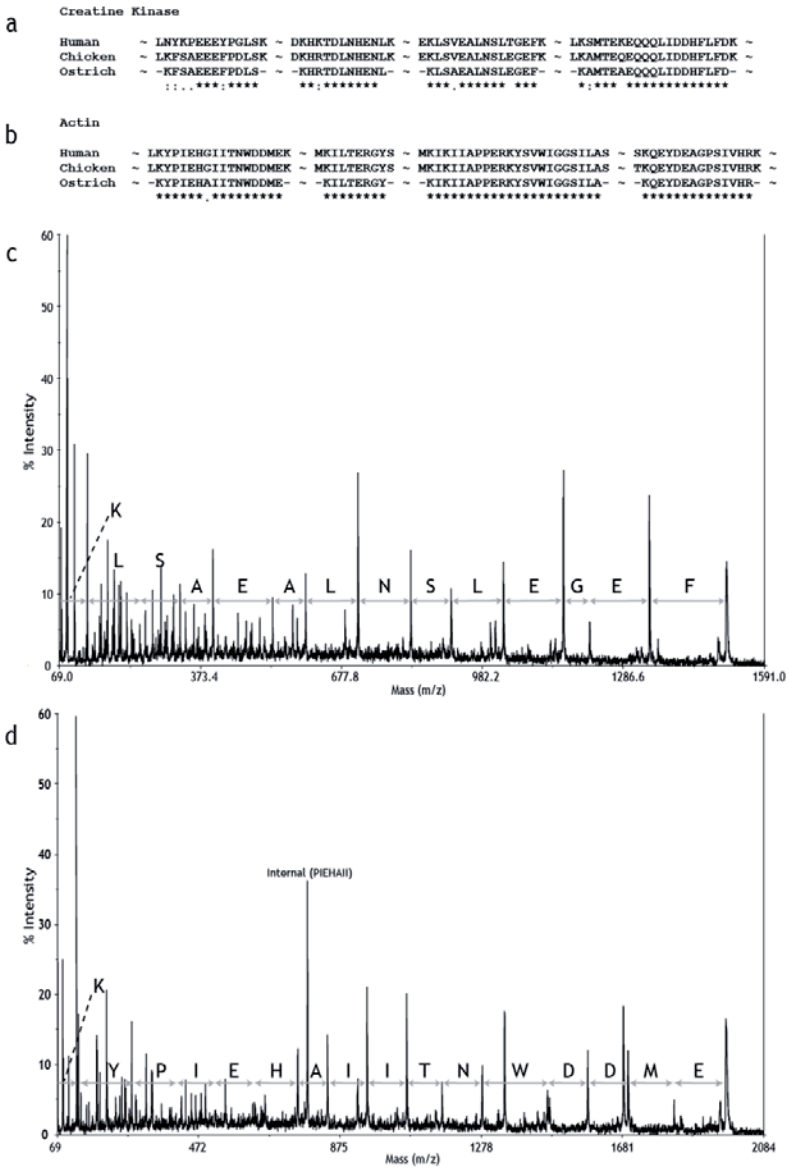

Figure 5: Lys-N facilitates *de novo* sequencing of ostrich creatine kinase and actin. (a) Multiple sequence alignment of *de novo* sequenced ostrich peptides with partial creatine kinase sequences of human (P06732) and chicken (P00565). (b) Multiple sequence alignment of de novo sequenced ostrich peptides with partial actin sequences of human (P62736) and chicken (P68139). (c) MALDI-CID spectrum of creatine kinase peptide KLSAEALNSLEGEF (*m/z* 1507.71, 1+) which has a V to A mutation of the fourth residue compared to chicken and human creatine kinase. (d) MALDI-CID spectrum of actin peptide KYPIEHAIITNWDDME (*m/z* 1974.84, 1+) which has a G to A mutation of the seventh residue compared to chicken and human actin.

Discussion

In the present study, we show that in MALDI-CID fragmentation of Lys-N peptides the basic N-terminus has a strong influence on fragment ion formation and leads to clean *b*-ion ladder series. These ladders are easily deciphered as the presence of interfering ion-series is significantly reduced or are altogether missing. We show that this is clearly different from MALDI-CID spectra of peptides with a C-terminal lysine, generated by a Lys-C protease, where a mixture of both *b*- and *y*-ion series was detected, as expected because both N-terminus and C-terminus contain basic entities. Furthermore, MALDI provides informative lower *m/z* ions including the b_1-ion (see Figure 3). An additional advantage is the good mass accuracy and resolution for these spectra since analyses were performed with a TOF mass analyzer. It should be noted that b_1-ions are typically missing in tandem MS of tryptic peptides[35]. The presence of the b_1-ion for Lys-N peptides is likely to originate through a distinct, lysine specific, cleavage pathway as it cannot be achieved via the regular b_x/y_z fragmentation pathway[36].

Evidently, CID of doubly charged Lys-N peptide ions is different from CID of singly charged Lys-N peptide ions. In ESI-CID MS/MS of the identical, Lys-N generated, doubly charged peptide ions, a mixture of *b*- and *y*-ions is detected with the *b*-ion series being somewhat more intense[24, 28, 30]. During fragmentation of doubly charged peptide ions, one of the protons will be sequestered by lysine. The other, mobile proton will then be less prone to protonate the N-terminus and thus the possibility of the formation of also C-terminal ions is increased.

In order to isolate and obtain a statistically significant number of Lys-N peptides with a single basic residue an initial low pH SCX chromatographic step was added[31]. However, the fact that we chose to isolate peptides with a single N-terminal lysine does not mean that the remaining peptides, that contain for example more than one basic residue, are of no value. Tandem mass spectra can likewise be obtained from Lys-N peptides with additional basic residues, where the fragmentation will follow a pattern similar to that achieved with tryptic peptides containing mis-cleavages. Although, containing *b*- and *y*-ion series, they are not as straightforward to interpret manually as the spectra from peptides with a single N-terminal lysine. These spectra have similar appearances to those originating from tryptic or Lys-C peptides (Supplemental Figure 3).

Scrutinizing spectra for additional trends, we observed that the MALDI-CID spectra of proline containing peptides contained intense peaks corresponding to fragment ions with an N-terminal proline, a well described phenomenon[4, 37-39]. The intensity of these peaks has been

explained by the extraordinary structure of proline, with its side chain forming a five-membered ring with the peptide backbone. This hinders cleavage C-terminal of proline, as it would involve the generation of a strained bicyclic structure[39]. Cleavage on the N-terminal side causes the formation of the proline's secondary amine group which can sequester the mobile proton, causing the generation of the C-terminal fragment ion (see Figure 4b)[38]. Our data suggest that a combination of both effects prompt the emergence of intense peaks of *y*-ions with an N-terminal proline. Firstly, the fact that the intensity of a *b*-ion with a C-terminal proline is significantly lower than other *b*-ions indicates that the formation of a *b*-ion with a bicyclic C-terminal proline structure is unfavored[39]. Secondly, the high gas-phase proton affinity of the proline amine group is reflected in the intensity of the *y*-ion with an N-terminal proline being substantially larger than the intensity of the *b*-ion that is generated N-terminal of the same proline. It appears thus that the two amino groups at the N-terminus of the potential *b*-ion exert less influence than the secondary amine of the proline present on the *y*-ion at the point of cleavage[38]. On average, *b*-ions in MALDI CID spectra of proline containing peptides account for 71% of the total intensity of *b*- and *y*-ions. When the intense *y*-ions corresponding to fragments with an N-terminal proline were removed, this percentage of *b*-ions increased to 81%. This is slightly lower than for proline free peptides. Despite the occurrence of *y*-ions, the remainder of such Lys-N MALDI-CID spectra is still dominated by sequence ladders of *b*-ions. About 56% of the peptides we identified with a single N-terminal lysine contained a proline residue. This is similar to a recent study in which, the occurrence of at least one proline in a tryptic peptide was determined, both theoretically (using the IPI-Human database) and practically, to be around 50%[40].

Efficient and facile *de novo* sequencing requires good quality straightforward mass spectra. Lys-N proteolysis allows significant portions of a protein to generate simplified MALDI-CID spectra without any derivatization steps. Therefore, as an example of the applicability of Lys-N proteolysis for *de novo* sequencing, we strived to identify the protein contents of a 1D SDS-PAGE gel band of ostrich muscle tissue. Clear, simplified spectra were generated facilitating *de novo* sequencing. Sequenced peptides could be aligned to chicken (the phylogenetically closest species to ostrich with a genome that is sequenced) creatine kinase and actin. Of these peptides, at least two would not have been identified if traditional database search strategy had been performed since these peptides slightly differ in sequence compared to sequences available in genomic databases. For example, in the ostrich creatine kinase peptide KLSAEALNSLEGEF, the third residue, an alanine, is a valine in chicken and human, which is probably a DNA point mutation as the translation codons of valine and

alanine only differ by one nucleotide. The same is true for the actin peptide KYPIEHAIITNWDDME, where a single DNA point mutation could explain the conversion of glycine in chicken and human to alanine in ostrich. Furthermore, modifications to this peptide were found in other mass spectra, which could easily be detected, including methionine oxidation, tryptophan hydroxylation and decomposed carboxymethylated methionine. Although methionine oxidation is generally included as a variable modification in database searching, for tryptophan hydroxylation and decomposed carboxymethylated methionine this is generally not the case thus underscoring the potential of our *de novo* sequencing approach. Similarly, we expect that other modifications can be easily identified and located using the MS/MS sequence ladders by combining Lys-N digestion with MALDI-MS/MS analysis. Labile modifications such as phosphorylation might result in neutral loss dominated spectra as fragmentation is performed by CID. However, these spectra will still be simpler than those achieved by tryptic peptides.

As discussed before, an apparent drawback of Lys-N MALDI-CID is that the clear *b*-ion dominated spectra will exclusively be observed for peptides containing a single N-terminal basic residue. Depending on the position in the peptide sequence, an extra basic residue might lead to the generation of an additional (*y*-) ion series thereby complicating manual interpretation. Chemical derivatization that aid *de novo* sequencing such as CAF and SPITC can be applied but it is necessary to have a basic residue at the C-terminus thus tryptic peptides are necessary. In theory, the number of peptides with a single basic residue in a tryptic digest is approximately three fold higher than in a Lys-N digest. However, internal basic residues negatively affect these derivatization strategies in a similar way to Lys-N, i.e. generating additional fragment ion series[11, 16, 18, 19]. Furthermore, the sulfonyl group, important for the CAF and SPITC approaches, reduces peptide signal intensities due to the intrinsic negative charge. Also these chemical derivatization steps potentially lead to sample loss and further signal reduction. Finally, side reactions (often) hamper the analysis by increasing the complexity of the sample with uninformative peptides. Through the use of Lys-N *de novo* sequencing peptides can be attained without chemical derivatization.

