# Expansion of the toolbox to decipher the (phospho)proteome

Marco L. Hennrich

dedicated to my wife
Hanka

ISBN: 978-90-393-5758-3

The research presented in this thesis was performed in the Biomolecular Mass Spectrometry and Proteomics group at the Bijvoet Center for Biomolecular Research at Utrecht University.

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

Printed by: Proefschriftmaken.nl || Printyourthesis.com

# Expansion of the toolbox to decipher the (phospho)proteome

Nieuwe technieken voor het ontcijferen van het (fosfo)proteoom

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op woensdag 11 april 2012 des middags te 12.45 uur

door

Marco Leo Hennrich

geboren op 16 oktober 1976 te Buchen, Duitsland

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

Co-promotor: Dr. S. Mohammed

**TABLE OF CONTENTS**

CHAPTER 1:
**GENERAL INTRODUCTION..... 7**

CHAPTER 2
**RECENT ADVANCES**
**IN PEPTIDE SEPARATION BY MULTIDIMENSIONAL**
**LIQUID CHROMATOGRAPHY FOR PROTEOME ANALYSIS ..... 35**

CHAPTER 3
**DIMETHYL ISOTOPE LABELING**
**ASSISTED DE NOVO PEPTIDE SEQUENCING..... 71**

CHAPTER 4
**EFFECT OF CHEMICAL MODIFICATIONS**
**ON PEPTIDE FRAGMENTATION BEHAVIOR**
**UPON ELECTRON TRANSFER INDUCED DISSOCIATION ..... 85**

CHAPTER 5
**IMPROVING DEPTH IN PHOSPHOPROTEOMICS BY**
**USING A STRONG CATION EXCHANGE-WEAK ANION EXCHANGE-**
**REVERSED PHASE MULTIDIMENSIONAL SEPARATION APPROACH ..... 103**

CHAPTER 6
**ULTRA ACIDIC STRONG CATION EXCHANGE ENABLING**
**THE EFFICIENT ENRICHMENT OF BASIC PHOSPHOPEPTIDES..... 117**

CHAPTER 7
**SUMMARY**
**SAMENVATTING**
**ZUSAMMENFASSUNG**
**PUBLICATIONS**
**CURRICULUM VITAE**
**ACKNOWLEDGEMENT ..... 131**

# Chapter 1

## General Introduction

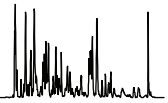


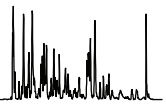
**TABLE OF CONTENTS**

GENERAL INTRODUCTION ..... 9
IONIZATION TECHNIQUES ..... 10
ESI ..... 10
MALDI ..... 12
MASS ANALYSERS ..... 12
Quadrupoles ..... 13
Quadrupole and Linear Ion Trap ..... 13
Time-of-Flight (TOF) ..... 14
Orbitrap ..... 15
FRAGMENTATION TECHNIQUES ..... 16
Collision-Induced Dissociation ..... 16
Electron Capture and Electron Transfer Dissociation ..... 17
OPTIMIZING PEPTIDES FOR FRAGMENTATION ..... 19
Choice of Protease ..... 19
Chemical Modification and MALDI-CID ..... 20
Isotopic Identifiers ..... 23
PEPTIDE AND PROTEIN IDENTIFICATION ..... 23
SEPARATION TECHNIQUES ..... 25
Reversed Phase Liquid Chromatography ..... 25
Ultrahigh-Pressure Liquid Chromatography ..... 26
Multidimensional Separation ..... 26
PHOSPHOPROTEOMICS ..... 27
OUTLINE ..... 28
REFERENCES ..... 30

## GENERAL INTRODUCTION

In almost all cellular processes proteins play pivotal roles. The field of protein science, proteomics, deals with the determination of the primary sequence, structure, chemical modifications, quantification, regulation, and the interaction between proteins or with other intra- or extracellular molecular components. The large number of proteins of a proteome, meaning all proteins of a cell, tissue or organism at a given time, and the dynamics of their abundance makes this a challenging task. In addition, the primary sequence of a protein can be chemically modified which can have ramifications on all aspects of a protein. Examples of post translational modifications (PTMs) are phosphorylation of serine, threonine or tyrosine, acetylation of the protein N-terminus or disulfide bond formation.[1] Protein phosphorylation in particular has gained a lot of interest, due to its high occurrence and its association with many cellular functions like signal transduction in apoptosis or in the regulation of cell division. It has been postulated that perhaps 30% of all human proteins are phosphorylated.[2] Additionally, a number of diseases are caused by aberrant phosphorylation, which include diabetes, Alzheimer's disease,[3] and a number of cancers.[4, 5] In mammals, phosphorylation predominantly occurs at serine, threonine and tyrosine residues, which constitute approximately 86%, 12% and 2%, respectively.[6] The introduction of the highly negative charge of the phosphoryl group might lead to, enable or hinder binding to a ligand, cofactor or substrate or might induce conformational changes that alter the function of the protein. This reversible modification is introduced by a kinase with adenosine triphosphate (ATP) as phosphoryl donor. More than 500 protein kinases have been identified in the human genome[7] of which approximately 100 are tyrosine kinases and the rest serine/threonine kinases.[8] The dephosphorylating enzymes are referred to as protein phosphatases. In contrast to over 400 serine/threonine kinases only approximately 30 catalytic subunits of serine/threonine specific phosphatases are known, whereas the number of tyrosine phosphatases identified in the genome is approximately equivalent to the number of tyrosine kinases.[8] In order to study processes associated with phosphorylation, the analysis of the proteome including the phosphorylation sites is essential.

Currently, an armada of techniques are applied to address the challenges of proteomics. However, liquid chromatography coupled to mass spectrometry (LC-MS) is the method of choice for the final step and is necessary when complex protein mixtures such as total proteomes are analyzed.[9-11] The technology has improved by leaps and bounds over the last decade, now offering the opportunity to analyze thousands of peptides from a cell lysate in a few hours.[12] However, mammalian cell lysate digests consist of millions of peptides. The complexity of such samples needs to be reduced prior to the mass spectrometric analysis. Complexity reduction can be accomplished by many chromatographic technologies,[13] which are further described in detail in chapter 2. In addition to complexity reduction through chromatography, it is often necessary to apply some kind of enrichment if one is, for instance, interested in a post translational modification. For example, phosphorylation



occurs at sub-stoichiometric levels and, being negatively charged, has poor response factors when being analyzed by MS. This has led to the development of a wide variety of enrichment techniques, including a number of chelation based techniques.[14-17] A new strategy to enrich for phosphopeptides with multiple basic residues based on SCX is presented in chapter 6. In chapter 5, a multidimensional separation strategy has been applied to further separate phosphopeptides with a single basic residue in order to increase the identifiable phosphopeptides.[18]

In order to be analyzed in the mass spectrometer the analytes need to be converted into gas phase ions. This can be accomplished by a variety of ionization processes and the two most commonly applied to larger biomolecules are described in the first part of the introduction. The common types of mass analyzers used in proteomics are subsequently described. Mass determination of a peptide is insufficient to identify the sequence and so fragmentation is often applied and is described in the third part of the introduction. Fragmentation is heavily influenced by the chemistry of the fragmented ion. Thus, by changing the chemistry of a peptide by derivatizing specific chemical groups one can influence the fragmentation behavior. A comprehensive study on the changes in fragmentation behavior upon electron transfer dissociation (ETD) after derivatization of peptides with different tags is presented in chapter 4.[19] The resulting fragmentation spectra of peptides can be interpreted based on protein databases. If a protein is not present in the database, *de novo* sequencing is necessary. Also, in respect to *de novo* sequencing, chemical derivatization has been shown to help. A new approach to facilitate *de novo* sequencing with the help of chemical labeling is described in chapter 3.[20] The mentioned methods following in the chapters 3 to 6 will further expand the toolbox to decipher the (phospho)proteome.

## IONIZATION TECHNIQUES

A major prerequisite for mass spectrometry is that the analyte is present as a gas phase ion. A number of methods exist for the conversion of solids, liquids and gases into gas phase ions. The majority of these techniques are not appropriate for proteins, since they are too energetic (e.g. electron impact ionization) or provide poor sensitivity (e.g. fast atom bombardment). The two most common techniques applied in proteomics are matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI).[21-24]

### ESI

Nowadays, electrospray ionization (ESI) is arguably the most important of the two ionization techniques applied in proteomics. Therefore, it is astonishing that the principle of electrospray (ES) had already been described, albeit theoretically, in 1882 by Lord Rayleigh.[25] It took until 1968, when the group of Malcolm Dole used ES to create charged polystyrene molecules in the gas phase, before it became associated with mass spectrometry.[26] Still, almost another 20 years passed by before the Fenn group, in the early 1980s, developed ESI into the technique we know now and showed its power for the study

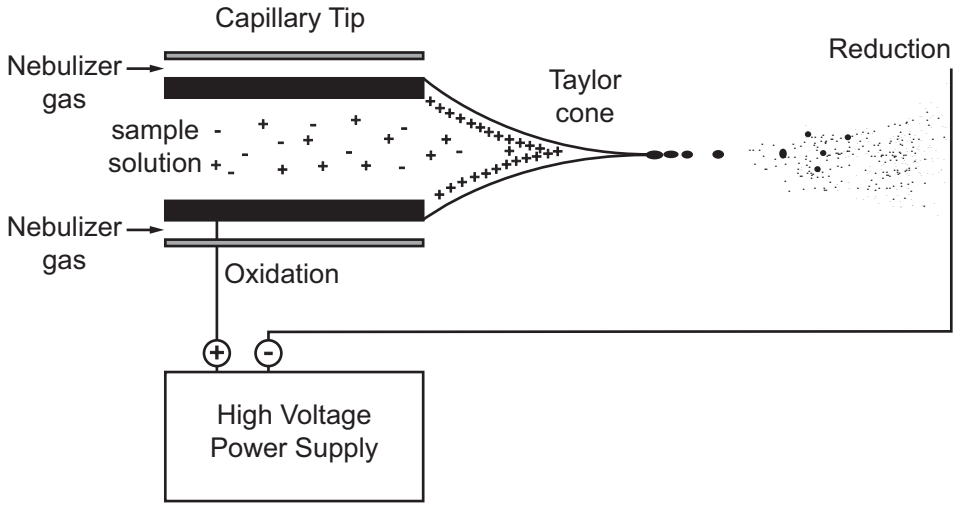



**Figure 1.** Illustration of the instrumental setup to initiate electrospray. A high voltage is applied between the sample tip and a counter electrode, inducing an electrospray. A nebulizer gas assists the formation of smaller droplets at the spray tip. Adapted from Kebarle and Verkerk[27]

of biomolecules.[23, 24] The importance of this development led to John Fenn being awarded the Nobel Prize for chemistry in 2002 (www.nobelprize.org).

ESI is initiated by applying a high voltage between the conductive tip of a capillary or directly to the liquid and a counter electrode, as depicted in Figure 1. The processes taking place can be divided into three major steps, which are first the creation of charged droplets at the tip of the ES capillary. The high voltage induces an accumulation, via repulsion, of charged analyte ions at the tip of the capillary. The charge accumulation deforms the surface of the solvent and with increasing voltage results in a cone shape (Taylor cone) that releases charged droplets (Figure 1).[28] In the second step the droplets shrink due to evaporation of the solvent. Additional heat or a nebulizer gas (applied around the spray tip in the same direction as the spray) is sometimes needed to enhance the evaporation especially when liquid chromatography is coupled with ESI-MS and high flow rates are applied. At the point where the coulombic repulsion of the charges in the droplet overcomes the surface tension of the droplet, the so called Rayleigh limit, the droplet ejects smaller droplets. This process is called coulombic fission.[27] The final step is the generation of gas-phase ions. The exact mechanism how the gas-phase ions are released is not known, but two models have been proposed. The first model called the charge residue model (CRM) proposes that the droplets undergo coulombic fission and evaporation until only one charged molecule remains.[26] This is the most likely mechanism for the creation of gas-phase peptide and protein ions.[27] The ion evaporation model (IEM) on the contrary proposes that singular gas-phase ions can be ejected directly from a droplet before the Rayleigh limit is reached.[29] This is thought to be the favorable mechanism for small (in)organic ions.[27] The charged molecules are then extracted into the mass spectrometer.[30]

## MALDI

Matrix-assisted laser desorption/ionization (MALDI) is another important ionization method in proteomics. In the late 1980s MALDI was introduced by the groups of Tanaka[31] and Hillenkamp/Karas.[21, 32] In 2002, Tanaka, alongside Fenn, was honored with the Nobel Prize for their development of soft desorption ionisation methods for mass spectrometric analyses of biological macromolecules.(www.nobelprize.org) Tanaka used fine cobalt powder suspended in glycerol[31] and Karas aromatic organic compounds as matrix.[21, 32] Nowadays, common matrices applied in proteomics are 2,5-dihydroxybenzoic acid (DHB),[33] sinnapic acid, and *alpha*-cyano-4-hydroxycinnamic acid.[22, 34]
Initially, a matrix solution is mixed with the peptide solution often acidified with trifluoroacetic acid. Also other acids have been applied, for instance phosphoric acid in combination with DHB for improved phosphopeptide detection.[35] The matrix analyte mixture is subsequently co-crystallized on a metal plate (MALDI plate). The main functions of the matrix are to isolate the analytes from each other to prevent cluster formation and to absorb the laser light at its specific wavelength leading to the ionization of the sample. The most common laser applied in MALDI are UV-lasers, e.g. $N_2$ lasers (337 nm). The matrix/sample crystals are irradiated by very short laser pulses (nanoseconds) which desorb and ionize the sample and matrix molecules (Figure 2). The ionization process itself is very complex and poorly understood, but was proposed to be a chemical ionization-like process that induces protonation of the analyte by proton transfer reactions.[22]

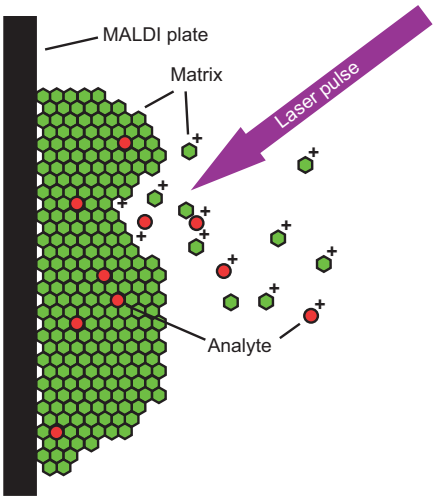



**Figure 2.** Simplified illustration of the MALDI process. Co-crystallized analyte-matrix molecules are irradiated by a laser pulse (violet). The energy of the laser is absorbed by the matrix (green), which leads to ionization of the matrix (green) and the analyte (red).

## MASS ANALYSERS

In general mass analyzers employ dynamic or static electric or magnetic fields in order to separate analytes of different *m/z* values. The strengths and weaknesses of different types of mass analyzers are predominantly determined by their analysis/scan speed, transmission, resolution, mass range and mass accuracy. The following types of mass analyzers are of major importance for shotgun proteomics: quadru-/multipoles, ion traps, time-of-flight (TOF) analyzers and the orbitrap.

## Quadrupoles

The principles of a quadrupole were first described by Paul and Steinwegen in 1953.[36] Quadrupoles consist of 4 perfectly parallel, ideally hyperbolic rods. Opposite rods are connected and the combination of a static and oscillating electric field is applied (Figure 3).[36-38] Gas-phase ions are introduced along the axis into the middle of the radially positioned rods. The same potentials are applied to opposite rods, while adjacent rods have exactly the opposite potential being applied. This dynamic electric field creates a turbulent environment in which ions may or may not possess a stable trajectory. Ions of differing *m/z* values require different combinations of electric fields in order to stably traverse the quadrupole. If the ion has no stable trajectory, it leaves the ion path and/or reaches one of the rods, which leads to the ion being discharged and, as a consequence, the analyte is not detected. If the ion has a stable trajectory through the quadrupole, it will move with a circlular motion in the x- and y-axis, resulting in a corkscrew kind of motion through the rods. Through manipulation of the electric fields, it is possible to isolate ions of a particular *m/z* value or obtain an entire spectrum by scanning through all *m/z* values. When the DC potential is set to ground and only the RF-potential is applied, the quadrupole functions as a wide band mass filter. In this mode, quadrupoles are used as ion guides. Furthermore, quadrupoles can be used as reaction chambers, for instance in collision induced dissociation (CID). CID is further described in a distinct section of the introduction. If instead of four rods, six or eight rods are used, the instrument is called hexapole or octopole, respectively.

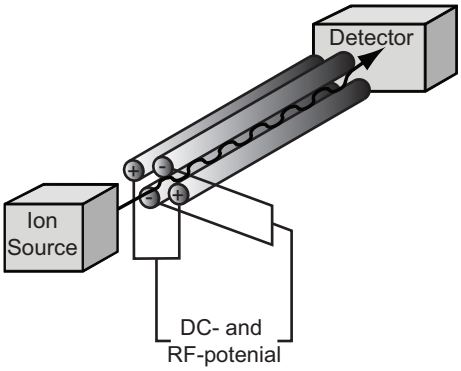



**Figure 3.** Simplified illustration of a quadrupole mass spectrometer.

## Quadrupole and Linear Ion Trap

The Paul ion trap, also called 3D-trap, quadrupole ion trap (QIT), or quadrupolar ion storage (QUISTOR) trap was invented in 1953 by Paul.[37] The trap consists of a circular electrode (ring electrode) and two hyperbolic caps/electrodes (endcaps), as depicted in Figure 4. An oscillating electric field is applied to the ring electrode and the end-cap electrodes are kept at ground potential. Since there is no DC potential applied, the trap will allow ions of a wide mass to charge range to have stable trajectories in 3 dimensions which effectively means the ions are trapped. Discrimination of *m/z* is performed using a changing

amplitude of the RF field and resonance excitation. Often, helium is used as a buffer gas within the trap to focus/trap the ions more efficiently through collisional cooling.[39] Furthermore, careful control of resonance excitation can be used to increase the kinetic energy of the ions and allows collision induced dissociation (CID) to be performed (for CID see section 'fragmentation techniques').[40]

A different type of an ion trap is the 2D-trap, also called linear ion trap (LIT). An LIT consists of a quadrupole with two end electrodes, which can be built of two plates or short quadrupoles as illustrated in Figure 5. The ions are trapped in an axial motion and can be ejected radially through slots in the rods.[41] The design of 2D-traps enables to increase the ion storage volume compared to 3D-traps. A benefit from an increased ion storage volume are reduced space charge effects. Space charge effects limit the maximum load of an ion trap, due to repulsion of ions, which in consequence disrupts their stable trajectories.[41]

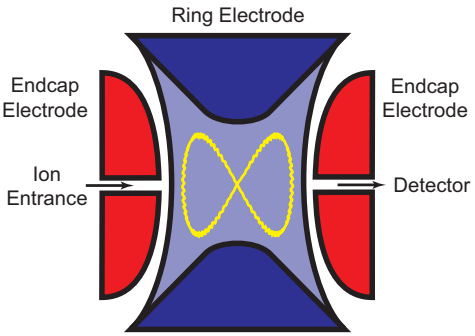



**Figure 4.** Schematic picture of a Paul trap. The ring electrode is depicted in blue and the two endcap electrodes are colored red. Both endcap electrodes contain a hole in order to enable the entrance and exit of the ions. The predicted ion trajectory in the trap is depicted in yellow.

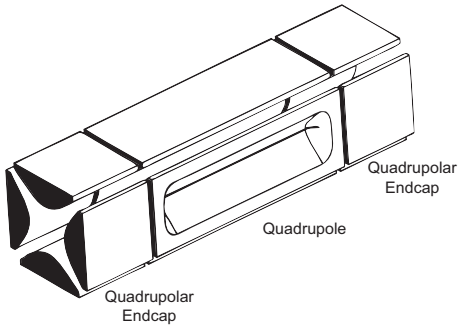



**Figure 5.** Schematic picture of an example of a linear ion trap. In this example the trap consists of three quadrupoles. The small first and last quadrupole function as end plates; endcaps and enable axial trapping. The ions can be ejected either axial or through a slot in one of the rods. Adapted from Schwartz et al.[41]

## Time-of-Flight (TOF)

The concept of time of flight (TOF) was first introduced in 1946 by Stephens.[42] The ions are initially, and ideally, placed in the same location and are at rest. Subsequently, the ions are accelerated to the same kinetic energy using an electric field and are then allowed to traverse a field free region of known length. The square of the velocity ($v$) of the ion is now inversely proportional to the mass ($m$) over charge ($z$),

$$v^2 = \frac{2zeV}{m}$$

where *e* is the fundamental unit of charge and *V* the acceleration voltage. Thus, ions with a high *m/z* are slower and reach the detector later than lower *m/z* ions. Since the distance and time can be accurately measured, one can obtain the *m/z* value of the ion. In order to overcome issues with different initial kinetic energies, an electrostatic ion mirror, a so called reflectron, is often used in TOF analyzers.[43, 44]

**Orbitrap**

The orbitrap is the most recent mass analyzer widely used in proteomics and was first described by Makarov in 2000.[45] However, the concept of orbital trapping was first described by Kingdon in 1923, whose Kingdon trap consisted of a central filament of only 0.01 cm inner diameter which runs axially through a cylindrical electrode with two closed ends.[46] The orbitrap consists of a central spindle-shaped electrode and a surrounding barrel-shaped electrode that is split in half, separated by an insulating ring. Ions are trapped by an electrostatic field applied to the orbitrap and move axially around the central electrode in a harmonic orbit. The oscillating ions induce an image current that is detected with the help of a differential amplifier between the halves of the outer electrode of the orbitrap.[45] The signals are converted into frequencies with the help of Fourier Transform. The mass spectrum can then be calculated by frequency-to-*m/z* conversion.[44] The resolution achieved in the orbitrap is relative high, but is dependent on the time span the ions oscillate around the central electrode. Resolving powers (at *m/z* 400) are approximately 7,500 for 0.2 second read out and up to 100,000 for an 1.6 second read out time.[47]

Modern mass spectrometers are often hybrid instruments, which means that they combine several different mass analyzers in one instrument. The orbitrap is generally combined with a linear ion trap as exemplified in Figure 6. In this example multipoles are used in RF-only mode to transfer and focus the ions. The ions are transmitted into a dual linear ion trap with different gas pressures. The first trap accumulates ions and induces fragmentation at relative higher pressure (compared to the second trap) of approximately $5x10^{-3}$ torr. The scanning of the fragments is conducted in the second trap at a relative lower pressure of approximately $3.5x10^{-4}$ torr. The ions are ejected through slits in the rods in order to reach the detectors.[48] If fragmentation in the dual ion trap is not intended or if the fragments should be read out in the orbitrap, the ions are guided by a multipole to another "linear" ion trap bound in a banana shape, the so called C-trap. The C-trap is used to accumulate and focus ions in order to introduce them into the orbitrap. The orbitrap analyzer is then utilized to read out the *m/z* with high resolution.[48] The mass spectrometer in Figure 6 has an additional multipole for beam-type fragmentation (for details see fragmentation section), the higher-energy collisional dissociation cell (HCD collision cell).[49]

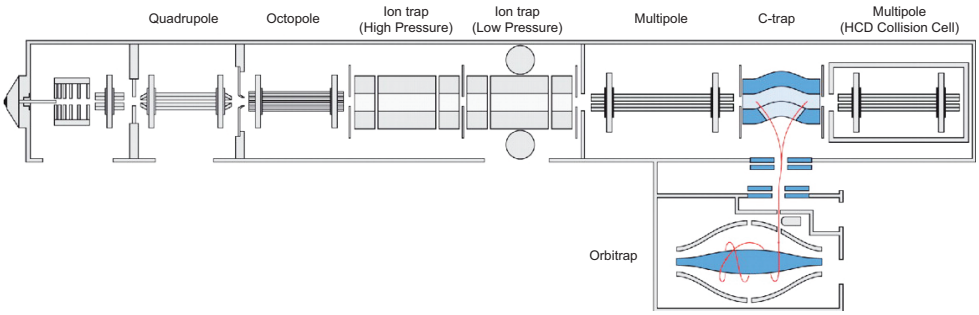



**Figure 6.** Scheme of a modern hybrid mass spectrometer; the LTQ Velos Orbitrap according to Olsen et al.[48] The instrument combines quadru-, and multipoles, three linear ion traps, and an orbitrap. The quadru- and multipoles are functioning as ion guides and as higher-energy C-trap dissociation (HCD) collision cell. The high pressure ion trap is used to capture and fragment the precursor ions, whereas the low pressure ion trap is used to scan the fragment ions. The third ion trap, the C-trap, is used to collect and inject ions into the orbitrap. The orbitrap functions as a high resolution read out device.

## FRAGMENTATION TECHNIQUES

### Collision-Induced Dissociation

Collision-induced dissociation (CID), also called collision-activated dissociation (CAD), is the most popular fragmentation technique for proteomics today.[9] In CID, the analyte ions (generally called precursor or parent ions) are accelerated to a specific kinetic energy by an

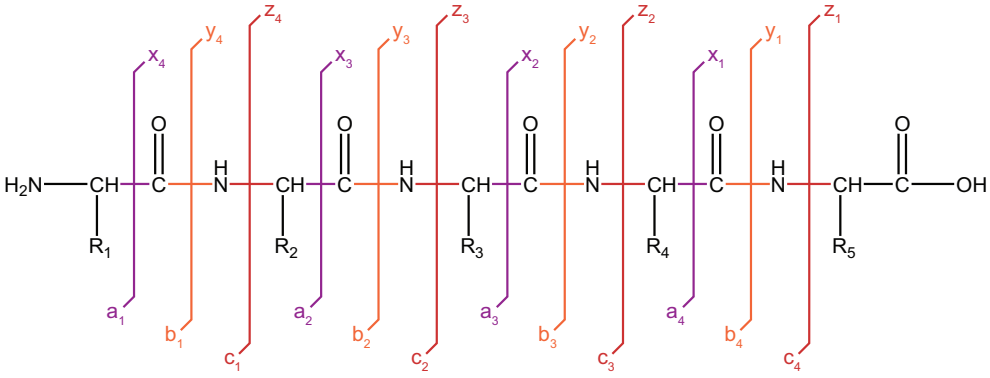



**Figure 7.** Nomenclature of peptide fragment ions as proposed by Roepstorff and Fohlman.[54, 55] Fragments resulting from the dissociation of the Cα-$C_1$-bond (purple) are the N-terminal a-ion fragments and the C-terminal x-ion fragments. Fragments resulting from the breakage of the peptide bond (orange) are the N-terminal b-ion fragments and the C-terminal y-ion fragments. b- and y-ions are predominantly induced by collision-induced dissociation (CID). The fragments resulting from the cleavage of the N-Cα-bond (red) are the N-terminal c-ion fragments and C-terminal z-ion fragments, which are the predominant ion species created in electron based dissociation techniques.

electric field and are subsequently subjected to collisions with an inert gas. Most common gasses include, nitrogen, helium and argon. Upon collision of the precursor with the gas the kinetic energy of the precursor is partially converted into internal vibrational energy that is, eventually, randomized amongst the peptide. Three different collision energy regimes are generally distinguishable and can be linked to a type of mass spectrometer. In TOF/TOF instruments, high-energy CID is applied. The name comes from the initial high kinetic energy (>1 keV) of the precursor ion. High-energy CID usually involves single collision and leads to fragmentation in low microseconds. In triple quadrupole instruments the kinetic energy of the precursor is in the range of 10-200 eV (beam type CID). Fragmentation in beam type CID occurs after multiple collisions with the gas and also occurs on a microsecond time scale but longer compared to high energy CID. The term higher-energy collisional dissociation (HCD) also refers to beam type CID. In ion trap CID (IT-CID) the kinetic energy of the precursor is generated using resonance and corresponds to a few eV. Here, the precursor undergoes multiple collisions with gas molecules, which increases the internal energy of the ion until dissociation occurs. The timeframe of activation in IT-CID is usually several milliseconds.

The fragmentation of protonated peptides is mainly due to charge directed reactions. The 'mobile proton model' introduced by the group of Wysocki[50] and equivalently described by the group of Gaskell[51, 52] explains the dissociation of protonated peptides upon excitation. The model states that upon excitation proton(s) migrate to different amide locations along the backbone of the peptide, as long as they are not sequestered by a basic amino acid side chain. Migration to an amide nitrogen leads to weakening of the bond and combined with the energy provided via CID ultimately leads to dissociation.[50, 53]

A nomenclature of N- and C-terminal peptide fragments has been proposed by Roepstorff and Fohlman[54] (Figure 7) and was further modified by Biemann.[55] The N-terminal fragment ion generated upon cleavage of the amide bond is referred to as a b-ion, while the C-terminal ion is referred to as a y-ion. The choice of which amide bond in a peptide will cleave is semi-random. This fortuitous behavior leads to CID spectra containing ions corresponding to cleavage of most of the amide bonds. Considering that the mass difference between two adjacent amide bonds is essentially dictated by the amino acid and that 18 of the 20 amino acids have a unique mass, one can use CID spectra to obtain peptide sequence information.[53]

**Electron Capture and Electron Transfer Dissociation**

In proteomics, electron capture (ECD) and electron transfer dissociation (ETD) are relatively new dissociation techniques. The technique ECD was first described in 1998 by the group of McLafferty.[56] This process is predominantly conducted in Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers, which allow the trapping of electrons and positively charged peptide ions. Within the cell the two ion populations are forced to overlap and under such conditions an electron can be captured by the charged peptide. The group of McLafferty further postulated that the captured electron populates an excited Rydberg level of the ion. Subsequently, the excited Rydberg state undergoes

relaxations with eventual release of a hydrogen atom, that could attack a disulfide bond or a backbone carbonyl oxygen. In the case of the disulfide bond the attack of the hydrogen atom will lead to a cleavage resulting in an -S• radical and -SH. In the case of the attack of a carbonyl oxygen the transfer leads to the formation of a radical center at the carbonyl carbon ($C_1$), resulting in the cleavage of the adjacent N-$C_\alpha$ bond (Figure 8a).[56, 57] This explains why the N-$C_\alpha$ bonds and not the peptide bonds are cleaved in ECD. The described mechanism is often called the Cornell mechanism and was the first one postulated.[58]

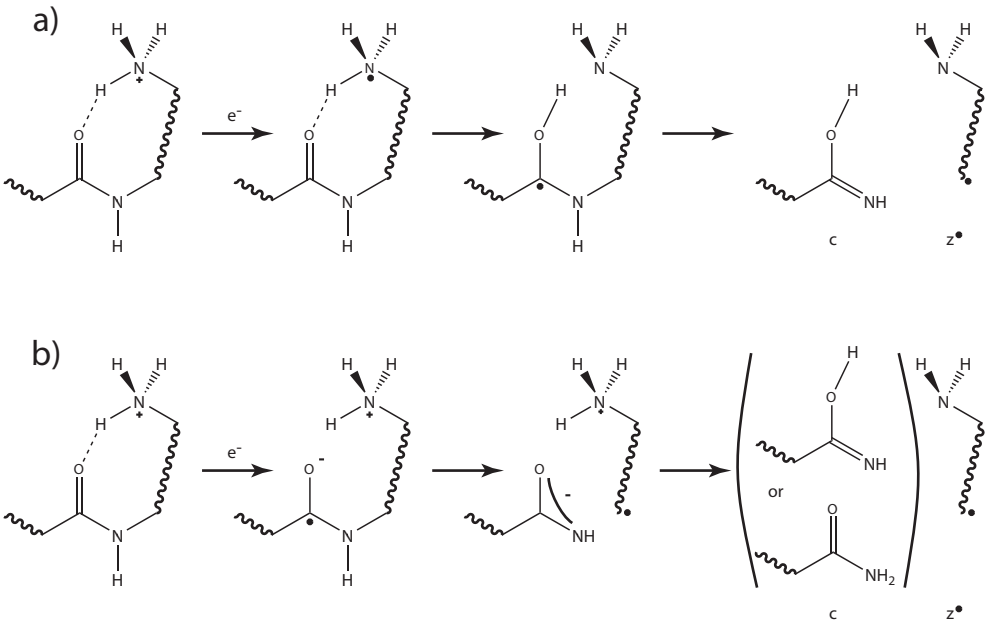



**Figure 8.** Proposed mechanisms of electron based fragmentation. a) Cornell mechanism and b) Utah-Washington mechanism. Adapted from Simons[58]

In contradiction to the Cornell mechanism, it has been shown (by replacing peptide-$[H^+]_x$ by peptide-$[Na^+]_x$) that hydrogen atoms may not be necessary for induction of bond cleavage. These findings resulted in the so called Utah-Washington mechanism, that is depicted in Figure 8b. Here, it is postulated that the electron is either directly captured into an exited S-S or amide orbital or it is captured in an exited Rydberg orbit on a positive site and is transferred (intra-molecular) to the $R_1$-S-S-$R_2$ orbit or amid orbit during its relaxation. In the following, the disulfide bond is cleaved into an $R_1$-S• radical and $R_2$-$S^-$ anion. The N-$C_\alpha$ bond cleavage results in the formation of a p-delocalized R-$^-$OC=NH anion. In the following, intra-molecular proton transfer from the protonated sites to the anions might occur. This would result in the same products as described by the Cornell mechanism (Figure 8).[58] The final products of the cleavage at N-$C_\alpha$ bonds are N-terminal c-type ions and C-terminal z-type radical ions as depicted in Figure 7 and 8.[57, 59]

More recently, a similar technique to ECD called electron transfer dissociation (ETD) has been described by Hunt and coworkers.[60] Here, a negatively charged species is generated, via chemical ionization. A common chemical used as electron transfer reagent is

fluoranthene.[61] The reagent anion is subsequently injected into an ion trap together with the positively charged peptides. Here, an electron is transferred from the negatively charged reagent to the positively charged peptide, which consequently leads to fragmentation of the peptide similar to ECD. Thus, the predominant ions found in ETD spectra are also c-type fragment ions and z-type fragment radical ions (Figure 7). One big difference between ECD and ETD, which also led to the development of ETD, is that in contrast to ECD, ETD enables the use of ion traps for electron based dissociation.

Studies in ECD revealed, that the electron capture cross section is proportional to the square of the charge state of the precursor. This explains why ECD leads to a more efficient fragmentation for higher charged peptides and to the poor dissociation of doubly charged peptides. Unsurprisingly, similar observations have been made with ETD. In order to increase the dissociation efficiency one can excite the nondissociative electron transfer products with supplemental collisional activation (ETcaD).[61] In contrast, CID efficiently fragments doubly charged peptides and struggles more and more with increasing charge state. In essence, ETD and CID are complementary fragmentation techniques with respect to charge state. As modern high resolution mass spectrometers can determine the charge state of the precursor (on the fly) based on its isotopic distribution and can also possess the ability to perform both fragmentation techniques, they can be programmed to choose the most appropriate technique automatically.[62, 63]

A current advantage of electron based dissociation techniques over CID is that ECD and ETD can keep certain acid labile post translational modifications (e.g. phosphorylation) intact, enabling PTM site determination to be easier.[64-66] In addition, the collision based and electron based fragmentation techniques can be complementary for certain sequence situations. The most striking example is the fragmentation behavior of proline residues. In CID a dominant cleavage on the N-terminal side will occur, whereas ETD exhibits no cleavage due to the requirement of two bonds to be broken. Thus, when peptides are fragmented by both techniques more complete information can be gained, which is of special interest when *de novo* sequencing approaches are required.[67]

## OPTIMIZING PEPTIDES FOR FRAGMENTATION

### Choice of Protease

The easiest and most common way of creating peptides with defined properties for improved fragmentation is the use of specific proteases. The best example is trypsin, which is also the most popular protease used in proteomics. Trypsin cleaves adjacent to arginine and lysine residues to create peptides with a basic residue at the C-terminus. Since the N-terminus is a primary amine, regular tryptic peptides contain basic centers, one at each terminus. Upon electrospray ionization, regular peptides often possess 2 protons. When these doubly charged tryptic peptides are fragmented by CID, the proton that is loosely associated with the N-terminus becomes ‘mobile’.[50] Eventually, spectra dominated by C-terminal y-ion fragments and to a lesser extent N-terminal b-ion fragments are created. The

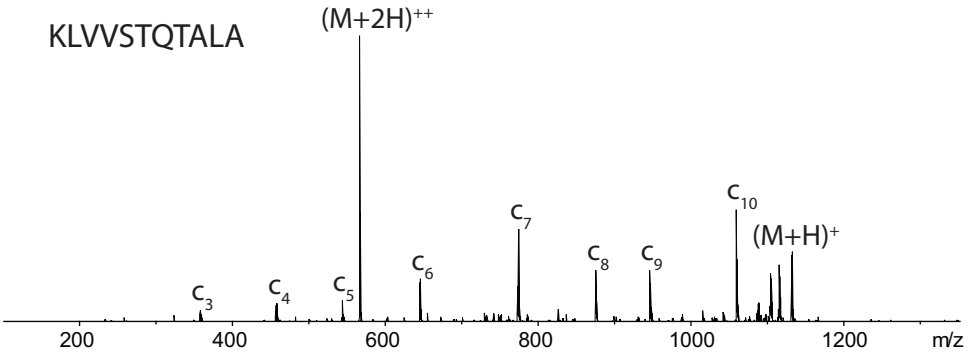



**Figure 9.** ETD spectrum of the Lys-N peptide KLVVSTALA. The dominant c-ion fragment series enables almost complete sequence determination by comparing the mass differences between the c-ions with the known masses of the individual amino acids.

reason being the preferred residence of protons are at sites with high gas-phase basicity. Manipulation of the location of the basic moieties within a peptide can modify the level of b- and y-ions observed and their charge states. Through manipulation, a spectrum containing only one but complete ion series can be potentially generated. Such a spectra would enable facile unambiguous sequence readout. One methodology to create such spectra involves the use of the protease Lys-N. This metalloendoprotease has cleavage specificity N-terminal of lysine residues. The resulting peptides have a basic cluster at the N-terminus, due to the basic lysine being present at the N-terminus. If a mass spectrometric technique is chosen that results in singly charged peptides (like MALDI), the proton will be heavily attracted towards the basic cluster at the N-terminus. The same case will apply upon fragmentation with CID. Thus, spectra of Lys-N peptides generated by MALDI-CID analysis are dominated by N-terminal b-ion fragments, enabling easy sequence determination.[68] The caveat is that this situation is only applicable as long as no other basic residues are present in the sequence. Another strategy is based on the fragmentation of doubly charged Lys-N peptides using ETcaD. The 2+ charge state is the preferred charge state in ESI of regular Lys-N peptides, as long as no additional (besides the terminal lysine) basic residues are present in the sequence. When these doubly charged peptides are subjected to ETcaD, the transferred electron reduces the charge state of the peptide prior to fragmentation, i.e. a single proton remains. The residual proton is directed towards the basic cluster at the N-terminus leading to spectra dominated by N-terminal c-ions. An example of such a spectrum is shown in Figure 9.[19, 69]

**Chemical Modification and MALDI-CID**

Many attempts have been made for the simplification of MS/MS spectra via chemical modification of peptides. A prerequisite for a successful labeling strategy are complete labeling of the chemical group of interest within the peptide and a high specificity of the

chemical reaction. Quick and easy one step reactions with high specificity and an almost complete labeling exist, for instance, for modification of primary amines, and thiol groups. Figure 10 shows the reaction schemes of three common labeling reagents (2, 4 and 7). N-hydroxysuccinimide (NHS) esters (2) are widely used in proteomics to label primary amines.[70] An example is the CAF (chemically assisted fragmentation) reagent which is the NHS ester of 3-sulfopropionic acid.[71] The N-hydoxysuccinimide group is an excellent leaving group making such esters highly susceptible to nucleophilic attack. Importantly, NHS esters (2) preferentially react with the N-terminal amine and the ε-amine of lysine (1) to release NHS and create amides (3) at the peptide. Disadvantages are side reactions with

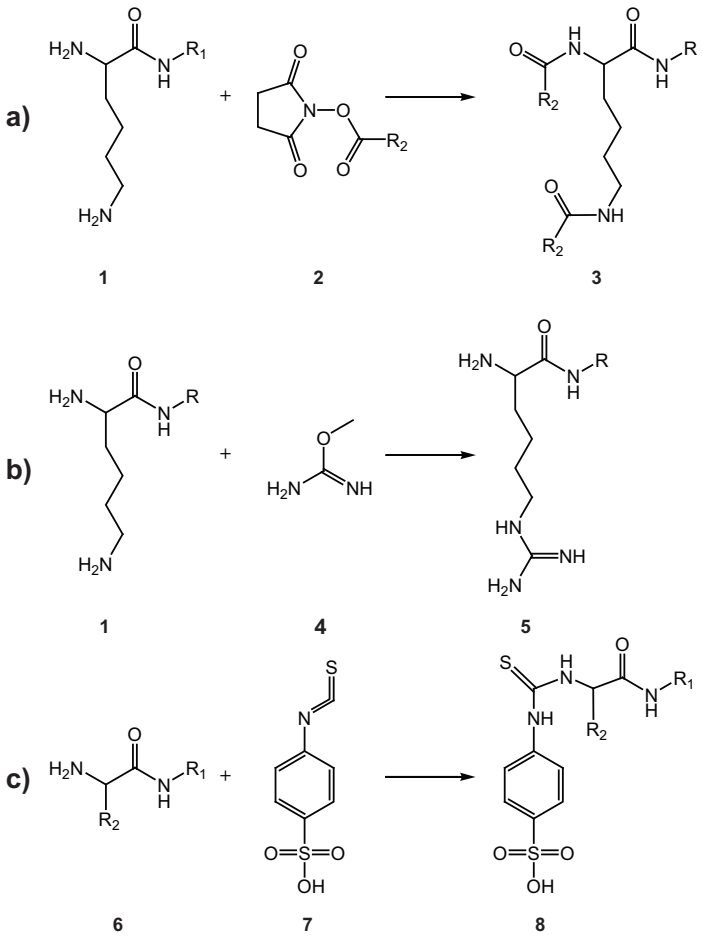
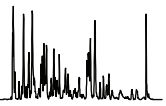



**Figure 10.** Chemical modification of primary amines of peptides. A) Derivatization of the primary amines of the N-terminus and the ε-amine of lysine of a peptide (1) with an reactive NHS-ester (2) to form an amide (3). In CAF $R_2$ is sulfoethyl. B) Specific derivatization of the ε-amine of lysine of a peptide (1) by O-methyl isourea (4) to form a guanidine group (5). C) The primary amine of a peptide (6) is derivatized with 4-sulfophenyl isothiocyanate (7) to form the thiourea derivative (8).



serine and tyrosine residues and the increasing hydrolysis rate at elevated pH especially above pH 8.5.[70, 72] *O*-methyl isourea (4) is commonly used to specifically guanidinate the ε-amine of lysine containing peptides (1).[73-75] The resulting homoarginine (5) differs from an arginine by an increased side chain length of one carbon. High concentrations (~ 1-2 M) of *O*-methyl isourea are usually applied for modification, which requires removal of the reagent prior to MS analysis.[19, 76] The technique is highly specific and minimal side reactions with the N-terminus have been reported.[76] 4-sulfophenyl isothiocyanate (SPITC) (6) specifically derivatizes primary amines (7) to form the thiourea derivative (8). The reaction is very specific, but complete modification is challenging and the reaction conditions generally require desalting, prior to MS analysis.[77, 78]

These chemical modifications can be accomplished with high efficiency either on the protein level[79] or on the peptide level.[19, 77, 80, 81] In order to observe a complete ion series, peptides need a site with a high gas phase basicity at the corresponding terminus. Tryptic peptides contain the basic amino acid lysine or arginine at the C-terminus, but also the N-terminus. In order to observe only C-terminal y-ions, Keough et al. sulfonated the N-terminus of tryptic peptides. The highly acidic sulfonate group at the N-terminus is counterbalanced by the basic residue at the C-terminus, probably by salt bridge formation, resulting in a net charge of zero, prior to ionization. Ionization by MALDI adds an additional proton that is mobile upon excitation as the most basic center (C-terminal arginine or lysine) is occupied.[82] The major dissociation products are C-terminal y-ions, that will dominate the spectrum, and neutral N-terminal fragments, that are not detected by the mass spectrometer.[82, 83] Sulfonic acid derivatives resulting from the reaction with the popular reagent 4-sulfophenyl isothiocyanate (SPITC) (Figure 10 c) show a similar fragmentation behavior.[77, 78]

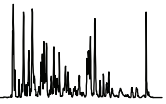
Furthermore, it has been realized that tryptic peptides with a C-terminal lysine residue (lysine is less basic than arginine) ionize less compared to arginine terminated peptides[73, 74, 76] and often show fragments from ion series other than y-ions.[83] In addition, sulfonic acid derivatives have a decreased ionization efficiency. In order to improve the fragmentation and ionization of lysine terminated tryptic peptides, the ε-amine of lysines can be guanidinated(Figure 10 b). The modification leaves the N-terminus to be the only left primary amine, which enables specific sulfonation.[71] The guanidination results in an increased basicity at the derivatized lysine comparable to arginine and thus gives similar results to arginine terminated peptides.[83] In addition, it has been shown that guanidination of lysine residues has a beneficial effect on the MS level, due to the increased gas phase basicity of the homoarginine.[73] Also the nicotinylation of the N-terminus has a positive influence on the MS level, owing to the high gas phase basicity of the nicotinyl group. In addition, N-terminally nicotinylated peptides resulting from an Asp-C digest and containing no additional basic residues lead to predominantly N-terminal b-ions in MALDI-CID.[81]

**Isotopic Identifiers**

Another approach to ease the *de novo* sequencing of peptides is to introduce a label that enables the retrieval of the origin for the fragment. The label has to be introduced at the N- or the C-terminus in order to trace back N- or C-terminal fragments, respectively. It was shown that the C-terminus can be labeled (into pairs) by using a mixture of $H_2{}^{16}O$ and $H_2{}^{18}O$ during tryptic digestion. The different isotopologues are then co-fragmented in order to get specific isotopic patterns or they are individually fragmented and the spectra are compared in order to identify the C-terminal fragments.[84-87] Another method involves the specific labeling of the N-terminus. Munchbach et al. described an approach that includes succinylation of lysine residues at the protein level, followed by Asp-C digestion and nicotinylation of the peptide N-termini with a mixture of conventional 1-(nicotinoyloxy) succinimide ester (Nic-NHS) and $D_4$-Nic-NHS. Co-fragmentation of the isotopologues resulted in N-terminal fragments with a defined isotopic pattern.[81] This pattern enables the identification of N-terminal ion series and as a consequence allows to gain more information and easier interpretation of the spectra. In another approach lysine terminated tryptic peptides were derivatized with either $^{14}N_2$- or $^{15}N_2$-*O*-methylisourea. Simultaneous fragmentation of the resulting isotopologues led to the creation of isotopic patterns in MS/MS that facilitates *de novo* sequencing.[75]

**PEPTIDE AND PROTEIN IDENTIFICATION**

Nowadays, tandem MS is the method of choice to analyze protein digests. The recorded MS/MS spectra provide direct information about the peptide sequence. However, the huge amount of spectra recorded in a standard LC-MS analysis causes manual interpretation to become a futile exercise. The predominant method to assign sequences to tandem MS spectra is database searching.[88] Database search algorithms like Mascot[89] first generate a candidate list of theoretical peptides that coincides the mass of the precursor, within the user defined mass window. This mass window depends on the accuracy and precision of the mass spectrometer. The generation of the theoretical peptides is accomplished by *in silico* digestion of proteins from a protein database, based on user defined search parameters. The size of the database as well as the specificity and efficiency of the digest and possible modifications (both *in vivo* and *in vitro*) of amino acids of the peptide can affect the search space, the theoretical number of peptides per unit of mass. A very unspecific protease will lead to a huge amount of theoretical peptides, created by the *in silico* digestion, as each possible cleavage site is taken into account. In contrast, a protease with specificity for a few amino acids will lead to fewer peptides per protein and thus results in a smaller search space. The efficiency of the digest has a similar influence. If the expected efficiency is poor, peptides containing missed cleavage sites need to be taken into account and, as a consequence, the amount of *in silico* derived theoretical peptides increase the search space. Therefore, proteases, like trypsin, Lys-C, or Lys-N, with a high specificity and a relative high efficiency are preferred.



