Advancing liquid chromatography-mass spectrometry based technologies for proteome research

ISBN: 978-90-393-5265-6

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

Financial support from the J.E. Jurriaanse Stichting and UIPS for printing this thesis is kindly acknowledged.

Advancing liquid chromatography-mass spectrometry based technologies for proteome research

Avanceren van op vloeistofchromatografie-massaspectrometrie gebaseerde technologieën voor proteoom onderzoek

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. J.C. Stoof, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op woensdag 27 januari 2010 des middags te 12.45 uur

door

Paul Jonathan Boersema

geboren op 22 mei 1982 te Assen

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

Co-promotor: Dr. S. Mohammed

Table of contents

Chapter 1:
General introduction

Paul J. Boersema[1,2], Shabaz Mohammed[1,2] and Albert J. R. Heck[1,2]

[1] 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

[2] Netherlands Proteomics Centre

GENOMICS, TRANSCRIPTOMICS, PROTEOMICS

Proteins carry out or are involved in most, if not all, processes in an organism such as the catalysis of biochemical reactions, cell signaling, DNA and RNA translation and transcription, cell adhesion, metabolism, etc. Proteins are polymeric chains of amino acids that fold into globular forms depending on their amino acid sequence, which is encoded in the sequence of a gene. The structure is important as it often dictates the function. Post translational modifications (PTMs) can affect the structure and, for example, switch the protein into an active state. The molecular analysis of protein and cellular functioning can potentially be performed at the encryption level of the genomic code, at the transcription level, *i.e.* mRNA, or the translation level, *i.e.* the actual proteins.

Traditionally, biological analyses have focused on the study of the role of one or a few components. Yet, more and more, researchers have become aware that a 'systems biology' perspective could potentially provide a more comprehensive insight in cellular processes as most often these processes are orchestrated by several components and networks of proteins rather than a single protein [1]. The allure of the systems biology concept has led to the development of techniques for large scale and high throughput cellular analysis at DNA, RNA and protein level. The analysis at DNA level is generally referred to as genomics and is aimed at determining DNA sequences. Variations in genomic sequences can be determined and some of these polymorphisms have been related to particular diseases or carcinogenesis. A genome represents a stable set of genes that is more or less identical in all cells of an organism, which facilitates the comprehensive sequencing. Full DNA sequences of several (model-)organisms, including human, have been determined.

Transcriptomics, *i.e.* the analysis of relative mRNA expression levels, provides an indication of the genes that are expressed at a given time and could present a more specific image of protein expression in a biological sample in response to certain external environmental conditions. High throughput techniques have been developed, for example, based on DNA microarray platforms that can routinely analyze thousands of gene transcripts. Although the mRNA gives an indication of which genes are active, the mRNA levels have been found to not fully correlate with eventual protein expression levels due to differences in the extent of re-use of mRNA and processing of both mRNA and proteins [2, 3]. Techniques have therefore been developed that allow the direct analysis of proteins in a large scale fashion. This research field is referred to as proteomics and is aimed at determining the protein status ('proteome') for (part of) an organism at a certain time under a certain condition. The proteome can be very complex and dynamic in terms of protein contents, their expression levels and post translation modifications [4, 5]. This information cannot be acquired by genomics or transcriptomics approaches. However, the increase in sample complexity when moving from DNA to protein level analysis poses challenges on the comprehensiveness that can be achieved. Over the last decade, advances in liquid chromatography (LC) and mass spectrometry (MS) have enabled and facilitated large scale protein sequencing. Nowadays, complex biological samples can be analyzed almost routinely and several hundreds to thousands of proteins can be identified from a sample. The fundamentally different way of detecting proteins by sequencing by MS removes the limits of, for example, the availability and efficiency of antibodies. Although peptide sequencing was available before by Edman sequencing [6], which can still be useful in certain applications, the sequencing speed of MS allows the large scale analysis of entire and complex protein complements to acquire a deeper biological insight into proteins and their interacting partners and to study their function and their role in cellular processes.

Proteomics workflow

Proteomics is a rather general term that describes the large scale analysis of proteins and there is no such thing as a 'typical' proteomics workflow. The exact approach and sample handling steps depend on the specific research question and available instrumentation. Some 'shotgun' approaches try to comprehensively map protein contents of a sample, while other experiments can be more targeted by precipitating proteins of interest using antibodies or tandem affinity purification tags. Yet others aim at intact proteins and/or complexes. Although some proteomics approaches have been described where no MS is used for the detection of proteins (for example the use of flow cytometry as a readout [7]) in this thesis, proteomics is predominantly considered to be LC-MS based. In Figure 1, an attempt is made to outline a generic proteomics workflow that would apply to most experiments and to my work. Even so, each proteomic experiment described in the following chapters strays from this workflow indicating the level of method design tailoring required for most successful analysis.

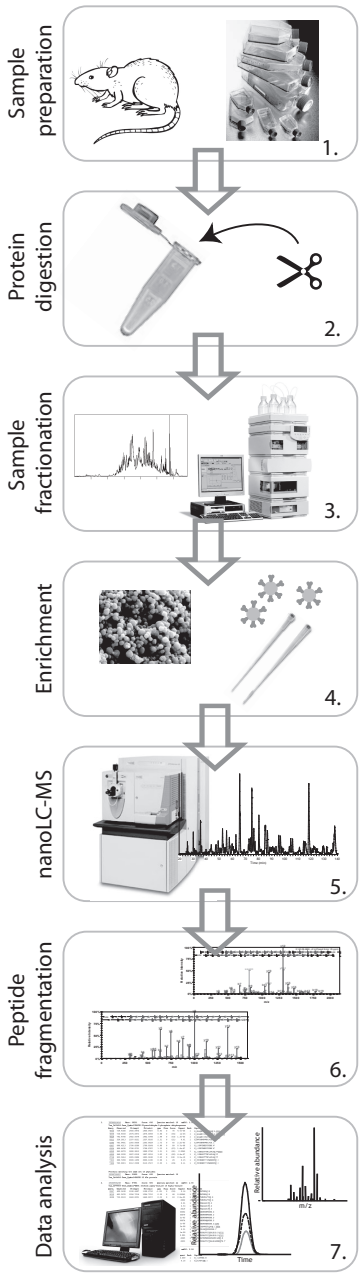


Figure 1. *Typical proteomics workflow*

The sample preparation (Step 1) in a proteomics experiment depends largely on the type of sample and the ultimate research goal. The sample can be tissue, isolated primary cells or cells from laboratory cultures, plant material etc., while the goal could be a comprehensive profiling of the proteome of the sample, or a more targeted analysis on certain subcellular compartments or the detection of certain PTMs. Also, whether or not the study should be quantitative will have a great impact on the chosen path. In general, the sample is homogenized and lysed in urea or SDS containing buffers. The use of such reagents also aid protein denaturation and solubilisation. The next step is the digestion of the proteins with proteolytic enzymes to generate peptides that are in the molecular weight window for optimal MS analysis (Step 2). The reduction of proteins to peptides has a number of analytical benefits such as the removal of protein interactions. Furthermore, liquid chromatography based separations of peptides are superior to protein separations [8]. One of the most often used proteolytic enzymes is trypsin. This enzyme cleaves proteins very specifically and efficiently at the C-terminal side of Lys and Arg residues [9]. These basic residues are generally adequately distributed over a protein amino acid sequence and their frequency results in peptides with a close to ideal length for subsequent MS analysis. Also, peptides containing a basic residue at a terminus fragment better in MS/MS. Other enzymes used in proteomics are Lys-C (cleaves C-terminal of Lys), which also works at higher urea concentrations than trypsin; V8 (C-terminal of Asp

and Glu), chymotrypsin, (C-terminal of Tyr, Trp and Phe), and Lys-N (N-terminal of Lys).
A biological sample typically consists of tens of thousands of proteins. Digesting these proteins to peptides causes the complexity of the sample to be enormously increased, which might pose a problem for MS. A mass spectrometer isolates and sequences peptides one by one, which is most effectively performed when the complexity of peptides injected at a time is low. The complexity of a sample should therefore be reduced for the mass spectrometer to analyze as many peptides as possible. Currently, this is often performed by fractionation of the sample using LC (Step 3). Several different LC modes have been applied to fractionate peptides, for example, based on hydrophobicity, charge state, size, hydrophilicity or phosphorylation state. A more detailed discussion of different LC phases and their respective applications can be found in section 1-B.
It can be important to determine phosphorylation sites and the levels of phosphorylation in, for instance, signaling pathway studies. These levels are generally substoichiometric, which might result in phosphorylated peptides in complex samples being masked by more abundant and ubiquitous non-modified peptides. The poorer detectability by LC-MS of phosphorylated peptides compared to their non-modified counterparts augments the challenge. To be able to detect and sequence phosphorylated peptides, several enrichment strategies have been developed to isolate these peptides either after or before sample fractionation (Step 4). Often, these enrichment strategies are based on chelation such as Immobilized Metal Affinity Chromatography (IMAC, often using Fe^{3+} or Ga^{3+}) [10, 11] and metal oxide affinity chromatography (TiO_2 [12] or ZrO_2 [13]). A discussion of phosphopeptide enrichment strategies can be found in section 1-B-III.
In the next step, LC-MS (Step 5), peptides are separated by reversed phase chromatography before being sprayed into the mass spectrometer. The mass spectrometer determines the m/z of the peptides eluting from the LC column and after being electrosprayed into the MS. The instrument can then select and isolate a peptide ion for fragmentation (Step 6) in order to establish the amino acid sequence. The peptide coverage that can be obtained from a biological sample depends on the complexity of the sample. However, the specifications of LC-MS instrumentation such as power and length of the LC separation and the sensitivity, dynamic range of detection, mass accuracy, scanning speed and fragmentation type of the mass spectrometer also determine the comprehensiveness one can achieve. A more elaborate discussion on MS, fragmentation and types of mass spectrometers can be found in section 1-C.
Quantification in proteomics is, nowadays, often performed by stable isotope labeling (see section 1-E). These stable isotopes can be incorporated at several moments in the proteomics pipeline, either metabolically, for example, during cell culturing or chemically at the protein or peptide level. In section 1-E technical aspects of stable isotope labeling for quantitative LC-MS are discussed.
In the final step, data analysis (Step 7), bioinformatics tools are employed to identify proteins by matching mass spectra to peptides, perform quantitative analysis, carry out clustering and pathway/network analyses etc. Peptide sequences can be automatically determined by matching the observed fragmentation spectra with theoretical fragmentation spectra generated *in silico* from genomic databases. As these databases can be rather large, robust statistical tools need to be applied in the form of peptide matching scores, confidence thresholds and false discovery rates (FDR; see section 1-D). Also, MS based quantification requires bioinformatic tools that can detect peaks and extract and integrate ion current chromatograms from the MS data. Furthermore, datasets that are acquired in proteomics studies are generally very large and therefore require specialized tools that can handle these amounts of data and preferably aid in making biological sense out of the identified proteins, their expression

dynamics and observed PTMs. A plethora of tools exists for clustering and pathways analyses, annotation of proteins and database management systems for easier bookkeeping.

In the following section several steps in the proteomics workflow are discussed in more detail as advances to these steps have been developed in the work presented here or they are key for the understanding of the impact of work described in this thesis.

(TWO DIMENSIONAL-) LIQUID CHROMATOGRAPHY

(nano)-liquid chromatography

LC is routinely used in proteomics to reduce the complexity of peptide mixtures. Several stationary phases have been used for the separation of peptides such as reversed phase (RP), strong cation exchange (SCX), strong anion exchange (SAX) and hydrophilic interaction liquid chromatography (HILIC). These phases separate peptides based on hydrophobicity, positive charges, negative charges and hydrophilicity, respectively. RP is most often used in proteomics studies as RP has the highest resolving power available and it is directly compatible with ESI-MS or indirectly with MALDI when the eluent is mixed with a MALDI matrix [14].

LC separations in current LC-MS configurations are often performed at nano-scale as nanoliter per minute flow rates (20-200 nl/min) are vital for an efficient electrospray process [15]. The LC systems have been miniaturized to allow for such low flow rates by using smaller inner diameter columns of typically 75 μm. In our hands 50 μm id columns have shown to perform optimally, while in some laboratories columns with id as small as 25 μm are used. To accommodate typical sample volumes of several microliters on a nanoLC system, a precolumn can be used to concentrate the sample at flow rates of several μL/min. Afterwards the sample is eluted from the precolumn onto the analytical column at a nanoliter flow rate [16].

Two dimensional liquid chromatography

A typical proteomics sample is often too complex to be sufficiently fractionated by a single RP separation. Several attempts have been made to improve single-dimension separations by using longer columns (of up to 70 cm) and increasing the gradient length (up to 10 h) [17] to increase the peak capacity of the LC separation. However, these approaches typically require LC instrumentation that operate at ultrahigh pressures (in excess of 1000 bar). An alternative way to further reduce the complexity of a sample is to use different LC separations in tandem, also called two dimensional LC (2D-LC). One of the most often used 2D-LC systems in proteomics combines SCX in the first dimension with RP in the second dimension and is often referred to as multidimensional protein identification technology (MudPIT)[18, 19] when a column is packed with consecutively SCX and RP material. Peptides bind initially to the SCX phase and are eluted onto the RP section by steps of increasing salt concentration. The peptides bound to the RP section are then separated using a conventional gradient of increasing acetonitrile concentration and the peptides are sprayed directly into the mass spectrometer. This setup of 2D-LC is referred to as 'online' since the SCX and RP separation are directly connected to the MS. 'Offline' hyphenation of SCX and RP has also been extensively used. The sample is first fractionated over the SCX column after which the fractions are analyzed separately by RP LC-MS (Figure 2). While online hyphenation has certain advantages over offline hyphenation, such as automation, minimal sample loss, no sample dilution and no vial contamination [20, 21], offline 2D-LC allows more flexibility. For example, LC phases requiring solvents that are not directly compatible with each other

or with a mass spectrometer can be used. Additionally, certain fractions can be selected for further and in depth analysis and also a conventional gradient can be used for optimal chromatographic performance in the first dimension [22, 23].

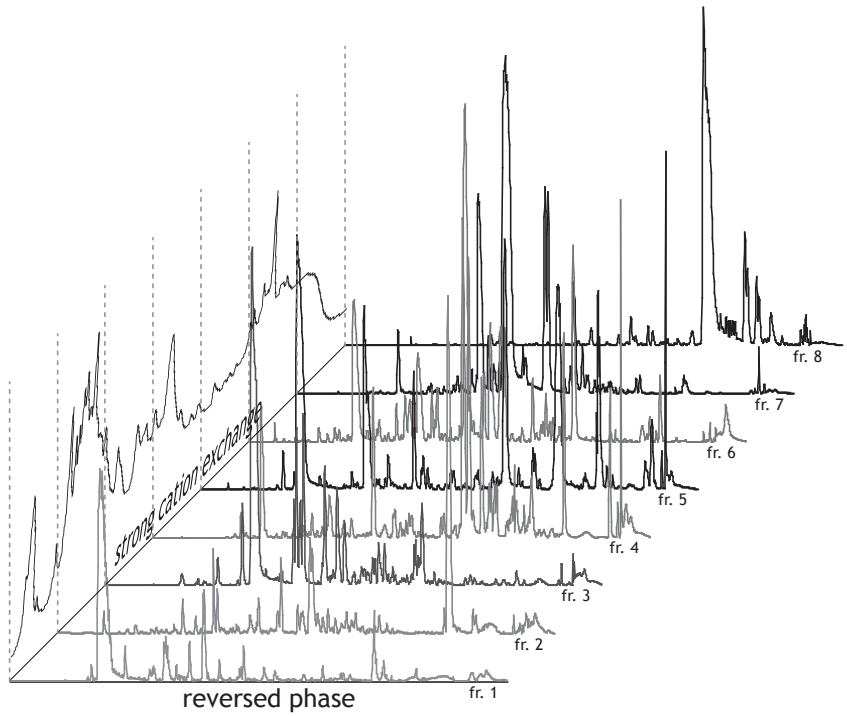


Figure 2. *2D-LC separation of a complex proteomics sample. The sample is fractionated in the first dimensional SCX separation. The fractions are then analyzed separately by LC-MS to increase the proteomics coverage.*

Several other phases have been used in 2D-LC configurations, such as size exclusion chromatography (SEC)[24], strong anion exchange (SAX)[25], RP at pH 10 [26] and hydrophilic interaction liquid chromatography (HILIC)[27, 28]. These LC phases are typically used in the first dimension while the second dimension is RP. There are several key factors for the selection of the phases in a 2D-LC system. The most important factor is the orthogonality of separation of the two different phases [28]. The modes of separation should be non-correlative in order to obtain the highest peak capacity. Another factor that determines the success of a 2D-LC system is the separation power [29]. When the separation power of an LC-phase is too low in the first dimension, peptides might elute over several fractions which will result in the same peptides being identified several times while lower abundant peptides will not be sequenced. SCX shows an adequate separation power and the mode of separation is orthogonal to RP, *i.e.* the separation is based on positive charges and hydrophobicity, respectively. It is therefore widely used in proteomics. Furthermore, it has been found that low pH SCX can enrich for phosphorylated and N-acetylated peptides [30-32]. At pH 3, the only entities on a peptide that carry a positive charge are the basic Lys, Arg and His residues and the peptide N-terminus. Acidic residues are protonated at this pH, but the pK_a of a phosphorylated residue is lower than Asp and Glu and phosphorylated residues are therefore negatively charged. The negative charge of the phosphate group reduces the net charge of the peptide with one charge and phosphopeptides elute therefore typically before the bulk of

non-phosphorylated peptides. N-acetylation neutralizes the basic N-terminus and decreases thereby also the net charge of the peptide and also reduces the retention of N-acetylated peptides. Another LC phase that has been shown to be interesting for use as a first dimension in 2D-LC is HILIC. The separation mechanism by HILIC is considered to be a mixed mode of electrostatic and polar interactions [27]. Peptides are found to elute more evenly dispersed over the whole run, unlike SCX where the non-modified peptides with typically a charge state of 2+ or 3+ cluster together. Furthermore, the additional negative charge of the phosphate group causes phosphopeptides to be more hydrophilic and increases their retention on a HILIC column. HILIC has therefore been shown to be useful in the enrichment of phosphopeptides [33]. N-acetylated peptides with the N-terminal basic group neutralized are less hydrophilic and these peptides elute before the non-modified peptides [27].

Phosphopeptide enrichment

The detection of phosphopeptides by LC-MS can be challenging because phosphorylated peptides in complex samples can be masked by more abundant non-modified peptides due to the low stoichiometry of phosphorylation and the poorer detectability of phosphopeptides by LC-MS [34, 35]. One way to accommodate these challenges is to specifically enrich for phosphopeptides. SCX and HILIC have been mentioned above as ways to enrich phosphopeptides [30, 33]. However, in SCX, phosphopeptides containing an additional basic residue still elute with the bulk of non-modified peptides. Also, both N-acetylated peptides and protein C-terminal peptides (that do not contain Lys or Arg residues) have a similar net charge as phosphopeptides and co-elute. Several affinity enrichment strategies have been developed that isolate phosphopeptides of which IMAC [11] and TiO_2 [12] are presently the most often utilized methods. The selectivity of IMAC is based on the efficient chelation of the phosphate group to a metal (for example Fe^{3+} or Ga^{3+}) immobilized to the resin. Phosphopeptides are loaded onto the IMAC material at low pH and can be eluted again by increasing the pH [10]. As most non-phosphorylated peptides are not retained on the IMAC material, the elution is highly enriched for phosphopeptides. However, clusters of acidic amino acid residues can also chelate to some degree to the IMAC metal and therefore acidic and non-phosphorylated peptides might bind and co-elute. It has been shown that O-methylesterification of the Asp and Glu residues reduces some of this non-specific co-enrichment of acidic peptides [11]. However, the derivatization step often results in significant sample loss and has potentially detrimental side effects, such as incomplete derivatization and modification of other amino acid residues.

The specificity of TiO_2 for phosphopeptides has been suggested to be based on Lewis acid–base interactions and is considered to be more specific than IMAC [36]. Unspecific binding of non-phosphorylated peptides can be further reduced by using, for example, 2,5-dihydroxybenzoic acid, phthalic acid, glycolic acid or formic acid in the loading buffer [37]. IMAC and TiO_2 enrichment is most often performed offline using microcolumns, but online approaches have also been described for TiO_2 as binding to TiO_2 is fast and buffers are compatible with RP LC-MS [12, 38]. In the optimized online approach a triple stage precolumn is fabricated that consists of a RP part, a TiO_2 section and again a RP section (Figure 3)[38]. In such a configuration, all peptides in a sample are first trapped on the first RP section. Elution from the first RP section is achieved by increasing the acetonitrile concentration moving the phosphorylated peptides to the TiO_2 section while non-phosphorylated peptides flow through and are separated by the analytical column and analyzed by MS. The phosphorylated peptides are then eluted from the TiO_2 section onto the second RP part by increasing the solvent pH to above 9 followed by an acetonitrile gradient to separate the phosphorylated peptides on

the analytical column. The online affinity enrichment diminishes sample loss and allows automation of the LC-MS analysis and is therefore particularly suited for the analysis of large batches of samples, for example, originating from SCX fractions.

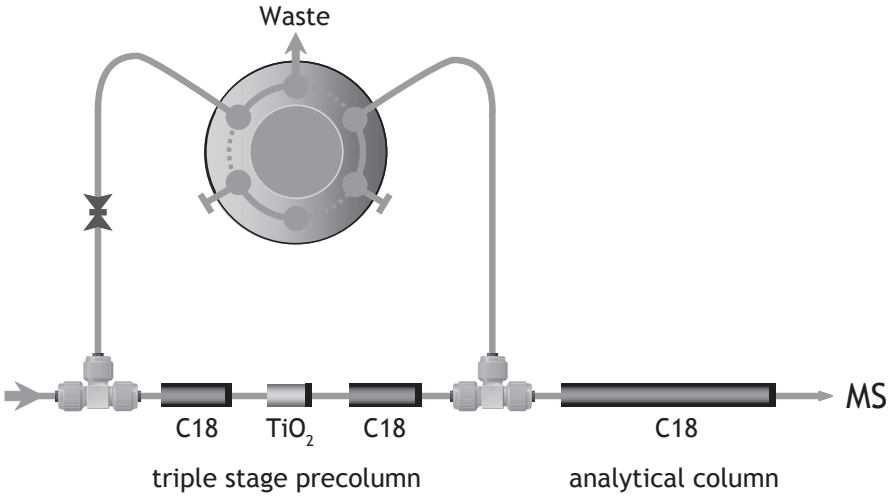


Figure 3. *Illustration of an online automated TiO2 system using as triple stage precolumn comprising consecutively a C18, TiO2 and another C18 precolumn.*

The analysis of Tyr phosphorylation is considered more challenging compared to phosphorylated Ser and Thr because of the much lower cellular levels of Tyr phosphorylation. However, Tyr phosphorylation plays a very important role in several cellular processes and is often involved in the initiation of cell signaling pathways [39]. Strategies have therefore been developed to enrich for Tyr phosphorylation by immunoprecipitation using antibodies with specificity for phosphorylated Tyr [40-42]. Similar antibodies have also been developed for the immuno-affinity enrichment of Ser and Thr phosphorylated proteins and peptides. However, these antibodies typically lack an adequate specificity to efficiently pull down only these species from a complex sample [43].

Immunoprecipitation at the protein level followed by digestion and LC-MS analysis can be used to determine which proteins are Tyr phosphorylated. However, digestion of these precipitated proteins results in a mixture of Tyr phosphorylated and non-phosphorylated peptides. It can therefore be difficult to determine the exact sites of phosphorylation and particularly the dynamics of individual phosphorylation sites as the peptides containing the phosphorylation might not always be detected. Also, proteins that bind to Tyr phosphorylated proteins might be co-precipitated. Immunoprecipitation is therefore nowadays often performed at the peptide level in order to derive a sample that predominantly contains Tyr phosphorylated peptides [44, 45]. Another advantage of performing the immunoprecipitation at the peptide level is that chemical stable isotope labeling for quantitation can be performed prior to immunoprecipitation so that potential variation caused by this sample handling step can be diminished [46].

MASS SPECTROMETRY

MS has become the method of choice for the identification of peptides –and therefore proteins- in proteomics. Based on the *m*/*z* of the peptide ion combined with the *m*/*z* of ions obtained by fragmenting the precursor ion in MS/MS a peptide can be identified since the

products of the fragmentation process are generally predictable. The detection and sequencing of proteins is not dependent on the availability of, for example, antibodies. Theoretically any protein can therefore be identified by MS making it an essential instrument in proteomics to comprehensively map the proteomic contents of a sample.
A mass spectrometer can be described to consist of four parts. A sample is introduced in the mass spectrometer through the ionization source. The compounds are separated according to their *m*/*z* in the mass analyzer and captured by the detector. Finally, compounds can be fragmented in a collision cell that can be tandem-in-space or –time, after which the product ions are measured in the mass analyzer and captured by the detector.

Ionization
Analysis in a mass spectrometer is performed under vacuum. To introduce a biological sample into a mass spectrometer the compounds must be converted to gas-phase ions which, in proteomics, is typically performed by matrix assisted laser desorption/ionization (MALDI) or electrospray ionization (ESI).
In MALDI, the biological sample is mixed with a matrix and co-crystallized on a MALDI target plate by vaporization of the solvents [47, 48]. A laser is fired at the sample to ionize the matrix molecules. The charge is (partially) transferred to the compounds which results in predominantly singly protonated peptide ion species.
In ESI, the analyte is dissolved in a liquid solvent and dispersed to a fine aerosol through electrostatic charging [49]. In this stage, the volatile solvent in the droplet evaporates, sometimes aided by heating or use of an inert nebulizing gas. As the solvent evaporates, the charge density of the droplets increases until the so called Rayleigh point is reached. At this juncture, Coulombic repulsion exceeds droplet surface tension and fission occurs [50]. Two models have been proposed to explain the final generation of gas-phase ions: the Ion Evaporation Model (IEM)[51] and the Charged Residue Model (CRM)[52]. According to the IEM, field desorption occurs at the surface of the droplet when the field strength of the decreasing droplet is large enough. In CRM, the initial droplet undergoes several cycles of fission until a state is reached at which droplets contain on average one analyte ion or less. The gas-phase ion is formed when all the solvent is evaporated and the ion obtains the charge that the droplet carried. ESI typically produces multiply charged ions, depending on the size of the ion and the amino acid sequence.
The advantage of MALDI over ESI is that MALDI is relatively more tolerant to salts, contaminations and detergents. Also, a sample can be reanalyzed as non-ionized compounds are conserved in the matrix spot. The largest advantage of ESI over MALDI and the reason why it is now the most prominent ionization technique in proteomics, is that it can be used to hyphenate LC with MS. The throughput is thereby increased. More importantly, however, the constant sample flow of ESI as compared to the pulsed generation of ions in MALDI, allows maximal exploitation of the sequencing speed of current mass spectrometers.

Mass analyzers
In the mass analyzer, ions are separated according to their *m*/*z*. Several types of mass analyzers exist that determine *m*/*z* in different ways. Here, time-of-flight (TOF), linear quadrupole ion trap (LTQ), Fourier transform ion cyclotron resonance (FTICR) and Orbitrap types of mass analyzers will be discussed since these types of mass spectrometers have been used extensively in the following chapters to perform MS analyses.

Conceptually the most straightforward mass analyzer is the TOF type analyzer [53, 54]. In

a TOF analyzer, packages of ions are accelerated by an electrostatic field to the same kinetic energy. The velocity the ions achieve is dependent on the *m*/*z* of the ion. The higher the *m*/*z* of the ion, the longer it takes to fly through the TOF tube and to arrive at the detector. Several parameters can affect the resolution of mass analysis, such as spatial and initial-energy distributions and metastable ion formation. Implementation of an ion reflectron in the fly path has solved the effect partly [55]. The reflectron consists of a decelerating and reflecting field and compensates for small differences in kinetic energy between ions with the same *m*/*z* thereby achieving mass resolution of up to 50.000 FWHM on current generation TOFs with a mass accuracy of down to 5 ppm [56]. MALDI is often connected to TOF types of mass analyzers as they are well suited for analyzing ion packages such as those generated by the pulsed laser beam (Figure 4A).

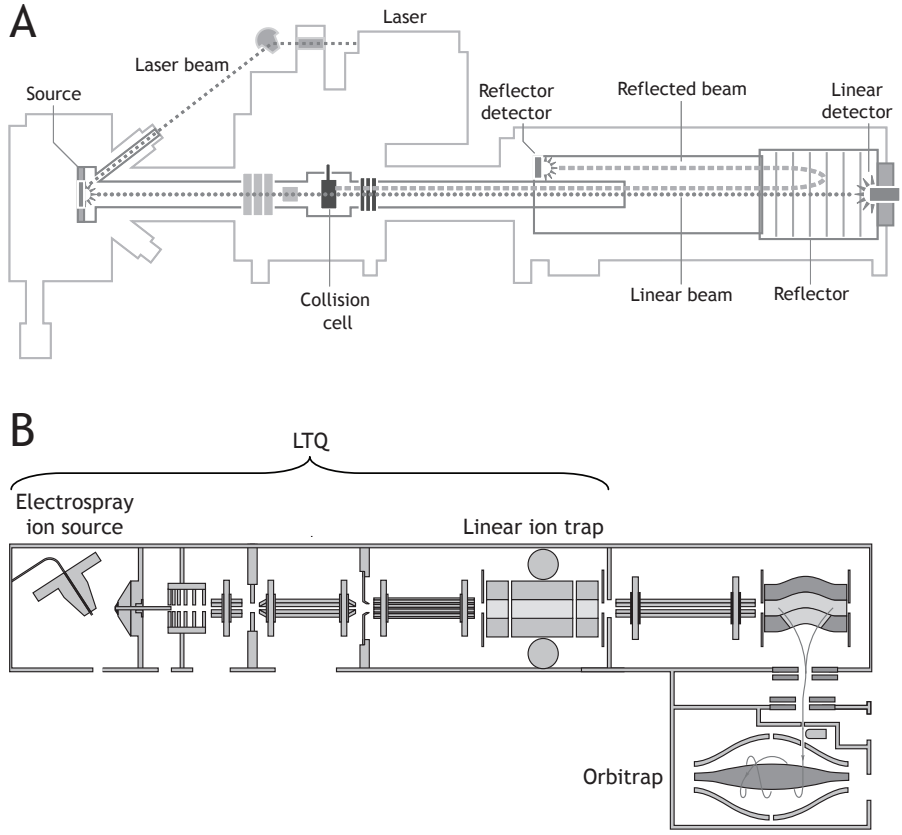


Figure 4. *Schematic representation of A) an ABI 4700 MALDI-TOF/TOF and B) a Thermo LTQ-Orbitrap.*

In an ion trap, ions are captured through a combination of electric or magnetic fields. In an LTQ, for example, that consists of four hyperbolic rods, ions are confined radially in a two-dimensional radio-frequency (RF) field and trapped axially by a static electric potential on end electrodes. Ions can be sequentially ejected from the trap by ramping the RF voltage. Thereby, ions are ejected on the z-axis at a specific RF which depends on their *m*/*z* [57].
Higher mass resolution and accuracy measurements can be performed by using, for example, FTICR or Orbitrap based instruments. The FTICR determines the *m*/*z* of ions based on their cyclotron frequency in a fixed magnetic field [58, 59]. Internal energy transfer to ions

depends on their *m*/*z* and makes the ions move in larger orbits by an oscillating electric field. They are then sensed by the detector plates and the image current is Fourier transformed to obtain a mass spectrum. By FTICR a mass resolution of more than 100.000 FWHM can be achieved with sub-ppm mass accuracy [58, 60]. An Orbitrap consists of an outer barrel-like electrode and a central spindle-like electrode along the axis [61, 62]. Ions cycle radially around this central electrode, but also move back and forth along its axis. The frequency of these axial oscillations is inversely proportional to the square root of the *m*/*z* of the ion. Also here, the detected image current is a superimposition of frequencies of individual ion species and a Fourier transform has to be performed to deconvolute the image and obtain a mass spectrum. In an Orbitrap, a mass resolution of up to 100.000 FWHM can be achieved with a mass accuracy lower than 2 ppm or even in the ppb range when an internal calibration is applied [63]. In proteomics, FTICR and Orbitrap mass spectrometers are typically hybridized with a linear ion trap such as the LTQ (Figure 4B)[60, 64]. MS measurements, where high mass resolution and accuracy is key, are performed in the FTICR cell or Orbitrap, while MS/MS measurements are performed in the LTQ which has a faster scanning speed.

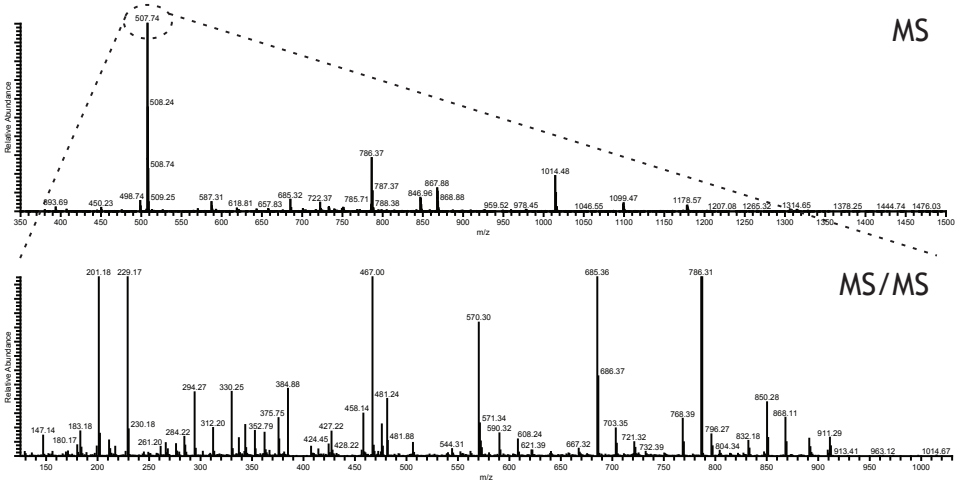


Figure 5. *Tandem MS. In a first round of MS, precursor ion masses are determined. One precursor ion is isolated (m/z 507.74 in this example) and fragmented after which the m/z of the fragment ions is determined in a second round of MS (MS/MS).*

Fragmentation

Nowadays, for accurate peptide sequencing, tandem MS (or MS/MS or MS^2) is performed by isolating the peptide precursor ion followed by fragmenting the peptide ion and analyzing the fragment ions in a second round of MS (Figure 5). The fragmentation products are predictable as the process follows more or less certain 'rules'. By measuring not only the intact mass but also the mass of the fragments of a peptide, the amino acid sequence can be determined and isobaric/isomeric peptides can be distinguished.