Conclusion

In summary, we evaluated here MALDI-CID fragmentation of singly charged Lys-N generated peptide ions. With a lysine residue on the peptide N-terminal, protonation of N-terminal fragments of these peptides is favored, resulting in the detection of dominant, nearly complete series of *b*-ions, rendering Lys-N a useful protease aiding in the unambiguous sequencing of peptides in MALDI-MS/MS. Lys-N can be applied on a proteome scale using a low pH SCX MudPIT strategy where one can isolate and separate single lysine containing peptides that represent the whole protein content. Equally, Lys-N is also applicable in a 1D or 2D-gel strategy, where a significant portion of generated peptides for each protein can be *de novo* sequenced in a similar manner to CAF and SPITC.

Acknowledgements

We thank Dr. M.M. Drugan for performing statistical analyses. This work was supported by the Netherlands Proteomics Centre (www.netherlandsproteomicscentre.nl).

Supplementary data

Supplementary Figure 1: Comparison of in-gel digestion with trypsin and Lys-N.

Supplementary Figure 2: Additional MALDI-CID spectra to figure 5. (A-E) Creatine Kinase peptide sequences

Supplementary Figure 3: MALDI-CID spectra of similar Lys-C (upper panels) and Lys-N (lower panels) peptides containing an internal basic residue

References

1. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** 422, (6928), 198-207.
2. Chalmers, M. J.; Gaskell, S. J., Advances in mass spectrometry for proteome analysis. *Curr. Opin. Biotechnol.* **2000,** 11, (4), 384-390.
3. Olsen, J. V.; Ong, S.-E.; Mann, M., Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues. *Mol. Cell. Proteomics* **2004,** 3, (6), 608-614.
4. Paizs, B.; Suhai, S., Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.* **2005,** 24, (4), 508-548.
5. Burlet, O.; Yang, C. Y.; Gaskell, S. J., Influence of Cysteine to Cysteic Acid Oxidation on the Collision-Activated Decomposition of Protonated Peptides - Evidence for Intraionic Interactions. *J. Am. Soc. Mass Spectrom.* **1992,** 3, (4), 337-344.
6. Harrison, A. G.; Yalcin, T., Proton mobility in protonated amino acids and peptides. *Int. J. Mass. Spectrom.* **1997,** 165, 339-347.
7. Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A., Special feature: Commentary - Mobile and localized protons: a framework for understanding peptide dissociation. *J. Mass Spectrom.* **2000,** 35, (12), 1399-1406.
8. Roepstorff, P.; Fohlman, J., Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass. Spectrom.* **1984,** 11, (11), 601.
9. Wattenberg, A.; Organ, A. J.; Schneider, K.; Tyldesley, R.; Bordoli, R.; Bateman, R. H., Sequence dependent fragmentation of peptides generated by MALDI quadrupole time-of-flight (MALDI Q-TOF) mass spectrometry and its implications for protein identification. *J. Am. Soc. Mass Spectrom.* **2002,** 13, (7), 772-783.
10. Khatun, J.; Ramkissoon, K.; Giddings, M. C., Fragmentation Characteristics of Collision-Induced Dissociation in MALDI TOF/TOF Mass Spectrometry. *Anal. Chem.* **2007,** 79, (8), 3032-3040.
11. Chen, W.; Lee, P. J.; Shion, H.; Ellor, N.; Gebler, J. C., Improving de Novo Sequencing of Peptides Using a Charged Tag and C-Terminal Digestion. *Anal. Chem.* **2007,** 79, (4), 1583-1590.
12. Palagi, P. M.; Hernandez, P.; Walther, D.; Appel, R. D., Proteome informatics I: Bioinformatics tools for processing experimental data. *Proteomics* **2006,** 6, (20), 5435-5444.
13. Zaia, J., Charged Derivatives for Peptide Sequencing Using a Magnetic Sector Instrument. *Methods. Mol. Biol.* **1996,** 61, 29-41.
14. Spengler, B.; Luetzenkirchen, F.; Metzger, S.; Chaurand, P.; Kaufmann, R.; Jeffery, W.; Bartlet-Jones, M.; Pappin, D. J. C., Peptide sequencing of charged derivatives by postsource decay MALDI mass spectrometry. *Int. J. Mass Spectrom.* **1997,** 169, 127-140.
15. Roth, K. D. W.; Huang, Z.-H.; Sadagopan, N.; Watson, J. T., Charge derivatization of peptides for analysis by mass spectrometry. *Mass Spectrom. Rev.* **1998,** 17, (4), 255-274.
16. Keough, T.; Youngquist, R. S.; Lacey, M. P., A method for high-sensitivity peptide sequencing using postsource decay matrix-assisted laser desorption ionization mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **1999,** 96, (13), 7131-7136.
17. Flensburg, J.; Tangen, A.; Prieto, M.; Hellman, U.; Wadensten, H., Chemically-assisted fragmentation combined with multi-dimensional liquid chromatography and matrix-assisted laser desorption/ionization post source decay, matrix-assisted laser desorption/ionization tandem time-of-flight or matrix-assisted laser desorption/ionization

tandem mass spectrometry for improved sequencing of tryptic peptides. *Eur. J. Mass. Spectrom.* **2005,** 11, (2), 169-179.
18. Brancia, F. L.; Oliver, S. G.; Gaskell, S. J., Improved matrix-assisted laser desorption/ionization mass spectrometric analysis of tryptic hydrolysates of proteins following guanidination of lysine-containing peptides. *Rapid Commun. Mass. Spectrom.* **2000,** 14, (21), 2070-2073.
19. Peters, E. C.; Horn, D. M.; Tully, D. C.; Brock, A., A novel multifunctional labeling reagent for enhanced protein characterization with mass spectrometry. *Rapid Commun. Mass. Spectrom.* **2001,** 15, (24), 2387-2392.
20. Gevaert, K.; Demol, H.; Martens, L.; Hoorelbeke, B.; Puype, M.; Goethals, M.; Van Damme, J.; De Boeck, S.; Vandekerckhove, J., Protein identification based on matrix assisted laser desorption/ionization-post source decay-mass spectrometry. *Electrophoresis* **2001,** 22, (9), 1645-1651.
21. Marekov, L. N.; Steinert, P. M., Charge derivatization by 4-sulfophenyl isothiocyanate enhances peptide sequencing by post-source decay matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Mass Spectrom.* **2003,** 38, (4), 373-377.
22. Keough, T.; Lacey, M. P.; Youngquist, R. S., Derivatization procedures to facilitate de novo sequencing of lysine-terminated tryptic peptides using postsource decay matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass. Spectrom.* **2000,** 14, (24), 2348-2356.
23. Regnier, F. E.; Samir, J., Primary amine coding as a path to comparative proteomics. *Proteomics* **2006,** 6, (14), 3968-3979.
24. Taouatas, N.; Drugan, M. M.; Heck, A. J. R.; Mohammed, S., Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase. *Nat. Methods* **2008,** 5, (5), 405-407.
25. Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E.; Coon, J. J., Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal. Chem.* **2007,** 79, (2), 477-85.
26. Meiring, H. D.; van der Heeft, E.; ten Hove, G. J.; de Jong, A., Nanoscale LC-MS(n): technical design and applications to peptide and protein analysis. *J. Sep. Sci.* **2002,** 25, (9), 557-568.
27. Takio, K., In *Handbook of Proteolytic Enzymes*, 2nd ed.; Elsevier Scientific Publishing Company, Amsterdam, the Netherlands: 2004; pp 788-791.
28. Rao, K. C. S.; Palamalai, V.; Dunlevy, J. R.; Miyagi, M., Peptidyl-Lys Metalloendopeptidase-catalyzed 18O Labeling for Comparative Proteomics: Application to Cytokine/Lipolysaccharide-treated Human Retinal Pigment Epithelium Cell Line. *Mol. Cell. Proteomics* **2005,** 4, (10), 1550-1557.
29. Dormeyer, W.; Mohammed, S.; Breukelen, B.; Krijgsveld, J.; Heck, A. J., Targeted analysis of protein termini. *J. Proteome Res.* **2007,** 6, (12), 4634-45.
30. Taouatas, N.; Altelaar, A. F. M.; Drugan, M. M.; Helbig, A. O.; Mohammed, S.; Heck, A. J. R., SCX-based fractionation of Lys-N generated peptides facilitates the targeted analysis of post-translational modifications. *Mol. Cell. Proteomics* **2008**, M800285-MCP200.
31. Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J. X.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P., Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U.S.A.* **2004,** 101, (33), 12130-12135.

32. Pinkse, M. W. H.; Mohammed, S.; Gouw, L. W.; van Breukelen, B.; Vos, H. R.; Heck, A. J. R., Highly robust, automated, and sensitive on line TiO2-based phosphoproteomics applied to study endogenous phosphorylation in Drosophila melanogaster. *J. Proteome Res.* **2008,** 7, (2), 687-697.
33. Boersema, P. J.; Divecha, N.; Heck, A. J. R.; Mohammed, S., Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. *J. Proteome Res.* **2007,** 6, (3), 937-946.
34. Jones, M. D.; Merewether, L. A.; Clogston, C. L.; Lu, H. S., Peptide Map Analysis of Recombinant Human Granulocyte Colony Stimulating Factor: Elimination of Methionine Modification and Nonspecific Cleavages. *Anal. Biochem.* **1994,** 216, (1), 135-146.
35. Fu, Q.; Li, L. J., De novo sequencing of neuropeptides using reductive isotopic methylation and investigation of ESI QTOF MS/MS fragmentation pattern of neuropeptides with N-terminal dimethylation. *Anal. Chem.* **2005,** 77, (23), 7783-7795.
36. Yalcin, T.; Harrison, A. G., Ion chemistry of protonated lysine derivatives. *J. Mass Spectrom.* **1996,** 31, (11), 1237-1243.
37. Breci, L. A.; Tabb, D. L.; Yates, J. R.; Wysocki, V. H., Cleavage N-Terminal to Proline: Analysis of a Database of Peptide Tandem Mass Spectra. *Anal. Chem.* **2003,** 75, (9), 1963-1971.
38. Hunt, D. F.; Yates, J. R.; Shabanowitz, J.; Winston, S.; Hauer, C. R., Protein Sequencing by Tandem Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **1986,** 83, (17), 6233-6237.
39. Vaisar, T.; Urban, J., Probing the proline effect in CID of protonated peptides. *J. Mass Spectrom.* **1996,** 31, (10), 1185-1187.
40. Van Hoof, D.; Pinkse, M. W. H.; Ward-Van Oostwaard, D.; Mummery, C. L.; Heck, A. J. R.; Krijgsveld, J., An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. *Nat. Methods* **2007,** 4, (9), 677-678.

Chapter 5

Evaluation of metalloendopeptidase Lys-N protease performance under different sample handling conditions

Nadia Taouatas, Albert J.R. Heck and Shabaz Mohammed

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

Based on J Proteome Res. 2010, (8), 4282-4288

Abstract

Trypsin, the most widely used enzyme in proteomics, has a few caveats as it does not perform well under certain harsh sample handling conditions and creates relatively short peptides less amenable to for instance electron transfer dissociation. There is, thus, room for improvement using alternative proteases. Here, we evaluate the performance of such an alternative protease, the metalloendopeptidase Lys-N, in sample preparation for proteomic analyses under various experimental conditions. The experimental parameters we evaluated were protein-to-protease ratio, incubation time, temperature, and several concentrations of denaturing modifiers often used in proteomics sample handling. Our data reveal that Lys-N is still very efficient under some very harsh (denaturing) conditions (e.g. 8M urea, 80% acetonitrile), and at temperatures as low as 4 ^{o}C and up to 80 ^{o}C, but severely hampered by guanidine hydrochloride and methanol. These rather unique features make Lys-N a good candidate for a variety of applications, such as membrane proteomics and possibly H/D exchange mass spectrometry. Additionally, we show that Lys-N is capable of, in contrast to trypsin or Lys-C, cleaving adjacent to mono- and di-methylated lysines, making it a good candidate for targeted epigenetic analysis of for instance histones.