Modifications (both *in vivo* and *in vitro*) of amino acids are very common. If a modification is possibly present, i.e. variable, then the search engine takes multiple versions of the same peptide into account, where each contains one of the permutations. A classic example is the phosphorylation of serine, threonine or tyrosine residues. If a peptide contains 2 serines (no threonines and tyrosines) then four possible versions of the peptide will need to be generated. This example illustrates that search space rapidly increases upon the introduction of variable modifications.

Once the algorithm has chosen the peptide candidates, it creates theoretical spectra, generated based on known fragmentation rules. However, each search algorithm places different weightings on each fragment ion series. The experimental spectrum is then compared with the theoretical spectra of the candidates. The matching of experimental and theoretical fragment ion peaks is determined by the user defined MS/MS mass window, which depends on the mass accuracy and mass precision in MS/MS mode. For each spectrum peptide match (PSM) a score is reported, which is often based on how much signal can be matched deducted for experimental peaks that are not matched. Again, each search algorithm scores matches and calculates penalties differently.

However, incorrect PSMs are present in every dataset and might even represent the majority of PSMs, especially when low mass accuracy instrumentation has been utilized.[88] Thus, additional scoring models to assess the confidence of PSMs are generally applied, for example by generating a statistical measure like the *p*-value or *E* value (expectation value).[88, 90] A way of calculating the *p*-value is to estimate a null distribution from a frequency histogram of the scores from all but the best candidate matched against the experimental spectrum. In the next step the search engine references the best score against the null distribution.[88] The *E* value corresponds to the number of times the score is expected only by chance. It can be calculated based on the user defined *p*-value (*p*), score of the peptide (*score*) and the number of theoretical peptides in the candidate list (*n*). The *E* value in Mascot for instance is defined as[91]

$$E\ value = p \cdot 10^{\left(-10 \cdot log_{10}\left(20 \cdot \frac{p}{n}\right) - score\right)/10}$$

From this equation follows, that the larger the number of candidate peptides the higher a score needs to be to obtain the same expect value.

In addition to these scoring models, global approaches to compute optimized cut-off scores for individual datasets based on a false discovery rate (FDR) are often calculated. In order to calculate the FDR, a second search is performed against a database with non-sense sequences, a decoy database. Such databases can be generated by reversing, randomizing or scrambling the sequences of the original or 'target' database. The results of the search against the decoy database are considered true negatives and are assumed to be equal to the number of false positives in the original search. The FDR can then be calculated for a conservative threshold score by dividing the number of decoy PSMs by the total number of PSMs.[92]

## SEPARATION TECHNIQUES

### Reversed Phase Liquid Chromatography

The most important type of chromatography for proteomics is reversed phase (RP) high performance liquid chromatography (HPLC). RP-HPLC is a rather old technology, which was first used for peptide separation in 1976.[93] Still, as a result of many further developments it is now the technique of choice to be combined with ESI-MS for the analysis of peptides. RP-HPLC has the advantage that the buffers applied for separation are

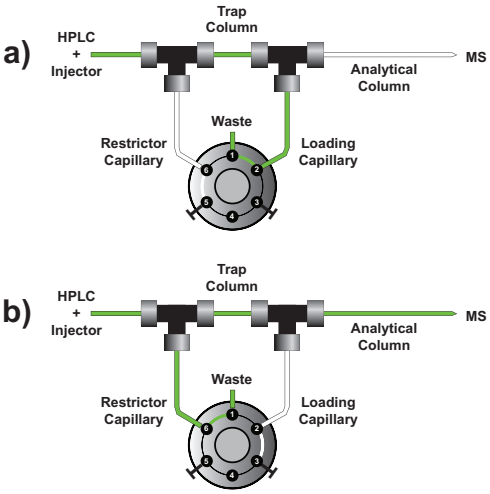
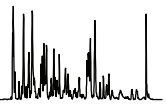



**Figure 11:** Example of an online desalting system, consisting of two T-pieces, a restrictor capillary, a loading capillary, a switching valve, the trap column and the analytical column. The flow path is indicated with green color. a) First the sample is loaded onto the trap column by closing the restrictor capillary and opening the loading capillary. The backpressure of the analytical column prevents the flow to go into the MS. The peptides are captured on the trap column and are desalted. b) The restrictor capillary is opened and the loading capillary is closed. The restrictor capillary functions as a split flow. The backpressure due to the restrictor capillary and the flow rate is high enough that the flow goes over the trap and analytical column into the MS.

compatible with mass spectrometry. Furthermore, RP-HPLC possesses one of the best separation powers of any technique, today.[94, 95] The separation is based on interactions of the analyte with the hydrophobic stationary phase and the mobile phase and can be modeled using partition theory, where the stationary phase is essentially considered a hydrophobic liquid.[96] The mobile phase generally consists of water, an organic solvent and often additives to adjust the pH or influence interactions between the analyte and the stationary phase. Initially, peptides are dissolved in an aqueous acidic solution and are exposed to the hydrophobic stationary phase. Subsequently, the concentration of an organic solvent is gradually increased in order to elute the peptides. A common organic solvent applied in RP for peptide separation is acetonitrile, as it is compatible with ESI and leads to a superior separation power compared to methanol and 2-propanol.[97] The hyphenation with mass spectrometry determines several parameters of the chromatographic system. One parameter is the flow rate. It is normally kept very low (50-200 nL/min). The reason why is that RP-HPLC is typically used in combination with ESI and this ionization technique is concentration dependent.[30] Thus, the sensitivity in MS increases when the flow rate decreases. The low flow rates further determine the dimensions of the columns and



capillaries used, since the separation is dependent on the linear velocity through the column. Common inner diameters (i.d.s) are in the range of 50-100 µm. Another property of ESI has an impact on the chromatographic system: ESI is compromised by nonvolatile salts, which are often present after a protein digestion. Therefore, trap columns are often integrated in RP systems in order to enable online desalting prior to RP separation.[98] Often a switching valve is integrated after the trap column. The switch directs the flow either into waste (desalting) or to the analytical column. Another configuration designed in order to minimize dead volumes and to allow to split the flow rate is shown in Figure 11.[99]

**Ultrahigh-Pressure Liquid Chromatography**

One development that had and still has a great impact on RP chromatography is ultrahigh-pressure liquid chromatography (UHP-LC).[100] The possibility to work at elevated pressures enables the use of smaller particle sizes (sub 2 micron) or longer columns, both of which increase the separation power of a system. A plate number (a measure for the column efficiency of an analyte) of one million has been reported by utilizing a stack of monolithic columns that have a total length between 11-12 meters.[94] Furthermore, peak capacities (measure of the theoretical amount of analytes that can be baseline separated with the chosen method) of up to 1500 could be generated with a 2 meter long column packed with 3 µm particles with a pore size of 300 Å during an 2000 minute gradient working at a pressure of more than 1300 bar.[95] In order to improve their chromatographic separation, more and more proteomic labs are nowadays changing to ultrahigh-pressure systems.

**Multidimensional Separation**

Even though RP is an extremely powerful tool by itself, it reaches its limits when total cell lysate digests are analyzed. Hence, additional separation techniques have been combined with RP. In these systems two separation techniques with different specificities are carried out consecutively. Fractions of the elution of the first separation/dimension are transferred to the second dimension in order to further separate components that are not separated by the first dimension. In such a two dimensional approach a high orthogonality is desired, which means that the analytes in the individual fractions from the first dimension should be evenly distributed in the second dimension. A great advantage of two dimensional systems are the high peak capacities that can be achieved, as the peak capacities of the individual dimensions multiply (assuming complete orthogonality). Thus, peak capacities of up to several thousand can be generated with standard equipment.[101] The most common technique combined with RP in a two dimensional separation is strong cation exchange (SCX) chromatography.[101] In SCX, the analytes are predominantly separated based on their net positive charge, due to coulombic interactions between the positive charges of the molecules and the negatively charged stationary phase.[102-105] As a result of the different modes of separation in SCX and RP, this combination results in a high orthogonality. Various instrumental setups and analytical conditions have been published[12, 106-112] and are further described in detail in chapter 2. Besides SCX, other techniques like hydrophobic

interaction liquid chromatography (HILIC) have been combined with RP. HILIC has been shown to have a high orthogonality with RP, which makes it especially useful for two dimensional separations.[113, 114] Furthermore, RP has been combined with RP in a two dimensional separation approach.[113, 115-117] The difference between the first and second RP separation is the pH of the separation buffer. The pH influences the properties of charged residues in peptides due to protonation or deprotonation. Neutralization of charged residues leads to a decreased hydrophilicity (increased hydrophobicity) and as a consequence changes the retention of peptides in RP. When the difference in pH between the two dimensions are big enough, for instance pH 10 in the first and pH 2,6 in the second dimension, the orthogonality of a two dimensional RP-RP separation reaches the values of SCX-RP approaches.[113] This and other separation techniques and multidimensional separation strategies are reviewed in detail in chapter 2.

## PHOSPHOPROTEOMICS

Phosphorylation of proteins is an important regulation process in mammalian cells. In order to study such processes, the identification and exact localization of the phosphorylation site is needed. In general, proteins are first digested and the resulting peptides are analyzed by LC-MS, as described above. Unfortunately, many phosphorylated proteins are very low abundant and in consequence the resulting phosphopeptides are rarely detected solely by LC-MS.

Thus, several enrichment methods have been developed to overcome the dynamic range problem. Many common methods are based on the chelation of metal ions or metal oxides with phosphopeptides. Common metal ions used in immobilized metal ion affinity chromatography (IMAC) are $Fe^{3+}$, $Zr^{4+}$, and $Ti^{4+}$.[15, 16, 118, 119] The phosphopeptides chelate to the metal ions under low pH conditions, whereas most of the non-phosphorelated peptides do not. Peptides with several acidic amino acids frequently bind to the metal as well, since the carboxylic side chains can also chelate with the metal. Modification of the acidic residues by *O*-methyl esterification enhances the specificity of IMAC,[120] but complete derivatization is challenging.[121] Elution of phosphopeptides and unspecific bound peptides is conducted under basic conditions.

Another methodology for phosphopeptide enrichment is metal oxide affinity chromatography (MOAC). Here, titanium dioxide ($TiO_2$) is by far the most common metal oxide applied.[14] The binding of phosphopeptides to the $TiO_2$ beads is conducted under acidic conditions and the peptides are eluted under basic conditions. The specificity of $TiO_2$ can be further enhanced by washing with solutions containing 2,5-dihydroxybenzoic acid, salicylic acid, phthalic acid, or benzoic acid.[121] Other methods that rely on affinity are based on immunoprecipitation. Most of the immunoprecipitation methods focus on the enrichment of tyrosine phosphorylated peptides[4, 122] or proteins.[123]

In addition to the mentioned affinity based enrichment strategies, several methods have been developed that are based on the liquid chromatographic (LC) separation of phosphopeptides from non-phosphorylated peptides.[124] Strong cation exchange (SCX) is the most common LC-based method for the enrichment of phosphorylated peptides

today.[125] In SCX, the separation is mainly based on the coulombic interaction between a positively charged analyte and the negatively charged stationary phase. In order to suppress unspecific hydrophobic interactions, organic solvents need to be added to the separation buffers.[126] Generally, acetonitrile in a concentration of 25-30% is added to the buffers.[18, 101, 113] When the pH is kept at approximately 3 the acidic amino acids, glutamic and aspartic acid are protonated and do not contribute to the net charge of the peptide any more. At this pH only basic residues and highly acidic residues like phosphorylation contribute to the net charge. Due to the influence of phospho-groups on the net charge, it is possible to separate distinct groups of phosphopeptides from the bulk of non-phosphorylated peptides.[125] A further detailed explanation on this topic is given in chapter 2 and chapter 6.

**OUTLINE**

Nowadays, the method of choice to analyze proteins and peptides is mass spectrometry. Still, several issues in the analysis of especially large pools of proteins like total cell lysates need to be solved to enable complete coverage of all proteins present in such a sample. In general, the proteins are first digested due to peptides being more amenable to sequencing in the mass spectrometer and better suited for many separation techniques.

The most common liquid chromatography based separation techniques for proteomics are reversed phase (RP), ion exchange and hydrophilic interaction liquid chromatography (HILIC). A detailed overview over the principles of these separation techniques is given in chapter 2. This chapter further focuses on the combination of the mentioned techniques for multidimensional separations applied in shotgun proteomics. Furthermore, different instrumental setups and conditions are reviewed and several examples of the application of multidimensional setups are described.

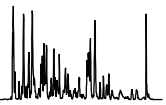
Today, most mass spectrometric data of peptide analyses are interpreted by comparison with *in silico* generated spectra derived from protein or genomic databases. If the sequence information of the protein/peptide of interest is not available in the database, *de novo* sequencing is the method of choice to interpret the recorded spectra. In chapter 3, we describe a novel method for *de novo* sequencing of peptides. The method is based on the digestion of the proteins of interest with Lys-N metalloendoprotease in order to generate peptides with all primary amines at the N-terminus. Dimethyl labeling of the N-terminus with two different isotopes of formaldehyde introduces an isotopic label solely on the N-terminus. When the isotopologues are fragmented together, the fragments resulting from the N-terminus show an isotopic pattern and can be identified as N-terminal fragments. This enables facile *de novo* sequencing. By this methodology, we were able to partially sequence a previously unknown protein from avocado fruit.

Another methodology for *de novo* sequencing is based on the analysis of doubly charged Lys-N peptides with ETD.[69] In chapter 4, we further show, that derivatization of Lys-N peptides with some commonly applied labels lead to better sequence coverage, when analyzed with ETD. Furthermore, we demonstrate that the increase in gas phase basicity by

imidazolinylation of the ε-amine of the C-terminal lysine of doubly charged tryptic peptides leads to predominantly C-terminal fragments, enabling facile sequence readout. We also found that not all labeling strategies used for CID are applicable for ETD. Nicotinylation of peptides, for instance, suppresses the backbone fragmentation under ETD conditions. As a consequence most of the sequence information is lost, making this labeling strategy inappropriate for ETD based sequencing strategies.

In order to enable the analysis of thousands of peptides by mass spectrometry, separation techniques are needed to reduce the complexity prior to the introduction of the peptide ions into the mass spectrometer. In addition to the complexity, some pools of peptides, like phosphorylated peptides, are often present in much lower concentration as the bulk of regular peptides. Therefore, enrichment strategies like SCX chromatography have been developed. SCX has been described in detail in chapter 2. In chapter 5, we applied weak anion exchange (WAX) to further decrease the complexity of the enriched phosphopeptide fractions from SCX. As a result, we were able to identify more than 10,000 phosphopeptides out of a single SCX fraction.

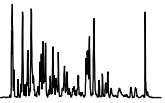
Unfortunately, phosphorylated peptides that contain multiple basic residues elute with the bulk of regular peptides in SCX conducted at a pH of 3. Thus, further enrichment is needed to enable detection of these mainly low abundant pool of phosphorylated peptides. We developed a facile method to enrich for these pools of phosphorylated peptides by conducting an additional separation with SCX at a pH of 1. This method is presented in chapter 6. Here, individual fractions from the first SCX separation at a pH of 3 are separated with SCX at a pH of 1. At a pH of 3 the phospho-group is still negatively charged, whereas at a pH of 1 it is protonated. This changes the net charge of the phosphorylated peptides, whereas the charge of the regular peptides stays the same. Thus, phosphorylated peptides retain differently at a pH of 1 compared to the regular peptides that coelute at a pH of 3, enabling separation of regular and phosphorylated peptides in this tandem SCX approach. By this method, we were able to identify more than 10,000 phosphopeptides out of only 500 µg of protein. We further found that many phosphopeptides identified with tandem SCX contain motives of basophilic kinases, emphasizing the importance of the identified phosphopeptide pools.



## REFERENCES


(1) Mann, M.; Jensen, O. N. *Nat Biotech* **2003**, *21*, 255-261.
(2) Cohen, P. *Nature Cell Biology* **2002**, *4*, E127-E130.
(3) Pereira, S. R.; Vasconcelos, V. M.; Antunes, A. *Critical Reviews in Toxicology* **2011**, *41*, 83-110.
(4) Rush, J.; Moritz, A.; Lee, K. A.; Guo, A.; Goss, V. L.; Spek, E. J.; Zhang, H.; Zha, X. M.; Polakiewicz, R. D.; Comb, M. J. *Nature Biotechnology* **2005**, *23*, 94-101.
(5) Gu, T. L.; Nardone, J.; Wang, Y.; Loriaux, M.; Villen, J.; Beausoleil, S.; Tucker, M.; Kornhauser, J.; Ren, J. M.; MacNeill, J.; Gygi, S. P.; Druker, B. J.; Heinrich, M. C.; Rush, J.; Polakiewicz, R. D. *Plos One* **2011**, *6*, e19169.
(6) Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. *Cell* **2006**, *127*, 635-648.
(7) Schutkowski, M.; Reineke, U.; Reimer, U. *Chembiochem* **2005**, *6*, 513-521.
(8) Wurzenberger, C.; Gerlich, D. W. *Nature Reviews Molecular Cell Biology* **2011**, *12*, 469-482.
(9) Yates, J. R.; Ruse, C. I.; Nakorchevsky, A. *Annual Review of Biomedical Engineering* **2009**, *11*, 49-79.
(10) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198-207.
(11) Domon, B.; Aebersold, R. *Science* **2006**, *312*, 212-217.
(12) Kocher, T.; Swart, R.; Mechtler, K. *Analytical Chemistry* **2011**, *83*, 2699-2704.
(13) Fournier, M. L.; Gilmore, J. M.; Martin-Brown, S. A.; Washburn, M. P. *Chemical Reviews* **2007**, *107*, 3654-3686.
(14) Pinkse, M. W. H.; Uitto, P. M.; Hilhorst, M. J.; Ooms, B.; Heck, A. J. R. *Analytical Chemistry* **2004**, *76*, 3935-3943.
(15) Zhou, H.; Low, T. Y.; Hennrich, M. L.; van der Toorn, H.; Schwend, T.; Zou, H.; Mohammed, S.; Heck, A. J. R. *Molecular & Cellular Proteomics* **2011**, *10*.
(16) Zhou, H. J.; Xu, S. Y.; Ye, M. L.; Feng, S.; Pan, C.; Jiang, X. G.; Li, X.; Han, G. H.; Fu, Y.; Zou, H. *Journal of Proteome Research* **2006**, *5*, 2431-2437.
(17) Zhou, H. L.; Watts, J. D.; Aebersold, R. *Nature Biotechnology* **2001**, *19*, 375-378.
(18) Hennrich, M. L.; Groenewold, V.; Kops, G.; Heck, A. J. R.; Mohammed, S. *Analytical Chemistry* **2011**, *83*, 7137-7143.
(19) Hennrich, M. L.; Boersema, P. J.; van den Toorn, H.; Mischerikow, N.; Heck, A. J. R.; Mohammed, S. *Analytical Chemistry* **2009**, *81*, 7814-7822.
(20) Hennrich, M. L.; Mohammed, S.; Altelaar, A. F. M.; Heck, A. J. R. *Journal of the American Society for Mass Spectrometry* **2010**, *21*, 1957-1965.
(21) Karas, M.; Bachmann, D.; Bahr, U.; Hillenkamp, F. *International Journal of Mass Spectrometry and Ion Processes* **1987**, *78*, 53-68.
(22) Karas, M.; Gluckmann, M.; Schafer, J. *Journal of Mass Spectrometry* **2000**, *35*, 1-12.
(23) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64-71.
(24) Whitehouse, C. M.; Dreyer, R. N.; Yamashita, M.; Fenn, J. B. *Analytical Chemistry* **1985**, *57*, 675-679.
(25) Rayleigh, L. *Philosophical Magazine Series 5* **1882**, *14*, 184-186.
(26) Dole, M.; Mack, L. L.; Hines, R. L. *Journal of Chemical Physics* **1968**, *49*, 2240-&.
(27) Kebarle, P.; Verkerk, U. H. *Mass Spectrometry Reviews* **2009**, *28*, 898-917.
(28) Taylor, G. *Proceedings of the Royal Society of London Series a-Mathematical and Physical Sciences* **1964**, *280*, 383-+.
(29) Iribarne, J. V.; Thomson, B. A. *Journal of Chemical Physics* **1976**, *64*, 2287-2294.
(30) Banks, J. F.; Whitehouse, C. M. *Methods in Enzymology* **1996**, *270*, 486-519.
(31) Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T. *Rapid Communications in Mass Spectrometry* **1988**, *2*, 151-153.
(32) Karas, M.; Hillenkamp, F. *Analytical Chemistry* **1988**, *60*, 2299-2301.
(33) Strupat, K.; Karas, M.; Hillenkamp, F. *International Journal of Mass Spectrometry and Ion Processes* **1991**, *111*, 89-102.
(34) Jaskolla, T. W.; Papasotiriou, D. G.; Karas, M. *Journal of Proteome Research* **2009**, *8*, 3588-3597.
(35) Kjellstrom, S.; Jensen, O. N. *Analytical Chemistry* **2004**, *76*, 5109-5117.
(36) Paul, W.; Steinwedel, H. *Zeitschrift Fur Naturforschung Section a-a Journal of Physical Sciences* **1953**, *8*, 448-450.
(37) Paul, W.; Steinwedel, H., H01J49/42 ed.: German Patent, 1956; Vol. DE 944900.
(38) Paul, W.; Reinhard, H. P.; Vonzahn, U. *Zeitschrift Fur Physik* **1958**, *152*, 143-182.
(39) March, R. E. *Journal of Mass Spectrometry* **1997**, *32*, 351-369.
(40) March, R. E. *Mass Spectrometry Reviews* **2009**, *28*, 961-989.

(41) Schwartz, J. C.; Senko, M. W.; Syka, J. E. P. *Journal of the American Society for Mass Spectrometry* **2002**, *13*, 659-669.
(42) Stephens, W. E. *Physical Review* **1946**, *69*, 691-691.
(43) Mamyrin, B. A.; Karataev, V. I.; Shmikk, D. V.; Zagulin, V. A. *Soviet Physics - Journal of Experimental and Theoretical Physics* **1973**, *37*, 45-48.
(44) Marshall, A. G.; Hendrickson, C. L. In *Annual Review of Analytical Chemistry*; Annual Reviews: Palo Alto, 2008; Vol. 1, pp 579-599.
(45) Makarov, A. *Analytical Chemistry* **2000**, *72*, 1156-1162.
(46) Kingdon, K. H. *Physical Review* **1923**, *21*, 408-418.
(47) Scigelova, M.; Makarov, A. *Proteomics* **2006**, 16-21.
(48) 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. *Molecular & Cellular Proteomics* **2009**, *8*, 2759-2769.
(49) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. *Nature Methods* **2007**, *4*, 709-712.
(50) Dongre, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H. *Journal of the American Chemical Society* **1996**, *118*, 8365-8374.
(51) Cox, K. A.; Gaskell, S. J.; Morris, M.; Whiting, A. *Journal of the American Society for Mass Spectrometry* **1996**, *7*, 522-531.
(52) Summerfield, S. G.; Gaskell, S. J. *International Journal of Mass Spectrometry* **1997**, *165*, 509-521.
(53) Paizs, B.; Suhai, S. *Mass Spectrometry Reviews* **2005**, *24*, 508-548.
(54) Roepstorff, P.; Fohlman, J. *Biomedical Mass Spectrometry* **1984**, *11*, 601.
(55) Biemann, K. *Methods in Enzymology* **1990**, *193*, 886-887.
(56) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *Journal of the American Chemical Society* **1998**, *120*, 3265-3266.
(57) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Analytical Chemistry* **2000**, *72*, 563-573.
(58) Simons, J. *Chemical Physics Letters* **2010**, *484*, 81-95.
(59) Zubarev, R. A. *Current Opinion in Biotechnology* **2004**, *15*, 12-16.
(60) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 9528-9533.
(61) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E. P.; Coon, J. J. *Analytical Chemistry* **2007**, *79*, 477-485.
(62) Frese, C. K.; Altelaar, A. F. M.; Hennrich, M. L.; Nolting, D.; Zeller, M.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. *Journal of Proteome Research* **2011**, *10*, 2377-2388.
(63) Swaney, D. L.; McAlister, G. C.; Coon, J. J. *Nature Methods* **2008**, *5*, 959-964.
(64) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **2006**, *1764*, 1811-1822.
(65) Kjeldsen, F.; Giessing, A. M. B.; Ingrell, C. R.; Jensen, O. N. *Analytical Chemistry* **2007**, *79*, 9243-9252.
(66) Boersema, P. J.; Mohammed, S.; Heck, A. J. R. *Journal of Mass Spectrometry* **2009**, *44*, 861-878.
(67) Bertsch, A.; Leinenbach, A.; Pervukhin, A.; Lubeck, M.; Hartmer, R.; Baessmann, C.; Elnakady, Y. A.; Muller, R.; Bocker, S.; Huber, C. G.; Kohlbacher, O. *Electrophoresis* **2009**, *30*, 3736-3747.
(68) Boersema, P. J.; Taouatas, N.; Altelaar, A. F. M.; Gouw, J. W.; Ross, P. L.; Pappin, D. J.; Heck, A. J. R.; Mohammed, S. *Mol Cell Proteomics* **2009**, *8*, 650-660.
(69) Taouatas, N.; Drugan, M. M.; Heck, A. J. R.; Mohammed, S. *Nature Methods* **2008**, *5*, 405-407.
(70) Sinz, A. *Mass Spectrometry Reviews* **2006**, *25*, 663-682.
(71) Keough, T.; Youngquist, R. S.; Lacey, M. P. *Analytical Chemistry* **2003**, *75*, 156A-165A.
(72) Kalkhof, S.; Sinz, A. *Analytical and Bioanalytical Chemistry* **2008**, *392*, 305-312.
(73) Brancia, F. L.; Oliver, S. G.; Gaskell, S. J. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 2070-2073.
(74) Brancia, F. L.; Openshaw, M. E.; Kumashiro, S. *Rapid Communications in Mass Spectrometry* **2002**, *16*, 2255-2259.
(75) Warwood, S.; Mohammed, S.; Cristea, I. M.; Evans, C.; Whetton, A. D.; Gaskell, S. J. *Rapid Communications in Mass Spectrometry* **2006**, *20*, 3245-3256.
(76) Beardsley, R. L.; Reilly, J. P. *Analytical Chemistry* **2002**, *74*, 1884-1890.
(77) Gevaert, K.; Demol, H.; Martens, L.; Hoorelbeke, B.; Puype, M.; Goethals, M.; Van Damme, J.; De Boeck, S.; Vandekerckhove, J. *Electrophoresis* **2001**, *22*, 1645-1651.
(78) Wang, D. X.; Kalb, S. R.; Cotter, R. J. *Rapid Communications in Mass Spectrometry* **2004**, *18*, 96-102.
(79) Schmidt, A.; Kellermann, J.; Lottspeich, F. *Proteomics* **2005**, *5*, 4-15.




(80) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nature Biotechnology* **1999**, *17*, 994-999.
(81) Munchbach, M.; Quadroni, M.; Miotto, G.; James, P. *Analytical Chemistry* **2000**, *72*, 4047-4057.
(82) Keough, T.; Youngquist, R. S.; Lacey, M. P. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 7131-7136.
(83) Keough, T.; Lacey, M. P.; Youngquist, R. S. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 2348-2356.
(84) Takao, T.; Gonzalez, J.; Yoshidome, K.; Sato, K.; Asada, T.; Kammei, Y.; Shimonishi, Y. *Analytical Chemistry* **1993**, *65*, 2394-2399.
(85) Qin, J.; Herring, C. J.; Zhang, X. L. *Rapid Communications in Mass Spectrometry* **1998**, *12*, 209-216.
(86) Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. *Rapid Communications in Mass Spectrometry* **1997**, *11*, 1015-1024.
(87) Uttenweiler-Joseph, S.; Neubauer, G.; Christoforidis, A.; Zerial, M.; Wilm, M. *Proteomics* **2001**, *1*, 668-682.
(88) Nesvizhskii, A. I. *Journal of Proteomics* **2010**, *73*, 2092-2123.
(89) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551-3567.
(90) Fenyo, D.; Beavis, R. C. *Analytical Chemistry* **2003**, *75*, 768-774.
(91) Brosch, M.; Swamy, S.; Hubbard, T.; Choudhary, J. *Molecular & Cellular Proteomics* **2008**, *7*, 962-970.
(92) Weatherly, D. B.; Atwood, J. A.; Minning, T. A.; Cavola, C.; Tarleton, R. L.; Orlando, R. *Molecular & Cellular Proteomics* **2005**, *4*, 762-772.
(93) Gruber, K. A.; Stein, S.; Brink, L.; Radhakrishnan, A.; Udenfriend, S. *Proceedings of the National Academy of Sciences of the United States of America* **1976**, *73*, 1314-1318.
(94) Miyamoto, K.; Hara, T.; Kobayashi, H.; Morisaka, H.; Tokuda, D.; Horie, K.; Koduki, K.; Makino, S.; Nunez, O.; Yang, C.; Kawabe, T.; Ikegami, T.; Takubo, H.; Ishihama, Y.; Tanaka, N. *Analytical Chemistry* **2008**, *80*, 8741-8750.
(95) Shen, Y. F.; Zhang, R.; Moore, R. J.; Kim, J.; Metz, T. O.; Hixson, K. K.; Zhao, R.; Livesay, E. A.; Udseth, H. R.; Smith, R. D. *Analytical Chemistry* **2005**, *77*, 3090-3100.
(96) Dorsey, J. G.; Cooper, W. T. *Analytical Chemistry* **1994**, *66*, A857-A867.
(97) Guo, D. C.; Mant, C. T.; Taneja, A. K.; Parker, J. M. R.; Hodges, R. S. *Journal of Chromatography* **1986**, *359*, 499-517.
(98) Raijmakers, R.; Berkers, C. R.; de Jong, A.; Ovaa, H.; Heck, A. J. R.; Mohammed, S. *Molecular & Cellular Proteomics* **2008**, *7*, 1755-1762.
(99) Meiring, H. D.; van der Heeft, E.; ten Hove, G. J.; de Jong, A. *Journal of Separation Science* **2002**, *25*, 557-568.
(100) Jorgenson, J. W. In *Annual Review of Analytical Chemistry, Vol 3*; Yeung, E. S., Zare, R. N., Eds.; Annual Reviews: Palo Alto, 2010; Vol. 3, pp 129-150.
(101) Gilar, M.; Daly, A. E.; Kele, M.; Neue, U. D.; Gebler, J. C. *Journal of Chromatography A* **2004**, *1061*, 183-192.
(102) Isobe, T.; Takayasu, T.; Takai, N.; Okuyama, T. *Analytical Biochemistry* **1982**, *122*, 417-425.
(103) Cachia, P. J.; Vaneyk, J.; Chong, P. C. S.; Taneja, A.; Hodges, R. S. *Journal of Chromatography* **1983**, *266*, 651-659.
(104) Mant, C. T.; Hodges, R. S. *Journal of Chromatography* **1985**, *327*, 147-155.
(105) Alpert, A. J.; Andrews, P. C. *Journal of Chromatography* **1988**, *443*, 85-96.
(106) Takahashi, N.; Takahashi, Y.; Putnam, F. W. *Journal of Chromatography* **1983**, *266*, 511-522.
(107) Fränzel, B.; Wolters, D. A. *Proteomics* **2011**, *11*, 3651-3656.
(108) Wolters, D. A.; Washburn, M. P.; Yates, J. R. *Analytical Chemistry* **2001**, *73*, 5683-5690.
(109) Peng, J. M.; Elias, J. E.; Thoreen, C. C.; Licklider, L. J.; Gygi, S. P. *Journal of Proteome Research* **2003**, *2*, 43-50.
(110) Washburn, M. P.; Wolters, D.; Yates, J. R. *Nature Biotechnology* **2001**, *19*, 242-247.
(111) Ballif, B. A.; Villen, J.; Beausoleil, S. A.; Schwartz, D.; Gygi, S. P. *Molecular & Cellular Proteomics* **2004**, *3*, 1093-1101.
(112) Bayer, E. M.; Bottrill, A. R.; Walshaw, J.; Vigouroux, M.; Naldrett, M. J.; Thomas, C. L.; Maule, A. J. *Proteomics* **2006**, *6*, 301-311.
(113) Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. *Analytical Chemistry* **2005**, *77*, 6426-6434.
(114) Boersema, P. J.; Mohammed, S.; Heck, A. J. R. *Analytical and Bioanalytical Chemistry* **2008**, *391*, 151-159.
(115) Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. *Journal of Separation Science* **2005**, *28*, 1694-1703.
(116) Song, C. X.; Ye, M. L.; Han, G. H.; Jiang, X. N.; Wang, F. J.; Yu, Z. Y.; Chen, R.; Zou, H. F. *Analytical Chemistry* **2010**, *82*, 53-56.

(117) Zhou, F.; Sikorski, T. W.; Ficarro, S. B.; Webber, J. T.; Marto, J. A. *Analytical Chemistry* **2011**, *83*, 6996-7005.
(118) Nuhse, T. S.; Stensballe, A.; Jensen, O. N.; Peck, S. C. *Molecular & Cellular Proteomics* **2003**, *2*, 1234-1243.
(119) Villen, J.; Gygi, S. P. *Nature Protocols* **2008**, *3*, 1630-1638.
(120) Ficarro, S. B.; McCleland, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M. *Nature Biotechnology* **2002**, *20*, 301-305.
(121) Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jorgensen, T. J. D. *Molecular & Cellular Proteomics* **2005**, *4*, 873-886.
(122) Boersema, P. J.; Foong, L. Y.; Ding, V. M. Y.; Lemeer, S.; van Breukelen, B.; Philp, R.; Boekhorst, J.; Snel, B.; den Hertog, J.; Choo, A. B. H.; Heck, A. J. R. *Molecular & Cellular Proteomics* **2010**, *9*, 84-99.
(123) Steen, H.; Kuster, B.; Fernandez, M.; Pandey, A.; Mann, M. *Journal of Biological Chemistry* **2002**, *277*, 1031-1039.
(124) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J. X.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 12130-12135.
(125) Mohammed, S.; Heck, A. J. R. *Current Opinion in Biotechnology* **2010**, *22*, 9-16.
(126) Burke, T. W. L.; Mant, C. T.; Black, J. A.; Hodges, R. S. *Journal of Chromatography* **1989**, *476*, 377-389.

# Chapter 2

## Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis

Serena Di Palma*,†,‡ Marco L. Hennrich*,†,‡ Albert J. R. Heck,†,‡ and Shabaz Mohammed,†,‡

*These authors contributed equally to this work
†Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
‡Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

Submitted

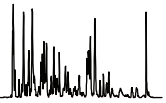


**TABLE OF CONTENTS**

ABSTRACT 36
INTRODUCTION 37
Orthogonality and peak capacity 38
On-line and off-line set-ups 39
Multidimensional separation combined with mass spectrometry 40
REVERSED-PHASE CHROMATOGRAPHY 42
RP in multidimensional strategies 45
ION EXCHANGE CHROMATOGRAPHY 47
Strong Cation Exchange-RP 47
SCX-RP for the Enrichment of Post Translational Modified Peptides 51
Anion Exchange-RP 52
HYDROPHILIC INTERACTIONS LIQUID CHROMATOGRAPHY (HILIC) 55
HILIC in multidimensional strategies 57
HILIC for the analysis of PTMs 59
Alternative separation based on HILIC: ERLIC 61
CONCLUSIONS AND OUTLOOK 64
ACKNOWLEDGEMENTS 64
REFERENCES 65

**ABSTRACT**

Shotgun proteomics dominates the field of proteomics. The foundations of the strategy consist of multiple rounds of peptide separation where chromatography provides the bedrock. Initially, the scene was relatively simple with the majority of strategies based on some type of ion exchange and reversed phase chromatography. The thirst to achieve comprehensivity, when it comes to proteome coverage and the global characterization of post translational modifications, has led to the introduction of several new separations. In this review, we attempt to provide a historical perspective to separations in proteomics as well as indicating the principles of their operation and rationales for their implementation. Furthermore, we provide a guide on what are the possibilities for combining different separations in order to increase peak capacity and proteome coverage. We aim to show how separations enrich the world of proteomics and how further developments may impact the field.

## INTRODUCTION

The proteome describes the entire compliment of proteins in a given biological system expressed at a certain time under particular conditions. The field of proteomics includes the systematic investigations of distribution, abundance, modification, interactions and function for a protein or a set of proteins.[1-3] The vast interest in proteins has led to a significant and persistent effort in development of analytical strategies for proteome analysis.[4] The majority of techniques developed so far require enzymatic digestion of protein mixtures followed by analysis at the peptide level. The identification of the protein is then performed by the peptides being subjected to tandem mass spectrometry[5] (MS/MS) and interpretation of the acquired spectrum by an algorithm which compares the experimental data with *in silico* derivations. Subsequently, the protein is inferred from the identified peptide sequences.[6] This peptide-centric approach, also referred to as the bottom-up strategy[7, 8] in order to distinguish it from the top-down protein-centric strategy,[9-11] has demonstrated to be the most powerful strategy and has given an enormous contribution to the proteomics field. The reduction of proteins to peptides has a number of analytical benefits, such as the denaturation and removal of proteins interactions, and the relatively small size of peptides, which render them easy to handle biochemically. Then, peptides can be more readily induced to fragment and their masses are easier to determine with high accuracy. Furthermore, peptides possess greater solubility in a wider range of solvents, and liquid chromatography (LC) based separations of peptides are superior to methods based on proteins separation.[3, 12]

However, the preferred handling of peptides over proteins faces challenges concerning the higher complexity of the proteome sample and the wider dynamic range, where thousands of peptides, with very similar mass to charge (m/z) ratios, are present at different levels of abundance.[13] Considering that the human proteome (containing approximately 20,000 genes) yields millions of peptides, it becomes clear that this task is not a trivial exercise.[14] Therefore, methods exhibiting highly resolving power are required to maximize the separation of peptides before the mass spectrometric analysis to allow the detection of low-abundance proteins.

Not surprisingly, despite the rapid and massive advances in separation and instrument technology, current analytical strategies are still not sufficient to fully resolve complex biological samples.[15-18] Most approaches in proteomics end with peptide sequencing via nanolitre flow rate reversed phase chromatography and tandem mass spectrometry (nanoLC-MS). This step is relatively fast, reproducible and effective but its use in proteomics is hampered by the limited capability to resolve highly complex samples, such as cell lysate digests. In fact, up to date, no separation method is capable of resolving all components in a single analytical dimension.[19, 20] Therefore, the combination of two or more orthogonal separation procedures is required to increase the peak capacity, and thus the overall resolving power.[21] In the last few years, several multidimensional chromatographic methods, coupled either on-line or off-line, have been introduced to

enhance in-depth proteome analysis and these approaches are now considered routine parts of shotgun proteomics pipeline. The obvious need for separation/fractionation is to minimize the number of co-eluting peptides introduced into the MS at any time during the analysis, allowing the detection of a greater number of peptides.

The aim of this review is to highlight some of the advances and new developments that have been made in the area of liquid chromatographic-based separations, especially their combinations in hyphenated multidimensional formats, to maximize the resolution, reduce sample complexity, widen the overall dynamic range and consequently increase the proteome coverage.

**Orthogonality and peak capacity**

Multidimensional separations are used to address the high complexity of shotgun proteomic samples and required the combination of orthogonal separation techniques to increase the overall resolving power. Giddings first formalized the concept of multidimensional chromatographic separations in 1984.[22] When two separations systems based on different (non-correlative) retention mechanisms are coupled, the resulting 2D system has a higher resolving power than each single dimension. This requires that the selectivity of each separation step towards the analytes must differ substantially in order to maximize the orthogonality, and the resolution achieved in the first separation should be maintained or improved during the second dimension.

Amino acids possess a number of physicochemical properties that can be used as basis for a mode of separation. Peptides in a mixture greatly differ in their physicochemical properties and so separation of peptides can involve differences in polarity (for RP and hydrophilic liquid interaction chromatography, HILIC), charge (ion exchange chromatography) or other peptide properties such as isoelectric point (isoelectric focusing). Some separations employ more than one property for selectivity, and a mixed-mode separation can be achieved when different mechanisms of interaction interplay with each other. Electrostatic repulsion liquid chromatography (ERLIC) and zwitterionic-HILIC are two examples of chromatographic separation where the selectivity for peptides is based on the combination of charge and hydrophilic interactions. Figure 1 illustrates schematically the selectivity of six different types of stationary phases

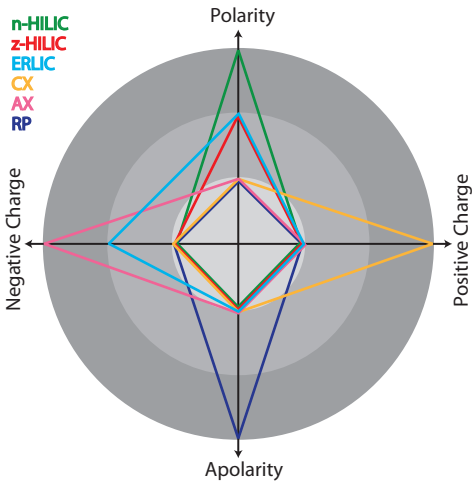



**Figure 1.** Schematic of the selectivity of neutral HILIC (n-HILIC), zwitterionic HILIC (z-HILIC), ERLIC, cation exchange (CX), anion exchange (AX) and reversed phase (RP) chromatography. The light inner circle indicates low and the dark outer circle high selectivity towards positive charge, negative charge, apolarity or polarity, respectively.

on the basis of polariry, hydrophobicity and charge (negative and positive) of peptides, anticipating their orthogonality in multidimensional combinations.[23] One can expect that ion exchange chromatography, which separate peptides based on charge and only partly on hydrophobicity, is more orthogonal to RP than the combination RP-RP in a 2D strategy which is mainly related to hydrophobicity in both dimensions. On the same line HILIC, employing a zwitterionic stationary phase, or ERLIC (a HILIC variant), using anion exchange stationary phase, can potentially display a higher orthogonality than anion or cation exchange towards RP since the peptide selectivity is driven by a combination of polar and ionic interactions.

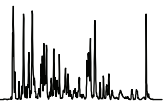
The resolving power of any chromatographic technique is measured by the peak capacity which is defined as the maximum number of peaks that can be separated within a gradient time by a specific system.[24] The peak capacity in multidimensional strategies can be maximized by combining methods that separate peptides on the basis of different physicochemical properties; the final peak capacity for 2D separation results from the linear combination of peak capacities in both separation dimension.[25] Theoretically, the overall separation power becomes the product of each separation dimension, effectively creating a second opportunity for co-eluting analytes to be resolved in the orthogonal second dimension. In reality, several practical considerations limit the achievable peak capacity in 2D-LC. First of all, a situation of perfect orthogonality can rarely be achieved due to the fact that existing LC separation methods display a mixed character of interactions with the stationary phase, increasing the overlap between different techniques.[23] Second, the inefficient transfer of separated zones from one dimension to the other can produce the remixing of separated peaks,[22, 25] resulting in a lower peak capacity than expected.

**On-line and off-line set-ups**

A multidimensional separation can be carried out in an on-line fashion, which involves a direct transfer of the eluent from the first dimension onto the next one, with no flow interruption;[26] or in an off-line way, based on fraction collection in the first dimension and their analysis in the following dimension.[27]

Each approach presents advantages and disadvantages. Off-line set-ups are often more simple in design and operation, and each dimension can be fully and independently optimized.[28] The collection of multiple fractions from an efficient separation/fractionation in the first dimension, in concert with a fast and comprehensive analysis in the second LC dimension, is a crucial step and, when carefully optimized, can offer a significant advantage in terms of analysis time, reducing number of fractions and content of overlap between adjacent fractions.

Off-line approaches offer greater flexibility than on-line separations in the choice of LC modes, especially in the case of incompatibility of mobile phases between consecutive dimensions. In addition, fractions can be manipulated, i.e., diluted, concentrated or dissolved in a different solvent, chemically modified and, if necessary, reanalyzed. But the major limitations of off-line set-ups are related to higher risk of sample loss and contamination, due to sample handling, and lack of automation of the system. However,



performing an experiment in an off-line fashion is not necessarily inferior, and, in real life, is the most common and favorable approach since it allows more combinations than on-line approaches with a non-compromised optimization of each single dimension.
When using on-line set-ups, sample handling is notably reduced since there is continuity in the sample transfer between consecutive dimensions realized via the use of switching valves, additional pumps and trapping columns. On-line configurations are usually designed to avoid dead volume and prevent sample loss, although the sample can still be exposed to additional column connections that may lead to carryover or loss.[29, 30] On-line hyphenation possesses certain advantages compared to offline designs, for instance it can decrease the overall analysis time, but has also more stringent requirements, such as the solvent use for the elution in the first dimension must be a weak eluent in the second dimension, and the second dimension needs to be relatively fast in order not to lose the achieved resolution in the first dimension. In addition, independent of whatever on-line or off-line set-up, the column used in the second dimension must have a small internal diameter in order not to compromise the sensitivity of the mass spectrometric detection.