The most popular means of peptide ion fragmentation in proteomics is collision-induced dissociation (CID; also: collisionally activated dissociation, CAD)[65]. In CID, protonated peptides are accelerated by an electric potential in the vacuum of the mass spectrometer and then allowed to collide with an inert neutral gas (typically helium, nitrogen or argon). Through the collisions, the kinetic energy of the peptide ion is partially converted into internal energy that is distributed over the molecule, disrupting bonds and causing the peptide

ion to fragment. This dissociation is typically initiated by a 'mobile' proton that weakens amide bonds in the peptide backbone [66]. CID can be performed in a collision cell that is tandem-in-space or in an ion trap that is tandem-in-time [67]. Fragmentation in a collision cell typically occurs on a short timescale (microseconds) and can be at a 'high' collision energy (from tens of eV in Q-TOF instruments up to several keVs in MALDI-TOF/TOF). In an ion trap, fragmentation occurs through slow heating whereby the activation time is typically several milliseconds, but at only a few eV collision energy.

CID typically results in the disruption of single amide bonds. However, dependent on the amino acid sequence and the type of mass spectrometer, also additional fragment ions can arise such as immonium ions, internal fragments or peaks that represent the neutral loss of ammonia, water or a phosphate group from a phosphorylated residue. Furthermore, some bonds are more likely to be broken, such as those N-terminal of proline or between acidic residues [68]. CID spectra can therefore be rather complex with incomplete ion series and different ion signal intensities and require bioinformatics tools for interpretion (see next section).

Alternative, electron-driven, dissociation methods such as electron capture dissociation (ECD)[69] and particularly electron transfer dissociation (ETD)[70] are becoming more and more popular in proteomics as they leave PTMs such as phosphorylation and glycosylation intact. In ECD and ETD, the N-Cα bond of the amino acid backbone is cleaved (a more elaborate description of ETD and ECD can be found in Chapter 6).

TANDEM MS-BASED IDENTIFICATION

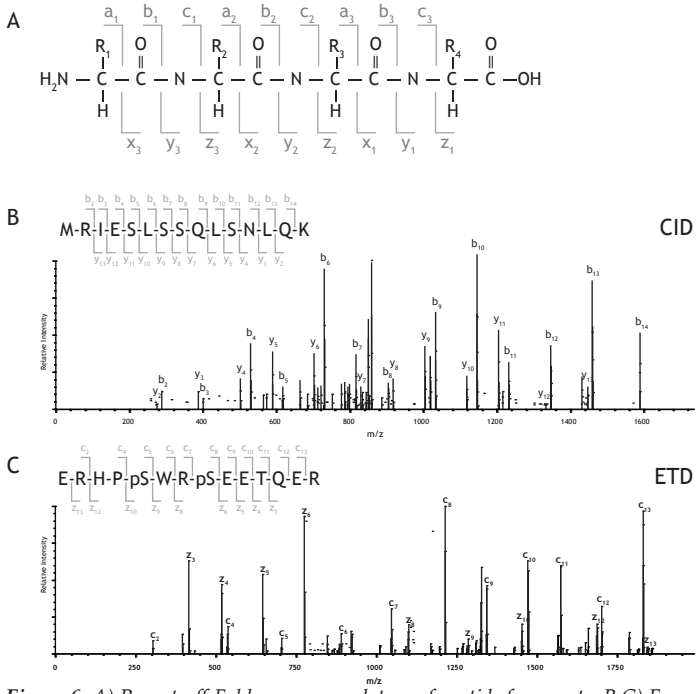


Figure 6. *A) Roepstorff-Fohlman nomenclature of peptide fragments. B,C) Example of a CID (B) and ETD (C) spectrum annotated according to the Roepstorff-Fohlman nomenclature.*

As MS based fragmentation of a peptide is overall rather predictable and occurs more or less randomly over the peptide backbone, a fragmentation spectrum provides generally se-

quence information of the peptide sequence. The amino acid sequence can be established by determining the mass difference between the fragment ions. Confirmation of the amino acid sequence can be obtained as typically both N- and C-terminal fragment ion series can be observed in a fragmentation spectrum. A nomenclature has been established to indicate for the fragment ion whether it is an N- or C-terminal fragment, which bond was broken and at what position in the sequence (see Figure 6)[71, 72]. N-terminal fragments, that can be observed when the N-terminal side has retained a charge after fragmentation, are indicated with a, b or c while C-terminal fragments are indicated with x, y and z. In CID spectra, mostly b and y-ions are observed since most often the amide bond is broken while in contrast mainly c- and z-ions arise in ECD and ETD. A subscript number is used to indicate at what position in the amino acid sequence the bond was cleaved.

Unfortunately, not all fragmentation spectra are straightforward to interpret since spectra can contain many more peaks than just those representing backbone cleavages. Therefore, *de novo* sequencing –*i.e.* determining the amino acid sequence based on the fragmentation spectrum alone- can be a tedious undertaking. To aid in the identification of peptides that generate such spectra and to automate the interpretation of the thousands of spectra that are typically obtained in an LC-MS analysis, several software tools have been developed to match the obtained spectra with theoretical spectra generated from a genomic database. Mascot [73] and Sequest [74] are examples of such software tools that have been often used. More recent tools include X!tandem [75], Phenyx [76] and OMSSA [77]. The principles and assumptions of each algorithm and their particular scoring functions are slightly different for these programs. However, they all are fundamentally exploiting the same known rules regarding peptide fragments and thus generate similar results. Proteins in a genomic database are *in silico* digested to obtain a set of theoretically observable peptides. Based on the intact mass of the peptide determined by MS, the set of potentially matching candidates can be reduced to those that fall within the mass window that was chosen. The theoretical fragmentation spectra of these potential matches are compared with the obtained spectrum. The number of matching fragment ions and therefore the likeliness of the match is then translated into a peptide identification confidence score.
These database searching programs are all probability based. Usually, a score threshold is therefore used in order to include only reliable peptide identifications in the results. To determine a proper threshold that is at the same time low enough to include most of the true positive identifications, but also high enough to reduce the number of false positives, a search against a database that consists of non-sense sequences can be performed [78, 79]. This so-called decoy database can be constructed for example by reversing or scrambling the protein amino acid sequences in a genomic database. The rationale is that any hit that is returned when searching against this decoy database is most likely a false positive hit in the normal database. The number of false positives at a certain threshold score divided by the number of hits in the normal database gives a FDR. Typically, a peptide score threshold is chosen where the FDR is 1%. However, debate is still ongoing about how to exactly calculate this FDR, and whether these FDRs are really reliable.

MS-BASED QUANTIFICATION

The quantitative analysis of proteomes is becoming more and more feasible in proteomics [80]. In general, it is not the presence of certain proteins that is biologically relevant but how their abundance changes upon a certain effector or stimulus. To reduce the effect of variation in sample preparation and LC-MS conditions, samples should ideally be combined and

analyzed simultaneously. Stable isotope labeling has become the method of choice to allow a concurrent analysis of different samples [81]. The principle behind this technique is that isotopomers behave similarly in sample preparation and LC separation but they can be distinguished in the mass spectrometer based on their mass difference. Relative abundances can be determined by integrating the intensities of the different ion signal intensities. Stable isotopes that have been used include ^{2}H, ^{13}C, ^{15}N and ^{18}O [81]. Often, two samples are compared in one experiment, but multiplex labeling can be advantageous as multiple (in some cases up to eight) samples can be directly compared in a single experiment.

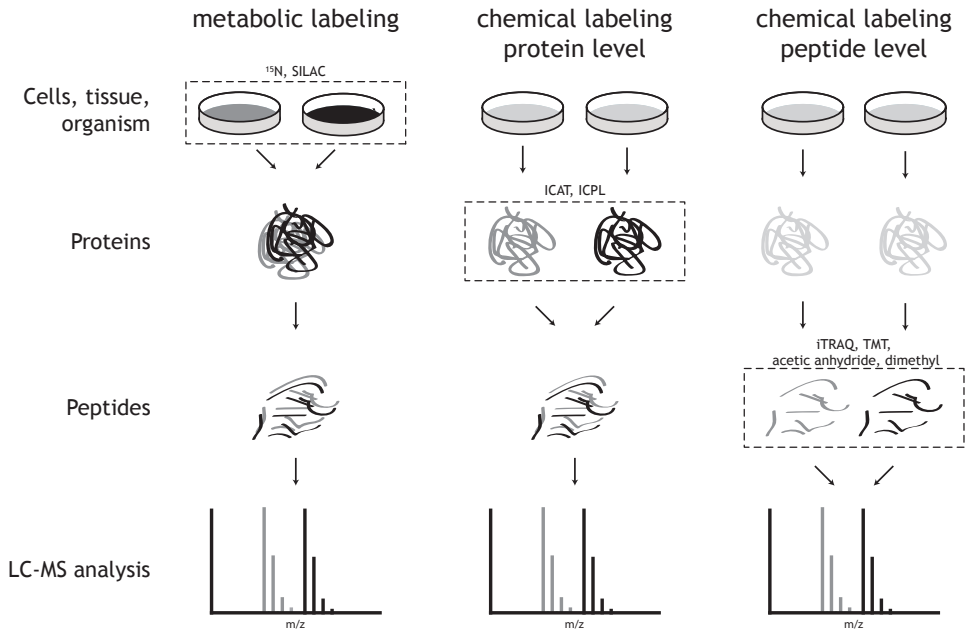


Figure 7. *Introduction of stable isotopes at different levels in the proteomics workflow as indicated by the square boxes. Note: iTRAQ and TMT are isobaric tagging techniques and would therefore not generate peak pairs in the mass spectrum.*

Metabolic labeling

Stable isotope labels can be introduced at several levels in the proteomics workflow (Figure 7). The earliest possible time point is during cell growth when proteins are labeled metabolically by supplying heavy isotopes to a cell culture in such a way that they are exclusively incorporated into newly synthesized proteins [82, 83]. Differentially labeled cell cultures can then be mixed at the earliest stage of sample treatment. Potential variation in lysis, proteolysis, fractionation and LC separation can thus all be minimized. A very popular way of metabolic labeling is by stable isotope labeling with amino acids in cell culture (SILAC)[84]. Hereby, cells are grown in minimal culture that lack essential amino acid(s). These amino acids can then be provided in their natural form or an isotopically labeled form (typically with ^{13}C and/or ^{15}N) in order to obtain *in vivo* incorporation of the isotopically labeled amino acid. Several different amino acids have been used for SILAC, but most popular are Lys and Arg as labeling of these amino acids in combination with trypsin digestion results in all peptides being quantifiable [85]. The mass difference between the differentially labeled peptides depends on the number of isotopically labeled amino acids the peptide sequence contains and which isotopes were applied.

Another way of metabolic labeling is based on ^{15}N [83, 86]. For example, *E. coli* and *S. cerevisiae* can be grown on media containing near exclusively the ^{15}N isotope of nitrogen. In this way, ^{15}N is incorporated into all amino acids. By feeding *D. melanogaster* or *C. elegans* with these respective ^{15}N labeled organisms also higher organisms can be metabolically labeled [86]. Rats have also been metabolically labeled by feeding them with ^{15}N labeled algal cells [87]. In ^{15}N labeling, the difference between the differentially labeled peptides depends on the number of nitrogen atom the peptide contains. Typically, for the interpretation of tandem mass spectra a database is constructed in which all the nitrogens are replaced for ^{15}N to be able to identify the heavy labeled peptides [88].
Metabolic labeling is generally performed on cell systems that are grown in culture, although as mentioned above also nematode worms, fruit flies and rats have been isotopically labeled. However metabolic labeling of higher organisms can be cost prohibitive and also, stable isotope labeling in cell cultures can be expensive as it usually takes several generations to obtain complete labeling. Furthermore, often the growth media have to be changed to allow maximal incorporation of the isotopes, which might have a profound effect on growth and other biological processes [85].

Chemical labeling
An alternative way to incorporate isotopes into a sample and one that is applicable to virtually any sample including human tissue samples, is performing chemical labeling with isotopomeric or isobaric tags [81]. Proteins can, for example, be derivatized using an isotope-coded affinity tag (ICAT)[89]. The advantage is that differently labeled samples can be mixed prior to proteolysis. However, as ICAT is aimed at cysteine residues, only cysteine containing peptides can be used for quantification. Other labeling techniques target the protein and/or peptide N-terminus and the ε-amino group of Lys residues as for example isotope coded protein label (ICPL)[90], acetic anhydride [91] and stable isotope dimethyl labeling [92][93, 94]. The latter option is a particularly cost friendly chemical stable isotope labeling approach. The isotopomeric dimethyl labels can be generated by performing quick and straightforward chemistry using readily available chemicals. Using formaldehyde and cyanoborohydride, the primary amines on the N-terminus of peptides and Lys side chains are dimethyl labeled. By using combinations of isotopomers of formaldehyde and cyanoborohydride triplex labeling can be performed with labels that differ 4 Da in mass [93]. To diminish sample loss, peptides can be labeled on LC columns or online with LC-MS, thereby combining several sample handling steps [94, 95]. Also, because the chemicals are inexpensive, there are no financial constraints with regard to the amount of sample that is to be labeled. Up to several milligrams of sample have been successfully labeled with stable isotope dimethyl labeling [94].
Another popular way to perform chemical isotope labeling is by isobaric tagging of peptides by using labels that consist of a reporter moiety, a balance moiety to counterbalance the differential weight of the reporter moiety and a N-hydroxy-succinimide ester to link the label to the primary amines of a peptide. This approach is used in isobaric tagging for relative and absolute quantification (iTRAQ)[96] and tandem mass tags (TMTs) [97]. In MS, the differentially labeled peptides cannot be distinguished because the balance moiety makes them isobaric. However, upon MS/MS, the reporter moiety is cleaved off and functions as a reporter ion to determine relative abundances. The advantage of the isobaric labels is that the complexity of the mass spectrum is not increased which allows multiplexing of up to 4 and even 8 samples [98] at a time and quantification is straightforward as only the intensities

of the reporter ions have to be compared. However, to observe the reporter ions, specific MS instrumentation is required that is capable of measuring fragment ions at low *m*/*z*, typically TOF mass spectrometers or current day ion traps equipped with options to perform pulsed Q dissociation or higher energy collision induced dissociation [99]. Furthermore, quantification is based on just one tandem fragmentation spectrum. Moreover, the commercially marketed labels are rather expensive.

OUTLINE OF THESIS

The field of proteomics has evolved rapidly over the last decade with substantial advances in particularly nanoLC and MS instrumentation and bioinformatic software tools. However, improvements are still necessary to increase the sensitivity and dynamic range of LC-MS in order to dig deeper into the proteome to identify lower abundant proteins, to advance MS based quantification and to enhance throughput. In this thesis, several technological developments are described to advance proteomics and their applicability is demonstrated in several different research lines.

In chapter 2, the use of HILIC for separation in multidimensional chromatography approaches in proteomics is reviewed. In chapter 3, the advantages of HILIC are evaluated and further optimized by using zwitterionic HILIC (ZIC-HILIC) in a 2D-LC setup. A mixed mode separation was observed for ZIC-HILIC consisting of both electrostatic and polar interactions between the peptides and stationary phase and which can be altered by adjusting the pH of the solvent. Also, less clustering of the typically ubiquitous +2 and +3 charged peptides was observed compared to SCX, rendering ZIC-HILIC an attractive alternative to SCX in multidimensional peptide separations.

In chapter 4, the development of a cost effective triplex stable isotope dimethyl labeling approach is reported. By using different isotopomers of formaldehyde and cyanoborohydride, three different dimethyl labels can be generated in order to simultaneously analyze peptides from three different samples by LC-MS. Several practical aspects, such as the effect of deuterium labeling on the retention on RP and ZIC-HILIC columns, are examined. The chemical labeling method was applied in a typical 'shotgun' proteomics approach and in a more targeted study wherein the method was used to distinguish specific from aspecific cAMP binding proteins when performing pull-down experiments using immobilized cAMP.

In chapter 5, the use of the metalloendopeptidase Lys-N in MALDI-MS/MS proteomics applications is investigated. Lys-N produces peptides with an N-terminal Lys residue. Fragmentation in MALDI of peptides where this N-terminal Lys is the only basic group results in straightforward CID spectra with often complete sequence ladders of b-ions. The extent of this effect was statistically supported by performing Lys-N digestion on a complex cellular lysate. Furthermore, it was found that these straightforward spectra facilitate *de novo* sequencing.

Phosphorylated peptides typically fragment differently compared to non-modified peptides. In CID for example, fragmentation of phosphorylated peptides can result in spectra that are dominated by a single peak that represents the neutral loss of the phosphate group. In chapter 6, this neutral loss effect and ways to accommodate challenges of MS based analysis of phosphopeptides are reviewed. Several alternative fragmentation approaches such as MS^3, multistage activation, ECD and ETD are discussed and their theoretical and actual

benefits for large scale phosphopeptide analysis.

In chapter 7, triplex stable isotope dimethyl labeling is combined with phosphopeptide immuno-affinity purification using phosphoTyr specific antibodies to profile Tyr phosphorylation after pervanadate or EGF stimulation. The stable isotope dimethyl labeling approach allowed the labeling of several milligrams of sample after which an optimized immunoprecipitation procedure purified Tyr phosphorylated peptides. In this way, with just a single LC-MS run, a rather complete qualitative and quantitative image can be generated of Tyr phosphorylation signaling events. In chapter 8, the quantitative immunoprecipitation method is applied to study the role of FGF-2 in the maintenance of the pluripotency state of human embryonic stem cells. Several hundred Tyr phosphorylation sites could be identified and quantified of which some 140 showed a differential regulation upon FGF-2 stimulation. These regulated sites were found on proteins that include FGF receptors, members of the Src family, proteins involved in actin polymerization and cyclin dependent kinases.

REFERENCES

[1] Abbott, A., A post-genomic challenge: learning to read patterns of protein synthesis. *Nature* 1999, *402*, 715-720.
[2] Hegde, P. S., White, I. R., Debouck, C., Interplay of transcriptomics and proteomics. *Curr Opin Biotechnol* 2003, *14*, 647-651.
[3] Gygi, S. P., Rochon, Y., Franza, B. R., Aebersold, R., Correlation between Protein and mRNA Abundance in Yeast. *Mol Cell Biol* 1999, *19*, 1720-1730.
[4] Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., *et al.*, From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Amino Acid Analysis. *Nat Biotechnol* 1996, *14*, 61-65.
[5] Pandey, A., Mann, M., Proteomics to study genes and genomes. *Nature* 2000, *405*, 837-846.
[6] Edman, P., Method for determination of the amino acid sequence in peptides. *Acta Chem Scand* 1950, *4*, 283-293.
[7] Newman, J. R. S., Ghaemmaghami, S., Ihmels, J., Breslow, D. K., *et al.*, Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* 2006, *441*, 840-846.
[8] Aebersold, R., Mann, M., Mass spectrometry-based proteomics. *Nature* 2003, *422*, 198-207.
[9] Olsen, J. V., Ong, S.-E., Mann, M., Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues. *Mol Cell Proteomics* 2004, *3*, 608-614.
[10] Andersson, L., Porath, J., Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography. *Anal Biochem* 1986, *154*, 250-254.
[11] Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., *et al.*, Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat Biotechnol* 2002, *20*, 301-305.
[12] Pinkse, M. W. H., Uitto, P. M., Hilhorst, M. J., Ooms, B., Heck, A. J. R., Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-nanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem* 2004, *76*, 3935-3943.
[13] Kweon, H. K., Hakansson, K., Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Anal Chem* 2006, *78*, 1743-1749.
[14] Shen, Y., Smith, R. D., Proteomics based on high-efficiency capillary separations. *Electrophoresis* 2002, *23*, 3106-3124.
[15] Chervet, J. P., Ursem, M., Salzmann, J. P., Instrumental Requirements for Nanoscale Liquid Chromatography. *Anal Chem* 1996, *68*, 1507-1512.
[16] Meiring, H. D., van der Heeft, E., ten Hove, G. J., de Jong, A., Nanoscale LC-MS(n): technical design and applications to peptide and protein analysis. *J Sep Sci* 2002, *25*, 557-568.
[17] Luo, Q. Z., Shen, Y. F., Hixson, K. K., Zhao, R., *et al.*, Preparation of 20-mu m-i.d. silica-based monolithic columns and their performance for proteomics analyses. *Anal Chem* 2005, *77*, 5028-5035.
[18] Washburn, M. P., Wolters, D., Yates, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001, *19*, 242-247.
[19] Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., *et al.*, Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 1999, *17*, 676-682.
[20] Peng, J. M., Elias, J. E., Thoreen, C. C., Licklider, L. J., Gygi, S. P., Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: The yeast proteome. *J Proteome Res* 2003, *2*, 43-50.
[21] Wagner, K., Miliotis, T., Marko-Varga, G., Bischoff, R., Unger, K. K., An automated on-line multidimensional HPLC system for protein and peptide mapping with integrated sample preparation. *Anal Chem* 2002, *74*, 809-820.
[22] Vollmer, M., Horth, P., Nagele, E., Optimization of two-dimensional off-line LC/MS separations to improve resolution of complex proteomic samples. *Anal Chem* 2004, *76*, 5180-5185.
[23] Nagele, E., Vollmer, M., Horth, P., Vad, C., 2D-LC/MS techniques for the identification of proteins in highly complex mixtures. *Exp Rev Proteomics* 2004, *1*, 37-46.
[24] Opiteck, G. J., Jorgenson, J. W., Anderegg, R. J., Two-dimensional SEC/RPLC coupled to mass spectrometry for the analysis of peptides. *Anal Chem* 1997, *69*, 2283-2291.
[25] Dai, J., Jin, W. H., Sheng, Q. H., Shieh, C. H., *et al.*, Protein phosphorylation and expression profiling by Yin-Yang multidimensional liquid chromatography (Yin-Yang MDLC) mass spectrometry. *J Proteome Res* 2007, *6*, 250-262.
[26] Gilar, M., Olivova, P., Daly, A. E., Gebler, J. C., Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J Sep Sci* 2005, *28*, 1694-1703.
[27] Boersema, P. J., Divecha, N., Heck, A. J. R., Mohammed, S., Evaluation and optimization of ZIC-

HILIC-RP as an alternative MudPIT strategy. *J Proteome Res* 2007, *6*, 937-946.
[28] Gilar, M., Olivova, P., Daly, A. E., Gebler, J. C., Orthogonality of separation in two-dimensional liquid chromatography. *Anal Chem* 2005, *77*, 6426-6434.
[29] Gilar, M., Daly, A. E., Kele, M., Neue, U. D., Gebler, J. C., Implications of column peak capacity on the separation of complex peptide mixtures in single- and two-dimensional high-performance liquid chromatography. *J Chromatogr A* 2004, *1061*, 183-192.
[30] Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., *et al.*, Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* 2004, *101*, 12130-12135.
[31] Gauci, S., Helbig, A. O., Slijper, M., Krijgsveld, J., *et al.*, Lys-N and Trypsin Cover Complementary Parts of the Phosphoproteome in a Refined SCX-Based Approach. *Anal Chem* 2009, *81*, 4493-4501.
[32] Dormeyer, W., Mohammed, S., Breukelen, B., Krijgsveld, J., Heck, A. J., Targeted analysis of protein termini. *J Proteome Res* 2007, *6*, 4634-4645.
[33] McNulty, D. E., Annan, R. S., Hydrophilic Interaction Chromatography Reduces the Complexity of the Phosphoproteome and Improves Global Phosphopeptide Isolation and Detection. *Mol Cell Proteomics* 2008, *7*, 971-980.
[34] Reinders, J., Sickmann, A., State-of-the-art in phosphoproteomics. *Proteomics* 2005, *5*, 4052-4061.
[35] Ishihama, Y., Wei, F. Y., Aoshima, K., Sato, T., *et al.*, Enhancement of the efficiency of phosphoproteomic identification by removing phosphates after phosphopeptide enrichment. *J Proteome Res* 2007, *6*, 1139-1144.
[36] Pinkse, M. W. H., Heck, A. J. R., Essential enrichment strategies in phosphoproteomics. *Drug Discov Today* 2006, *3*, 331-337.
[37] Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., Jorgensen, T. J. D., Highly Selective Enrichment of Phosphorylated Peptides from Peptide Mixtures Using Titanium Dioxide Microcolumns. *Mol Cell Proteomics* 2005, *4*, 873-886.
[38] Pinkse, M. W. H., Mohammed, S., Gouw, L. W., van Breukelen, B., *et al.*, Highly robust, automated, and sensitive on line TiO2-based phosphoproteomics applied to study endogenous phosphorylation in Drosophila melanogaster. *J Proteome Res* 2008, *7*, 687-697.
[39] Blume-Jensen, P., Hunter, T., Oncogenic kinase signalling. *Nature* 2001, *411*, 355-365.
[40] Steen, H., Kuster, B., Fernandez, M., Pandey, A., Mann, M., Tyrosine phosphorylation mapping of the epidermal growth factor receptor signaling pathway. *J Biol Chem* 2002, *277*, 1031-1039.
[41] Hinsby, A. M., Olsen, J. V., Bennettt, K. L., Mann, M., Signaling initiated by overexpression of the fibroblast growth factor receptor-1 investigated by mass spectrometry. *Mol Cell Proteomics* 2003, *2*, 29-36.
[42] Pandey, A., Podtelejnikov, A. V., Blagoev, B., Bustelo, X. R., *et al.*, Analysis of receptor signaling pathways by mass spectrometry: Identification of Vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors. *Proc Natl Acad Sci U S A* 2000, *97*, 179-184.
[43] McLachlin, D. T., Chait, B. T., Analysis of phosphorylated proteins and peptides by mass spectrometry. *Curr Opin Chem Biol* 2001, *5*, 591-602.
[44] Zhang, Y., Wolf-Yadlin, A., Ross, P. L., Pappin, D. J., *et al.*, Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics* 2005, *4*, 1240-1250.
[45] Rush, J., Moritz, A., Lee, K. A., Guo, A., *et al.*, Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat Biotechnol* 2005, *23*, 94-101.
[46] Boersema, P. J., Foong, L. Y., Ding, V. M. Y., Lemeer, S., *et al.*, In depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immuno-affinity purification and stable isotope dimethyl labeling. *Mol Cell Proteomics* 2009, *in press*.
[47] Tanaka, K., Waki, H., Ido, Y., Akita, S., *et al.*, Protein and polymer analyses up to *m/z* 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 1988, *2*, 151-153.
[48] Karas, M., Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 1988, *60*, 2299-2301.
[49] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* 1989, *246*, 64-71.
[50] Li, K.-Y., Tu, H., Ray, A. K., Charge Limits on Droplets during Evaporation. *Langmuir* 2005, *21*, 3786-3794.
[51] Iribarne, J. V., Thomson, B. A., On the evaporation of small ions from charged droplets. *J Chem Phys* 1976, *64*, 2287-2294.
[52] Malcolm, D., Mack, L. L., Hines, R. L., Mobley, R. C., *et al.*, Molecular Beams of Macroions. *J Chem Phys* 1968, *49*, 2240-2249.
[53] Cameron, A. E., D. F. Eggers, Jr., An Ion ``Velocitron''. *Rev Sci Instrum* 1948, *19*, 605-607.

[54] Wiley, W. C., McLaren, I. H., Time-of-Flight Mass Spectrometer with Improved Resolution. *Rev Sci Instrum* 1955, *26*, 1150-1157.
[55] Alikhanov, S. G., A new impulse technique for ion mass measurement. *Sov Phys JETP* 1957, *4*, 452.
[56] Marshall, A. G., Hendrickson, C. L., High-Resolution Mass Spectrometers. *Annu Rev Anal Chem* 2008, *1*, 579-599.
[57] Schwartz, J. C., Senko, M. W., Syka, J. E. P., A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom* 2002, *13*, 659-669.
[58] Marshall, A. G., Hendrickson, C. L., Jackson, G. S., Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom Rev* 1998, *17*, 1-35.
[59] Amster, I. J., Fourier Transform Mass Spectrometry. *J Mass Spectrom* 1996, *31*, 1325-1337.
[60] Syka, J. E. P., Marto, J. A., Bai, D. L., Horning, S., *et al.*, Novel Linear Quadrupole Ion Trap/FT Mass Spectrometer: Performance Characterization and Use in the Comparative Analysis of Histone H3 Post-translational Modifications. *J Proteome Res* 2004, *3*, 621-626.
[61] Makarov, A., Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal Chem* 2000, *72*, 1156-1162.
[62] Qizhi, H., Robert, J. N., Hongyan, L., Alexander, M., *et al.*, The Orbitrap: a new mass spectrometer. *J Mass Spectrom* 2005, *40*, 430-443.
[63] Olsen, J. V., de Godoy, L. M. F., Li, G., Macek, B., *et al.*, Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. *Mol Cell Proteomics* 2005, *4*, 2010-2021.
[64] Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., *et al.*, Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. *Anal Chem* 2006, *78*, 2113-2120.
[65] McLuckey, S. A., Principles of collisional activation in analytical mass spectrometry. *J Am Soc Mass Spectrom* 1992, *3*, 599-614.
[66] Paizs, B., Suhai, S., Fragmentation pathways of protonated peptides. *Mass Spectrom Rev* 2005, *24*, 508-548.
[67] Sleno, L., Volmer, D. A., Ion activation methods for tandem mass spectrometry. *J Mass Spectrom* 2004, *39*, 1091-1112.
[68] Frank, A. M., Predicting Intensity Ranks of Peptide Fragment Ions. *J Proteome Res* 2009, *8*, 2226-2240.
[69] Zubarev, R. A., Electron-capture dissociation tandem mass spectrometry. *Curr Opin Biotechnol* 2004, *15*, 12-16.
[70] Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* 2004, *101*, 9528-9533.
[71] Roepstorff, P., Fohlman, J., Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* 1984, *11*, 601.
[72] Biemann, K., Contributions of mass spectrometry to peptide and protein structure. *Biomed Environ Mass Spectrom* 1988, *16*, 99-111.
[73] Perkins, D. N., Pappin, D. J. C., Creasy, D. M., Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, *20*, 3551-3567.
[74] Eng, J. K., McCormack, A. L., Yates, J. R., An approach to correlate tandem mass-spectral data of peptide with amino-acid-sequence in a protein database. *J Am Soc Mass Spectrom* 1994, *5*, 976-989.
[75] Craig, R., Beavis, R. C., TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 2004, *20*, 1466-1467.
[76] Colinge, J., Masselot, A., Giron, M., Dessingy, T., Magnin, J., OLAV: Towards high-throughput tandem mass spectrometry data identification. *Proteomics* 2003, *3*, 1454-1463.
[77] Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., *et al.*, Open Mass Spectrometry Search Algorithm. *J Proteome Res* 2004, *3*, 958-964.
[78] Higgs, R. E., Knierman, M. D., Freeman, A. B., Gelbert, L. M., *et al.*, Estimating the statistical significance of peptide identifications from shotgun proteomics experiments. *J Proteome Res* 2007, *6*, 1758-1767.
[79] Weatherly, D. B., Atwood, J. A., Minning, T. A., Cavola, C., *et al.*, A heuristic method for assigning a false-discovery rate for protein identifications from mascot database search results. *Mol Cell Proteomics* 2005, *4*, 762-772.
[80] Heck, A. J. R., Krijgsveld, J., Mass spectrometry-based quantitative proteomics. *Exp Rev Proteomics* 2004, *1*, 317-326.
[81] Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* 2007, *389*, 1017-1031.
[82] Veenstra, T. D., Martinovic, S., Anderson, G. A., Pasa-Tolic, L., Smith, R. D., Proteome analysis using selective incorporation of isotopically labeled amino acids. *J Am Soc Mass Spectrom* 2000, *11*, 78-82.

[83] Oda, Y., Huang, K., Cross, F. R., Cowburn, D., Chait, B. T., Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A* 1999, *96*, 6591-6596.
[84] Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., *et al.*, Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol Cell Proteomics* 2002, *1*, 376-386.
[85] Harsha, H. C., Molina, H., Pandey, A., Quantitative proteomics using stable isotope labeling with amino acids in cell culture. *Nat Protoc* 2008, *3*, 505-516.
[86] Krijgsveld, J., Ketting, R. F., Mahmoudi, T., Johansen, J., *et al.*, Metabolic labeling of C. elegans and D. melanogaster for quantitative proteomics. *Nat Biotechnol* 2003, *21*, 927-931.
[87] Wu, C. C., MacCoss, M. J., Howell, K. E., Matthews, D. E., Yates, J. R., Metabolic Labeling of Mammalian Organisms with Stable Isotopes for Quantitative Proteomic Analysis. *Anal Chem* 2004, *76*, 4951-4959.
[88] Gouw, J. W., Pinkse, M. W. H., Vos, H. R., Moshkin, Y., *et al.*, In Vivo Stable Isotope Labeling of Fruit Flies Reveals Post-transcriptional Regulation in the Maternal-to-zygotic Transition. *Mol Cell Proteomics* 2009, *8*, 1566-1578.
[89] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., *et al.*, Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 1999, *17*, 994-999.
[90] Schmidt, A., Kellermann, J., Lottspeich, F., A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* 2005, *5*, 4-15.
[91] Che, F.-y., Fricker, L. D., Quantitation of Neuropeptides in Cpefat/Cpefat Mice Using Differential Isotopic Tags and Mass Spectrometry. *Anal Chem* 2002, *74*, 3190-3198.
[92] Hsu, J. L., Huang, S. Y., Chow, N. H., Chen, S. H., Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem* 2003, *75*, 6843-6852.
[93] Boersema, P. J., Aye, T. T., van Veen, T. A. B., Heck, A. J. R., Mohammed, S., Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. *Proteomics* 2008, *8*, 4624-4632.
[94] Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S., Heck, A. J. R., Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protoc* 2009, *4*, 484-494.
[95] Raijmakers, R., Berkers, C. R., de Jong, A., Ovaa, H., *et al.*, Automated Online Sequential Isotope Labeling for Protein Quantitation Applied to Proteasome Tissue-specific Diversity. *Mol Cell Proteomics* 2008, *7*, 1755-1762.
[96] Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., *et al.*, Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. *Mol Cell Proteomics* 2004, *3*, 1154-1169.
[97] Thompson, A., Schafer, J., Kuhn, K., Kienle, S., *et al.*, Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal Chem* 2003, *75*, 1895-1904.
[98] Pierce, A., Unwin, R. D., Evans, C. A., Griffiths, S., *et al.*, Eight-channel iTRAQ Enables Comparison of the Activity of Six Leukemogenic Tyrosine Kinases. *Mol Cell Proteomics* 2008, *7*, 853-863.
[99] Bantscheff, M., Boesche, M., Eberhard, D., Matthieson, T., *et al.*, Robust and Sensitive iTRAQ Quantification on an LTQ Orbitrap Mass Spectrometer. *Mol Cell Proteomics* 2008, *7*, 1702-1713.