Keywords: protease efficiency, sample preparation, digestion, Lys-N, methylated lysine

Introduction

Sample preparation is one of the most crucial processes in proteomics research. At present proteome analyses typically starts with the proteolytic digestion of all proteins present in the lysate, where after the resulting peptide mixture is separated by chromatographic means and analyzed by mass spectrometry (MS)[1]. A commonly used MS-based proteomics approach employs a combination of trypsin proteolysis for peptide generation and collision induced dissociation (CID) for peptide ion fragmentation and identification[2]. Alternative peptide ion fragmentation methods have recently been introduced, such as electron capture dissociation (ECD)[3] and electron transfer dissociation (ETD)[4]. They are quite complementary to CID and are particularly useful for the analyses of highly basic peptides[5] and for locating post translational modifications[6-8]. Although, trypsin is by far the most common enzyme used, a few groups are actively exploring alternative proteases, which can result in cleavage products more suitable for mass spectrometry analysis using ETD or ECD[9-12]. Furthermore, combinations of different enzymes have been implemented, which can lead to a more comprehensive coverage of protein sequences[13-18].

Although trypsin has been used for many years its activity/purity/selectivity is still being investigated and optimized. For instance, Pham *et al.* studied the heat stability of trypsin and found that glycated trypsin greatly increase the thermostability of the enzyme compared to the native enzyme[19]. This can be useful when working with native proteins that do not readily digest without first denaturing the protein to achieve efficient digestion. Recent studies performed by Petritis *et al.* found that boiling trypsin results in a complete deactivation of the enzyme, which can be used to an advantage to prevent oxygen back-exchange when using ^{18}O-labeled samples. Interestingly, their data indicates that the boiling procedure does not completely quench trypsin activity when organic co-solvents were present[20]. Different trypsin based digestion conditions were studied by Chen *et al.* to improve proteolytic processing of complex protein mixtures, whereby several MS-compatible detergents were found to be useful to increase the total number of identified peptides[21]. Although trypsin is still the *de facto* choice of enzyme for proteomics it has limitations as it is somewhat inefficient when dealing with certain subgroups of proteins and is not efficient in certain solvents[22]. A few proteases that are more tolerant to a wider range of solvents (and or organic additives) have therefore been introduced into the field of proteomics to circumvent the latter problem.

Many of these efforts have focused on membrane proteins attempting to improve the accessibility for digestion, which is hindered by their trans-membrane regions. Rietschel *et*

al. described a protocol optimized for membrane proteomics where the combination of elastase and methanol for a liquid chromatographic (LC)-based membrane proteome analysis resulted in an increased identification of trans-membrane peptides[23]. The enzyme pepsin was used by Han and Schey for the characterization of an aquaporin[24, 25].

Alternative proteases to trypsin have also been of interest in phospho-proteomics studies[10, 13, 14]. For instance, Schlosser *et al.* combined elastase digestion with phosphopeptide enrichment and tandem mass spectrometry for improved identification of phosphorylation sites[26].

Here we focus on a relatively little explored protease in the field of proteomics, which is a metalloendopeptidase referred to as Lys-N[27]. The metalloendopeptidase Lys-N is unique in the sense that it has enzymatic specificity for the N-terminal side of lysine residues and has only been implemented in a few proteomics applications[9, 11, 28-31]. Gauci *et al.* evaluated the complementarities of Lys-N, Lys-C and trypsin for their ability to contribute to the global analysis of the phosphoproteome[15]. In addition, Lys-N has been described to aid in the *de novo* sequencing of peptides by mass spectrometry as Lys-N generated peptides, that contains a single basic residue, fragmented by electron transfer dissociation (ETD) mass spectrometry or MALDI CID, display very simple fragmentation patterns containing predominantly N-terminal fragment ions[9, 30].

Here, we evaluated, in detail, the performance of Lys-N for the use in different proteomic sample preparation protocols. We used a protein standard (bovine serum albumin, BSA) as a model system for Lys-N digestion and analyzed the resulting peptides by mass spectrometry. We explored the stability of Lys-N during different digestion conditions by subjecting the protease to different solvents and temperatures. We also evaluated the ability of Lys-N to cleave modified lysine residues, using synthetic peptides containing mono, - di, - or tri-methylated lysines.

Materials and methods

Materials.

The metalloendopeptidase was isolated from Grifola Frondosa (Lys-N) and was obtained from Seikagaku Corporation (Tokyo, Japan). Iodoacetamide was obtained from Sigma-Aldrich (Steinheim, Germany). DL-Dithiothreitol was obtained from Fluka Biochemical (Steinheim, Germany). HPLC-S gradient grade acetonitrile was purchased from Biosolve (Valkenwaard, The Netherlands). Acetic acid was obtained by MERCK KGaA (Damstadt, Germany) and high purity water obtained from Milli-Q system (Millipore, Bedford, MA).

Sample preparation.

BSA in-solution digestion.

The protein Bovine serum albumin (BSA) was used as a model system for the digestion experiments where 10 µg of protein was used for each of the digestion experiments performed. All the BSA samples were reduced with 45 mM dithiothreitol (50 °C, 15 min) followed by alkylation using 110 mM iodoacetamide (dark, RT, 15min). The appropriate solvents (urea, guanidine hydrochloride, methanol, acetonitrile, ammonium bicarbonate, acetic acid or ammonium acetate) were used to dilute each sample to a final volume of 200 µl. The protein samples were digested with Lys-N where Lys-N was added at a ratio of 1/50 to 1/1000 (w/w). The samples were incubated from 1 min to 24 hours and incubated from 4°C to 80°C. At selected time-points the digestion was brought to an end by diluting the sample into cold (4°C) quenching solution (1% acetic acid) to impede the proteolytic activity of Lys-N.

Peptide in-solution digestion of Lysine methylated peptides.

Four different versions of the synthetic peptide H3 (EIAQDFKTDLR, wherein the lysine was modified with either 0, 1, 2 or 3 methylations) were digested with Lys-N. Lys-N was added to the peptides at a ratio of 1/100 (w/w) and incubated over night at 37°C.

Mass spectrometry and Database searches.

Digested BSA was subjected to nanoscale liquid chromatography tandem mass spectrometry (nano-LC-MS/MS) analysis, performed on an Agilent 1100 HPLC system (Agilent Technologies) connected to a hybrid mass spectrometer consisting of a linear ion trap (LTQ) and a FT-ICR mass spectrometer from Thermo Scientific, Inc. (Waltham, MA) essentially as

described previously [32]. Briefly, the samples were diluted in 5 % formic acid and injected (20 fmol of BSA) on the trap column (Aqua C18 (phenomenex, Torrance, CA)), 20 mm x 100 µm I.D.) at a flow rate of 5 µL / min. The peptides were transferred with a split-reduced flow rate of 100 nL / min solvent A (0.1 M acetic acid) onto an analytic column (Reprosil C18 RP (Dr Maisch, Germany), 20 cm x 50 µm I.D.). Elution of peptides from digested BSA was achieved with a linear gradient from 0 to 40% B (acetonitrile/water (v/v) containing 0.1 M acetic acid) in 30 min. The column effluent was directly introduced into the ESI source of the MS.

The LTQ-FT-ICR mass spectrometer was operated in positive ion mode, from 350 to 1500 *m/z* in MS mode and with an AGC value of 5.00e+05. The two most intense parent ions (with a threshold above 500) were isolated and fragmented by CID in data-dependent mode with an AGC value of 1.00e+04. Ions were fragmented in the linear ion trap using CID with normalized collision energy of 35 and 30 ms activation time.

Database search.

Raw MS data were converted to peak lists using Bioworks Browser software, version 3.1.1. After MS measurements data was analyzed with the MASCOT software version 2.2.0 (www.matrixscience.com). The database search was made using mono-isotopic masses with parameters set to consider a peptide tolerance of ± 30 ppm, a fragment tolerance of ± 0.5 Da, a static modification of + 57.021464 Da on cysteine residues (Carbamidomethylation) and a differential modification of + 15.994915 Da on methionine (oxidation). CID spectra of digested BSA were searched in a BSA database.

Data analysis.

Spectral counting was used to evaluate the protease activity in the BSA digestion experiments. Normalization of the data was performed by using the experiment generating the highest number of MS/MS spectra identified as the maximum result[33]. Spectral counting was performed by allowing 2 Missed Cleavages and using all spectral identifications with an ions score of 20. All results are presented in the form of a scaffold file, which can be downloaded from the ProteomeCommons.org Tranche network using the following hash:

eHA9Xf7Ia45k5lWGvHjWN9U+asmvim/JAFI7073tUDOwUa98E5ce+SzHVIuWJAy8ldtYt
youxi5CHKnX/0aHx2E1giEAAAAAAAADsg==.

Results and discussions

The protease Lys-N belongs to the zinc metalloendopeptidase superfamily. Following initial isolation by Nonaka *et al*., the protease was characterized for specificity and efficiency, whereby it was found that it had several remarkable properties including specific cleavage of N-terminal acyl-lysine bonds. Furthermore, they assessed the thermostability, and binding properties to β-glycans and its optimal activity in the alkaline range of pH 9-10[27, 34, 35]. Here, we explore the properties of Lys-N for use in several alternative proteomics workflows. We characterize the performance of the enzyme when subjecting it to more 'extreme' conditions such as high incubation temperatures, and high concentration of organic solvents and denaturing agents. For our model system we used bovine serum albumin (BSA) as it is an easily digestible protein and because it is commonly used as a performance indicator for proteomics workflows.

Evaluation of optimal enzyme/substrate ratio.

First, the optimal enzyme to substrate ratio was evaluated. We therefore applied 8 different enzyme to substrate ratios (1/50, 1/85, 1:100, 1/200, 1/400, 1/600, 1/800 and 1/1000) for the digestion experiments. All the samples were incubated for 4 hours at the standard digestion temperature of 37°C. A fraction of each sample (5µl) was transferred, after 4 hours of digestion, to 100 µl of cold (4°C) quench solution containing 1% acetic acid (see Figure 1 for scheme). The performance metric chosen for the BSA digestion experiments was spectral counting[33, 36, 37]. Normalization was performed, to obtain a scaled protease efficiency index, using the experiment generating the highest number of MS/MS spectra identified as the maximum result. The same scaled protease efficiency index is used throughout the further experiments described here. The performance of Lys-N at the different protease to protein ratios is presented in Supplementary Figure 1. The performance of Lys-N was found not to change significantly at ratios above 1/100. Using this enzyme to substrate ratios a BSA sequence coverage of more than 50% could be obtained using ion trap CID. An example of a typical BSA LC chromatogram is given in Figure 1. A significant drop in protease efficiency was observed when diluting the enzyme to substrate ratio below 1/100. For the subsequent experiments we chose a protease to substrate ratio of 1/200, partly to restrict the enzyme amounts available for BSA digestion, enabling us to monitor differences in Lys-N performance more easily under the chosen different experimental conditions.