**Multidimensional separation combined with mass spectrometry**

The successful combination of multidimensional separations with mass spectrometry for proteins and peptides analysis was achieved with the advent of the soft ionization techniques MALDI[31, 32] and ESI,[33-35] which paved the way for the modern bench-top MS proteomics, assisted by the continuous development of new powerful MS instrumentations.[36] While MALDI is usually combined with gel-based separations, high-pressure liquid chromatography (HPLC) separations are more frequently coupled to instrument with an ESI source requiring a continuous flow. This has allowed HPLC to become a standard front end for many proteomic applications, giving rise to several LC-MS set-ups.[5, 21, 37, 38]
Although MS technology has and will continue to rapidly improve with the advent of higher speed, higher resolution and more sensitive instrumentations,[39, 40] even the best dynamic range capabilities tolerated by MS are lower than the wide dynamic range of protein present in samples. Thus, the deeper and comprehensive characterization of proteomes still largely relies on efficient separation technologies prior to MS. Then, mass spectrometry essentially adds an extra dimension thanks to its capability to discriminate single components in co-eluting peaks.
Several chromatographic techniques have been employed in 2D-LC configurations, whereby mostly the second dimension is performed by RP[41, 42] due to its better compatibility with electrospray ionization (ESI) MS,[43] high resolving power and the advantageous sample desalting when the first dimension requires salt gradients. The vast majority of 2D-LC systems developed up-to-date utilize SCX coupled to RP in both on-line[44] and off-line[45] modes, mainly because of the good orthogonality of these two separations. Other 2D-LC strategies have been emerging in recent years as promising alternative to traditional approaches, including size exclusion chromatography (SEC),[46] affinity purification chromatography (AFC),[47] different types or combinations of ion

exchangers,[48, 49] anion and cation mixed-bed exchange,[50] and hydrophilic interaction liquid chromatography (HILIC). [17, 23, 51-53]

In this review, we choose to focus on LC methods which, according to our personal opinion, represent some of the major breakthroughs over the last few years and whose applicability will rapidly increase in the proteomic field; namely, RP, ion-exchange chromatography (IEX), including anion and cation exchangers, and hydrophilic interaction chromatography (HILIC), with different type of stationary phases which we define in this review as neutral (n-HILIC), zwitterionic (z-HILIC) and charged in ERLIC mode. A schematic representation of these chromatographic methods is illustrated in figure 2, where a special emphasis is given to the different modes of interaction between tryptic peptides and these stationary phases.

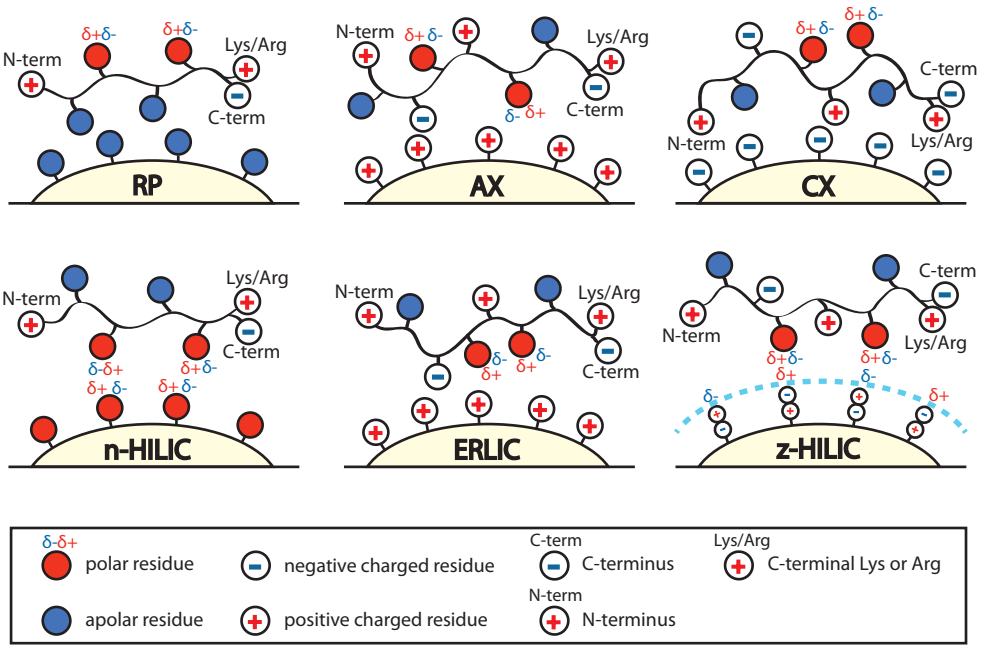
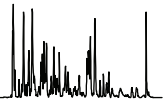



**Figure 2.** Schematic illustration of interactions between polar, apolar, negatively charged and positively charged sites of a peptide and different stationary phases.



## REVERSED-PHASE CHROMATOGRAPHY

Reversed-phase chromatography was introduced as a peptide separation in 1976.[54] Though other types of chromatographic separations have undergone significant developments over the last decade, nowadays reversed-phase (RP) chromatography is still the most widely used method for sample preparation and, as well, separation of peptides prior to MS analysis.

The separation principle of RP is based on the partitioning of the analytes between a hydrophobic stationary phase (which can be visualized as a hydrophobic solvent) and a polar hydrophilic mobile phase. Peptides are loaded onto a RP column under low-organic-solvent conditions, which cause the peptides to partition into the RP material. Salts and the majority of components used in digestion protocols prefer to remain in the low organic solvent and, thus, RP is often used to 'desalt' and concentrate the sample. Separation (or elution of the peptide) is then achieved by increasing the organic modifier content in the mobile phase and, if the mobile phase is of sufficient hydrophobicity, the peptide will start to partition in and out of the two phases as it moves down the column. The order of the eluting peptides relates to the strength of their hydrophobic interactions with the stationary phase. The most common organic modifier employed in RP-LC-MS for peptides separation is acetonitrile, since it offers the best compromise in hydrophobicity, volatility, viscosity and MS compatibility.[55]

Columns often use silica particles as supporting material. The silanol groups are derivatised with aromatic or aliphatic compounds to generate a hydrophobic surface. One drawback of silica materials is the interaction of residual silanol groups with the positive charges of peptides. This effect can be minimized lowering the pH below 4, as silanol groups then become protonated. Often, manufacturers attempt to reduce the number of residual silanol groups with improved derivatisation procedures or attempt to 'cap' these groups post-derivatisation. The choice of these chemistries often have significant effects on the selectivity of columns. In classical peptide separation by LCMS, the RP separation is usually performed under acidic conditions. At those pH values, carboxylic groups of peptides are also protonated and neutral, while basic amino acidic residues (lysine, arginine and histidine) and free N-termini are positively charged. This charge can actually improve retention of peptides if one chooses the correct reagent for acidification. Acetic acid, formic acid and trifluoracetic acid (TFA) reduce the pH and then form a salt bridge with the protonated residues; thus, the protonated group is not strictly charged but becomes a hydrophobic group. The process is referred to as ion pairing. A balance between retention and ionization needs to be made and formic acid is often the acid of choice in LC-MS. Even though TFA would be a better ion pairing reagent, its use is limited since it causes ionization suppression due to the strong salt bridge formation.

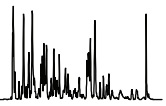
Nowadays, the most often used RP stationary phases are C18 resins, silica derivatised to possess an alkane chain containing 18 carbons, which can also be referred to as an octadecasyl group (ODS). Columns are capillary-scale (internal diameters are below 100

μm) and combined with nano-ESI-MS.[29, 35, 56, 57] The mobile phase employed in RP (a mixture of acetonitrile and water with the acidic ion-pairing reagent) has an excellent compatibility with ESI-MS.[43] The efficiency of the electrospray process is dictated by the size of the initial droplet size, which is dependent on the flow rate and the internal/external diameter of the electrospray needle.[35, 58] One of the most important parameters that effects separation efficiency is the linear velocity (cm/s), as a natural consequence of the van Deemter equation. Thus, there is a drive to reduce flow rates for improving the electrospray process and, in order to maintain separation efficiency (the linear velocity), the column i.d.s are reduced (see below for more details).

Significant efforts have been placed to increase the resolving power,[59] sensitivity[60-63] and speed of analysis[64] under single dimension RP separation. Several parameters are of special interest in order to achieve this goal, such as length and inner diameter (i.d.) of the column, velocity of the solvent, flow rate, particle size, temperature and pressure. Of these parameters, the choice of solvents and the range of flow rate are partially limited due to the hyphenation with ESI-MS. In contrast, the column length and i.d., and the particle size can be varied and have a strong influence on the resolving power. In theory, increasing the length of the column or reducing the particle size will lead to an increase in separation power, although both will be at the cost of an increase in pressure. Therefore, the pressure often represents the limiting factor.[15] The majority of experiments are still performed at room temperature, albeit it is well known that higher temperatures decrease the viscosity of the mobile phase, leading to a decrease in pressure. This can allow more easily the use of longer columns and/or smaller particles, improving the separation power.[65]

A measure for the separation power is the peak capacity, described as the maximum number of peaks that can theoretically be separated in a defined system. It is calculated by dividing the gradient time by the peak width at 4σ (= peak width at 13.4% height). State-of-the-art RP columns provide a peak capacity of several hundreds,[19, 41, 66] depending on their length and the gradient slope. A predicted maximum achievable peak capacity in single-dimensional (1D) RP-LC is within the range of 1400-1600.[67]

Various strategies have been proposed for LC peak capacity improvements in one-dimensional approaches: (i) decreasing the gradient slope (when column length is fixed), (ii) increasing the column length (L) with proportional increase in gradient time, and (iii) employing columns with smaller internal diameter and packed with smaller sorbent particles.[67] The first two strategies have diminishing returns, because the gains in peak capacity are not linear with the increase in analysis time and column length.[19] The third strategy could be more powerful since sufficiently small i.d. columns can provide higher separation efficiencies as well as optimized ESI sensitivity.[41] However, their use presents several technical issues that have to date effectively precluded the routine use of <50 microns i.d. LC columns. First of all, the fabrication of columns with i.d.'s less than 30 μm is extremely difficult. One issue associated with the fabrication of small i.d. columns is the agglomeration of particles, which can prevent successful packing.[41] Effective preparation of small i.d. columns requires smaller diameter particles (below 3 μm) that are highly uniform. Since nonporous particles have lower particle size deviation compared to porous particles, they represent the material of choice when very small i.d.s are chosen, at the cost of



reduced sample loading capacity. The second major obstacle related to the use of smaller i.d. columns packed with smaller particle sizes, as mentioned above, is the increased back pressure, also during the packing procedure. In general, the back pressure created by a column is inversely proportional to the cube of the packing material diameter, thus smaller-diameter columns can easily exceed the maximum operating pressure of conventional HPLC systems (400 bar). The prospects of increasing the separation power using smaller particles was the main driving force to develop high pressure systems capable of working in the regimen of 1000 bar. Special pumping equipment is required to operate such systems, called ultra-performance liquid chromatography (UPLC). We refer to the excellent review of Jorgenson for more details.[68] Even though the use of columns with small i.d.s and small particles working at high pressures is not yet standard, several approaches have proven high potential in terms of separation power.[19, 69] Already, more than ten years ago, the group of Jorgenson used 30 µm i.d. columns of up to 66 cm length packed with 1.5 µm nonporous C18 silica based particles, resulting in highly efficient separations. In order to achieve an optimum flow rate for the separation, they chose to operate the system at approximately 1400 bar, while applying the maximum pressure achievable (up to 4100 bar) they reduced the analysis time, but partly compromised the separation power.[70] In a follow-up study, they used 33 µm i.d. columns with a 27 cm length, packed with 1.0 µm nonporous C18 silica based particles, and demonstrated a peak capacity of 300 for a 30 min analysis of a tryptic ovalbumin digest.[71] A peak capacity of over 1000 has been shown by Shen et al. after 180 min analysis of a tryptic yeast protein digest on a 87 cm x 19.8 µm i.d. column packed with 3 µm particles with a pore size of 300 Å.[41] A few years later the same group analyzed a tryptic digest of Shewanella oneidensis and reached a peak capacity of 1500 in 2000 min (33.3 h) by the use of a 200 cm x 50 µm i.d. column packed with 3 µm particles with a pore size of 300 Å.[72] All these promising applications show that this approach probably represents a bright future in peptide separation strategies.

An RP separation using 75 µm or 100- µm columns packed with 3-5 µm particles at nanoliter flow rate regime (200-300 nL/min) provides, at present, the predominant approach for miniaturized systems. A set of parameters that can be relatively easily achieved packing in-house columns. Such columns are now also becoming available in chip form[73, 74] which allows a broader group of researchers to effortlessly build and operate nanoLC systems.

An appealing alternative that potentially avoids the difficulties of packing traditional C18 columns is the use of silica-based monolithic columns thanks to their robustness, versatile surface chemistry, high column efficiency and fast separation. The permeability of monolithic columns is much higher than for packed C18 columns as a consequence of the presence of large macropores, which permit increased flows through the monolith and, thus, lower pressure requirements.[75] Their use has great advantages when increasing column length and decreasing inner diameter. Luo et al. reported the preparation and applicability of a long (70 cm) silica-based monolithic columns with i.d. of 20 microns[76] and, subsequently, a shorter 25 cm with 10 microns i.d. column,[77] providing high efficiency in separation at low back pressure. In contrast, the group of Tanaka used extreme long columns with larger i.d.s. They connected three C18 silica based monolithic columns of 100 µm i.d. (total length of 11-12 meters), reaching 1 million theoretical plates. They

further reported a peak capacity of 380 in 215 min by the use of a 3 meter column.[78] Recently, Iwasaki et al. used a 3.5 meter-long silica based monolythic column (100 µm i.d.) at ambient temperature, with pressures less than 200 bar, resulting in a peak capacity over 600 during a 41h gradient.[79] However, monolithic columns suffer from some important limitations: first, each synthesis is unique and not easily standardized, resulting in a low column-to-column reproducibility; second, monolithic columns can be easily overloaded, which may represent a serious obstacle when high sample amounts are being injected.

**RP in multidimensional strategies**

Although relatively complex mixtures can be handled by RP because of its excellent resolving power, reproducibility and efficiency, peptides analysis in shotgun experiments is still limited when using one-dimensional RP-based approaches. An efficient way to address the limited peak capacity is to integrate RP as part of a multidimensional separation strategy. Various combinations of orthogonal separation methods have been employed for more comprehensive analyses, as reported in other excellent reviews[21, 80]

The use of two identical chromatographic separations does not represent an ideal approach to increase the overall resolving power. Thus, the combination of RP as both first and second dimensions in 2D-LC strategies does not look promising at first glance, due to the primarily hydrophobic-based retention in both dimensions. However, the selectivity for peptides under RP separation, among hydrophobicity, can be modulated by varying other factors, such as type of stationary phase, mobile phase pH, choice of organic modifier and ion pairing reagent. Some of these parameters might be helpful to achieve a higher degree of orthogonality in 2D-LC configurations. For instance, the effect of changing the type of stationary phase has been investigated using different functionalities, such as C18, phenyl, or pentafluorophenyl (PFP).[23]

The evaluation of their selectivity showed a high correlation, thus the combination of two different types of RP columns provides only a limited orthogonality and is not suitable for 2D-LC applications.[23] In contrast, the impact of the mobile phase pH has a more pronounced effect on altering peptides selectivity than the stationary phase.[81-84] Since

**Table 1.** Internal $pK_a$ values of the side chains of different amino acids and the N- and C-terminus as measured by different groups.[86]

| | $pK_{int}$ (Nozaki&Tanford)[85] | $pK_{int}$ (Thurlkill)[86] | $pK_{int}$ (Gurd Lab)[87] | $pK_{int}$ (Wuthrich Lab)[88] |
|---|---|---|---|---|
| **α-Carboxyl** | **3.8** | **3.7** | **3.3** | **x** |
| **Asp** | **4** | **3.7** | **3.9** | **3.9** |
| **Glu** | **4.4** | **4.25** | **4.2** | **4.2** |
| **His** | **6.3** | **6.5** | **6.8** | **6.9** |
| **a-Amino** | **7.5** | **8** | **8.1** | **x** |
| **Lys** | **10.4** | **10.4** | **10.5** | **11** |
| **Arg** | **12** | **x** | **x** | **x** |

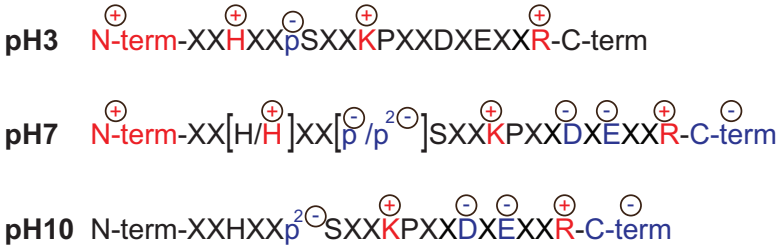



**Figure 3.** Charge states of individual residues in the peptide XXHXXpSXXKPXXDXEXXR at pH3, pH7 and pH10. Red indicates a positive charge and blue indicates a negative charge. P in lowercase refers to phosphorylation. Letters in brackets indicate that the charge state is undefined. The charges were drawn based on the internal pKa values of the individual residues.

peptides are charged molecules comprised of ionizable basic and acidic functional groups, the change of mobile phase pH also affects their retention behavior. Neutralization of a charged residue decreases its hydrophilicity or increases its hydrophobicity and, consequently, leads to a better retention.

Thus, the pH is a potent tool to obtain more orthogonal separations and can be integrated in a 2D RP-RP approach based on the use of different pH values of the mobile phase in the first and second dimension. Gilar et al.[23] evaluated the impact of pH on peptide selectivity under RP separation using different pH conditions, ranging from 2.6 to 10. They showed that acidic peptides are more retained at low pH where the carboxylic acids are protonated/neutralized (increased hydrophobicity), while basic peptides are better retained at high pH, due to the deprotonation/neutralization of basic residues. In this respect the internal pKa values of the side chain functionalities play a major role, as they indicate at which pH value a moiety is charged or not (see table 1).

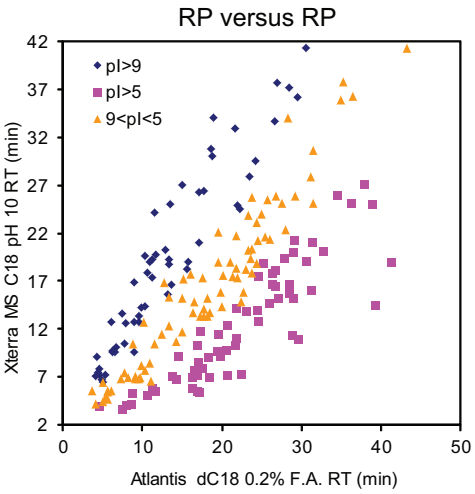



**Figure 4**: Normalized retention time plot showing the impact on the pH on achieving orthogonality in RP-RP separation. Adapted from Gilar et al.[81]

The comparison, by Gilar et al., between LC systems operating at pH 2.6 versus 7.8, 8.5, or 10 showed that greater orthogonality was reached when a wider pH gap was used. This might be the result of the change in the protonation state of the primary amine of the N-terminus (pKa between 8 and 10) and histidine residues (pKa app. 6). The total deprotonation of histidine and the partial up to total deprotonation of the N-terminus at a pH of 10 change the hydrophobicity, whereas at a pH of 7.8 or 8.5 the N-termini may be still protonated

(see figure 3). Taking into account these results, Gilar et al. designed a RP-RP bidimensional system using pH 10 in the first and pH 2.6 in the second dimension, employing columns with identical (C18) packing material, and confirming that the orthogonality was solely dictated by the pH effect.[23, 81]

Nevertheless, the orthogonality of RP-RP in multidimensional systems needs further improvements. A strategy to address this problem was recently proposed by the group of Zou.[89] This mainly theoretical approach is based on 2D high pH-low pH RP-RP combined with a new fractionation plus pooling scheme where the early fractions are pooled with late fractions and then analyzed in the second RP dimension. In this way the separation space in the second dimension is used more efficiently and, as a consequence, the orthogonality is increased.[89]

Each RP-RP approach suffers from certain limitations, for instance the incompatibility of mobile phases due to the high organic content employed for the elution in the first dimension, which makes it difficult to design the 2D system in an on-line format. Therefore, off-line strategies are usually favorable, based on a partial evaporation of the collected fractions to decrease the organic content prior to the second dimension in order to minimize sample losses. A major concern related to the use of alkaline pH with silica-based stationary phase is the long-term stability of the column and the loss of performance during extended use, which may be tackled by using more pH-resistant polymer-based stationary phases, such as PS-DVB[83] or specifically end-capped silica particles.[23]

In summary, the 2D RP-RP approach has in our opinion great potential to play a major role in the future, due to the high resolving power of RP in both dimensions and the relative high orthogonality (comparable to SCX) as long as the difference of pH in the two dimensions is kept high (pH 10 and pH 2.6).[23]

## ION EXCHANGE CHROMATOGRAPHY

Ion exchange chromatography (IEX) has been used for peptide separation on HPLC systems for a number of decades.[90-94] The separation is mainly based on Coulombic interactions between charged groups of the analyte and the charged stationary phase. In order to elute the peptides salts are generally mixed into the mobile phase. The salt's cation or anion population competitively displaces the peptide. A higher concentration of salt is required for stronger bound (more highly charged) peptides. Elution can also be conducted by a change of the mobile phase pH in order to neutralize or invert the charge of the analyte and/or the stationary phase. Principally, there are two main forms of IEX; cation exchange (CX) and anion exchange (AX).

### Strong Cation Exchange-RP

Cation exchange (CX) materials were proven applicable for separating peptides in the 1980's[91-94] and the first attempts in two dimensional peptide separations containing an CX step were also conducted in these times.[95] Over the last decade, the combination of CX with RP has been the most popular multidimensional approach for peptide separation. In CX the

stationary phase has an anionic functional group that enables binding of molecules with cationic groups. In strong cation exchange (SCX) chromatography the functional groups are strong acids, which enables to work at a wide pH range. The most common functional groups in SCX materials are sulfonic acid derivatives. Due to the low $pK_a$ value of the sulfonic acid group, these materials can be utilized at very low pH without losing their binding capacities. The application of weak cation exchange (WCX) materials on the contrary is restricted to a smaller pH range (above approx. pH 5), which explains its rare use in shotgun proteomics. Working at a low pH of approximately 3 is common in SCX of peptides. Here, the carboxylic acid of the side chains of aspartic and glutamic acid, as well as the C-terminal carboxylic acid, are neutralized by protonation and the basic sites are positively charged, as visualized in Figure 3. This results in most of the tryptic peptides having a positive net charge, which enables binding to the anionic SCX material.

IEX and thus SCX separate predominantly based on charge, whereas RP-LC separation is based on hydrophobicity. The two different modes of separation lead to a high orthogonality and thus make these two techniques good partners for two dimensional separation approaches.[23] When SCX is combined with RP, in general, SCX is used as the first dimension. The two main reasons are the excellent compatibility of RP-LC with mass spectrometry and that the method with the highest separation power is applied as last dimension.[67, 96]

Figure 5 illustrates that the separation power of RP is better than SCX and illustrates that peptides elute in clusters depending on their charge state in SCX whereas they are more evenly distributed in RP. Determination of the separation power of SCX and RP-LC has been conducted by Gilar et al.[23] For SCX they determined peak capacities of 63, 85 and 113 using 20, 40, and 80 min gradients, respectively. In Gilar's study RP-LC reached over 350 for a 100 min gradient,[67] and, furthermore, higher values can be reached when working with ultra-high pressures. For example Köcher et al. managed a separation of 700 when utilizing 2 µm material in combination with ultra-high pressure and gradients of up to 10 hours.[19] Further information on the achievements in RP chromatography are given in the previous section about RP chromatography.

SCX-RP has been applied to a great number of research topics. Numerous proteomes or subproteomes of bacteria,[97] fungi,[29, 44, 98] animals[73] and plants[99] have been analyzed by

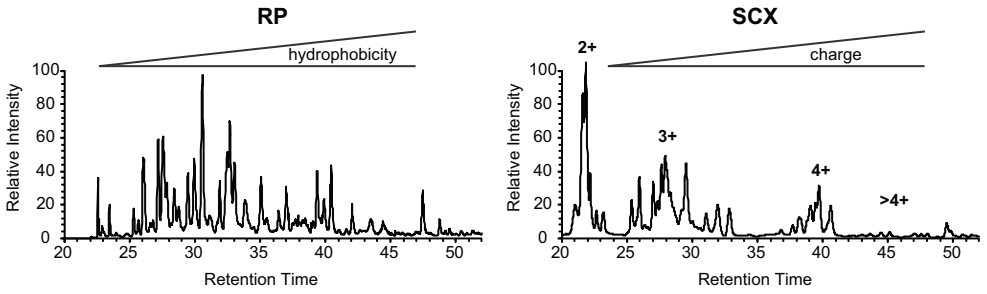



**Figure 5.** Chromatograms of a mixture of bovine serum albumin and bovine casein separated by RP and SCX. The gradient time was approximately 30 minutes for each separation.

SCX-RP and the identification of thousands of proteins is now standard, as exemplified by the analysis of human embryonic stem cells, which resulted in the identification of over 5000 proteins even using a highly conservative Mascot score cutoff of greater than 35.[100]
The easiest way of combining SCX and RP is by constructing an offline setup like the original from Takahashi et al..[95] Here, the sample is first separated by SCX and fractions are collected, as depicted in Figure 6. As a consequence there is a great variability, as SCX separation and RP are conducted completely independent from each other. Thus, there are no limitations in the SCX separation. The SCX column dimensions can be chosen depending on the sample amount to be analyzed and the optimum flow rate can be used. Also, all common buffer system can be utilized for SCX, due to the possibility to remove unwanted constituents prior to RP-LC-MS analysis. Thus, non-MS-compatible phosphate buffers and salts like sodium or potassium chloride are often applied for elution and the collected fractions are desalted prior to MS analysis. Furthermore, acetonitrile is commonly added to the mobile phases of the SCX separation. The addition of acetonitrile to the buffers reduces the unspecific binding to the support material. As a consequence the influence of the hydrophobicity of the peptide on the retention can be minimized,[101] and thus the separation is mainly based on the net charge of the analyte, which increases the orthogonality in SCX-RP. Thus, often, more than 20% acetonitrile is part of the SCX buffers in offline setups.[23, 29, 49, 67] After the SCX separation the acetonitrile can be removed by concentrating the sample by vacuum centrifugation.

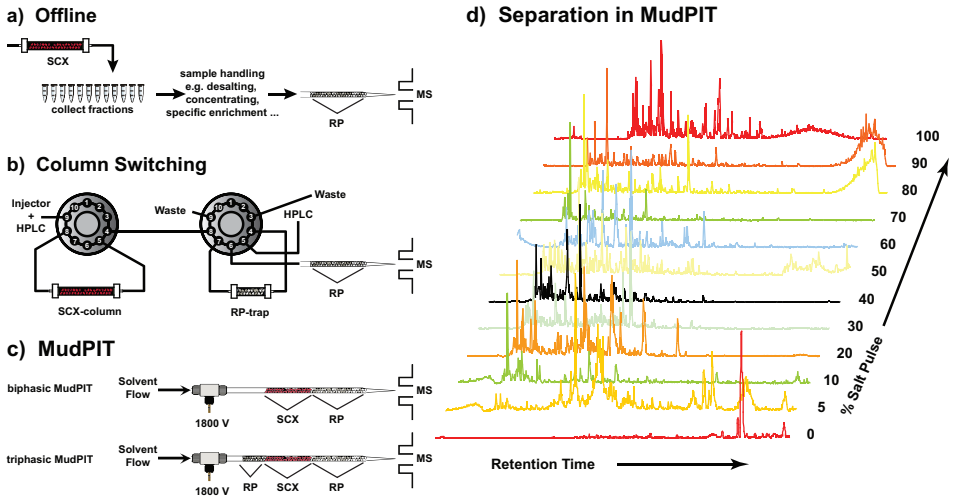
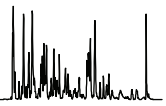



**Figure 6.** Different instrumental setups for two dimensional separations applying SCX and RP. a) In an offline setup the sample is first separated by SCX and fractions are collected. The fractions can be processed if needed and are subsequently separated by RP-LC and analyzed by MS. b) An example of an online column switching setup. The sample is first loaded onto the SCX column and step wise eluted onto the trap column. The sample is first desalted and subsequently eluted onto the analytical RP column followed by MS analysis. c) In the MudPIT approach the SCX and RP material are in one capillary that also functions as spray tip for direct MS analysis. In the triphasic setup an additional RP phase is packed before the SCX and functions as a trap for desalting the sample prior to SCX-RP-MS. d) Chromatograms of a MudPIT analysis. Each color indicates the RP separation after one salt step. c) and d) adapted from Yates et al.[44]



Furthermore, additional steps like phosphopeptide enrichment can be conducted between SCX and RP.[73] An additional great advantage of an offline approach is that selected fractions of interest can be concentrated, further processed, analyzed, reanalyzed or stored until further needed. The biggest disadvantage of the offline approach are the sample losses, due to the sample handling between the different dimensions.

In addition to the offline setup, various online setups have been used for SCX-RP of peptides. The online-approaches can be further divided into column-switching and multidimensional protein identification technology (MudPIT) systems. In the late nineties the group of Yates provided the blueprint of two dimensional chromatography for proteomics, when they introduced MudPIT.[44, 98] Here, the SCX column and the subsequent RP column are combined in one fused silica capillary needle that is directly connected to the mass spectrometer and also acts as an ESI emitter (Figure 6c). In MudPIT the sample is first trapped on the SCX phase. The peptides are partially eluted from the SCX onto the RP phase by increasing salt steps. In between the salt steps, the peptides are eluted into the mass spectrometer by the use of an acetonitrile gradient (Figure 6c and d).

MudPIT has the major advantage that the whole procedure is automated and in general this reduces the time of the total analysis. In addition, sample loss is reduced as the sample is directly transferred from the first dimension to the second. One disadvantage of MudPIT is the limited sample amount that can be loaded. Furthermore, the loaded sample needs to be desalted prior to loading to the system. In order to desalt online a triphasic column (RP-SCX-RP) as depicted in Figure 6c can be utilized. The additional RP material acts as a trap and is used to desalt the sample prior to the SCX-RP separation.[102] In the initial description of MudPIT in which it was named integrated direct analysis of large protein complexes (integrated DALPC), nonvolatile salts ($K_2HPO_4$ and KCl) were used,[103] which turned out to be problematic in combination with MS.[44] Soon, they were substituted by volatile salts like ammonium acetate.[44, 69, 98, 102] Besides the advantage to be the easiest online setup, the major drawback is the limited lifetime of the MudPIT columns and the general risk of online approaches that if an error occurs during the long run a great part of the sample might be lost. The original setup has been further developed in the last few years by creating a triphasic column (RP-SCX-RP) as mentioned.[102] Another development is the combination of ultra-high pressure (UHP) LC, operating at >1300 bar, with MudPIT (UHP-MudPIT).[69] Still, UHP is a relative new technique and not easy to operate. Furthermore, operating a triphasic UHP-MudPIT setup at elevated temperatures has been shown beneficial compared to operating at room temperature. The most identifications were achieved at a temperature of 45°C. By analyzing the cytosolic and membrane fraction using a 45 cm column and a 20 hour 6-step MudPIT at 45°C, 46% of the proteome of Escherichia coli could be identified which corresponds to almost 1900 proteins.[97]

In addition to pure SCX materials, mixed bed columns have been tested in MudPIT for their advantages over standard SCX-RP. Mixed bed columns combine at least two different modes of separation in one column, by mixing particles with different selectivity. Motoyama et al. investigated the use of weak anion exchange (WAX) and SCX material in a mixed bed.[50] They used a triphasic MudPIT setup and compared solely SCX (RP-SCX-RP) and WAX (RP-WAX-RP) with different blends of WAX and SCX (RP-WAX/SCX-

RP). Blends of WAX and SCX material in a ratio of 1:2, 1:1, 2:1 and 5:1 were tested. In addition, WAX and SCX columns were connected in tandem and compared against the triphasic MudPIT setups. Out of all the 2:1 WAX:SCX blend led to the greatest increase in peptide identifications. In addition, they observed an improved orthogonality with the mixed bed compared to solely SCX. They stated that the increased orthogonality most likely contributes significantly to the improved identification rate.[50]

In column-switching based online approaches, the SCX and the C18 column are connected via switching valves, as depicted in the example in Figure 6b. Opiteck et al. were one of the first to build an SCX-RP online system to analyze whole proteins coupled to MS.[46] Since then, diverse setups were constructed to connect SCX and RP-LC-MS. Often in these setups trap columns are integrated between the SCX and RP-LC. The trap columns enable desalting and in consequence the use of common non-volatile buffers and salts for elution in the initial SCX. In addition, the integration of trapping columns enables the concentration of the sample and thus gives more freedom in the flow rates applied for the initial SCX separation.[104] Furthermore, results can differ when a pump is used to deliver the buffers for gradient elution from the SCX[105] or step elution is conducted by injection of SCX elution buffer with defined salt concentrations. In any case the two separation techniques need to be adapted to each other which makes these setups very complicated and not appropriate for routine applications. Still, the advantages of reduced sample loss, relative fast analysis due to automation and the applicability to minor amounts of sample, led to these exciting developments.

**SCX-RP for the Enrichment of Post Translational Modified Peptides**

In addition to the separation of regular peptides, SCX has been optimized to enrich for certain classes of post translationally modified peptides. The most prominent example is the enrichment of sub groups of phosphorylated tryptic peptides, which was first conducted with SCX by the group of Gygi.[73] Here, the pH of the SCX buffers are kept at approximately 3.

In this pH range carboxylic acid residues, being predominantly protonated as their $pK_a$ value is higher than three (Table 1),[86] do not contribute to the net charge of the peptide, whereas the phospho-sites are still negatively charged and reduce the net charge by one in comparison to their unphosphorylated counterparts, as depicted in Figure 3. The difference in net charge state enables the separation of phosphopeptides from their non-phosphorylated counterparts. Thus, enrichment of pools of phosphopeptides that do not co-elute with lower charged regular peptides is possible.[49, 73, 105, 106] These pools are multiply phosphorylated peptides and singly phosphorylated peptides, that contain only a single basic amino acid. Singly phosphorylated peptides with more than one basic amino acid have the same net charge as non-phosphorylated peptides with one basic residue less. Thus, these classes of phosphopeptides co-elute with regular peptides and need further enrichment where immobilized metal affinity chromatography (IMAC)[107, 108] or $TiO_2$[109-111] is most common.

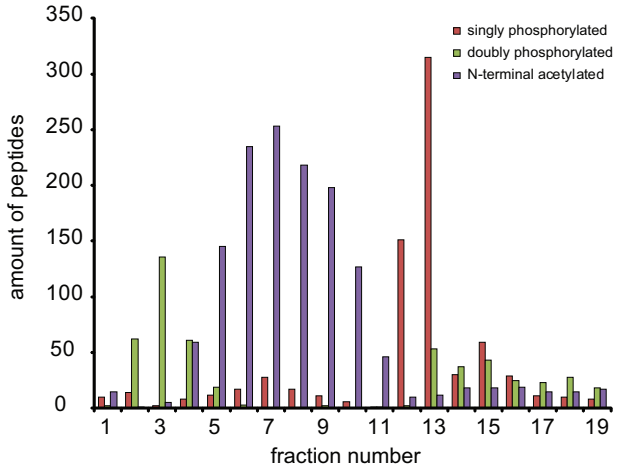



**Figure 7.** Distribution of singly (red) and doubly (green) phosphorylated and N-terminal acetylated (violet) peptides in the early fractions of an SCX separation at pH 3.[106, 115]

The enrichment of N-terminal acetylated peptides is also possible with SCX at a pH of 3.[106, 112-115] N-terminal acetylated peptides have a reduced net charge due to the neutralization of the basic N-terminus by acetylation. Thus, N-terminal acetylated peptides with only one basic amino acid can be separated in the same way, as described for the phosphorylated peptides. It is even possible to separate the N-terminal acetylated peptides from singly phosphorylated peptides, as depicted in Figure 7, even though they have the same net charge.[106]

One explanation is, that the orientation of the peptide also has an influence in the separation in SCX. Alpert et al. demonstrated that the carboxyl-group of the C-terminus interacts with the C-terminal basic residue of tryptic peptides and in consequence the N-terminal amino group has a higher exposed basicity. Thus, the N-terminus of regular and phosphorylated peptides is oriented towards the stationary phase.[116] In the case of N-terminal acetylated peptides, it can be assumed that the orientation changes, due to the loss of basicity at the N-terminus. Now the less basic C-terminus is orientated towards the stationary phase, which would be a good explanation why N-terminal acetylated peptides retain less compared to phosphorylated peptides (as depicted in Figure 7).

**Anion Exchange-RP**

The second form of IEX is anion exchange (AX). AX was also proven applicable for separating peptides already in the 1980s.[90] Anion exchange (AX)-RP has also been used for multidimensional approaches since the 1980s[117, 118] and was soon been automated and miniaturized.[119]

Inversely to CX, in AX the stationary phase contains positively charged residues that undergo Coulombic interactions with anionic sites of the analytes. Even though peptides have acidic residues the separation of peptides by anion exchange is less common. This might be due to the fact that theoretically over 29% of human tryptic peptides will have a neutral or basic net charge at a pH lower than 8.5 and thus, these peptides are not expected

to be retained by AX.[120] The reason for this are the high $pK_a$ values of the side chains of lysine and arginine, as given in Table 1.[86] Thus, when working at a pH of 8.5, lysine and arginine are still positively charged. If no additional acidic residues (besides the C-terminus) are present in the peptide the C-terminus is the only acidic residue present and the net charge of the peptide is zero or even positive when the peptide has missed cleavages. As a consequence these peptides do not bind in anion exchange.

Similarly to CX, AX materials are divided into strong anion exchange (SAX) and weak anion exchange (WAX) materials, where the functional group is a strong base in SAX, such as, trimethyl amine, while in WAX bases like diethylaminoethyl are used as functional groups. SAX is in general more popular, due to it being applicable also at high pH values, whereas WAX has a more limited pH range.

AX is a very useful tool for specific separation and enrichment of acidic peptides, like from acidic proteins from brain[118] or phosphorylated peptides.[49, 120-124] The phospho-group adds up to two negative charges to the peptide, depending on the pH. The additional negative charges causes the net charge of phosphopeptides (without missed cleavages) to be negative and as a consequence enables their retention on AX materials (Figure 8). Several phosphopeptide enrichment and separation methods have been developed that mainly apply SAX materials in combination with pH step[121] or gradient elution.[120, 122, 124] Due to the obvious bias of AX towards acidity, AX has been combined with SCX in various ways in order to complement each other.[48, 49, 122, 125] Dai and coworkers have used SCX as a first dimension in their offline approach. The elution was conducted by different pH steps ranging from a pH of 3 up to a pH of 10. The SCX flow through (pH 2.5) was further separated with SAX followed by RP of all SCX and SAX fractions. The parameters for each step were carefully chosen, as the solvent conditions for the SCX separation were chosen in a way that the flow through of the SCX is dominated by acidic peptides including most of the identified phosphopeptides and SAX is well suited to separate this pool of peptides. In this way they improved their SCX dimension by further separating a pool of low resolved peptides with SAX.[48] Recently, Hennrich et al. showed that weak anion exchange (WAX) at a low pH is especially suited to further separate phosphopeptides from early SCX fractions. In contrast to the approach from Dai et al. Hennrich et al. used a refined SCX which was conducted at a pH of 3 and applied a salt gradient for elution. This setup leads to some fractions enriched for singly phosphorylated peptides (Figure 7).[106] These two to three fractions contain thousands of phosphopeptides. One of these fractions was further separated by WAX in order to overcome issues of dynamic range and complexity. The WAX was conducted at a constant pH of approximately 3 and a salt gradient was applied for elution. Using this offline SCX-WAX-RP approach over 11,000 unique phosphopeptides could be identified only from this subpopulation of the total phosphoproteome.[49]

A complete online 3 dimensional system has been developed by Zhou et al. (Figure 8).[126] They combined RP at high pH with SAX and RP at low pH and compared it against RP-RP high-low pH. The use of SAX as intermediary step also allowed the solvent incompatibilities of the two RP steps to be circumvented. They observed an increased identification rate with RP-SAX-RP compared to RP-RP when analyzing complex

proteomes of *Escherichia coli* and *Saccharomyces cerevisiae*. In addition, they found that sampling 101 instead of 51 fractions to the final RP dimension, adds less than 5% protein and 6% unique peptide identifications.[126] This is very valuable information and should be tested when a multidimensional separation is first constructed, in order to identify the optimum compromise between experiment time and proteome coverage. In a follow up study the same group applied the RP-SAX-RP approach to phosphopeptide analysis. They achieved a peak capacity of over 3500 in 126.5 hours leading to approximately 12,000 phosphopeptide identifications.[127]

In addition AX-RP mixed mode columns have been applied for peptide separation. Mixed mode columns combine at least two different modes of separation in one column. In mixed mode columns two or more chemical groups with different binding specificities are combined in the uniform stationary phase.[128] Phillips and coworkers used a commercially available mixed mode column in a two dimensional setup. They applied the mixed mode column as first dimension and compared it to a standard SCX column. The mixed mode column combined an anion exchange (AX) and RP functionality and was used in an offline setup. As the separation with the RP/AX stationary phase is based on hydrophobic and electrostatic interactions, a gradient with an increase in acetonitrile (3%-80%) in conjunction with a decrease in buffer concentration (20 mM ammonium formate to 2 mM ammonium formate) and pH (6.5-2.5) was applied. They observed an increase in peptide separation efficiency and a more homogeneous distribution of tryptic peptides with the RP/AX column when compared to SCX. In addition, the RP/AX setup resulted in an increase of at least 50% on unique peptide identification.[128]

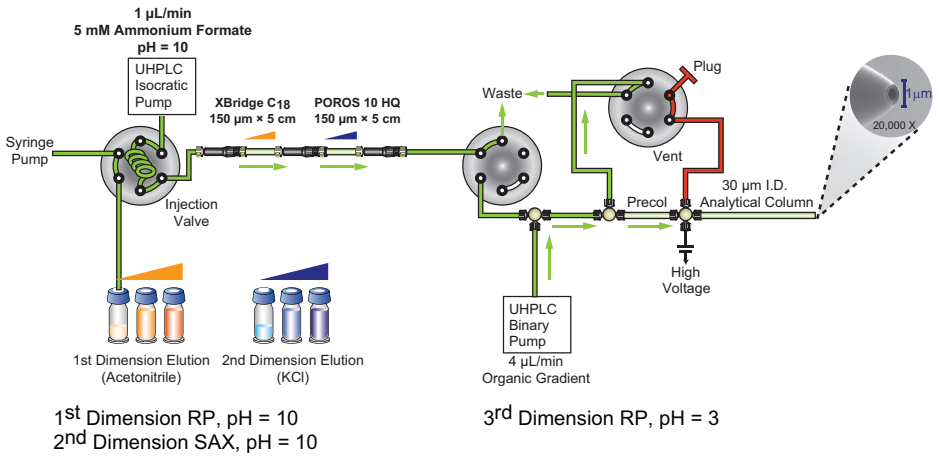



**Figure 8**: Schematic diagram of an automated, online nanoflow RP-SAX-RP system. The autosampler first loads the sample and is then used to inject the elution buffers for the first- (acetonitrile, orange) or the second- (KCl, blue) dimension, respectively. Second dimension fractions are diluted (4:1) and acidified with reversed phase solvent A (0.1% formic acid, 3% acetonitrile) introduced by an ultrahigh-pressure (UHP) binary pump. An additional six-port valve provides a vented third dimension column configuration and allows concentration of the peptides on the precolumn. An organic gradient is delivered by the UHP binary pump to elute peptides from the third dimension column for MS/MS analysis. Adapted from Zhou et al.[126]

## HYDROPHILIC INTERACTIONS LIQUID CHROMATOGRAPHY (HILIC)

The term HILIC was coined by Alpert in 1990[129] to describe normal-phase chromatography with water miscible mobile phases to separate hydrophilic compounds, such as proteins and peptides, but separations using these conditions were originally described by Linden et al. already in 1975, where a polar stationary phase and a relatively high water containing mobile phase were employed for carbohydrate separation.[130]

The exact separation mechanism in HILIC is still not fully understood but the hydrophilic partitioning model can be used to adequately rationalize many trends observed in analyte retention.[129] A sufficiently polar stationary phase and a typically low aqueous mobile phase (5-20 % water in ACN)[129] are employed in order to create a water-enriched liquid layer around the polar stationary phase. The separation is achieved by partitioning of analytes from the dynamic mobile phase into this stagnant aqueous layer around the stationary phase.[129, 131] The elution is then obtained by increasing the hydrophilicity of the mobile phase through increasing the water content. Factors governing the retention are both hydrogen bonding,[132] the extent of which depends on the acidity and basicity of the peptides, and dipole-dipole interactions, which depend on the dipole moments and the polarizability of the analytes.[129] In HILIC, the retention of peptides increases with increasing polarity or hydrophilicity of peptides and stationary phases, opposite to the trend observed in RP. However, the order of elution upon HILIC separation is not simply the reverse of RP, showing elution trends related to a set of physicochemical properties rather than simply hydrophilicity. Several papers and reviews have attempted to examine the variables involved in peptide retention under HILIC conditions,[131-134] but the exact retention mechanisms of HILIC are largely determined by the individual stationary and mobile phases in use, the type and concentration of salts and the pH.[135, 136]

The most common stationary phases applied for peptide separation in HILIC mode include underivatized silica that contain functional groups such as siloxane, silanols and a small quantity of metals, [137, 138] and derivatized silica, which can be neutral stationary phases, such as TSKgel Amide-80;[134] ionic stationary phases, such as the weak cation echanger Polycat A,[139] the cation exchanger polysulphoethyl A,[129] the weak anion exchanger PolyWAX;[123] zwitterionic stationary phases, such as ZIC-HILIC[140, 141] and ZIC-cHILIC.[53] Each of these materials is capable of generating a semi-immobilized aqueous layer on their polar surface.[129] However, different stationary and mobile phases display different retention characteristics, and the type of salt employed can also influence the retention behavior.[142]

When peptide separation is performed using a neutral hydrophilic TSKgel Amide-80 stationary phase, the primary interaction is hydrogen bonding and the retention is mostly based on the overall hydrophilicity of the analytes. [134, 143] The main advantage of this kind of stationary phase is the use of salt-free buffers, since the elution is obtained by solely increasing the polarity of the mobile phase (higher ratio water/organic solvent). Thus, this separation mode is directly compatibility with MS detection or any further step of separation in multidimensional strategies when HILIC is used as first step.