Chapter 2:
Hydrophilic interaction liquid chromatography (HILIC) in proteomics

Paul J. Boersema[1,2], Shabaz Mohammed[1,2] and Albert J. R. Heck[1,2]

[1] 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
[2] Netherlands Proteomics Centre

Based on *Anal. Bioanal. Chem.* 2008, *391*, 151-159

ABSTRACT

In proteomics, nanoflow multidimensional chromatography is now the gold standard for the separation of complex mixtures of peptides as generated by in-solution digestion of whole-cell lysates. Ideally, the different stationary phases used in multidimensional chromatography should provide orthogonal separation characteristics. For this reason, the combination of strong cation exchange chromatography (SCX) and reversed-phase (RP) chromatography is the most widely used combination for the separation of peptides. Here, we review the potential of hydrophilic interaction liquid chromatography (HILIC) as a separation tool in the multidimensional separation of peptides in proteomics applications. Recent work has revealed that HILIC may provide an excellent alternative to SCX, possessing several advantages in the area of separation power and targeted analysis of protein posttranslational modifications.

INTRODUCTION

Recent years have seen a significant increase of interest in hydrophilic interaction liquid chromatography (HILIC). Although HILIC was first introduced in the 1970s [1], it took until the start of this century for a considerable number of HILIC applications to appear highlighting its versatile nature [2-7]. A major cause of this increase is the growing need for the analysis of polar compounds that do not bind to reversed-phase (RP) materials and the constantly increasing complexity of samples [8]. Additionally, the exceptional growth of liquid chromatography-electrospray ionisation-mass spectrometry (LC-ESI-MS) for analytical analyses, has lead to a widening search of acceptable chromatographic materials. In this search it has been noted that the buffer conditions that are used for HILIC are highly compatible with MS and the high organic contents of these buffers can, potentially, increase sensitivity in ESI-MS [9, 10]. Moreover, an HILIC separation is 'orthogonal' to RP separation which makes it viable for multidimensional separation of complex samples [11-13]. Finally, HILIC has been found to allow enrichment and the targeted analysis of post-translational modifications (PTMs) such as glycosylation [14], N-acetylation [12] and phosphorylation [15] in proteomics applications.

Although HILIC applications cover a large range of bio-active compounds, in this review we will focus on the use of HILIC in proteomics. We will describe how HILIC can be used for the separation of peptides in multidimensional chromatography and provide insights into its sensitivity, selectivity, separation power and "orthogonality". We will also discuss how HILIC has recently been used for the enrichment of post translational modifications at both peptide and protein levels.

HILIC

HILIC is characterized by the use of a hydrophilic stationary phase and a hydrophobic organic mobile phase [16] which was originally described by Linden *et al.* in 1975 [1]. The order of elution is reversed relative to reversed phase chromatography (RP), with hydrophilic compounds being retained longer than hydrophobic compounds. Therefore, HILIC can simply be seen as a form of normal-phase (NP) chromatography. However, the acronym HILIC was suggested to distinguish it from NP, as NP is typically performed with non-aqueous, non-water-miscible solvent buffers, while HILIC is performed with water-miscible solvents

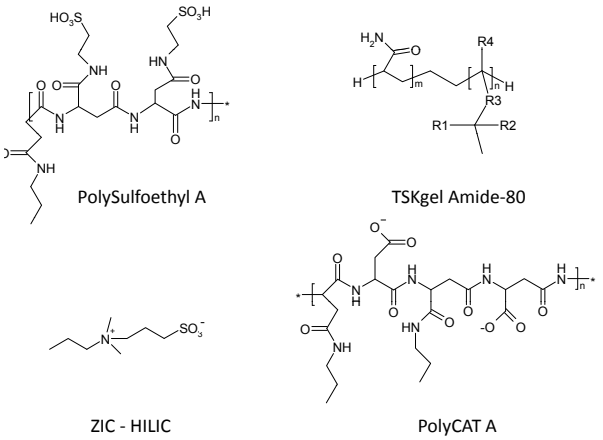


Figure 1. *Chemical structures of the functional groups in common HILIC stationary phases.*

and elution is achieved by a water gradient [2, 9, 17, 18].
In recent years, several stationary phases have emerged that are specifically made for HILIC approaches. Popular phases include underivatized silica stationary phases that contain functional groups such as siloxanes, silanols with (or without) a small quantity of metals [2, 9, 19, 20], derivatized silica, such as the cation exchanger polysulfoethyl A [17, 18], the weak cation exchanger Polycat A [21], the weak anion exchanger PolyWAX [22, 23], TSKgel amide 80 [24, 25], zwitterionic (ZIC)-HILIC [14] and 'click'-saccharides [26]; see Figure 1 for structures of typical HILIC phases. Each of these materials display different retention characteristics and separation selectivity and require distinct buffer constitutions for optimal results [10].
A HILIC buffer typically contains more than 70% acetonitrile [17]. Other eluents have been tested, for instance methanol or isopropanol, but resulted in poor chromatography or no analyte retention [2]. It is believed that the hydrophilic stationary phase enriches water from the buffer and thus generates an aqueous layer [17]. This allows for hydrophilic analytes to partition between hydrophilic layer and the hydrophobic elution buffer. Elution is obtained through increasing the hydrophilicity of the mobile phase by increasing the water content. The final separation mechanism of elution however, is most probably a superpositioning of partitioning and electrostatic interactions or hydrogen bonding to the stationary phase [2]. The extent to which each mechanism dominates is dependent on the actual type of stationary phase and the buffer conditions, including the level and type of organic content, the type and concentration of salts and the pH [10]. Charged stationary phases, such as the above mentioned anion or cation exchangers and deprotonated silanol groups, are most likely to display some degree of electrostatic interaction [2]. Buffering salts in the mobile phase can decrease these electrostatic interactions through disruption. The choice of salts is however limited, due to the highly organic conditions of HILIC buffers, which makes it difficult to dissolve some salts in them [10]. Typically, ammonium acetate and formate are chosen because of their compatibility with MS, but also ammonium bicarbonate, triethylamine phosphate (TEAP), sodium perchlorate and sodium methylphosphonate (Na-$MePO_4$) have been found to be applicable [2, 23]. Additionally, the same salt based disruption can decrease the retention of analytes and can be of aid in elution [17]. However, in some instances an increase in retention was seen upon increase of salt concentration [10]. This was rationalized by the suggestion of a mechanism in which salt is enriched in the aqueous layer, which in turn increases the hydrophilicity of this liquid layer around the stationary phase [10].
Unsurprisingly, the type of salt that is used can also influence the retention behavior. It was observed that basic peptides were retained longer with TEAP in the buffer, while with Na-$MePO_4$ these were the first peptides to elute with the weak anion exchanger PolyWAX as stationary phase [23]. It was reasoned that TEAP, as a counter-ion, will generate a sublayer where one negative charge of the salt is attracted to the positively charged stationary phase, while the second negative charge is free to attract basic peptide residues. In the case of Na-$MePO_4$ there is only one negative charge that interacts with either the stationary phase or basic residues, leaving enough stationary phase available for interaction with acidic peptides[23].
A special type of HILIC called electrostatic repulsion-hydrophilic interaction chromatography (ERLIC), actually specifically utilizes the electrostatic interactions in HILIC [23]. In ERLIC, a stationary phase is chosen that has a similar charge as the analytes to be separated. Analytes are on the one hand repelled by the stationary phase but on the other retained in the aqueous layer around the stationary phase. These opposing interactions allow isocratic resolution [23]. ERLIC can also be exploited in the selective isolation of phosphopeptides from a tryptic digest [23, 27].

Another factor that influences the retention characteristics in HILIC is the pH of the buffer. Whether the buffer pH is above or below the pK_a of the analyte determines its charge state, which in turn affects the hydrophilicity of the analyte and likewise the interaction with the stationary phase [10, 12]. For example, large differences were found in the retention profile of a highly complex tryptic digest run over the same column at pH 3 and pH 8 [12]. At pH 3, acidic peptide residues are protonated and so the overall hydrophilicity of peptides containing these residues is decreased, leading to poorer retention of these peptides, with the effect being magnified with higher number of acidic residues [12].
For a more detailed overview of different HILIC stationary phases and their applications outside the field of proteomics, see Hëmstrom *et al.* [2].

HILIC IN MULTIDIMENSIONAL PEPTIDE SEPARATION

Shotgun proteomics involves the analysis of entire proteomes in single experiments. Analysis of such samples raises a number of issues, including enormous complexity, large dynamic range and low levels of available analyzable material. Often, SDS PAGE based [28] or LC separation techniques are used to reduce the sample complexity followed by RP LC-MS as the final readout. Major advances have been made in miniaturization of LC-MS since reducing the flow rate improves mass spectrometric sensitivity [29], reduces nonspecific adsorption onto separation devices and improves chromatographic resolving power [30], but improvements are still required. The complexity of a sample can be further reduced prior to mass spectrometric analysis by the addition of extra dimensions of separation. At present, the method of choice is the combination of strong cation exchange (SCX) with RP [31-34]. Over time, a number of alternative configurations have been developed, including size exclusion chromatography (SEC)-RP [35, 36] and RP at pH 10 followed by RP at pH 2.6 [37]. However, the low separation power of SEC [36] and the correlative separation of RP at different pH levels has not yielded the optimal two dimensional (2D)-LC system yet.
Recently, HILIC was introduced as a dimension for 2D-LC-MS and shown to remove a few of the disadvantages of existing techniques [11, 12]. Evidently, RP and HILIC buffers are not directly compatible and so online hyphenation of these stationary phases will be difficult. On-line hyphenation possesses certain advantages such as minimal loss of sample, no vial contamination, and no sample dilution [38, 39]. However, performing an experiment in an off-line fashion is not necessarily inferior since with off-line fractionation overloading of the second dimension is no longer an issue as it is for on-line approaches. Moreover, in an off-line mode, columns and/or buffers that are normally not compatible with the two consecutive separations can be used [38, 40]. Finally, a conventional gradient for peptide separation can be used in off-line setups, which has been shown to be generally superior to step elution [40].
Gilar *et al.* separated and analyzed a mixture of approximately 200 peptides using different stationary phases, including SEC, HILIC, RP at pH 2.6 and pH 10 and SCX. HILIC was shown to have a separation power superior to both SCX and SEC [11]. The orthogonality – *i.e.,* how different the separation mechanisms are- of the stationary phases was determined by plotting the peptide retention times of two dimensions against each other (see Figure 2). Although this study did not actually experimentally connect the different stationary phases, since all LC-MS runs were one dimensional, it revealed that HILIC is slightly more orthogonal to RP than SCX [11]. A genuine two-dimensional HILIC-RP system was presented recently [12]. A microliter flow-scale ZIC-HILIC system was connected off-line with a nanoliter flow-scale RPLC-MS. For semi-automation, a microliter-scale fraction collector was connected to the HILIC output to mix the eluent with RP- and MS-compatible aqueous

buffer and to allow direct consecutive analysis of the HILIC fractions by MS. In this study, a level of orthogonality to RP was found that matched the theoretical picture presented by Gilar *et al.* [11]. SCX separation can be limited by the fact that similarly charged peptides cluster in narrow elution windows (see also Figure 2) and it was shown that HILIC did not exhibit this behavior [11, 12]. Analysis of data acquired using this two-dimensional system shed further light on how HILIC can be orthogonal to RP. Indeed, a separation mechanism was observed that is based on both retention by hydrophilicity due to partitioning with the aqueous sublayer and retention by charge due to electrostatic interaction with the stationary phase [11, 12]. These electrostatic interactions ensure that HILIC separation is more than merely the reverse of RP, while the presence of hydrophilic interactions made similarly charged peptides elute over a wider window [12]. It was further noted that orthogonality of ZIC-HILIC with RP is dependent on the buffer pH. At pH 3, a higher orthogonality was seen than at pH 6.8 and pH 8 [12].

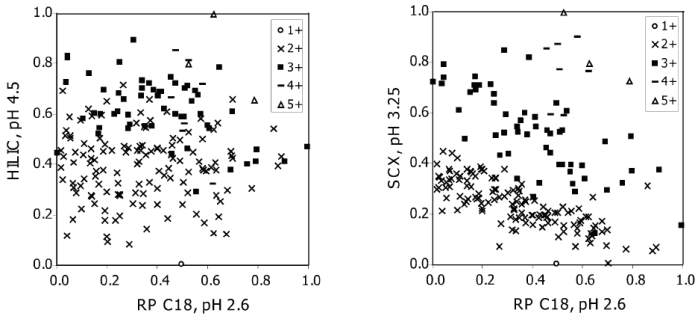


Figure 2. *Two-dimensional plots of normalized peptide retention times in HILIC, SCX and RP separation. Both HILIC and SCX have separation mechanisms orthogonal to RP, but the clustering of similarly charged peptides in SCX makes this a less optimal first dimension. From the work of Gilar et al. [11] and reproduced with permission from the American Chemical Society, copyright (2005).*

The studies discussed so far were all performed with HILIC columns packed with particles. A further improvement of separation could be obtained by using monolithic structures instead of porous particles. The higher porosity of monolithic materials enables the use of longer columns and/or higher flow rates, increasing the separation efficiency without increasing the backpressure [41]. Recently, attempts to develop HILIC columns that possess monolithic characteristics have been reported [41-45]. Horie et al. demonstrated the use of HILIC monoliths for peptide separations [42]. The authors applied increased flow rates and confirmed that short gradients of 3-10 minutes were possible with 20-cm-long 200 μm ID columns, while the peak capacities were comparable to a similar monolithic RP setup. The poor peptide sequencing speed of a mass spectrometer can be a potential drawback in the direct coupling of such columns to MS. Currently, mass spectrometers can sequence 2-3 peptides per second which is significantly lower than the number of peptides that will be delivered by the monolith column. However, such columns are highly applicable as a first dimension, where time is not a parameter.

HILIC is proving to be an attractive choice in the range of separation methods available for the proteomics researcher. The combinations chosen so far for 2D-LC-MS, as reported in the literature, have been limited to HILIC-RP, but the versatility of the chromatographic material will also allow combinations such as SCX-HILIC and three-dimensional approaches can be envisaged. The high organic levels utilized in HILIC buffers compared to RP buffers

have the additional benefit that they improve electrospray ionization, which increases the sensitivity of ESI-MS [9], thus suggesting that it may be an interesting final dimension [11].

HILIC FOR THE ANALYSIS OF PTMS

The analysis of protein PTMs are a very important albeit challenging task in proteomics. PTMs are generally present at much lower levels than their non-modified counterparts, which makes it very easy to 'miss them' in untargeted analyses, such as those performed using shotgun proteomics techniques [46]. Protein modifications such as phosphorylation and glycosylation have proven to be difficult to investigate in a routine fashion. This is mainly because these moieties are prone to elimination and adsorption if care is not taken with sample preparation and MS analysis protocols [14, 47]. Recently, some of these MS problems have been overcome by alternative MS peptide fragmentation techniques, such as electron capture dissociation (ECD) [48] and electron transfer dissociation (ETD) [49]. Several approaches have been developed to counter some of the problems presented by the low abundance and highly labile nature of PTM peptides. The use of chelation [50-52] or specific chromatographic materials [47] have proven to be successful strategies and in the latter category HILIC has proven to be a worthy addition.

HILIC in the targeted analysis of phosphorylated peptides

Phosphorylation is probably the most studied protein post-translational modification in proteomics; it provides an important function in signal transduction, metabolic maintenance and cell division [53]. The development of techniques to enrich for phosphopeptides has involved major efforts in phosphoproteomics. One of the more successful approaches targets the enrichment of phosphopeptides by SCX at low pH [47]. At pH3, acidic residues such as glutamic and aspartic acid are neutral while phospho-serine/threonine/tyrosine will still be negative. Such a distinction allows proteolytic tryptic phosphopeptides to be separated from 'normal' tryptic peptides due to an earlier elution. However, other peptide subgroups with reduced net charge such as N-acetylated tryptic peptides will also exhibit a shorter retention time, and thus a second level of enrichment is required to separate coeluting non-phosphorylated peptides from phosphorylated peptides [54]. Furthermore, trypsinization of proteins is less efficient for regions where phosphomoeities are present [55]. Thus a large fraction of proteolytic phosphopeptides will contain miscleavages (*i.e.*, additional basic residues), and will not be enriched. Therefore, other techniques for the targeted enrichment of phosphopeptides are required and generally performed after SCX, such as those based on immobilized metal affinity chromatography (IMAC) [50, 54] or the metal oxide TiO_2 [51, 52, 56]. One other limitation of SCX for phosphopeptide enrichment is that it is based on poor retention. This means that multiply phosphorylated peptides, another large sub-group, are difficult to retain or even lost. The use of strong anion exchange (SAX) stationary phases immediately following SCX retention can circumvent some of these problems [57]. The generally lower pIs of phosphopeptides will allow stronger retention on SAX than 'normal' peptides. However, the extra LC separation increases the analysis time and the likelihood of sample loss.

Finally, a mixed-bed ion exchange method with both cation and anion exchange was presented recently and was shown to recover more normal tryptic peptides than SCX and to exhibit orthogonal separation towards RP [58]. Phosphopeptides could be eluted by applying a single final salt step. However, in this step only a subset of phosphopeptides -the acidic- could be resolved.

HILIC is presented as an alternative enrichment technique as it exploits the increase in hy-

drophilicity of peptides by the attachment of a phosphogroup [15, 22, 23, 27]. McNulty *et al.* applied a TSKgel Amide-80 based HILIC system to phosphopeptide analysis of a Calyculin A-treated HeLa cell lysate digest. The authors demonstrated that the system, under optimal conditions, allowed phosphorylated peptides to be separated from nonphosphorylated peptides and acidic peptides that might interfere with subsequent IMAC enrichment. Interestingly, the phosphopeptides eluted in the middle of the HILIC separation and would allow a considerable level of fractionation. This rudimentary analysis permitted more than 1000 unique phosphopeptides to be identified after further IMAC enrichment and LC-MS analysis [15].

Ytterberg *et al.* evaluated the use of HILIC in a preparative separation set-up using HILIC material packed in micro tips [22]. Different combinations of SCX, RP and two types of HILIC – Polyhydroxyethyl aspartamide (PHEA) and PolyWAX- were tested for the enrichment of phosphopeptides from a saliva digest. As expected, most of the phosphorylated peptides were retained strongly under typical HILIC analysis conditions. Furthermore, the separation power of the stepwise elution was sufficient to separate peptides with differing numbers of phosphorylated residues. The conditions used with PolyWAX are ERLIC conditions and seem to be very useful in phosphopeptide enrichment [23]. Normal tryptic peptides are simultaneously repelled by the weak anion exchange material and attracted by hydrophilic interactions, while phosphopeptides are attracted by both. This mode of separation is very responsive to salt contents and the pH of the buffer and can thus be tweaked to enrich for highly phosphorylated peptides containing up to six phosphorylated residues [22].

In contrast to these results obtained with TSKgel Amide-80, PHEA and PolyWAX, enrichment of phosphopeptides was not observed with ZIC-HILIC [12]. This might be explained by the zwitterionic nature of this chromatographic material: the negative charge could potentially repel the phosphate group, resulting in phosphopeptides eluting with nonphosphorylated peptides.

The micro-column approach is a simple and undemanding preparative method. However, HPLC-mode HILIC has the potential for automation and fractionation, as the HILIC buffers are directly compatible with IMAC and TiO_2. It might be a valuable alternative to SCX, with improved fractionation and the better retention of multiply phosphorylated peptides, and so it potentially could allow a more comprehensive analysis of the phosphoproteome. While it should be noted that the application of HILIC to phosphoproteomics is still in its infancy, it is likely that the actual power of HILIC in this field will become clear in the next few years.

HILIC in the targeted analysis of N-terminally acetylated peptides

The positive N-terminal charge of a protein can be neutralized by different post-translational modifications each potentially influencing protein function, stability and interaction with other biomolecules [59]. Amongst the biologically relevant N-terminal modifications are acetylation, methylation and myristoylation [60]. As mentioned above, tryptic N-acetylated peptides can be enriched by SCX [33, 54, 61]. The acetylation neutralizes the N-terminal charge, lowering the net charge compared to the unmodified counterpart. Therefore, N-acetylated tryptic peptides will, like phosphopeptides, elute in the first few SCX fractions.

ZIC-HILIC has been proposed as another means of enriching N-acetylated tryptic peptides [12]. When it contains a neutralized N-terminus, the hydrophilicity of an N-acetylated peptide is decreased. This reduction in polarity is even more pronounced at pH 3, where only basic peptide residues are charged and acidic residues are protonated. Through the use of a ZIC-HILIC separation and fractionation, Boersema *et al.* showed that this subgroup of

trypsinized peptides that are N-terminally blocked could be enriched due to shorter retention than 'normal' trypsinized peptides (see Figure 3) [12]. Similarly, other subgroups of N-terminally modified peptides, such as those that contain an N-terminal pyroglutamic acid, were also found primarily in the first ZIC-HILIC fractions [12].
Unfortunately, as trypsin cannot cleave acetylated lysine residues, peptides containing this modification could not be enriched. The misscleavage-containing K-acetylated tryptic pep-

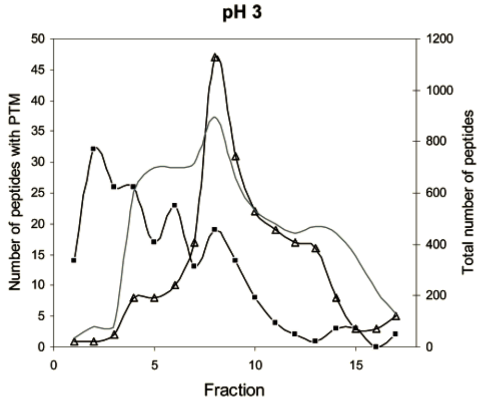


Figure 3. *Distribution of phosphorylated and N-acetylated peptides over ZIC-HILIC fractions. First dimension: ZIC-HILIC, 200 µm x 160 mm, 3.5 µm, 200 Å. Flow rate 1.5 µL/min, 1 min. fractions. Number of peptides: (–) total, (-▪-) N-acetylated and (-Δ-) phosphorylated. ZIC-HILIC provides clear enrichment of N-acetylated peptides in the initial fractions. From the work of Boersema et al. [12] with permission from the American Chemical Society copyright (2007).*

tide will possess a net charge that is the same as 'normal' tryptic peptides [12].
Interestingly, and in contrast with SCX [33], no significant enrichment of C-terminal peptides was observed in this ZIC-HILIC set-up, although the hydrophilicity of these peptides is also different than "normal" tryptic peptides, due to the absence of a terminal arginine or lysine residue. It was hypothesized that this might be explained by the difference in charge distribution after losing an N-terminal positive charge compared to lacking a C-terminal positive charge. A C-terminal peptide would as a result be more polar –and thus more hydrophilic- compared to an N-terminally blocked peptide [12]. As suggested above, the ZIC-HILIC method will enrich for any peptide containing a neutralized N-terminus and thus would be applicable as a first step to enrich formylated, carboxylated, hydroxylated, palmitoylated and myrisotylated peptides.

HILIC in the targeted analysis of glycosylated peptides
Glycosylation, the attachment of a carbohydrate moiety to a protein, is a frequent but very heterogeneous PTM. It modulates the physicochemical and biological properties of proteins and serves as a recognition determinant between molecules, molecules and cells, or between cells [62]. Glycosylation of outer cell wall proteins are, for instance, involved in host-pathogen interactions. It has been estimated that more than half of all proteins in a biological system are glycoproteins [63]. The heterogeneity of glycan structures makes a global proteomics characterization of glycosylation extremely difficult, as the signal intensity of glycosylated peptides with the same amino acid backbone is dispersed over peptides with different oligosaccharide chains [64]. The proteomics analysis of glycopeptides is made even more tedious by the fact that MS/MS data rarely allow the confident identification of peptide sequences due to the relatively large molecular weights of glycosylated peptides [65] and the complex fragmentation that originates from both peptide and glycan cleavages [14]. To overcome some of these issues in proteomics analysis, the labile sugar groups are often removed by deglycosylation prior to MS analysis [66]. The problem of low glycopeptides abundance can be tackled by enrichment techniques, mostly those based on lectin-mediated

affinity capture [66]. Recently, HILIC was introduced as a promising additional enrichment step for glycopeptides [14].
The hydrophilicity of glycopeptides makes them ideal candidates for separation by HILIC. A first glycopeptide application for HILIC was found in the analysis of differentially sialylated glycopeptides from interferon-γ [67]. Fractions of glycopeptides separated by RP were further separated by HILIC. Their orthogonal retention on the HILIC column was shown to correlate well with the number of sialyl groups. Further studies by ESI-MS revealed that the interaction of N-glycans with ZIC-HILIC is based on a partitioning mechanism while separation of differently sialylated N-glycans could be explained by an electrostatic repulsion interaction mechanism [68]. Recently, more global proteomics approaches were reported targeting protein glycosites using HILIC for enrichment of glycopeptides [14, 66, 69-72]. Generally, in these approaches glycoproteins were first selectively captured by lectin-mediated affinity chromatography, followed by SDS-PAGE and in-gel digestion. The peptides obtained were then further enriched for glycosylation using ZIC-HILIC micro-columns. The glycopeptides that were primarily found in the later fractions were then enzymatically degycosylated prior to MS analysis. Figure 4 shows the relative intensities after solid-phase extraction (SPE) of a tryptic digest of TIMP-1 with RP (R2) and HILIC micro-columns. The relative abundance of glycopeptides was significantly increased following enrichment by HILIC. SPE by RP caused the matrix assisted laser desorption/ionisation (MALDI)-spectrum to be dominated by non-glycosylated peptides, while, in comparison, SPE with HILIC resulted in a spectrum crowded by glycosylated peptides [70].
Small diversions to this general protocol have been reported and include the lectin-mediated affinity capture at the peptide level [69] and the further separation of glycopeptides by SCX

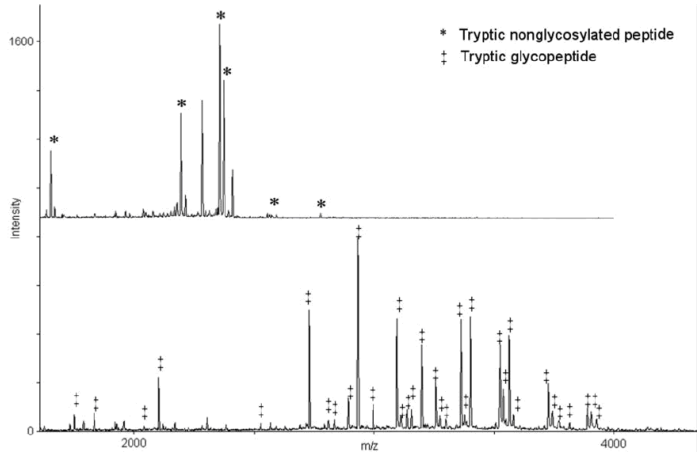


Figure 4. *MALDI-TOF spectra after SPE of a tryptic digest of TIMP-1 by RP (R2, upper spectrum) and HILIC (lower spectrum) microcolumns. HILIC clearly enriches for glycopeptides. Figure kindly provided by Dr. P. Højrup (similar to [70]).*

[66]. It was shown that the unambiguous localization of glycosylation sites is facilitated by leaving a single GlcNac residue on the site [14] or tagging the site using ^{18}O isotope labeling [66, 69].
Using the HILIC method, 62 glycosylation sites could be identified for 37 glycoproteins in human plasma [14], while a further study of the Cohn IV fraction of human plasma revealed 103 N-glycosylation sites and 23 fucosylated N-glycosylation sites [66]. 1465 N-glycosylated sites were found on 829 proteins in *C. elegans* [72] However, this study was performed using

three times 50-200 mg sample for affinity chromatography on three different lectin columns and each LC-MS analysis was performed in triplicate [72], unlike the Cohn IV work where a 1 mg fraction was used [66].
The glycosylation sites were determined in these studies, but not the actual glycan structure. A few studies have used HILIC for desalting and enrichment [65, 70] or separation [73] of intact glycopeptides. However, these studies all used relatively simple protein mixtures or even single proteins, indicating that the unambiguous characterization of peptide glycan structures at the proteome level is still rather challenging.

These early experiments show that HILIC can be used as an effective and relatively simple tool for the targeted analysis of protein glycosylation. Future studies are likely to follow, wherein HILIC might also have an important role in not only the glycosite elucidation, but also in the compositional analysis of glycan structures, as HILIC was already found to be convenient for desalting and enrichment [64] or even separation [74] of glycans themselves.

HILIC in the targeted analysis of histone modifications
The regulation of chromatin structure, and therefore DNA transcription and replication, is driven by post-translational modifications of the core histones. These modifications, including methylation, acetylation, phosphorylation, sumoylation and ubiquitination, are primarily found at the histone N-terminal tails [75]. Different modifications can occur at the same histone, and it is emerging that not just single site modification but rather the interdependence of these modifications is very important in regulation [76]. The analysis of histone modifications on intact histone proteins by MS is very demanding, due to the enormous heterogeneity of the modifications [77]. The separation of differentially modified proteins prior to MS is therefore essential.
HILIC has been used in the last decade for the analysis acetylated and methylated histones [5, 7, 21, 78-80]. In these studies, RP and HILIC phases are generally combined to separate different histones. In analyses where LC-MS analysis is used for the characterization of histone modifications, this is performed on peptide level. However, information on interdependence of modifications is lost due to the hydrolysis step, because different modifications will end up at different peptides [5, 6, 79-81]. To preserve this information, a HILIC separation of intact histones was recently performed using a top-down MS approach, *i.e.* at the level of intact proteins [77, 82]. Histone H4 was purified by RP while HILIC was used for further separation of differentially modified forms. Subsequently, these H4 HILIC frac-

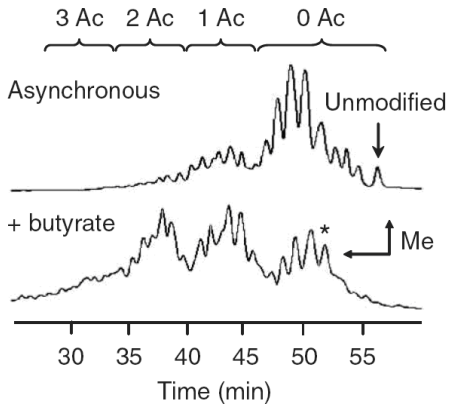


Figure 5. *HILIC separation of by Glu-C generated long N-terminal peptides from histone H3.2. HILIC primary separation is controlled by the number of acetylations (the three bigger peaks for the +butyrate sample), and within these peaks a secondary separation is observed relating to the number of methylations (the narrower peaks). Adapted by permission from Macmillan Publishers Ltd: Nature Methods [77] copyright (2007)*

tions were analyzed by ESI-Q-FTMS/MS with ECD fragmentation [82]. A similar but semi top-down approach was performed for the analysis of H3.2. As H3.2 is not very soluble in the high organic HILIC buffers, GluC digestion was employed to generate a 50 amino acid long N-terminal peptide that contains most of the modification sites [77]. Figure 5 clearly shows the separation power of HILIC in the analysis of histones. It simultaneously separates acetylated and methylated histones, greatly reducing the complexity of the sample. Acetylation is the major determinant for separation. Separation of proteins of which acetylation was induced by butyrate resulted in more peaks at a lower retention time. The smaller peaks within these larger peaks represent proteins with different levels of methylation.
In these histone analyses, HILIC delivered the extra separation step that was needed to sufficiently reduce the complexity of the samples and to separate the most important histone modifications.

CONCLUDING REMARKS

Recent years have witnessed an increased interest in HILIC. More versatile and diverse stationary phases have become available, leading to reports of an exciting and broad range of applications. The unique separation and orthogonality of HILIC towards RP (the most commonly used peptide separation method) make it an ideal method for multidimensional chromatography to extend separation power. As far as selectivity is concerned, HILIC can compete well with RP and SCX which are the two main chromatographic techniques applied today. Focusing solely on proteomics applications, HILIC has shown to be very versatile in the analyses of protein modifications. The retention of hydrophilic compounds can be effectively exploited in the enrichment of phosphorylated, N-terminally blocked and glycosylated peptides. Moreover, in combination with RP it allows separation of complex differentially modified histones, an incredibly challenging area. It is therefore expected that the number of applications of HILIC in the proteomics field will increase significantly in the years to come and the development of novel HILIC stationary phases and/or monolithic columns will continue. Much progress is expected.