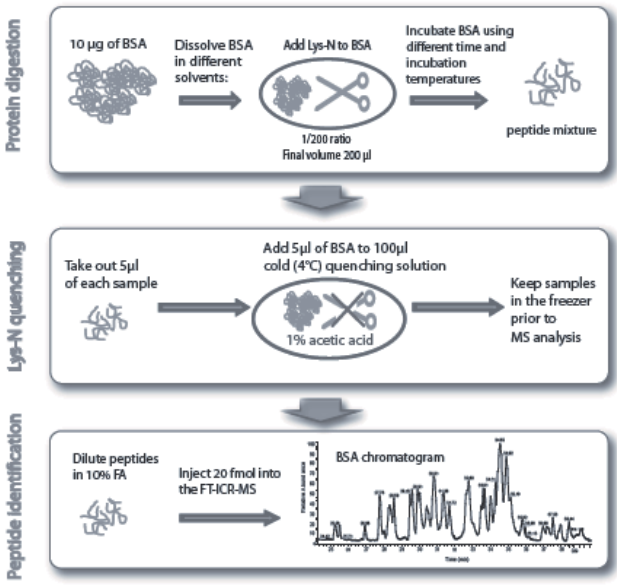

Figure 1: Overview of experiments. The model system protein BSA (10 µg) was reduced, alkylated and diluted in different solvents. Lys-N was added at a ratio of 1/200 to a final volume of 200 µl and incubated at different temperatures and at different incubations times. A certain volume (5 µl) was taken out from each sample after a specific time point and added to a cold (4°C) quench solution containing 1% acetic acid. Each sample was diluted in 10% formic acid (FA) and 20 fmol was injected for mass spectrometry analysis.

Subsequently, we evaluated the proteolytic activity of Lys-N with respect to optimal incubation time. The experiment was performed by incubating the enzyme with BSA (ratio 1/200) at 37°C for either 1 min, 10 min, 30 min, 1 hour, 2 hours, 4 hours, 6 hours, 12 hours or 24 hours (Figure 2). The number of peptides identified varied with the different incubation times where the highest number of peptides identified was after 2 hours of incubation. A decrease in peptide identification is observed when the digestion time was reduced to 1 hour, 30 minutes or 10 minutes and hardly any peptides of BSA were generated using only 1 minute of digestion. What is surprising is the decreasing number of peptides identified after increasing the incubation beyond 2 hours, i.e. up to 24 hours. comparable observations were made by Klammer and MacCoss in their similar study on the performance of trypsin using different digestions protocols[38]. Their rationale was that proteins could be degraded to peptides of a molecular weight lower than detectable by standard mass spectrometry analysis partially caused by a higher level of unspecific cleavages. Performing a no-enzyme database search on the BSA dataset indicated a modest number of unspecific cleavages, which increased over time (Supplementary Table 1). The 'unspecific' cleavages occurred at predominantly alanine, serine and arginine residues, in agreement with a previous report[28]. Cleavages adjacent to acidic residues were also observed but this is more likely to be related to chemical hydrolysis. Additional possibilities include peptide adsorption and peptide

instability which may lead to the decreased detection of BSA peptides at longer incubation times[39].

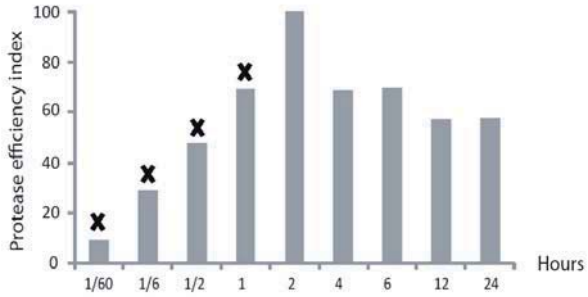

Figure 2: Protease efficiency of Lys-N at different incubation times. BSA digestion was performed using a protease to protein ratio of 1/200 and incubating the sample at 37°C. The protease efficiency index is depicted on the y-axis (see text for description). According to the results obtained here Lys-N showed maximum activity using 2 hours of incubation time at 37°C. The 4 crosses on the figure indicate the incubation times used for the subsequent experiments.

Exploring the thermo stability and activity of Lys-N.

Nonaka *et al.* found that Lys-N is rather thermo-stable and able to maintain enzymatic activity up to 80°C[34]. The experiments described above indicated that an enzyme/substrate ratio of 1/200 and an incubation time below 1 hour created experimental conditions with rate limiting conditions that are relatively easy to monitor. We therefore used this enzyme/substrate ratio to evaluate the performance of Lys-N at elevated temperatures. BSA was digested using an incubation times of 1 min, 10 min, 30 min or 1 hour at 7 different incubation temperatures (4°C, 25°C, 37°C, 50°C, 60°C, 70°C and 80°C). Each temperature generated a unique picture with regard to number of peptides detected (see Figure 3). With an incubation of only 1 minute hardly any peptides were observed. Temperatures below 37°C showed an increase in performance with increasing incubation time. From 70°C to 80°C a dip or plateau in the number of peptides is observed, which may also be related to peptide degradation. The trends at high temperatures become a little unclear, which is most likely due to our quenching procedure. An aliquot (5µl) of the digest is placed in 100 µl of quench solution, which is stored at 4°C. It is possible at the high temperatures that cooling is less efficient leading to spurious results. Nevertheless, a 10 minute digestion at 70°C produced an equivalent result to 1 hour at 25°C, indicating that Lys-N has potential to be used for digestion even at these elevated temperatures. Pham *et al.* found similar results, although only by using glycolated trypsin, where incubation times could be reduced by increasing the

temperature[19]. Lys-N can maintain activity over such a broad temperature without the need to modify the enzyme to improve thermo-stability[19].

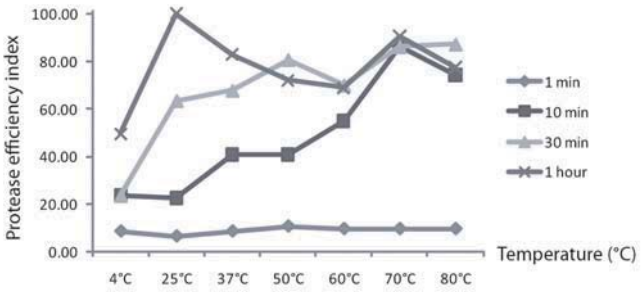

Figure 3: Lys-N activity remains optimal at elevated incubation temperatures. BSA was digested with Lys-N using 4 different incubation times and 7 different incubation temperatures at a protease to protein ratio of 1/200.

Effects of various solvents on Lys-N.

We further explored the effect of different solvents and denaturing agents on the enzymatic activity of Lys-N. Denaturing agents such as urea and guanidine hydro chloride (GHCL) are sometimes very useful in in-solution digestion protocols. Unfortunately, they have shown to have a rather negative effect on the enzymatic activity for a number of enzymes including trypsin[40]. We evaluated the potential of using some of the most used denaturing agents for Lys-N in-solution digestions. Here, we resuspended BSA in 0 to 8 M urea or 0 to 8 M GHCL and added Lys-N at a ratio of 1/200 prior to incubating the samples for 1 or 4 hours at 37°C (Figure 4A and B). Interestingly, the performance of Lys-N remains similarly high from 0 up to 8 M urea, as illustrated in Figure 4A. These results further expand on previous observations where it was found that Lys-N is able to retain enzymatic activity in the presence of 4 M urea[15, 34]. In contrast, GHCL had a more destructive effect on the enzymatic activity of Lys-N. The number of peptides observed severely decreases with increasing concentration of GHCL, as illustrated in Figure 4B. Next, we assessed the activity of Lys-N in the organic solvents acetonitrile and methanol. BSA was dissolved in 0 to 80% of acetonitrile or 0 to 80% methanol and incubated with Lys-N at a ratio of 1/200 for 1 and 4 hours at 37°C. The number of peptides identified remains relatively constant with the different levels of acetonitrile, indicating that Lys-N's activity is not severely affected by this solvent (Figure 4C). In contrast, elevated concentrations of methanol were shown to have a negative effect on Lys-N's activity, Figure 4D.

Additional studies were performed where Lys-N was subjected to different concentrations of ammonium acetate (0.1 M, 0.3 M and 0.5 M with pH values of 6.5, 7 and 7.5 respectively)

and acetic acid (0.1%, 0.5% and 1% with pH values of 4, 3.5 and 3 respectively) (Supplementary Figure 2). Ammonium bicarbonate (AMBIC) was used as a positive control. These results revealed that Lys-N retains a high enzymatic activity in 0.1M and 0.3 M of ammonium acetate though lower than observed for 50 mM AMBIC. When the pH is lowered further (acetic acid) the number of peptides identified decreased dramatically. Lys-N is essentially inactivated when the acetic acid amount is increased (0.5% and 1%), which justifies the use of 1% acetic acid as our quench solvent.

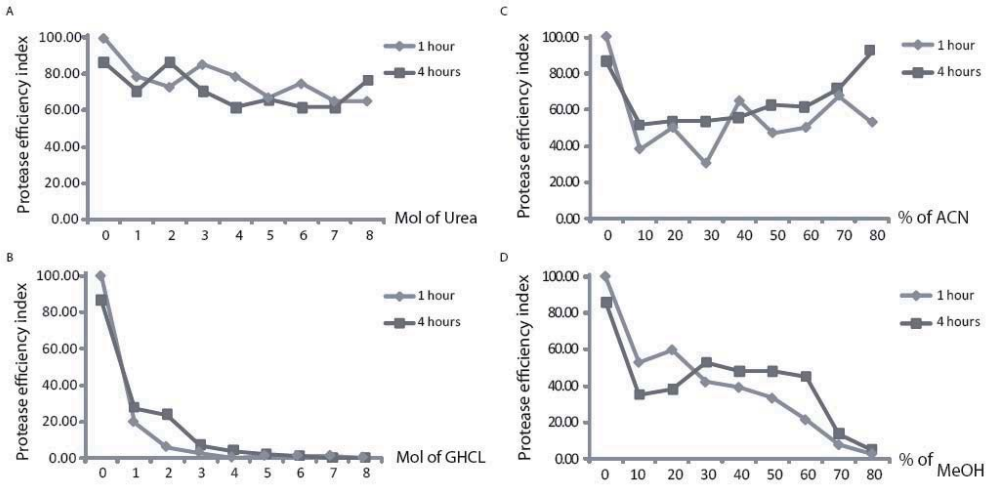

Figure 4: Proteolytic activity of Lys-N at elevated concentrations of A) urea B) guanidine hydro chlorate C) acetonitrile and D) methanol. All the BSA digestions were performed by incubation the samples for either 1 or 4 hours at 37°C using an enzyme to protein ratio of 1/200.