With ionic stationary phases, besides polarity, electrostatic interactions can also play an important role in increasing the selectivity for peptides,[129, 131] but the downside of those ionic interactions is the need of higher salt concentration in the buffer to disrupt these interactions and to obtain the elution,[137] which may cause dramatic ionization suppression and negative effects with MS detection. The choice of salts is however limited by their low solubility in highly organic buffers,[136] and ammonium formate and acetate are typically chosen for their better compatibility with MS. A novel method utilizing 'saltless' pH gradient with a WAX-HILIC chromatography was described for the automated on-line separation of hypermodified histone peptides, directly introduced via nanoelectrospray into the mass spectrometer, improving the ionization and sensitivity during the MS analysis.[144]

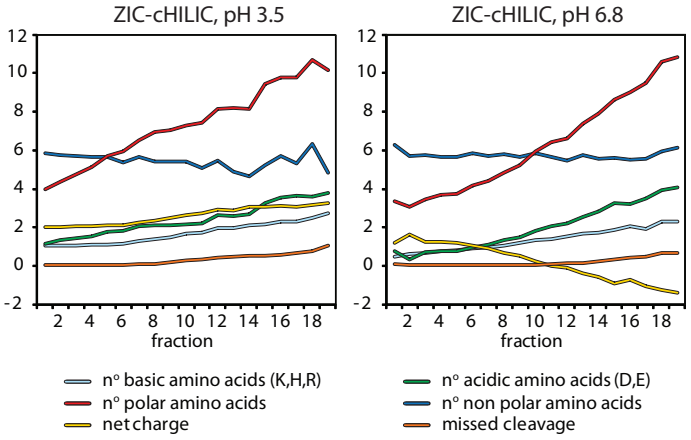



**Figure 9**: Analyses of peptide characteristics observed in ZIC-cHILIC fractions at a pH of 3.5 and 6.8 analyzing the effect of basicity, acidity, net charge, polarity, and number of missed cleavages on HILIC retention time. Adapted from Di Palma et al.[53]

When zwitterionic materials are employed, a good compromise between selectivity and compatibility can be achieved since peptide separation involves both hydrophilic and electrostatic interactions,[145] but the latter interactions are weaker compared to normal ionic exchangers and do not require high salt concentration for the elution.[131] An advantageous property of zwitterionic stationary phases relies on the amphoteric character of their functional moieties, containing two oppositely charged groups in close proximity at a stoichiometric ratio, the charges of which are permanent and unaffected by pH changes. Hence, the optimization of the mobile phase pH is solely dictated by the analytes charge and represents a potent factor to modulate the interactions between charged peptides and the zwitterionic stationary phase. Changing pH within a specific range influences the ionization of peptides due to protonation or deprotonation of certain amino acids, affecting their hydrophilicity and, consequently, their retention in HILIC mode.[51] Recently, two pH conditions, one neutral and one acidic, have been extensively evaluated with respect to their influence on peptides retention using a zwitterionic stationary phase (ZIC-cHILIC).[53] It was shown that ZIC-cHILIC selectivity at pH 6.8 was mostly driven by the peptide hydrophilicity, consistently with the hydrophilic partitioning model, whereas at pH 3.5 the retention was governed by a stronger mixed-mode mechanism based on balanced hydrophilic and electrostatic interactions (Figure 9).

**HILIC in multidimensional strategies**

HILIC is a relatively new player in the proteomics field, providing particular advantages in comparison to more established techniques. First, the high organic content of the buffer, together with the use of low salt concentration or volatile salts, makes HILIC highly compatible with ESI-MS, potentially increasing its sensitivity[142, 146, 147] and suggesting that HILIC can be an interesting final dimension.[148] Second, the ideal applicability for polar compounds allows the analysis of highly hydrophilic species that would be otherwise lost under RP analysis.[149] Third, the high orthogonality or complementarity to RP makes HILIC a suitable candidate in multidimensional approaches for the analysis of complex samples.[23, 51] All these reasons have contributed to an exponential increase in the application of HILIC chromatography in proteomics, and especially in multidimensional strategies. Furthermore, HILIC can represent an excellent alternative to SCX as a first dimension, providing higher resolution[53] and potentially allowing the enrichment of polar protein post translation modifications (PTMs),[52] such as glycosylation,[51, 150] phosphorylation[143, 151] and N-acetylation.[51]

The successful combination of HILIC with RP in two-dimensional systems relies on their opposite direction of polarity on peptide retention. Intuitively, it would seem that these two modes of separation act simply with reverse mechanisms. However, HILIC chromatography is strongly affected by other factors, such as the ion-pairing effect, composition of the mobile phase, pH, stationary phase, and more than solely polarity contributes to the analytes' retention.[136] In fact, Gilar et al.[23] have shown that HILIC has one of the highest degrees of orthogonality to RP of all commonly used peptide separation modes, such as SCX, RP at high pH and SEC. Boersema et al.[51] further confirmed the competitive level of orthogonality for HILIC when compared to SCX, though a different stationary phase was employed, showing less clusters of similarly charged peptides, as highlighted in Figure 10. This behavior was explained on the basis of the mixed-mode interaction with the zwitterionic stationary phase, where peptides are on one side retained by charge, due to electrostatic interactions, and on the other side by hydrophilic partitioning, allowing a larger spread in the distribution of peptides with similar charge.

Although aqueous-organic mobile phases of similar composition (a mixture of ACN and water) are employed for HILIC and RP, compatibility issues in 2D approaches would arise from differences in the elution strengths in the two modes. Mobile phases with high concentration of acetonitrile (ACN) employed with HILIC are strong eluents for RP; thus the hyphenation of HILIC with RP in on-line formats is not frequent as it would demand the use of larger second dimension columns in order to dilute the eluent from the first dimension. To overcome this problem, we[53] recently designed an off-line 2D system based on the combination of zwitterionic HILIC and RP in nanoscale format, which represents a refinement from an earlier set-up developed by Boersema et al.[51] In the new design, as schematically shown in figure 11, the HILIC eluent is directly collected during the

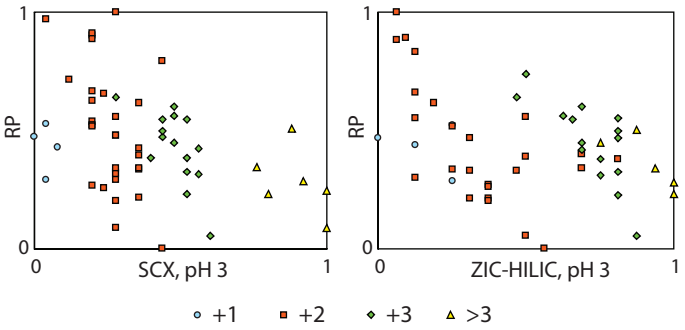



**Figure 10**: Normalized retention time plots for SCX-RP and ZIC-HILIC-RP. The net charge of the peptides are indicated with (circle) 1+, (square)2+, (rhomb) 3+, and (triangle) >3+. (Left panel) SCX pH 3 versus RP; (right panel) ZIC-HILIC pH 3 versus RP. Adapted from Boersema et al.[51]

separation as 1-min fractions into a 96 well plate already containing an acidified water solvent. Since the HILIC separation is performed at nanoliters per minute, the volume necessary to dilute the HILIC eluent is on the microliter scale and the fractions are sufficiently aqueous and large to avoid evaporation. This strategy could be also defined as in-line, as it is optimized to reduce sample handling, which cause sample loss and contamination, and to allow a direct analysis of the fractions with the subsequent RP-LC-MS/MS. This 2D-ZIC-HILIC-RP approach combines excellent resolution in both first and second dimension with minimal sample loss, and allows for a more sensitive proteome analysis in comparison to current methods that require more input material. In fact, the analysis can be performed on a mere few micrograms of starting material, achieving proteome coverage comparable to 'large scale' strategies. To probe the sensitivity of this strategy with a real low scale-amount sample, we[17] analyzed 10,000 FACS-sorted colon stem cells, directly after the extraction from the mouse intestine, allowing the identification of 15,775 unique peptides, originating from 3,775 proteins. A further comparison between the obtained proteome data and previous microarray experiments from the same colon stem cells confirmed the quality of the method, showing that 95% of the proteins detected in this study were also found to be expressed at the mRNA level.

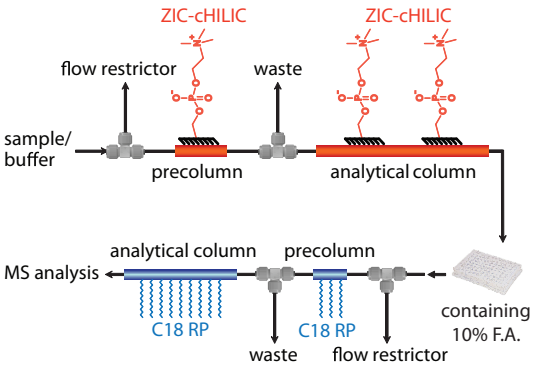
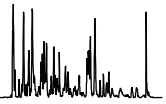



**Figure 11**. Schematic design of a 2 dimensional HILIC-RP approach.[17] The sample is concentrated on a ZIC-cHILIC precolumn and subsequently separated at nanoliter flowrates. One minute fractions are directly collected into a 96 well plate already containing an acidified water solvent. Subsequently, the fractions are directly analyzed by RP-LC-MS with an implemented RP precolumn for desalting and enrichment. Adapted from Di Palma et al.[53]

**HILIC for the analysis of PTMs**

Given the suitability for the separation of highly polar molecules, HILIC can be ideal in phosphoproteomics and glycoproteomics approaches, or in the targeted analysis of other PTMs such as acetylation, where it can represent an alternative enrichment technique or a chromatographic step in order to selectively reduce the sample complexity. For instance, Young et al. developed an effective on-line HILIC-MS method for the analysis of complex mixture of modified histone forms enabling to chromatographically distinguish isobaric modifications like trimethylation and acetylation and allowing the complete characterization of the human histones H3.2 and H4. [144] Since a general overview of HILIC applications for PTMs analysis was provided in an earlier review,[52] here we report solely the latest endeavours in implementing HILIC in the phosphoproteomics and glycoproteomics fields.

Phosphopeptides are in theory good candidates because of the increased overall hydrophilicity and their higher negative charge in comparison to regular peptides. McNulty et al.[143] evaluated the contribution of HILIC in phosphoproteomics showing that this chromatography can be an optimal first dimension in multidimensional systems for the targeted analysis of phosphopeptides. They compared the combinations of HILIC (using the TSKgel Amide-80 stationary phase) and IMAC in different orders (HILIC-IMAC versus IMAC-HILIC) to establish which approach would be more effective for the selective enrichment of phosphopeptides. Though approximately the same number of peptides were identified with the two approaches, in the IMAC-HILIC experiment only 60% of the total peptides were phosphorylated, while the HILIC-IMAC strategy showed a very selective phospho-enrichment (higher than 99%). However, phosphopeptides were evenly distributed throughout the HILIC fractions indicating that HILIC chromatography alone, employing TSKgel Amide-80 stationary phase, cannot be used as an efficient enrichment strategy, but represents an ideal fractionation prior to any more specific phospho-enrichment approach to improve the selectivity. We[53] investigated the use of two zwitterionic materials, ZIC-HILIC and ZIC-cHILIC, in phosphoproteomics. These two stationary phases exhibit an opposite charge spatial arrangement of their zwitterionic groups, with ZIC-HILIC possessing a negative charge as a distal moiety on the surface, while ZIC-cHILIC a positive charge. In theory, the more exposed positive charge on the surface of ZIC-cHILIC should electrostatically interact with the negatively charged phosphogroup of peptides, enhancing their binding and their retention. Since the phophopeptides were evenly spread all over the fractions in both cases, no obvious differences were found between the two zwitterionic materials, demonstrating that in this case the different charge arrangement on stationary phases did not affect the separation. In fact, the surface charge properties of this two materials have been studied through zeta-potential measurements showing that they both exhibit a negative surface charge at a wide range of pH, although their spatial arrangement should favor generating a positive charge.[141] This result confirmed that HILIC itself is not an efficient enrichment approach for phosphoproteomics, requiring the combination with



more specific strategy. We also believe that it may be possible to increase the phospho-selectivity by the addition in the mobile phase of hydrophobic ion-pairing reagents, such as TFA, in order to preferentially reduce the interactions of regular peptides with the stationary phase.

Due to the high complexity and heterogeneity of protein glycosylation, the analysis of this PTMs is at the present a challenging task. The investigation of complex glycopeptide pools and the insightful glycan structural elucidation require good chromatographic separations in combination with glycoenrichment strategies.[152] It is also common to enzymatically remove the glycans attached to the peptides or proteins prior to the MS analysis, thereby reducing the complexity and facilitating glycosylation site determination. Hägglund et al.[153] described an analytical method based on glycopeptide enrichment by HILIC (using microcolumns packed with ZIC-HILIC materials), followed by two different enzymatic deglycosylation strategies for a global proteome analysis of N- and O-glycosylation site from human plasma proteins. A modified protocol was reported by Thaysen-Andersen et al.[154] to extend this technique for site-specific characterization of low amounts of N-glycosilated proteins after immunoaffinity chromatography, followed by gel electrophoresis separation and in-gel digestion. The peptides obtained were further enriched for glycosylation using HILIC microcolumns, allowing a comprehensive glycoprofiling of the tissue inhibitor of metalloproteinases-1.

For glycopeptides separation, Takegawa et al. already highlighted the potential of HILIC.[155, 156] They showed that ZIC-HILIC columns allowed a high resolution separation of isomeric glycoforms based on structural recognition. Different glycopeptides containing neutral and sialylated N-glycans as well as released N-glycans were well separated.[155, 156] Wohlgemuth et al.[148] recently described a scheme for glycopeptides profiling based on a ZIC-HILIC pre-enrichment followed by a 2D RP-ZIC-cHILIC separation employing monolithic capillary columns in both dimensions, showing that this two chromatographic techniques are complementary for glycopeptide separation and their combination considerably improved the site-specific elucidation of glycans.

Mysling et al.[157] described an efficient glycoproteomic approach based on ZIC-HILIC separation on a microcolumn format. They explored the use of different mobile phases in order to improve the selectivity of the method and highlighted that the use of an ion-pairing reagent such as TFA in the mobile phase notably increased the efficiency of glycopeptide enrichment (see figure 12). The use of TFA in the mobile phase acts as an ion pairing agent for the analytes as well as for the stationary phase. The effect on the zwitterionic stationary phase is a higher contribution of hydrophilic partitioning in the HILIC retention since the electrostatic interactions are suppressed by the ion pairing reagent. The effect on the analytes will be a larger hydrophilicity difference between glycosylated and non-glycosylated peptides,[158] with a preferential decreasing in hydrophilicity for 'regular' peptides. In spite of that, this ion-pairing-HILIC approach is solely applicable in off-line format due to the presence of TFA that reduces the MS sensitivity and significantly lower the pH of the mobile phase, affecting the long term stability of ZIC-HILIC silica based resins.

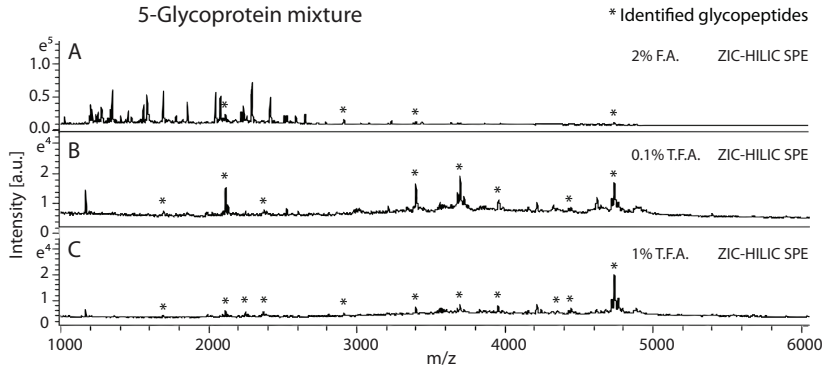


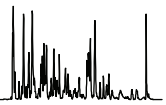

**Figure 12**: Glycopeptide enrichment from a peptide mixture derived from a tryptic digest of five glycoproteins (RNase B, ovalbumin, serotransferrin, fetuin, and R-1-acid glycoprotein). The digest was divided in three equal fractions and applied to IP2% FA, IP0.1% TFA, and IP1% TFA ZIC-HILIC SPE and the retained fractions analyzed using MALDI-TOF MS (A-C). Numerous signals in the low m/z region (m/z 1000-2500) were observed for the IP2% FA ZIC-HILIC SPE retained fraction, which most likely corresponded to nonglycosylated peptides. These analytes were depleted in the IP 0.1% TFA and IP 1% TFA ZIC-HILIC SPE retained fraction, and a number of signals in the higher m/z region were observed (m/z 3000-5000) corresponding to glycopeptides (assigned with asterisks (*)). Adapted from Mysling et al.[157]

## Alternative separation based on HILIC: ERLIC

A special variant of HILIC that departs from merely partitioning and exploits additionally superimposed electrostatic interactions is called ERLIC (electrostatic repulsion and hydrophilic interaction liquid chromatography).[123] An ion-exchange stationary phase is used with a highly organic mobile phase, similar to HILIC, in order to generate hydrophilic interactions with polar analytes. In this way, all peptides in a mixture would be retained through hydrophilic partitioning, despite charged peptides being repelled to some extent by similar charges present on the stationary phase. Since hydrophilic interactions, on one side, and electrostatic repulsions, on the other side, have opposite effects on peptide retention, an isocratic resolution of heterogeneous peptides can be achieved.

The ERLIC separation, based on anion exchange chromatography in combination with hydrophilic interactions, has shown good resolving power and high orthogonality to RP; thus, it is especially suitable for first dimensional fractionation of peptides in 2D-LC settings. Hao et al.[159] introduced an ERLIC-RP method for fractionating peptides based on both p*I* and polarity using a salt-free pH gradient of increasing water content. They compared this strategy to the traditional SCX-RP for the analysis of rat kidney tissue and showed that ERLIC outperformed SCX identifying 4821 proteins and 30 659 unique peptides. The dataset seemed to have a better representation of highly hydrophobic and basic peptides. They subsequently confirmed their results extending the strategy in a quantitative setting with iTRAQ labeled peptides.[159]

The ERLIC technique has been recently introduced as a potential phosphopeptide enrichment method using a WAX column[123] as an alternative to immobilized metal affinity



chromatography (IMAC) and Lewis acids, such as titania,[160] and zirconia.[161] Peptides containing phosphate groups maintain their negative charge even at pH values low enough to deprotonate acidic amino acids. This characteristic can be used to enhance the electrostatic interaction of negatively charged phosphopeptides with the positively charged ERLIC stationary phase, allowing their selective isolation from a peptide mixture. In addition, the retention is further improved by the use of a high concentration of organic solvents (ACN), promoting hydrophilic interactions of the polar phosphate group with the column. A salt and aqueous gradient is then used to gradually elute phosphopeptides.

A comparative phosphoproteomic study of ERLIC versus SCX-IMAC has been reported for the evaluation of isolation/enrichment of phosphopeptides showing that, unlike SCX-IMAC, the ERLIC approach achieved both sufficient enrichment of phosphopeptides and fractionation in one step, though SCX-IMAC outperformed ERLIC in the enrichment efficiency. The results revealed that only 12% of phosphopeptides were commonly identified in both strategies. The ERLIC approach enriched more selectively for multiply phosphorylated peptides with acidophilic motifs, while SCX-IMAC covered a wider variety of motifs, including acidophilic and basophilic. This study highlighted that the two methods are complementary to each other and should be used in combination to obtain a more comprehensive phosphoproteome coverage.[162] To provide a more comprehensive view of different fractionation techniques in phosphoproteome analyses, Zarei et al.[163] systematically compared 3 different chromatographic techniques, ERLIC, HILIC and SCX, coupled to $TiO_2$ as a specific phosphopeptide enrichment, evaluating the fractionation performance of each workflow based on the number of detected phosphopeptides, percentage of singly and multiply phosphorylated peptides, as well as their distribution over the applied gradients and number of incubation steps with $TiO_2$. On the basis of all three methods, more than 5,000 nonredundant phosphopeptides were identified after one step of incubation with $TiO_2$, of which 51.6% were detected only by SCX, 16.3% only by ERLIC, and 9.7% only by HILIC. The majority of phosphopeptides in SCX and HILIC were singly phosphorylated, 63% and 68%, respectively, whereas in ERLIC 69% were multiphosphorylated, confirming that ERLIC has a higher efficiency to bind and fractionate multiphosphorylated peptides versus SCX and HILIC. This can be attributed to the positively charged anion-exchange ERLIC stationary phase that binds multiphosphorylated peptides more strongly than monophosphorylated peptides. While the number of detected phosphopeptides in SCX-$TiO_2$ was 2 times higher than in ERLIC–$TiO_2$, the latter strategy can be applied as an optimal complementary method to SCX, as it proved to be especially suited for the fractionation of multiphosphorylated peptides. In contrast, a very low enrichment is observed by HILIC, as already shown in previous studies.[53, 143] One reason might be the broad distribution of nonphosphorylated peptides over the gradient compared to ERLIC and SCX where nonphosphorylated peptides can be removed at the front and at the end of the gradient, respectively.

Theoretically ERLIC can be applied for the enrichment of other types of protein post-translation modifications, as far as they display sufficient charges, either negative or positive. The same group reported an optimized ERLIC-based protocol for the

simultaneous enrichment of glyco- and phosphopeptides from mouse brain membrane, providing the opportunity to study the interaction of two different PTMs.[164] As phospho- and glycopeptides differ in their charge and hydrophilicity, their retention and elution profile would be different. Indeed, they found that phosphopeptides mainly eluted with an organic amount of approximately 70-60%, while the majority of glycopeptides did not elute until reaching a lower amount of organic solvent (30%). Thus, increasing the gradient slope with a wider range of organic solvent (from 70 to 25 % of acetonitrile) and optimizing the elution gradient, they allowed not only the enrichment of both phospho- and glycopeptides, but also a differential distribution of these two modified peptides in different fractions, whereas the phosphopeptides eluted earlier and the glycopeptides eluted mainly in later fractions (see figure 13). This approach resulted in the identification of 519 glycoproteins and 337 phosphorylation sites from a relatively small amount of sample (3 mg). In addition, the optimization of the sample preparation in combination with the efficient fractionation allows the identification of membrane phosphoproteins, which are particularly difficult to identify due to their intrinsic higher hydrophobicity and low abundance of phosphopeptides. An interesting finding was that more than 40% of these membrane proteins were simultaneously phosphorylated and glycosylated, demonstrating that these two PTMs can modulate protein function in a cooperative way.

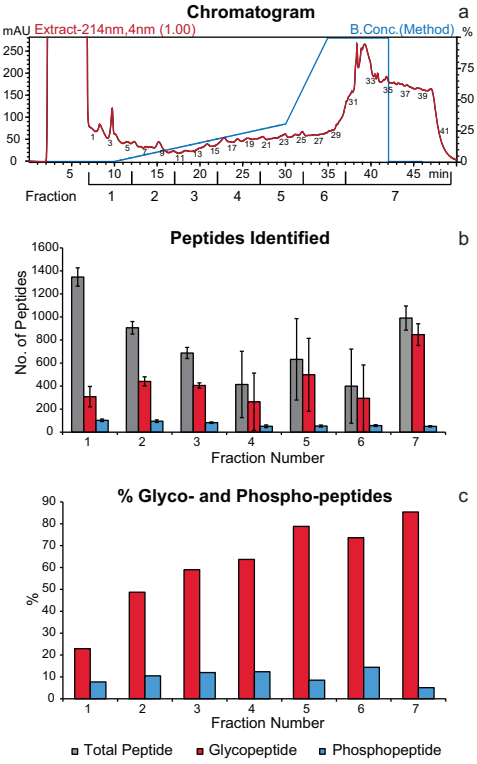
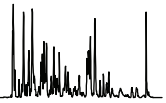



**Figure 13**. a, ERLIC chromatogram of 1mg of mouse brain tryptic digest. Forty fractions from 7 to 50 min were collected and then combined to seven final fractions as shown for b and c. Distribution of peptides in the seven collected fractions. b, average number of peptides, glycopeptides, and phosphopeptides identified in each fraction from six replicates. Error bars are the S.D. c, percentage of glyco- and phosphopeptides identified in each fraction. Adapted from Zhang et al.[164]



## CONCLUSIONS AND OUTLOOK

In conclusion, many different methods for peptide separation have been described in the last 30 years. Furthermore, all these separations are dramatically evolving. The question remains which approach is most suitable for an experiment. The choice is dependent on the analytical question, the available equipment, the amount of sample and analysis time available and the experience of the operators. Thus, it is not possible to suggest a general method for shotgun proteomics. Still, the following points need particular attention for creating a successful multidimensional setup:

- a high orthogonality between the individual dimensions
- a high peak capacity in each dimension
- maintaining the peak capacity of the early dimensions
- minimizing sample loss throughout the procedure

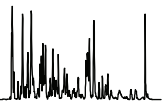
We believe that the multidimensional setups of the future will focus on optimization of these points in order to further increase the total peak capacity of coming methods. The simplest multidimensional system to construct is an RP-RP, with SCX-RP being not far behind. The major reason for this situation is that these two separations are incredibly well established with vendors providing excellent choices. However, in recent years the promise shown by ERLIC, HILIC and mixed bed material have raised awareness and so the choice of multidimensional strategies will expand and become less straightforward.
It should be mentioned that the majority of configurations discussed in this review will provide excellent results regarding proteome coverage and it is essentially the question of what sub proteome populations one wants to observe that will guide for a preferred strategy. Finally, massive improvements in the next few years are expected especially in the field of peptide separation because of availability of LC systems capable of operating at UPLC pressures and columns exploiting such equipment. We expect that the power of separations and the proteomics results from their use will continue to improve at an exponential rate.

## ACKNOWLEDGEMENTS

The authors acknowledge all members of the Heck-group. We'd also like to thank the groups of Dr. Martin Gilar, Dr. Jarrod Marto, Dr. John Yates, Dr. Peter Hojrup and Dr. Siu Kwan Sze for providing figures. This work was in parts supported by the PRIME-XS project, Grant Agreement Number 262067, funded by the European Union Seventh Framework Program; The Netherlands Proteomics Centre, embedded in The Netherlands Genomics Initiative, and the Centre for Biomedical Genetics; and The Netherlands Organization for Scientific Research (NWO) with the VIDI Grant 700.10.429 for SM.

## REFERENCES

(1) Chait, B. T. *Annual Review of Biochemistry* **2011**, *80*, 239-246.
(2) Yarmush, M. L.; Jayaraman, A. *Annual Review of Biomedical Engineering* **2002**, *4*, 349-373.
(3) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198-207.
(4) Yates, J. R.; Ruse, C. I.; Nakorchevsky, A. In *Annual Review of Biomedical Engineering*, 2009; Vol. 11, pp 49-79.
(5) Yates, J. R. *Annual Review of Biophysics and Biomolecular Structure* **2004**, *33*, 297-316.
(6) Shteynberg, D.; Deutsch, E. W.; Lam, H.; Eng, J. K.; Sun, Z.; Tasman, N.; Mendoza, L.; Moritz, R. L.; Aebersold, R.; Nesvizhskii, A. I. *Molecular & Cellular Proteomics* **2011**, *10*.
(7) McDonald, W. H.; Yates, J. R. *Current Opinion in Molecular Therapeutics* **2003**, *5*, 302-309.
(8) McDonald, W. H.; Yates, J. R. *Disease Markers* **2002**, *18*, 99-105.
(9) Reid, G. E.; McLuckey, S. A. *Journal of Mass Spectrometry* **2002**, *37*, 663-675.
(10) Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M.; Wu, C.; Sweet, S. M. M.; Early, B. P.; Siuti, N.; LeDuc, R. D.; Compton, P. D.; Thomas, P. M.; Kelleher, N. L. *Nature* **2011**, *480*, 254-258.
(11) Kellie, J. F.; Tran, J. C.; Lee, J. E.; Ahlf, D. R.; Thomas, H. M.; Ntai, I.; Catherman, A. D.; Durbin, K. R.; Zamdborg, L.; Vellaichamy, A.; Thomas, P. M.; Kelleher, N. L. *Molecular BioSystems* **2010**, *6*, 1532-1539.
(12) Cox, J.; Mann, M. In *Annual Review of Biochemistry, Vol 80*; Kornberg, R. D. R. C. R. H. R. J. E. T. J. W., Ed., 2011; Vol. 80, pp 273-299.
(13) Anderson, N. L.; Anderson, N. G. *Molecular & cellular proteomics : MCP* **2002**, *1*, 845-867.
(14) Tyers, M.; Mann, M. *Nature* **2003**, *422*, 193-197.
(15) Tolley, L.; Jorgensen, J. W.; Moseley, M. A. *Analytical Chemistry* **2001**, *73*, 2985-2991.
(16) Nagaraj, N.; Wisniewski, J. R.; Geiger, T.; Cox, J.; Kircher, M.; Kelso, J.; Paabo, S.; Mann, M. *Mol Syst Biol* **2011**, *7*.
(17) Di Palma, S.; Stange, D.; van de Wetering, M.; Clevers, H.; Heck, A. J. R.; Mohammed, S. *Journal of Proteome Research* **2011**, *10*, 3814-3819.
(18) Beck, M.; Schmidt, A.; Malmstroem, J.; Claassen, M.; Ori, A.; Szymborska, A.; Herzog, F.; Rinner, O.; Ellenberg, J.; Aebersold, R. *Mol Syst Biol* **2011**, *7*.
(19) Köcher, T.; Swart, R.; Mechtler, K. *Analytical Chemistry* **2011**, *83*, 2699-2704.
(20) Nagaraj, N.; Kulak, N. A.; Cox, J.; Neuhaus, N.; Mayr, K.; Hoerning, O.; Vorm, O.; Mann, M. *Molecular & Cellular Proteomics* **2011**.
(21) Fournier, M. L.; Gilmore, J. M.; Martin-Brown, S. A.; Washburn, M. P. *Chemical Reviews* **2007**, *107*, 3654-3686.
(22) Giddings, J. C. *Analytical Chemistry* **1984**, *56*, 1258-&.
(23) Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. *Anal Chem* **2005**, *77*, 6426-6434.
(24) Giddings, J. C. *Analytical Chemistry* **1967**, *39*, 1027-&.
(25) Giddings, J. C. *Journal of High Resolution Chromatography* **1987**, *10*, 319-323.
(26) Dixon, S. P.; Pitfield, I. D.; Perrett, D. *Biomedical Chromatography* **2006**, *20*, 508-529.
(27) Majors, R. E. *Journal of Chromatographic Science* **1980**, *18*, 393-393.
(28) Vollmer, M.; Horth, P.; Nagele, E. *Analytical Chemistry* **2004**, *76*, 5180-5185.
(29) Peng, J.; Elias, J. E.; Thoreen, C. C.; Licklider, L. J.; Gygi, S. P. *Journal of Proteome Research* **2003**, *2*, 43-50.
(30) Wagner, K.; Miliotis, T.; Marko-Varga, G.; Bischoff, R.; Unger, K. K. *Analytical Chemistry* **2002**, *74*, 809-820.
(31) Karas, M.; Hillenkamp, F. *Analytical Chemistry* **1988**, *60*, 2299-2301.
(32) Tanaka, K. *Angewandte Chemie-International Edition* **2003**, *42*, 3860-3870.
(33) Whitehouse, C. M.; Dreyer, R. N.; Yamashita, M.; Fenn, J. B. *Analytical Chemistry* **1985**, *57*, 675-679.
(34) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64-71.
(35) Wilm, M.; Mann, M. *Analytical Chemistry* **1996**, *68*, 1-8.
(36) Makarov, A.; Scigelova, M. *Journal of Chromatography A* **2010**, *1217*, 3938-3945.
(37) Domon, B.; Aebersold, R. *Science* **2006**, *312*, 212-217.
(38) Romijn, E. P.; Krijgsveld, J.; Heck, A. J. R. *Journal of Chromatography A* **2003**, *1000*, 589-608.
(39) 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. *Molecular & Cellular Proteomics* **2009**, *8*, 2759-2769.




(40) Andrews, G. L.; Simons, B. L.; Young, J. B.; Hawkridge, A. M.; Muddiman, D. C. *Analytical Chemistry* **2011**, *83*, 5442-5446.
(41) Shen, Y.; Smith, R. D. *Electrophoresis* **2002**, *23*, 3106-3124.
(42) Emmett, M. R.; Caprioli, R. M. *Journal of the American Society for Mass Spectrometry* **1994**, *5*, 605-613.
(43) Shen, Y.; Smith, R. D.; Unger, K. K.; Kumar, D.; Lubda, D. *Analytical Chemistry* **2005**, *77*, 6692-6701.
(44) Wolters, D. A.; Washburn, M. P.; Yates, J. R., 3rd *Anal Chem* **2001**, *73*, 5683-5690.
(45) Wang, N.; Xie, C.; Young, J. B.; Li, L. *Analytical Chemistry* **2009**, *81*, 1049-1060.
(46) Opiteck, G. J.; Jorgenson, J. W.; Anderegg, R. J. *Analytical Chemistry* **1997**, *69*, 2283-2291.
(47) Geng, M.; Ji, J.; Regnier, F. E. *Journal of chromatography. A* **2000**, *870*, 295-313.
(48) Dai, J.; Jin, W.-H.; Sheng, Q.-H.; Shieh, C.-H.; Wu, J.-R.; Zeng, R. *Journal of Proteome Research* **2006**, *6*, 250-262.
(49) Hennrich, M. L.; Groenewold, V.; Kops, G. J. P. L.; Heck, A. J. R.; Mohammed, S. *Analytical Chemistry* **2011**, *83*, 7137-7143.
(50) Motoyama, A.; Xu, T.; Ruse, C. I.; Wohlschlegel, J. A.; Yates, J. R., 3rd *Analytical Chemistry* **2007**, *79*, 3623-3634.
(51) Boersema, P. J.; Divecha, N.; Heck, A. J.; Mohammed, S. *J Proteome Res* **2007**, *6*, 937-946.
(52) Boersema, P. J.; Mohammed, S.; Heck, A. J. *Anal Bioanal Chem* **2008**, *391*, 151-159.
(53) Di Palma, S.; Boersema, P. J.; Heck, A. J. R.; Mohammed, S. *Analytical Chemistry* **2011**, *83*, 3440-3447.
(54) Gruber, K. A.; Stein, S.; Brink, L.; Radhakrishnan, A.; Udenfriend, S. *Proceedings of the National Academy of Sciences* **1976**, *73*, 1314-1318.
(55) Guo, D. C.; Mant, C. T.; Taneja, A. K.; Parker, J. M. R.; Hodges, R. S. *Journal of Chromatography* **1986**, *359*, 499-517.
(56) Chervet, J. P.; Ursem, M.; Salzmann, J. P. *Analytical Chemistry* **1996**, *68*, 1507-1512.
(57) Meiring, H. D.; van der Heeft, E.; ten Hove, G. J.; de Jong, A. P. J. M. *Journal of Separation Science* **2002**, *25*, 557-568.
(58) Schmidt, A.; Karas, M.; Dulcks, T. *Journal of the American Society for Mass Spectrometry* **2003**, *14*, 492-500.
(59) Shen, Y.; Zhang, R.; Moore, R. J.; Kim, J.; Metz, T. O.; Hixson, K. K.; Zhao, R.; Livesay, E. A.; Udseth, H. R.; Smith, R. D. *Analytical Chemistry* **2005**, *77*, 3090-3100.
(60) Wilkins, J. A.; Xiang, R.; Horvath, C. *Analytical Chemistry* **2002**, *74*, 3933-3941.
(61) Xiang, R.; Horvath, C.; Wilkins, J. A. *Analytical Chemistry* **2003**, *75*, 1819-1827.
(62) Ficarro, S. B.; Zhang, Y.; Lu, Y.; Moghimi, A. R.; Askenazi, M.; Hyatt, E.; Smith, E. D.; Boyer, L.; Schlaeger, T. M.; Luckey, C. J.; Marto, J. A. *Analytical Chemistry* **2009**, *81*, 3440-3447.
(63) Masuda, T.; Sugiyama, N.; Tomita, M.; Ishihama, Y. *Analytical Chemistry* **2011**, *83*, 7698-7703.
(64) Wang, X. L.; Stoll, D. R.; Carr, P. W.; Schoenmakers, P. J. *Journal of Chromatography A* **2006**, *1125*, 177-181.
(65) Dolan, J. W. *Journal of chromatography. A* **2002**, *965*, 195-205.
(66) Thakur, S. S.; Geiger, T.; Chatterjee, B.; Bandilla, P.; Frohlich, F.; Cox, J.; Mann, M. *Molecular & Cellular Proteomics* **2011**, *10*.
(67) Gilar, M.; Daly, A. E.; Kele, M.; Neue, U. D.; Gebler, J. C. *Journal of Chromatography A* **2004**, *1061*, 183-192.
(68) Jorgenson, J. W. *Annual Review of Analytical Chemistry* **2010**, *3*, 129-150.
(69) Motoyama, A.; Venable, J. D.; Ruse, C. I.; Yates, J. R. *Analytical Chemistry* **2006**, *78*, 5109-5118.
(70) MacNair, J. E.; Lewis, K. C.; Jorgenson, J. W. *Analytical Chemistry* **1997**, *69*, 983-989.
(71) MacNair, J. E.; Patel, K. D.; Jorgenson, J. W. *Analytical Chemistry* **1999**, *71*, 700-708.
(72) Hsu, J. L.; Huang, S. Y.; Chen, S. H. *Electrophoresis* **2006**, *27*, 3652-3660.
(73) Ballif, B. A.; Villen, J.; Beausoleil, S. A.; Schwartz, D.; Gygi, S. P. *Molecular & Cellular Proteomics* **2004**, *3*, 1093-1101.
(74) Raijmakers, R.; Berkers, C. R.; de Jong, A.; Ovaa, H.; Heck, A. J.; Mohammed, S. *Mol Cell Proteomics* **2008**, *7*, 1755-1762.
(75) Tanaka, N.; Kobayashi, H.; Nakanishi, K.; Minakuchi, H.; Ishizuka, N. *Analytical Chemistry* **2001**, *73*, 420 A-429 A.
(76) Luo, Q.; Shen, Y.; Hixson, K. K.; Zhao, R.; Yang, F.; Moore, R. J.; Mottaz, H. M.; Smith, R. D. *Analytical Chemistry* **2005**, *77*, 5028-5035.
(77) Luo, Q.; Tang, K.; Yang, F.; Elias, A.; Shen, Y.; Moore, R. J.; Zhao, R.; Hixson, K. K.; Rossie, S. S.; Smith, R. D. *Journal of Proteome Research* **2006**, *5*, 1091-1097.

(78) Miyamoto, K.; Hara, T.; Kobayashi, H.; Morisaka, H.; Tokuda, D.; Horie, K.; Koduki, K.; Makino, S.; Nunez, O.; Yang, C.; Kawabe, T.; Ikegami, T.; Takubo, H.; Ishihama, Y.; Tanaka, N. *Analytical Chemistry* **2008**, *80*, 8741-8750.
(79) Iwasaki, M.; Miwa, S.; Ikegami, T.; Tomita, M.; Tanaka, N.; Ishihama, Y. *Analytical Chemistry* **2010**, *82*, 2616-2620.
(80) Sandra, K.; Moshir, M.; D'hondt, F.; Tuytten, R.; Verleysen, K.; Kas, K.; François, I.; Sandra, P. *Journal of Chromatography B* **2009**, *877*, 1019-1039.
(81) Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. *Journal of separation science* **2005**, *28*, 1694-1703.
(82) Delmotte, N.; Lasaosa, M.; Tholey, A.; Heinzle, E.; Huber, C. G. *Journal of Proteome Research* **2007**, *6*, 4363-4373.
(83) Toll, H.; Oberacher, H.; Swart, R.; Huber, C. G. *Journal of chromatography. A* **2005**, *1079*, 274-286.
(84) Manadas, B.; English, J. A.; Wynne, K. J.; Cotter, D. R.; Dunn, M. J. *Proteomics* **2009**, *9*, 5194-5198.
(85) Nozaki, Y.; Tanford, C. In *Methods in Enzymology*; Hirs, C. H. W., Ed.; Academic Press, 1967; Vol. Volume 11, pp 715-734.
(86) Thurlkill, R. L.; Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. *Protein Science* **2006**, *15*, 1214-1218.
(87) Keim, P.; Vigna, R. A.; Morrow, J. S.; Marshall, R. C.; Gurd, F. R. N. *Journal of Biological Chemistry* **1973**, *248*, 7811-7818.
(88) Richarz, R.; Wuthrich, K. *Biopolymers* **1978**, *17*, 2133-2141.
(89) Song, C.; Ye, M.; Han, G.; Jiang, X.; Wang, F.; Yu, Z.; Chen, R.; Zou, H. *Analytical Chemistry* **2010**, *82*, 53-56.
(90) Dizdaroglu, M. *Journal of Chromatography* **1985**, *334*, 49-69.
(91) Isobe, T.; Takayasu, T.; Takai, N.; Okuyama, T. *Analytical Biochemistry* **1982**, *122*, 417-425.
(92) Cachia, P. J.; Vaneyk, J.; Chong, P. C. S.; Taneja, A.; Hodges, R. S. *Journal of Chromatography* **1983**, *266*, 651-659.
(93) Mant, C. T.; Hodges, R. S. *Journal of Chromatography* **1985**, *327*, 147-155.
(94) Alpert, A. J.; Andrews, P. C. *Journal of Chromatography* **1988**, *443*, 85-96.
(95) Takahashi, N.; Takahashi, Y.; Putnam, F. W. *Journal of Chromatography* **1983**, *266*, 511-522.
(96) Issaq, H. J.; Chan, K. C.; Janini, G. M.; Conrads, T. P.; Veenstra, T. D. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2005**, *817*, 35-47.
(97) Fränzel, B.; Wolters, D. A. *Proteomics* **2011**, *11*, 3651-3656.
(98) Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd *Nat Biotechnol* **2001**, *19*, 242-247.
(99) Bayer, E. M.; Bottrill, A. R.; Walshaw, J.; Vigouroux, M.; Naldrett, M. J.; Thomas, C. L.; Maule, A. J. *Proteomics* **2006**, *6*, 301-311.
(100) Van Hoof, D.; Munoz, J.; Braam, S. R.; Pinkse, M. W.; Linding, R.; Heck, A. J.; Mummery, C. L.; Krijgsveld, J. *Cell Stem Cell* **2009**, *5*, 214-226.
(101) Burke, T. W. L.; Mant, C. T.; Black, J. A.; Hodges, R. S. *Journal of Chromatography* **1989**, *476*, 377-389.
(102) McDonald, W. H.; Ohi, R.; Miyamoto, D. T.; Mitchison, T. J.; Yates, J. R. *International Journal of Mass Spectrometry* **2002**, *219*, 245-251.
(103) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R. *Nature Biotechnology* **1999**, *17*, 676-682.
(104) Davis, M. T.; Beierle, J.; Bures, E. T.; McGinley, M. D.; Mort, J.; Robinson, J. H.; Spahr, C. S.; Yu, W.; Luethy, R.; Patterson, S. D. *Journal of Chromatography B* **2001**, *752*, 281-291.
(105) Lim, K. B.; Kassel, D. B. *Analytical Biochemistry* **2006**, *354*, 213-219.
(106) Gauci, S.; Helbig, A. O.; Slijper, M.; Krijgsveld, J.; Heck, A. J.; Mohammed, S. *Analytical Chemistry* **2009**, *81*, 4493-4501.
(107) Gruhler, A.; Olsen, J. V.; Mohammed, S.; Mortensen, P.; Faergeman, N. J.; Mann, M.; Jensen, O. N. *Molecular & Cellular Proteomics* **2005**, *4*, 310-327.
(108) Villen, J.; Gygi, S. P. *Nature Protocols* **2008**, *3*, 1630-1638.
(109) Lemeer, S.; Pinkse, M. W. H.; Mohammed, S.; van Breukelen, B.; den Hertog, J.; Slijper, M.; Heck, A. J. R. *Journal of Proteome Research* **2008**, *7*, 1555-1564.
(110) Benschop, J. J.; Mohammed, S.; O'Flaherty, M.; Heck, A. J. R.; Slijper, M.; Menke, F. L. H. *Molecular & Cellular Proteomics* **2007**, *6*, 1198-1214.
(111) Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. *Cell* **2006**, *127*, 635-648.
(112) Mischerikow, N.; Heck, A. J. R. *Proteomics* **2011**, *11*, 571-589.
(113) Aivaliotis, M.; Gevaert, K.; Falb, M.; Tebbe, A.; Konstantinidis, K.; Bisle, B.; Klein, C.; Martens, L.; Staes, A.; Timmerman, E.; Van Damme, J.; Siedler, F.; Pfeiffer, F.; Vandekerckhove, J.; Oesterhelt, D. *Journal of Proteome Research* **2007**, *6*, 2195-2204.
(114) Dormeyer, W.; Mohammed, S.; van Breukelen, B.; Krijgsveld, J.; Heck, A. J. R. *Journal of Proteome Research* **2007**, *6*, 4634-4645.

(115) Helbig, A. O.; Gauci, S.; Raijmakers, R.; van Breukelen, B.; Slijper, M.; Mohammed, S.; Heck, A. J. R. *Molecular & Cellular Proteomics* **2010**, *9*, 928-939.

(116) Alpert, A. J.; Petritis, K.; Kangas, L.; Smith, R. D.; Mechtler, K.; Mitulovic, G.; Mohammed, S.; Heck, A. J. R. *Analytical Chemistry* **2010**, *82*, 5253-5259.

(117) Takahashi, N.; Ishioka, N.; Takahashi, Y.; Putnam, F. W. *Journal of Chromatography* **1985**, *326*, 407-418.

(118) Matsuoka, K.; Taoka, M.; Isobe, T.; Okuyama, T.; Kato, Y. *Journal of Chromatography* **1990**, *515*, 313-320.

(119) Holland, H. A.; Jorgenson, J. W. *Analytical Chemistry* **1995**, *67*, 3275-3283.

(120) Dai, J.; Wang, L. S.; Wu, Y. B.; Sheng, Q. H.; Wu, J. R.; Shieh, C. H.; Zeng, R. *Journal of Proteome Research* **2009**, *8*, 133-141.

(121) Nuhse, T. S.; Stensballe, A.; Jensen, O. N.; Peck, S. C. *Molecular & Cellular Proteomics* **2003**, *2*, 1234-1243.

(122) Nie, S.; Dai, J.; Ning, Z. B.; Cao, X. J.; Sheng, Q. H.; Zeng, R. *Journal of Proteome Research* **2010**, *9*, 4585-4594.

(123) Alpert, A. J. *Anal Chem* **2008**, *80*, 62-76.

(124) Han, G. H.; Ye, M. L.; Zhou, H. J.; Jiang, X. N.; Feng, S.; Jiang, X. G.; Tian, R. J.; Wan, D. F.; Zou, H. F.; Gu, J. R. *Proteomics* **2008**, *8*, 1346-1361.