ACKNOWLEDGEMENTS

We thank Dr. A.F.M. Altelaar for critically reviewing this manuscript. This work was supported by the Netherlands Proteomics Centre (www.netherlandsproteomicscentre.nl)

REFERENCES

[1] Linden, J. C., Lawhead, C. L., Liquid chromatography of saccharides. *J Chromatogr A* 1975, *105*, 125-133.
[2] Hemström, P., Irgum, K., Hydrophilic interaction chromatography. *J Sep Sci* 2006, *29*, 1784-1821.
[3] Strege, M. A., Stevenson, S., Lawrence, S. M., Mixed mode anion-cation exchange/hydrophilic interaction liquid chromatography-electrospray mass spectrometry as an alternative to reversed phase for small molecule drug discovery. *Anal Chem* 2000, *72*, 4629-4633.
[4] Risley, D. S., Strege, M. A., Chiral separations of polar compounds by hydrophilic interaction chromatography with evaporative light scattering detection. *Anal Chem* 2000, *72*, 1736-1739.
[5] Sarg, B., Koutzamani, E., Helliger, W., Rundquist, I., Lindner, H. H., Postsynthetic trimethylation of histone H4 at lysine 20 in mammalian tissues is associated with aging. *J Biol Chem* 2002, *277*, 39195-39201.
[6] Sarg, B., Helliger, W., Hoertnagl, B., Puschendorf, B., Lindner, H., The N-terminally acetylated form of mammalian histone H1(o), but not that of avian histone H5, increases with age. *Arch Biochem Biophys* 1999, *372*, 333-339.
[7] Lindner, H., Sarg, B., Grunicke, H., Helliger, W., Age-dependent deamidation of H1(0) histones in chromatin of mammalian tissues. *J Cancer Res Clin Oncol* 1999, *125*, 182-186.
[8] Xu, R. N., Fan, L., Rieser, M. J., El-Shourbagy, T. A., Recent advances in high-throughput quantitative bioanalysis by LC-MS/MS. *J Pharm Biomed Anal* 2007, *44*, 342-355.
[9] Naidong, W., Bioanalytical liquid chromatography tandem mass spectrometry methods on underivatized silica columns with aqueous/organic mobile phases. *J Chromatogr B Analyt Technol Biomed Life Sci* 2003, *796*, 209-224.
[10] Guo, Y., Gaiki, S., Retention behavior of small polar compounds on polar stationary phases in hydrophilic interaction chromatography. *J Chromatogr A* 2005, *1074*, 71-80.
[11] Gilar, M., Olivova, P., Daly, A. E., Gebler, J. C., Orthogonality of separation in two-dimensional liquid chromatography. *Anal Chem* 2005, *77*, 6426-6434.
[12] Boersema, P. J., Divecha, N., Heck, A. J. R., Mohammed, S., Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. *J Proteome Res* 2007, *6*, 937-946.
[13] Wang, X. D., Li, W. Y., Rasmussen, H. T., Orthogonal method development using hydrophilic interaction chromatography and reversed-phase high-performance liquid chromatography for the determination of pharmaceuticals and impurities. *J Chromatogr A* 2005, *1083*, 58-62.
[14] Hagglund, P., Bunkenborg, J., Elortza, F., Jensen, O. N., Roepstorff, P., A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J Proteome Res* 2004, *3*, 556-566.
[15] McNulty, D. E., Annan, R. S., *55th ASMS Conference on Mass Spectrometry*, Indianapolis, Indiana 2007.
[16] Alpert, A. J., Andrews, P. C., Cation-exchange chromatography of peptides on poly(2-sulfoethyl aspartamide)-silica. *J Chromatogr* 1988, *443*, 85-96.
[17] Alpert, A. J., Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J Chromatogr* 1990, *499*, 177-196.
[18] Zhu, B.-Y., Mant, C. T., Hodges, R. S., Hydrophilic-interaction chromatography of peptides on hydrophilic and strong cation-exchange columns. *J Chromatogr A* 1991, *548*, 13-24.
[19] Nikolov, Z. L., Reilly, P. J., Retention of carbohydrates on silica and amine-bonded silica stationary phases: application of the hydration model. *J Chromatogr A* 1985, *325*, 287-293.
[20] Kirkland, J. J., Dilks, C. H., Destefano, J. J., Normal-Phase High-Performance Liquid-Chromatography with Highly Purified Porous Silica Microspheres. *J Chromatogr* 1993, *635*, 19-30.
[21] Lindner, H., Sarg, B., Meraner, C., Helliger, W., Separation of acetylated core histones by hydrophilic-interaction liquid chromatography. *J Chromatogr A* 1996, *743*, 137-144.
[22] Ytterberg, J. A., Ogorzalek-Loo, R. R., Boontheung, P., Loo, J. A., *55th ASMS Conference on Mass Spectrometry*, Indianapolis, Indiana 2007.
[23] Alpert, A. J., Electrostatic Repulsion Hydrophilic Interaction Chromatography for Isocratic Separation of Charged Solutes and Selective Isolation of Phosphopeptides. *Anal Chem* 2007.
[24] Yoshida, T., Peptide separation in normal phase liquid chromatography. *Anal Chem* 1997, *69*, 3038-3043.
[25] Tomiya, N., Awaya, J., Kurono, M., Endo, S., *et al.*, Analyses of N-linked oligosaccharides using a two-dimensional mapping technique. *Anal Biochem* 1988, *171*, 73-90.
[26] Guo, Z., Lei, A., Zhang, Y., Xu, Q., *et al.*, "Click saccharides": novel separation materials for hydrophilic interaction liquid chromatography. *Chem Commun (Camb)* 2007, 2491-2493.

[27] Alpert, A. J., Gygi, S. P., Shookla, A. K., *55th ASMS Conference on Mass Spectrometry*, Indianapolis, Indiana 2007.
[28] Rabilloud, T., Two-dimensional gel electrophoresis in proteomics: Old, old fashioned, but it still climbs up the mountains. *Proteomics* 2002, *2*, 3-10.
[29] Wang, H., Hanash, S., Multi-dimensional liquid phase based separations in proteomics. *J Chromatogr B Analyt Technol Biomed Life Sci* 2003, *787*, 11-18.
[30] Nagele, E., Vollmer, M., Horth, P., Vad, C., 2D-LC/MS techniques for the identification of proteins in highly complex mixtures. *Expert Rev Proteomics* 2004, *1*, 37-46.
[31] Dong, M.-Q., Venable, J. D., Au, N., Xu, T., *et al.*, Quantitative Mass Spectrometry Identifies Insulin Signaling Targets in C. elegans. *Science* 2007, *317*, 660-663.
[32] Brunner, E., Ahrens, C. H., Mohanty, S., Baetschmann, H., *et al.*, A high-quality catalog of the Drosophila melanogaster proteome. *Nat Biotech* 2007, *25*, 576-583.
[33] Dormeyer, W., Mohammed, S., Breukelen, B. v., Krijgsveld, J., Heck, A. J. R., Targeted Analysis of Protein Termini. *J Proteome Res* 2007.
[34] Washburn, M. P., Wolters, D., Yates, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotech* 2001, *19*, 242-247.
[35] Lemieux, L., Amiot, J., Application of Reversed-Phase High-Performance Liquid-Chromatography to the Separation of Peptides from Phosphorylated and Dephosphorylated Casein Hydrolysates. *J Chromatogr* 1989, *473*, 189-206.
[36] Opiteck, G. J., Jorgenson, J. W., Anderegg, R. J., Two-dimensional SEC/RPLC coupled to mass spectrometry for the analysis of peptides. *Anal Chem* 1997, *69*, 2283-2291.
[37] Gilar, M., Olivova, P., Daly, A. E., Gebler, J. C., Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J Sep Sci* 2005, *28*, 1694-1703.
[38] Peng, J. M., Elias, J. E., Thoreen, C. C., Licklider, L. J., Gygi, S. P., Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: The yeast proteome. *J Proteome Res* 2003, *2*, 43-50.
[39] Wagner, K., Miliotis, T., Marko-Varga, G., Bischoff, R., Unger, K. K., An automated on-line multidimensional HPLC system for protein and peptide mapping with integrated sample preparation. *Anal Chem* 2002, *74*, 809-820.
[40] Vollmer, M., Horth, P., Nagele, E., Optimization of two-dimensional off-line LC/MS separations to improve resolution of complex proteomic samples. *Anal Chem* 2004, *76*, 5180-5185.
[41] Pack, B. W., Risley, D. S., Evaluation of a monolithic silica column operated in the hydrophilic interaction chromatography mode with evaporative light scattering detection for the separation and detection of counter-ions. *J Chromatogr A* 2005, *1073*, 269-275.
[42] Horie, K., Ikegami, T., Hosoya, K., Saad, N., *et al.*, Highly efficient monolithic silica capillary columns modified with poly(acrylic acid) for hydrophilic interaction chromatography. *J Chromatogr A* 2007, *1164*, 198-205.
[43] Ikegami, T., Fujita, H., Horie, K., Hosoya, K., Tanaka, N., HILIC mode separation of polar compounds by monolithic silica capillary columns coated with polyacrylamide. *Anal Bioanal Chem* 2006, *386*, 578-585.
[44] Ikegami, T., Horie, K., Jaafar, J., Hosoya, K., Tanaka, N., Preparation of highly efficient monolithic silica capillary columns for the separations in weak cation-exchange and HILIC modes. *J Biochem Biophys Methods* 2007, *70*, 31-37.
[45] Ye, F., Xie, Z., Wong, K. Y., Monolithic silica columns with mixed mode of hydrophilic interaction and weak anion-exchange stationary phase for pressurized capillary electrochromatography. *Electrophoresis* 2006, *27*, 3373-3380.
[46] McLachlin, D. T., Chait, B. T., Analysis of phosphorylated proteins and peptides by mass spectrometry. *Curr Opin Chem Biol* 2001, *5*, 591-602.
[47] Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., *et al.*, Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* 2004, *101*, 12130-12135.
[48] Hakansson, K., Chalmers, M. J., Quinn, J. P., McFarland, M. A., *et al.*, Combined electron capture and infrared multiphoton dissociation for multistage MS/MS in a Fourier transform ion cyclotron resonance mass spectrometer. *Anal Chem* 2003, *75*, 3256-3262.
[49] Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 2004, *101*, 9528-9533.
[50] Posewitz, M. C., Tempst, P., Immobilized gallium(III) affinity chromatography of phosphopeptides.

Anal Chem 1999, *71*, 2883-2892.
[51] Pinkse, M. W. H., Uitto, P. M., Hilhorst, M. J., Ooms, B., Heck, A. J. R., Selective Isolation at the Femtomole Level of Phosphopeptides from Proteolytic Digests Using 2D-NanoLC-ESI-MS/MS and Titanium Oxide Precolumns. *Anal Chem* 2004, *76*, 3935-3943.
[52] Pinkse, M. W. H., Mohammed, S., Gouw, J. W., van Breukelen, B., *et al.*, Highly Robust, Automated, and Sensitive Online TiO2-Based Phosphoproteomics Applied To Study Endogenous Phosphorylation in Drosophila melanogaster. *J Proteome Res* 2007.
[53] Reinders, J., Sickmann, A., State-of-the-art in phosphoproteomics. *Proteomics* 2005, *5*, 4052-4061.
[54] Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., *et al.*, Quantitative Phosphoproteomics Applied to the Yeast Pheromone Signaling Pathway. *Mol Cell Proteomics* 2005, *4*, 310-327.
[55] Molina, H., Horn, D. M., Tang, N., Mathivanan, S., Pandey, A., Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A* 2007, *104*, 2199-2204.
[56] Benschop, J. J., Mohammed, S., O'Flaherty, M. C., Heck, A. J. R., *et al.*, Quantitative phospho-proteomics of early elicitor signalling in Arabidopsis. *Mol Cell Proteomics* 2007, *6*, 1198-1214.
[57] Dai, J., Jin, W. H., Sheng, Q. H., Shieh, C. H., *et al.*, Protein phosphorylation and expression profiling by Yin-Yang multidimensional liquid chromatography (Yin-Yang MDLC) mass spectrometry. *J Proteome Res* 2007, *6*, 250-262.
[58] Motoyama, A., Xu, T., Ruse, C. I., Wohlschlegel, J. A., Yates, J. R., Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. *Anal Chem* 2007, *79*, 3623-3634.
[59] Polevoda, B., Sherman, F., Composition and function of the eukaryotic N-terminal acetyltransferase subunits. *Biochem Biophys Res Commun* 2003, *308*, 1-11.
[60] Polevoda, B., Sherman, F., N-terminal Acetyltransferases and Sequence Requirements for N-terminal Acetylation of Eukaryotic Proteins. *J Mol Biol* 2003, *325*, 595-622.
[61] Aivaliotis, M., Gevaert, K., Falb, M., Tebbe, A., *et al.*, Large-Scale Identification of N-Terminal Peptides in the Halophilic Archaea Halobacterium salinarum and Natronomonas pharaonis. *J Proteome Res* 2007, *6*, 2195-2204.
[62] Lis, H., Sharon, N., Protein Glycosylation - Structural and Functional-Aspects. *Eur J Biochem* 1993, *218*, 1-27.
[63] Apweiler, R., Hermjakob, H., Sharon, N., On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* 1999, *1473*, 4-8.
[64] Yu, Y. Q., Gilar, M., Kaska, J., Gebler, J. C., A rapid sample preparation method for mass spectrometric characterization of N-linked glycans. *Rapid Commun Mass Spectrom* 2005, *19*, 2331-2336.
[65] Yu, Y. Q., Fournier, J., Gilar, M., Gebler, J. C., Identification of N-linked glycosylation sites using glycoprotein digestion with pronase prior to MALDI tandem time-of-flight mass spectrometry. *Anal Chem* 2007, *79*, 1731-1738.
[66] Hagglund, P., Matthiesen, R., Elortza, F., Hojrup, P., *et al.*, An Enzymatic Deglycosylation Scheme Enabling Identification of Core Fucosylated N-Glycans and O-Glycosylation Site Mapping of Human Plasma Proteins. *J Proteome Res* 2007, *6*, 3021-3031.
[67] Zhang, J. F., Wang, D. I. C., Quantitative analysis and process monitoring of site-specific glycosylation microheterogeneity in recombinant human interferon-gamma from Chinese hamster ovary cell culture by hydrophilic interaction chromatography. *J Chromatogr B Biomed Sci Appl* 1998, *712*, 73-82.
[68] Takegawa, Y., Deguchi, K., Ito, H., Keira, T., *et al.*, Simple separation of isomeric sialylated N-glycopeptides by a zwitterionic type of hydrophilic interaction chromatography. *J Sep Sci* 2006, *29*, 2533-2540.
[69] Kaji, H., Yamauchi, Y., Takahashi, N., Isobe, T., Mass spectrometric identification of N-linked glycopeptides using lectin-mediated affinity capture and glycosylation site-specific stable isotope tagging. *Nat Protoc* 2006, *1*, 3019-3027.
[70] Thaysen-Andersen, M., Thogersen, I. B., Nielsen, H. J., Lademann, U., *et al.*, Rapid and Individual-specific Glycoprofiling of the Low Abundance N-Glycosylated Protein Tissue Inhibitor of Metalloproteinases-1. *Mol Cell Proteomics* 2007, *6*, 638-647.
[71] Omaetxebarria, M. J., Hagglund, P., Elortza, F., Hooper, N. M., *et al.*, Isolation and characterization of glycosylphosphatidylinositol-anchored peptides by hydrophilic interaction chromatography and MALDI tandem mass spectrometry. *Anal Chem* 2006, *78*, 3335-3341.
[72] Kaji, H., Kamiie, J.-i., Kawakami, H., Kido, K., *et al.*, Proteomics reveals N-linked glycoprotein diversity in Caenorhabditis elegans and suggests an atypical translocation mechanism for integral membrane proteins. *Mol Cell Proteomics* 2007, *6*, 2100-2109.
[73] Wuhrer, M., Koeleman, C. A. M., Hokke, C. H., Deelder, A. M., Protein glycosylation analyzed by

normal-phase nano-liquid chromatography-mass spectrometry of glycopeptides. *Anal Chem* 2005, *77*, 886-894.
[74] Zhao, J., Qiu, W., Simeone, D. M., Lubman, D. M., N-linked glycosylation profiling of pancreatic cancer serum using capillary liquid phase separation coupled with mass spectrometric analysis. *J Proteome Res* 2007, *6*, 1126-1138.
[75] Gilmore, J. M., Washburn, M. P., Deciphering the combinatorial histone code. *Nat Methods* 2007, *4*, 480-481.
[76] Coon, J. J., Ueberheide, B., Syka, J. E. P., Dryhurst, D. D., *et al.*, Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc Natl Acad Sci U S A* 2005, *102*, 9463-9468.
[77] Garcia, B. A., Pesavento, J. J., Mizzen, C. A., Kelleher, N. L., Pervasive combinatorial modification of histone H3 in human cells. *Nat Methods* 2007, *4*, 487-489.
[78] Lindner, H., Sarg, B., Hoertnagl, B., Helliger, W., The microheterogeneity of the mammalian H1(0) histone. Evidence for an age-dependent deamidation. *J Biol Chem* 1998, *273*, 13324-13330.
[79] Sarg, B., Green, A., Soderkvist, P., Helliger, W., *et al.*, Characterization of sequence variations in human histone H1.2 and H1.4 subtypes. *Febs J* 2005, *272*, 3673-3683.
[80] Sarg, B., Helliger, W., Talasz, H., Forg, B., Lindner, H. H., Histone H1 phosphorylation occurs site-specifically during interphase and mitosis: identification of a novel phosphorylation site on histone H1. *J Biol Chem* 2006, *281*, 6573-6580.
[81] Sarg, B., Helliger, W., Talasz, H., Koutzamani, E., Lindner, H. H., Histone H4 hyperacetylation precludes histone H4 lysine 20 trimethylation. *J Biol Chem* 2004, *279*, 53458-53464.
[82] Pesavento, J. J., Mizzen, C. A., Kelleher, N. L., Quantitative analysis of modified proteins and their positional isomers by tandem mass spectrometry: Human histone H4. *Anal Chem* 2006, *78*, 4271-4280.

Chapter 3:
Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy

Paul J. Boersema[1,2], Nullin Divecha[3], Albert J.R. Heck[1,2] and Shabaz Mohammed[1,2]

[1] 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
[2] Netherlands Proteomics Centre
[3] Division of Cellular Biochemistry, The Netherlands Cancer Institute, Amsterdam, The Netherlands

Based on *J. Proteome Res.* 2007, *6*, 937-946.

ABSTRACT

In proteomics a digested cell lysate is often too complex for direct comprehensive mass spectrometric analysis. To reduce complexity, several peptide separation techniques have been introduced including very successful two-dimensional liquid chromatography (2D-LC) approaches. Here, we assess the potential of zwitterionic Hydrophilic Interaction Liquid Chromatography (ZIC-HILIC) as a first dimension for the analysis of complex peptide mixtures. We show that ZIC-HILIC separation is dramatically dependent on buffer pH in the range from 3 to 8, due to deprotonation of acidic amino acids. ZIC-HILIC exhibits a mixed-mode effect consisting of electrostatic and polar interactions. We developed a 2D-LC system that hyphenates ZIC-HILIC off-line with Reversed Phase (RP). The two dimensions are fairly orthogonal, and the system performs very well in the analysis of minute amounts of complex peptide mixtures. Applying this method to the analysis of 10 μg of a cellular nuclear lysate, we were able to confidently identify over 1000 proteins. Compared to strong cation exchange chromatography (SCX), ZIC-HILIC shows better chromatographic resolution and absence of clustering of prevalent +2 and +3 charged peptides. At pH 3, ZIC-HILIC separation allows best orthogonality with RP and resembles conventional SCX separation. A significant enrichment of N-acetylated peptides in the first fractions is observed at these conditions. ZIC-HILIC separation at high pH (6.8 and 8) however, enables better chromatography, resulting in more comprehensive data acquisition. With this extended flexibility we conclude that ZIC-HILIC is a very good alternative for the more conventional SCX in multidimensional peptide separation strategies.

INTRODUCTION

Two of the main challenges facing comprehensive proteomics analyses are the sheer complexity of the proteome and the huge dynamic range in protein expression [1]. To accommodate these challenges, a number of protein and peptide separation strategies are implemented to reduce the complexity for the final step of mass spectrometric detection and protein identification. Depending on what level the first dimension is performed, most separation techniques can be placed in one of two categories: protein level or peptide level. On protein level, techniques such as SDS-PAGE, isoelectric focusing (IEF), size exclusion chromatography (SEC), ion exchange chromatography (IEC) and reversed-phase (RP) chromatography have been implemented [2]. On peptide level, digestion is performed immediately after cell lysis, and peptides are then subjected to a number of, most usually, chromatographic separations. Recently, also IEF and Capillary Electrophoresis (CE) approaches have been shown to be suitable for peptide-level first dimension separation [3-6].

The end point of most proteomics experiments is a peptide mixture separated on RP at nanoliter flow rates on a liquid chromatography (LC) system coupled to a mass spectrometer (nanoLC-MS) [2]. The use of peptides rather than proteins allows for an easier automation of the process, minimizes nonspecific adsorption onto separation devices and allows a higher chromatographic resolving power [7]. Although RP chromatography has one of the highest separation powers available, it alone cannot sufficiently reduce the complexity of most peptide mixtures for comprehensive analysis by MS. A number of laboratories have attempted to improve single dimension separation by using longer columns (over 50 cm) and increasing gradients (up to 10 hours) [8]. However, such systems typically operate at ultra-high pressures, *i.e.,* in excess of 1000 Bar and, thus, cannot be implemented using routine HPLC instrumentation.

The use of chromatographic techniques in tandem to allow a higher separation power and improve comprehensiveness of an analysis has proven to be successful. The most prominent strategy to use multiple chromatographic separations is often referred to as Multidimensional Protein Identification Technology (MudPIT)[9]. Originally called direct analysis of large protein complexes (DALPC)[10], the typical approach is to create a column containing two phases (strong cation exchange (SCX) and RP) to increase resolving power. The peptides bind initially on SCX material and are stepwise eluted onto the RP part, where each fraction is further separated and analyzed by mass spectrometry [9]. Recently, MudPIT has been implemented on a ultrahigh-pressure scale allowing improvements in separation [11]. Compared to an off-line system, online hyphenation displays advantages such as minimal loss of sample, no vial contamination and no sample dilution [12, 13]. Despite the convenience brought by automation by online 2D-LC, it has been found to be rigid and a compromise on separation efficiency. Off-line fractionation removes some of these limitations, *i.e.,* loading an appropriate level of material for the second dimension is no longer an issue as it is for online approaches, where overloading of the second dimension can easily occur. Peptide separations have also been shown to be generally superior for off-line coupling, since a conventional gradient can be implemented [14]. It is also possible to perform selective, in-depth analysis of certain SCX fractions allowing to focus on a subgroup of peptides such as those that are phosphorylated [15]. The pH of the SCX elution buffer can also be modified from the necessary low pH for electrospray, which normally causes the majority of tryptic peptides to elute in a narrow window [16]. Finally, an additional advantage of off-line coupling is the use of different column materials that are not directly compatible with each other in terms of required solvents [12, 14].

Over time, a number of 2D-LC configurations have been developed, including the off-line [17] and online [18] coupling of SEC in the first dimension and RP in the second. Although SEC is compatible with RP and separation may be orthogonal, this combination is not very widespread due to the low separation power of SEC [18]. The combination of RP at pH 10 in the first dimension and RP at pH 2.6 in the second is an interesting setup for its compatibility with MS and surprisingly high orthogonality [19]. However, because the separation is correlative, this is still not the optimal 2D-LC system. It has been suggested that Hydrophilic Interaction LIquid Chromatography (HILIC) is a good alternative candidate as first dimension of a hyphenated system, showing an orthogonal separation to RP, with a separation power similar to RP and compatibility with MS [20, 21]. HILIC is characterized by the use of a hydrophilic stationary phase and a hydrophobic organic mobile phase, conditions that have been in use since 1975 [22]. The order of elution of peptides is reversed to RP, but rather than choosing the historical name Normal Phase (NP), the descriptive acronym HILIC was used to differ NP from HILIC, since NP is performed with nonaqueous, non-water-miscible solvent buffers, whereas HILIC is performed with water-miscible solvents and elution is achieved by a water gradient [23-26]. Although discussion still exists about the exact separation mechanism, it is generally accepted that an aqueous layer is formed around the hydrophilic functional groups of the HILIC material. The separation of peptides can be explained by a partitioning mechanism between the aqueous layer and the hydrophobic buffer [23], by hydrogen bonding with the HILIC-material [27] or a mechanism somewhere in between, with both partitioning and hydrogen bonding [28].
At present, a number of different materials are commercially available for HILIC, including underivatised silica that contains functional groups such as siloxanes, silanols and a small quantity of metals [24, 25, 29, 30], and derivatized silica, such as Polysulfoethyl A [23, 26], the weak anion exchanger Polycat A [31], TSKgel amide 80 [27, 32] and ZIC-HILIC [33]. They are all capable of generating an aqueous layer around their functional groups. The charge most functional groups carry contributes to the hydrophilicity of the stationary phase [23]. For a more comprehensive and detailed overview of different HILIC materials, see Hemström and Irgum [24].

In this paper, we report on the evaluation and optimization of off-line hyphenation of ZIC-HILIC with RP for the separation of minutes amounts of (complex) peptide mixtures for subsequent MS analysis. The optimized system is then used to characterize a cellular nuclear lysate.

MATERIALS AND METHODS

Chemicals and reagents

Formic acid, acetic acid and sodium chloride were purchased from Merck (Darmstadt, Germany). Acetonitrile was purchased from Biosolve (Valkenswaard, Netherlands). Dithiothreitol and ammonium bicarbonate were purchased from Fluka (Buchs, Switzerland). Trypsin and endoprotease LysC were purchased from Roche Diagnostics GmbH (Mannheim, Germany). Water used in these experiments was obtained from a Milli-Q purification system (Millipore, Bedford, MA). All other chemicals were from Sigma (St. Louis, MO).

Preparation of protein standard mixture

For the optimization and evaluation of ZIC-HILIC we prepared a model standard peptide mixture consisting of combined protein digests from bovine serum albumin (BSA) and α and β casein. Each protein was digested separately. A total of 31 μL of 45 mM DTT was added to 125 μL of 4 μg/μL dissolved protein and incubated at 50°C for 30 min. A total of 31 μL 110 mM Iodoacetamide was added, and the mixture was kept at room temperature for 30 min. The mixture was diluted 5 times with 50 mM ammonium bicarbonate and 10 μg of trypsin was added for overnight digestion at 37°C. Digests were subsequently mixed 1:1:1 and the mixture was dried *in vacuo* ('Speedyvac', Thermo, CA) and reconstituted in buffer A for subsequent LC separation.

Cell culture, subcellular fractionation, sample preparation

Murine Erythroleukemia (MEL) cells were maintained in routine culture using 10% FCS in Dulbecco's modified Eagle medium (DMEM). Nuclei isolation was performed as published previously [34, 35]. Briefly, cells were fractionated by using a detergent-free hypotonic buffer. Separation using a sucrose cushion yields a mixed cytosol/membrane fraction and intact nuclei. A pellet of 1200 μg of nuclei was dissolved in 300 μL 8 M urea, 400 mM ammonium bicarbonate by two 5 s sonication bursts on ice. A total of 40 μg of sample was reduced in 2.5 μL 45 mM DTT for 30 min at 50°C. A total of 2.5 μL of 100 mM Iodoacetamide was added, and the mixture was kept at room temperature for 30 min. LysC (0.5 μg) was added, and the mixture was incubated at 37°C for 4 hours. The mixture was diluted 7 times, 1 μg trypsin was added and the mixture was incubated at 37°C overnight. Sample was desalted using C18 ZipTip (Millipore, Bellerica, MA), filled with extra Aqua C18 beads. The eluent was dried *in vacuo* and reconstituted in buffer A of which one-fourth (~10 μg) was used for subsequent LC separation.

LC buffers

pH 3 buffer A was 80% acetonitrile (ACN), 0.05% formic acid (FA), and buffer B was 40% ACN, 0.05% FA. pH 4.5, 6.8 and 8 buffer A was 80% ACN/ 20% water, 20 mM ammonium acetate (overall), buffer B was 40% ACN/ 60% water, 20 mM ammonium acetate (overall). SCX buffer A was 30% ACN, 0.05% FA; buffer B was 30% ACN, 0.05% FA, 0.5 M NaCl. Buffer A for nanoLC-LTQ-Orbitrap analysis was 0.5% acetic acid, buffer B was 80% ACN, 0.5% acetic acid.

LC-MS

One dimensional ZIC-HILIC separation was performed using an Agilent 1100 series LC system, with a ZIC-HILIC column (SeQuant, Umeå, Sweden) 1.0 mm x 150 mm, 3.5 μm, 200 Å. The flow was passively split from 800 μL/min to 40 μL/min before the column. Gradient elution was performed for all pH conditions similarly: 0-100% B in 43 min. Column output

was coupled to a Q-TOF Micromass spectrometer (Micromass UK Ltd., Manchester, UK) via a passive split reducing the flow to 250 nL/min. Nanospray was achieved using a distally coated fused silica emitter (New Objective, Cambridge, MA) (360 μm OD / 20 μm ID / 10 μm tip ID) biased to 2.5 kV. The mass spectrometer was operated in the positive ion mode with a resolution of 4500-5500 full-width half-maximum (FWHM) using a source temperature of 80 °C and a counter current nitrogen flow rate of 150 L/h. Data dependent analysis was employed (three most abundant ions in each cycle): 1 second MS (*m/z* 350-1500) and max 2 seconds MS/MS (*m/z* 50-2000, continuum mode), 30 seconds dynamic exclusion. A charge state recognition algorithm was employed to determine optimal collision energy for low energy CID MS/MS of peptide ions. External mass calibration using NaI resulted in mass errors of less than 50 ppm, typically 5-15 ppm in the m/z range 50-2000.

2D-LC-MS, first dimension

First dimensional ZIC-HILIC separation was performed on a Famos/Ultimate LC instrument (LC Packings, Naarden, Netherlands), using a vented column set-up. [36] The trapping column was ZIC-HILIC, 200 μm x 5 mm, 3.5 μm, 200 Å; analytical column ZIC-HILIC, 200 μm x 160 mm, 3.5 μm, 200 Å. For SCX separation the trapping column was polysulfoethyl A (PolyLC, Columbia, MD) 200 μm x 32 mm, 5 μm, 200 Å; the analytical column was polysulfoethyl A 200 μm x 105 mm, 5 μm, 200 Å. All columns were packed in-house. Trapping was performed at 2 μL/min for 10 min, analytical separation at 1.5 μL/min, passively split from 250 μL/min. Gradient elutions were chosen such that peptides elute in a similar timeframe, generally 0-80% B in 30 min, 80-100% B in 2 min, 100% for 5 min. One-minute fractions were collected in a titer plate using a Probot Microfraction collector (LC Packings), adding 40 μL of 5% FA per fraction. No additional sample modification was performed before second dimensional LC-MS.

2D-LC-MS, second dimension

A volume of 8 μL of collected fractions was used for subsequent nanoLC-LTQ-Orbitrap-MS (Thermo, San Jose, CA). An Agilent 1100 series LC system was equipped with an Aqua (Phenomenex, Torrance, CA), 50 μm x 10 mm, 5 μm, 120Å trapping column and a Reprosil (Dr. Maisch GmbH, Ammerbuch, Germany), 50 μm x 254 mm, 3 μm, 120 Å analytical column. Trapping was performed at 5 μL/min for 10 min; elution was achieved with a gradient of 0-45% B in 45 min, 45-100% B in 1 min, 100% B for 4 min, with a flow rate of 0.4 ml/min passively split to 100 nl/min. Nanospray was achieved using a distally coated fused silica emitter (New Objective, Cambridge, MA) (360 μm OD / 20 μm ID / 10 μm tip ID) biased to 1.8 kV. The mass spectrometer was operated in the data dependent mode to automatically switch between MS, MS/MS. Survey Full scan MS spectra (from *m/z* 350 – 1500) were acquired in the FT-Orbitrap with resolution R= 60 000 at *m/z* 400 (after accumulation to a target value of 500 000 in the linear ion trap). The two most intense ions were fragmented in the linear ion trap using collisionally induced dissociation at a target value of 10 000.

Protein identification

Data analysis was carried out using the Mascot (version 2.1.0) software platform (Matrix Science, London, UK). Q-TOF Micromass spectra from the protein standard mixture were searched against the UniProt-SwissProt 50.4 database with taxonomy: other mammalia, trypsin with maximal 2 missed cleavages, carbamidomethyl (C) as fixed modification and oxidation (M), N-acetylation (N-terminus) and phosphorylation (S, T, Y) as variable modifications. Peptide tolerance was set to 50 ppm with 1+, 2+ and 3+ peptide charges and MS/

MS tolerance 0.9 Da.
LTQ Orbitrap spectra from protein standard mix and nuclear lysate were respectively search against UniProt-SwissProt 50.4 database with taxonomy: other mammalia and IPI-Mouse 3.19 database. Further settings: trypsin with 2 missed cleavages, carbamidomethyl (C) as fixed modification, oxidation (M), N-acetylation (N-terminus) and phosphorylation (S, T, Y) as variable modifications. Peptide tolerance was set to 5 ppm for 2+ and 3+ charged peptides and MS/MS tolerance was 0.9 Da. A minimum peptide score was set to 20 and expect value ≤0.05. If a peptide was found in more than one fraction, the retention time was determined to 0.5 min precision, calculated from peptide intensities. Minimum Mascot protein score of 60 was used for confident identification.

RESULTS

1D ZIC-HILIC-MS

For this research, ZIC-HILIC was chosen over alternative HILIC materials due to its zwitterionic functional group. It has been suggested that such a material creates weaker ionic interactions when compared to charged HILIC-materials such as polysulfoethyl A and silica, allowing the use of buffers with lower ionic strengths [37]. The nature of ZIC-HILIC has an additional benefit in that between pH 3 and 8 the charge of the chromatographic material will not change [24]. However, changing pH within this range does have an effect on the charge state of peptide residues. Table 1 lists the charge of certain amino acids at the pH levels used in this study. Below pH 4, only basic residues and N-termini (when not acetylated) are charged, while above pH 4, acidic residues are deprotonated and become negatively charged. Increasing pH over 6 will cause hisitidines to be deprotonated and lose their charge. To assess the effect of pH and peptide charge on peptide separation, the two extreme pH conditions the silica-based ZIC-HILIC could handle (pH 3 and 8) and two intermediate pH levels (pH 4.5 and 6.8) were selected.

	pH 3	pH 4.5	pH 6.8	pH 8
K	+	+	+	+
R	+	+	+	+
H	+	+	0	0
D	0	0/-	-	-
E	0	0/-	-	-
N-term	+	+	+	+
N-term acetylated	0	0	0	0
pS, pT, pY	-	-	-	-

Table 1. *Charge state of peptide residues at the different pH values used in this paper.*

A peptide mixture consisting of BSA and α and β casein in a 1:1:1 ratio was applied on a ZIC-HILIC column, 1.0 mm x 150 mm, separated under the four different pH conditions and analyzed by MS. Peptide elution was performed with a 80-40% ACN gradient in approximately 40 minutes. LC-MS chromatograms are depicted in Figure 1a, and additional information on retention times of individual peptides is provided in the Supporting Information Table 1[1]. Examination of the retention times of the peptides shows a trend of longer retention with increasing pH. This is probably due to the fact that the interaction with ZIC-HILIC material is increased upon deprotonation of acidic residues. To further investigate

1 Supporting information for this chapter is available on http://pubs.acs.org/doi/suppl/10.1021/pr060589m

trends relating to peptide net charge changes, retention times of 'model peptides' from each sub-group were tracked with buffer pH change with their extracted ion chromatograms being shown in Figure 1b. The phosphopeptide FQpSEEQQQTEDELQDK (peptide 4) carries three charges at pH 3: the N-terminus, a Lysine and the phospho-group. When the buffer reaches above pH 4, the aspartic acid and glutamic acid residues are deprotonated, adding six extra charges to the peptide. The interaction with the chromatographic material is thus improved, consequently increasing the retention time dramatically. On the other hand, the peptide GPFPIIV (peptide 1) does not contain any acidic groups and therefore its retention is not affected by pH of the buffer. The basic peptide LKECCDKPLLEK (peptide 5) carries four positive charges at pH 3, but the addition of three extra charges when increasing the pH up to 8 seems not to affect the retention of this peptide dramatically.