To summarize the results obtained Lys-N proved to be thermostable and maintained a high performance even at 80°C. It was found that at elevated temperature the incubation time for digestion could be drastically decreased. The performance of Lys-N remains relatively constant at urea concentrations of up to 8 M urea and acetonitrile levels of up to 80%.

Effect of methylated lysines on Lys-N activity

Trypsin, Lys-C and Lys-N have in common that they all cleave adjacent to lysine residues. In proteins lysine amino acids can be modified either by acetylation or methylation. Lysine methylation is an important modification, functionally best characterized in histones[41]. Lysine methylation can occur to different degrees, where 1, 2 or 3 methyl groups can be attached to

the lysine side-chain[42, 43]. Trypsin and Lys-C are very inefficient at cleaving proteins at lysine residues when these are methylated and completely ineffective when lysines are acetylated. Such modification will therefore be difficult to detect due to miss-cleavages. We, therefore, tested the ability of Lys-N to cleave proteins containing modified lysine residues. Lys-N, similar to Lys-C and trypsin was unable to cleave N-terminal of lysine residues that were acetylated (data not shown). More interestingly when Lys-N was tested for its ability to digest peptides containing methylated lysines (Figure 5) it became apparent that it cleaves efficiently adjacent to mono-methylated lysine, and to some extent di-methylated lysine. Therefore, the synthetic peptide H3 with the sequence EIAQDFKTDLR was used, wherein the internal lysine residue was either mono, - di or tri- methylated. Our data indicate that unlike trypsin **(**Supplementary Figure 3) and Lys-C, Lys-N is able to cleave N-terminally of both mono- and di-methylated lysine residues.

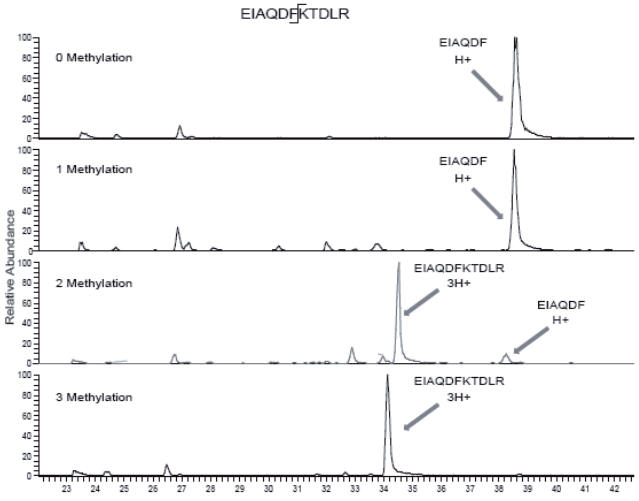

Figure 5: Lys-N can cleave adjacent to mono- and dimethylated lysine residues. The synthetic peptide H3, with the sequence EIAQDFKTDLR, was used to test the ability of Lys-N to digest methylated lysines. The internal lysine in the synthetic peptide was either mono, - di, - or tri-methylated. The methylated peptides where digested with Lys-N together with a non-methylated version of the peptide as a positive control (top-chromatogram).

In the top chromatogram is a large peak corresponding to the fragment peak EIAQDF at 38.54 min, which shows that Lys-N is able to digest the synthetic peptide containing no modifications. In the chromatogram below (mono-methylated lysines) the same large fragmentation peak is observed at 38.52 min, proving that Lys-N is able to cleave adjacent to mono-methylated lysines. In the third chromatogram (for the peptide with a di-methylated lysine) a smaller peak corresponding to the fragment peak EIAQDF is observed at 38.25 minutes. However, a large peak is also observed at 34.52 minutes from the precursor peptide, which illustrates that Lys-N is able to cleave adjacent to di-methylated lysines, although with decreased proteolytic activity. A large peak corresponding to the precursor of the tri-methylated lysine containing peptide is observed in the bottom chromatogram at 34.13 minutes.

Conclusion

Although, trypsin is still the enzyme of choice in proteomics, there is a growing need for expanding the tool-box with new proteases that for instance can perform better in membrane proteomics or can result in cleavage products more suitable for mass spectrometry analysis using ETD. In that respect, the protease Lys-N has unique characteristics making it a complementary enzyme of choice to use next to trypsin. Lys-N is unique in the sense that it has enzymatic specificity for the N-terminal side of lysine. Through in-depth evaluation using BSA we show, here, that Lys-N exhibits many useful characteristics. It is highly thermo-stable, and tolerant towards denaturing agents such as urea and acetonitrile. The tolerance to urea may be very helpful to improve digestion of so-called insoluble proteins including membrane proteins, whereby possible elevating the incubation temperature may help as well. Optimal digestion times were shown to be much shorter at these elevated temperatures, possibly preventing aggregation. Moreover, we also showed the unique capacity (compared to trypsin and Lys-C) of Lys-N to cleave adjacent to lysine residues modified by a single or doubly methylated lysine, making it useful in detecting these naturally occurring post-translational modification.
Overall, the enzyme Lys-N is shown here to be a very stable and efficient enzyme, making it a valuable tool in proteomics sample preparation.

Acknowledgements.

This work was supported by the Netherlands Proteomics Centre, embedded in the Netherlands Genomics Initiative.

Supplementary data.

Supplementary Figure 1: Effect of different protease to substrate ratios on the proteolytic activity of Lys-N.

Supplementary Figure 2: The effect of different solvents on the proteolytic activity of Lys-N.

Supplementary Figure 3: The synthetic peptide H3 with the sequence EIAQDFKTDLR was used to test the ability of trypsin to digest methylated lysines.

Supplementary Table 1: Non-specific peptides generated from a Lys-N digestion using different incubation times.

Supplementary Table 2: Number of spectra (queries) identified for each data point in each set of experiments. This material is available free of charge via the Internet at http://pubs.acs.org.

References

1. Yates, J. R., 3rd; Gilchrist, A.; Howell, K. E.; Bergeron, J. J., Proteomics of organelles and large cellular structures. *Nat Rev Mol Cell Biol* **2005,** 6, (9), 702-14.
2. Delahunty, C.; Yates, J. R., 3rd, Protein identification using 2D-LC-MS/MS. *Methods* **2005,** 35, (3), 248-55.
3. Zubarev, R. A., Electron-capture dissociation tandem mass spectrometry. *Curr Opin Biotechnol* **2004,** 15, (1), 12-6.
4. Coon, J. J.; Shabanowitz, J.; Hunt, D. F.; Syka, J. E., Electron transfer dissociation of peptide anions. *J Am Soc Mass Spectrom* **2005,** 16, (6), 880-2.
5. Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A., Mobile and localized protons: a framework for understanding peptide dissociation. *J Mass Spectrom* **2000,** 35, (12), 1399-406.
6. Leitner, A.; Foettinger, A.; Lindner, W., Improving fragmentation of poorly fragmenting peptides and phosphopeptides during collision-induced dissociation by malondialdehyde modification of arginine residues. *J Mass Spectrom* **2007,** 42, (7), 950-9.
7. Wu, S. L.; Huhmer, A. F.; Hao, Z.; Karger, B. L., On-line LC-MS approach combining collision-induced dissociation (CID), electron-transfer dissociation (ETD), and CID of an isolated charge-reduced species for the trace-level characterization of proteins with post-translational modifications. *J Proteome Res* **2007,** 6, (11), 4230-44.
8. Han, H.; Xia, Y.; Yang, M.; McLuckey, S. A., Rapidly alternating transmission mode electron-transfer dissociation and collisional activation for the characterization of polypeptide ions. *Anal Chem* **2008,** 80, (9), 3492-7.
9. Taouatas, N.; Drugan, M. M.; Heck, A. J.; Mohammed, S., Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase. *Nat Methods* **2008,** 5, (5), 405-7.
10. Molina, H.; Horn, D. M.; Tang, N.; Mathivanan, S.; Pandey, A., Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A* **2007,** 104, (7), 2199-204.
11. Chalkley, R. J.; Medzihradszky, K. F.; Lynn, A. J.; Baker, P. R.; Burlingame, A. L., Statistical analysis of Peptide electron transfer dissociation fragmentation mass spectrometry. *Anal Chem* **2010,** 82, (2), 579-84.
12. Kalli, A.; Hakansson, K., Comparison of the electron capture dissociation fragmentation behavior of doubly and triply protonated peptides from trypsin, Glu-C, and chymotrypsin digestion. *J Proteome Res* **2008,** 7, (7), 2834-44.
13. Mohammed, S.; Lorenzen, K.; Kerkhoven, R.; van Breukelen, B.; Vannini, A.; Cramer, P.; Heck, A. J., Multiplexed proteomics mapping of yeast RNA polymerase II and III allows near-complete sequence coverage and reveals several novel phosphorylation sites. *Anal Chem* **2008,** 80, (10), 3584-92.
14. Swaney, D. L.; Wenger, C. D.; Coon, J. J., Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics. *J Proteome Res* **2010**.
15. Gauci, S.; Helbig, A. O.; Slijper, M.; Krijgsveld, J.; Heck, A. J.; Mohammed, S., Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. *Anal Chem* **2009,** 81, (11), 4493-501.
16. Mischerikow, N.; Spedale, G.; Altelaar, A. F.; Timmers, H. T.; Pijnappel, W. W.; Heck, A. J., In-depth profiling of post-translational modifications on the related transcription factor complexes TFIID and SAGA. *J Proteome Res* **2009,** 8, (11), 5020-30.
17. Gatlin, C. L.; Eng, J. K.; Cross, S. T.; Detter, J. C.; Yates, J. R., 3rd, Automated identification of amino acid sequence variations in proteins by HPLC/microspray tandem mass spectrometry. *Anal Chem* **2000,** 72, (4), 757-63.
18. MacCoss, M. J.; McDonald, W. H.; Saraf, A.; Sadygov, R.; Clark, J. M.; Tasto, J. J.; Gould, K. L.; Wolters, D.; Washburn, M.; Weiss, A.; Clark, J. I.; Yates, J. R., 3rd, Shotgun identification of protein