(125) Wagner, Y.; Sickmann, A.; Meyer, H.; Daum, G. *Journal of the American Society for Mass Spectrometry* **2003**, *14*, 1003-1011.

(126) Zhou, F.; Sikorski, T. W.; Ficarro, S. B.; Webber, J. T.; Marto, J. A. *Analytical Chemistry* **2011**, *83*, 6996-7005.

(127) Ficarro, S. B.; Zhang, Y.; Carrasco-Alfonso, M. J.; Garg, B.; Adelmant, G.; Webber, J. T.; Luckey, C. J.; Marto, J. A. *Molecular & Cellular Proteomics* **2011**, *10*.

(128) Phillips, H. L.; Williamson, J. C.; van Elburg, K. A.; Snijders, A. P. L.; Wright, P. C.; Dickman, M. J. *Proteomics* **2010**, *10*, 2950-2960.

(129) Alpert, A. J. *J Chromatogr* **1990**, *499*, 177-196.

(130) Linden, J. C.; Lawhead, C. L. *Journal of Chromatography* **1975**, *105*, 125-133.

(131) Hemstrom, P.; Irgum, K. *J Sep Sci* **2006**, *29*, 1784-1821.

(132) Yoshida, T. *J Biochem Biophys Methods* **2004**, *60*, 265-280.

(133) Yoshida, T. *Analytical Chemistry* **1997**, *69*, 3038-3043.

(134) Yoshida, T. *Journal of Chromatography A* **1998**, *808*, 105-112.

(135) Guo, Y.; Gaiki, S. *Journal of Chromatography A* **2011**, *1218*, 5920-5938.

(136) Guo, Y.; Gaiki, S. *Journal of Chromatography A* **2005**, *1074*, 71-80.

(137) Jandera, P. *Analytica Chimica Acta* **2011**, *692*, 1-25.

(138) Jandera, P. *Journal of Separation Science* **2008**, *31*, 1421-1437.

(139) Lindner, H.; Sarg, B.; Meraner, C.; Helliger, W. *Journal of Chromatography A* **1996**, *743*, 137-144.

(140) Intoh, A.; Kurisaki, A.; Fukuda, H.; Asashima, M. *Biomedical Chromatography* **2009**, *23*, 607-614.

(141) Jiang, W.; Irgum, K. *Analytical Chemistry* **2002**, *74*, 4682-4687.

(142) Fountain, K. J.; Xu, J.; Diehl, D. M.; Morrison, D. *Journal of Separation Science* **2010**, *33*, 740-751.

(143) McNulty, D. E.; Annan, R. S. *Mol Cell Proteomics* **2008**, *7*, 971-980.

(144) Young, N. L.; DiMaggio, P. A.; Plazas-Mayorca, M. D.; Baliban, R. C.; Floudas, C. A.; Garcia, B. A. *Molecular & Cellular Proteomics* **2009**, *8*, 2266-2284.

(145) Di Palma, S.; Raijmakers, R.; Heck, A. J. R.; Mohammed, S. *Analytical Chemistry* **2011**, *83*, 8352-8356.

(146) Weng, N. *Journal of Chromatography B* **2003**, *796*, 209-224.

(147) Grumbach, E. S.; Diehl, D. M.; Neue, U. D. *Journal of Separation Science* **2008**, *31*, 1511-1518.

(148) Wohlgemuth, J.; Karas, M.; Jiang, W.; Hendriks, R.; Andrecht, S. *Journal of Separation Science* **2010**, *33*, 880-890.

(149) Xu, R. N.; Boyd, B.; Rieser, M. J.; Ei-Shourbagy, T. A. *Journal of Separation Science* **2007**, *30*, 2943-2949.

(150) Palmisano, G.; Lendal, S. E.; Engholm-Keller, K.; Leth-Larsen, R.; Parker, B. L.; Larsen, M. R. *Nature Protocols* **2010**, *5*, 1974-1982.

(151) Wu, C.-J.; Chen, Y.-W.; Tai, J.-H.; Chen, S.-H. *Journal of Proteome Research* **2011**, *10*, 1088-1097.

(152) Scott, N. E.; Parker, B. L.; Connolly, A. M.; Paulech, J.; Edwards, A. V. G.; Crossett, B.; Falconer, L.; Kolarich, D.; Djordjevic, S. P.; Højrup, P.; Packer, N. H.; Larsen, M. R.; Cordwell, S. J. *Molecular & Cellular Proteomics* **2011**, *10*.

(153) Hägglund, P.; Matthiesen, R.; Elortza, F.; Højrup, P.; Roepstorff, P.; Jensen, O. N.; Bunkenborg, J. *Journal of Proteome Research* **2007**, *6*, 3021-3031.

(154) Thaysen-Andersen, M.; Thøgersen, I. B.; Nielsen, H. J.; Lademann, U.; Brünner, N.; Enghild, J. J.; Højrup, P. *Molecular & Cellular Proteomics* **2007**, *6*, 638-647.

(155) Takegawa, Y.; Deguchi, K.; Keira, T.; Ito, H.; Nakagawa, H.; Nishimura, S. *Journal of Chromatography A* **2006**, *1113*, 177-181.
(156) Takegawa, Y.; Ito, H.; Keira, T.; Deguchi, K.; Nakagawa, H.; Nishimura, S.-I. *Journal of Separation Science* **2008**, *31*, 1585-1593.
(157) Mysling, S.; Palmisano, G.; Hojrup, P.; Thaysen-Andersen, M. *Analytical Chemistry* **2010**, *82*, 5598-5609.
(158) Ding, W.; Nothaft, H.; Szymanski, C. M.; Kelly, J. *Mol Cell Proteomics* **2009**, *8*, 2170-2185.
(159) Hao, P.; Qian, J.; Ren, Y.; Sze, S. K. *Journal of Proteome Research* **2011**, *10*, 5568-5574.
(160) Zhou, H.; Low, T. Y.; Hennrich, M. L.; van der Toorn, H.; Schwend, T.; Zou, H.; Mohammed, S.; Heck, A. J. R. *Molecular & Cellular Proteomics* **2011**, *10*.
(161) Kweon, H. K.; Håkansson, K. *Analytical Chemistry* **2006**, *78*, 1743-1749.
(162) Gan, C. S.; Guo, T.; Zhang, H.; Lim, S. K.; Sze, S. K. *Journal of Proteome Research* **2008**, *7*, 4869-4877.
(163) Zarei, M.; Sprenger, A.; Metzger, F.; Gretzmeier, C.; Dengjel, J. *Journal of Proteome Research* **2011**, *10*, 3474-3483.
(164) Zhang, H.; Guo, T.; Li, X.; Datta, A.; Park, J. E.; Yang, J.; Lim, S. K.; Tam, J. P.; Sze, S. K. *Molecular & Cellular Proteomics* **2010**, *9*, 635-647.

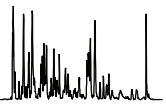
# Chapter 3

## Dimethyl Isotope Labeling Assisted de novo Peptide Sequencing

Marco L. Hennrich,†,‡ Shabaz Mohammed,†,‡ A. F. Maarten Altelaar,†,‡ and Albert J. R. Heck,†,‡

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

Published in the Journal of the American Society for Mass Spectrometry 2010; volume, 21 issue 12: page 1957-1965.




## ABSTRACT

Here, we explore a *de novo* sequencing strategy in which we combine Lys-N protein digestion with differential isotopic dimethyl labeling to facilitate the (*de novo*) identification of multiply charged peptides in ESI-MS, both under CID and ETD conditions. For a large fraction of the Lys-N generated peptides all primary amines are present at the N-terminal lysine, enabling specific labeling of the N-terminus. Differential derivatization of only the peptide N-terminus in combination with the simultaneous fragmentation of the corresponding isotopologues allows the straightforward distinction of N-terminal fragments from C-terminal and internal fragments. Furthermore, also singly and multiply charged N-terminal fragments can easily be distinguished due to the mass differences of the isotope labeled fragment pairs. As a proof of concept we applied this approach to proteins isolated from an avocado fruit, and were able to partially *de novo* sequence and correctly align, with green plant homologues, a previously uncharacterized avocado ascorbate peroxidase.

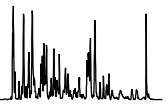

## INTRODUCTION

In mass spectrometry (MS) based proteomics, protein identification is typically accomplished by proteolytic digestion with trypsin followed by tandem MS (MS/MS) of the resultant proteolytic peptides.[1] The peptides are typically separated by nanoflow liquid chromatography and fragmented by collision induced dissociation (CID), followed by matching of the generated fragment spectra against *in silico* derived spectra from large protein sequence databases.[2-4] This conventional proteomics workflow reaches its boundaries when unexpected peptide modifications are present or when the required databases are either incomplete, error-prone, or not present at all. These boundaries can be overcome by *de novo* sequencing, where spectra are interpreted based on known fragmentation rules. However, the complexity of the obtained peptide fragmentation spectra poses a severe challenge in particular for *de novo* sequencing strategies[5, 6] since the unknown origin (N-, C-terminal, internal) and charge state of the observed fragments generally complicates correct peak annotation. Therefore, many chemical strategies to modify peptides have been introduced in order to simplify MS/MS spectra. The goal is, often, to make the fragmentation spectra consist of a single but complete ion series allowing easier interpretation via *de novo* sequencing approaches.[7-12]

Alternative to the above described chemical approaches we recently introduced a biochemical approach to manipulate the basicity of the peptide termini, using the metalloendopeptidase Lys- N.[13-15] This enzyme has cleavage specificity N-terminal of lysine residues,[16] creating peptides with a strong basicity on the N-terminus caused by the presence of the N-terminal lysine. Fragmentation of singly charged Lys-N peptides that contain a lysine as the only basic residue generates spectra dominated by N-terminal fragment ions. The simplification of the spectra is caused by the strong basicity of the N-

terminus, attracting the single proton. This feature is quite apparent in MALDI-CID spectra, wherein the fragment spectra are dominated by b-ions, as recently reported by Boersema et al..[13] Interestingly, doubly charged Lys-N peptides after ESI are also amenable to N-terminal directed dissociation,[14] after charge reduction and electron transfer induced dissociation (ETD).[17-19] We speculated that in single lysine containing Lys-N peptides the remaining single proton is directed by the strong basicity of the N-terminus, in a similar way as known for MALDI-CID, generating dominant, easy to interpret c-ion ladders in ESI-ETD.[14] This phenomena can be further enhanced by chemically increasing the basicity of the N-terminus by imidazolinylation or guanidination of the lysine side chain or dimethylation of the N-terminal primary amines.[20] Alternative strategies to facilitate *de novo* sequencing are based on differential labeling of the N- or C-terminus, providing mass signatures in MS/MS. Several of such methods have been described using specific labeling of the N- or C-terminus to create specific isotopic patterns in MS/MS spectra. An easy way of specific C-terminal labeling has been accomplished by proteolytic $^{18}O$ incorporation with the protease trypsin. Fragmentation of the $^{16}O$ and $^{18}O$ labeled peptides, either individually and/or simultaneously, enables the identification of fragments from the $^{16}O/^{18}O$ C-terminus and has been used for *de novo* sequencing.[21-24] Most of the N-terminal labeling strategies target primary amines, as several specific derivatization procedures are known.[10, 11, 25] These strategies are complicated by the fact that primary amines are not only present at the N-terminus but also on lysines. For specific N-terminal labeling two modification steps are needed, first the modification of the ε-amine of lysine at the protein level and then modification at the peptide level of the proteolytically formed N-termini. Examples are modification of lysine residues with succinic anhydride followed by digestion with Asp-C and N-terminal derivatization with 1-([H4/D4]-nicotinoyloxy)succinimide.[11] Other strategies use guanidination of lysine followed by N-terminal derivatization by reductive amination[26] or derivatization using sulfonate groups.[10, 27] The most recent developments in *de novo* sequencing strategies are often MS based, i.e. recording of MS/MS spectra with high resolution[28, 29] or applying complementary fragmentation techniques like CAD and ECD/ETD.[30, 31]

Here, we explore stable isotope dimethyl labeling in combination with Lys-N protein digestion in order to facilitate *de novo* sequencing. Since, all primary amines after Lys-N cleavage are clustered at the N-terminus, specific terminal labeling can be achieved in a single step by reductive amination. Reductive amination with formaldehyde and sodiumcyanoborohydride is facile, highly selective, allows near complete derivatization and has proven popular for quantitative proteomics.[25, 32-34] This derivatization introduces two methyl groups to each primary amine and is therefore often also called dimethylation or dimethyl labeling. Labeling of Lys-N peptides leads to the introduction of four methyl groups to the N-terminus of each Lys-N peptide. As described before,[20] an additional benefit of dimethyl labeling of Lys-N peptides is the increased basicity of the N-terminus. We make use of a specific differential isotopic dimethyl labeling of the N-terminus of Lys-N peptides, as this leads to a characteristic isotopic pattern on all N-terminal fragments of the MS/MS spectra. The isotopic pattern is observed for all Lys-N peptides, independent of charge state and in both CID and ETD fragmentation spectra. This strategy enables us to



both identify N-terminal fragments and to distinguish between singly and doubly charged N-terminal fragments even with low resolution MS/MS using a conventional ion trap for fragmentation.

## MATERIAL AND METHODS

**Materials.** Bovine Serum Albumin (BSA), iodoacetamide, formaldehyde (37% solution in $H_2O$), triethylammonium bicarbonate (1 M solution), 3-[(3-Cholamidopropyl)dimethylammonio]-1- propanesulfonate (CHAPS) and thiourea were supplied by Sigma-Aldrich (Steinheim, Germany). Formaldehyde 99% $^{13}C$ (20% solution in water) was purchased from Isotec (Miamisburg Ohio U.S.A.). Acetonitrile was supplied by Biosolve B.V. (Valkenswaard, The Netherlands). Ammonia (25% solution in water), acetone, urea and formic acid were obtained from Merck (Darmstadt, Germany). Ammonium bicarbonate, dithiothreitol (DTT), poly(vinylpolypyrrolidone) (PVPP) and sodium cyanoborohydrate were purchased from Fluka (Buchs, Switzerland). Precast Criterion XT gels, XT sample buffer 4x and XT MOPS running buffer 20x were purchased from Bio-Rad Laboratories (Hercules, CA, U.S.A.) and GelCode Blue Stain Reagent was purchased from Thermo Scientific (Rockford, IL, U.S.A.). The water used in these experiments was obtained from a Milli-Q purification system (Millipore, Bedford, MA, U.S.A.). Avocado (Persea americana Mill.) fruit cv. Ryan was purchased at the local store. Metalloendopeptidase Lys-N (Grifola frondosa) was obtained from Seikagaku Corporation (Tokyo, Japan).

**Protein extraction and digestion.** BSA was reduced and carbamidomethylated using DTT and iodoacetamide followed by digestion with Lys-N at 37°C overnight. A protein extract of an avocado fruit mesocarp was prepared as described previously.[35] Briefly, the mesocarp was ground to a fine powder in liquid nitrogen. The powder was suspended in lysis buffer and after lysis proteins were extracted by acetone precipitation following separation by SDS-PAGE. Two bands at approximately 26 kDa and 31 kDa were excised. After reduction and carbamidomethylation using DTT and iodoacetamide the proteins were digested in-gel with Lys-N at 37°C overnight.[14]

**Labeling of Peptides.** The samples were dried down, reconstituted in 200 µl triethylammonium bicarbonate and split in two. Either conventional or $^{13}C$ formaldehyde (4 µl of a 4 % solution) and 4 µl of 0.6 M cyanoborohydrate solution were added to 100 µL peptide solution and was left shaking for 1 h at room temperature. The reaction was terminated by addition of 16 µL 1 % ammonia solution.

**NanoLC-ESI-CID/ETD-MS/MS.** Peptides were subjected to nanoLC-MS/MS analysis, performed on an Agilent 1200 HPLC system (Agilent technologies, Waldbronn, Germany) connected to an Orbitrap XL Mass Spectrometer equipped with an ETD source, Thermo Fisher (Bremen, Germany). The instrument was equipped with a 20 mm x 100 µm i.d. Reprosil C18 trap column and a 400 mm x 50 µm i.d. Reprosil C18 RP analytical column (Dr Maisch, Ammerbuch-Entringen, Germany). Trapping and washing of the sample was performed at 5 µL/min for 10 min with 100 % solvent A, and elution was achieved during a 52 min gradient from 13 % to 32 % solvent B at 350 µL/min (with solvent B being 0.1 M

acetic acid in 80 % ACN and solvent A being 0.1 M acetic acid). The flow rate was passively split from 350 µL/min to 100 nL/min and the column effluent was directly introduced into the ESI source of the MS using an in-house pulled fused silica emitter, gold-coated by a Scancoat six Pirani 501 sputter coater (Edwards Laboratories, Crawley, UK), 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 automatic gain control (AGC) value of $5.00e^{+05}$ and a max injection time of 250 ms. Parent ions were fragmented by CID and ETD in data dependent mode with an AGC value of $1.00e^{+04}$ and a max. injection time of 100 ms. ETD fragmentation was performed with supplemental activation for doubly charged species, fluoranthene was used as reagent anion and ion/ion reaction time in the ion trap was charge state dependent with 50 ms reaction for 2+ ions, 33.3 ms for 3+ ions and 25 ms for 4+ ions.

**Data analysis.** *De novo* sequencing of peptides from Avocado fruit was performed manually. The obtained sequences were searched against green plants and the whole NCBI database using BLAST. Consensus sequences were identified between the peptides of the 26 kDa band and ascorbate peroxidases, chitinases, proteasome subunit alpha of different plants and autocleavage products of Lys-N. Subsequently, all sequenced peptides were aligned against ascorbate peroxidases from green plants using ClustalX 2.0.[36] BSA peptides were used as a test bed. MS/MS spectra were manually *de novo* sequenced and additionally converted to DTA files by Bioworks 3.3.1 followed by searching against a homemade BSA database using the Mascot search engine software version 2.2.0. with carbamidomethylcysteine as a fixed modification and light and heavy dimethylation and oxidized methionine as variable modification. Lys-N was used as protease and one missed cleavage was allowed.

## RESULTS

Here, we present a method in which we aim at the simplification of peptide fragmentation spectra using the protease Lys-N and a dimethyl isotope labeling strategy to introduce a mass signature in MS/MS. Our approach is based on the simultaneous fragmentation of isotopologues of differential isotopic N-terminal labeled Lys-N peptides.

In LC-MS/MS simultaneous fragmentation is only possible, when the corresponding differential labeled peptides are close in mass and exhibit indentical chromatographic retention behavior. Therefore, carbon isotopologues were chosen for differential isotopic dimethyl labeling, as they are known to have nearly identical behavior in reversed phase liquid chromatography.[37] The protein standard Bovine serum albumin (BSA) was used as a test sample to set up the method. The protein was digested with Lys-N and subsequently dimethyl labeled by either $^{12}C$ formaldehyde or $^{13}C$ formaldehyde in presence of cyanoborohydride. The peptides labeled with $^{12}C$ formaldehyde will be referred to as "light" peptides, whereas the peptides labeled with $^{13}C$ formaldehyde are further called "heavy" peptides. Extracted ion chromatograms of the monoisotopic mass of three corresponding light and heavy peptides from BSA are shown in Figure 1. As expected, the retention times and peak shapes of the corresponding peptides are identical. This confirms the identical behavior of the corresponding light and heavy peptides in our reversed phase

liquid chromatography system. On an ETD enabled Orbitrap XL mass spectrometer, simultaneous fragmentation was achieved by applying a broad precursor ion isolation width of 6 m/z units. This isolation width has been chosen after testing a precursor ion isolation width of 3, 4, 6 and 8 m/z units. Isolation widths smaller than 6 m/z units led to unequal intensities of the corresponding fragment pairs, while a greater isolation width was avoided in order to minimize potential unwanted co-fragmentation of co-eluting peptides. Applying an isolation window of 6 m/z, we observed that it is beneficial when the heavy peptide is chosen as precursor. When the light peptide was chosen, the heavy fragments were often observed with decreased intensities. Therefore, the amount of heavy peptides in the mixture was increased by 10% relative to the light peptides to enhance the possibility that the data dependent precursor selection software chooses the heavy peptide as precursor. This resulted in the clear observation of fragment ion pairs with differences of four (singly charged) and two m/z units (doubly charged).

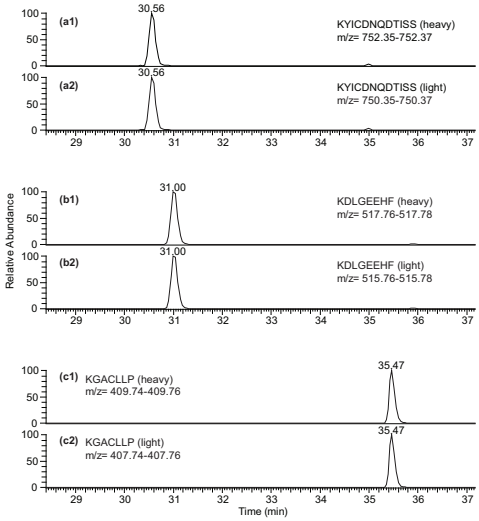



**Figure 1:** Extracted ion chromatograms of the monoisotopic precursors of three doubly charged peptides (a-c) from a Lys-N digestion of BSA. The peptides are either (1) heavy or (2) light dimethyl labeled at the N-terminus.

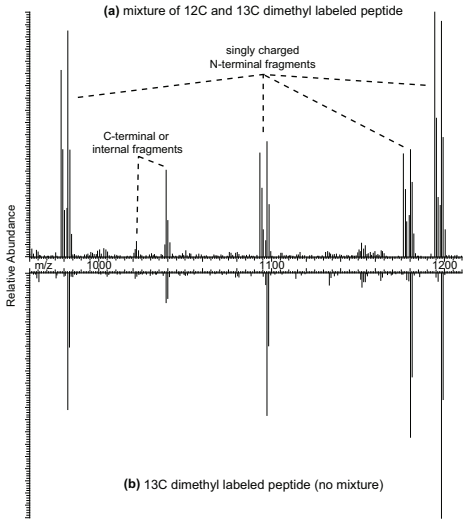



**Figure 2:** Comparison of a part of the CID fragmentation spectra of (a) a mixture of the light and heavy form and (b) only the heavy form of the doubly charged peptide KYICDNQDTISS from BSA.

In Figure 2 a zoomed in part of the CID spectrum resulting from the simultaneous fragmentation of a mixture of the light and heavy form of the doubly charged peptide KYICDNQDTISS from BSA (Fig. 2a) is compared with the same part of the spectrum of only the heavy form of the same peptide (Fig. 2b). C-terminal and internal fragments of the corresponding light and heavy peptide have identical masses and are observed as identical peaks. N-terminal fragments contain the isotope label and are therefore observed as fragment pairs, due to the light and heavy form of the fragment, as visualized in Figure 2. N-terminal fragments can be easily distinguished from others, facilitating annotation and spectral interpretation.

Next, we explored the potential of our method for the *de novo* sequencing of proteins from the avocado fruit for which currently no genome is available. The proteins from avocado fruit mesocarp were extracted and separated by SDS-PAGE. Extraction was carried out by acetone precipitation after grinding in liquid nitrogen to a powder and reconstitution in lysis buffer as described before.[35] It was observed, that addition of polyvinyl polypyrrolidone (PVPP) to the lysis buffer was crucial to remove polyphenols, and get sharp bands in SDS-PAGE. The extracted proteins were separated by SDS-PAGE and two gel bands, one at approximately 26 kDa and one at approximately 31 kDa were excised. The proteins were further processed using the workflow depicted in Figure 3. Briefly, the proteins were reduced, alkylated and digested in gel using Lys-N, followed by differential labeling. The light and heavy peptides are mixed in a 10:11 ratio respectively and the mixture is analyzed by LC-MS/MS using alternating CID and ETD for every precursor selected. Applying an isolation width of 6 m/z led to spectra resulting from the simultaneous fragmentation of light and heavy peptide.

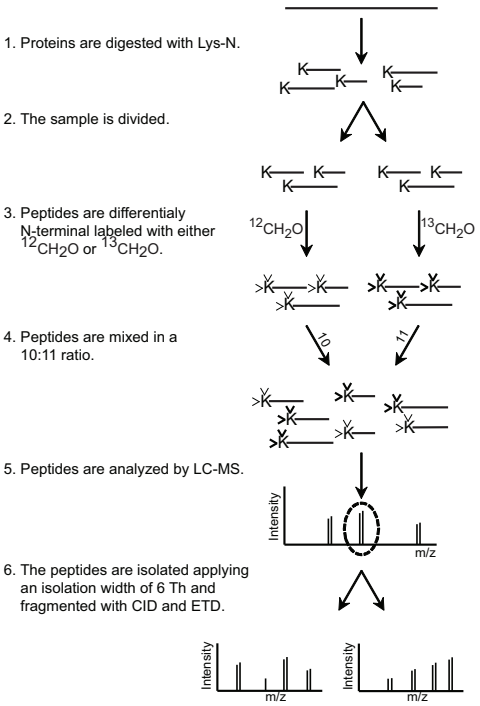
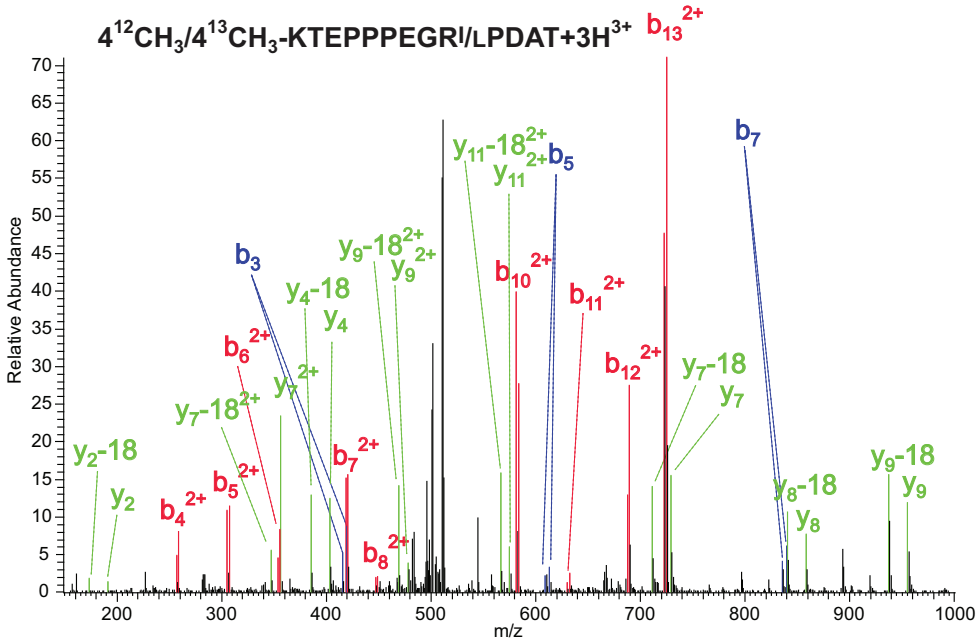



**Figure 3:** Schematic overview of the workflow. The proteins are first digested with Lys-N. The resulting peptides are divided in two, dimethyl labeled with either $^{12}C$ formaldehyde or $^{13}C$ formaldehyde and combined in a 10:11 ration. The mixture is analyzed by LC-MS and the peptide isotopologues are fragmented simultaneously using both CID and ETD for each precursor.

A typical CID spectrum obtained from one of the peptides originating from the avocado sample is depicted in Figure 4. Although this spectrum is quite complicated it could be manually interpreted using the unique signatures of the N-terminal fragments. Also, ion pairs with a mass difference of 4 m/z and 2 m/z could be easily assigned as singly and doubly charged N-terminal fragments, respectively. The complex spectrum in Figure 4 is dominated by doubly charged b-ions and near complete sequence coverage was obtained solely on the basis of N-terminal fragment ions. After identification of b-ions the masses of the corresponding singly and doubly charged y-ions as well as associated fragments resulting from water or ammonia loss could be calculated and assigned. All other single peaks were, subsequently, interpreted as C-terminal or internal fragments. Since, for every precursor selected for CID fragmentation an alternating ETD spectrum was recorded we used this complementary fragmentation data to increase the spectral annotation. Despite the presence of four proline residues in the sequence which led to a relatively poor ETD



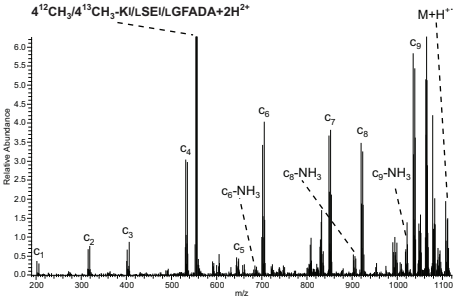

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $b_1$ | $b_2$ | $b_3$ | $b_4$ | $b_5$ | $b_6$ | $b_7$ | $b_8$ | $b_9$ | $b_{10}$ | $b_{11}$ | $b_{12}$ | $b_{13}$ | |
| $b_1^{2+}$ | $b_2^{2+}$ | $b_3^{2+}$ | $b_4^{2+}$ | $b_5^{2+}$ | $b_6^{2+}$ | $b_7^{2+}$ | $b_8^{2+}$ | $b_9^{2+}$ | $b_{10}^{2+}$ | $b_{11}^{2+}$ | $b_{12}^{2+}$ | $b_{13}^{2+}$ | |
| **K** | **T** | **E** | **P** | **P** | **P** | **E** | **G** | **R** | **I/L** | **P** | **D** | **A** | **T** |
| | $y_{13}$ | $y_{12}$ | $y_{11}$ | $y_{10}$ | $y_9$ | $y_8$ | $y_7$ | $y_6$ | $y_5$ | $y_4$ | $y_3$ | $y_2$ | $y_1$ |
| | $y_{13}^{2+}$ | $y_{12}^{2+}$ | $y_{11}^{2+}$ | $y_{10}^{2+}$ | $y_9^{2+}$ | $y_8^{2+}$ | $y_7^{2+}$ | $y_6^{2+}$ | $y_5^{2+}$ | $y_4^{2+}$ | $y_3^{2+}$ | $y_2^{2+}$ | $y_1^{2+}$ |

**Figure 4:** CID spectrum of simultaneous fragmentation of a mixture of light and heavy triply charged peptide KTEPPPEGR(I/L)PDAT from ascorbate peroxidase of avocado fruit with a table of the b- and y-ion fragments of the peptide. The colored b- and y-ions in the table have been identified in the spectrum.

spectrum,[17, 19] sufficient sequence information was observed to complete the sequence of the peptide. The two isobaric amino acids isoleucine and leucine could not be distinguished, whereby the sequence of the peptide could be identified as KTEPPPEGR(I/L)PDAT.

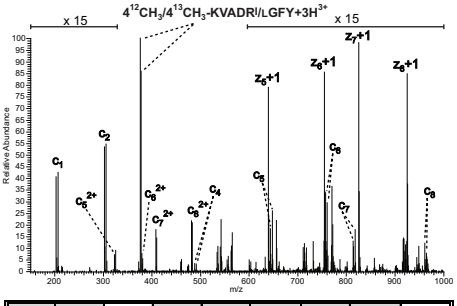
A typical ETD spectrum of the light and heavy form of a doubly charged avocado peptide generated by Lys-N is shown in Figure 5. Also, in this case, full sequence coverage could be achieved now primarily on the basis of the complete and dominant c-ion series, leading to the sequence K(I/L)SE(I/L)GFADA. Furthermore, the use of a dimethyl label at the N-terminus lead to a more complete series of c-ions.[20] Also c6, c8 and c9-ions with a loss of ammonia (formally b-ions) could be identified. ETD spectra of triply charged peptides were found to be more complex, as doubly charged fragments and z-ions are more abundant. Figure 6 is an example of such an ETD spectrum of a triply charged peptide from the digest of the 31 kDa band. Also, in-line with the CID data, this spectrum could be fully annotated on basis of the fragment pairs and the peptide was sequenced to completion, leading to the

sequence KVADR(I/L)GFY. Our approach directly revealed that c1, c2 and c4 are present as singly charged fragments, whereas c5-c8 are present as singly and doubly charged fragments, due to the presence of an internal arginine in the latter. The z-ions could be annotated on basis of the identified c-ions. The gap in the sequence because of the missing c3-ion was compensated by the corresponding z-ion being present. The result from the ETD spectrum was confirmed by the corresponding CID spectrum (data not shown) and the peptide sequence matches to a 32 kDa avocado endochitinase.[38] The presented examples clearly demonstrate that the introduced mass signatures facilitate annotation and consequently sequence determination. The mass difference of four Dalton between corresponding light and heavy peptides has the advantage that even with low resolution MS/MS, it is still possible to distinguish between singly and doubly charged fragments. Especially for annotation of fragments from triply and higher charged peptides this is a major advantage. All the obtained information was used to sequence the peptides and the sequence information of corresponding CID and ETD spectra were combined to complement one another in order to achieve the highest sequence coverage.




| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **$c_1$** | **$c_2$** | $c_3$ | **$c_4$** | $c_5$ | **$c_6$** | $c_7$ | **$c_8$** | **$c_9$** | |
| $c_1$-$NH_3$ | $c_2$-$NH_3$ | $c_3$-$NH_3$ | $c_4$-$NH_3$ | $c_5$-$NH_3$ | **$c_6$-$NH_3$** | $c_7$-$NH_3$ | **$c_8$-$NH_3$** | **$c_9$-$NH_3$** | |
| **K** | **I/L** | **S** | **E** | **I/L** | **G** | **F** | **A** | **D** | **A** |
| | $z_9$+1 | $z_8$+1 | $z_7$+1 | $z_6$+1 | $z_5$+1 | $z_4$+1 | $z_3$+1 | $z_2$+1 | $z_1$+1 |

**Figure 5:** ETD spectrum of simultaneous fragmentation of a mixture of light and heavy doubly charged peptide K(I/L)SE(I/L)GFADA from ascorbate peroxidase from avocado fruit with a table of the c-, c-NH3 and z+1-ion fragments of the peptide. The fragment ions depicted in bold on a grey background have been identified in the spectrum.




| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **$c_1$** | **$c_2$** | $c_3$ | **$c_4$** | $c_5$ | **$c_6$** | $c_7$ | **$c_8$** | |
| $c_1^{2+}$ | $c_2^{2+}$ | $c_3^{2+}$ | $c_4^{2+}$ | **$c_5^{2+}$** | **$c_6^{2+}$** | $c_7^{2+}$ | **$c_8^{2+}$** | |
| **K** | **V** | **A** | **D** | **R** | **I/L** | **G** | **F** | **Y** |
| | **$z_8$+1** | **$z_7$+1** | **$z_6$+1** | **$z_5$+1** | $z_4$+1 | $z_3$+1 | $z_2$+1 | $z_1$+1 |

**Figure 6:** ETD spectrum of simultaneous fragmentation of a mixture of light and heavy triply charged peptide KVADR(I/L)GFY from avocado fruit with a table of the c-, doubly charged c- and z+1-ion fragments of the peptide. The low and high mass range is 15 fold magnified as indicated. The fragment ions depicted in bold on a grey background have been identified in the spectrum.

After *de novo* sequencing a homology-search of the identified peptides from two bands, 26 kDa and 31 kDa, of avocado fruit was performed using NCBI-BLAST. The sequences were searched against the entire NCBI database and, in a second search, against the NCBI database limited to green plants. Out of the sequenced peptides from the 26 kDa band, five peptides were annotated that had high sequence homology or were even identical with stretches from ascorbate peroxidases of different plants. One of these five is the peptide

KTEPPPEGR(I/L)PDAT presented in Figure 4. The N-terminal amino acids of this peptide are in a low conserved region. Hence, this peptide cannot be identified in a standard database search, as the sequence is not present in any known ascorbate peroxidase in green plants. With our method it was possible to *de novo* sequence this peptide and identify it as belonging to an ascorbate peroxidase, by homology-searching. Clustal X[36] alignment of all *de novo* sequenced peptides from the 26 kDa band against ascorbate peroxidases from green plants, resulted in the alignment of two additional semi-sequenced peptides. The multiple sequence alignment of the total of these seven *de novo* sequenced peptides and cytosolic ascorbate peroxidases from papaya, cotton, maize, soybean, tobacco, and arabidopsis is presented in Figure 7.

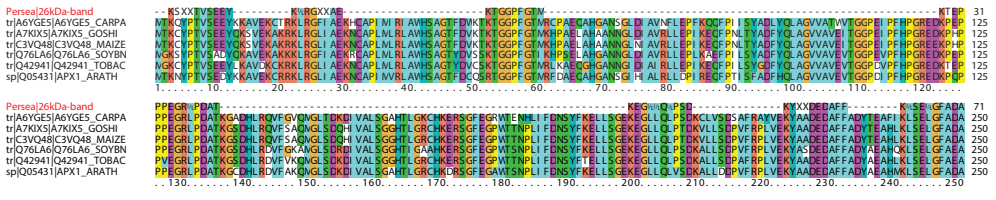



**Figure 7:** Multiple sequence alignment of *de novo* sequenced peptides from the 26 kDa band of avocado fruit and ascorbate peroxidase (a.p.) of papaya (CARPA), cytosolic a.p. 1 of cotton (GOSHI), a.p. of maize (MAIZE), cytosolic a.p. 2 of soybean (SOYBN), a.p. of tobacco (TOBAC) and cytosolic L-a.p. 1 of arabidopsis (ARATH).

The seven peptides that had homology with stretches of ascorbate peroxidases, match to cytosolic as well as to thylakoid bound and stromal ascorbate peroxidases. Most of the known cytosolic ascorbate peroxidases have a length of approximately 250 amino acids and their average weight is between 26 and 28 kDa. The protein in the gel band has an expected weight of approximately 26 kDa, which corresponds with cytosolic ascorbate peroxidases. Thylakoid bound and stromal ascorbate peroxidases are longer and none of the peptides align with the regions which are specific for thylakoid bound or stromal ascorbate peroxidases. Therefore, we conclude that the peptides from the 26 kDa band belong to an as yet undescribed cytosolic ascorbate peroxidase from avocado. The alignment gave further insights into the unidentified stretches of the sequenced peptides. The combined mass of the two unidentified amino acids (xx) in the first peptide with the sequence KSxxTVSEEY equals the mass of tyrosine plus proline, which is a conserved motif in cytosolic ascorbate peroxidases of green plants. Therefore, we suggest the sequence of this peptide to be KSYPTVSEEY. As mentioned above, the two isobaric amino acids leucine and isoleucine could not be distinguished with our method. However, in peptide 2 (K(I/L)RGxxAE) leucine at position 2 and isoleucine at position 6 are conserved and in the peptide 4, 5 and 7 leucine is dominant among the species. The unidentified amino acids at position 5 and 6 of peptide 2 have the mass of phenylalanine plus leucine/isoleucine. Based on the homology the suggested sequence of peptide 2 (K(I/L)RGxxAE) is KLRGFIAE, which fits the mass of the precusor. In peptide 6 (KYxxDEDAFF) the sum of the two unidentified amino acids equals the sum of alanine plus valine. This is an uncommon combination upon cytosolic

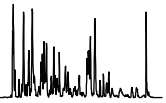
ascorbate peroxidases, but as hydrophobic amino acids are common at this position in other species such a variation is not exceptional.
In total, we were able to achieve a sequence coverage of over 23 % by *de novo* sequencing and identified the protein as a cytosolic ascorbate peroxidase from avocado fruit. Besides the seven peptides from ascorbate peroxidase, we *de novo* sequenced peptides from proteins in the 26 kDa band that have high sequence homology either to chitinases or to a proteasome subunit alpha. The abundant peptide KVADR(I/L)GFY from the 31 kDa band (Figure 6) had sequence homology with a peptide of a known and already described 32 kDa endochitinase from avocado.[38]

## CONCLUSION

Here, we have evaluated the use of Lys-N, differential isotopic dimethyl labeling and simultaneous peptide fragmentation of the corresponding light and heavy forms for *de novo* peptide sequencing. The resulting mass signature observed in the fragmentation spectra facilitates easy assignment of N-terminal fragments in both CID and ETD and allows distinguishing singly and doubly charged N-terminal fragments even with low resolution mass spectrometers. As a proof-of-concept, we were able to identify a previously unreported cytosolic ascorbate peroxidase of avocado fruit and resolve more than 23 % of the sequence of this protein. The presented method is uncomplicated, cheap and facilitates *de novo* sequencing.

## ACKNOWLEDGEMENTS

We like to acknowledge the Heck-group members and in particular Pieter B. T. Neerincx for bioinformatic support. The Netherlands Proteomics Centre, embedded in the Netherlands Genomics Initiative, is kindly acknowledged for financial support.



## REFERENCES


(1) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198-207.
(2) Domon, B.; Aebersold, R. *Science* **2006**, *312*, 212-217.
(3) Cravatt, B. F.; Simon, G. M.; Yates, J. R. *Nature* **2007**, *450*, 991-1000.
(4) Tang, J.; Gao, M. X.; Deng, C. H.; Zhang, X. M. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2008**, *866*, 123-132.
(5) Roepstorff, P.; Fohlman, J. *Biomedical Mass Spectrometry* **1984**, *11*, 601.
(6) Tang, X. J.; Thibault, P.; Boyd, R. K. *Analytical Chemistry* **1993**, *65*, 2824-2834.
(7) Spengler, B.; Luetzenkirchen, F.; Metzger, S.; Chaurand, P.; Kaufmann, R.; Jeffery, W.; Bartlet-Jones, M.; Pappin, D. J. C. *International Journal of Mass Spectrometry* **1997**, *169*, 127-140.
(8) Keough, T.; Youngquist, R. S.; Lacey, M. P. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 7131-7136.
(9) Brancia, F. L.; Oliver, S. G.; Gaskell, S. J. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 2070-2073.
(10) Keough, T.; Lacey, M. P.; Youngquist, R. S. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 2348-2356.
(11) Munchbach, M.; Quadroni, M.; Miotto, G.; James, P. *Analytical Chemistry* **2000**, *72*, 4047-4057.
(12) Standing, K. G. *Current Opinion in Structural Biology* **2003**, *13*, 595-601.
(13) Boersema, P. J.; Taouatas, N.; Altelaar, A. F. M.; Gouw, J. W.; Ross, P. L.; Pappin, D. J.; Heck, A. J. R.; Mohammed, S. *Mol Cell Proteomics* **2009**, *8*, 650-660.
(14) Taouatas, N.; Drugan, M. M.; Heck, A. J. R.; Mohammed, S. *Nature Methods* **2008**, *5*, 405-407.
(15) Taouatas, N.; Altelaar, A. F. M.; Drugan, M. M.; Helbig, A. O.; Mohammed, S.; Heck, A. J. R. *Molecular & Cellular Proteomics* **2009**, *8*, 190-200.
(16) Nonaka, T.; Hashimoto, Y.; Takio, K. *Journal of Biochemistry* **1998**, *124*, 157-162.
(17) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *Journal of the American Chemical Society* **1998**, *120*, 3265-3266.
(18) Zubarev, R. A. *Current Opinion in Biotechnology* **2004**, *15*, 12-16.
(19) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 9528-9533.
(20) Hennrich, M. L.; Boersema, P. J.; van den Toorn, H.; Mischerikow, N.; Heck, A. J. R.; Mohammed, S. *Analytical Chemistry* **2009**, *81*, 7814-7822.
(21) Takao, T.; Gonzalez, J.; Yoshidome, K.; Sato, K.; Asada, T.; Kammei, Y.; Shimonishi, Y. *Analytical Chemistry* **1993**, *65*, 2394-2399.
(22) Qin, J.; Herring, C. J.; Zhang, X. L. *Rapid Communications in Mass Spectrometry* **1998**, *12*, 209-216.
(23) Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. *Rapid Communications in Mass Spectrometry* **1997**, *11*, 1015-1024.
(24) Uttenweiler-Joseph, S.; Neubauer, G.; Christoforidis, A.; Zerial, M.; Wilm, M. *Proteomics* **2001**, *1*, 668-682.
(25) Hsu, J. L.; Huang, S. Y.; Chow, N. H.; Chen, S. H. *Analytical Chemistry* **2003**, *75*, 6843-6852.
(26) Ji, C. J.; Guo, N.; Li, L. *Journal of Proteome Research* **2005**, *4*, 2099-2108.
(27) Madsen, J. A.; Brodbelt, J. S. *Analytical Chemistry* **2009**, *81*, 3645-3653.
(28) Frank, A. M.; Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A.; Pevzner, P. A. *Journal of Proteome Research* **2007**, *6*, 114-123.
(29) 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. *Journal of Proteome Research* **2010**, *9*, 2713-2724.
(30) Savitski, M. M.; Nielsen, M. L.; Kjeldsen, F.; Zubarev, R. A. *Journal of Proteome Research* **2005**, *4*, 2348-2354.
(31) Bertsch, A.; Leinenbach, A.; Pervukhin, A.; Lubeck, M.; Hartmer, R.; Baessmann, C.; Elnakady, Y. A.; Muller, R.; Bocker, S.; Huber, C. G.; Kohlbacher, O. *Electrophoresis* **2009**, *30*, 3736-3747.
(32) Boersema, P. J.; Aye, T. T.; van Veen, T. A. B.; Heck, A. J. R.; Mohammed, S. *Proteomics* **2008**, *8*, 4624-4632.
(33) Raijmakers, R.; Berkers, C. R.; de Jong, A.; Ovaa, H.; Heck, A. J. R.; Mohammed, S. *Molecular & Cellular Proteomics* **2008**, *7*, 1755-1762.
(34) Boersema, P. J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J. R. *Nat. Protocols* **2009**, *4*, 484-494.
(35) Barraclough, D.; Obenland, D.; Laing, W.; Carroll, T. *Postharvest Biology and Technology* **2004**, *32*, 175-181.

(36) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. *Bioinformatics* **2007**, *23*, 2947-2948.
(37) Zhang, R. J.; Regnier, F. E. *Journal of Proteome Research* **2002**, *1*, 139-147.
(38) Sowka, S.; Hsieh, L. S.; Krebitz, M.; Akasawa, A.; Martin, B. M.; Starrett, D.; Peterbauer, C. K.; Scheiner, O.; Breiteneder, H. *Journal of Biological Chemistry* **1998**, *273*, 28091-28097.

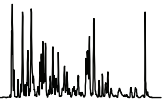
# Chapter 4

## Effect of Chemical Modifications on Peptide Fragmentation Behavior upon Electron Transfer Induced Dissociation

Marco L. Hennrich,†,‡ Paul J. Boersema,†,‡ Henk van den Toorn,†,‡ Nikolai Mischerikow,†,‡ Albert J. R. Heck,†,‡,§ and Shabaz Mohammed,†,‡

†Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
‡Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands
§ Centre for Biomedical Genetics

Published in Analytical Chemistry 2009; volume 81, issue 18: pages 7814-7822.