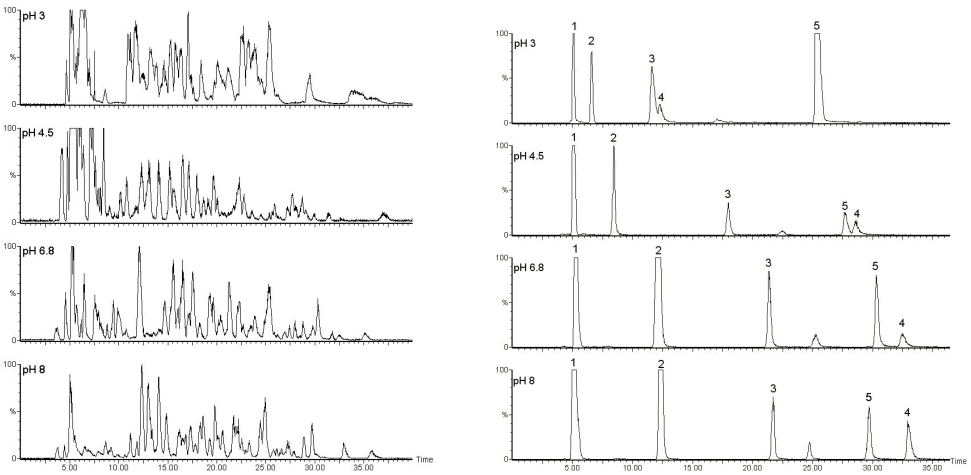


Figure 1. *LC-MS chromatograms of a standard peptide mixture using a ZIC-HILIC, 1.0 mm x 150 mm, 3.5 µm, 200 Å column online connected to a Q-TOF Micromass mass spectrometer. Buffer pH was as indicated. (a) base peak chromatograms. (b) extracted ion chromatograms of peptides (1) GPFPIIV, (2) LVTDLTK, (3) ADLAK, (4) FQpSEEQQQTEDELQDK and (5) LKECCDKPLLEK.*

The chromatographic performance of ZIC-HILIC is visibly improved upon changing the buffer. Through the use of ammonium acetate buffers at higher pH, peaks are generally sharper and more peaks are baseline separated as can be seen in Figure 1a. This improved separation can be partly explained by the nature of the tryptic peptide pool used for evaluation. At pH 3, acidic residues are neutral and so regular tryptic peptides generally carry only two or three positive charges. At higher pH levels the acidic residues are charged and so the variation in net charge of peptides increases and consequently coelution is less likely to happen, improving chromatographic performance of ZIC-HILIC.

In all four conditions, the peak intensity diminished over the run, which is probably due to the decreasing ACN gradient. The magnitude of this effect is most likely related to the suboptimal performance of the electrospray conditions applied in the test bed system [25]. In other words, lower signals for late eluting peptides are not due to the ZIC-HILIC column or separation but due to the ionization process as could be confirmed by UV detection (data not shown).

Figure 2 compares peptide retention times at different pH values. Little difference is observed between pH 6.8 and 8 which is not surprising since no pK_a or pK_b is passed. A large difference in retention times was observed between pH 3 and pH 8, confirming the effect of changing net charge state of peptides. Peptides with a longer retention time at pH 8 usually

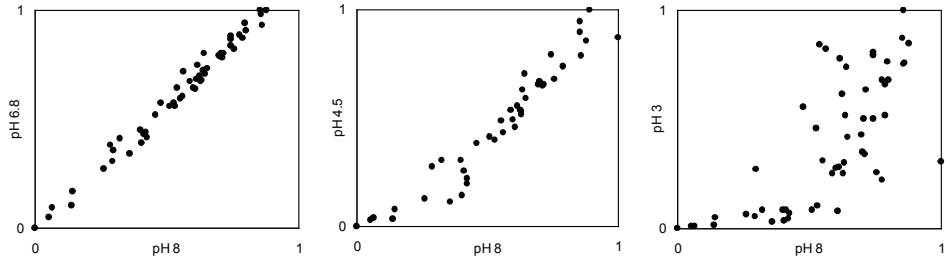


Figure 2. *Normalized peptide retention times of all peptides detected in figure 1 at different pH conditions. All peptides eluted between 5 and 40 minutes.*

have at least two acidic residues, while peptides with a higher retention time at pH 3 generally contain more than two basic residues or lack acidic residues.

2D ZIC-HILIC-RP LC-MS

Generally, the separation power of 2D-LC is superior over 1D-LC, since the peak capacity of the system is increased by the addition of an extra column. A common way to gauge the theoretical peak capacity of a 2D-LC system is by multiplying the peak capacity of both columns [38]. It can also be visualized as a dotplot with the two chromatographic dimensions making up the axes. In an ideal situation dots will cover the whole dotplot area. However, this will not occur since the separation of peptides is not carried out in an 'orderly fashion' [39]. Furthermore, the peak capacity in 2D-LC is dependent on dimensionality of the sample

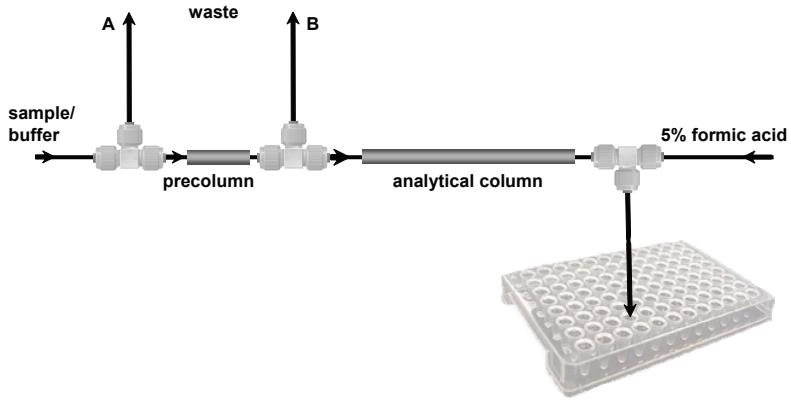


Figure 3. *Schematic presentation of the first-dimensional ZIC-HILIC separation/fractionation system. During sample loading, waste line A is closed, and the flow is 2 µL/min. During analysis waste line B is closed, and the flow is split by the first T-piece from 250 µL/min to 1.5 µL/min. The eluent is mixed with 5% formic acid and one-minute fractions are collected in a titer plate which can be directly used for subsequent nanoRP-LC-MS analysis.*

and column and the orthogonality of the two columns [20, 39].

The high organic content of the buffers used in ZIC-HILIC separations does not allow a direct hyphenation with RP. To overcome this problem, we developed an off-line 2D-LC system, based on the vented column design [36], in which peptides are initially separated over a home-made 200 µm i.d. ZIC-HILIC column operating at a flow rate of 1-2 µL/min. Instead of directly collecting the peptides, the eluent was diluted via a T-piece with a make-up solution of 5% formic acid operating at 80 µL/min (Figure 3) Thus, 1 min fractions would

now be sufficiently large and aqueous to reduce evaporation and become compatible with nanoRP chromatography. Additionally, the acidified fractions were directly placed in a 96 well plate, reducing sample handling. The second dimension was also based on a vented column system, where peptides are trapped first on a RP precolumn and then analyzed at nano flow rates. It is expected that certain hydrophilic peptides such as those that are glycosylated

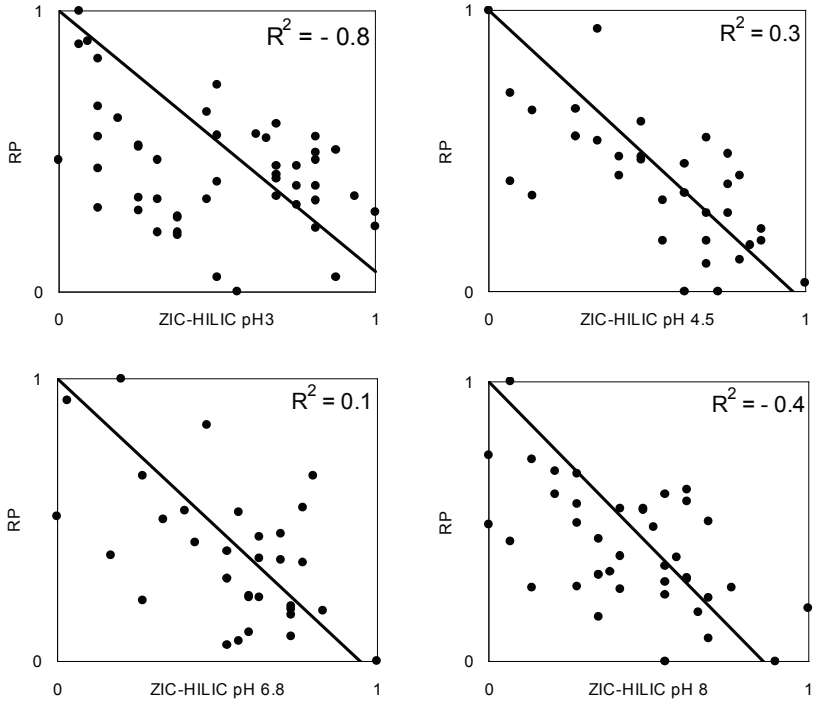


Figure 4. *Normalized peptide retention time plots, comparing ZIC-HILIC versus RP at different pH conditions. Peptides elute between 10 and 50 minutes from ZIC-HILIC and between 5 and 40 minutes from RP. The downward diagonal is the trend line, and R2 is the correlation coefficient.*

might have been retained on ZIC-HILIC, are not retained on a RP column [40].

As shown above, changing buffer pH influences the retention times of peptides. The impact of this variation on the orthogonality of ZIC-HILIC with RP was investigated by applying the peptide mixture to the 2D-LC system. The results, presented as dotplots (Figure 4), provide an impression of the system's orthogonality at different ZIC-HILIC pH conditions. A system that merely has a reverse elution pattern compared to RP would show a downward diagonal from left to right. The apparent deviation from such a downward trend line indicates that ZIC-HILIC is not simply the reverse of RP. In fact, at pH 3 and 8 the correlation coefficient (R^2) is negative, demonstrating that the trend line was forced to be a diagonal and is far from the best fit. In conclusion, ZIC-HILIC seems to perform well as a first dimension at all four pH conditions, displaying the highest orthogonality at pH 3.

As implied by the results from the one-dimensional ZIC-HILIC experiments, ZIC-HILIC at pH 3 shows an elution profile that resembles that of an SCX system. To allow detailed comparison of ZIC-HILIC with SCX, the same peptide mix was analyzed using the off-line fractionation system in which the ZIC-HILIC column was replaced by a polysulfoethyl A column. Under the modified conditions using SCX, a distinctive separation of the charged

peptide sub-groups is observed, with nearly no overlap. Peptides with a net charge of 1+, including three phosphopeptides, elute first, followed by peptides with a net charge of respectively 2+, 3+ and finally those that contain more than 3+ charges. When performing ZIC-HILIC at pH 3, peptides with a net charge of 1+ also elute first, followed by peptides with a net charge of 2+, 3+ and more than 3+. Figure 5 highlights the resemblance of ZIC-HILIC at pH 3 with SCX. However, with ZIC-HILIC, separation with respect to peptide charge is less distinct and more overlap between differently charged peptides occurs. The three phosphopeptides with a net charge of 1+ are co-eluting with 2+ peptides. A hydrophilic effect is also visible: within a group of similarly charged peptides a reverse correlation is apparent. This results in a positive effect on the distribution of similar peptides over the chromatogram. When using SCX conditions, peptides with a net charge of 2+ and 3+ elute during about one third of the gradient time, whereas with ZIC-HILIC, those peptides elute during almost the whole gradient. Since those peptides are most prevalent, the possibility of

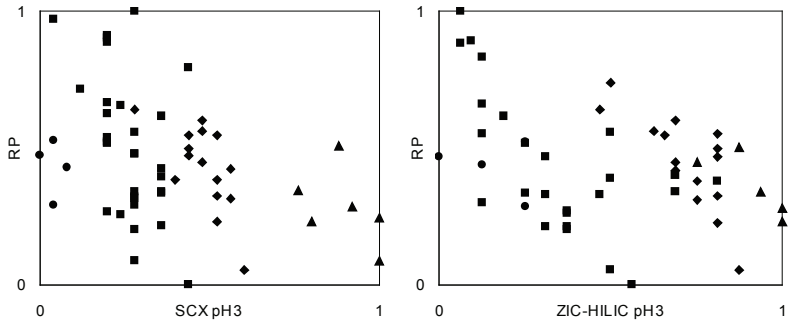


Figure 5. *Normalized retention time plots for SCX-RP and ZIC-HILIC-RP. The net charge of the peptides are indicated with (●) 1+, (■) 2+, (♦) 3+ and (▲) >3+. (Left panel) SCX pH 3 versus RP, (right panel) ZIC-HILIC pH 3 versus RP.*

separating them in more fractions, as made possible by ZIC-HILIC, is preferred.

Analysis of a nuclear fraction of a cellular lysate

One of the challenges of proteomics is to identify proteins in a very large dynamic range that can be more than 10 orders of magnitude [41]. Nuclear fractions represent one of the most difficult proteomes to analyze. One way to cope with this complexity is to use extensive separation methods. We examined the practicality of the optimized 2D-ZIC-HILIC-RP system to investigate the nuclear proteome. Nuclei were isolated from Murine erythroleukemia-cells and proteins were digested in-solution with trypsin, yielding a sample that is complex in the number of proteins and the dynamic range at which they are expressed, with histones being highly abundant. Separation of 10 μg of nuclear proteins was performed in the first dimension using ZIC-HILIC at pH 3 (showing best orthogonality with RP) and pH 6.8 (showing best separation), while the second dimension consisted of a 60 min gradient analysis on a nanoRP-LC-MS. The use of ZIC-HILIC at pH 3 as a first dimension allowed the identification of 1040 proteins with 4973 unique peptides and 1284 proteins with 6625 peptides at pH 6.8 (MS/MS spectra and information on protein and peptide scores are available at https://bioinformatics.chem.uu.nl/supplementary/boersema_jpr). Applying the same sample solely on the second RP dimension, only 367 proteins could be confidently identified with 1230 unique peptides. Comparison between the analyses at the two different pHs an overlap of 772 proteins (50%) was observed (Figure 6). In total, 268 proteins were exclusively identified at pH 3 and 512 exclusively at pH 6.8. This might be explained by either undersampling or separation power or both. It has been shown many times that the overlap be-

tween two subsequent 'MudPIT' type experiments is poor, often below 50%. Such a low level of reproducibility is related to the speed of sequencing by the mass spectrometer and the overwhelmingly complex proteolytic analyte [42]. For instance, the reproducibility and level

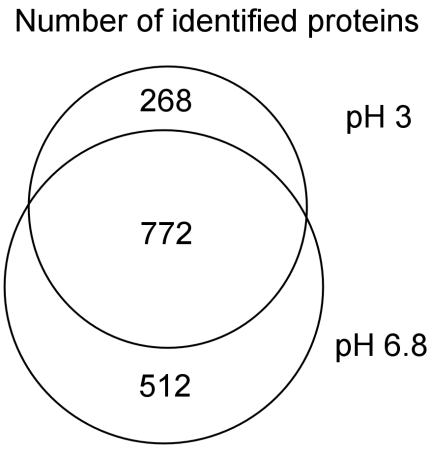


Figure 6 *Venn diagram showing the number of proteins identified under either or both pH ZIC-HILIC separation conditions.*

of random sampling was investigated for MudPIT LCQ-MS analysis of proteins of a yeast cell lysate [42]. 24% of proteins were found in just one analysis and only 35.4% of proteins were identified in all of the analyses. Although this situation of low overlap between subsequent MudPIT experiments is constantly improving with new and faster mass spectrometers with better dynamic range as Shen *et. al*. [43] (amongst others) have shown, improvements are still required. However, the increase in identifications observed by increasing the pH of the 2D-LC system cannot be solely explained by undersampling. The identification of nearly an extra 1700 peptides is significant and most likely an effect of the change of conditions. As shown in Figure 1, the peptide standard mixture revealed a better separation with the higher pH conditions. Such an improvement in chromatography will help analysis of complex samples by reducing the emphasis on the mass spectrometer, therefore allowing more peptides to be sequenced and identified. The observations suggest that ZIC-HILIC at pH 6.8 is better than pH 3 for a more comprehensive analysis. As discussed previously, RP has one of the highest separation powers. A 2D system combining RP at different pH levels would also result in comprehensive results. However, in such a system a pH level as high as 10 needs to be used [19]; under such conditions, there is a risk peptides such as those that are phosphorylated may start to degrade.

Apart from the identification of highly abundant proteins such as histones, tubulin, actin and an almost complete set of ribosomal proteins and proteins involved in the citric acid cycle, also lower abundant proteins were identified. Those lower expressed proteins include transcription factors and members of the SWI-SNF complex and the phosphatidyl inositol signaling pathway. The observation of these proteins indicate that the here presented ZIC-HILIC-RP method allows a reasonable dynamic range coverage. Functional annotation and categorization of more than 70% of the identified proteins was carried out with the online tool DAVID Bioinformatic Resources 2006 (Supplemental figure 1; http://david.abcc.ncifcrf.gov/) [44]. As expected, most proteins identified are involved in DNA or protein binding,

including histones and ribosomal proteins, which are the most abundant proteins in the nucleus. On the other hand, also a number of proteins with a more specialized or rare functionality are identified.

The complexity of biological samples in proteomics is further increased by the existence of post-translational modifications (PTMs) of proteins. Identification of these typically lower abundant modifications can be critical for the elucidation of cellular processes. Recently, the ability of SCX to isolate peptides with a specific net charge has been exploited for the enrichment of phosphorylated peptides, based on the fact that the net charge of a phosphorylated peptide is lower than its analogous non-phosphorylated peptide [15]. The nuclear fraction analysis provides a far larger dataset for elution trend analysis than the peptide standard mixture. Analysis of the effect of peptide net charge on retention time once again demonstrates the similarity of ZIC-HILIC at pH 3 to SCX (figure 7). Although ZIC-HILIC separation resembles SCX separation, no obvious enrichment of phosphorylated peptides was observed at either of the pH conditions (Figure 8), although it must be stated that phosphopeptides in our study represent only a few percent of the total peptide amount. The functional group of SCX material only contains a negative charge, repelling the similarly charged phospho-group. ZIC-HILIC, however, also contains a positive group that can attract the phospho-group. Moreover, the phospho-moiety increases the hydrophilicity of the peptide, enhancing the retention instead of decreasing it. Total phosphopeptide statistics need to be improved before firm conclusions can be made.

Another PTM that influences the charge of a peptide is N-acetylation. The positive charge of the N-terminus of a peptide is removed by the addition of an acetyl group, resulting in a reduction of both the hydrophilicity and the number of charges on the peptide. At pH 3, N-acetylated peptides tend to cluster in the first fractions of the ZIC-HILIC separation, indicat-

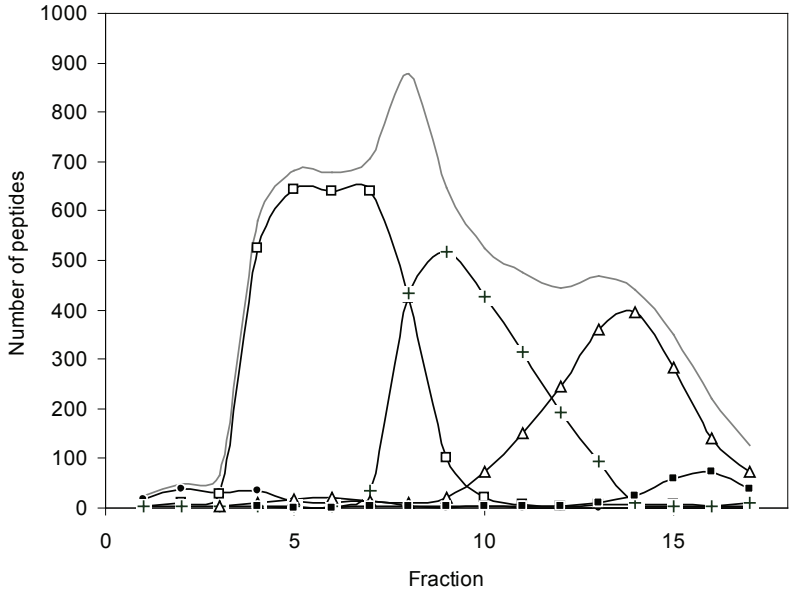


Figure 7. *Distribution of peptides after ZIC-HILIC fractionation at pH 3 of MEL-cells nuclear extract digest. The number of peptides per fraction are plotted as indicated: (–) total, net charge: (-●-) 1+; (-□-) 2+; (-+-) 3+; (-Δ-) 4+; and (-■-) >4+.*

ing that enrichment can be performed. Sixty percent of the peptides in the first two fractions are N-acetylated peptides. At pH 6.8, such a fractional enrichment for N-acetylated peptides disappears. It is possible that deprotonated acidic residues could compensate for the loss of one charge by N-acetylation at this pH. The use of ZIC-HILIC appears more suitable for analysis of such N-terminal peptides than SCX, since in ZIC-HILIC, N-acetylated peptides are more exclusively separated while with SCX similar elution profiles for both phosphorylated and N-acetylated peptides are achieved.

The formation of N-terminal pyroglutamic acid is a peptide modification that is analogous to N-acetylation in that it also removes the N-terminal charge. As was expected these modified peptides also show up in the first fractions of the ZIC-HILIC separation (Supporting Information Figure 2).

C-terminal peptides lack a basic residue and thus carry the same charge as an N-acetylated peptide. However, it seems that C-terminal peptides are not enriched in the very first fractions with admittedly, too low a number of identified C-terminal peptides for reliable statistics (supplemental figure 2). One can hypothesize that the loss of a positive charge in the case of N-terminal acetylation is different from the lack of one positive charge of a tryptic C-terminal peptide and that the charge distribution on a C-terminal peptide makes it more polar than an N-acetylated peptide.

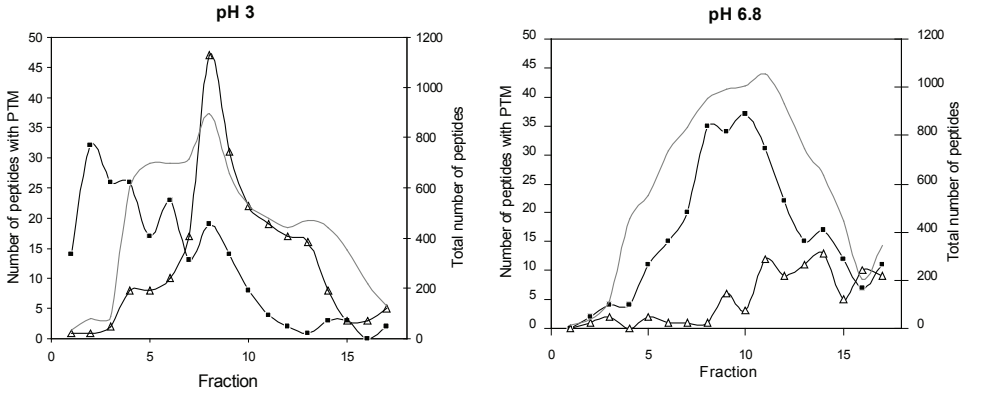


Figure 8. *Distribution of phosphorylated and N-acetylated peptides over ZIC-HILIC fractions. First dimension: ZIC-HILIC, 200 μm x 160 mm, 3.5 μm, 200 Å. Flow rate 1.5 μL/min after splitting, one-minute fractions. Number of peptides: (–) total; (-■-) N-acetylated; and (-Δ-) phosphorylated.*

DISCUSSION

Our results obtained for the protein standard mixture and the cellular nuclear lysate reveal that ZIC-HILIC separation is showing a mixed-mode effect of both polar and electrostatic interactions. An obvious indicator that electrostatic interactions were playing a role was the need to remove all salts before analysis; otherwise, little retention was observed. A similar mixed-mode effect was reported before, with the use of the SCX material polysulfoethyl A at high (>50%) ACN conditions, where elution was achieved using a sodium perchlorate gradient [45]. Separation with this material is only based on hydrophilicity and positive charges, but was believed to be rivaling RP for the separation of peptides. Furthermore, a setup that combines both a weak anion exchanger and weak cation exchanger was used for separation at high ACN (>50%) conditions. Such a system required a combined aqueous and pH gradient for elution [46]. Here, we report a system that shows a mixed-mode HILIC/ion exchange separation with an aqueous gradient which is sufficient for elution with a superior separation to SCX using a salt gradient.

The use of SCX in 2D-LC systems has been shown to be useful for the separation of complex peptide mixtures [9, 11, 14]. However, the separation power of SCX is not optimal, and also unwanted clustering of the most prevalent peptides (net charge +2 and +3) is observed [19]. The peak capacity of HILIC is shown to be higher than that of SCX [20], and we demonstrated here that the +2 and +3 charged peptides are eluting over a wider time window due to the mixed-mode way of separation of ZIC-HILIC at pH 3. With ZIC-HILIC at pH 6.8 or 8, this effect is nearly abolished, whereby chromatographic resolution has improved. All this increases the separation power of the 2D-LC set-up and decreases the eventual coelution into the mass spectrometer, allowing more comprehensive data acquisition.
The separation on a ZIC-HILIC column shows dependency on the pH of the buffer, caused by changes in peptide composition rather than chromatographic material. The effect of acidic residues underlying the improvement in separation, has been previously noted [47]. A sudden decrease in retention was observed when the pH of the buffer dropped below the pK_a of the acid, an effect that was also observed with different HILIC materials, including ZIC-HILIC. We also observed that using ZIC-HILIC the basic residues appear to have a larger effect on retention time than acidic residues, possibly due to the fact that the sulfonic group on ZIC-HILIC is a distal charged moiety, giving the material a low negative excess charge [24]. It is possible that reversing the positioning of the acidic and basic components on the silica particles can further enhance the acidic influenced better separation of tryptic peptides at pH 8.

CONCLUSIONS

Although the zwitterionic charge of ZIC-HILIC is pH independent, the ability to separate peptides is influenced by buffer pH. The primary reason for pH influence is the different peptide net charge distributions at different pH levels. The change in peptide net charge not only changes the interaction of the peptide with the chromatographic surface but has the knock-on effect of producing a different separation. At higher pH conditions (pH 6.8 and 8), separation power is highest; at pH 3 orthogonality with RP is best. The off-line hyphenation of ZIC-HILIC with RP showed successful in the analysis of the nuclear proteome. Although the most comprehensive analysis is achieved at pH 6.8 in the first dimension, separation at pH 3 can be used for the enrichment of certain PTMs. N-acetylated peptides clustered in the first few fractions of ZIC-HILIC. We can hypothesize that enrichment for also other PTMs can be achieved by ZIC-HILIC. Peptides that are formylated, carboxylated, hydroxylated (all N-terminal charge removed) and palmitoylated, myrisotylated (N-terminal charge removed and hydrophobic group attached) will all show up in the first fractions of a ZIC-HILIC run at pH 3.
Although ZIC-HILIC, unlike SCX, cannot be coupled directly to RP, it shows better peak capacity, and no unwanted clustering of most prevalent peptides is occurring. Consequently, ZIC-HILIC-RP should allow more comprehensive data acquisition from complex peptide mixtures.
Therefore we conclude that the presented ZIC-HILIC-RP LC-MS setup is a useful alternative for 2D-LC in proteomics, with separation that to some extent can be tailored to the research question. Most comprehensive results can be obtained by using ZIC-HILIC at pH 6.8, but pH 3 can be used when the interest is PTMs.

ACKNOWLEDGEMENTS
We thank Dr. Bas van Balkom for critically reviewing this manuscript. This work was supported by The Netherlands Proteomics Centre (http://www.netherlandsproteomicscentre.nl/).

Supporting Information Available: A table listing peptides identified in one dimensional ZIC-HILIC-MS including retention times at different pH conditions, a pie chart displaying the functional annotation of proteins identified in the MEL cells nuclear lysate and a figure displaying the distribution of C-terminal and N-terminal pyroglutamic peptides over ZIC-HILIC fractions at pH 3. This material is available free at http://pubs.acs.org/doi/suppl/10.1021/pr060589m.
MS/MS spectra are available as a Scaffold-file at https://bioinformatics.chem.uu.nl/supplementary/boersema_jpr. This file includes protein and peptide scoring and information on PTMs. The Scaffold viewer is also available for download.

REFERENCES

[1] Rabilloud, T., Two-dimensional gel electrophoresis in proteomics: Old, old fashioned, but it still climbs up the mountains. *Proteomics* 2002, *2*, 3-10.

[2] Wang, H., Hanash, S., Multi-dimensional liquid phase based separations in proteomics. *J Chromatogr B* 2003, *787*, 11-18.

[3] Krijgsveld, J., Gauci, S., Dormeyer, W., Heck, A. J. R., In-gel isoelectric focusing of peptides as a tool for improved protein identification. *J Proteome Res* 2006, *5*, 1721-1730.

[4] Simpson, D. C., Smith, R. D., Combining capillary electrophoresis with mass spectrometry for applications in proteomics. *Electrophoresis* 2005, *26*, 1291-1305.

[5] Wall, D. B., Kachman, M. T., Gong, S. Y., Hinderer, R., *et al.*, Isoelectric focusing nonporous RP HPLC: A two-dimensional liquid-phase separation method for mapping of cellular proteins with identification using MALDI-TOF mass spectrometry. *Anal Chem* 2000, *72*, 1099-1111.

[6] Cargile, B. J., Talley, D. L., Stephenson, J. L., Immobilized pH gradients as a first dimension in shotgun proteomics and analysis of the accuracy of pI predictability of peptides. *Electrophoresis* 2004, *25*, 936-945.

[7] Nagele, E., Vollmer, M., Horth, P., Vad, C., 2D-LC/MS techniques for the identification of proteins in highly complex mixtures. *Expert Rev Proteomics* 2004, *1*, 37-46.

[8] Luo, Q. Z., Shen, Y. F., Hixson, K. K., Zhao, R., *et al.*, Preparation of 20-mu m-i.d. silica-based monolithic columns and their performance for proteomics analyses. *Anal Chem* 2005, *77*, 5028-5035.

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

[10] Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., *et al.*, Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 1999, *17*, 676-682.

[11] Motoyama, A., Venable, J. D., Ruse, C. I., Yates, J. R., Automated ultra-high-pressure multidimensional protein identification technology (UHP-MudPIT) for improved peptide identification of proteomic samples. *Anal Chem* 2006, *78*, 5109-5118.

[12] Peng, J. M., Elias, J. E., Thoreen, C. C., Licklider, L. J., Gygi, S. P., Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: The yeast proteome. *J Proteome Res* 2003, *2*, 43-50.

[13] Wagner, K., Miliotis, T., Marko-Varga, G., Bischoff, R., Unger, K. K., An automated on-line multidimensional HPLC system for protein and peptide mapping with integrated sample preparation. *Anal Chem* 2002, *74*, 809-820.

[14] Vollmer, M., Horth, P., Nagele, E., Optimization of two-dimensional off-line LC/MS separations to improve resolution of complex proteomic samples. *Anal Chem* 2004, *76*, 5180-5185.

[15] Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., *et al.*, Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* 2004, *101*, 12130-12135.

[16] Gilar, M., Daly, A. E., Kele, M., Neue, U. D., Gebler, J. C., Implications of column peak capacity on the separation of complex peptide mixtures in single- and two-dimensional high-performance liquid chromatography. *J Chromatogr A* 2004, *1061*, 183-192.

[17] Lemieux, L., Amiot, J., Application of Reversed-Phase High-Performance Liquid-Chromatography to the Separation of Peptides from Phosphorylated and Dephosphorylated Casein Hydrolysates. *J Chromatogr* 1989, *473*, 189-206.

[18] Opiteck, G. J., Jorgenson, J. W., Anderegg, R. J., Two-dimensional SEC/RPLC coupled to mass spectrometry for the analysis of peptides. *Anal Chem* 1997, *69*, 2283-2291.

[19] Gilar, M., Olivova, P., Daly, A. E., Gebler, J. C., Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J Sep Sci* 2005, *28*, 1694-1703.

[20] Gilar, M., Olivova, P., Daly, A. E., Gebler, J. C., Orthogonality of separation in two-dimensional liquid chromatography. *Anal Chem* 2005, *77*, 6426-6434.

[21] Wang, X. D., Li, W. Y., Rasmussen, H. T., Orthogonal method development using hydrophilic interaction chromatography and reversed-phase high-performance liquid chromatography for the determination of pharmaceuticals and impurities. *J Chromatogr A* 2005, *1083*, 58-62.

[22] Linden, J. C., Lawhead, C. L., Liquid chromatography of saccharides. *J Chromatogr A* 1975, *105*, 125-133.

[23] Alpert, A. J., Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J Chromatogr* 1990, *499*, 177-196.

[24] Hemström, P., Irgum, K., Hydrophilic interaction chromatography. *J Sep Sci* 2006, *29*, 1784-1821.