modifications from protein complexes and lens tissue. *Proc Natl Acad Sci U S A* **2002,** 99, (12), 7900-5.
19. Pham, V. T.; Ewing, E.; Kaplan, H.; Choma, C.; Hefford, M. A., Glycation improves the thermostability of trypsin and chymotrypsin. *Biotechnol Bioeng* **2008,** 101, (3), 452-9.
20. Petritis, B. O.; Qian, W. J.; Camp, D. G., 2nd; Smith, R. D., A simple procedure for effective quenching of trypsin activity and prevention of 18O-labeling back-exchange. *J Proteome Res* **2009,** 8, (5), 2157-63.
21. Chen, E. I.; Cociorva, D.; Norris, J. L.; Yates, J. R., 3rd, Optimization of mass spectrometry-compatible surfactants for shotgun proteomics. *J Proteome Res* **2007,** 6, (7), 2529-38.
22. Fischer, F.; Poetsch, A., Protein cleavage strategies for an improved analysis of the membrane proteome. *Proteome Sci* **2006,** 4, 2.
23. Rietschel, B.; Arrey, T. N.; Meyer, B.; Bornemann, S.; Schuerken, M.; Karas, M.; Poetsch, A., Elastase digests: new ammunition for shotgun membrane proteomics. *Mol Cell Proteomics* **2009,** 8, (5), 1029-43.
24. Han, J.; Schey, K. L., Proteolysis and mass spectrometric analysis of an integral membrane: aquaporin 0. *J Proteome Res* **2004,** 3, (4), 807-12.
25. Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Yates, J. R., 3rd, A method for the comprehensive proteomic analysis of membrane proteins. *Nat Biotechnol* **2003,** 21, (5), 532-8.
26. Schlosser, A.; Bodem, J.; Bossemeyer, D.; Grummt, I.; Lehmann, W. D., Identification of protein phosphorylation sites by combination of elastase digestion, immobilized metal affinity chromatography, and quadrupole-time of flight tandem mass spectrometry. *Proteomics* **2002,** 2, (7), 911-8.
27. Nonaka, T.; Dohmae, N.; Hashimoto, Y.; Takio, K., Amino acid sequences of metalloendopeptidases specific for acyl-lysine bonds from Grifola frondosa and Pleurotus ostreatus fruiting bodies. *J Biol Chem* **1997,** 272, (48), 30032-9.
28. Hohmann, L.; Sherwood, C.; Eastham, A.; Peterson, A.; Eng, J. K.; Eddes, J. S.; Shteynberg, D.; Martin, D. B., Proteomic analyses using Grifola frondosa metalloendoprotease Lys-N. *J Proteome Res* **2009,** 8, (3), 1415-22.
29. Rao, K. C.; Carruth, R. T.; Miyagi, M., Proteolytic 18O labeling by peptidyl-Lys metalloendopeptidase for comparative proteomics. *J Proteome Res* **2005,** 4, (2), 507-14.
30. Boersema, P. J.; Taouatas, N.; Altelaar, A. F.; Gouw, J. W.; Ross, P. L.; Pappin, D. J.; Heck, A. J.; Mohammed, S., Straightforward and de novo peptide sequencing by MALDI-MS/MS using a Lys-N metalloendopeptidase. *Mol Cell Proteomics* **2009,** 8, (4), 650-60.
31. Taouatas, N.; Altelaar, A. F.; Drugan, M. M.; Helbig, A. O.; Mohammed, S.; Heck, A. J., Strong cation exchange-based fractionation of Lys-N-generated peptides facilitates the targeted analysis of post-translational modifications. *Mol Cell Proteomics* **2009,** 8, (1), 190-200.
32. Pinkse, M. W.; Uitto, P. M.; Hilhorst, M. J.; Ooms, B.; Heck, A. J., Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem* **2004,** 76, (14), 3935-43.
33. Carvalho, P. C.; Hewel, J.; Barbosa, V. C.; Yates, J. R., 3rd, Identifying differences in protein expression levels by spectral counting and feature selection. *Genet Mol Res* **2008,** 7, (2), 342-56.
34. Nonaka, T.; Ishikawa, H.; Tsumuraya, Y.; Hashimoto, Y.; Dohmae, N., Characterization of a thermostable lysine-specific metalloendopeptidase from the fruiting bodies of a basidiomycete, Grifola frondosa. *J Biochem* **1995,** 118, (5), 1014-20.
35. Nonaka, T.; Hashimoto, Y.; Takio, K., Kinetic characterization of lysine-specific metalloendopeptidases from Grifola frondosa and Pleurotus ostreatus fruiting bodies. *J Biochem* **1998,** 124, (1), 157-62.
36. Choi, H.; Fermin, D.; Nesvizhskii, A. I., Significance analysis of spectral count data in label-free shotgun proteomics. *Mol Cell Proteomics* **2008,** 7, (12), 2373-85.
37. Liu, H.; Sadygov, R. G.; Yates, J. R., 3rd, A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **2004,** 76, (14), 4193-201.

38. Klammer, A. A.; MacCoss, M. J., Effects of modified digestion schemes on the identification of proteins from complex mixtures. *J Proteome Res* **2006,** 5, (3), 695-700.
39. van Midwoud, P. M.; Rieux, L.; Bischoff, R.; Verpoorte, E.; Niederlander, H. A., Improvement of recovery and repeatability in liquid chromatography-mass spectrometry analysis of peptides. *J Proteome Res* **2007,** 6, (2), 781-91.
40. Delaage, M.; Lazdunski, M., Trypsinogen, trypsin, trypsin-substrate and trypsin-inhibitor complexes in urea solutions. *Eur J Biochem* **1968,** 4, (3), 378-84.
41. Watanabe, H.; Soejima, K.; Yasuda, H.; Kawada, I.; Nakachi, I.; Yoda, S.; Naoki, K.; Ishizaka, A., Deregulation of histone lysine methyltransferases contributes to oncogenic transformation of human bronchoepithelial cells. *Cancer Cell Int* **2008,** 8, 15.
42. Wang, Y.; Jia, S., Degrees make all the difference: the multifunctionality of histone H4 lysine 20 methylation. *Epigenetics* **2009,** 4, (5), 273-6.
43. Sarg, B.; Koutzamani, E.; Helliger, W.; Rundquist, I.; Lindner, H. H., Postsynthetic trimethylation of histone H4 at lysine 20 in mammalian tissues is associated with aging. *J Biol Chem* **2002,** 277, (42), 39195-201.

Chapter 6

Summary

Sammenvatting in het Nederlands

Curriculum vitae

List of publications

Acknowledgments

Summary

In chapter 1, is given a general overview of typical workflows used for protein identification and characterization in mass spectrometry based proteomics. Different mass spectrometers and parts of the instruments are described, as well as different fragmentation techniques. Next, general proteolytic enzymes used for mass spectrometry based proteomics are described, including a general description of the data analysis used for the generated mass spectra of the proteolytic peptides. The metalloendopeptidase Lys-N is likewise introduced, where the specificity and property of the enzyme is described. The protease is a general link between all chapters in the thesis as the unique specificity and tolerance is explored, in combination with different mass spectrometry based techniques, in chapter 2 to 5. Last, liquid chromatography mass spectrometry is described where general peptide pre-fractionation and purification techniques are mentioned, for instance, as an approach to reduce sample complexity and increase peptide identification.

In chapter 2, the metalloendopeptidase Lys-N is introduced as an alternative method for sequencing of peptides by mass spectrometry. The fragmentation technique electron transfer dissociation (ETD) is used for sequencing of doubly charged Lys-N generated proteolytic peptides and compared to peptide sequencing by collision induced dissociation (CID). The Lys-N generated spectra are further validated by comparing them to spectra of Lys-C produced peptides, likewise analyzed by CID and ETD. We show that the combination of Lys-N and ETD based sequencing generates unique easy interpreted spectra dominated by *c*-type fragment ions. Lastly, we demonstrate that it is possible to enrich for these doubly charged Lys-N generated peptides by using low-pH strong cation exchange (SCX).

In chapter 3, we further evaluate the combination of Lys-N generated peptides and SCX fractionation mentioned in chapter 2. We show that it is possible to separate and fractionate Lys-N generated peptides, by SCX, into different functional categories as, acetylated N-terminal peptides, singly phosphorylated peptides, peptides with a single basic residue and peptides with multiple basic residues. In general, we show that the analysis of the SCX separated peptide groups, by LC-MS/MS using both CID and ETD induced fragmentation, is a great method for global proteome and phosphoproteome analysis.

In chapter 4, we looked into the potential of using Lys-N for MALDI-MS/MS applications. Both Lys-N digested BSA and a cells lysate were used to show that we could generate simple, straightforward spectra containing complete *b*-ion series. Additional, we showed that

the simple spectra generated by the combination of Lys-N proteolytic doubly charges peptides and MALDI CID can be used for de novo sequencing

In chapter 5, we wanted to evaluate the proteolytic performance of Lys-N under various experimental conditions. We used the standard protein BSA as a model system for the evaluation of Lys-N in sample preparation for proteomic analyses. We show that Lys-N maintains the activity in harsh conditions such as 8M urea and 80% acetonitrile, and even when subjected to temperatures as low as 4 ^{o}C and as high as 80 ^{o}C. We further demonstrated that by increasing the digestion temperature a decrease in incubation time can be achieved. The proteolytic performance of Lys-N was found to be hampered when subjected to guanidine hydrochloride and methanol. Last, we show that Lys-N is able to cleave adjacent to mono- and di-methylated lysines, in contrast to other more general proteases as trypsin and Lys-C. This may be valuable when analyzing naturally occurring post-translational modifications.

Sammenvatting in het Nederlands

Hoofdstuk 1 geeft een algemeen overzicht van de verschillende strategieën voor eiwit identificatie en karakterisering met behulp van op massaspectrometrie gebaseerde proteomics. Verschillende massaspectrometers en onderdelen daarrvan worden uitgelegd, evenals de verschillende fragmentatiemethoden. Daarnaast worden de verschillende proteolytische enzymen beschreven die voor proteomics experimenten worden gebruikt en wordt een beschrijving gegeven van de data analyse voor de interpretatie van de massaspectra van de proteolytische peptiden.

Het metalloendopeptidase Lys-N word geïntroduceerd, met een nadruk op de specificiteit en andere eigenschappen van het enzym. Dit protease vormt een gemeenschappelijk factor tussen alle hoofdstukken in dit proefschrift, waarin de unieke specificiteit en tolerantie van dit eiwit worden bestudeerd in hoofdstukken 2 tot en met 5. Tot slot wordt vloeistofchromatografie behandeld, waarbij vooral fractionerings- en zuiveringsmethoden voor peptiden aan bod komen, bijvoorbeeld om de complexiteit van monsters te verminderen en het aantal geïdentificeerde peptiden te verhogen.

In hoofdstuk 2 wordt het metalloendopeptidase Lys-N wordt geïntroduceerd als een alternatieve methode voor het sequencen van peptiden met behulp van massaspectrometrie. De fragmentatietechniek *electron transfer dissociation* (ETD) wordt gebruikt voor het sequencen van dubbel geladen peptiden afkomstig van Lys-N klieving en wordt vergeleken met sequencing met behulp van *collision induced dissociation* (CID). De met Lys-N gegenereerde spectra worden verder gevalideerd door ze te vergelijken met spectra van peptiden die met Lys-C gemaakt zijn, die ook geanalyseerd worden met CID en ETD.

We laten zien dat uit de combinatie van Lys-N met op ETD gebaseerde sequencing eenvoudig te interpreteren spectra ontstaan die gedomineerd worden door *c*-type fragment ionen. Ten slotte bewijzen we dat het mogelijk is om deze dubbel geladen Lys-N gegenereerde peptiden te verrijken met behulp van strong cation exchange (SCX) bij lage pH.

In hoofdstuk 3 evalueren we de combinatie van met Lys-N gemaakte peptiden en SCX fractionering uit hoofdstuk 2 verder. We laten zien dat het mogelijk is om de Lys-N proteolytische peptiden door middel van SCX in verschillende functionele categorieën te scheiden, zoals N-terminale peptiden, peptiden met een enkele fosfaatgroep, peptiden met een enkel basisch aminozuur en peptiden met meerdere basische aminozuren. Ook laten we zien dat de analyse van met SCX gescheiden peptiden door middel van LC-MS/MS met zowel

CID als ETD fragmentatie een uitstekende methode is voor het globaal analyseren van een proteoom en fosfoproteoom.