## ABSTRACT

In proteomics, proteolytic peptides are often chemically modified to improve MS analysis, peptide identification, and/or to enable protein/peptide quantification. It is known that such chemical modifications can alter peptide fragmentation in collision induced dissociation MS/MS. Here, we investigated the fragmentation behavior of such chemically modified peptides in MS/MS using the relatively new activation method electron transfer dissociation (ETD). We generated proteolytic peptides using the proteases Lys-N and trypsin and compared the fragmentation behavior of the unlabeled peptides with that of their chemically modified cognates. We investigated the effect of several commonly used modification reactions, namely, guanidination, dimethylation, imidazolinylation, and nicotinylation (ICPL). Of these guanidination and imidazolinylation specifically target the ε-amino groups of lysine residues in the peptides, whereas dimethylation and nicotinylation modify both N-termini and ε-amino groups of lysine residues. Dimethylation, guanidination, and particularly imidazolinylation of doubly charged Lys-N peptides resulted in a significant increase in peptide sequence coverage, resulting in more reliable peptide identification using ETD. This may be rationalized by the increased basicity and resulting positive charge at the N-termini of these peptides. Nicotinylation of the peptides, on the other hand, severely suppressed backbone fragmentation, hampering the use of this label in ETD based analysis. Doubly charged C-terminal lysine containing tryptic peptides also resulted in an enhanced observation of a single type of fragment ion series when guanidinated or imidazolinylated. These labels would thus facilitate the use of *de novo* sequencing strategies based on ETD for both arginine and lysine containing tryptic peptides. Since isotopic analogues of the labeling reagents applied in this work are commercially available, one can combine quantitation with improved ETD based peptide sequencing for both Lys-N and trypsin digested samples.

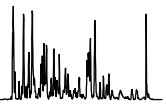

## INTRODUCTION

In recent years, mass spectrometry (MS) has become the method of choice for protein identification.[1] Currently, sequencing and identification is often performed with peptides with a molecular weight below 3 kDa. One of the main rationales for reducing a protein to peptides is the superior sequencing of peptides possible by mass spectrometry alongside an advanced armory of MS compatible peptide separation strategies.[2-4]

The protease most often used for digestion of proteins into peptides is trypsin. The tryptic peptides can then be separated by chromatography and subsequently subjected to sequencing by collision induced dissociation (CID) based tandem mass spectrometry (MS/MS). Currently, the most popular strategy to obtain sequence information from the spectra is by employing database search algorithms. These algorithms align experimental data against *in silico* derived spectra to obtain the most likely candidate sequence.[1, 5]

Typically under CID conditions tryptic peptide ions generate a number of different fragment ion series wherein N-terminal (b-ions) and C-terminal (y-ions) fragment ions dominate the tandem mass spectra.[6-8] However, these mixtures of different fragment ion series can create rather complex spectra that contain several layers of potentially redundant information. Reducing spectra to minimal information, ideally originating from a single but complete ion series, will allow easier interpretation either manually or by automated search algorithms. Over the years, a number of strategies have been developed to influence peptide ion fragmentation creating simplified spectra. Under normal MALDI and ESI conditions, peptide fragment ion formation and observation is ultimately linked to protons. The "mobile" proton weakens the peptide bonds, helps to induce bond cleavage, and allows observation of fragment ions.[9-12] Gas phase basicity of the various functional groups present in peptides and their fragments dictate the location of the proton(s) to a major extent. The average residence site of the proton(s) controls which fragment ions are observed. One strategy to create simplified spectra involves modifying functional groups on peptides. Changing the basicity or acidity will affect proton localization and thus the peptide fragmentation. When the protons can be sequestered at a single terminus of the peptide, a single ion series of sequence diagnostic fragment ions may be observed.[13] Most chemical modification strategies prefer MALDI since it generates peptides with only a single proton, allowing proton localization and thus fragmentation to be more easily controlled. Two commonly used methods that simplify spectra by removing the presence of N-terminal fragments are chemically assisted fragmentation (CAF)[14, 15] and 4-sulfophenyl isothiocyanate (SPITC).[14] In both these reactions, the N-terminal amino group is modified to a sulfonate derivative which is highly acidic/negatively charged. The sulfonate moiety leads to N-terminal fragments that are either negatively charged or neutral, eliminating their presence in the CID spectrum, which causes the fragmentation spectra to be dominated by C-terminal fragment ions (y-ions). However, the acidic nature of the N-terminus leads to a loss of sensitivity in positive ion mode since peptide ionization is compromised. Other examples of chemical modifications of tryptic peptides to increase ionization efficiency include the derivatization of C-terminal lysines to enhance the basicity/positive charge of the C-terminus. In this case, C-terminal fragments (y-ions) are preferably formed upon CID fragmentation.[16-20] However, the increased basicity of the peptide has also a side effect; the collision energy needs to be increased for efficient fragmentation. The increased collision energy often leads to "overfragmentation" and when insufficient energy is applied low-energy pathways such as cleavage adjacent to aspartates and glutamates dominate.[21, 22] The results of the N- and C-terminal labeling methods described above are complementary as they both lead to the increased formation of C-terminal fragment ions, and thus it is not surprising that the two strategies have also been combined.[14, 23]

Recently, we showed that fragmentation behavior of peptide ions cannot only be chemically regulated but also biochemically using the proteolytic enzyme Lys-N which generates peptides that exhibit two basic entities at the N-terminus.[24-27] This concentration of basicity directs the protons/charge to the N-terminus and leads to fragmentation spectra dominated by b-ions in MALDI-CID.[25] We also showed that using ESI-MS in combination with



electron transfer dissociation (ETD) for the doubly charged Lys-N peptides that contain solely a single basic residue give rise to simple fragment ion spectra dominated by N-terminal c-ion fragments.[26, 27]

At present, little is known about the fragmentation behavior of chemically modified peptides following ETD. Moreover, only a few chemical labeling strategies have been investigated in combination with ETD for quantification (e.g., using the isobaric tags, isobaric tag for relative and absolute quantitation (iTRAQ) and tandem mass tags (TMT)),[28-30] but these reports focus on the fragmentation of the label and not on the changed behavior of the bound peptide. Furthermore, both labels were compromised in their ability to allow quantitation with ETD based tandem mass spectrometry. We set out to investigate the effect of other commonly used chemical labels on the fragmentation behavior of Lys-N and tryptic peptides under ETD conditions. The four commonly used peptide modifications we examined are guanidination,[17-19] dimethylation,[31-33] imidazolinylation (Lys Tag 4H),[20] and nicotinylation (ICPL).[34, 35] Since these labels can also be made with differential stable isotopes, they have been previously applied in quantitative proteomics applications, albeit by using CID for fragmentation.

## EXPERIMENTAL SECTION

**Materials.** Bovine serum albumin, iodoacetamide, formaldehyde (37% solution in $H_2O$), and triethylammonium bicarbonate (1 M solution) were supplied by Sigma-Aldrich (Steinheim, Germany). Ammonia (25% solution in $H_2O$) and formic acid were obtained from Merck (Darmstadt, Germany). Ammonium bicarbonate, dithiothreitol (DTT), sodium cyanoborohydrate, and *O*-methyl isourea hemisulfate were purchased from Fluka (Buchs, Switzerland). Protease inhibitor cocktail was supplied by Roche Diagnostics (Mannheim, Germany). The water used in these experiments was obtained from a Milli-Q purification system (Millipore, Bedford, MA). Human embryonic kidney (HEK) 293 cells were provided by Dr. Pantelis Hatzis and Dr. Tokameh Mahmoudi from The Netherlands Institute for Developmental Biology, (Utrecht, The Netherlands). Sep-Pak Vac tC18 1 $cm^3$ cartridges were obtained from Waters Corporation (Milford, MA). Nicotine N-hydroxy succinimide ester (Nic-NHS ester) in DMSO was a gift from Dr. Keidel, Max-Plank-Institute for Biochemistry (Martinsried, Germany). 2-Methoxy-4,5-dihydro-1H-imidazole (Lys Tag 4H) was a gift from Agilent Technologies. Metalloendopeptidase Lys-N (Grifola Frondosa) was obtained from Seikagaku Corporation (Tokyo, Japan).

**Cell Lysing and Digestion of Proteins.** A protein extract of a HEK293 cell lysate was prepared as described previously.[26] The protein concentration was estimated by a Bradford assay. HEK293 proteins and bovine serum albumin (BSA) were reduced and carbamidomethylated using DTT and iodoacetamide. Subsequently, the proteins were digested with Lys-N or trypsin at 37°C overnight.

**Strong Cation Exchange (SCX) Chromatography.** The HEK293 peptide solution was desalted prior to SCX chromatography using a C18 Sep-Pak cartridge (Waters Corporation, Milford, MA), dried in a vacuum centrifuge, and resuspended in 10% formic acid. SCX

was performed on a SCX system as described previously.[26] Collected fractions were desalted and dried in a vacuum centrifuge.

**Labeling of Peptides.** For guanidination a 5 M solution of *O*-methyl isourea hemisulfate in water was prepared immediately before usage. Peptides (1-10 µg) were resuspended in 30 µL of 1 M $Na_2CO_3$ and mixed with 6 µL of 5 M *O*-methyl isourea hemisulfate. A pH of 10.5-11 was estimated by the use of indicator strips. The solution was incubated for 30 min at 65°C. The reaction was terminated by adding 14 µL of 50% formic acid.

For imidazolinylation, a 2 M solution of 2-methoxy-4,5-dihydro-1H-imidazole in water was prepared immediately before usage. Peptides (1-10 µg) were resuspended in 10 µL of 1 M $Na_2CO_3$ and mixed with 45 µL of 2 M 2-methoxy-4,5-dihydro-1H-imidazole. A pH of 11-12 was estimated by the use of indicator strips. The solution was incubated for 3 h at 55°C. The reaction was terminated by addition of 5 µL of 50% formic acid.

For nicotinylation, 25 µL of a 33 mg/mL nicotine N-hydroxy succinimide ester in DMSO solution was added to 1-10 µg of peptides resuspended in 5 µL of 50% MeCN. This mixture was left at room temperature for 2 h. Additional 10 µL of nicotine N-hydroxy succinimide ester solution was added after 15, 30, and 45 min.

For dimethylation, 1 µg of peptides was dissolved in 100 µL of triethylammonium bicarbonate. Formaldehyde (4 µL of a 4% solution) and 4 µL of 0.6 M cyanoborohydrate solution were added to the peptides, and the solution was left shaking for 1 h at room temperature. The reaction was terminated by addition of 1% ammonia solution.

**NanoLC-ESI-CID/ETD-MS/MS.** The peptides were subjected to NanoLC-MS/MS analysis with CID and ETD or ETcaD for doubly charged peptides. The NanoLC-MS/MS consists of a HCT ultra PTM Discovery system (Bruker Daltonics) equipped with an Agilent 1200 system HPLC. The columns were made in-house. The material used for the precolumn is Aqua C18 (Phenomenex, Torrance, CA). Reprosil-pur C18 3 µm (Dr. Maisch, Ammerbuch-Entringen, Germany) was used for the 25 cm analytical capillary column with 50 µm inner diameter. Mobile-phase buffer A consisted of 0.1 M acetic acid in water, and mobile-phase buffer B consisted of 0.1 M acetic acid in 80% acetonitrile.

**Protein Identification.** Raw MS data were converted to peak lists using Data Analysis software (Bruker Daltonics, version 4.0) Spectra were searched against the IPI (International Protein Index) Human database version 3.36 (69 012 sequences; 29 002 682 residues) or a homemade BSA database (8 sequences; 2973 residues) using Mascot software version 2.2.0 (www. matrixscience.com), with Lys-N or trypsin 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 ±1.2 Da, allowing 1 missed cleavage, carbamidomethyl (C) as fixed modification, oxidation (M), and the appropriate chemical labeling as a variable modification. Tandem mass spectra assigned with a Mascot score ≥25 ($p$-value ≤0.05) were accepted.

**Statistical Analysis.** The sequence coverage for each peptide from the BSA digests was calculated by dividing the observed number of ions for a particular fragment ion series by the maximum theoretical number of observable ions. The differences in sequence coverage between labeled and unlabeled peptides were determined, and the average and the standard deviation of these differences were calculated.

The number of ions for the individual fragment ion series of the peptides from the HEK 293T cell lysate digests were calculated and exported from Mascot. Also, for the HEK 293T cell lysate the differences in the number of ions between labeled peptides and their unlabeled cognates were determined and the average and standard deviation of the differences were calculated. *p*-values were calculated based on the number of compared peptides and the corresponding average and standard deviation using a two sample one-sided *t* test. In all cases, differences were considered to be significant at *p*-values < 0.05.

## RESULTS AND DISCUSSION

Electron transfer dissociation of doubly charged peptides leads to the generation of an entire population of product species that are singly charged.[36] The generation of this singly charged population has been exploited by our group through the use of Lys-N in order to predominantly produce a single type of fragment ion series in ETD. The underlying principle for creating these simplified spectra is the manipulation of basicity where Lys-N creates peptides with all amino groups placed at the N-terminus. We observed that creating such peptides lead to preferentially N-terminal fragment ions. To carry out a more in-depth analysis of the effect of basicity on peptide fragmentation we modulated the basicity of the amino groups.

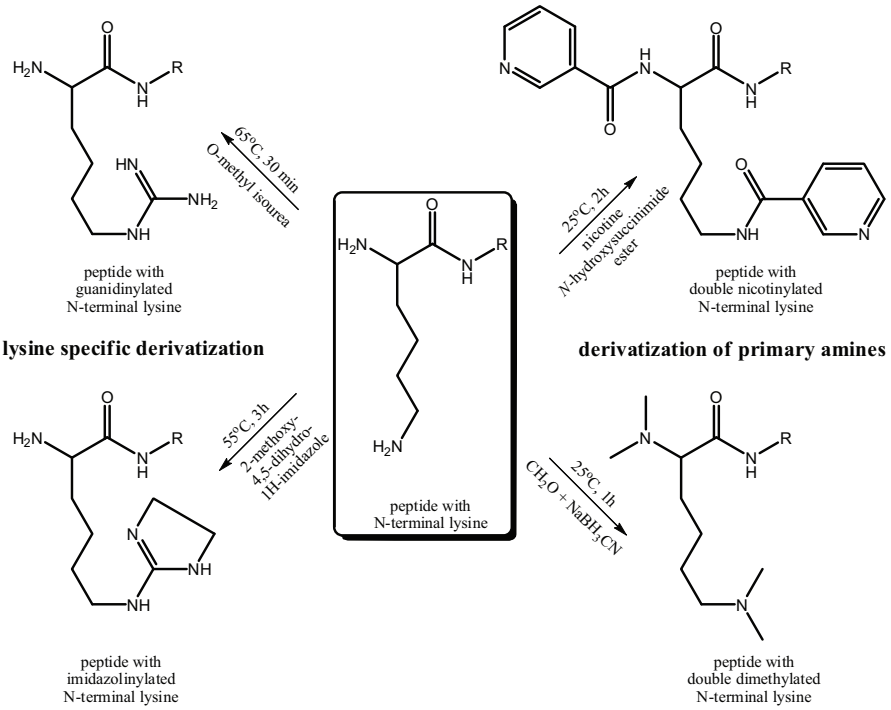



**Figure 1.** Overview illustrating the lysine (guanidination and imidazolinylation) and amino group (nicotinylation and dimethylation) specific peptide derivatization chemistries.

A number of labeling strategies exist within proteomics for modifying amino groups.[13, 17, 20, 31, 32, 34, 35, 37] For specific derivatization of the ε-amine of lysine, we utilized the two reagents: *O*-methyl isourea[17] and 2-methoxy-4,5-dihydro-1H-imidazole[20] (Figure 1). The latter reagent creates a moiety with a higher basicity than the original ε-amino and the guanidino group produced by *O*-methyl isourea. Additionally, two further labeling strategies were characterized for derivatization of all primary amines, nicotine N-hydroxy succinimide ester[34, 35] and formaldehyde in the presence of cyanoborohydrate[31, 32, 38, 39] (Figure 1). For both of these reagents the local basicity of the modified residues is also enhanced upon derivatization.

In order to provide a platform for the labeling comparison, a test bed of peptides was generated by digesting bovine serum albumin (BSA) with either trypsin or Lys-N. In the case of Lys-N, the lysine residue would be present at the N-terminus and thus the labels would exclusively modify this region of the peptide. In the cases of guanidination and imidazolinylation, a single label will be placed on the lysine. Dimethylation and nicotinylation will modify both amino groups and thus two labels will be placed at the N-terminus (Figure 1). Tryptic peptides will possess an α-amino group at the N-terminus and potentially an ε-amino group (lysine) at the C-terminus. Guanidination and imidazolinylation will label only a subset of tryptic peptides exclusively at the C-terminal lysine. Dimethylation and nicotinylation will label all tryptic peptides on the N-terminus, and those containing a C-terminal lysine will also be labeled at the C-terminus.

Proteolytic peptides of BSA were generated by trypsin or Lys-N, and aliquots were labeled using the four different reagents. The labeled peptides and their unlabeled counterparts were analyzed by CID and ETD on a three-dimensional quadrupole ion trap. Furthermore, doubly charged peptide ions were subjected to automated

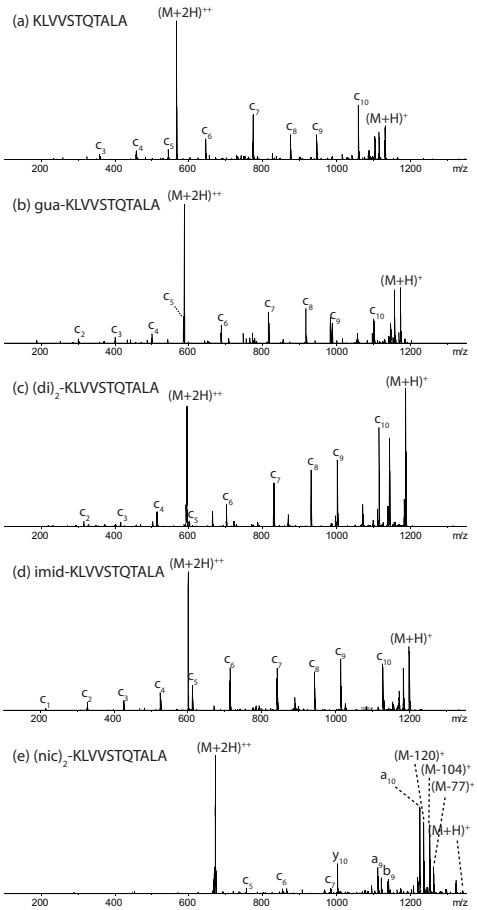



**Figure 2.** Comparison of ETcaD fragmentation spectra of the doubly charged Lys-N peptide KLVVSTQTALA derived from bovine serum albumin (BSA): (a) unmodified, (b) guanidinated (gua), (c) double dimethylated (di), (d) imidazolinylated (imid), and (e) double nicotinylated (nic).

supplemental collisional activation (ETcaD) in ETD experiments.[40] As previously reported and mentioned above, ETD spectra of unlabeled doubly charged Lys-N peptides containing one basic amino acid are dominated by c-type fragment ions.[27] The intensities in this series of fragment ions are greatest in the high mass range and decrease with mass. Typical examples of such spectra are shown in Figure 2. The peptide modification reactions chosen for this study (guanidination, dimethylation, nicotinylation, and imidazolinylation) are all known to increase basicity of either or both α- and ε-amino groups present in a peptide. Consequently, the gas phase basicity driven effect of observing solely c-type fragments is still present. Nicotinylated peptides showed nonsequence informative ETcaD fragmentation, which will be discussed in detail below.

The main difference between the spectra of unmodified and guanidinated, dimethylated, and imidazolinylated Lys-N peptides was found in the relative intensities for the c-ions observed. All three labels, albeit to different extents, reduced the generally observed signal drop from high to low mass for the c-ion series. Strikingly, additional low mass c-ions could now be observed in the majority of spectra of guanidinated, dimethylated, and imidazolinylated peptides when compared to their unlabeled cognate. The appearance of low mass c-ions was most pronounced for imidazolinylated peptides relative to guanidinated and dimethylated, correlating well with the gas phase basicity of these derivatives (Figure 2). Performing a manual fragment-ion count based on the spectra annotated by scaffold for the doubly charged peptides from the Lys-N digest of BSA and their labeled counterparts revealed that the sequence coverage of the guanidinated and imidazolinylated peptides is significantly greater in the modified than in the unlabeled peptides (Figure 3). On the basis of a two sample one-sided *t* test, the probabilities for the observed increase in c-ions generated by guanidination and imidazolinylation were $p < 0.02$ and $p < 0.001$, respectively (Supplemental Table 1 in the Supporting Information).

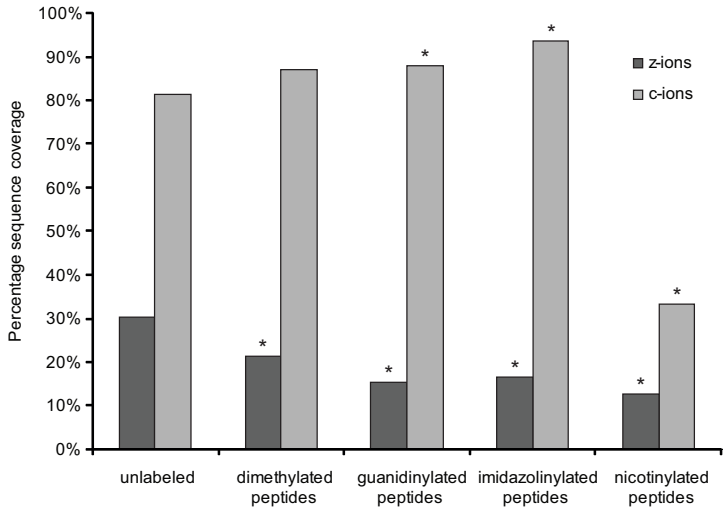



**Figure 3.** Average percentage of peptide sequence coverage based on c- and z-ions found in spectra of peptides resulting from a Lys-N digest of BSA and their labeled counterparts. All these BSA peptides (n = 8) contain a single lysine residue at the N-terminus. An asterisk indicates significant ($p < 0.05$) increase in c-ions and decrease in z-ions, respectively, between labeled peptides and their unlabeled cognates.

Next we also evaluated the effect of the different labels on the ETD induced fragmentation of tryptic peptides (Figure 4). As expected, doubly charged tryptic peptides with a single lysine residue resulted in MS/MS spectra containing a mixture of c- and z-ions. The observed distribution of c- and z-ions can be rationalized by the location of the basic moieties at each of the two termini of these peptides. ETD induced fragmentation of peptides containing a C-terminal arginine resulted in dominant z-ion series reflecting the higher basicity of this residue versus the primary amine present at the N-terminus (see the Supporting Information).[27, 41-43]

The ETD spectra of the guanidinated or imidazolinylated tryptic peptides became generally more dominated by z-ions. This may be rationalized as *O*-methyl isourea or 2-methoxy-4,5-dihydro-1H-imidazole only labels the ε-amino group of lysines, increasing the basicity at the C-terminus of lysine containing tryptic peptides. The reduction in signal observed with fragment ion mass for the z-ion series of the labeled peptides is abated with respect to unmodified peptides (Figure 4) in line with the trends observed for Lys-N peptides. The relative intensities of c-ions against z-ions are lower compared to their unlabeled cognates providing further evidence that basicity is directing product ion detection in ETD spectra of doubly charged ions. In contrast to the other two labels, dimethylation of tryptic peptides showed no significant difference compared to unmodified peptides (Supplemental Table 1 in the Supporting Information). Dimethylation modifies both α- and ε-amino groups, which in the case of lysine containing tryptic peptides, will mean that both the N- and C-terminus are derivatized. Performing a manual fragment ion-count for tryptic BSA peptides with a C-terminal lysine revealed that the guanidinated and imidazolinylated peptides had a statistically significantly greater peptide sequence

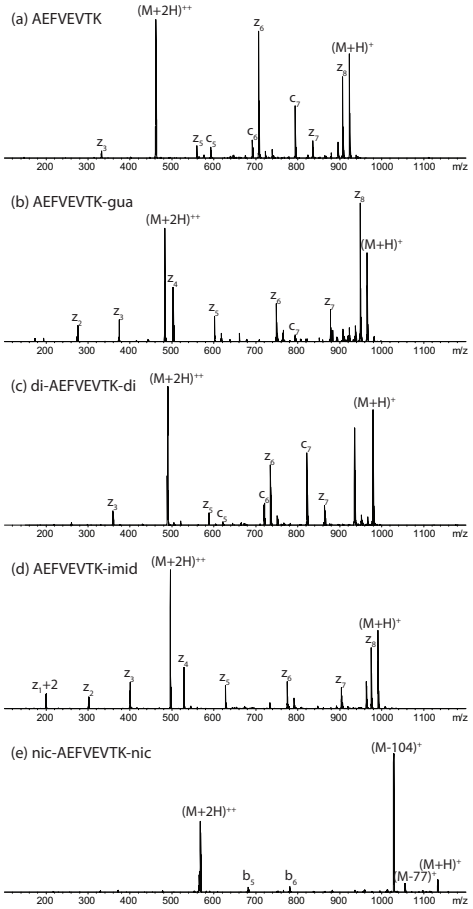



**Figure 4.** Comparison of ETcaD fragmentation spectra of the doubly charged tryptic peptide AEFVEVTK derived from BSA: (a) unmodified, (b) C-terminal guanidinated (gua), (c) N- and C-terminal dimethylated (di), (d) C-terminal imidazolinylated (imid), and (e) N- and C-terminal nicotinylated (nic).

coverage based on only z-ions compared to the unlabeled counterpart (Figure 5), each with *p*-values < 0.03 (Supplemental Table 1 in the Supporting Information). Strikingly, the imidazolinylation of lysine containing tryptic peptides generated ETD spectra containing extensive z-ion series (i.e., few c-ions) that may facilitate *de novo* sequencing strategies based on ETD (Figure 5 and Supplemental Table 1 in the Supporting Information).

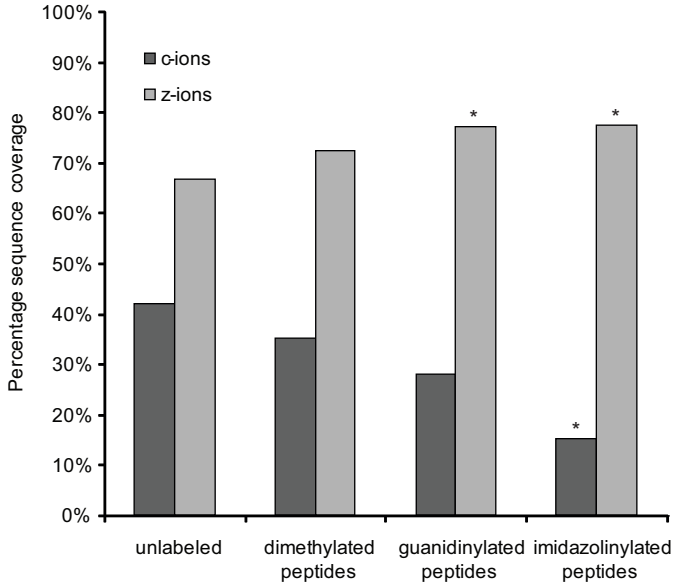



**Figure 5.** Average percentage of peptide sequence coverage based on c- and z-ions found in spectra of peptides resulting from a tryptic digest of BSA and their labeled counterparts. All peptides (n = 6) contain a single lysine at the C-terminus. An asterisk indicates significant ($p < 0.05$) decrease in c-ions and increase in z-ions, respectively, between labeled peptides and their unlabeled cognates.

In order to test our findings in a proteomics environment, we progressed to a more complex sample which in our case was created by the use of a human cell lysate. We evaluated both trypsin and Lys-N with dimethylation, guanidination, or imidazolinylation. Proteolytic peptides were, initially, separated by low pH strong cation exchange (SCX) chromatography.[26] At pH 2.7, all carboxylic acid residues will be neutral and charge will be dictated mostly by protonated basic moieties.[44-46] Under such conditions, SCX enables separation of peptides on the basis of the number of basic residues present. Fractions containing peptides with a single basic amino acid or two basic amino acids were selected and chemically modified. Appearance of fragment ions for the labeled peptides were compared against their unlabeled cognates using a two sample one-sided *t* test to calculate the probability for the difference in c- and z-ion populations. Pleasingly, the trends observed mirrored those reported above for the BSA peptides. The number of c-ions detected in the guanidinated and imidazolinylated doubly charged Lys-N peptides were significantly greater compared to their unlabeled cognates with a corresponding *p*-value < 0.0001 (Table 1). In this larger data set we could also confirm that more c-ions are observed by double dimethylated doubly charged Lys-N peptides compared to their unlabeled cognates with a corresponding *p*-value < 0.05. In agreement with the results obtained for the peptides originating from the BSA digest, in the cell lysate derived data we also observed significantly more z- and less c-ions for C-terminal guanidinated and imidazolinylated tryptic peptides compared to their unlabeled cognates (*p*-values < 0.0001),

whereas no significant changes in c- or z-ion amounts could be observed for tryptic peptides dimethylated at both termini. All *p*-values for all combinations of protease and chemical labels are summarized in Table 1.

**Table 1. Statistical Significance of the Observed Trends in the ETD Fragment Ion Spectra of Lys-N and Tryptic Labeled and Unlabeled Peptides[a]**

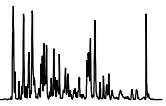
$KX_n$-peptides[b] (Lys-N)

| HEK 293T cell lysate | | | |
|---|---|---|---|
| | **increased** number of annotated c-ions | **decreased** number of annotated z-ions | number of labeled peptides compared against unlabeled peptides |
| guanidination | $p < 0.00001$ | $p < 0.0001$ | 206 |
| imidazolinylation | $p < 0.00001$ | $p < 0.001$ | 108 |
| dimethylation | $p < 0.05$ | $p < 0.001$ | 186 |

$X_nK$-peptides[c] (tryptic)

| HEK 293T cell lysate | | | |
|---|---|---|---|
| | **decreased** number of annotated c-ions | **increased** number of annotated z-ions | number of labeled peptides compared against unlabeled peptides |
| guanidination | $p < 0.00001$ | $p < 0.00001$ | 127 |
| imidazolinylation | $p < 0.00001$ | $p < 0.00001$ | 75 |
| dimethylation | $p = 0.182338$ | $p = 0.39796$ | 187 |

[a] *p*-values derived from a *t*-test based on the number of c- and z-ions observed for labeled versus unlabeled Lys-N or tryptic peptides (HEK 293T cell lysate). *P*-values smaller than 0.05 indicate a significant increase or decrease of ions for the labeled peptides compared to their unlabeled cognates.
[b] peptides with a N-terminal lysine where X is any amino acid besides H/K/R.
[c] peptides with a C-terminal lysine where X is any amino acid besides H/K/R.



Peptides containing more than one basic amino acid are predominantly triply or higher charged in ESI, and these triply charged peptides, in contrast to doubly charged peptides, produce spectra with almost complete complementary c- and z-ion series upon ETD. In some cases, these peptides can generate doubly charged precursors. In the data sets from the cell lysate, we could observe the effect of labeling on doubly charged peptides containing a histidine or an arginine. A "typical" example with labeled analogues is shown in Figure 6. Unlabeled doubly charged Lys-N or tryptic peptides with one histidine/arginine residue produce both c- and z-ions. Exact populations of different fragment ion series are dependent on the location of the histidine/arginine residue. Compared to the unlabeled cognate, guanidination and imidazolinylation of these peptides reduce the signal drop from high mass to low mass for the c-ion series with Lys-N peptides (Figure 6) and analogues for the z-ion series with C-terminal labeled tryptic peptides. On average, more than 1.5 additional c-ions and z-ions can be observed for guanidinated or imidazolinylated Lys-N and tryptic peptides, respectively, in comparison to their unlabeled cognate (Supplemental Table 1 in the Supporting Information). It follows that the sequence coverage obtainable with a single ion series is increased upon chemical labeling. Additionally, the relative intensities of dominant z-ions compared to the c-ions present are decreased in guanidinated and imidazolinylated Lys-N peptides and vice versa the number of c-ions is decreased for the guanidinated and imidazolinylated tryptic peptides (the corresponding *p*-values are summarized in Supplemental Table 1 in the Supporting Information).

As mentioned above, the majority of the nicotinylated peptides, when subjected to ETcaD, did not yield spectra revealing significant fragmentation (Figure 7). In our experiments, CID and ETD spectra were recorded for each precursor ion and thus characterization of ETD spectra was possible via CID spectra (Figure 7). A characteristic pattern was observed in several spectra of doubly nicotinylated peptides consisting of fragments with 77, 104, and 120 Da lower than the peptide mass. We hypothesize that these fragments are generated by the loss of pyridine-, dehydropyridinecarboxaldehyde-, and nicotinamide-species. Apart from these losses, b- and y-ions are also observed in spectra of nicotinylated Lys-N peptides. The b- and y-ions are most likely generated by the supplemental collisional activation. Nicotinylated peptides generated by both trypsin and Lys-N exhibit these same tendencies in ETD (Figures 2 and 4). In line with these findings, Li et al. observed decreased backbone and enhanced side chain cleavage upon labeling peptides with 2,4,6,-trimethylpyridine (TMP) in ECD.[47] Moreover, Belyayev et al. found decreasing backbone and enhanced side chain fragmentation upon labeling peptides with coumarin derivatives.[48] Both groups reasoned that their tags provide sites where a radical can be resonance stabilized and therefore might function as a radical trap. Further evidence is provided by Jones et al. who used a tag that contains a pyridinyl group and obtained similar results to those described here.[49] Additionally, the imidazole moiety of histidine has also been shown to have a subtle negative impact on dissociation after electron capture.[50] Nicotinic acid is a member of the heterocyclic aromatic family along with the above-mentioned reagents. We hypothesize that the nicotinic tag is functioning as a radical trap leading to the observed suppression of backbone fragmentation in ETD.

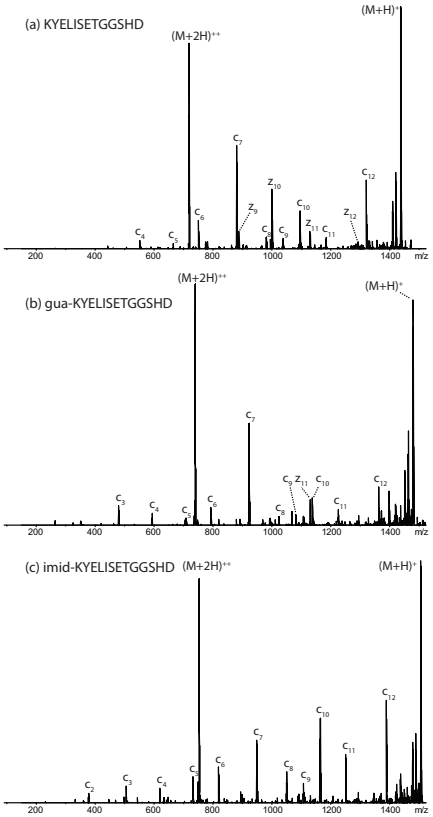



**Figure 6.** Comparison of ETcaD fragmentation spectra of the histidine containing doubly charged peptide KYELISETGGSHD identified from a Lys-N digested HEK293T cellular lysate: (a) unmodified, (b) guanidinated (gua), and (c) imidazolinylated (imid). Note that almost a full peptide sequence coverage is obtained on the basis of c-ions for the imidazolinylated peptide, whereby the z-ions observed for the unlabeled peptide are suppressed.

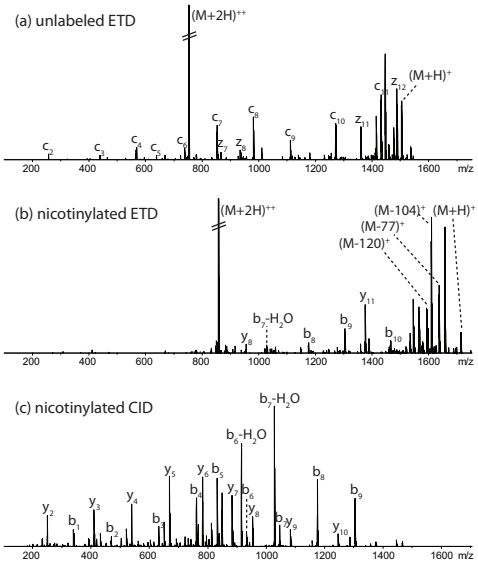
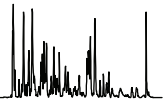



**Figure 7.** Tandem mass spectra of the Lys-N peptide KEYEATLEEccA derived from BSA. Cysteines of the peptide are carbamidomethylated. (a) ETcaD spectrum of the unlabeled peptide. (b) ETcaD spectra of the nicotinylated peptide. (c) CID spectrum of the nicotinylated peptide.

Quantification of proteins is becoming an increasingly important aspect of proteomics. It is most often carried out by comparison of differential metabolic or chemical isotope labeled samples.[37] The quantitative labels used at present have almost exclusively been used on CID instrumentation. To our knowledge only ETD experiments with iTRAQ[28, 29] and TMT[30] have been performed up to now. These tandem mass spectrometric based quantitation strategies were both compromised when used with ETD. The labels we investigated in this work and their isotopes are routinely used for differential isotopic labeling in CID.[19, 20, 31, 32, 35] In ETD we observed the guanidination, N-dimethylation, and imidazolinylation to be well suited particularly in combination with Lys-N peptides. Furthermore, all labels except nicotinylation improve the level of peptide sequence information obtainable by ETD based tandem mass spectrometry where imidazolinylation



demonstrated the largest improvement. In fact, labeling in a quantification experiment can potentially also improve the *de novo* sequencing aspect of the experiment.

C-terminal lysine containing tryptic peptides when guanidinated or imidazolinylated exhibited simplified spectra that contained nearly exclusively z-ions when no other basic residue was present. Both these reagents would allow tryptic pools of peptides when combined with low-pH SCX and ETD to create a *de novo* sequencing proteomic experiment similar to that previously proposed for Lys-N.[26, 27] Doubly charged arginine containing tryptic peptides already exhibited a strong z-ion series, and here we show that chemical labeling enables similar signatures for lysine containing peptides. Unlike peptides generated by Lys-N, tryptic peptides require chemical modification in order to allow the generation of simplified spectra. However, trypsin creates, potentially, a larger pool of peptides for *de novo* sequencing. The fourth modification we investigated, nicotinylation, is also used for quantification in CID.[34, 35] In ETD, nicotinylation of Lys-N and tryptic peptides hampered peptide identification since it suppresses peptide backbone fragmentation. Thus nicotinylation is inappropriate for any strategy that will involve ETD based sequence readouts.

## CONCLUSIONS

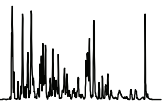
In recent years it has become apparent that ETD is a very useful and complementary activation method to fragment peptides in mass spectrometry. Here, we investigated the fragmentation behavior in ETD of chemically modified doubly charged peptides. Such chemical modification is performed quite often in proteomics to improve MS analysis, peptide identification, and/or to enable protein/peptide quantification. We used four different generally available labels, all targeting N-terminal amino groups and/or the ε-amino group of lysine. We found that guanidinated, dimethylated, and imidazolinylated peptides generated by Lys-N resulted upon ETD in even more extensive simple sequence ladders, potentially further facilitating *de novo* sequencing of such peptides. Guanidination and imidazolinylation of lysine terminated tryptic peptides also showed similar simplified spectra to those achieved by Lys-N peptides and arginine terminated tryptic peptides. In general the most striking positive effects were observed for imidazolinylated peptides. Modification of peptides by nicotinylation suppressed peptide backbone fragmentation in ETD, greatly reducing the number of successful peptide identifications. As all the chemical modifications that were studied in this work are amenable for a quantification strategy based on the incorporation of stable isotopes they except for nicotinylation can provide additional advantages to proteomic comparative experiments based on ETD sequencing.

**ACKNOWLEDGMENT**

This work was supported by The Netherlands Proteomics Centre (http://www.netherlandsproteomicscentre.nl/), a program embedded in The Netherlands Genomics Initiative. We thank Chris Miller (Agilent Technologies) and Dr. Eva-Maria Keidel (Max-Plank-Institute for Biochemistry; Martinsried, Germany) for donation of chemicals and Dr. Pantelis Hatzis and Dr. Tokameh Mahmoudi (Netherlands Institute for Developmental Biology, Utrecht, The Netherlands) for donation of HEK293T cells. We also thank Dr. Bas v. Breukelen (Biomolecular Mass Spectrometry and Proteomics Group, Utrecht, The Netherlands).

**SUPPORTING INFORMATION AVAILABLE**

Supplementary data set containing one excel file (Supplemental Table 1). This material is available free of charge via the Internet at http://pubs.acs.org (Two scaffold files including most of the data used in this study are available free of charge at https://proteomecommons.org/tranche/data-downloader.jsp?fileName=1dJj54qjNLW6%2FJg%2FSh9ryNFqf%2F10sAXkUqd0eiRkpH%2FwrpNSPsvvKIp9OBpzvgoJwuSTl8GUjdsft5fWfTnqgIMTmUoAAAAAAAAHSw%3D%3D; the passphrase is effpepchem1. The software to view the files (Scaffold viewer) is available free of charge at http://www.proteomesoftware.com/Proteome_software_prod_Scaffold_download-main.html. The first scaffold file contains the results of the tryptic and Lys-N digestion of BSA, and the second file contains the results of the Lys-N digestion of the cell lysate).



## REFERENCES


(1) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198-207.
(2) Sandra, K.; Moshir, M.; D'Hondt, F.; Verleysen, K.; Kas, K.; Sandra, P. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2008**, *866*, 48-63.
(3) Tang, J.; Gao, M. X.; Deng, C. H.; Zhang, X. M. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2008**, *866*, 123-132.
(4) Cravatt, B. F.; Simon, G. M.; Yates, J. R. *Nature* **2007**, *450*, 991-1000.
(5) Domon, B.; Aebersold, R. *Science* **2006**, *312*, 212-217.
(6) Johnson, R. S.; Martin, S. A.; Biemann, K. *International Journal of Mass Spectrometry and Ion Processes* **1988**, *86*, 137-154.
(7) Steen, H.; Mann, M. *Nature Reviews Molecular Cell Biology* **2004**, *5*, 699-711.
(8) Roepstorff, P.; Fohlman, J. *Biomedical Mass Spectrometry* **1984**, *11*, 601.
(9) Burlet, O.; Orkiszewski, R. S.; Ballard, K. D.; Gaskell, S. J. *Rapid Communications in Mass Spectrometry* **1992**, *6*, 658-662.
(10) Dongre, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H. *Journal of the American Chemical Society* **1996**, *118*, 8365-8374.
(11) Tsaprailis, G.; Nair, H.; Somogyi, A.; Wysocki, V. H.; Zhong, W. Q.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J. *Journal of the American Chemical Society* **1999**, *121*, 5142-5154.
(12) Paizs, B.; Suhai, S. *Mass Spectrometry Reviews* **2005**, *24*, 508-548.
(13) Leitner, A.; Lindner, W. *Proteomics* **2006**, *6*, 5418-5434.
(14) Keough, T.; Lacey, M. P.; Youngquist, R. S. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 2348-2356.
(15) Keough, T.; Youngquist, R. S.; Lacey, M. P. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 7131-7136.
(16) Spengler, B.; Luetzenkirchen, F.; Metzger, S.; Chaurand, P.; Kaufmann, R.; Jeffery, W.; Bartlet-Jones, M.; Pappin, D. J. C. *International Journal of Mass Spectrometry* **1997**, *169*, 127-140.
(17) Brancia, F. L.; Oliver, S. G.; Gaskell, S. J. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 2070-2073.
(18) Beardsley, R. L.; Reilly, J. P. *Analytical Chemistry* **2002**, *74*, 1884-1890.
(19) Warwood, S.; Mohammed, S.; Cristea, I. M.; Evans, C.; Whetton, A. D.; Gaskell, S. J. *Rapid Communications in Mass Spectrometry* **2006**, *20*, 3245-3256.
(20) Peters, E. C.; Horn, D. M.; Tully, D. C.; Brock, A. *Rapid Communications in Mass Spectrometry* **2001**, *15*, 2387-2392.
(21) Sullivan, A. G.; Brancia, F. L.; Tyldesley, R.; Bateman, R.; Sidhu, K.; Hubbard, S. J.; Oliver, S. G.; Gaskell, S. J. *International Journal of Mass Spectrometry* **2001**, *210*, 665-676.
(22) Qin, J.; Chait, B. T. *International Journal of Mass Spectrometry* **1999**, *191*, 313-320.
(23) Joss, J. L.; Molloy, M. P.; Hinds, L. A.; Deane, E. M. *International Journal of Peptide Research and Therapeutics* **2006**, *12*, 225-235.
(24) Nonaka, T.; Hashimoto, Y.; Takio, K. *Journal of Biochemistry* **1998**, *124*, 157-162.
(25) Boersema, P. J.; Taouatas, N.; Altelaar, A. F. M.; Gouw, J. W.; Ross, P. L.; Pappin, D. J.; Heck, A. J. R.; Mohammed, S. *Mol Cell Proteomics* **2009**, *8*, 650-660.
(26) Taouatas, N.; Altelaar, A. F. M.; Drugan, M. M.; Helbig, A. O.; Mohammed, S.; Heck, A. J. R. *Molecular & Cellular Proteomics* **2009**, *8*, 190-200.
(27) Taouatas, N.; Drugan, M. M.; Heck, A. J. R.; Mohammed, S. *Nature Methods* **2008**, *5*, 405-407.
(28) Han, H. L.; Pappin, D. J.; Ross, P. L.; McLuckey, S. A. *Journal of Proteome Research* **2008**, *7*, 3643-3648.
(29) Phanstiel, D.; Zhang, Y.; Marto, J. A.; Coon, J. J. *Journal of the American Society for Mass Spectrometry* **2008**, *19*, 1255-1262.
(30) Viner, R. I.; Zhang, T.; Second, T.; Zabrouskov, V. *Journal of Proteomics* **2009**, *72*, 874-885.
(31) Boersema, P. J.; Aye, T. T.; van Veen, T. A. B.; Heck, A. J. R.; Mohammed, S. *Proteomics* **2008**, *8*, 4624-4632.
(32) Hsu, J. L.; Huang, S. Y.; Chow, N. H.; Chen, S. H. *Analytical Chemistry* **2003**, *75*, 6843-6852.
(33) Raijmakers, R.; Berkers, C. R.; de Jong, A.; Ovaa, H.; Heck, A. J. R.; Mohammed, S. *Molecular & Cellular Proteomics* **2008**, *7*, 1755-1762.
(34) Munchbach, M.; Quadroni, M.; Miotto, G.; James, P. *Analytical Chemistry* **2000**, *72*, 4047-4057.
(35) Schmidt, A.; Kellermann, J.; Lottspeich, F. *Proteomics* **2005**, *5*, 4-15.