[25] Naidong, W., Bioanalytical liquid chromatography tandem mass spectrometry methods on underi-

vatized silica columns with aqueous/organic mobile phases. *J Chromatogr B Analyt Technol Biomed Life Sci* 2003, *796*, 209-224.
[26] Zhu, B.-Y., Mant, C. T., Hodges, R. S., Hydrophilic-interaction chromatography of peptides on hydrophilic and strong cation-exchange columns. *J Chromatogr A* 1991, *548*, 13-24.
[27] Yoshida, T., Peptide separation in normal phase liquid chromatography. *Anal Chem* 1997, *69*, 3038-3043.
[28] Berthod, A., Chang, S. S. C., Kullman, J. P. S., Armstrong, D. W., Practice and mechanism of HPLC oligosaccharide separation with a cyclodextrin bonded phase. *Talanta* 1998, *47*, 1001-1012.
[29] Nikolov, Z. L., Reilly, P. J., Retention of carbohydrates on silica and amine-bonded silica stationary phases: application of the hydration model. *J Chromatogr A* 1985, *325*, 287-293.
[30] Kirkland, J. J., Dilks, C. H., Destefano, J. J., Normal-Phase High-Performance Liquid-Chromatography with Highly Purified Porous Silica Microspheres. *J Chromatogr* 1993, *635*, 19-30.
[31] Lindner, H., Sarg, B., Meraner, C., Helliger, W., Separation of acetylated core histones by hydrophilic-interaction liquid chromatography. *J Chromatogr A* 1996, *743*, 137-144.
[32] Tomiya, N., Awaya, J., Kurono, M., Endo, S., *et al.*, Analyses of N-linked oligosaccharides using a two-dimensional mapping technique. *Anal Biochem* 1988, *171*, 73-90.
[33] Idborg, H., Zamani, L., Edlund, P. O., Schuppe-Koistinen, I., Jacobsson, S. P., Metabolic fingerprinting of rat urine by LC/MS Part 1. Analysis by hydrophilic interaction liquid chromatography-electrospray ionization mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* 2005, *828*, 9-13.
[34] Clarke, J. H., Letcher, A. J., D'Santos, C. S., Halstead, J. R., *et al.*, Inositol lipids are regulated during cell cycle progression in the nuclei of murine erythroleukaemia cells. *Biochem J* 2001, *357*, 905-910.
[35] Cocco, L., Gilmour, R. S., Ognibene, A., Letcher, A. J., *et al.*, Synthesis of polyphosphoinositides in nuclei of Friend cells. Evidence for polyphosphoinositide metabolism inside the nucleus which changes with cell differentiation. *Biochem J* 1987, *248*, 765-770.
[36] Meiring, H. D., van der Heeft, E., ten Hove, G. J., de Jong, A., Nanoscale LC-MS(n): technical design and applications to peptide and protein analysis. *J Sep Sci* 2002, *25*, 557-568.
[37]
[38] Giddings, J. C., Concepts and comparisons in multidimensional separation. *J High Resolut Chrom* 1987, *10*, 319-323.
[39] Giddings, J. C., Sample dimensionality: A predictor of order-disorder in component peak distribution in multidimensional separation. *J Chromatogr A* 1995, *703*, 3-15.
[40] Hagglund, P., Bunkenborg, J., Elortza, F., Jensen, O. N., Roepstorff, P., A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J Proteome Res* 2004, *3*, 556-566.
[41] Anderson, N. L., Anderson, N. G., The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 2002, *1*, 845-867.
[42] Liu, H. B., Sadygov, R. G., Yates, J. R., A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 2004, *76*, 4193-4201.
[43] Shen, Y. F., Zhang, R., Moore, R. J., Kim, J., *et al.*, Automated 20 kpsi RPLC-MS and MS/MS with chromatographic peak capacities of 1000-1500 and capabilities in proteomics and metabolomics. *Anal Chem* 2005, *77*, 3090-3100.
[44] Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., *et al.*, DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol* 2003, *4*.
[45] Zhu, B. Y., Mant, C. T., Hodges, R. S., Mixed-Mode Hydrophilic and Ionic Interaction Chromatography Rivals Reversed-Phase Liquid-Chromatography for the Separation of Peptides. *J Chromatogr* 1992, *594*, 75-86.
[46] Strege, M. A., Stevenson, S., Lawrence, S. M., Mixed mode anion-cation exchange/hydrophilic interaction liquid chromatography-electrospray mass spectrometry as an alternative to reversed phase for small molecule drug discovery. *Anal Chem* 2000, *72*, 4629-4633.
[47] Guo, Y., Gaiki, S., Retention behavior of small polar compounds on polar stationary phases in hydrophilic interaction chromatography. *J Chromatogr A* 2005, *1074*, 71-80.

Chapter 4:
Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates

Paul J. Boersema[1,2], Thin Thin Aye[1,2], Toon A.B. van Veen[3], Albert J.R. Heck[1,2] and Shabaz Mohammed[1,2]

[1] 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

[2] Netherlands Proteomics Centre

[3] Department of Medical Physiology, University Medical Center Utrecht, Yalelaan 50, 3584 CM, Utrecht, The Netherlands

Based on *Proteomics* 2008, *8*, 4624-4632

ABSTRACT

Stable isotope labeling is at present one of the most powerful methods in quantitative proteomics. Stable isotope labeling has been performed at both the protein as well as the peptide level using either metabolic or chemical labeling. Here, we present a straightforward and cost-effective triplex quantification method that is based on stable isotope dimethyl labeling at the peptide level. Herein, all proteolytic peptides are chemically labeled at their α- and ε-amino groups. We use three different isotopomers of formaldehyde to enable the parallel analysis of three different samples. These labels provide a minimum of 4 Da mass difference between peaks in the generated peptide triplets. The method was evaluated based on the quantitative analysis of a cell lysate, using a typical 'shotgun' proteomics experiment. While peptide complexity was increased by introducing three labels, still more than 1300 proteins could be identified using 60 μg of starting material, whereby more than 600 proteins could be quantified using at least 4 peptides per protein. The triplex labeling was further utilized to distinguish specific from aspecific cAMP binding proteins in a chemical proteomics experiment using immobilized cAMP. Thereby, differences in abundance ratio of more than two orders of magnitude could be quantified.

INTRODUCTION

The aspiration of most proteomics experiments is to differentially quantify protein abundance. To achieve this, several quantification techniques based upon introducing stable isotopes into the samples of interest have been introduced [1-3]. Stable isotope labeling of proteins or peptides allows samples to be analyzed and compared in a single analytical experiment [1, 4, 5]. The isotope label can be introduced into the sample via metabolic or chemical labeling. A popular form of metabolic labeling is stable isotope labeling by amino acids in cell culture (SILAC), which utilizes *in vivo* incorporation of specific auxotrophic amino acids [4]. Alternatively, the stable isotope label can be introduced chemically as a tag on specific amino acids at the protein (*e.g.* ICAT [6]) or at the peptide level (*e.g.* iTRAQ [7]). Multiplex labeling, *i.e.,* the use of more than two different isotopomeric labels, facilitates the comparison of multiple states and allows the analysis of, for example, dose-response curves or time courses [7-9]. Triplex SILAC has been shown in a time-course study [9], but is rather expensive and somewhat limited to studies of biological systems that can be grown in culture. In iTRAQ, a strategy that isotope labels at the peptide level quantification is performed at the MS/MS level evaluating four [7] or even eight [10] samples in parallel. However, this commercial labeling reagent is cost-prohibitive and not very stable chemically due to it being an NHS-ester. Moreover, this method generally requires instruments capable of measuring MS/MS fragments at low *m*/*z* which narrows the choice of mass spectrometers.

Here, we introduce a triplex stable isotope labeling technique that is affordable, fast, sensitive, and accurate and can be applied to virtually any proteomics experiment, including those on mammalian tissues or body fluids. The method is an adaptation of the dimethyl labeling strategy introduced by Hsu *et al.* [11] and utilizes isotopomers of formaldehyde and sodium cyanoborohydride to incorporate dimethyl labels at the α- and ε-amino groups of all proteolytic peptides. Using the three different formaldehyde isotopomers, peptide triplets are generated that exhibit a mass difference of at least 4 Da. To illustrate its applicability in different proteomics applications we performed a traditional shotgun experiment, including 2D-LC, using a mixture of 60 µg of total murine erythroleukemia (MEL) cell lysate. We show that, even though MS complexity is increased upon introducing three labels, more than 1300 proteins could still be confidently identified, whereby more than 600 proteins could be accurately quantified using a threshold of at least 4 peptide triplets per protein. We further demonstrate the generic applicability of the method in a quantitative chemical proteomics experiment, whereby the labeling was used to distinguish specific from aspecific cAMP binding proteins using immobilized cAMP. Thereby, differences in abundance ratio of more than two orders of magnitude could be quantified.

MATERIALS AND METHODS

Materials

The 8-(6-Aminohexyl)aminoadenosine-3',5'-cyclic monophosphate coupled agarose beads were purchased from Biolog (Bremen, Germany). Protease inhibitor cocktail, trypsin and endoprotease lysine-C were purchased from Roche Diagnostics GmbH (Mannheim, Germany). DTT, ammonium bicarbonate and sodium cyanoborohydride were purchased from Fluka (Buchs, Switzerland). Iodoacetamide, triethyl ammonium bicarbonate (TEAB) and formaldehyde (37%) were purchased from Sigma (St.Louis, MO). CD_2O (98% D, 20% wt), $^{13}CD_2O$ (99% ^{13}C, 98% D, 20% wt) and sodium cyanoborodeuteride (96% D) were purchased from Isotec (Miamisburg, OH). Water used in these experiments was obtained from a Milli-Q purification system (Millipore, Bedford, MA). All other chemicals were purchased from commercial sources and were of analysis grade.

Cell culture and chromatin enriched CEF preparation

CEF of MEL cells was isolated as described before [12]. Three times 20 μg of CEF was reduced in 20 μL 2.5 mM DTT for 30 min at 56°C. Iodoacetamide (2.5 μL 100 mM) was added and the mixture was kept at room temperature for 30 min. Lys-C (0.20 μg) was added and the mixture was incubated at 37°C for 4 hours. The mixture was diluted seven-fold, 0.4 μg trypsin was added and the mixture was incubated at 37°C overnight. Sample was desalted using C18 ZipTip (Millipore, Bellerica, MA), filled with extra Poros R3 beads. The eluent was dried *in vacuo* and reconstituted in 100mM TEAB for isotope labeling.

Tissue preparation

Skeletal muscle from 6 months old Wistar rats was frozen in liquid nitrogen and stored at -80°C until use. For protein isolation, approximately 100 mg of skeletal muscle was pulverized in a custom made mortar which was pre-cooled with liquid nitrogen. The powdered tissue was then transferred to 1 mL of ice-cold lysis buffer (50mM K_2HPO_4, 150mM NaCl. 0.1% Tween 20, and Protease Inhibitor cocktail) and left at RT for 5 minutes. After centrifugation at 14,000 rpm for 10 min, the soluble fraction yielded approximately 30 mg of protein as determined by Bradford assay.

Pull-down assay and digestion

Prior to each pull-down, 50 μL dry volume of immobilized cAMP beads (~300 nmol of cAMP) was washed with 1 mL of PBS buffer (50 mM K_2HPO_4, 150 mM NaCl). Prior to the pull-down assays, tissue lysates were incubated with 0, 0.5 or 10 mM ADP/GDP for 30 minutes at 4 °C. cAMP agarose beads were added to the lysate in a volume ratio of 1:100 beads to lysate. The lysate-bead suspension was incubated for 2 hours at 4 °C by rotary shaking. After spinning down at 700 rpm, the supernatant was removed and the beads were washed with 12 mL of lysis buffer to further reduce non-specific binding. Then, 30 μL of 8 M Urea and 50 mM ammonium bicarbonate was added to the beads and reduction in 2mM DTT at 56 °C was performed, followed by alkylation in 4mM iodoacetamide. 0.2 μg Lys-C was added to the pull-down and incubated for 4 hours at 37 °C. After which further digestion by 0.2 μg trypsin was performed in a 4 times diluted solution.

Samples were desalted using C18 ZipTip (Millipore, Bellerica, MA), filled with extra Poros R3 beads. The eluent was dried *in vacuo* and reconstituted in 100 mM TEAB for dimethyl labeling. A quarter of the sample was eventually analyzed by LC-MS

Dimethyl labeling
Samples (BSA digest, CEF and pull-down) were dissolved in 100 µL 100mM TEAB buffer. CH_2O (4 µL, 4%, 'light', 0 mM ADP/GDP) or 4% CD_2O ('intermediate', 0.5 mM ADP/GDP) or 4% $^{13}CD_2O$('heavy', 10 mM ADP/GDP) was added followed by 4 µL of 600 mM $NaBH_3CN$ (light and intermediate) or 4 µL of 600 mM $NaBD_3CN$ (heavy). The mixture was incubated for 1 h at room temperature. The reaction was quenched with 16 µL of 1% ammonia. Finally, 8 µL formic acid was added and the three differentially labeled samples were pooled and desalted using C18 ZipTip (Millipore, Bellerica, MA), filled with extra Poros R3 beads.

ZIC-HILIC fractionation
ZIC-HILIC was performed as described before [13] using pH 6.8 buffers. One-minute fractions (27) were collected in a titer plate using a Probot Microfraction collector (LC Packings), adding 40 µL 5% formic acid per fraction. No additional sample modification was performed before second dimensional LC-MS.

SCX fractionation
Strong cation exchange (SCX) was performed using a Zorbax BioSCX-Series II column (0.8 mm id × 50 mm length, 3.5 µm), a FAMOS autosampler (LC-packings, Amsterdam, The Netherlands), a Shimadzu LC-9A binary pump and a SPD-6A UV-detector (Shimadzu, Tokyo, Japan). SCX buffer A was 20% acetonitrile, 0.05% formic acid; SCX buffer B was 500 mM KCl 20% acetonitrile, 0.05% formic acid. After injection, a linear gradient of 1%/min solvent B was performed. A total of 30 SCX fractions (1 min each, 50 µL elution volume) were collected and dried in a vacuum centrifuge.

nanoLC-LTQ-Orbitrap-MS
Dimethyl labeled BSA digest, ZIC-HILIC fractions and the resuspended SCX fractions were subsequently analyzed by nano-LC-LTQ-Orbitrap-MS (Thermo, San Jose, CA). An Agilent 1100 series LC system was equipped with an Aqua (Phenomenex, Torrance, CA), 50 µm x 10 mm, 5 µm, 120 Å trapping column and a Reprosil (Dr. Maisch GmbH, Ammerbuch, Germany), 50 µm x 254 mm, 3 µm, 120 Å analytical column. Trapping was performed at 5 µL/min for 10 min, elution was achieved with a gradient of 0-45% B in 45 min, 45-100% B in 1 min, 100% B for 4 min, with a flow rate of 0.4 mL/min passively split to 100 nL/min. Nanospray was achieved using a distally coated fused silica emitter (New Objective, Cambridge, MA; 360 µm od / 20 µm id / 10 µm tip id) biased to 1.8 kV. The mass spectrometer was operated in the data dependent mode to automatically switch between MS and MS/MS. Survey full scan MS spectra (from *m/z* 350 – 1500) were acquired in the Orbitrap mass spectrometer with resolution R=60,000 at *m/z* 400 (after accumulation to a target value of 500,000 in the linear IT). The two most intense ions were fragmented in the linear IT using collisionally induced dissociation at a target value of 10,000.

Protein identification
Peak lists were created by Bioworks 3.3.1 while further data analysis was carried out using the Mascot (version 2.1.0) protein identification platform (Matrix Science, London, UK). MS/MS spectra of pull-down were searched against IPI-Rat version 3.36 database, CEF was searched against IPI Mouse 3.25. Further settings included; trypsin with two missed cleavages, carbamidomethyl (C) and light-dimethyl (K and N-term) as fixed modifications and oxidation (M), intermediate-dimethyl (K and N-term) and heavy-dimethyl (K and N-term) as variable modifications. Peptide tolerance was set to 20 ppm for 2+ and 3+ charged pep-

tides and MS/MS tolerance was 0.9 Da.

Quantification

MSQuant 1.4.2a13 was customized for triplex dimethylation and used for extraction and integration of ion chromatograms of all peptides. For evaluation of standard deviations, a threshold of at least 4 peptides per protein with a total Mascot score of at least 60 was used. MSQuant output was imported in the in-house developed StatQuant 1.1.0 [14] for further statistics and normalization to the median of peptide ratios.

RESULTS AND DISCUSSION

Establishment of the triplex stable isotope dimethyl labeling method

Stable isotope dimethyl labeling is a simple, fast, affordable and efficient strategy for quantitative proteomics [8, 11]. As labeling is performed after proteolysis, the method is implementable in virtually any bottom-up proteomics experiment. Dimethyl labeling is based on the reductive amination reaction of formaldehyde with the peptide N-terminus or the ε-amino group of lysine residues. The Schiff base that is formed is reduced by sodium cyanoborohydride. Using CD_2O and normal cyanoborohydride the first stable isotope label generated the intermediate labeled peptide, 4 Da heavier than the light label (see Figure 1). The new heavy label that we introduce here using $^{13}CD_2O$ and $NaBD_3CN$ generates peptides that are 4 Da heavier than the 'intermediate' label (see Figure 1). We believe this mass difference of at least 4 Da minimizes overlap of peptide isotopes and is essential for proper quantification [2, 11]. Peptide labeling was allowed to proceed for 1 h which resulted in complete labeling of all available amino groups, while no unwanted side products were observed.

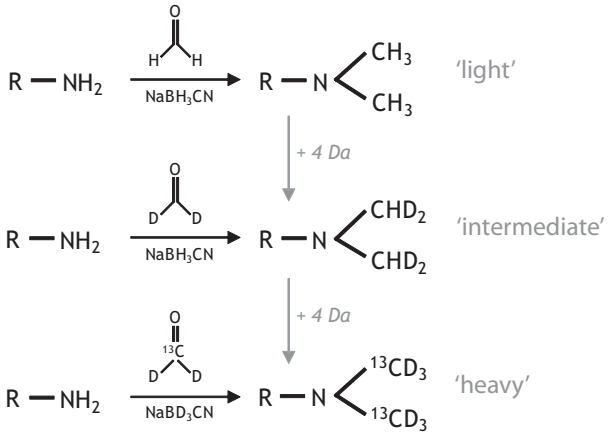


Figure 1. *Labeling schemes of triplex stable isotope dimethyl labeling to derive three different dimethyl labels with 4 Da mass shifts.*

We first tested the method on a simple BSA tryptic digest. Illustrative triplet peaks for a number of BSA peptides are shown in Figure 2 all exhibiting the expected 1:1:1 ratio between light, intermediate and heavy labels. In Figure 2A a triplet is displayed that represents a peptide that lacks a lysine residue and thus, only the N-terminus is available for labeling, resulting in a 4 Da difference between the differentially labeled peptides. Had there been a smaller mass difference, the isotope envelopes of the differentially labeled peptide would have overlapped, complicating quantification, thus illustrating the need for a minimal mass shift of 4 Da between isotope labels [2, 11]. Figure 2B and 2C represent peptides with one and two lysine residues, respectively. The mass difference is accordingly larger through the extra labeling sites. It has been stated that this labeling of multiple sites on tryptic peptides would complicate the quantification [15]. However, in our approach different numbers of labeling sites are not a significant problem as quantification is performed after peptide identification and the mass difference can be correctly calculated based on the peptide sequence. Also, as Mascot considers the N-terminus a separate entity, dimethyl labeled peptides with an N-terminal lysine (due to missed cleavages), which would involve two modifications at one amino acid residue, can be identified and quantified too. An average ratio of 1:1.06(+/-0.2):1.01(+/-0.2) was found for the differentially labeled BSA digests using 17 identified peptides.

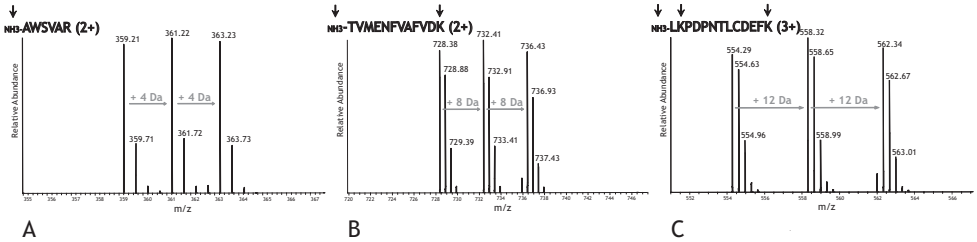


Figure 2. *Triplex dimethyl labeling of BSA mixed 1:1:1. Labeling sites are indicated with arrows. Respectively, 4, 8 and 12 Da mass shifts are obtained for peptides containing 0, 1 or 2 lysine residues.*

Triplex dimethyl labeling of a cellular lysate

In principal, increasing multiplexity of isotope labeling augments its potential to analyze more samples in parallel, however, at the cost of increased complexity when mass difference labels are used [7], potentially limiting the number of protein identifications. To evaluate the effect of triplex dimethyl labeling on analytical complexity and its consequences for protein identification efficiency, we analyzed a triplexed differentially dimethyl labeled complex lysate digest under typical shotgun proteomics conditions. As a proof of principal, 3 times 20 μg of trypsin digested CEF from MEL cells were labeled 1:1:1 with light, intermediate and heavy dimethyl stable isotope labels. We employed our recently introduced 2D-LC system, combining ZIC-HILIC fractionation in the first dimension with RP-LTQ-Orbitrap-MS analysis to allow for a comprehensive analysis [13]. Adequate separation of dimethyl labeled peptides is important as the isotope envelope of a "triplet peak" of co-eluting peptides with different sequences but a similar mass might overlap. This possibility is increased when three isotope labels are introduced causing concerns regarding quantification. Interestingly, we observed, very little mass overlap of triplets of co-eluting peptides suggesting that the separation power of the ZIC-HILIC-RP system is more than sufficient for a lysate digestion. We evaluated the consequences of triplex dimethyl labeling on protein identification. We hypothesized that all abundant peptides appearing in triplets may be selected three times as often for MS/MS fragmentation leaving little MS/MS time for the lower abundant peptides to be sequenced. However, we were able to identify 1379 proteins using 60 μg starting material (the set threshold values for identification were for a protein Mascot score > 60 and at least two peptides per protein; MS/MS spectra and identifications are available at the PRIDE database, (http://www.ebi.ac.uk/pride; accession number 3534.).

Quantification was performed with MSQuant. An in-house developed program StatQuant [14] was used for further normalization and evaluation of individual peptide and protein statistics. Although protein quantification by stable isotope dimethyl labeling has been performed before using as little as one or two peptides per protein [8, 16], here, we chose to set a more stringent threshold of four peptides per protein for precise quantification and to allow standard deviations to be calculated. These restrictions reduced the number of candidates applicable for quantification to 620 proteins (see Table S1 of Supporting Information[2]). The differentially labeled samples should exhibit a ratio of 1:1:1. In agreement, the measured average protein $\log_2$ ratios of heavy/light (Figure 3) and intermediate/light were very close to zero; -0.021 (+/- 0.233) and -0.007 (+/- 0.137), respectively.

2 Supporting Information for this chapter is available at http://www3.interscience.wiley.com/journal/121448057/suppinfo

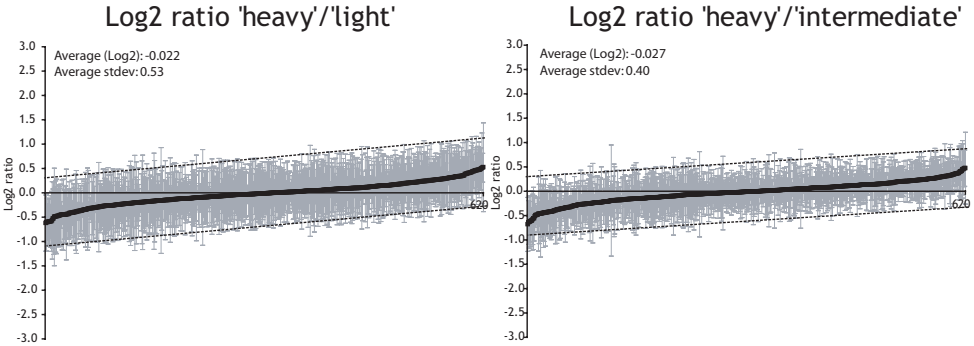


Figure 3. *Log_2 ratios and standard deviation of heavy divided by light and heavy divided by intermediate of proteins identified and quantified in mouse CEF. 20 µg of CEF was labeled 1:1:1 with light, intermediate and heavy dimethyl. After pooling, the sample was initally separated by ZIC-HILIC and analyzed by RP-LTQ-Orbitrap MS. Average Log_2 ratios and average standard deviations in Log_2 are indicated.*

Isotope effect on peptide LC retention and influence of quantification

It is well known that the introduction of stable isotopes into peptides may affect the retention time of peptides under RP-LC conditions [17], especially when the hydrogen atoms are replaced by deuterium. The mass shift between the different dimethyl labels here is obtained by deuterium and ^{13}C. Deuterium is slightly more hydrophilic than hydrogen, which might lead to LC separation of differentially labeled peptides [17]. Peptide separation in the first dimension was performed by ZIC-HILIC, which is mainly based on hydrophilicity [13]. Here, differentially labeled peptides could, potentially, end up in different fractions, and thus skew peptide ratios measured in the second RP dimension. Evaluation of the raw data led to the identification of multiple examples of peptides differentially separated over two ZIC-HILIC fractions. In the first fraction where the peptide is identified, typically more light dimethyl labeled peptide was found, while in the following fraction the heavy dimethyl peptide dominated the triplet (see Figure 4). The difference between the used heavy label and light label is six D's, while between the heavy label and intermediate label the difference is only two D's. Thus, the isotope effect is likely to be larger between the heavy label and the light than the heavy label and the intermediate label. This becomes clear from Figure 4, where the heavy labeled peptides show a larger difference in abundance ratio between consecutive ZIC-HILIC fractions than the intermediate labeled peptides. This also explains why the average standard deviation in the abundance ratio of heavy/light in Figure 3 is higher than in the abundance ratio of heavy/intermediate.

Guo *et al*. [18] recently compared the isotope effect of dimethyl labeling on LC separation by RP and HILIC. Interestingly, in their work, the retention shift between differentially labeled isoleucine and leucine that was observed with RP was abolished using HILIC. This apparent discrepancy between their and our observations is probably related to the specific HILIC phase used exhibiting differing separation mechanisms [19]. Guo *et al.* performed HILIC using non-ionic TSKgel Amide 80, while in this study zwitterionic ZIC-HILIC was applied. Although separation power and 'orthogonality' is near optimal for ZIC-HILIC at pH 6.8 [13] it appears that the separation mechanism in operation will resolve hydrogens and deuteriums and thus is not an appropriate first dimension. Therefore, in the next experiment, we used SCX as first dimension. The separation mechanism in SCX is largely based on peptide charge, and thus unlikely to be influenced by D-isotope effects, potentially making peptide quantification more manageable. The triplex dimethyl labeled sample was fractionated by

SCX and as expected, no complications relating to isotope effects were observed (see Figure 4).

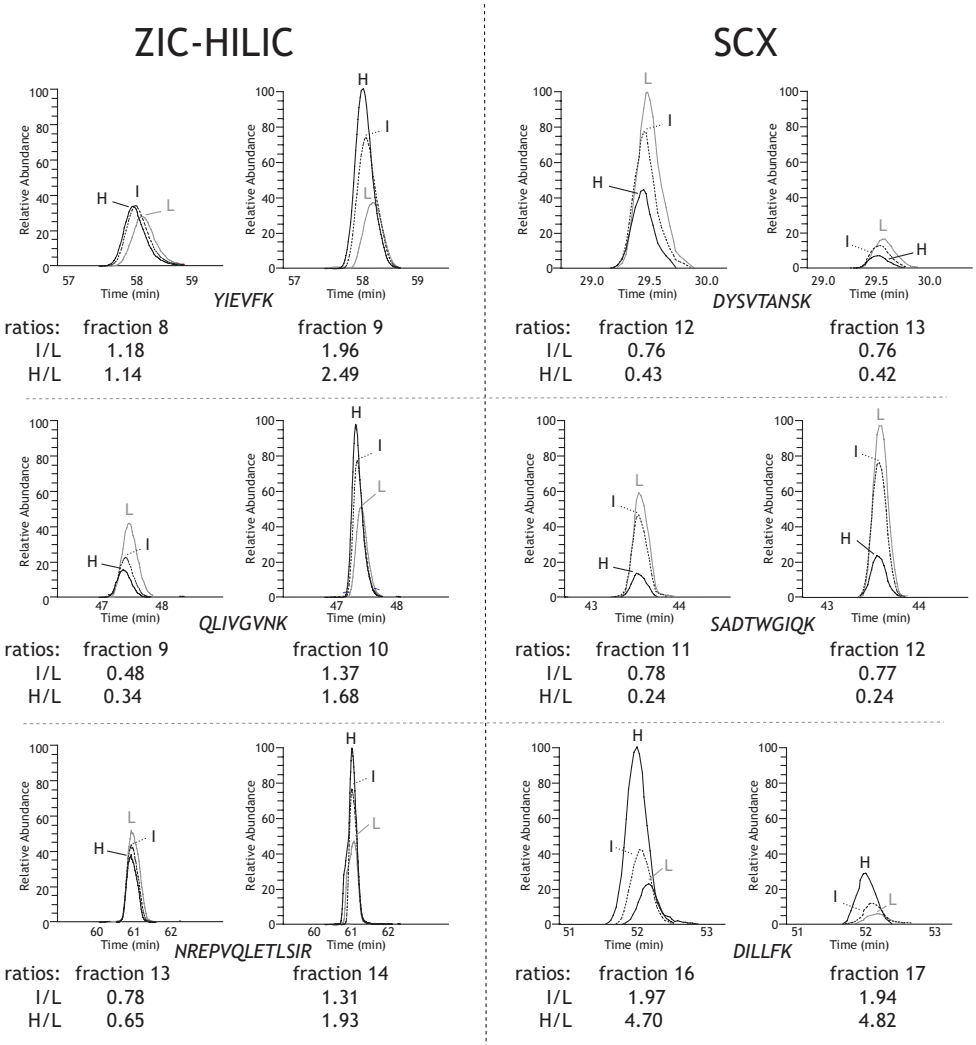


Figure 4. *Comparison of the isotope effect when fractionating triple dimethyl labeled samples with ZIC-HILIC or SCX. Shown are the RPLCMS extracted ion chromatograms of the same peptide from two consecutive ZIC-HILIC or SCX fractions. The peptide sequence is indicated in italics. Note: abundance ratios differ from one ZIC-HILIC fraction to the next due to the isotope effect, while they are similar in consecutive SCX fractions. L: light label (grey line); I: intermediate label (dashed black line); H: heavy label (solid black line).*

In the second dimension RP-LC separation, we found a retention time shift of typically around 1–5 s between light, and heavy labeled peptides (see also Figure 5). This is comparable to what has been observed previously [8, 11]. The limited extent of the effect might be attributed to the dimethyl isotope labels being placed on hydrophilic functional groups, thus minimizing the interaction of deuterium with the stationary phase [17]. Extraction of peptide peak areas suggested that the isotope effect in the second RP dimension is not a problem for quantification [2].

Quantification of protein binding to cAMP immobilized beads

Isotope labeling has also been used in experiments on protein complexes and other interactome studies to aid in evaluation whether detected proteins are genuine interactors of the bait protein/small molecule, or non-specific binders [9, 20]. In our laboratory, we have used cAMP immobilized on agarose beads to study proteins that are indirect or direct binders of cAMP, such as protein kinase A (PKA) [21, 22]. However, in such experiments typically hundreds of proteins are pulled-down. We noted that a lot of abundant AMP/ADP/ATP or GMP/GDP/GTP binding proteins are co-purified when no stringent washes are implemented [21]. We introduced an approach to remove these less-selective binding proteins from the pull-down by co-incubating these samples with ADP and GDP. These nucleotides bind to abundant AMP/ADP/ATP- or GMP/GDP/GTP-binding proteins in the lysate preventing them to bind to the beads. Here, we used the triplex dimethyl stable isotope labeling to evaluate the effect of different concentrations of ADP/GDP (*i.e.* 0, 0.5 and 10 mM ADP/GDP) used during the co-incubation.

We performed our cAMP pull-down experiments essentially as described previously [21, 22] and evaluated the cAMP "interactome" of rat skeletal muscle tissue, from 6 months old Wistar rats. Pulled down proteins were digested with Lys-C and trypsin and the three samples (containing 0, 0.5 and 10 mM ADP/GDP) were labeled with light, intermediate and heavy dimethyl labels, respectively. The labeled digests were then pooled and separated by SCX followed by nanoLC-LTQ-Orbitrap analysis. Protein identification was performed by Mascot. MS/MS spectra and identifications are available at the PRIDE database (http://www.ebi.ac.uk/pride; accession number 3533). In total, 142 proteins were identified in all three pooled cAMP pull-downs with at least two unique peptides and a Mascot score of at least 60. Markedly, among these were the known high-affinity cAMP binding proteins PKA-RIα and PKA-RIIα and several of their A-kinase anchoring proteins (AKAP) binding partners: AKAP1, AKAP2, AKAP4, AKAP11 and AKAP13. Using a more stringent threshold of four unique peptides left 70 proteins that could be well differentially quantified (see Table S2a and b of Supporting Information).

In Figure 5, several different peptide abundance profiles are highlighted to illustrate the extent of variation that can be quantified by our triplex dimethyl labeling. The relative abundance of IVDVIGEK and SLEMSER, peptides that belong to PKA-RIIα, are increased upon co-incubation with increasing concentration ADP/GDP. The inverse effect is observed for VMLGETNPADSKPGTIR and GDFCIQVGR, belonging to nucleoside diphosphate kinase B (Nme2), which is known to bind to ADP/ATP, with the heavy labeled peptide being barely present. Peptides LTGAVMHYGNMK and TPGAMEHELVLHQR, belonging to myosin 4, do not show much difference in abundance in the three samples, and this protein thus be considered as non-specifically binding to the beads. Notably, myosin is a high abundant protein in skeletal muscle.