In hoofdstuk 4 hebben we gekeken naar de mogelijkheden van Lys-N in MALDI-MS/MS toepassingen. Zowel met Lys-N gedigesteerd BSA als met cellysaat laten we zien dat de resulterende peptiden leidden tot simpele, eenvoudig te analyseren spectra die complete series *b*-ionen bevatten. Daarnaast bleek dat de simpele spectra die ontstaan uit de combinatie van Lys-N gegenereerde dubbel geladen peptiden en MALDI CID gebruikt kunnen worden voor *de novo sequencing.*

In hoofdstuk 5 wilden we de proteolytische activiteit van Lys-N onder verschillende omstandigheden bestuderen. We gebruikten het standaard eiwit BSA als een model systeem voor de evaluatie van Lys-N in de bereiding van proteomics monsters. We laten zien dat Lys-N zijn activiteit behoud onder extreme omstandigheden, zoals 8M ureum of 80% acetonitril en zelfs terwijl het wordt blootgesteld aan extreme temperaturen als 4 ^{o}C of 80 ^{o}C. We demonsteren dat door het verhogen van de digestie temperatuur een aanzienlijke verkorting in de incubatietijd kan worden bereikt. De proteolytische activiteit van Lys-N bleek wel verminderd te worden door de aanwezigheid van guanidine hydrochloride of methanol. Tenslotte laten we zien dat Lys-N in staat is om te knippen naast mono- en di-gemethyleerde lysine residuen, in tegenstelling tot veel andere, vaak gebruikte, proteases zoals trypsine en Lys-C, wat van belang is voor de analyse van natuurlijk voorkomende post-translationele modificaties.

Chapter 6

Curriculum vitae

Name:
Nadia Taouatas

Place and date of birth:
18.05.1978, Vejle, Denmark

Education:
2005: M.Sc. in Genetechnology. Department of Biochemistry and Molecular Biology (BMB), University of Southern Denmark, 5230 Odense M, Denmark. Project: Studies of phosphopeptides by immobilized metal affinity chromatography, surface plasmon resonance and mass spectrometry (Supervisor: Ole Nørregaard Jensen, Protein research (PR) Group).

2006 - 2010: Ph.D. at the Utrecht Institute for Pharmaceutical Sciences and Bijvoet Center for Biomolecular Research, Utrecht University, Netherlands. Project: Enhancing protein- and peptide separation and selective enrichment by capillary based nanoflow chromatography (Supervisors: Albert J. R. Heck and Shabaz Mohammed, Biomolecular Mass Spectrometry and Proteomics Group).

Professional Experience:
2006 (January to July): Research assistant in the Protein research (PR) Group, Department of Biochemistry and Molecular Biology (BMB), the University of Southern Denmark, 5230 Odense M, Denmark.

List of publications

1. Mahmoudi T, Boj S, Hatzis P, Li V, Taouatas N, Vries R, Teunissen H, Begthel H, Korving J, Mohammed S, Heck A, Vlevers H, The Leukemia-associated Mllt10/Af10- Dot1l are β-catenin coactivators essential for intestinal homeostasis, PLoS Biology, Accepted.

2. Taouatas N, Mohammed S, Heck AJ., Exploring new proteome space: combining Lys-N proteolytic digestion and strong cation exchange (SCX) separation in peptide centric MS-driven proteomics, Methods in Molecular Biology. Accepted.

3. Aye TT*, Scholten A*, Taouatas N, Vos MA, van Veen T, Heck AJ. Proteome-wide protein concentrationsin the human heart, Molecular BioSystems. 6(10), p1917-27, **2010**.

4. Taouatas N, Heck AJ, Mohammed S., Evaluation of metalloendopeptidase Lys-N protease performance under different sample handling conditions., Journal of Proteome Research. 9(8), p4282-8, **2010.**

5. van Breukelen B, Georgiou A, Drugan MM, Taouatas N, Mohammed S, Heck AJ., LysNDeNovo: an algorithm enabling de novo sequencing of Lys-N generated peptides fragmented by electron transfer dissociation., Proteomics. 10(6), p1196-201, **2010**.

6. Mahmoudi T*, Li VS*, Ng SS, Taouatas N, Vries RG, Mohammed S, Heck AJ, Clevers H., The kinase TNIK is an essential activator of Wnt target genes., EMBO J**.** 28(21), p3329-40, **2009**.

7. Boersema PJ*, Taouatas N*, Altelaar AF, Gouw JW, Ross PL, Pappin DJ, Heck AJ, Mohammed S. Straightforward and de novo peptide sequencing by MALDI-MS/MS using a Lys-N metalloendopeptidase., Mol Cell Proteomics. 8(4), p650-60, **2009**.

8. Taouatas N*, Altelaar AF*, Drugan MM, Helbig AO, Mohammed S, Heck AJ.Strong cation exchange-based fractionation of Lys-N-generated peptides facilitates the targeted analysis of post-translational modifications., Mol Cell Proteomics. 8(1), p190-200, **2009**.

9. Taouatas N, Drugan MM, Heck AJ, Mohammed S., Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase., Nat Methods. 5(5), p405-7, **2008**.

Acknowledgments

I would like to first thank my supervisor, Albert, for giving me the opportunity to continue my interest in mass spectrometry by giving me the change to come to Utrecht as a PhD student. Working in your group has taught me a lot, thank you for this opportunity.

Shabaz, I especially would like to thank you for being my co-promoter during my PhD. I have known you for some time now, first in Odense and then Utrecht and I have really learned a lot from you during these years. It has been tough at times but I know you mean well…most of the time ☺ Thanks for all your help, ideas and fast corrections. Thanks also for taking an interest in my future plans and for support. I wish you good luck with whatever your future plans are going to be.

Joost, thanks for all your help concerning MS, LC, SCX etc. You always take time to help and explain things to people and will be great to have in any lab. Thanks for the fun BBQ´s you organized at work. I have never made so many homemade drinks before I met you! Javier, thanks for having time to help and to chat once in a while, even though you are always very busy. It was also fun going on the skiing vacation with you, too bad I could not join last time. Kristina and Andreas, thanks for the dinners and social events at your home, it was very nice. Thanks Andreas for all the supportive conversations at the end of my PhD, I appreciated that. I hope you will both be very happy in Canada and wish you good luck with your new life there. Glen and Sarah two lovely people! You are both super energetic and fun people and I enjoyed a lot all the trips with you guys. I was a bit scared after the skiing holiday I must admit, especially of you Glen. I thought you guys were machines! No one can for sure match you Glen, when it comes to partying. I think you are both very valuable to have in a group as you organize a lot of social events and make sure to invite everyone. I wish you guys all the best and hope to see you again, either in Canada or Denmark. Good luck with brewing your beer in Canada Glen ☺ Bobby thanks for all your help and especially for helping making the nice cover for the MCP Lys-N article. Thanks to the bioinformatics people for all your help; Bas van Breukelen (thanks for all your help with computers etc and for the Lys-N work we did together), Madalina (Thanks for a successful collaboration), Henk (watch out for the scary telephone man!), Robert (It was a fun trip to the spoorweg museum with you and Silvia), Wim, Salvatore (good luck with your new project, being a father ☺), Pieter, Danny, Geert. Martje thanks for always being so helpful and for organizing social events as well. Werner thanks for inviting people when your band was playing in Utrecht. It was a fun evening, when we went to hear you and your band play. Good

luck with your music. Corine, thanks for always being so nice and helpful. I remember you coming very late in the evening to bring me my keys, when I just arrived in the country. Thanks for organizing social events in the group, they were always fun (well the running was more fun afterwards). Thanks also for your help with printing and sending my thesis. Hope you will be very happy in your new house and with your new little rabbit ☺ I would like to thank you Arjan for making my start in the group easier. You have always been friendly and helpful and I enjoyed very much working with you in the beginning (even though the project did not work out). I will miss joking with you and all the fun we had. Thanks for helping me move my little zoo to Beukstraat that was really nice of you ☺ Mirjam, thanks for all your help at work! It was fun skating with you and Corine and playing volleyball with you. Too bad it was cancelled this year! Jeffrey, I enjoyed a lot the time when Silvia, you and I worked late together and had dinners together. Also, thanks for taking us along with Harry and Joanne for movies, walks etc. Thanks for inviting us for dinner as well, even though your food is a bit too spicy for me ☺. Too bad we never managed to climb together. I wish you a lot of strength for the future and will miss you. If you and Joanne ever come to Copenhagen let me know. Remember Stockholm is not in Denmark! Onno, you are a very nice and supporting person. Thanks for inviting me to your and Marieke´s wedding. Thanks, to both of you, for inviting the girls to see cute little Max. That was a nice evening. Good luck with everything! Lars, thanks for the talks when you once in a while came to run samples and for joining our volleyball team, even though it did not end well for you. Good luck finishing your PhD. Maarten, one of the most relaxed people I have ever met! It was very nice and fun working with you on the ETD and for working with you on the MCP paper (what a nice cover! ☺). It was also fun (and crazy) with the boxing we tried at Olympus and I will never forget the dragon tail move we learned there. Thanks for taking time listen to talks, frustrations, for questions etc. Good luck with everything also your new project being a father ☺. Reinout, thank you for helping me translating the summary, and for listen to talks, and always being there for questions, advice etc. Christian senior, I will miss your lovely smile and happy personality. Seriously, thanks for a lot of fun nights out…. and again for being so happy all the time. Good luck finding a new job when the time comes and with sorting out all the future plans. By the way, are you still mixing Coke and Fanta? Arjen, it was short but nice collaborating with you. Good luck with finishing the project and with future projects. Ester you are a very caring and nice person. Thanks for inviting me to your wedding and for always being supportive. Papa Vincent (Halim), you are one of the most hardworking people I have ever met (and one of the nicest!). But luckily you are never too busy to talk to