(36) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 9528-9533.
(37) Gevaert, K.; Impens, F.; Ghesquiere, B.; Van Damme, P.; Lambrechts, A.; Vandekerckhove, J. *Proteomics* **2008**, *8*, 4873-4885.
(38) Lemeer, S.; Pinkse, M. W. H.; Mohammed, S.; van Breukelen, B.; den Hertog, J.; Slijper, M.; Heck, A. J. R. *Journal of Proteome Research* **2008**, *7*, 1555-1564.
(39) Boersema, P. J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J. R. *Nat. Protocols* **2009**, *4*, 484-494.
(40) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E. P.; Coon, J. J. *Analytical Chemistry* **2007**, *79*, 477-485.
(41) Kalli, A.; Hakansson, K. *Journal of Proteome Research* **2008**, *7*, 2834-2844.
(42) Cooper, H. J. *Journal of the American Society for Mass Spectrometry* **2005**, *16*, 1932-1940.
(43) Mohammed, S.; Lorenzen, K.; Kerkhoven, R.; van Breukelen, B.; Vannini, A.; Cramer, P.; Heck, A. J. R. *Analytical Chemistry* **2008**, *80*, 3584-3592.
(44) Ballif, B. A.; Villen, J.; Beausoleil, S. A.; Schwartz, D.; Gygi, S. P. *Molecular & Cellular Proteomics* **2004**, *3*, 1093-1101.
(45) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J. X.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 12130-12135.
(46) Pinkse, M. W. H.; Mohammed, S.; Gouw, L. W.; van Breukelen, B.; Vos, H. R.; Heck, A. J. R. *Journal of Proteome Research* **2008**, *7*, 687-697.
(47) Li, X. J.; Cournoyer, J. J.; Lin, C.; O'Connor, P. B. *Journal of the American Society for Mass Spectrometry* **2008**, *19*, 1514-1526.
(48) Belyayev, M. A.; Cournoyer, J. J.; Lin, C.; O'Connor, P. B. *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 1428-1436.
(49) Jones, J. W.; Sasaki, T.; Goodlett, D. R.; Turecek, F. *Journal of the American Society for Mass Spectrometry* **2007**, *18*, 432-444.
(50) Xia, Y.; Gunawardena, H. P.; Erickson, D. E.; McLuckey, S. A. *Journal of the American Chemical Society* **2007**, *129*, 12232-12243.

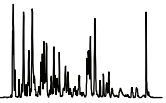
# Chapter 5

## Improving Depth in Phosphoproteomics by Using a Strong Cation Exchange-Weak Anion Exchange-Reversed Phase Multidimensional Separation Approach

Marco L. Hennrich,†,‡ Vincent Groenewold,§ Geert J. P. L. Kops,‡,§ Albert J. R. Heck,†,‡ and Shabaz Mohammed,†,‡

†Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
‡Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands
§Molecular Cancer Research and Cancer Genomics Centre, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands.

Published in Analytical Chemistry 2011; volume 83, issue 18: page 7137-7143.




## ABSTRACT

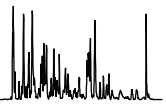
Several enrichment and separation strategies are available that allow nearly pure phosphopeptide pools to be created. These phosphopeptide pools are too complex to be completely unraveled by RP-LC-MS analysis alone. Here, we implement weak anion exchange (WAX) chromatography as an additional, complementary dimension to strong cation exchange (SCX) and reversed phase (RP). Initially, we used SCX to fractionate a human lysate digest to generate a fraction highly enriched for phosphopeptides. Analysis of this single fraction by RP-LC-MS with a 140 min gradient method allowed the identification of 4045 unique phosphopeptides (false discovery rate (FDR) < 1%; Mascot score > 20) using an Orbitrap Velos. Triplicate analysis (420 min total gradient time) of the same sample increased the total to just over 5000 unique phosphopeptides. When we separated the same sample by WAX and analyzed 14 WAX fractions by 30 min gradient RP-LC-MS (420 min total gradient time) we were able to identify 7251 unique phosphopeptides, an approximate increase of 40%, while maintaining the same total gradient time. We performed a more comprehensive, albeit also more time-consuming, analysis of the same 14 WAX fractions by the use of 140 min gradient LC-MS analyses, which resulted in the detection of over 11 000 unique phosphopeptides. Our results clearly demonstrate that additional separation dimensions are still necessary for in-depth phosphoproteomics and that WAX is a suitable dimension to be combined and sandwiched between SCX and RP chromatography.


## INTRODUCTION

Dynamic protein phosphorylation is a major regulation process in mammalian cells. A prerequisite to understand such processes is the exact identification of the phosphorylation sites and their occupancy. Nowadays, one of the fundamental technologies to analyze protein phosphorylation is reversed phase liquid chromatography coupled to mass spectrometry (LC-MS).[1-3] However, solely LC-MS cannot resolve the high complexity and the dynamic range associated with a mammalian cell phosphoproteome. Classically, the identification of low-abundance phosphopeptides requires highly efficient enrichment strategies. Currently, common strategies for phosphopeptide enrichment are based on affinity (either immunoprecipitation or chelation) or physicochemical properties through the use of chromatography. Chelation based methods rely on the use of metal ions (e.g., immobilized metal ion affinity chromatography, IMAC[4]) or metal oxides (metal oxide affinity chromatography, MOAC). The metal ions $Fe^{3+}$,[4] $Zr^{4+}$,[5] and $Ti^{4+}$ [6] are typical in IMAC, and the most common metal oxide used for phosphopeptide enrichment is titanium dioxide ($TiO_2$).[7-9] Immunoprecipitation methods are mainly focused on the specific enrichment of tyrosine phosphorylation either at the peptide[10, 11] or protein[12] level, although recently some progress has been made in the antibody-based enrichment of specific Ser/Thr phosphopeptides.[13] Liquid chromatographic based techniques being applied for enrichment

of phosphopeptides include strong cation exchange (SCX),[2, 14-16] strong anion exchange (SAX)[17, 18] and electrostatic repulsion hydrophilic interaction chromatography (ERLIC),[19, 20] among which SCX is one of the most commonly adopted.[2, 14, 15, 21, 22] The SCX separation relies predominantly on the peptide net charge, position of charged groups, and peptide length.[23] Typically, when SCX is applied for phosphopeptide enrichment, low-pH conditions are employed. Under these acidic conditions (pH ~2.7), glutamates and aspartates are predominantly protonated, and therefore neutral, whereas the phosphosite retains a negative charge. The negative charge associated with the phosphate groups on phosphopeptides leads to a lower net charge, when compared to nonphosphopeptides, and consequently leads to poorer retention on the SCX column, enabling a separation of phosphopeptides from the majority of nonphosphorylated peptides. When experimental conditions are controlled one can obtain the exclusive elution of a large uniform population of phosphopeptides.[15, 23] To improve the enrichment of all types of phosphopeptides, each SCX fraction can be further subjected to enrichment with either IMAC[21] or MOAC.[24-26]

In mammalian cells, there is a high occurrence of phosphorylation events that are spread across most proteins, thus the product of enrichment still exhibits an enormous dynamic range and high complexity. Analyses of proteolytic mixtures that have been fractionated and enriched will, almost always, still suffer from under-sampling and biases due to the sample composition. Thus, all aspects of proteomic workflows can still be further improved including, for instance, the final nanoliter flow rate reversed phase liquid chromatography step.[27-31] For instance, Köcher et al. have recently characterized a “long column” LC system which demonstrated a superior separation and led to an improved identification rate. Additionally, they described a linear correlation between the number of peptide identification and peak capacity in their LC-MS setup.[31] In multidimensional systems, the theoretical achievable peak capacity of the total system is almost equal to the product of the individual peak capacities, assuming complete orthogonality.[32-35] Thus, a great increase in peak capacity can be achieved by introducing an additional orthogonal dimension. Although complete orthogonality cannot be realized, a significant level of orthogonality can be achieved when the modes of separation of the chromatographic techniques are based on vastly different physicochemical properties of the peptides. Examples of two-dimensional systems with a high orthogonality are HILIC-RP-LC[36-39] and SCX-RP-LC, among which SCX-RP-LC is the most commonly used combination in proteomics.[33, 40] The multidimensional protein identification technology (MudPIT) for instance is a well-known application in which SCX and RP-LC are combined.[30, 41] In the Yin-Yang multidimensional liquid chromatography (MDLC) approach, Dai et al. applied SCX, SAX, and RP-LC-MS to study the mouse liver proteome. The flow through of the crude SCX separation was subjected to a pH step elution fractionation by SAX. This second step not only allowed further fractionation but also enrichment of singly and multiply phosphorylated peptides. Essentially, the authors used the complementary nature of SCX and SAX to fractionate a proteome.[17] Recently, we demonstrated that it is possible to separate multiply phosphorylated and singly phosphorylated peptides alongside nonmodified peptides by using SCX alone.[15]



In our quest to improve phosphopeptide identification rates we, in this work, implement and characterize weak anion exchange (WAX) chromatography as an intermediate step between the initial SCX fractionation and the nanoRP-LC-MS analysis. We demonstrate that the introduction of a complementary separation is able to dramatically increase the number of identified phosphopeptides, indicating that saturation in detection is still not reached in two-dimensional phosphoproteomics enrichment strategies.

## MATERIAL AND METHODS

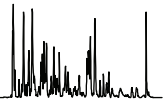
**Reagents and Materials.** Iodoacetamide was acquired from Sigma-Aldrich (Steinheim, Germany). Formic acid, sodium chloride, potassium chloride, and potassium dihydrogen phosphate were obtained from Merck (Darmstadt, Germany). Ammonium bicarbonate and dithiothreitol (DTT) were purchased from Fluka (Buchs, Switzerland). Protease inhibitor cocktail and PhosSTOP phosphatase inhibitor were supplied by Roche Diagnostics (Mannheim, Germany). Acetonitrile was purchased from Biosolve (Valkenswaard, The Netherlands). The water used in the experiments was obtained from a Milli-Q purification system (Millipore, Bedford, MA). Protein Assay Dye Reagent Concentrate was supplied by Bio-Rad Laboratories (Munich, Germany). Sep-Pak Vac tC18, 1 $cm^3$ cartridges were obtained from Waters Corporation (Milford, MA). Lys-C was obtained from Wako Chemicals (Richmond, VA) and trypsin from Promega (Madison, WI).

**Cell Lysis and Digestion of Proteins.** A protein extract of HelaS3 cells arrested in M-phase and of HEK293 cells was processed as described previously.[22] The protein concentration was estimated by a Bradford assay. HelaS3 and HEK293 proteins as well as a mixture of α and β casein were reduced and carbamidomethylated using DTT and iodoacetamide. Subsequently, the proteins were digested with Lys-C for 4 h at 37 °C followed by a 1:4 dilution with water and a digestion with trypsin at 37 °C overnight.

**Strong Cation Exchange (SCX) Chromatography.** The peptide solutions were desalted prior to SCX chromatography using a C18 Sep-Pak cartridge (Waters Corporation, Milford, MA), dried in a vacuum centrifuge, and resuspended in 10% formic acid. SCX was performed on an Agilent 1100 HPLC system (Agilent Technologies, Waldbronn, Germany) consisting of a nanopump, a micro well plate autosampler, a multiple wavelength detector and a fraction collector. The columns used were a strata-X 33 µm, 50 mm X 4.6 mm (Phenomenex, Torrance, CA) for trapping and a PolySULFOETHYL-A (PolyLC Inc., Columbia, MD) 200 mm X 2.1 mm analytical column with 5 µm particles. Buffers used for trapping were 0.05% formic acid in water (pH ~2.9) and 0.05% formic acid in 80% acetonitrile. Buffers used for SCX separation contained 0.05% formic acid, 5 mM $KH_2PO_4$, and 30% acetonitrile. The elution buffer additionally contained 350 mM KCl. The separation was performed by isocratic and gradient elution starting with a 2 min isocratic step followed by a 1 min gradient from 0% to 3% elution buffer. The elution buffer concentration was kept at 3% for 7 min followed by a 30 min gradient to 35% with subsequent washing and equilibration. Collected fractions were concentrated in a vacuum centrifuge.

**Weak Anion Exchange (WAX) Chromatography.** A total of 200 µg of the tryptic casein digest or the material of an individual SCX fraction enriched in phosphopeptides was separated by WAX chromatography. WAX was performed on an Agilent 1100/1200 HPLC system (Agilent Technologies, Waldbronn, Germany) consisting of an 1100 quaternary pump, an 1100 temperature controlled well plate autosampler, an 1100 variable wavelength detector, and a 1200 temperature controlled fraction collector. The columns used were an optilynx (Optimize Technologies, Oregon City, OR) trapping cartridge and a PolyWAX LP (PolyLC Inc., Columbia, MD) 50 mm X 1.0 mm analytical column with 3 µm particles. Buffers used for trapping were 0.05% formic acid in water (pH ~2.9) and 80% acetonitrile in water. Buffers used for separation contained 0.05% formic acid and 20% acetonitrile. The elution buffer additionally contained 500 mM NaCl. The separation was performed starting with a 6.25 min isocratic step followed by a 3.75 min gradient from 0% to 2% elution buffer, a 30 min gradient from 2% to 40% elution buffer, and subsequent washing and equilibration. Collected fractions were concentrated in a vacuum centrifuge to remove the acetonitrile.

**NanoLC-ESI-HCD-MS/MS.** The peptides were subjected to nanoLC-MS/MS analysis consisting of an Agilent 1200 HPLC system[42] (Agilent Technologies, Waldbronn, Germany) connected to an LTQ Orbitrap Velos Mass Spectrometer equipped with an electron transfer dissociation (ETD) source (Thermo Fisher, Bremen, Germany). The material used for the 20 mm X 100 µm trap column and 400 mm X 50 µm analytical column was Reprosil-pur C18 3 µm (Dr. Maisch, Ammerbuch-Entringen, Germany). Mobile phase buffer A consisted of 0.1 M acetic acid in water, and mobile-phase buffer B consisted of 0.1 M acetic acid in 80% acetonitrile. Trapping and washing of the sample was performed at 5 µL/min for 10 min with 100% mobile phase buffer A. The subsequent gradient varied in length between 17 and 140 min starting with 10% buffer B. Detailed information about the used gradients are given in Supplementary Table 1 in the Supporting Information. After the gradient, the column was washed by increasing the buffer B concentration to 100% followed by conditioning the system with 100% buffer A for at least 15 min. During the gradient and subsequent washing and conditioning of the column, the flow rate was passively split from 600 µL/min to 100 nL/min and the column effluent was directly introduced into the ESI source of the MS using an in-house pulled fused silica emitter, gold-coated by a Scancoat six Pirani 501 sputter coater (Edwards Laboratories, Crawley, U.K.), biased to 1.7 kV. The mass spectrometer was operated in positive ion mode, from 350 to 1500 m/z in MS mode. Parent ions were fragmented by HCD with a maximum injection time of 100 ms.[43, 44] The normalized collision energy was set to 40%. The signal threshold for triggering an MS/MS event was set to 500 counts. The 445.120 025 ion and 519.138 600 was used as lock mass for internal mass calibration. The low mass cutoff for HCD was set to 180 m/z. Charge state screening was enabled and precursors with unknown charge state or a charge state of 1 were excluded. For the WAX characterization experiments, the same RP-LC setup was connected to an LTQ Orbitrap XL mass spectrometer equipped with an ETD source (Thermo Fisher, Bremen, Germany). The mass spectrometer was operated in positive ion mode, from 350 to 1500 m/z in MS mode. Parent



ions were fragmented by collision induced dissociation (CID) with a maximum injection time of 100 ms. The normalized collision energy was set to 35%. The signal threshold for triggering an MS/MS event was set to 500 counts. No lock mass was used for internal mass calibration. Charge state screening was enabled and precursors with unknown charge state or a charge state of 1 were excluded.

**Protein Identification.** Raw MS data were converted to peak lists using Proteome Discoverer software (Thermo, version 1.3). The results were searched against the IPI (International Protein Index) Human database, version 3.37 (69 164 sequences; 29 064 824 residues) using Mascot software version 2.3.02 (www.ma- trixscience.com). The database search was made with trypsin cleavage specificity allowing 1 missed cleavage. The parameters were set to a peptide tolerance of ±50 ppm, a fragment tolerance of ±0.1 Da, carbamidomethyl (C) as a fixed modification and oxidation (M) and phosphorylation (STY) as a variable modification. A decoy search was enabled. Mascot results were filtered with the integrated Percolator based filter of Rockerbox software, version 1.2.6.[45] An FDR < 1% (based on peptide spectrum matches (PSMs)) was used as the cutoff in Percolator. Filtered tandem mass spectra assigned with a Mascot peptide score below 20 were discarded. An overview of the identified peptides used for the comparison is listed in Supplementary Table 2 in the Supporting Information. Additional information is available as a scaffold file on tranche at https://proteomecommons.org using the following hash: N7IavNwmExWOzBE+X7encybySfy/hksItuGJ+6KxcIOHBMh/LPxtNoPHS3Ug2GFCIpDJxv34l-tiBTMdx/MURvb11/2IAAAAAAAADqw==. The passphrase is SCXWAXRP911. The software to view the files (Scaffold viewer) is available free of charge at http://www.proteomesoftware.com/Proteome_software_prod_Scaffold_download-main.html.

## RESULTS

**Initial Evaluation of WAX.** A strategy to increase the peak capacity of a chromatographic system is to implement an additional orthogonal chromatographic dimension.[34, 35, 40] Since the peak capacity is related to the identification rate,[31] we decided to test an additional dimension in our analytical workflow to increase the identification rate. WAX chromatography was chosen because the mode of separation is complementary to SCX and RP-LC, which are the two dimensions present in our (and many others) proteomics workflow. Beneficially, WAX, unlike SAX, enables elution of phosphopeptides with a low salt concentration.

Initially, we configured a weak anion exchange (WAX) system with UV detection and fraction collection. The WAX system was set up in such a way that we were able to desalt the sample online prior to WAX separation. This was accomplished by the implementation of a C18 trap cartridge as depicted in the upper right corner of Figure 1. Furthermore, we chose to maintain 20% acetonitrile in the separation buffers to reduce unwanted unspecific hydrophobic interactions. In order to characterize the WAX separation, we analyzed a tryptic digest of 200 μg of an α and β casein protein mixture. The resultant UV-

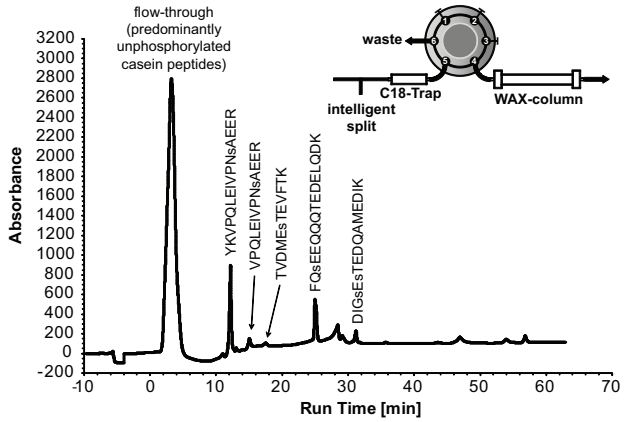



**Figure 1.** UV chromatogram of a mixture of α and β casein tryptic peptides separated by weak anion exchange (WAX). Small caps in the peptide sequence indicate phosphorylated residues. A schematic of the online desalting of the WAX system is depicted in the right corner.

chromatogram of the separation of this casein digest is shown in Figure 1. The peak widths at half height are less than 1 min, and the majority of peak widths were between 0.2 and 0.5 min wide (Figure 1). On the basis of the peak width and in order to keep the number of fractions reasonable, we chose to collect 1 min fractions. We then further characterized the WAX separation. An SCX sample highly enriched for phosphopeptides was chosen for this purpose, as our final goal is the implementation of WAX in a phosphoproteomic workflow. In order to create this sample, we chose SCX at low pH for enrichment of phosphopeptides from a tryptic digest of a HEK293 cell lysate, as described previously.[15] Separation of tryptic peptides by SCX leads to defined populations of peptides, which are characterized predominantly by their orientation and net charge.[23] The greatest population consists of unmodified peptides containing only one basic amino acid, which has been confirmed experimentally and by *in silico* tryptic digestion of the IPI human database.[14, 15] The phosphorylated counterparts of this population can be enriched by SCX leading to fractions almost pure in phosphopeptides.[15] One of these SCX fractions, further referred to as the SCX phosphofraction, was subsequently separated by WAX chromatography into 1 min fractions. The chosen SCX phosphofraction is dominated by phosphopeptides, which contain one basic amino acid and a free N-terminus. Phosphopeptides with a different net charge like multiply phosphorylated peptides or phosphopeptides with more than one basic amino acid elute in other fractions of the SCX separation and are not chosen to be separated by WAX. A number of consecutive WAX fractions were combined to create four pools containing roughly a similar number of peptides. The four pools were analyzed by RP-LC MS in order to obtain a rough picture of the WAX binding specificity. We expected retention of peptides to increase with increasing acidity, which is primarily controlled by the number of glutamic and aspartic acid residues in each peptide. As illustrated in Figure 2, a clear trend correlating with peptide acidity is observed. The first pool (including the flow-through) is dominated by phosphopeptides that contain no additional acidic amino acids (depicted in green). The consecutive pool (pool 2) was mostly (more than 90%) composed of phosphopeptides containing at least one glutamic or aspartic acid residue (light and dark yellow). In pool 3 there is a large population of phosphopeptides with 2 acidic residues (dark yellow) and pool 4 is dominated with phosphopeptides containing 3 or

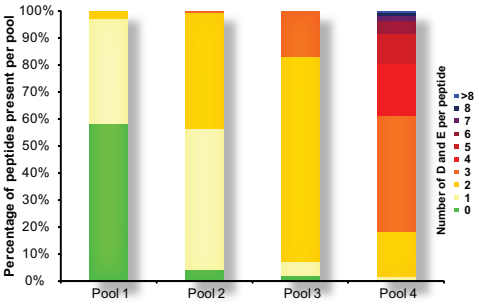



**Figure 2.** WAX separates primarily by the number of acidic residues present in the peptide sequence. In four subsequent pools of consecutive WAX fractions, the detected phosphopeptides were characterized by the number of glutamic acid (E) and aspartic acid (D) residues present in the sequence. In the early fractions (pool 1), nearly none or only one acidic residue was present. In contrast, the latest fractions (pool 4) were dominated by peptides having more than 3 acidic residues.

more acidic residues (orange and shades of red and blue). Resolving the data set further, we could observe that glutamic acid containing phosphopeptides eluted earlier than those peptides containing aspartic acid. This observation can be reasoned by the higher $pK_a$ value of the side chain carboxylic acid of the glutamic acid residue compared to the one of aspartic acid.[46] All these observations indicate that even though we are operating our system at low pH conditions and use solvents containing an acetonitrile concentration of 20% v/v, the mode of separation in our WAX system is dominated by anion exchange. Subsequently, we repeated the experiment and analyzed each fraction by RP-LC-MS. We observed that the phosphopeptides are not equally distributed over the fractions but elute in defined fractions of which 14 fractions contain most of the (phospho)peptides (see Supplementary Figure S1 in the Supporting Information and below).

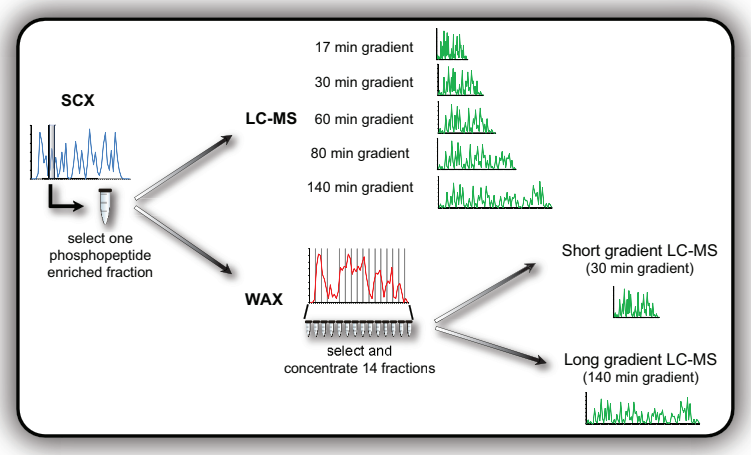



**Figure 3.** Schematic overview of the workflow used to directly compare SCX-RP-LC-MS analysis with SCX-WAX-RP-LC-MS. Proteins of a cell lysate were digested and subsequently first fractionated by using SCX. One of the SCX fractions, containing mostly singly phosphorylated peptides was taken for further analysis. Approximately 30% of this SCX fraction was used for several direct RP-LC-MS analyses using various gradients and number of replicates. The remaining 70% was first further fractionated by WAX. The WAX fractions were then analyzed by RP-LC-MS with either a 30 min gradient method or a 140 min gradient method.

**Comparison of RP-LC-MS with WAX-RP-LC-MS.** In order to explore the advantages and/or limitations of an additional dimension in the analysis of highly complex

phosphopeptide samples, we compared direct RP-LC-MS analysis with WAX separation followed by RP-LC-MS analysis, as illustrated in Figure 3. For this comparison we used another sample containing a high level of phosphorylation, Hela S3 cells arrested in M-phase. The starting material for the SCX separation was a tryptic digest of a lysate that contained approximately 2.5 mg of protein. The sample was separated by SCX, and the fraction corresponding to the highly enriched phosphopeptide pool "the SCX phospho-fraction" was utilized for the comparison.

**RP-LC-MS Analysis of a SCX Enriched Phosphofraction.** The RP-LC-MS system used for the comparison contains a short trap column for desalting and a 40 cm capillary analytical column that operates at a flow rate of 100 nL/min, as described previously.[42] The column was coupled to an Orbitrap Velos using higher energy collisional dissociation (HCD) for fragmentation. Our choice to use HCD is based on the superior results achieved by HCD when compared to ion trap collision induced dissociation (IT-CID) and electron transfer induced dissociation (ETD) for the analysis of doubly charged phosphopeptides.[43, 44, 47]

In order to characterize the RP-LC, we used different methods whose gradients varied in length from 17 to 140 min (Figure 3 and Supplementary Table 1 in the Supporting Information). For the analysis with the 30 and 140 min gradient method, replicates were also performed. We systematically increased the material used as the gradient time increased so as to maintain similar peak intensities for each analysis, a necessity since peaks broaden as the gradient increased in length. The amount loaded ranged from 0.33% to 2% of the phosphopeptide enriched SCX fraction, which approximately correlated with the increase in peak widths. For the gradient of only 17 min, we identified more than 1200 unique phosphopeptides with an FDR < 1% and a Mascot peptide score cutoff above 20. As expected, elongated gradient times led to more identifications (Figure 4 and Supplementary Table 1 in the Supporting Information). The longest gradient (140 min) applied enabled the identification of nearly 4 000 unique phosphopeptides. The total number of unique phosphopeptides identified using each different gradient are listed in Figure 4 and in Supplementary Tables 1 and 2 in the Supporting Information. We observed that the number of phosphopeptides identified showed a logarithmic relation with respect to the gradient time ($R^2 > 0.99$) (Figure 4). Thus, long gradients are seemingly not very efficient, with respect to the total time utilized (Supplementary Figure S2 in the Supporting Information). With regards to the total analysis time, very short gradients are also not the most efficient, as loading of the sample and the equilibration of the column does not lead to identifications (Supplementary Figure S3 in the Supporting Information). We calculated the number of phosphopeptides identified per total analysis time for each of the gradients and found the most efficient gradient in our system was 30 min (Supplementary Figure S3 in the Supporting Information).

Furthermore, we investigated whether a single analysis with a long gradient or repeated analyses with a short gradient lead to improved identifications of unique phosphopeptides. The combined data set of three replicates of the 30 min gradient resulted in the identification of 2669 uniquely identified phosphopeptides, while a single analysis using an

80 min gradient led to the identification of 3054 phosphopeptides. The combined data set of the triplicate analysis with the 140 min gradient resulted in 5090 unique phosphopeptides, indicating the sample is still more complex than the system is capable to handle. We calculated that more than 5200 unique phosphopeptides would be obtained from a single 420 min analysis. Our data suggest that more identifications are expected from single analysis with a long gradient compared to repeated analysis with short gradients. The comparison is also visualized in Figure 4. In addition, our finding that long gradients are more effective than repeated analysis using short gradients is in agreement with data from Köcher et al., who performed a similar more comprehensive comparison, albeit using unmodified peptides.[31]

**WAX-RP-LC-MS Analysis of an SCX Enriched Phosphofraction.** Next we subjected the same SCX phosphofraction, as used for the characterization of our RP-LC system, to an additional separation by WAX, as depicted in Figure 3. Initially, 70% of the SCX fraction was separated by WAX and 34 1 min fractions were collected. On the basis of preliminary analyses (Supplementary Figure S1 in the Supporting Information), we anticipated that the majority of phosphopeptides would be found in 14 of the 34 fractions. The WAX fractions were concentrated to reduce the acetonitrile concentration in the fractions, and 15 to 30% of each of these 14 fractions was subjected to RP-LC-MS analysis using the 30 min gradient (Figure 5). We chose the 30 min gradient RP-LC-MS analysis due to it being the most efficient method at identifying peptides with respect to time utilized (Supplementary Figure S3 in the Supporting Information). The analysis of the 14 WAX fractions with the 30 min gradient method generated 7251 unique phosphopeptides in the combined data set using the same criteria as applied to the single dimension RP-LC-MS analyses (Figure 5). In order to

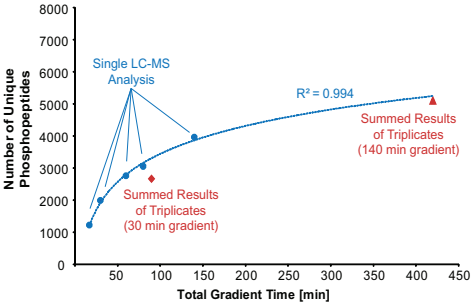



**Figure 4.** Number of phosphopeptides detected in the RP-LC-MS analysis of the phosphopeptide enriched SCX fraction using different gradient times and number of replicates. The number of unique phosphopeptides are plotted versus the used gradient analysis time. Blue circles represent the results of single RP-LC-MS analyses of the SCX fraction. Results obtained by using a triplicate RP-LC-MS analyses of the SCX fraction using a 30 min (diamond) or 140 min gradient (triangle) are colored red.

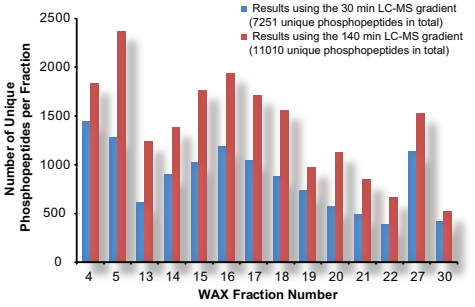



**Figure 5.** Number of phosphopeptides detected in each WAX fraction by RP-LC-MS analysis of the phosphopeptide enriched SCX fraction using two different gradient times. Data are shown of the RP-LC-MS analysis of the 14 WAX fractions which contained the most material. The number of unique phosphopeptides identified per fraction are plotted versus the fraction number. The blue bars represent the results by the use of a 30 min gradient, and the red bars represent the results by the use of a 140 min gradient.

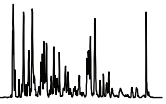
confirm the choice of fractions, the remaining 20 WAX fractions were also analyzed. The combined data set of all 34 analyzed WAX fractions led to 8004 unique phosphopeptides. The original 14 fractions contained over 90% of the identifications. To obtain an indication of what would be the maximum number of identifications we could reach from this phosphopeptide population, we subjected the 14 WAX fractions to 140 min gradient RP-LC-MS analyses (Figure 5). The 140 min gradient was chosen to test whether we had already reached saturation in phosphopeptide detection, realizing that this is at the expense of total analysis time. This latter analysis led to the identification of 11 010 unique phosphopeptides, indicating that saturation was not yet achieved.

## DISCUSSION

Here, we evaluated whether it is still beneficial to increase analytical power in phosphopeptide analysis through the addition of a weak anion exchange (WAX) separation in a SCX-WAX-RP configuration. Would addition of this extra dimension be beneficial and more time efficient when compared to using longer gradients in a two-dimensional SCX-RP approach? We therefore directly compared results of a RP-MS analysis with the results of the WAX-RP-MS analysis (Figure 3). Our RP-LC-MS analyses demonstrated indeed that analyses using longer gradients produced more identifications. The continuing increase confirmed that the complexity of this single SCX fraction is higher than the analytical power of our RP-LC-MS system. Increasing the number of identifications by elongating the gradient has evidently its limitations. The "efficiency" of the peptide identification decreases with longer gradient times (Supplementary Figure S2 in the Supporting Information), which can be reasoned by the logarithmic relationship between the number of phosphopeptide identifications and gradient time (Figure 4).[31] Furthermore, the results of Köcher and co-workers indicate that extremely long gradients might even lead to a decreasing number of identifications. The highest number identified by RP-LC-MS without WAX was around 4000 unique phosphopeptides with a single analysis (140 min gradient). Repeated analysis using the 140 min gradient method resulted in 5090 unique phosphopeptides in the combined data set of a triplicate analysis (total gradient time 420 min). In contrast to the direct RP-LC-MS measurements, the analysis of the 14 WAX fractions resulted in 7251 unique phosphopeptides, thereby using an identical total gradient time of 420 min. Over 40% more identifications could thus be achieved with the addition of a WAX separation stage.

## CONCLUSIONS

Our experiments highlight the immense complexity of the phosphoproteome. When starting from an SCX fraction that is highly enriched in phosphopeptides, single stage RP-LC-MS/MS analysis does not lead to saturation and/or comprehensiveness in phosphopeptide detection, not even when using rather long gradients and multiple replicate runs. We show that an additional dimension of separation is needed to get significantly closer to saturation. We explored the utility of WAX as a complementary chromatographic method to separate



phosphopeptides. As confirmed by our data, the retention of phosphopeptides in WAX under the described buffer conditions is primarily controlled by the number of glutamic and aspartic acid residues in the peptide sequence. We show that WAX may be efficiently combined with SCX chromatography and RP-LC-MS analysis. The analysis of one SCX fraction, highly enriched in phosphopeptides, using WAX-RP-LC-MS lead to more than 7200 unique phosphopeptide identifications in a total RP-LC-MS gradient time of 7 h, about 40% more than obtainable by SCX-RP-LC-MS alone in the same 7 h gradient time. To test if we had reached saturation, we conducted an even more comprehensive analysis by using long RP-LC gradients of 140 min. By this means, using an equivalent of half a milligram of starting material, we were able to identify 11 010 unique phosphopeptides (FDR < 1%) from only a part of a defined subpopulation of the phosphoproteome. In conclusion, although an additional (third) phase of separation may seem at first time inefficient, we show that it can be done, when using SCX-WAX-RP with appropriate experimental conditions, in an efficient manner enabling a significant increase in the amount of detected phosphopeptides.

## ASSOCIATED CONTENT

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

## ACKNOWLEDGMENT

The authors acknowledge the Heck-group members and in particular Henk van den Toorn for bioinformatic support. This work was in parts supported by the PRIME-XS project, Grant Agreement Number 262067, funded by the European Union Seventh Framework Programme; The Netherlands Proteomics Centre, embedded in The Netherlands Genomics Initiative, and the Centre for Biomedical Genetics; and The Netherlands Organization for Scientific Research (NWO) with the VIDI Grant 700.10.429.

## REFERENCES

(1) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198-207.
(2) Ballif, B. A.; Villen, J.; Beausoleil, S. A.; Schwartz, D.; Gygi, S. P. *Molecular & Cellular Proteomics* **2004**, *3*, 1093-1101.
(3) Cox, J.; Mann, M. *Annual Review of Biochemistry* **2011**, *80*, 273-299.
(4) Andersson, L.; Porath, J. *Analytical Biochemistry* **1986**, *154*, 250-254.
(5) Zhou, H. J.; Xu, S. Y.; Ye, M. L.; Feng, S.; Pan, C.; Jiang, X. G.; Li, X.; Han, G. H.; Fu, Y.; Zou, H. *Journal of Proteome Research* **2006**, *5*, 2431-2437.
(6) Zhou, H. J.; Ye, M. L.; Dong, J.; Han, G. H.; Jiang, X. N.; Wu, R. N.; Zou, H. F. *Journal of Proteome Research* **2008**, *7*, 3957-3967.
(7) Pinkse, M. W. H.; Uitto, P. M.; Hilhorst, M. J.; Ooms, B.; Heck, A. J. R. *Analytical Chemistry* **2004**, *76*, 3935-3943.
(8) Pinkse, M. W. H.; Mohammed, S.; Gouw, L. W.; van Breukelen, B.; Vos, H. R.; Heck, A. J. R. *Journal of Proteome Research* **2008**, *7*, 687-697.
(9) Lemeer, S.; Heck, A. J. R. *Current Opinion in Chemical Biology* **2009**, *13*, 414-420.
(10) Rush, J.; Moritz, A.; Lee, K. A.; Guo, A.; Goss, V. L.; Spek, E. J.; Zhang, H.; Zha, X. M.; Polakiewicz, R. D.; Comb, M. J. *Nature Biotechnology* **2005**, *23*, 94-101.

(11) Boersema, P. J.; Foong, L. Y.; Ding, V. M. Y.; Lemeer, S.; van Breukelen, B.; Philp, R.; Boekhorst, J.; Snel, B.; den Hertog, J.; Choo, A. B. H.; Heck, A. J. R. *Molecular & Cellular Proteomics* **2010**, *9*, 84-99.
(12) Steen, H.; Kuster, B.; Fernandez, M.; Pandey, A.; Mann, M. *Journal of Biological Chemistry* **2002**, *277*, 1031-1039.
(13) Gu, T. L.; Nardone, J.; Wang, Y.; Loriaux, M.; Villen, J.; Beausoleil, S.; Tucker, M.; Kornhauser, J.; Ren, J. M.; MacNeill, J.; Gygi, S. P.; Druker, B. J.; Heinrich, M. C.; Rush, J.; Polakiewicz, R. D. *Plos One* **2011**, *6*, e19169.
(14) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J. X.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 12130-12135.
(15) Gauci, S.; Helbig, A. O.; Slijper, M.; Krijgsveld, J.; Heck, A. J. R.; Mohammed, S. *Analytical Chemistry* **2009**, *81*, 4493-4501.
(16) Mohammed, S.; Heck, A. J. R. *Current Opinion in Biotechnology* **2010**, *22*, 9-16.
(17) Dai, J.; Jin, W. H.; Sheng, Q. H.; Shieh, C. H.; Wu, J. R.; Zeng, R. *Journal of Proteome Research* **2007**, *6*, 250-262.
(18) Han, G. H.; Ye, M. L.; Zhou, H. J.; Jiang, X. N.; Feng, S.; Jiang, X. G.; Tian, R. J.; Wan, D. F.; Zou, H. F.; Gu, J. R. *Proteomics* **2008**, *8*, 1346-1361.
(19) Alpert, A. J. *Analytical Chemistry* **2008**, *80*, 62-76.
(20) Gan, C. S.; Guo, T. N.; Zhang, H. M.; Lim, S. K.; Sze, S. K. *Journal of Proteome Research* **2008**, *7*, 4869-4877.
(21) Gruhler, A.; Olsen, J. V.; Mohammed, S.; Mortensen, P.; Faergeman, N. J.; Mann, M.; Jensen, O. N. *Molecular & Cellular Proteomics* **2005**, *4*, 310-327.
(22) Taouatas, N.; Altelaar, A. F. M.; Drugan, M. M.; Helbig, A. O.; Mohammed, S.; Heck, A. J. R. *Molecular & Cellular Proteomics* **2009**, *8*, 190-200.
(23) Alpert, A. J.; Petritis, K.; Kangas, L.; Smith, R. D.; Mechtler, K.; Mitulovic, G.; Mohammed, S.; Heck, A. J. R. *Analytical Chemistry* **2010**, *82*, 5253-5259.
(24) Benschop, J. J.; Mohammed, S.; O'Flaherty, M.; Heck, A. J. R.; Slijper, M.; Menke, F. L. H. *Molecular & Cellular Proteomics* **2007**, *6*, 1198-1214.
(25) Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. *Cell* **2006**, *127*, 635-648.
(26) Van Hoof, D.; Munoz, J.; Braam, S. R.; Pinkse, M. W. H.; Linding, R.; Heck, A. J. R.; Mummery, C. L.; Krijgsveld, J. *Cell Stem Cell* **2009**, *5*, 214-226.
(27) Tolley, L.; Jorgensen, J. W.; Moseley, M. A. *Analytical Chemistry* **2001**, *73*, 2985-2991.
(28) Swartz, M. E. *Journal of Liquid Chromatography & Related Technologies* **2005**, *28*, 1253-1263.
(29) Shen, Y. F.; Zhang, R.; Moore, R. J.; Kim, J.; Metz, T. O.; Hixson, K. K.; Zhao, R.; Livesay, E. A.; Udseth, H. R.; Smith, R. D. *Analytical Chemistry* **2005**, *77*, 3090-3100.
(30) Motoyama, A.; Venable, J. D.; Ruse, C. I.; Yates, J. R. *Analytical Chemistry* **2006**, *78*, 5109-5118.
(31) Kocher, T.; Swart, R.; Mechtler, K. *Analytical Chemistry* **2011**, *83*, 2699-2704.
(32) Giddings, J. C. *Analytical Chemistry* **1984**, *56*, 1258-&.
(33) Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. *Analytical Chemistry* **2005**, *77*, 6426-6434.
(34) Vivo-Truyols, G.; van der Wal, S.; Schoenmakers, P. J. *Analytical Chemistry* **2010**, *82*, 8525-8536.
(35) Motoyama, A.; Yates, J. R. *Analytical Chemistry* **2008**, *80*, 7187-7193.
(36) McNulty, D. E.; Annan, R. S. *Molecular & Cellular Proteomics* **2008**, *7*, 971-980.
(37) Boersema, P. J.; Divecha, N.; Heck, A. J. R.; Mohammed, S. *Journal of Proteome Research* **2007**, *6*, 937-946.
(38) Boersema, P. J.; Mohammed, S.; Heck, A. J. R. *Analytical and Bioanalytical Chemistry* **2008**, *391*, 151-159.
(39) Di Palma, S.; Boersema, P. J.; Heck, A. J. R.; Mohammed, S. *Analytical Chemistry* **2011**, *83*, 3440-3447.
(40) Sandra, K.; Moshir, M.; D'Hondt, F.; Tuytten, R.; Verleysen, K.; Kas, K.; Francois, I.; Sandra, P. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2009**, *877*, 1019-1039.
(41) Washburn, M. P.; Wolters, D.; Yates, J. R. *Nature Biotechnology* **2001**, *19*, 242-247.
(42) Raijmakers, R.; Berkers, C. R.; de Jong, A.; Ovaa, H.; Heck, A. J. R.; Mohammed, S. *Molecular & Cellular Proteomics* **2008**, *7*, 1755-1762.
(43) Boersema, P. J.; Mohammed, S.; Heck, A. J. R. *Journal of Mass Spectrometry* **2009**, *44*, 861-878.
(44) Frese, C. K.; Altelaar, A. F. M.; Hennrich, M. L.; Nolting, D.; Zeller, M.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. *Journal of Proteome Research* **2011**, *10*, 2377-2388.
(45) van den Toorn, H. W. P.; Munoz, J.; Mohammed, S.; Raijmakers, R.; Heck, A. J. R.; van Breukelen, B. *Journal of Proteome Research* **2011**, *10*, 1420-1424.
(46) Thurlkill, R. L.; Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. *Protein Science* **2006**, *15*, 1214-1218.
(47) Nagaraj, N.; D'Souza, R. C. J.; Cox, J.; Olsen, J. V.; Mann, M. *Journal of Proteome Research* **2010**, *9*, 6786-6794.

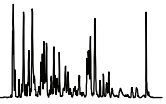
# Chapter 6

## Ultra Acidic Strong Cation Exchange Enabling the Efficient Enrichment of Basic Phosphopeptides

Marco L. Hennrich,†,‡ Henk van den Toorn,†,‡,§ Vincent Groenewold,‖
Albert J. R. Heck,†,‡ and Shabaz Mohammed,†,‡

†Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
‡Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands
§Netherlands Bioinformatics Centre, Geert Grooteplein 28, 6525 GA Nijmegen, The Netherlands
‖Molecular Cancer Research and Cancer Genomics Centre, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, the Netherlands

Published in Analytical Chemistry 2012; DOI: 10.1021/ac203303t




## ABSTRACT

We present a straightforward method to enrich phosphopeptides with multiple basic residues, an under-represented class in common enrichment strategies. Our method is based on a two dimensional strong cation exchange (SCX) strategy, operating at two different acidic pHs, enabling both separation and enrichment of different classes of phosphopeptides. The principle of enrichment is based on the change of net charge of phosphorylated peptides under strong acidic conditions in the second SCX, whereas the net charge of regular peptides remains unchanged, thus enabling separation based on net charge. Application of our tandem SCX approach to a modest amount of human cells allowed the identification of over 10000 unique 'basic' phosphopeptides of which many represent putative targets of basophilic kinases.


## INTRODUCTION

Dynamic protein phosphorylation is a key process in cells. In order to study this process, changes in the phosphorylation state and exact determination of the phosphorylation sites are necessary. Currently, the method of choice to sequence proteins is reversed phase liquid chromatography coupled to mass spectrometry (RP-LCMS). However, the low abundance, the high dynamic range, and the high complexity of the mammalian phosphoproteome requires enrichment and complexity reduction through fractionation.[1, 2] Several methods have been presented for the enrichment of phosphopeptides. These include immobilized metal ion affinity chromatography (IMAC) by the use of $Fe^{3+}$, $Zr^{4+}$, or $Ti^{4+}$ [3-6] and metal oxide affinity chromatography chelation (MOAC). Here, titanium dioxide ($TiO_2$) is by far the most common metal oxide applied.[7, 8] However, other materials such as niobium (V) oxide ($Nb_2O_5$)[9] or the rutile-form of $TiO_2$[10] have proven to be efficient for phosphopeptide enrichment and might be useful as complementary enrichment strategies, especially for the enrichment of basic phosphopeptides. Further methods include strong cation exchange (SCX),[11-13] immunoprecipitation,[6] or implementation of a sequence of chemical reactions.[14] However, these strategies require significant expertise in sample preparation and in some cases synthesis.[3, 14] In this work, we present a less demanding method involving solely SCX to isolate, in particular, a group of phosphopeptides that are typically underrepresented and hard to enrich as they contain multiple basic residues.