In Figure 6, the change of relative abundance levels of a selection of proteins after co-incubation with different concentrations of ADP/GDP is presented, whereby all detected peptide abundance ratios were averaged per protein. Abundance levels of Nme2 clearly drop with increase in concentration of ADP/GDP, confirming Nme2 to bind with high affinity to ADP/GDP. The same holds true for actin and its secondary and tertiary binding proteins troponin and tropomyosin. With the non-specifically binding proteins removed from the beads by adding ADP/GDP, more cAMP is available for the less abundant high affinity proteins. This is evidently the case for the cAMP binding protein PKA-RIIα, where relative abundance levels are increased upon co-incubation with ADP/GDP, especially at 10 mM ADP/GDP. A smaller effect is seen for PKA-RIα that probably binds cAMP immobilized beads well also

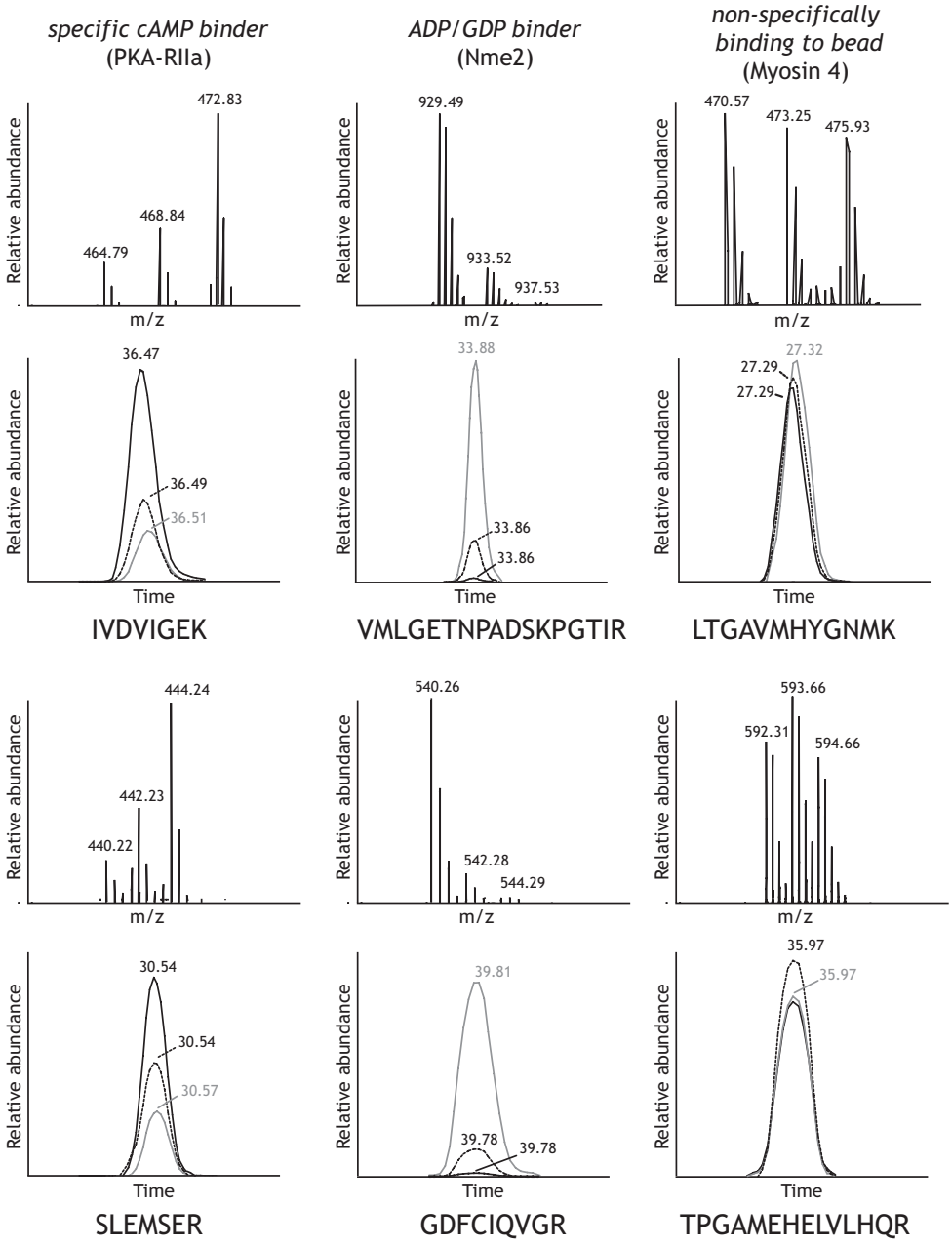


Figure 5. *MS spectra (upper panel) and extracted ion chromatograms (lower panel; light, solid grey line; intermediate, dashed black line; heavy, solid black line) of three peptides with illustrative different protein abundance patterns after cAMP pull-down with co-incubation with 0 (light), 0.5 (intermediate) and 10 mM ADP/GDP (heavy).*

when no ADP/GDP is present. AKAPs associate with different isoforms of PKA. AKAP1 [23], AKAP2 [24] have a dual specificity to both PKA-RI and PKA-RII. These AKAPs were found increased under conditions where PKA-RIIα and PKA-RIα showed higher abundance ratios, reconfirming their known interaction. Also, a group of proteins was found whose abundance levels neither raised nor dropped after co-incubation with ADP/GDP. These proteins

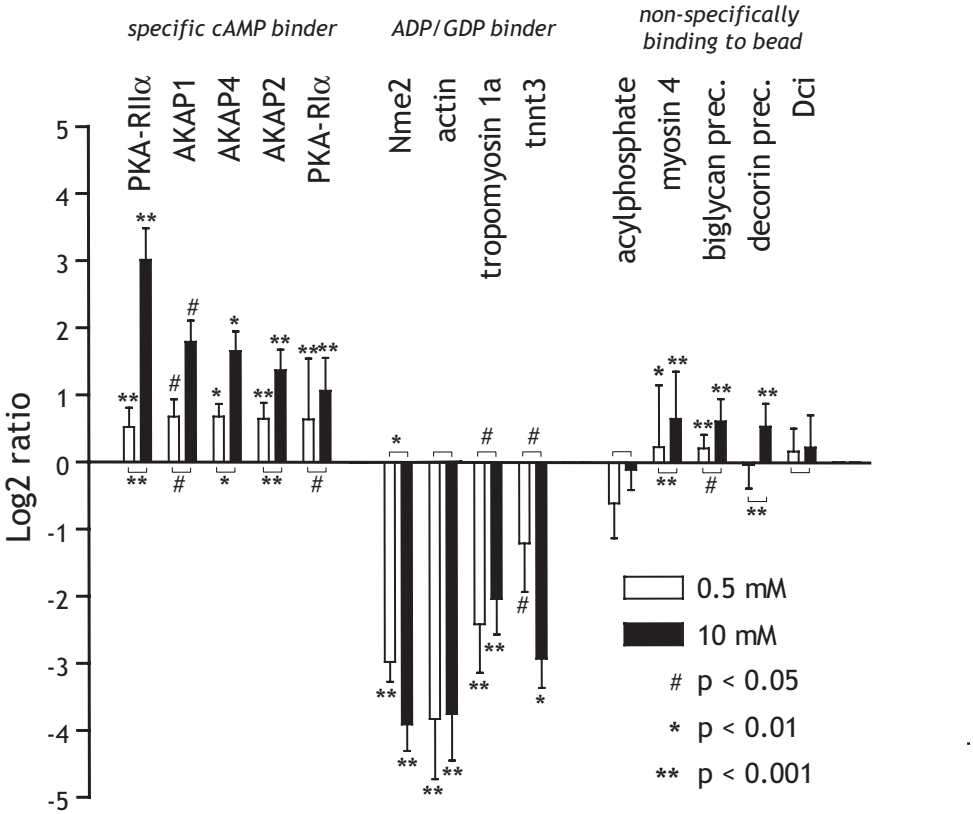


Figure 6. *Log2 abundance ratios with standard deviation for a selection of proteins. Protein abundance levels originating from the pull-down experiments with co-incubation with 0.5 and 10 mM ADP/GDP were divided by abundance levels from the 0 mM ADP/GDP experiment. Nme2: nucleoside disphospate kinase B; tnnt3: isoform 2 of troponin T; biglycan prec.: biglycan precursor; decorin prec.: decorin precursor; Dci: 3,2-trans-enoyl-CoA isomerase.*

probably bind non-specifically to the agarose beads itself rather than the functional group. Hence, their binding will not be affected by co-incubation with nucleotides.

In these experiments, 10 mM ADP/GDP seemed to be the preferred co-incubation condition to obtain a more selective pull-down. Especially, significantly more PKA and AKAP peptides were detected after co-incubation with 10 mM ADP/GDP. As expected, and noted in section 3.3, the abundance ratio of the same peptide eluting in two SCX fractions differed by less than 5% further confirming that SCX separation is not hampered by the D-isotope effect observed with ZIC-HILIC. In general, we show here that the triplex labeling can be used efficiently in affinity pull-down experiments and is generically applicable in any chemical proteomics or immunoprecipitation experiment, that all benefit from parallel analyses of more than two samples. So far, we have been able to successfully label protein amounts ranging from 1 μg up to 5 mg (unpublished data).

CONCLUDING REMARKS

Stable isotope dimethyl labeling of peptides is an affordable and easily implemented strategy for quantitative proteomics. Here, we extended the method introduced by Hsu *et al.* [8] to a triplex strategy by using new isotope labels based on the use of CD_2O and $NaBH_3CN$ and

$^{13}CD_2O$ and $NaBD_3CN$. We show that the labeling is applicable for the analysis of complex samples, *e.g.*, cell lysates and affinity purifications, and can be employed in a comprehensive proteomics experiment including those that require MudPIT methods. The simultaneous analysis of three samples appeared to have a minimal effect on protein identifications when adequate 2D-LC separation is performed. The triplex dimethyl labeling was also demonstrated in a chemical proteomics environment, in particular to study protein selectivity toward cAMP immobilized on agarose beads. Differences in abundance ratio ranging more than two orders of magnitude could be detected. The strategy will also be appropriate for dose-response and time course type studies such as that have recently been shown using methods such as SILAC [9]. As the labeling reaction is very quick using inexpensive reagents, application is not limited in terms of sample type or amount and any type of high resolution MS instrument can be used, we conclude that triplex dimethylation is a valuable strategy in shotgun proteomics as well as targeted chemical proteomics approaches.

ACKNOWLEDGEMENTS

We thank Dr. J.H. Ringrose and Dr. A. Scholten for critically reviewing this manuscript. We thank Nullin Divecha for the preparation of the chromatin enriched fractions of the MEL cells used here. This work was supported by The Netherlands Proteomics Centre (http://www.netherlandsproteomicscentre.nl/).

Supporting Information Available: A table listing the identified and quantified proteins from CEF and a table listing the identified and quantified proteins from the cAMP pull-down. This material is available free at http://www3.interscience.wiley.com/journal/121448057/suppinfo

REFERENCES

[1] Heck, A. J. R., Krijgsveld, J., Mass spectrometry-based quantitative proteomics. Expert Rev Proteomics 2004, 1, 317-326.

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

[3] Julka, S., Regnier, F., Quantification in proteomics through stable isotope coding: A review. J Proteome Res 2004, 3, 350-363.

[4] Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., et al., Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. Mol Cell Proteomics 2002, 1, 376-386.

[5] Julka, S., Regnier, F. E., Recent advancements in differential proteomics based on stable isotope coding. Brief Funct Genomic Proteomic 2005, 4, 158-177.

[6] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., et al., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 1999, 17, 994-999.

[7] Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., et al., Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. Mol Cell Proteomics 2004, 3, 1154-1169.

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

[9] Blagoev, B., Ong, S.-E., Kratchmarova, I., Mann, M., Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. Nat Biotechnol 2004, 22, 1139-1145.

[10] Pierce, A., Unwin, R. D., Evans, C. A., Griffiths, S., et al., Eight-channel iTRAQ enables comparison of the activity of 6 leukaemogenic tyrosine kinases. Mol Cell Proteomics 2008, 7, 853-863.

[11] Hsu, J. L., Huang, S. Y., Chen, S. H., Dimethyl multiplexed labeling combined with microcolumn separation and MS analysis for time course study in proteomics. Electrophoresis 2006, 27, 3652-3660.

[12] Jones, D. R., Bultsma, Y., Keune, W. J., Halstead, J. R., et al., Nuclear PtdIns5P as a transducer of stress signaling: An in vivo role for PIP4Kbeta. Mol Cell 2006, 23, 685-695.

[13] Boersema, P. J., Divecha, N., Heck, A. J. R., Mohammed, S., Evaluation and Optimization of ZIC-HILIC-RP as an Alternative MudPIT Strategy. J Proteome Res 2007, 6, 937-946.

[14] van Breukelen, B., van den Toorn, H. W. P., Drugan, M. M., Heck, A. J. R., StatQuant: A post quantification analysis toolbox for improving quantitative mass spectrometry. Bioinformatics 2009, 25, 1472-1473.

[15] Ji, C., Guo, N., Li, L., Differential Dimethyl Labeling of N-Termini of Peptides after Guanidination for Proteome Analysis. J Proteome Res 2005, 4, 2099-2108.

[16] Ji, C., Li, L., Gebre, M., Pasdar, M., Li, L., Identification and Quantification of Differentially Expressed Proteins in E-Cadherin Deficient SCC9 Cells and SCC9 Transfectants Expressing E-Cadherin by Dimethyl Isotope Labeling, LC-MALDI MS and MS/MS. J Proteome Res 2005, 4, 1419-1426.

[17] Zhang, R., Sioma, C. S., Thompson, R. A., Xiong, L., Regnier, F. E., Controlling Deuterium Isotope Effects in Comparative Proteomics. Anal Chem 2002, 74, 3662-3669.

[18] Guo, K., Ji, C., Li, L., Stable-Isotope Dimethylation Labeling Combined with LC-ESI MS for Quantification of Amine-Containing Metabolites in Biological Samples. Anal Chem 2007, 79, 8631-8638.

[19] Boersema, P. J., Mohammed, S., Heck, A. J. R., Hydrophilic interaction liquid chromatography (HILIC) in proteomics. Anal Bioanal Chem 2008.

[20] Mousson, F., Kolkman, A., Pijnappel, W. W. M. P., Timmers, H. T. M., Heck, A. J. R., Quantitative proteomics reveals regulation of dynamic components within TATA-binding protein (TBP) transcription complexes. Mol Cell Proteomics 2007, 7, 845-852.

[21] Scholten, A., Poh, M. K., vanVeen, T. A. B., vanBreukelen, B., et al., Analysis of the cGMP/cAMP Interactome Using a Chemical Proteomics Approach in Mammalian Heart Tissue Validates Sphingosine Kinase Type 1-interacting Protein as a Genuine and Highly Abundant AKAP. J Proteome Res 2006, 5, 1435-1447.

[22] Scholten, A., vanVeen, T. A. B., Vos, M. A., Heck, A. J. R., Diversity of cAMP-Dependent Protein Kinase Isoforms and Their Anchoring Proteins in Mouse Ventricular Tissue. J Proteome Res 2007, 6, 1705-1717.

[23] Huang, L. J. S., Durick, K., Weiner, J. A., Chun, J., Taylor, S. S., Identification of a novel protein kinase A anchoring protein that binds both type I and type II regulatory subunits. J Biol Chem 1997, 272, 8057-8064.

[24] Huang, L. J. S., Durick, K., Weiner, J. A., Chun, J., Taylor, S. S., D-AKAP2, a novel protein kinase A anchoring protein with a putative RGS domain. Proc Natl Acad Sci USA 1997, 94, 11184-11189.

Chapter 5: Straightforward and *de novo* peptide sequencing by MALDI-MS/MS using a Lys-N metalloendopeptidase

Paul J. Boersema[1,2,†], Nadia Taouatas[1,2,†], A. F. Maarten Altelaar[1,2], Joost W. Gouw[1,2], Philip L. Ross[3], Darryl J. Pappin[4], Albert J.R. Heck[1,2] and Shabaz Mohammed[1,2]

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

[2] Netherlands Proteomics Centre

[3] Applied Biosystems, 500 Old Connecticut Path, Framingham MA 01701, USA

[4] Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA

[†] PJB and NT contributed equally to this work and should be considered as joint first authors

Based on *Mol. Cell. Proteomics* 2009, 8, 650-660

ABSTRACT

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

INTRODUCTION

In proteomics, peptide sequencing is mainly performed by collision induced dissociation (CID) based tandem mass spectrometry [1, 2]. Generated peptide fragmentation spectra are matched against *in silico*-derived spectra from amino acid sequences in proteomic and genomic databases. Trypsin is the most frequently utilized protease as it generates peptides in the preferred mass range for effective fragmentation by CID [3]. Trypsin has high cleavage specificity and is stable under a wide variety of conditions generating peptides with a C-terminal arginine or lysine. This C-terminal positioning of the basic residue has consequences for fragment ion formation in CID. According to the 'mobile proton' model, dissociation upon excitation is initiated by a proton that weakens an amide bond in the peptide backbone [4-7]. The proton affinity/gas phase basicity of the two conjugate fragments will then dictate which fragment will inherit the amide-breaking proton, leading to the formation of b- or y-ions, respectively [8]. In MALDI-MS/MS of singly charged tryptic peptides, fragmentation results in complex spectra containing not only b- and y-ions, but also some a- and immonium ions, internal fragments and ions resulting from neutral loss of ammonia or water [9, 10]. Although all these fragment ions are used in typical database search strategies they often complicate and hamper *de novo* sequencing, *e.g.* sequencing of peptides from species of which no genome sequence is available [11, 12]. Therefore, several attempts have been made to simplify MALDI-CID spectra [11, 13-19]. For example, the peptide N terminus can be derivatized with sulfonic acid as in chemically assisted fragmentation (CAF) [16, 17] or with 4-sulfophenyl isothiocyanate (SPITC)[20, 21] to establish a fixed negative charge. After this reaction, primarily protonated C-terminal fragment ions are detected, giving rise to y-ion ladder series. The loss of sensitivity caused by adding a negative charge while performing positive ion mode analysis can, although only in part, be compensated for by increasing the basicity of lysine residues [22]. Another approach to simplify MALDI spectra is to add a fixed positively charged tag to the peptide N terminus [13] while modifying internal arginine residues [14] or removing the C-terminal lysine or arginine [11]. These modified peptides fragment to generate spectra with mainly a- and b-ions. One can also modify the basicity of the peptide to promote formation of a single series of ions. Lysine can be made more basic by guanidation [18] or treatment with 2-methoxy-4,5-dihydro-1H-imidazole [19]. In this way, the C-terminal fragment of a tryptic peptide is more likely to be protonated after fragmentation yielding spectra with more intense y-ions. Such techniques, however, require additional sample handling. Moreover, chemical derivatization of minute amounts of sample is more difficult and is often hampered by the formation of unwanted side products [23]. Recently, we explored a new method for mass spectrometry based sequencing of peptides using a little explored metalloendopeptidase with Lys-N cleavage specificity [24]. We showed that the combination of this protease with ESI-MS using electron transfer induced dissociation (ETD) for peptide fragmentation produced spectra that were completely dominated by c-type fragment ions, providing simple sequence ladders of the peptides of interest [24]. In ETD with supplemental collisional activation (commonly referred to as ETcaD), doubly charged peptide ions generated by ESI are charge reduced during the electron transfer process, resulting in the remainder of a single free proton [25]. As the N terminus of Lys-N peptides accommodates two basic entities, primarily N-terminal fragments are protonated, which therefore leads to the detection of mainly c-ions. However, ETD fragmentation requires multiply charged ions (i.e. ESI) and specific instrumentation that is not readily available. Therefore, in this paper, we comprehensively explored the use of the Lys-N metalloendopeptidase and CID fragmentation using MALDI-MS/MS. A HEK293 cellular lysate was digested by Lys-N and for a direct comparison also with Lys-C, which produces tryptic-like

peptides with the basic lysine at the C terminus. The resulting peptides were separated and enriched for peptides containing a single lysine residue by strong cation exchange (SCX) and analyzed by MALDI-TOF/TOF. The combination of Lys-N and MALDI-MS/MS resulted in spectra with clear and straightforward sequence ladders, consisting of almost exclusively b-ions. We also performed a direct comparison of the following four combinations: (i) Lys-C, SCX enrichment, and MALDI-MS/MS analysis; (ii) Lys-N, SCX enrichment, and MALDI-MS/MS analysis; (iii) Lys-N, SCX enrichment, and ESI-CID MS/MS analysis; and (iv) Lys-N, SCX enrichment, and ESI-ETD MS/MS analysis. The comparison clearly demonstrated that only the combinations of Lys-N generated peptides with MALDI-CID and ESI-ETD MS/MS provided very clear sequence ladders. Furthermore, the potential of this method for facilitating *de novo* sequencing was illustrated by the successful identification of proteins from an SDS-PAGE band of an ostrich tissue lysate where ostrich represents a species with an unsequenced genome.

MATERIALS AND METHODS

Materials

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

Sample preparation

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

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

In-solution digestion

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

In gel digest

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

Strong cation exchange (SCX)

SCX was performed using an Agilent 1100 series LC-system with a C18 Opti-Lynx (Optimize Technologies, Oregon OR) guard column and Polysulfoethyl A SCX column (PolyLC, Columbia, MD; 200 mm x 2.1 mm id). Sample was dissolved in 0.05% formic acid and loaded onto the guard column at 100 µL/min and consecutively eluted onto the SCX column with 80% ACN, 0.05% formic acid. SCX buffer A was 5 mM KH_2PO_4, 30% ACN, pH 2.7; SCX buffer B was 350 mM KCl, 5 mM KH_2PO_4, 30% ACN, pH 2.7. Gradient elution was performed

as follows: 0-85% B in 45 min, 85%-100% B in 6 min, and 100% B for 4 min. A total of 53 one-minute fractions were collected, and dried in a vacuum centrifuge.

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

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

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

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

charges (ESI), and MS/MS tolerance was 0.2 Da (MALDI) or 0.9 Da (ESI). Peptides were identified with a minimum Mascot score of 30, and at these settings the false discovery rate was less than 0.75% as estimated by using the Mascot decoy database function. For further data analysis, Mascot data were imported into Scaffold 1.7.

For the *de novo* sequencing of ostrich proteins the most abundant peptides in the MALDI MS spectra were fragmented. Manual annotation of CID spectra was performed using the mass differences between adjacent fragment ions. The obtained sequences were BLAST searched against a human and chicken UniProt database. Several assigned peptides were identical to peptides from chicken and/or human creatine kinase and actin. Other peptides were very similar, but revealed ostrich specific single or double amino acid differences.

RESULTS

In a typical proteomics experiment, digestion is performed by trypsin generating peptides with a C-terminal arginine or lysine. A mixture of N- and C-terminal fragment ions can be detected after CID of these peptides, but as the arginine and lysine residues are more basic than the α-amino group, generally y-ions are more abundant [10, 19]. This study evaluated the unique situation that arises when proteolysis is performed by Lys-N as peptides are yielded with an N terminus accommodating both α- and ε-amino basic entities. Fragmentation of Lys-N peptides is thus likely skewed towards the production of N-terminal ions (see Figure 1) as has been hinted at before when looking at a few individual peptides [27, 28]. To more comprehensively assess the fragmentation behavior of Lys-N-produced peptides, a whole cellular lysate was digested in parallel with Lys-N and Lys-C. Lys-C, like trypsin, generates peptides with a C-terminal basic amino acid residue. However, Lys-C has no specificity for arginine, and so comparing Lys-N with Lys-C is more appropriate than comparing it with trypsin. The digested cellular lysates were first subjected to low-pH SCX chromatography to enrich for and isolate peptides with a single basic lysine residue [24, 29, 30]. A few consecutive fractions of the SCX run will provide a set of peptides with a single C-terminal lysine residue for the Lys-C digested sample, and with a single N-terminal lysine from the Lys-N digest [24, 30-32]. Off-line nano-RP-LC separation was performed on these selected SCX fractions. The eluent was subsequently mixed postcolumn with α-cyano-4-hydroxycinnamic acid and automatically fractionated and spotted onto a MALDI target plate, an experimental set-up adopted form our previously described off-line ZIC-HILIC set-up [33]. The fractionated peptides were then subjected to analysis by MALDI-TOF/TOF

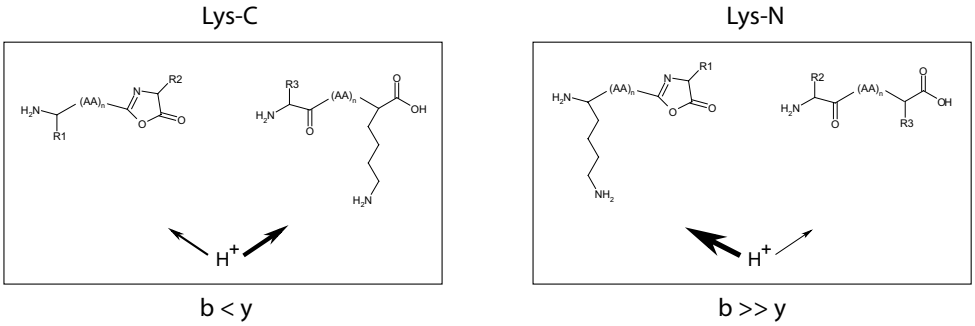


Figure 1. *Schematic representation of CID fragmentation of Lys-C and Lys-N derived singly charged peptide ions. Lys-C peptides have a basic N and C terminus; therefore, both termini will be protonated, leading to a mixture of b- and y-ions in CID. Lys-N peptides concentrate the basicity at the N terminus,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

based tandem mass spectrometry. As an illustrative result, fragmentation of the peptide KCQEVISWLDANTLAE, as depicted in Figure 2a, resulted in a CID spectrum that was typical for Lys-N proteolytic peptides (Figure 2b) and displayed a complete sequence ladder consisting of b-ions. As shown in Figure 2b, Lys-N peptide sequences can be easily read off as there are no significant 'interfering' ion-series present. As indicated in Figure 2, the b_1-ion is quite abundant, and additionally, the b_1 related ions a_1 and a_1-NH_3 were found to be often the base peak in the spectra. C-terminal fragment ions (y-ions) were detected at a very low frequency and/or intensity.

We further examined and compared the effect of the N- or C-terminal position of the lysine on the fragmentation of the peptides in CID. Typical peptide fragmentation spectra are shown in Figure 3 which incorporates MALDI-CID spectra of the same peptide with a lysine

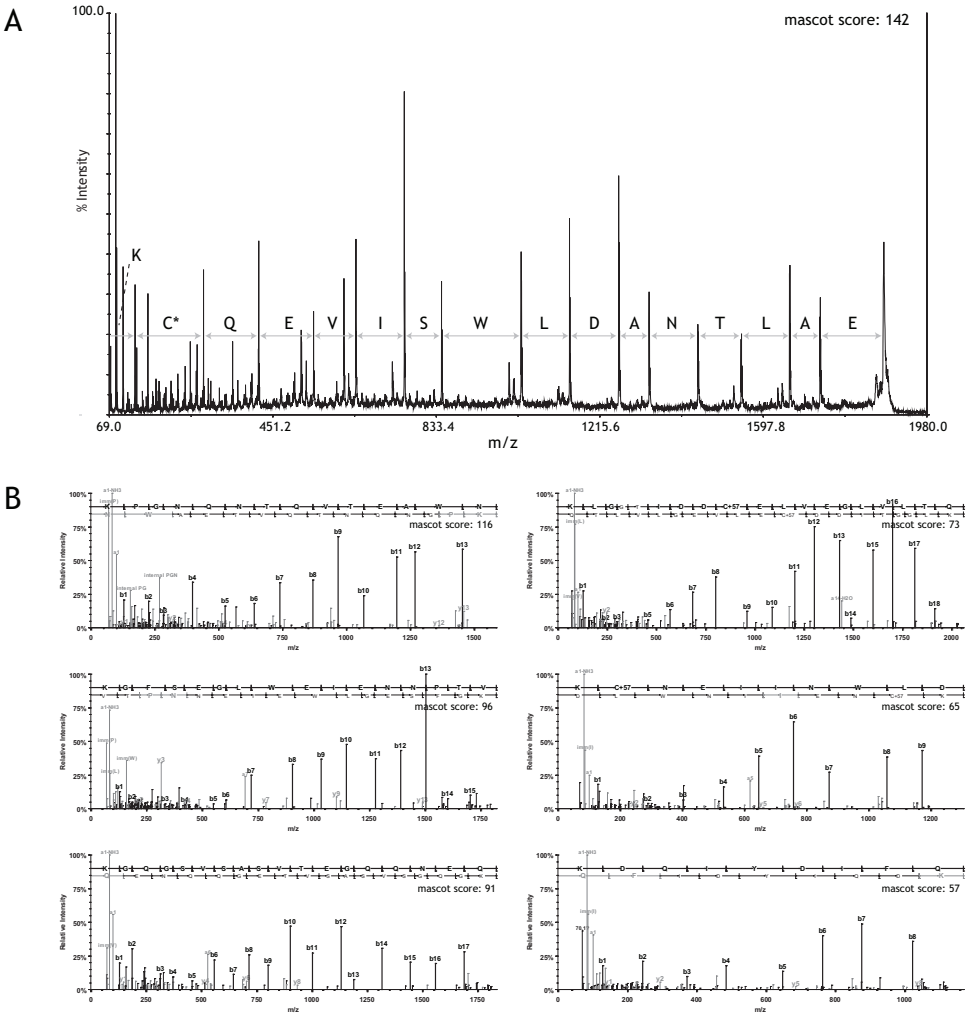


Figure 2. *Representative MALDI-CID spectra of peptides identified from a Lys-N digested HEK293 cellular lysate. (A) A clean b-series sequence ladder is detected for peptide KC*QEVISWLDANTLAE (m/z 1876.73, 1+; C*, carbamidomethylated cysteine) (B) Six typical CID spectra of Lys-N peptides annotated by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

either on the C-terminus or the N-terminus (respectively generated by Lys-C and Lys-N digestion) as well as spectra of the doubly charged ion of the same Lys-N peptide analyzed by ESI-CID MS/MS and ESI-ETcaD MS/MS. The MALDI tandem mass spectrum obtained for the peptide with a C-terminal lysine contains a mixture of b- and y-ions (and no complete series) and is clearly more complex than the spectrum from the analogous Lys-N peptide, which provides a nearly complete b-ion series. We found that fewer immonium ions were detected, and the number of non-informative background peaks seemed to be lower in the spectra of the Lys-N peptide, possibly related to the reduced number of fragment ion pathways available. The tandem mass spectrum of the doubly charged ion of the Lys-N peptide obtained by ESI-CID also shows a complex spectrum with both b- and y-ions which can be

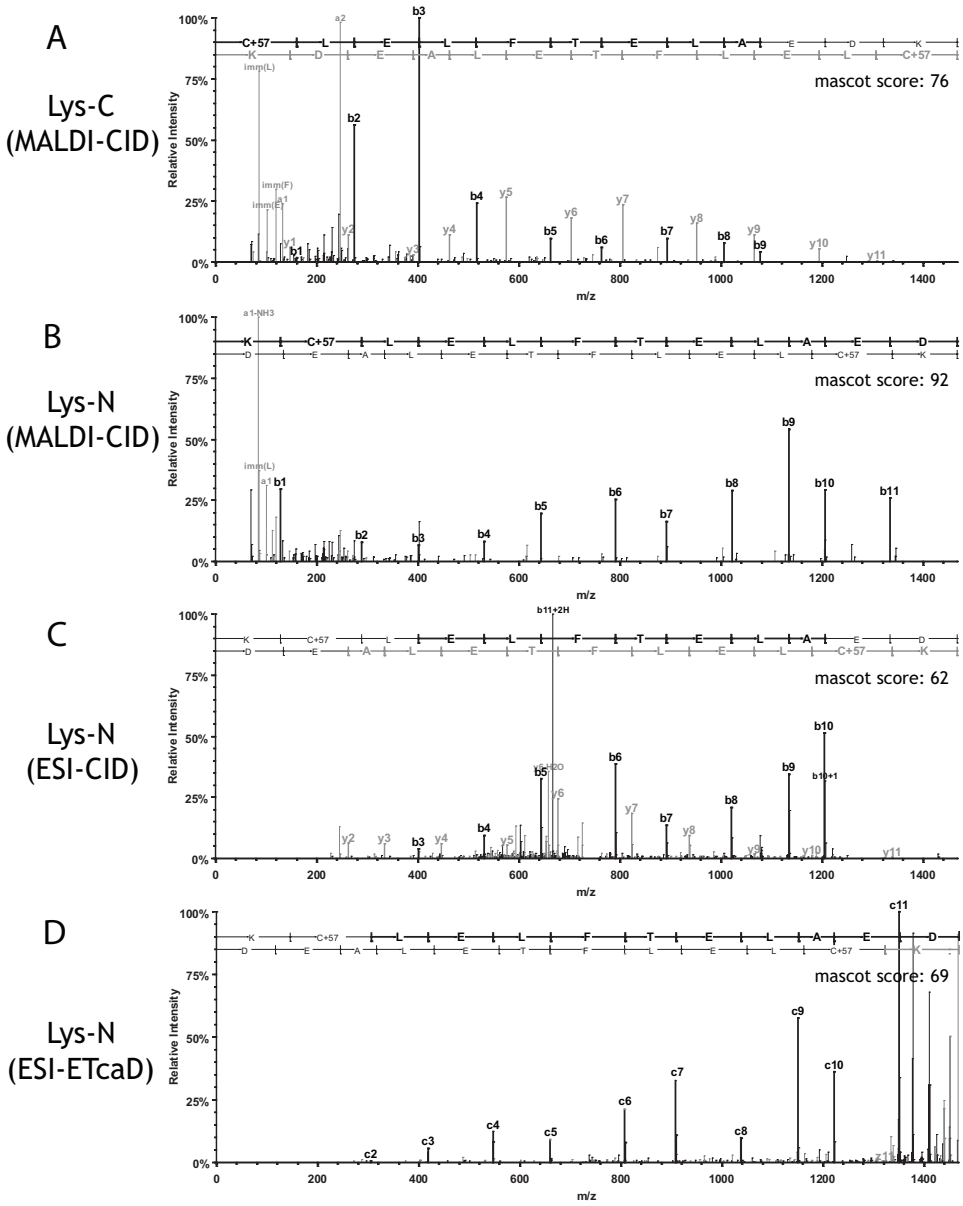


Figure 3. *Representative mass spectra annotated by Scaffold of the same peptide with lysine respectively on (a) the C or (b, c and d) the N terminus of the peptide. In (a) (m/z 1467.60, 1+) and (b) (m/z 1467.58, 1+) are shown the MALDI-CID spectra, in (c) (m/z 734.20, 2+) the ESI-CID spectrum obtained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

attributed to the availability of two protons for the fragment ions [24]. Finally the ETcaD spectrum of the doubly charged peptide ion is comprised of only c-ions, in agreement with our recent findings and in appearance similar to the MALDI-CID spectrum [24].