people. I wish you, Ivy and cute Jonathan all the best. Marco and Hanka, thanks for often inviting people to your (very green) home. You are very social people and it was always nice joining for social events with both of you. It was also very nice chatting with you Hanka, when you were waiting for Marco in the late evenings at work. Marco I hope the office will not be too quiet now that TT and I have left the office next door ☺ Good luck to you both with finishing your PhD. Teck, thanks for all the supportive talks in your office and for always having time for people. It was nice talking to you about science, job, travelling etc. I also enjoyed the sushi evenings, especially when a few of us went to Sumo in Amsterdam. Good luck with your future job, where it might be. Benjamin, you are a positive and fun person. Thanks for all the funny chats in the lab and for the fun evening when I went out dancing with you, Serena and Paul. Good luck with the rest of your PhD. Ayse, thanks for the nice trips we went on together and for the nice dinners. I wish you good luck with the rest of your PhD and hope to soon see pictures from your marriage ☺ Paul/Pauli/Paultje, what can I say other than I will miss you a lot! It was nice sharing an office with you and argue with you from time to time ☺ and for working with you on the MALDI paper. Friday beer sessions were always fun with you around. Good luck with your new job and life in Germany. Serena you really bring color to wherever you go. You are crazy (in a good way) and fun and I hope you will keep that way. I hope you will come to Copenhagen again so we can meet there ☺ Good luck finishing your PhD. Thomas, thanks for a lot of nice talks and for always being so social and helpful. I think you are a nice input for the group. Good luck with everything. Chhiti, I hope we did not scare you too much during our lunch conversations. Good luck with everything in the US. I hope you will be happy there with your husband. Nikolai, Thanks for all the conversations in Albert Heijn going home from work and for the more relaxing conversations at the "Buurten" and other cafés. Good luck with finishing your PhD…you are almost there! Also, good luck with finding a job and a new home. Houjang, my D90 buddy ☺ you are always so nice and sweet! I wish you all the best for the rest of your stay in Utrecht. Too bad I could not join for more sushi events ☺ Ioana, you are a straightforward, social and warm person. I am sure you fit well in your new job as a mum. Charlotte, Thanks for always inviting me to your birthday parties. It was also fun cycling with you to Lisse, even though it ended up with me getting into a little accident. You have always been very helpful and I wish you all the best in Sweden and hope you will be very happy there. The cold weather should at least not scare you ☺ Hongtao, thanks for always being so friendly and happy. It is always nice talking to you. It was also nice trying your delicious Chinese sweets ☺ Jessica, you are a super active and nice

person. Thanks for nice talks in the office. I wish you good luck with your PhD. Marco B, I only got to know you shortly but you were always very nice (sorry for stealing your chair so often)….I am sorry you had to witness all the girly talks in your office ☺ Good luck with your PhD as well. Jeroen Jansen, thanks for being so helpful in the lab and for always being in a good mood. It was also fun skiing with you last year. Good luck with your new house. Martina you are such a funny and warm person. I have never met anyone with so many funny stories; for sure your life is never boring ☺ It was always nice visiting your office for a chat. Good luck with your future plans for moving to Ireland. Tieneke, you are a strong and determine person. I enjoyed our conversations late in the evenings before your dance sessions. I will always remember our walk on queens' day (with queen, Silvia), that was a very nice day. Sharon, so caring and sweet…and a tiny bit stressed ☺ I wish you had lived a bit closer so we could have gone shopping together a bit more often. I will miss the movie nights at your place with the girls. It was also very nice spending New Years Eve with you and Cyril (even though I got a massive migraine attack), thanks for inviting Jan and I! I also wish we could have more dinners together the four of us, which was really nice. But then we just have to visit each other once in a while ☺ Thanks also for choosing me as your paranimf, it was a nice occasion to come back again to Holland and to see you again. Good luck with your new job. I wish all the best to Cyril, Alec and you. Adja, thanks for always being so helpful, supportive and a good friend! You were always there when things got hard at work. It was a nice and fun evening with you and Sharon in Breda and I will really miss being able to do that. I will always laugh when I hear someone talk about strong legs or see a painting hanging un-even ☺ . Thanks for your help printing the thesis and for many other things. Thanks a lot to both you and Sharon for being my paranimfs that means a lot to me. Poupak, the greatest artist in the group ☺ I very much enjoyed our talks in your office in the Went building. You are a very strong person and always kind and friendly. I enjoyed the dinners at you home very much. I wish you all the best. Simone you are a straightforward and very friendly person. I remember you from the time when I came with the Danish group in 2006. Joost and you gave a very good impression, as you were both super social and friendly. I hope you are happy in Munich I think you are, at least when you are not working with the ETD instrument. Christian Jr. (sunshine), it was always fun joking with you at work and nice talking to you. Good luck with your PhD. Eric, I think you are a very relaxed and fun person. It was nice flying with you and your girlfriend to Copenhagen. For sure my best journey home ☺ I wish you good luck with your PhD. Soenita you are one of the most caring person I have ever met. You are always very helpful and nice to everyone.

Thanks for inviting us for you birthday party. That was a great and fun party! Hope you will find someone else to hold your GPS in the car, now that I am not there ☺ Thanks for everything and all the best to you and your family. Harm, I guess you are happy when I leave, as you said, I was the only one that could see through you ☺ Thanks for all your help with the SCX and for the fun conversations during lunch. Thanks for inviting me to your wedding, Adja and I had a lot of fun. Becky and Tom thanks for all the parties and social events at your home. It was always fun and nice to go on trips together with you both, especially the skiing trip. I enjoyed very much our little talks in your office Becky, and to watch rugby with you (I will miss that a lot!). I am grateful that you took time reading and correcting my thesis that was very nice of you. It was also nice spending time with you during ASMS and in New York. Thanks also Becky for all the fun late nights at Mick O´Connells, Vogels etc. Good luck to both of you with everything. Hope you will come and visit us in Denmark. Shahram thanks for all your computer help and for coming in to chat with us in the office. It was always very funny conversations we had. Donna I will miss the smell of nice Thai food in the office and to try it ☺ . It was nice sharing the office with you and to follow all your vacation planning. I think you have been everywhere! Good luck finishing your PhD and organizing future trips. Pepijn it was a big change suddenly having you in the office, being used to only girls in there. However, it was nice and fun sharing the office with you and you were always very helpful. Good luck with you PhD and future plans. Dave, what an energetic and smiling person you are. We only shared an office together for a short time but it was nice having you there and to try your exotic snacks. Good luck with your new house in Singapore and your work there. If you ever come to Copenhagen again let me know. Erwin it was very nice sharing an office with you and all the talks we had about HTC gadgets, projectors, movies etc. I sometimes miss that. Good luck with your PhD. Thin Thin (TT) my special friend! I was lucky to share and office with you since the beginning of my PhD and I missed you a lot when you left (I think Marco and Teck as well, as they complained that it was suddenly too quiet next door). You were always there for me when needed and I missed that a lot in the last hard month. Thanks for choosing me as your paranimf it was so great seeing you again and spent time with you again. We always had a lot of fun on trips, in the office and when we had dinner with Silvia at my place. I will for sure miss all these things. Thanks also for bringing us along for the Burmese event in Utrecht, which was fun and nice trying out the different food they served. I have also learned a lot about Burma and Singapore, which has been interesting. Silvia you are for sure a queen. I am grateful for all the fun times and memories we share, (including rescuing a

bird and sheep, that was such a nice walk that day) I will never forget them. I also enjoyed visiting places in Holland with you and really missed our dinners and movie nights when you left. It was nice to visit you in Scotland as well and I will never forget our scary graveyard tour in Edinburgh. I will be prepared next time, if we should ever encounter a real ghost together, as I know now that you will push me in front of you :) . Thanks also for choosing me as a paranimf, it was actually great fun making your video ☺ Renske, we can definitely agree on one thing, Bond & Smolders is a great bakery! Thanks for all the chats and for always being so helpful. Kees thanks for always being helpful and take your time to explain things. I think people will miss you when you leave the group. Good luck with finding a new house. Eleonora, it was nice to have you in the group and they are lucky to have you back. Too bad you did not come earlier. But it was great meeting you and your family in Rome. It was a wonderful vacation with lots of delicious home cooked food! I will definitely come again! Good luck with your new PhD project and your life in Utrecht.

Thanks to:

Jeff and Rachel, thanks for introducing baseball to the group, that was a fun afternoon, and for inviting the group for your nice parties. Sara, too bad you joined the group again so late. Good luck with your PhD. Wilma (I enjoyed the talks in your and Silvia`s office a lot in the Went building), Siddarth, Bas N., Natasha, Vincent G, Hortense (It was nice with all the dinners in the Thai restaurants with you and Sylvain), Monique, Silvie, Nasrin (always happy and friendly), Manuel, Sara (Hope you will enjoy it in Utrecht. Good luck with your PhD), Luigi, Rui, nune, Alessandro (it was short but nice sharing an office with you), Joost, Marije, Jeroen Kool, Jeroen Krijgsveld, Natasja, Anca, Bas Tops, Gideon, and Vera.

I would like to thank Sylvia, Tokameh and Pantelis for great collaborations. I enjoyed it a lot and wish you all good luck in the future.

Thanks to my former supervisor Ole N. Jensen and current supervisor Jesper V. Olsen for giving me the chance to learn about mass spectrometry and to continue working in the field of proteomics.

Mina, thanks for nice and fun conversations at social events. Good luck with your work in Nijmegen and with everything. Yvette, thanks a lot for being a great neighbor. I enjoyed our talks in your lowly garden and taking care of Thomas the cat. Thanks a lot for taking so good care of all my animals when I went to Denmark. I wish you all the best. I would also like to thank Pim and Madelon for making my stay at Beukstraat very nice. Iris and Yves thanks for making my stay in Holland nice by inviting me to dinners, your beautiful wedding etc. I hope we can stay in touch and you are off course always welcome to come and visit us

in Denmark. I would like to thank Dees and the whole Utrecht women rugby team. I enjoyed the time with all of you a lot and I am happy to have been introduced to this great sport, it is the only fun sport to watch on TV ☺ Thank you Dees for always being understanding and supportive. Good luck to you all and remember to keep the tackles low ☺ Joyce, thanks for being a great housemate. I missed you a lot in the house when you left. Thanks for always taking good care of my little zoo, when I went back home, and it was wonderful coming back to homemade cakes. But it is nice that you still live relatively close. Thanks for choosing me as you paranimf as well. Wish you all the best in Sweden.

Tak til alle mine dejlige venner og veninder, specielt Lene, Line, Inge Lise, og Tina. Det var dejligt at vide, når det blev hårdt, at I var der for mig. Jeg er glad for at den lange afstand ikke har forhindret vores venskab og for den støtte I har givet mig. Line og Inge Lise, det var super dejlig at have besøg af jer. Jeg glæder mig til mange flere sjove, hyggelige og festlige stunder sammen med jer!

Shanna tak fordi du altid var en god og forstående veninde. Tak for alle minderne, jeg vil altid savne dig min ven. En ven som dig bliver aldrig glemt!

Tak til både Jan´s familie, Henning´s familie og min egen familie for altid at gøre det hyggeligt, når vi var sammen. Det var dejligt at komme hjem til. Specielt tak til Anders, Michael, Helle, Andi og Mikkel for hyggelige skiferier og middage.

Far, jeg ville ønske, du kunne have været her idag til at se mig afslutte mit studie. Tak for alt du har lært og givet mig. Farvel kære far, til vi ses igen.

Mormor, tak for altid at sende fødselsdagskort, julekort osv. til Holland. Det er så dejlig at modtage kort/breve, når man er langt hjemmefra, og vide at nogle har tænkt på en. Det var også altid hyggeligt at få julepynt og påskepynt sendt med og at komme hjem om sommeren og besøge dig.

Tak til min dejlige storebror Thomas og mine vidunderlige niecer, Freja og Rhea. Det var altid dejlig at komme hjem og besøge jer og have noget at glæde sig til.

Tak mor og Henning for al jeres støtte under mit ophold i Holland. Det var altid dejlig at vide, at I var der for mig. Det var også rigtig dejlig at have besøg af jer.

Basse (Jan), du siger, at du ikke behøver at blive takket i en bog, men du ved også, hvor meget du betyder for mig. Tak for alting!