## MATERIAL AND METHODS

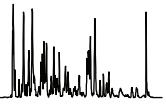
**Reagents and Materials.** Iodoacetamide, 1 M hydrochloric acid solution, and formaldehyde (37% solution in $H_2O$) were obtained from Sigma-Aldrich (Steinheim, Germany), and sodium cyanoborodeuterate was from Sigma-Aldrich (St. Louis, MO). Formic acid, sodium chloride, dipotassium hydrogen phosphate, and potassium dihydrogen phosphate were acquired from Merck (Darmstadt, Germany). Formaldehyde $^{13}CD_2O$ (20% solution in $D_2O$) was purchased from Isotec (Miamisburg, OH). Acetonitrile was obtained

from Biosolve (Valkenswaard, The Netherlands). Ammonium bicarbonate, 1 M sodium hydroxide solution, dithiothreitol (DTT), sodium cyanoborohydrate, and α-cyano-4-hydroxycinnamic acid were supplied from Fluka (Buchs, Switzerland). Protease inhibitor cocktail and PhosSTOP phosphatase inhibitor were purchased by Roche Diagnostics (Mannheim, Germany). Sep-Pak Vac tC18, 1 $cm^3$ cartridges were obtained from Waters Corporation (Milford, MA). Zip-Tip pipet tips were purchased from Millipore Corp. (Billerica, MA). Trypsin was obtained from Promega (Madison, WI) and Lys-C was from Wako Chemicals (Richmond, VA). The water used in the experiments was obtained from a Milli-Q purification system (Millipore, Bedford, MA).

**Cell Lysis and Digestion of Proteins.** Proteins of HelaS3 cells arrested in M-phase were extracted as described previously.[15] HelaS3 proteins as well as bovine serum albumine and a mixture of alpha and beta casein were reduced and carbamidomethylated using DTT and iodoacetamide. Subsequently, the proteins were digested with Lys-C for 4 h at 37 °C followed by a 1:4 dilution with water and a digestion with trypsin at 37 °C overnight.

**Strong Cation Exchange (SCX) Chromatography.** The samples were desalted prior to SCX chromatography using a C18 Sep-Pak cartridge, dried in a vacuum centrifuge, and resuspended in 10% formic acid. SCX was performed on an Agilent 1100/1200 HPLC system (Agilent Technologies, Waldbronn, Germany) consisting of an 1100 quaternary pump, an 1100 variable wavelength detector, an 1100 temperature controlled well plate autosampler, and an 1200 temperature controlled fraction collector. The columns used were 50 x 2.1 mm PL-SCX columns (Agilent Technologies, Shropshire, UK) with 5 µm particles and a pore size of 1000 Å. For all experiments at a pH of 1 two columns were used in series. For the comparison of pH = 1 and pH = 3 by the use of a mixture of BSA and casein, two columns were also used in series for the separation at a pH of 3; otherwise, only a single column was used for the separation at a pH of 3. Buffers for the SCX separation at a pH of 3 were 0.05% formic acid in 30% acetonitrile (buffer A) and, for elution, 0.05% formic acid and 500 mM NaCl in 30% acetonitrile (buffer B). Buffers for the SCX separation at a pH of 1 were 100 mM HCl in 30% acetonitrile (buffer A) and, for elution, 100 mM HCl and 500 mM NaCl in 30% acetonitrile (buffer B). The separation started isocratic at 100% buffer A for 1 min followed by a 1 min gradient from 0% to 10% buffer B, a 15 min gradient from 10% to 40% buffer B and a 3 min gradient from 40% to 100% buffer B. The column was subsequently washed for 5 min with 100% buffer B and equilibrated with 100% buffer A for 20 min. The flow rates were 100 µL/min for the two column set up and 200 µL/min for the single column set up. One minute fractions were collected. Fractions from SCX at a pH of 1 were neutralized by the addition of a NaOH solution. Subsequently, the fractions were concentrated in a vacuum centrifuge to remove the acetonitrile followed by desalting on Zip-Tip pipet tips prior to LC-MS analysis. Fractions from the SCX at a pH of 3 were concentrated in a vacuum centrifuge. The amount of HeLa peptides loaded on the SCX at a pH of 3 corresponded to 500 µg of protein. 50% of the resulting fractions, corresponding to 250 µg of protein, was separated at a pH of 1, and 2-50% thereof corresponding to 5-125 µg of protein was analyzed by LC-MS.

**NanoLC-MS/MS.** The samples were analyzed by NanoRP-LC-MS/MS analysis. The LC consisted of an Agilent 1200 HPLC system (Agilent Technologies, Waldbronn, Germany)



and was connected to an LTQ Orbitrap Velos Mass Spectrometer equipped with an ETD source (Thermo Fisher, Bremen, Germany) for analysis of the HeLaS3 cell lysate or was connected to an LTQ Orbitrap XL for the BSA/Casein analysis. The LC setup contained a 20 mm x 100 μm trap column and a 400 mm x 50 μm analytical column. The material used was Reprosil-pur C18 3 μm (Dr. Maisch, Ammerbuch-Entringen, Germany). Mobile phase buffers consisted of 0.1 M acetic acid in water (buffer A) and 0.1 M acetic acid in 80% acetonitrile (buffer B). Trapping and washing was performed at 5 μL/min for 10 min with 100% buffer A. Three different gradients (17 min, 90 min, and 150 min) were used. The 17 min gradient started from 10% buffer B and ended at 40% buffer B. The 90 min, 2 step gradient started with 10% buffer B and increased to 26% buffer B in 65 min followed by a 25 min gradient to 50% buffer B. The 150 min, 2 step gradient started with 10% buffer B and increased to 24% buffer B in 105 min followed by a 45 min gradient to 50% buffer B. Subsequently, the column was washed with 100% B and equilibrated with 100% buffer A. The flow rate was passively split from 600 μL/min to 100 nL/min during the gradient, and there was subsequent washing and conditioning of the column. The effluent was directly introduced into the ESI source of the MS. We used in-house pulled fused silica emitter, gold-coated by a Scancoat six Pirani 501 sputter coater (Edwards Laboratories, Crawley, UK). The ion spray voltage was set to 1.7 kV. The mass spectrometer was operated in positive ion mode, in a mass range from 350 to 1500 *m/z* in MS mode. Charge state screening was enabled. Precursors with a charge state of 1 or an unknown charge were not chosen for fragmentation. Precursors with a charge state of 2+ or higher were fragmented by HCD or ETD controlled by an in-house developed data dependent decision tree[16] or by CID for the Orbitrab XL. The normalized collision energy was set to 40% for HCD and 35% for CID. The low mass cutoff for HCD or FT-ETD was set to 180 *m/z*. 500 counts was set as threshold for triggering an MSMS event. The *m/z* of 445.120025 and 519.138600 was used as lock mass for internal mass calibration.

**Stability Test and MALDI Analysis.** Casein peptides solution was split in two and was either light or heavy dimethyl labeled on Sep-Pak Vac tC18 cartridges as described before.[17] Light and heavy peptides were split in two and dried in a vacuum centrifuge, and half of each labeled sample was reconstituted in either 30% acetonitrile (control) or in SCX buffer A (30% acetonitrile in 0.1 M HCl). The control and the SCX buffer samples were incubated for 1 h at room temperature. Subsequently, SCX buffer samples were neutralized with 0.05 M NaOH in 30% acetonitrile, and control samples were diluted with an equal volume of equally neutralized SCX buffer. Equal volumes of control (light labeled) and SCX buffer incubation (heavy labeled) were mixed. Control (heavy labeled) and SCX buffer incubation (light labeled) were mixed in the same way (label swop). The mixtures were desalted, cocrystallized with α-cyano-4-hydroxycinnamic acid and analyzed on an 4700 Proteomics Analyzer (Applied Biosystems, Foster City, CA) MALDI-TOF/TOF instrument. Spectra were acquired in a *m/z* range from 800 to 4000 in reflectron mode, and 20000 shots were averaged per spectrum. The instrument was equipped with a 200 Hz Nd:YAG laser operating at 355 nm.

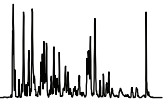
**Protein Identification.** Raw MS data were converted to peak lists using Proteome Discoverer software (Thermo Fisher Scientific Inc., version 1.3). Spectra were filtered for

the 6 most abundant peaks per 100 Da mass window. The results were searched against the swissprot database version 56.2 with taxonomy *Homo sapiens* (20407 sequences) by the use of Mascot software version 2.3. (www.matrixscience.com). The database search parameters were trypsin cleavage specific, allowing 2 missed cleavages, a peptide tolerance of ±50 ppm, a fragment tolerance of ±0.05 Da for orbitrap read out and ±0.9 Da for iontrap read out, carbamidomethyl (C) as fixed modification, oxidation (M), and phosphorylation (STY) as a variable modification. Mascot results were filtered with the integrated Percolator based filter using an FDR < 1% (based on PSMs). Only rank 1 peptides were allowed, and peptide hits with a mascot ion score below 20 and a peptide length <7 amino acids were discarded.

**Motif-x and iceLogo Analysis.** The 'neutral' phosphopeptide data set from a previous study[13] was reprocessed with the same parameters as the 'basic' phosphopeptide data set from the actual study. Phosphorylation site probability and scores were calculated by phosphoRS that is integrated in the Proteome Discoverer software (Thermo Fisher Scientific Inc., version 1.3). Phosphorylated peptides with a site probability <75% and a phosphoRS score <50 were discarded. The remaining unique further called type 1 phosphopeptides (Supplementary Table 2, Supporting Information) were loaded in Motif-x v1.2[18] with MS/MS as foreground format, 13 amino acids width, occurrence of 20, and a significance of 0.000001. For iceLogo v1.0.2 analysis,[19] the 'basic' phosphopeptides were compared with a reference set consisting of the 'neutral' phosphopeptide data set. We used a *p* value cutoff of 0.001.

## RESULTS

Our approach involves separating a sample using SCX at two different pH values. Our strategy was to exploit the change in charge of phosphate groups with pH in order to isolate phosphopeptides with multiple basic residues. The initial separation and fractionation is performed at a pH of 3, as successfully demonstrated many times,[11-13] where the only negatively charged group is the phosphate moiety which, alongside the basic residues, controls the separation. At this pH value the retention is primarily based on the net positive charge, leading to multiply phosphorylated peptides eluting first, followed by N-acetylated peptides, singly phosphorylated peptides, and finally regular peptides with a differing number of basic residues. In fact, we have shown that, using a refined SCX system at a pH of 3, one can efficiently enrich singly and multiply phosphorylated peptides, which contain only a single basic amino acid,[12] further called 'neutral' phosphopeptides. Unfortunately, phosphopeptides with two or more basic residues, further called 'basic' phosphopeptides, will have the same net positive charge as the bulk of the regular peptides and thus coelute during the SCX separation. Here, we experimentally explore, if a subsequent SCX separation, performed at an even more acidic pH, enables separation of these 'basic' phosphopeptides from the bulk of the regular peptides due to protonation and thus neutralization of the phosphate group. In practice, we take a fraction from the initial SCX separation at a pH of 3 containing 'basic' phosphopeptides and further acidify the solvent to a pH of 1, which renders the phosphate group neutral. Subsequently, we perform a second



SCX separation at a pH of 1. In this second separation, the unaffected regular tryptic peptides will elute first followed by the phosphopeptides, which now possess a higher net positive charge (Figure 1a). Several issues need to be considered and/or resolved for this approach to be successful. Our initial concern was the stability of regular and phosphorylated peptides at this low pH. Thus, we examined the stability of a typical set of peptides under such conditions. We opted for tryptic digests of caseins, since they covered all types of peptides. The digest was split into two equal aliquots and was differentially isotopically labeled.[17] Half of the peptides of each label was subjected to the solvent conditions required for SCX at a pH of 1, and the other half was used as a control. After incubation for 1 h at room temperature, the sample's pH was raised and mixed with the control. The analysis showed no obvious difference between the acidified and untreated sample (Supplementary Figure 1 and 2, Supporting Information). The second, maybe more important issue, is if one can perform an SCX separation with the required resolution at a pH of 1. Often classical chromatographic systems and consumables can fail under such conditions. Thus, we have specifically chosen commercial available LC equipment and SCX columns, that can operate at a pH of 3 and at a pH of 1. Using such a system, we obtained an average peak width at half height of approximately 0.5 min when separating a simple protein mixture digest at a pH of 3 (Supplementary Figure 3, Supporting Information). Pleasingly, the separation power of the same mixture at a pH of 1 is little affected, with an average peak width at half height of approximately 0.6 min (Supplementary Figure 3, Supporting Information). Furthermore, we compared the retention behavior of regular and phosphorylated peptides at the two pH conditions. Confirming our expectations, at a pH of 1, the phosphorylated peptides eluted later than the regular peptides that coeluted at a pH of 3 (Supplementary Figure 3 and Supplementary Table 1, Supporting Information).

In order to demonstrate the potential of two consecutive SCX separations (pH 3 and pH 1), which we call tandem SCX, we performed a full scale phosphoproteomics experiment focusing on the 'basic' phosphopeptides. HeLa cells arrested in M-phase were lysed and digested. Peptides corresponding to 500 μg of protein were first separated at a pH of 3, and 1 min fractions were collected. The 6 main fractions, containing the bulk of the doubly charged tryptic peptides, were each split into two, where one half was subjected to a second SCX separation at a pH of 1. One half of each of the main fractions from the second SCX separation (pH 1) was subjected to a nanoLC-MS analysis. The amount of material analyzed by nanoLC-MS corresponded to approximately 125 μg of starting material. One should keep in mind, that downscaling of the system (particularly column dimensions) might be advantageous, when the sample being analyzed is limited to low microgram amounts. We also analyzed the main fractions from the first SCX separation (pH 3) in order to create a reference to which we can calculate enrichment factors for the tandem SCX approach. We analyzed the type of peptides observed in each of the fractions created by the separation at a pH of 1. Two defined populations could be easily distinguished (Supplementary Figure 4, Supporting Information). The early fractions contain regular tryptic peptides, and the later fractions contain predominantly phosphorylated peptides.

Strikingly, the apexes of the two populations are several minutes apart, leading to close to near complete separation and enrichment of the 'basic' phosphopeptides.

We also performed direct nanoLC-MS analyses of the six initially collected SCX (pH 3) fractions. The direct analysis led to the identification of almost 18500 unique peptides but less than 550 unique phosphopeptides (Supplementary Table 2, Supporting Information). In contrast, the tandem SCX approach resulted in the identification of more than 31000 unique nonphosphorylated peptides and in addition more than 10000 unique 'basic' phosphopeptides (Figure 1b,c and Supplementary Table 2, Supporting Information). Compared to direct SCX-RP analysis, our tandem SCX approach thus resulted in a nearly 20-fold increase in identifications of 'basic' phosphopeptides (Figure 1b and Supplementary Figure 5, Supporting Information). The enrichment factor compares favorably to the best chelation based strategies.[3]

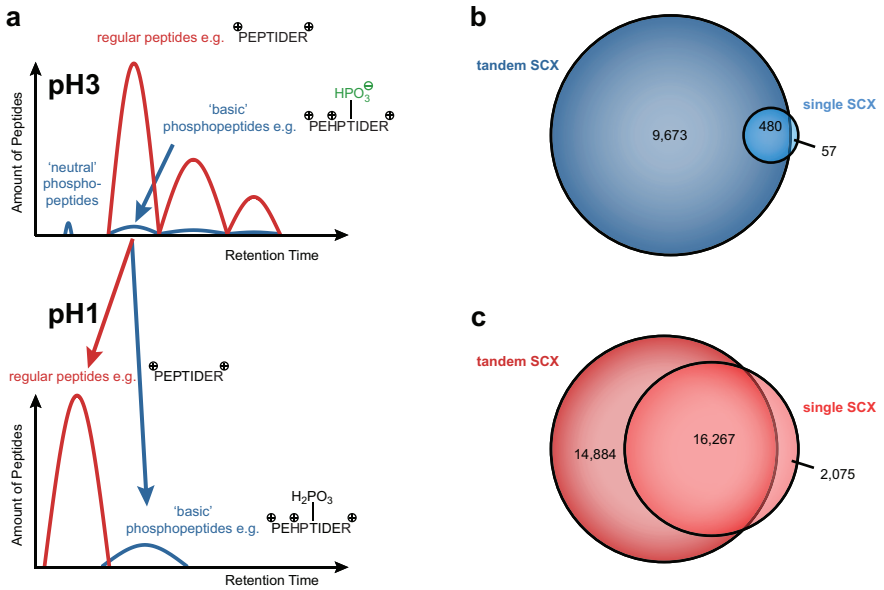
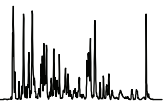



**Figure 1.** Overview of the principle and results of tandem SCX. **(a)** First peptides are separated with SCX at a pH of 3. The bulk of regular tryptic peptides (red) and 'basic' phosphopeptides (blue) both possess the same net positive charge and coelute. The regular peptides hinder the efficient analysis of the 'basic' phosphopeptides. Discrete fractions of the SCX separation at a pH of 3 are separated with SCX at a pH of 1. The net charge of phosphorylated peptides increases, due to protonation of the phosphorylation site at the pH of 1, as indicated by the example peptide. The change in net charge of the phosphopeptides leads to a relative increases in retention time, compared to the unaffected regular tryptic peptides. The net result is the separation of regular and phosphorylated peptides. **(b)** Venn diagram representing the overlap between unique phosphopeptides identified by tandem SCX (dark blue) or by direct analysis after SCX separation at a pH of 3 revealing a nearly 20-fold improvement in the number of detected phosphopeptides (light blue). **(c)** Venn diagram representing the overlap between unique regular peptides identified by tandem SCX (dark red) or after single SCX separation at a pH of 3 revealing only a 2-fold improvement in the number of detected regular peptides (light red).



The identified 'basic' phosphopeptides contain stretches that resemble targets of kinases with a preference for basic residues surrounding the phosphorylation site, so-called basophilic kinases. We compared the current tandem SCX results with a recently published similar sized data set of phosphopeptides which contain only a single basic amino acid and originate from the same cell line grown under similar conditions.[13] Using the iceLogo program,[19] we detected a significant relative overrepresentation of basic residues in our current tandem SCX data set (Supplementary Figure 6, Supporting Information). Similar overrepresentation of basic residues in phosphopeptides from late SCX fractions have also been found after enrichment with $TiO_2$ or $Ti^{4+}$-IMAC when attention is placed on the solvent systems applied (Supplementary Figure 6, Supporting Information).[3] Further analysis with the Motif-X program[18] pinpointed at several basic sequence logos only present in the tandem SCX data set (Supplementary Table 3, Supporting Information). Some of these, which are only identified in the tandem SCX data set, are motifs of well-known basophilic kinases including the PKA, DMPK, and the SGK families (Supplementary Table 3, Supporting Information).[20] Thus, the presented data set with several thousand phosphorylation sites might exhibit valuable information for researchers with an interest in basophilic kinases.

## CONCLUSIONS

We present a simple strategy based on a two-dimensional SCX strategy which we call tandem SCX. We show that tandem SCX is a facile method which allows the enrichment of 'basic' phosphopeptides. The separation and, as a consequence, the enrichment of 'basic' phosphopeptides are based on the change of the net charge in the second dimension (pH 1), whereas regular peptides keep the same net charge at both pHs (pH 3 and pH1). The difference in net charge in the second dimension allows the SCX separation of originally (first dimension) coeluting regular and phosphorylated peptides. Application of tandem SCX allowed us to identify over 10000 phosphorylated peptides from a subpopulation of the phosphoproteome of HeLa cells. Over 97% of the identified phosphopeptides had an additional basic residue, and many of them are potential targets of basophilic kinases. Thus, our approach is especially valuable for the research related to basophilic kinases. Additionally, tandem SCX combined with classical SCX at a pH of 3[11-13] will allow separation of all the main phosphopeptide families.

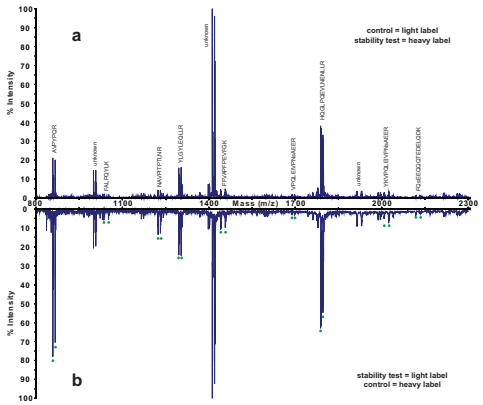
## ACKNOWLEDGEMENTS

The authors acknowledge all members of the Heck-group. This work was in part supported by the PRIME-XS project, Grant Agreement Number 262067, funded by the European Union Seventh Framework Program; The Netherlands Proteomics Centre, embedded in The Netherlands Genomics Initiative, and the Centre for Biomedical Genetics; and The Netherlands Organization for Scientific Research (NWO) with the VIDI Grant 700.10.429 for S.M.

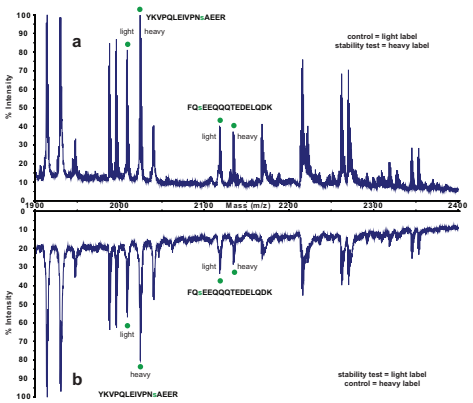
## SUPPORTING INFORMATION

Supplementary data set containing Supplementary Figures 1-6 and Supplementary Tables 1-3. This material is available free of charge via the Internet at http://pubs.acs.org. Additional information is available as .msf file on tranche at https://proteomecommons.org using the following hash:
bZZyQEga0NGTcZtVVPEBQlCWNqhwB5IMwsyobsXA560ZwfmYBHQtoL8DWZh5A
H940WjjURyHh/DaUq8nvnj7h+hbpawAAAAAAAB1fw== The passphrase is 2LSCX11. The software to view the files is Proteome Discoverer version 1.3 or the msf parser 1.0.3. The msf parser is available for free at http://code.google.com/p/thermo-msf-parser/

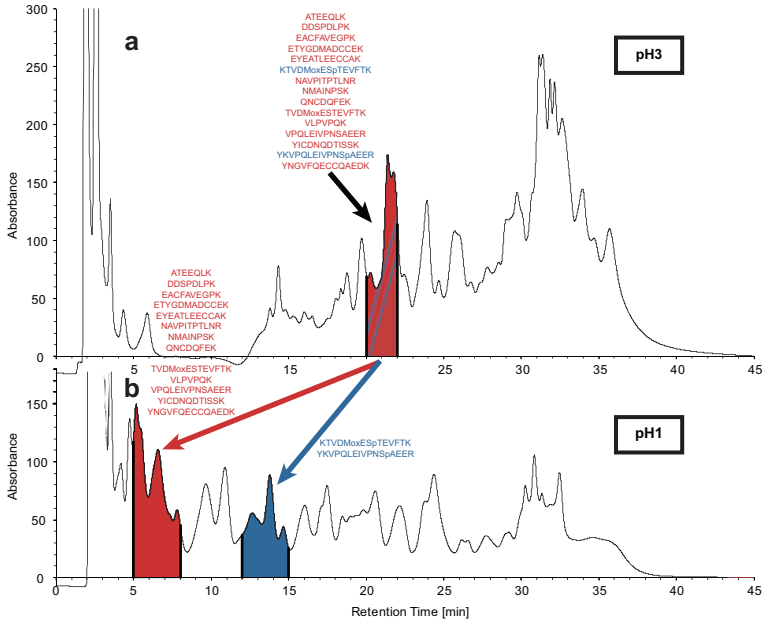
## SUPPLEMENTARY FIGURES



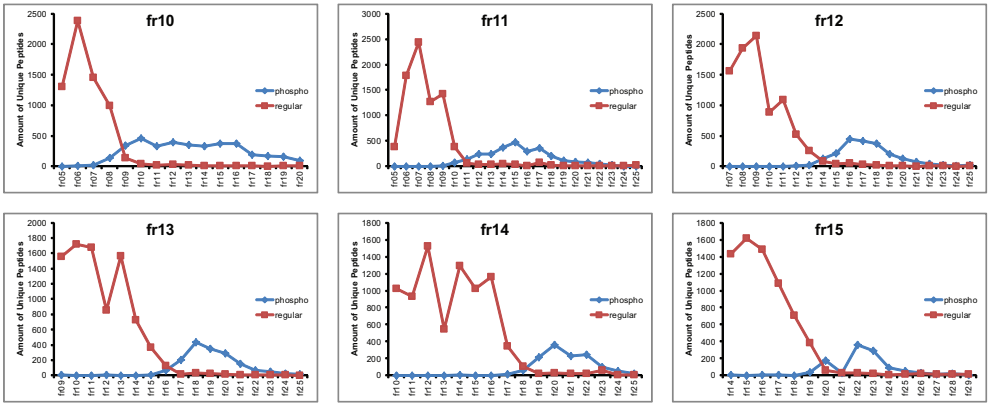

**Supplementary Figure 1.** MALDI spectra of differentially dimethyl labeled casein peptides originating from a stability test at a pH of 1. **(a)** Spectrum of a 1:1 mixture of light labeled control peptides and heavy labeled peptides incubated in SCX buffer at a pH of 1. **(b)** Spectrum of the label swop experiment, in which the light peptides were incubated in the SCX buffer at a pH of 1 and the heavy peptides served as control. The green dots indicate corresponding light and heavy forms of the same peptide. No obvious trend between the label swop experiments is visible, indicating no to little changes between control and stability test.



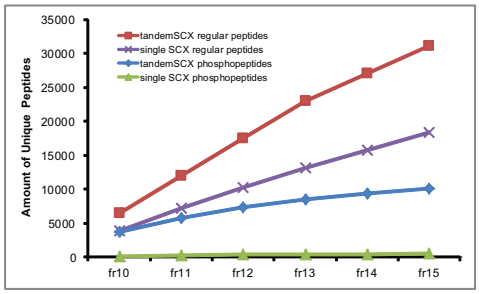

**Supplementary Figure 2.** Zoomed in region of MALDI spectra of differentially dimethyl labeled casein peptides from the stability test at a pH of 1. **(a)** Spectrum of a 1:1 mixture of light labeled control peptides and heavy labeled peptides incubated in SCX buffer at a pH of 1. **(b)** Spectrum of the label swop experiment, in which the light peptides were incubated in the SCX buffer at a pH of 1 and the heavy peptides served as control. The light and heavy form of the two phosphopeptides YKVPQLEIVPNsAEER and FQsEEQQQTEDELQDK are marked by a green dot. No obvious trend between the label swop experiments is visible, indicating no to little changes between control and stability test.



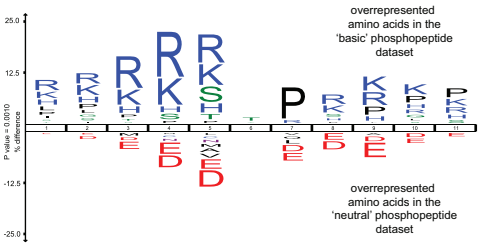

**Supplementary Figure 3.** UV chromatogram of SCX separations of a mixture of tryptic peptides from bovine serum albumin and caseins. Red colored sequences are non-phosphorylated peptides and blue colored sequences are 'basic' phosphopeptides. **(a)** SCX separation at a pH of 3. The peptides with the mentioned sequences coelute in the red plus blue colored region. **(b)** SCX separation at a pH of 1. The regular peptides elute in the red colored region and the 'basic' phosphopeptides elute later in the blue colored region.




**Supplementary Figure 4.** Comparison of the elution of regular peptides (red) and phosphopeptides (blue) after SCX separation at a pH of 1. Fraction 10 (fr10) to fraction 15 (fr15) from the SCX separation at a pH of 3 of digested HeLa lysate were subsequently separated by SCX at a pH of 1. The fraction numbers are plotted against the amount of unique peptides per SCX separation. Unique peptides (per SCX separation) were assigned to the fraction in which they exhibited the greatest peak area.




**Supplementary Figure 5.** Analysis of the accumulated amount of unique peptides in the single SCX and tandem SCX analysis. Most regular peptides were identified with tandem SCX (red squares). In fraction 10 almost the same amount of phosphopeptides were identified with the tandem SCX (blue diamonds) approach compared to regular peptides identified with the single SCX analysis (violet crosses). The amount of phosphorylated peptides detected with the single SCX analysis is negligible (green triangles) compared to the amount detected by tandem SCX (blue diamonds).

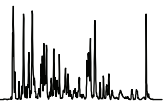



**Supplementary Figure 6.** An iceLogo created from the type 1 'basic' phosphopeptides dataset displays the under- or over-representation of amino acids surrounding the phosphorylation site. The position 6 is the phosphorylation site. The amino acid frequencies were compared with a reference set that contained type 1 'neutral' phosphopeptides.

## SUPPLEMENTARY TABLE

**Supplementary Table 1.** The table summarizes the detected peptides that elute in fraction 21 and 22 from an SCX separation (pH=3) of tryptic peptides from a bovine serum albumin and bovine casein digest. Letters in lower case indicate phosphorylation (p) or oxidation (ox). The shift in retention indicates how many minutes earlier the peptide elutes, when separated with SCX at a pH of 1, using the identical setup and salt gradient. The two 'basic' phosphopeptides (highlighted with green) elute later at a pH of 1 (smaller RT shift) compared to the non-phosphorylated peptides, as visualized in Supplementary Figure 3.

| Sequence | Protein | shift in retention [min] |
|---|---|---|
| ATEEQLK | Serum albumin | 15 |
| DDSPDLPK | Serum albumin | 14 |
| EACFAVEGPK | Serum albumin | 15 |
| ETYGDMADCCEK | Serum albumin | 15 |
| ETYGDMoxADCCEK | Serum albumin | 15 |
| EYEATLEECCAK | Serum albumin | 15 |
| KTVDMoxESpTEVFTK | Alpha-S2-casein | 8 |
| NAVPITPTLNR | Alpha-S2-casein | 15 |
| NMAINPSK | Alpha-S2-casein | 15 |
| QNCDQFEK | Serum albumin | 15 |
| TVDMoxESTEVFTK | Alpha-S2-casein | 15 |
| VLPVPQK | Beta-casein | 14 |
| VPQLEIVPNSAEER | Alpha-S1-casein | 15 |
| YICDNQDTISSK | Serum albumin | 15 |
| YKVPQLEIVPNSpAEER | Alpha-S1-casein | 7 |
| YNGVFQECCQAEDK | Serum albumin | 15 |



## REFERENCES


(1) Dunn, J. D.; Reid, G. E.; Bruening, M. L. *Mass Spectrometry Reviews* **2010**, *29*, 29-54.
(2) Thingholm, T. E.; Jensen, O. N.; Larsen, M. R. *Proteomics* **2009**, *9*, 1451-1468.
(3) Zhou, H.; Low, T. Y.; Hennrich, M. L.; van der Toorn, H.; Schwend, T.; Zou, H.; Mohammed, S.; Heck, A. J. R. *Molecular & Cellular Proteomics* **2011**, *10*.
(4) Zhou, H. J.; Xu, S. Y.; Ye, M. L.; Feng, S.; Pan, C.; Jiang, X. G.; Li, X.; Han, G. H.; Fu, Y.; Zou, H. *Journal of Proteome Research* **2006**, *5*, 2431-2437.
(5) Nuhse, T. S.; Stensballe, A.; Jensen, O. N.; Peck, S. C. *Molecular & Cellular Proteomics* **2003**, *2*, 1234-1243.
(6) Villen, J.; Gygi, S. P. *Nature Protocols* **2008**, *3*, 1630-1638.
(7) Pinkse, M. W. H.; Mohammed, S.; Gouw, L. W.; van Breukelen, B.; Vos, H. R.; Heck, A. J. R. *Journal of Proteome Research* **2008**, *7*, 687-697.
(8) Thingholm, T. E.; Jensen, O. N.; Robinson, P. J.; Larsen, M. R. *Molecular & Cellular Proteomics* **2008**, *7*, 661-671.
(9) Ficarro, S. B.; Parikh, J. R.; Blank, N. C.; Marto, J. A. *Analytical Chemistry* **2008**, *80*, 4606-4613.
(10) Kyono, Y.; Sugiyama, N.; Imami, K.; Tomita, M.; Ishihama, Y. *Journal of Proteome Research* **2008**, *7*, 4585-4593.
(11) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J. X.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 12130-12135.
(12) Gauci, S.; Helbig, A. O.; Slijper, M.; Krijgsveld, J.; Heck, A. J. R.; Mohammed, S. *Analytical Chemistry* **2009**, *81*, 4493-4501.
(13) Hennrich, M. L.; Groenewold, V.; Kops, G.; Heck, A. J. R.; Mohammed, S. *Analytical Chemistry* **2011**, *83*, 7137-7143.
(14) Zhou, H. L.; Watts, J. D.; Aebersold, R. *Nature Biotechnology* **2001**, *19*, 375-378.
(15) Taouatas, N.; Altelaar, A. F. M.; Drugan, M. M.; Helbig, A. O.; Mohammed, S.; Heck, A. J. R. *Molecular & Cellular Proteomics* **2009**, *8*, 190-200.
(16) Frese, C. K.; Altelaar, A. F. M.; Hennrich, M. L.; Nolting, D.; Zeller, M.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. *Journal of Proteome Research* **2011**, *10*, 2377-2388.
(17) Boersema, P. J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J. R. *Nat. Protocols* **2009**, *4*, 484-494.
(18) Schwartz, D.; Gygi, S. P. *Nature Biotechnology* **2005**, *23*, 1391-1398.
(19) Colaert, N.; Helsens, K.; Martens, L.; Vandekerckhove, J.; Gevaert, K. *Nature Methods* **2009**, *6*, 786-787.
(20) Miller, M. L.; Jensen, L. J.; Diella, F.; Jorgensen, C.; Tinti, M.; Li, L.; Hsiung, M.; Parker, S. A.; Bordeaux, J.; Sicheritz-Ponten, T.; Olhovsky, M.; Pasculescu, A.; Alexander, J.; Knapp, S.; Blom, N.; Bork, P.; Li, S.; Cesareni, G.; Pawson, T.; Turk, B. E.; Yaffe, M. B.; Brunak, S.; Linding, R. *Science Signaling* **2008**, *1*.

# Chapter 7

**Summary**

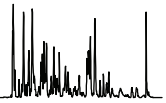
**Samenvatting**

**Zusammenfassung**

**Publications**

**Curriculum Vitae**

**Acknowledgement**



**SUMMARY**

Proteins play a central role in almost all cellular processes. One of the most important techniques to identify and quantify proteins is mass spectrometry. In general, proteins are first digested and the resulting peptides are subjected to analysis. When all proteins of a cell are digested, hundreds of thousands of different peptides are generated. A mass spectrometer cannot analyze all of them at once. Thus, efficient separation technologies are used to feed the mass spectrometer with peptides one after the other.

The major part of this thesis describes new methods to enrich for and separate peptides with specific modifications (phosphorylation) of high biological impact. These techniques enable to enrich, separate and detect over ten thousand phosphorylated peptides in only a subpopulation of all peptides. Besides mass determination of intact peptides mass spectrometers also allow to break chemical bonds by so called fragmentation techniques. The masses of the resulting fragments allow a further detailed characterization of the original molecule. Part of this thesis addresses the behavior of chemically modified (labeled) peptides, when they are subjected to a relative new fragmentation technique called ETD (electron transfer dissociation). We show which of the labels used lead to positive changes in fragmentation and which prevent fragmentation almost completely.

The new methods and fundamental research presented in this thesis will allow improved analysis of proteins and in particular proteins modified by phosphorylation.

**SAMENVATTING**

Eiwitten spelen een centrale rol in alle celgerelateerde processen. Een van the belangrijkste technieken om eiwitten te identificeren en te kwantificeren is massaspectrometrie. In het algemeen worden de eiwitten dan eerst gedigesteerd, waarna de resulterende peptiden worden geanalyseerd. Wanneer alle eiwitten in een cel worden gedigesteerd ontstaan honderdduizenden verschillende peptiden. Een massaspectrometer kan echter die niet allemaal tegelijk analyseren. Daarom zijn efficiënte scheidingstechnieken nodig om de peptiden een voor een met een massaspectrometer te kunnen analyseren.

Het grootste gedeelte van dit proefschrift beschrijft nieuwe methoden voor het verrijken en scheiden van peptiden met een specifieke modificatie (fosforylatie) die betrokken is bij veel biologische processen. Deze technieken maken het mogelijk om meer dan tienduizend gefosforyleerde peptiden te verrijken, scheiden en detecteren in een kleine populatie van alle peptiden. Naast het bepalen van de exacte massa van een peptide is het met massaspectrometrie ook mogelijk om chemische bindingen te verbreken met verschillende fragmentatietechnieken. De massa's van de resulterende fragmenten geven dan gedetailleerde informatie over het originele molecuul.

Een gedeelte van dit proefschrift beschrijft het gedrag van chemische gemodificeerde (gelabelde) peptiden. Deze worden geanalyseerd met een relatieve nieuwe fragmentatie techniek genaamd ETD (electron transfer dissociation). We laten zien welke labels een verbetering van de fragmentie geven en welke juist fragmentatie volledig verstoren.

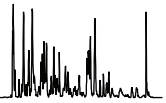
De nieuwe methodes en het fundamenteel onderzoek gepresenteerd in dit proefschrift zullen bijdragen aan een betere analyse van eiwitten en dan met name eiwitten die gemodificeerd zijn door fosforylatie.

**ZUSAMMENFASSUNG**

Eiweiße spielen eine zentrale Rolle in fast allen zellulären Prozessen. Einer der wichtigsten Techniken zur Identifikation und Quantifizierung von Eiweißen ist Massenspektrometrie. Im Allgemeinen werden Eiweiße zuerst verdaut (gespalten) und die daraus entstandenen Peptide gemessen. Wenn alle Proteine einer Zelle verdaut werden, entstehen hundert tausende von Peptiden. Ein Massenspektrometer kann diese nicht alle gleichzeitig messen und daher werden in der Regel hocheffiziente Trenntechnologien eingesetzt um das Massenspektrometer nacheinander mit den unterschiedlichen Peptiden zu füttern.

Ein Großteil dieser Doktorarbeit beschreibt neue Methoden von Trenntechniken um Peptide mit speziellen, biologisch sehr bedeutenden Modifikationen (Phosphorylierung) anzureichern und zu trennen. Hierbei ist es gelungen weit über zehntausend unterschiedliche phosphorylierte Peptide aus einer Teilgesamtheit aller Peptide anzureichern, zu trennen und zu detektieren. Neben der reinen Massenbestimmung von Peptiden erlauben gewisse Massenspektrometer auch chemische Bindungen mit Hilfe sogenannter Fragmentierungsmethoden zu lösen. Die Massen der resultierenden Bruchstücke (Fragmente) erlauben eine nähere Charakterisierung des Ausgangsstoffes. Ein weiterer Teil dieser Arbeit befasst sich damit, das Verhalten von chemisch veränderten (markierten) Peptiden zu untersuchen, wenn diese einer relativ neuen Fragmentierungsmethode mit Namen ETD (electron transfer dissociation) unterzogen werden. Ich konnte zeigen, welche der getesteten Modifikationen sich in welcher Form z.T. positiv auf die Fragmentierung auswirken oder das Trennen vieler Bindungen in Peptiden fast vollständig unterdrücken.

Die in dieser Arbeit entwickelten Methoden und Erkenntnisse werden zu einer verbesserten Analyse von Eiweißen und im Speziellen zur Analyse von phosphorylierten Eiweißen beitragen.



## PUBLICATIONS

Meijer, L. A. T.; Chan, O. Y.; Zhou H.; Altelaar, A. F. M.; **Hennrich, M. L.**; Mohammed, M.; Bos, J. L.; Heck, A. J. R. (2012). "Quantitative global phosphoproteomics of Human Umbilical Vein Endothelial Cells after activation of the Rap signaling pathway." Manuscript in preparation

Cristobal, A.*; **Hennrich, M. L.***; Giansanti P.; Goerdayal, S. S.; Heck, A. J. R.; Mohammed, S. (2012). "In-house construction of a UPLC system enabling the identification of over 4500 protein groups in a single analysis." Submitted

Di Palma, S.*; **Hennrich, M. L.***; Heck, A. J. R.; Mohammed, S. (2012). "Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis." Submitted

Altelaar, A. F. M.*; Frese, C. K.*; Preisinger, C.; **Hennrich, M. L.**; Schram, A. W.; Timmers, H. T. M.; Heck, A. J. R.; Mohammed S. (2012). "Benchmarking stable isotope labeling based quantitative proteomics." Submitted

**Hennrich, M. L.**; van den Toorn, H.; Groenewold, V.; Heck, A. J. R.; Mohammed, S. (2012). "Ultra Acidic Strong Cation Exchange Enabling the Efficient Enrichment of Basic Phosphopeptides." Analytical Chemistry. DOI 10.1021/ac203303t

Frese, C. K.; Altelaar, A. F. M.; **Hennrich, M. L.**; Nolting, D.; Zeller, M.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. (2011). "Improved Peptide Identification by Targeted Fragmentation Using CID, HCD and ETD on an LTQ-Orbitrap Velos." Journal of Proteome Research **10**(5): 2377-2388.

**Hennrich, M. L.**; Groenewold, V.; Kops, G.; Heck, A. J. R.; Mohammed, S. (2011). "Improving Depth in Phosphoproteomics by Using a Strong Cation Exchange-Weak Anion Exchange-Reversed Phase Multidimensional Separation Approach." Analytical Chemistry **83**(18): 7137-7143.

Zhou, H.; Low, T. Y.; **Hennrich, M. L.**; van der Toorn, H.; Schwend, T.; Zou, H.; Mohammed, S.; Heck, A. J. R. (2011). "Enhancing the Identification of Phosphopeptides from Putative Basophilic Kinase Substrates Using Ti (IV) Based IMAC Enrichment." Molecular & Cellular Proteomics **10**(10).

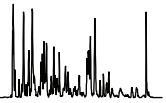
**Hennrich, M. L.**; Mohammed, S.; Altelaar, A. F. M.; Heck, A. J. R. (2010). "Dimethyl Isotope Labeling Assisted De Novo Peptide Sequencing." Journal of the American Society for Mass Spectrometry **21**(12): 1957-1965.

**Hennrich, M. L.**; Boersema, P. J.; van den Toorn, H.; Mischerikow, N.; Heck, A. J. R.; Mohammed, S. (2009). "Effect of Chemical Modifications on Peptide Fragmentation Behavior upon Electron Transfer Induced Dissociation." Analytical Chemistry **81**(18): 7814-7822.

*These authors contributed equally to this work.

## CURRICULUM VITAE

The author was born on the 16. October 1976 in Buchen, Germany. In 1996, he obtained the Abitur and started to study Nutritional Sciences at the Justus-Liebig-University of Giessen. In 1998, he additionally started to study Agricultural Sciences and Environmental Management at the same University. During his studies, he did internships at BASF (Limburgerhof, Germany), the central laboratory of Milupa (Friedrichsdorf, Germany), and on a pig farm (Boenen, Germany). In addition, he worked as musician, in the forestry, and for the University of Giessen at the Institute of Plant Nutrition, and at the Institute of Phytopathology and Applied Zoology. In 2003, he received his diploma in Nutritional Science and started to work as a chemical engineer at Altana Pharma AG in the Department of Metabolism and Pharmacokinetics. He graduated his bachelor of science in Agricultural Sciences and Environmental Management in 2004. In 2007, he left Altana Pharma AG and after starting a PhD position at Oncotest GmbH (Freiburg, Germany), and the University of Tübingen (Germany), he moved to Utrecht University, where he did his PhD training in the Biomolecular Mass Spectrometry and Proteomics group of Prof. Dr. A. Heck. Here, he did the research described in this thesis.



**ACKNOWLEDGEMENTS**

I want to thank the following people.

Monique Slijper
Onno Bleijerveld
Hans, Katrin, Johanna, Franz & Emil
Nina Hubner
Christian (Junior) Frese
Nadine Binai
Danny Navarro
Renske Raaphorst
**Corine Heuzer**
Nikolai Mischerikow
Sven & Ines
**Nadia Taouatas**
Marlies Paulitschke
Paul Boersema
Natalie Thompson
Tanja & Christian
Renske van Gestel
Eleonara Corradini
Patrick Wijten
Pepijn Burgers
Donna Kovanich
Tina & Tom
Pieter Neerincx
**Hanka**
Erik Graaf
Manrico & Evelyn
Piero Giansanti
Esther van Duijn
Poupak Dadvar
Rebecca Rose
**Glen Shoemaker**
Anton van Dellen
Reinout Raijmakers
Sarah Pelletier
Geert Mommen
Raphael, Heike & Marie
Gideon Oudgenoeg
Erwin Schoof
Salvatore Cappadona
Sara Rosati
Geert Hoogendijk
Hendrik Mertens
Harm Post
Jurriaan, Jorn & Zimon
Serena Di Palma
Henk van den Toorn
Shahram Sepehrkhouy
Gianluca Maddalo
Sharon Gauci
**Meine Eltern**
Houjiang Zhou
Fabio Marino
Hongtao Zhang
Christian (Senior) Preisinger
Geert Kops
Shabaz Mohammed

Joost Snijder
**Albert Heck**
Mahmoud El Azzouny
Ana Guerreiro
Werner Most
Arjen Scholten
Reinhard & Hannelore
Marco Benevento
Kristian & Eric
Arjan Barendregt
Mao Peng
Wilma, Peter, Lena & Daniel
Bernhard, Sylvia, Melanie & Mario
Adja Zoumaro-Djayoon
Martina O'Flaherty
Annegret & Marcus
Başak Kükrer
Benjamin Barasa
**Maria & Carsten**
Bas van Breukelen
Mirjam Damen
Martje Ebberink
Alba Cristobal
Vincent Halim
Arno, Diana & Maximilian
Charlotte Uetrecht
Joost Gouw
**Vincent Groenewold**
Wim Spee
Otto & Daniela
Andreas Helbig
Kristina Lorenzen
Kees Versluis
Lars Meijer
**Tobias, Sabrina, Magnus & Junis**
Ayse Polat
Lucrèce Matheron
Tieneke Schaaij-Visser
Ioana Barbu
Soenita Goerdayal
Maarten Altelaar
Jessica Bereszczak
Teck Low
Javier Muñoz
Thin Thin Aye
Thomas Schwend
Jeff Erickson
Jeffrey Ringrose
Tale Sliedrecht

If you did not find your name on the last two pages, I want to especially thank you as well.
Thanks for the great time!!!