As stated, MALDI-CID of peptides with a single N-terminal lysine yielded mass spectra dominated by b-ions and with only minor y-ions. However, when a proline is present in the sequence this straightforward fragmentation pattern becomes disrupted. Peptide cleavage on the N-terminal side of a proline leads to preferential formation of a y-ion (Figure 4a). An illustrative example of a CID spectrum of a peptide containing two prolines is shown in Figure 4b. The two most intense peaks correspond to y-ions with an N-terminal proline residue. Also, two intense internal fragments are detected with an N-terminal proline residue. Nevertheless an almost complete and clear b-ion series is still detected.
MALDI-CID analysis of 3 SCX fractions, enriched for peptides containing a single, N-terminal lysine, led to the identification of a total of 247 peptides (Mascot score ≥30, false dis-

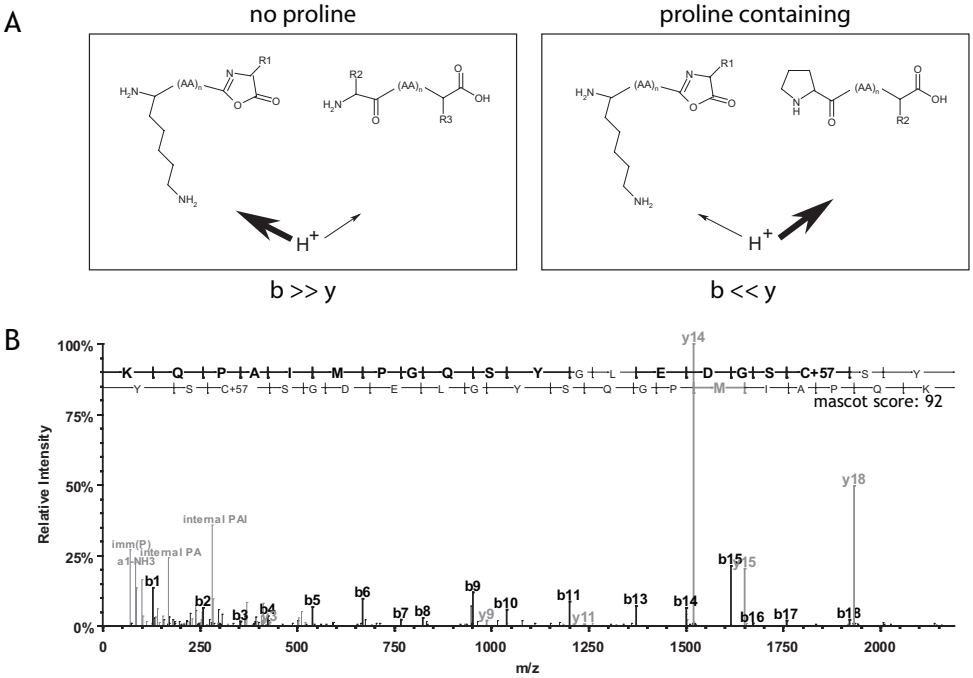


Figure 4. *CID fragmentation of proline-containing Lys-N peptides. (a) Schematic representation of CID fragmentation of Lys-N peptides that do not or do contain proline. Fragmentation of a proline-free peptide results in the detection of mainly b-ions, while cleavage N-terminal of a proline residue of a Lys-N peptide results in the detection of dominant y-ions. (b) Representative MALDI-CID spectrum of a proline-containing Lys-N peptide (KQPAIMPGQSYGLEDGSC*SY, m/z 2187.78, 1+) annotated by Scaffold. The two intense y-ions correspond to ions with an N-terminal proline. Also two intense internal fragments with an N-terminal proline are detected. Note that still a nearly full b ion series can be detected. AA, amino acid.*

covery rate, 0.75%) with an N-terminal lysine residue (tandem mass spectra are available as a PRIDE (Proteomics Identifications) database under accession number 3380). Of these, 36 contained an extra basic residue and were initially removed from further analysis. Of the remaining 211 peptides, 119 contained one or more proline residues. To account for the proline effect on ion formation, proline containing peptides were analyzed separately from proline free peptides. N-terminal b-ions represented in number 84% of the detected backbone frag-

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

To evaluate the applicability of Lys-N digestion for genuine *de novo* sequencing, muscle tissue from an ostrich, of which the genome has not (yet) been sequenced, was lysed and sepa-

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

Table 1. *Analysis of the frequency and normalized overall intensity of b-ions compared to y-ions in MALDI-CID spectra of Lys-N and Lys-C-derived peptides. The percent intensity is the intensity of all b-ions divided by the sum of intensity of all b- and y-ions. *After subtracting the intensity of y-ions with an N-terminal proline.*

rated by 1D SDS-PAGE. It should be noted that in our earlier work no significant difference in yield was found between in-gel digestion with trypsin or Lys-N [24] (Supplemental Figure 1[1] indicates the sequence coverage and peptide signal achieved by LC-MS for an in-gel digestion of equivalent amounts of BSA by Lys-N and trypsin). A band at ~42 kDa was excised from the ostrich gel, and its content was digested with Lys-N. These peptides were then separated by 'offline' nano-LC with the eluent being mixed post-column with α-cyano-4-hydroxycinnamic acid and automatically spotted onto a MALDI target plate. MALDI-TOF/TOF analysis was subsequently performed. CID spectra of ~ 20 of the most intense peptides were manually interpreted. Identified peptide sequences were homology searched against UniProt human and chicken databases using BLAST. Four unique peptides were found to align with creatine kinase (Figure 5a), whereas five unique peptides were found to align with actin (Figure 5b). Strikingly one of the actin peptides was found in four different forms. Alongside an unmodified form we observed sequences containing; methionine oxidation, methionine oxidation plus tryptophan hydroxylation and a decomposed carboxymethylated methionine [34]. Such modifications would be missed by a database search strategy if one does not, in advance, input these possibilities in the submission criteria. Three peptide sequences of creatine kinase and four sequences of actin were identical to human and chicken protein sequences allowing the identification of the protein. For both proteins we found peptides to have a sequence that differed slightly from the human and chicken sequences. In Figure 5c the CID spectrum of such a peptide of creatine kinase is depicted. The clear b-ion

1 Supplementary data to this chapter is available at http://www.mcponline.org/cgi/content/full/M800249-MCP200/DC1

A

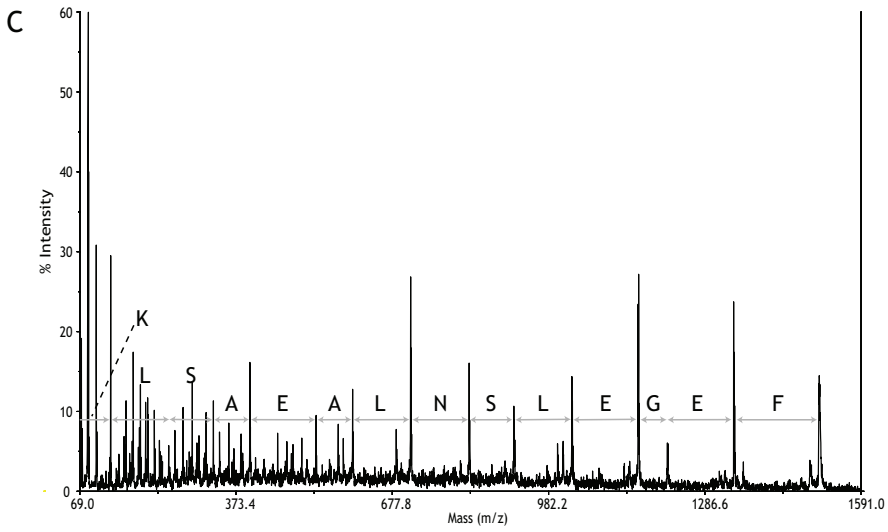
```
Creatine Kinase

Human     ~ LNYKPEEEYPGLSK ~ DKHKTDLNHENLK ~ EKLSVEALNSLTGEFK ~ LKSMTEKEQQQLIDDHFLFDK ~
Chicken   ~ LKFSAEEEFPDLSK ~ DKHRTDLNHENLK ~ EKLSVEALNSLEGEFK ~ LKAMTEQEQQQLIDDHFLFDK ~
Ostrich   ~ -KFSAEEEFPDLS- ~ -KHRTDLNHENL- ~ -KLSAEALNSLEGEF- ~ -KAMTEAEQQQLIDDHFLFD- ~
              ::..***:****     **:*******      ***.****** ***     *:*** *************
```

B

```
Actin

Human    ~ LKYPIEHGIITNWDDMEK ~ MKILTERGYS ~ MKIKIIAPPERKYSVWIGGSILAS ~ SKQEYDEAGPSIVHRK ~
Chicken  ~ LKYPIEHGIITNWDDMEK ~ MKILTERGYS ~ MKIKIIAPPERKYSVWIGGSILAS ~ TKQEYDEAGPSIVHRK ~
Ostrich  ~ -KYPIEHAIITNWDDME- ~ -KILTERGY- ~ -KIKIIAPPERKYSVWIGGSILA- ~ -KQEYDEAGPSIVHR- ~
           ******.*********     ********     **********************     **************
```



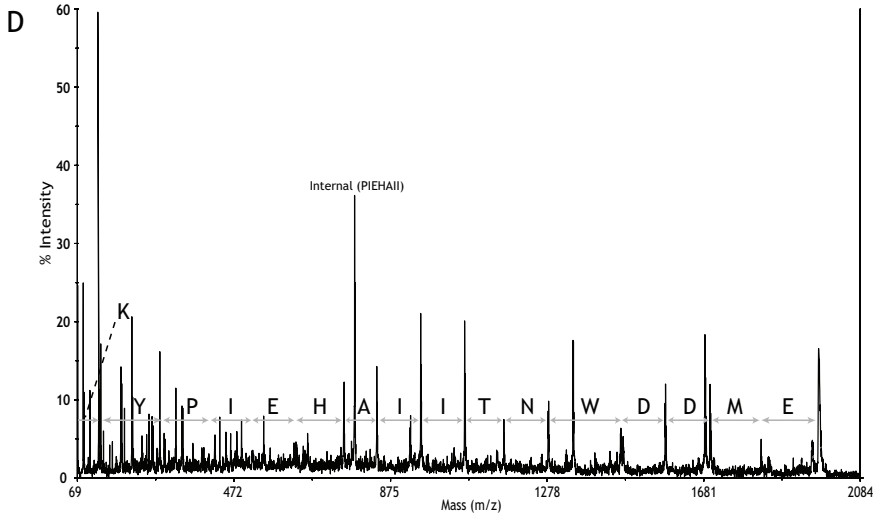


Figure 5 *Lys-N facilitates de novo sequencing of ostrich creatine kinase and actin. (a) Multiple sequence alignment of de novo sequenced ostrich peptides with partial creatine kinase sequences of human (P06732, UniProtKB/Swiss-Prot) and chicken (P00565). (b) Multiple sequence alignment of de novo sequenced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

series facilitated its *de novo* sequencing. The sequence deviates at the fourth (alanine) residue compared to chicken and human, whereas the 11th (glutamine) residue is similar to chicken, but different from human. In Figure 5d a CID spectrum is depicted of actin peptide KYP-IEHAIITNWDDME. This peptide aligns with human and chicken actin except for residue 7 (alanine). Additional annotated spectra can be found in supplemental Figure 2.

DISCUSSION

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

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

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

Scrutinizing spectra for additional trends, we observed that the MALDI-CID spectra of proline-containing peptides contained intense peaks corresponding to fragment ions with an N-terminal proline, a well described phenomenon [4, 37-39]. The intensity of these peaks has been explained by the extraordinary structure of proline with its side chain forming a five-membered ring with the peptide backbone. This hinders cleavage C-terminal of proline as it would involve the generation of a strained bicyclic structure [39]. Cleavage on the N-terminal side causes the formation of the secondary amine group of proline which can sequester the mobile proton, causing the generation of the C-terminal fragment ion (see Figure 4b)[38]. Our data suggest that a combination of both effects prompt the emergence of intense

peaks of y-ions with an N-terminal proline. First, the fact that the intensity of a b-ion with a C-terminal proline is significantly lower than other b-ions indicates that the formation of a b-ion with a bicyclic C-terminal proline structure is unfavored [39]. Second, the high gas phase proton affinity of the proline amine group is reflected in the intensity of the y-ion with an N-terminal proline being substantially larger than the intensity of the b-ion that is generated N-terminal of the same proline. It appears thus that the two amino groups at the N terminus of the potential b-ion exert less influence than the secondary amine of the proline present on the y-ion at the point of cleavage [38]. On average, b-ions in MALDI CID spectra of proline containing peptides account for 71% of the total intensity of b- and y-ions. When the intense y-ions corresponding to fragments with an N-terminal proline were removed, this percentage of b-ions increased to 81%. This is slightly lower than for proline free peptides. Despite the occurrence of y-ions, the remainder of such Lys-N MALDI-CID spectra is still dominated by sequence ladders of b-ions. About 56% of the peptides we identified with a single N-terminal lysine contained a proline residue. This is similar to a recent study in which the occurrence of at least one proline in a tryptic peptide was determined, both theoretically (using the IPI-Human database) and practically, to be around 50% [40].

Efficient and facile *de novo* sequencing requires good quality straightforward mass spectra. Lys-N proteolysis allows significant portions of a protein to generate simplified MALDI-CID spectra without any derivatization steps. An *in silico* digest of all proteins in the IPI Human database revealed that approximately one-third of all theoretical Lys-N peptides contains a single basic residue. Therefore, as an example of the applicability of Lys-N proteolysis for *de novo* sequencing we strived to identify the protein contents of a 1D SDS-PAGE gel band of ostrich muscle tissue. Clear, simplified spectra were generated facilitating *de novo* sequencing. Sequenced peptides could be aligned to chicken (the phylogenetically closest species to ostrich with a genome that is sequenced) creatine kinase and actin. Of these peptides, at least two would not have been identified if a traditional database search strategy had been performed since these peptides slightly differ in sequence compared to sequences available in genomic databases. For example, in the ostrich creatine kinase peptide KLSAEALNSLEGEF, the third residue, an alanine, is a valine in chicken and human; this is probably a DNA point mutation as the translation codons of valine and alanine only differ by one nucleotide. The same is true for the actin peptide KYPIEHAIITNWDDME where a single DNA point mutation could explain the conversion of glycine in chicken and human to alanine in ostrich. Furthermore, modifications to this peptide were found in other mass spectra that could easily be detected, including methionine oxidation, tryptophan hydroxylation and decomposed carboxymethylated methionine. Although methionine oxidation is generally included as a variable modification in database searching, for tryptophan hydroxylation and decomposed carboxymethylated methionine this is generally not the case, thus underscoring the potential of our *de novo* sequencing approach. Similarly we expect that other modifications can be easily identified and located using the MS/MS sequence ladders by combining Lys-N digestion with MALDI-MS/MS analysis. Labile modifications such as phosphorylation might result in neutral loss-dominated spectra as fragmentation is performed by CID. However, these spectra will still be simpler than those achieved by tryptic peptides.

As discussed before, an apparent drawback of Lys-N MALDI-CID is that the clear b-ion dominated spectra will exclusively be observed for peptides containing a single N-terminal basic residue. Depending on the position in the peptide sequence, an extra basic residue might lead to the generation of an additional (y-) ion series thereby complicating manual

interpretation. Chemical derivatization that aids *de novo* sequencing such as CAF and SPITC can be applied, but it is necessary to have a basic residue at the C-terminus thus tryptic peptides are necessary. In theory, the number of peptides with a single basic residue in a tryptic digest is ~3-fold higher than in a Lys-N digest. However, internal basic residues negatively affect these derivatization strategies in a similar way to Lys-N, *i.e.* generating additional fragment ion series [11, 16, 18, 19]. Furthermore the sulfonyl group which is key to CAF/SPITC, reduces peptide signal intensities due to the intrinsic negative charge. Also these chemical derivatization steps potentially lead to sample loss and further signal reduction. Finally, side reactions (often) hamper the analysis by increasing the complexity of the sample with uninformative peptides. Through the use of Lys-N *de novo* sequencable peptides can be attained without chemical derivatization.

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

ACKNOWLEDGEMENTS
We thank Dr. M.M. Drugan for performing statistical analyses. This work was supported by the Netherlands Proteomics Centre (http://www.netherlandsproteomicscentre.nl), a program embedded in the Netherlands Genomics Initiative.

Supplementary data is available free at http://www.mcponline.org/cgi/content/full/M800249-MCP200/DC1.

REFERENCES

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

[2] Chalmers, M. J., Gaskell, S. J., Advances in mass spectrometry for proteome analysis. *Curr Opin Biotechnol* 2000, *11*, 384-390.

[3] Olsen, J. V., Ong, S.-E., Mann, M., Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues. *Mol Cell Proteomics* 2004, *3*, 608-614.

[4] Paizs, B., Suhai, S., Fragmentation pathways of protonated peptides. *Mass Spectrom Rev* 2005, *24*, 508-548.

[5] Burlet, O., Yang, C. Y., Gaskell, S. J., Influence of Cysteine to Cysteic Acid Oxidation on the Collision-Activated Decomposition of Protonated Peptides - Evidence for Intraionic Interactions. *J Am Soc Mass Spectrom* 1992, *3*, 337-344.

[6] Harrison, A. G., Yalcin, T., Proton mobility in protonated amino acids and peptides. *Int J Mass Spectrom* 1997, *165*, 339-347.

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

[8] Roepstorff, P., Fohlman, J., Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* 1984, *11*, 601.

[9] Wattenberg, A., Organ, A. J., Schneider, K., Tyldesley, R., *et al.*, Sequence dependent fragmentation of peptides generated by MALDI quadrupole time-of-flight (MALDI Q-TOF) mass spectrometry and its implications for protein identification. *J Am Soc Mass Spectrom* 2002, *13*, 772-783.

[10] Khatun, J., Ramkissoon, K., Giddings, M. C., Fragmentation Characteristics of Collision-Induced Dissociation in MALDI TOF/TOF Mass Spectrometry. *Anal Chem* 2007, *79*, 3032-3040.

[11] Chen, W., Lee, P. J., Shion, H., Ellor, N., Gebler, J. C., Improving de Novo Sequencing of Peptides Using a Charged Tag and C-Terminal Digestion. *Anal Chem* 2007, *79*, 1583-1590.

[12] Palagi, P. M., Hernandez, P., Walther, D., Appel, R. D., Proteome informatics I: Bioinformatics tools for processing experimental data. *Proteomics* 2006, *6*, 5435-5444.

[13] Zaia, J., Charged Derivatives for Peptide Sequencing Using a Magnetic Sector Instrument. *Methods Mol Biol* 1996, *61*, 29-41.

[14] Spengler, B., Luetzenkirchen, F., Metzger, S., Chaurand, P., *et al.*, Peptide sequencing of charged derivatives by postsource decay MALDI mass spectrometry. *Int J Mass Spectrom* 1997, *169*, 127-140.

[15] Roth, K. D. W., Huang, Z.-H., Sadagopan, N., Watson, J. T., Charge derivatization of peptides for analysis by mass spectrometry. *Mass Spectrom Rev* 1998, *17*, 255-274.

[16] Keough, T., Youngquist, R. S., Lacey, M. P., A method for high-sensitivity peptide sequencing using postsource decay matrix-assisted laser desorption ionization mass spectrometry. *Proc Natl Acad Sci U S A* 1999, *96*, 7131-7136.

[17] Flensburg, J., Tangen, A., Prieto, M., Hellman, U., Wadensten, H., Chemically-assisted fragmentation combined with multi-dimensional liquid chromatography and matrix-assisted laser desorption/ionization post source decay, matrix-assisted laser desorption/ionization tandem time-of-flight or matrix-assisted laser desorption/ionization tandem mass spectrometry for improved sequencing of tryptic peptides. *Eur J Mass Spectrom* 2005, *11*, 169-179.

[18] Brancia, F. L., Oliver, S. G., Gaskell, S. J., Improved matrix-assisted laser desorption/ionization mass spectrometric analysis of tryptic hydrolysates of proteins following guanidination of lysine-containing peptides. *Rapid Commun Mass Spectrom* 2000, *14*, 2070-2073.

[19] Peters, E. C., Horn, D. M., Tully, D. C., Brock, A., A novel multifunctional labeling reagent for enhanced protein characterization with mass spectrometry. *Rapid Commun Mass Spectrom.* 2001, *15*, 2387-2392.

[20] Gevaert, K., Demol, H., Martens, L., Hoorelbeke, B., *et al.*, Protein identification based on matrix assisted laser desorption/ionization-post source decay-mass spectrometry. *Electrophoresis* 2001, *22*, 1645-1651.

[21] Marekov, L. N., Steinert, P. M., Charge derivatization by 4-sulfophenyl isothiocyanate enhances peptide sequencing by post-source decay matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J Mass Spectrom* 2003, *38*, 373-377.

[22] Keough, T., Lacey, M. P., Youngquist, R. S., Derivatization procedures to facilitate de novo sequencing of lysine-terminated tryptic peptides using postsource decay matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom* 2000, *14*, 2348-2356.

[23] Regnier, F. E., Samir, J., Primary amine coding as a path to comparative proteomics. *Proteomics* 2006, *6*, 3968-3979.

[24] Taouatas, N., Drugan, M. M., Heck, A. J. R., Mohammed, S., Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase. *Nat Methods* 2008, *5*, 405-407.
[25] Swaney, D. L., McAlister, G. C., Wirtala, M., Schwartz, J. C., *et al.*, Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal Chem* 2007, *79*, 477-485.
[26] Meiring, H. D., van der Heeft, E., ten Hove, G. J., de Jong, A., Nanoscale LC-MS(n): technical design and applications to peptide and protein analysis. *J Sep Sci* 2002, *25*, 557-568.
[27] Takio, K., *Handbook of Proteolytic Enzymes*, Elsevier Scientific Publishing Company, Amsterdam, the Netherlands 2004, pp. 788-791.
[28] Rao, K. C. S., Palamalai, V., Dunlevy, J. R., Miyagi, M., Peptidyl-Lys Metalloendopeptidase-catalyzed 18O Labeling for Comparative Proteomics: Application to Cytokine/Lipolysaccharide-treated Human Retinal Pigment Epithelium Cell Line. *Mol Cell Proteomics* 2005, *4*, 1550-1557.
[29] Dormeyer, W., Mohammed, S., Breukelen, B., Krijgsveld, J., Heck, A. J., Targeted analysis of protein termini. *J Proteome Res* 2007, *6*, 4634-4645.
[30] Taouatas, N., Altelaar, A. F. M., Drugan, M. M., Helbig, A. O., *et al.*, SCX-based fractionation of Lys-N generated peptides facilitates the targeted analysis of post-translational modifications. *Mol Cell Proteomics* 2008, *8*, 190-200.
[31] Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., *et al.*, Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* 2004, *101*, 12130-12135.
[32] Pinkse, M. W. H., Mohammed, S., Gouw, L. W., van Breukelen, B., *et al.*, Highly robust, automated, and sensitive on line TiO2-based phosphoproteomics applied to study endogenous phosphorylation in Drosophila melanogaster. *J Proteome Res* 2008, *7*, 687-697.
[33] Boersema, P. J., Divecha, N., Heck, A. J. R., Mohammed, S., Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. *J Proteome Res* 2007, *6*, 937-946.
[34] Jones, M. D., Merewether, L. A., Clogston, C. L., Lu, H. S., Peptide Map Analysis of Recombinant Human Granulocyte Colony Stimulating Factor: Elimination of Methionine Modification and Nonspecific Cleavages. *Anal Biochem* 1994, *216*, 135-146.
[35] Fu, Q., Li, L. J., De novo sequencing of neuropeptides using reductive isotopic methylation and investigation of ESI QTOF MS/MS fragmentation pattern of neuropeptides with N-terminal dimethylation. *Anal Chem* 2005, *77*, 7783-7795.
[36] Yalcin, T., Harrison, A. G., Ion chemistry of protonated lysine derivatives. *J Mass Spectrom* 1996, *31*, 1237-1243.
[37] Breci, L. A., Tabb, D. L., Yates, J. R., Wysocki, V. H., Cleavage N-Terminal to Proline: Analysis of a Database of Peptide Tandem Mass Spectra. *Anal Chem* 2003, *75*, 1963-1971.
[38] Hunt, D. F., Yates, J. R., Shabanowitz, J., Winston, S., Hauer, C. R., Protein Sequencing by Tandem Mass Spectrometry. *Proc Natl Acad Sci U S A* 1986, *83*, 6233-6237.
[39] Vaisar, T., Urban, J., Probing the proline effect in CID of protonated peptides. *J Mass Spectrom* 1996, *31*, 1185-1187.
[40] Van Hoof, D., Pinkse, M. W. H., Ward-Van Oostwaard, D., Mummery, C. L., *et al.*, An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. *Nat Methods* 2007, *4*, 677-678.

Chapter 6: Phosphopeptide fragmentation and analysis by mass spectrometry

Paul J. Boersema[1,2], Shabaz Mohammed[1,2] and Albert J.R. Heck[1,2,3]

[1]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.
[2]Netherlands Proteomics Centre
[3]Centre for Biomedical Genetics, the Netherlands.

Based on *J. Mass Spectrom.* 2009, 44, 861-878

ABSTRACT

Reversible phosphorylation is a key event in many biological processes and is therefore a much studied phenomenon. The mass spectrometric (MS) analysis of phosphorylation is challenged by the substoichiometric levels of phosphorylation and the lability of the phosphate group in collision-induced dissociation (CID). Here, we review the fragmentation behavior of phosphorylated peptides in MS and discuss several MS approaches that have been developed to improve and facilitate the analysis of phosphorylated peptides. CID of phosphopeptides typically results in spectra dominated by a neutral loss of the phosphate group. Several proposed mechanisms for this neutral loss and several factors affecting the extent at which this occurs are discussed. Approaches are described to interpret such neutral loss dominated spectra to identify the phosphopeptide and localize the phosphorylation site. Methods using additional activation, such as MS^3 and multistage activation (MSA), have been designed to generate more sequence-informative fragments from the ion produced by the neutral loss. The characteristics and benefits of these methods are reviewed together with approaches using phosphopeptide derivatization or specific MS scan modes. Additionally, electron driven dissociation methods by electron capture dissociation (ECD) or electron transfer dissociation (ETD) and their application in phosphopeptide analysis are evaluated. Finally, these techniques are put into perspective for their use in large scale phosphoproteomics studies.

INTRODUCTION

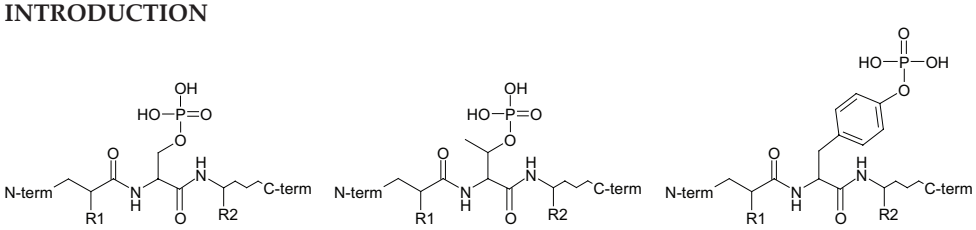


Figure 1. *Chemical structures of phosphoserine, phosphothreonine and phosphotyrosine peptides.*

Reversible phosphorylation is essential in the regulation of cellular proteins, their subcellular localization, degradation, complex formation and furthermore, it is the key event in cell signaling [1]. Phosphorylation in eukaryotes takes place on serine, threonine and tyrosine residues (Figure 1). To succeed in phosphopeptide detection and phosphosite localization by mass spectrometry (MS) one has to overcome several hurdles. The initial challenge is to observe the phosphopeptide of interest in a typically enormous pool of non-modified peptides [2-4]. These large numbers of peptides are created by the necessity to digest proteins to successfully sequence the resulting peptides by MS. Typically, phosphopeptides are present at substoichiometric levels when compared to non-modified peptides of the protein. Also, the ionization efficiency of phosphorylated peptides in positive ion mode MS –as typically is used for liquid chromatography (LC)-MS- is lower compared to non-phosphorylated peptides due to the negative phosphate group [5, 6]. Other challenges are the identification of the phosphopeptide and the localization of the site of phosphorylation, which can be impaired by the lability of the phosphate group in collision induced dissociation (CID)[2]. Typical fragmentation spectra of phosphopeptides are dominated by one peak that is related to the neutral loss of the phosphate group, hampering the level of backbone fragmentation observed, which is essential to determine the amino acid sequence. The obstacles relating to the complexity of the sample and the relatively low abundance of phosphopeptides observed by MS can be alleviated by applying various enrichment techniques, amongst others, based on: strong cation exchange (SCX)[7], TiO_2 chromatography [8, 9], immobilized metal affinity chromatography (IMAC)[10], barium [11] or calcium precipitation [12] and antibodies[13, 14] or combinations thereof, such as SCX followed by IMAC [15], SCX followed by TiO_2 [16] and hydrophilic interaction liquid chromatography followed by IMAC [17]. In this review, we focus, however, on the specific fragmentation of phosphopeptides in MS and how such data can be used to identify phosphopeptides/phosphoproteins and localize the site of phosphorylation. The fragmentation behavior of phosphopeptides in CID will be described and the neutral loss pathway discussed. Furthermore, the advantage of performing additional fragmentation steps will be reviewed as well as the benefits of alternative fragmentation techniques, such as electron capture dissociation (ECD) and electron transfer dissociation (ETD).

COLLISION INDUCED DISSOCIATION

In current proteomics approaches, identification of peptides is generally achieved by tandem MS. CID (also: collisionally activated dissociation; CAD) is the most popular means of gas phase fragmentation of peptides in proteomics. In CID, protonated peptides are accelerated by an electric potential in the vacuum of the mass spectrometer and then allowed to collide with an inert neutral gas (typically helium, nitrogen or argon). Through the collisions, the kinetic energy of the peptide ion is partially converted into internal energy that

is distributed over the molecule, disrupting bonds and causing the peptide ion to fragment. Usually, peptides are fragmented at the amide bonds along the backbone which results in the emergence of amino acid sequence informative b- and y-ions [18-20]. The eventual fragmentation pattern obtained by MS/MS depends on several parameters including the amino acid composition, size and charge state of the peptide ion, the excitation method of the mass spectrometer and the time scale of fragmentation and detection. The 'mobile proton model' is the most comprehensive model to describe the 'charge directed' dissociation of protonated peptides [20-22]. Upon excitation of the protonated peptide, the proton becomes 'mobile' and energetically less favored protonation sites become more populated, including the terminal amino group, amide oxygens and nitrogens and side chain groups [23]. Protonation of the amide nitrogen leads to weakening of the amide bond, and eventually, dissociation at the protonated site when there is sufficient internal energy present. As several different protonated species of the same peptide can coexist, dissociation occurs at different positions along the peptide backbone providing sequence information in the fragmentation spectrum. However, fragmentation generally occurs in a non-random fashion over the backbone because certain amino acids have an influence on the proton localization [22, 24, 25]. Furthermore, different fragmentation pathways give rise to, for example, a-ions or the neutral loss of water or ammonia, and may compete with the backbone cleavages which give rise to additional ions and/or differences in ion intensity in the fragmentation spectrum.

Neutral loss of the phosphate group from phosphorylated serine, threonine and tyrosine residues

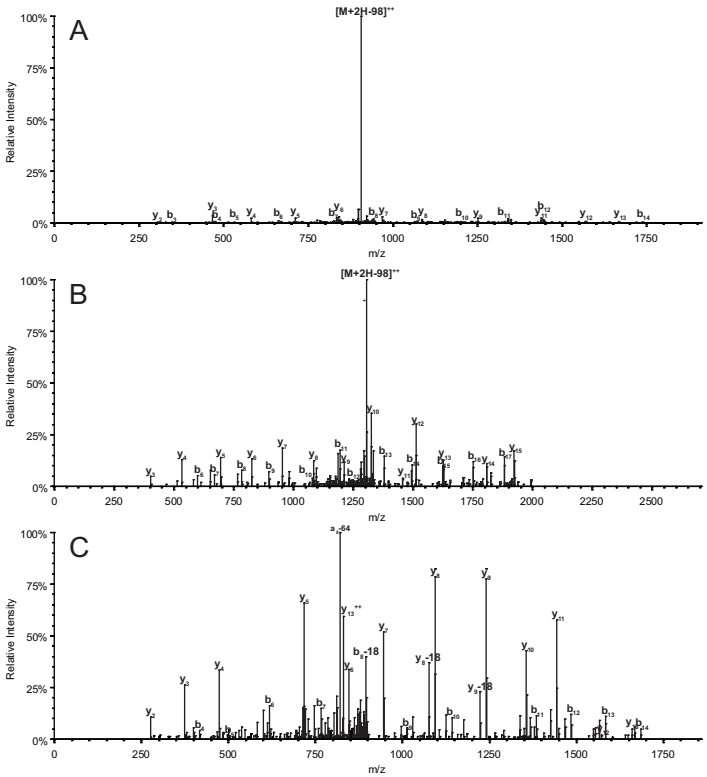
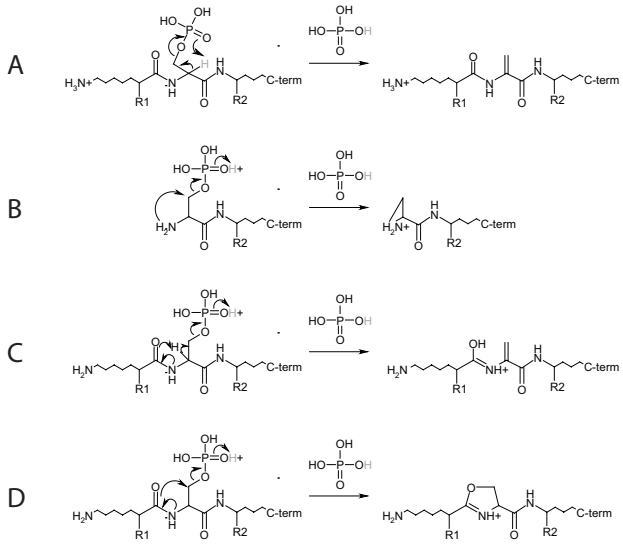


Figure 2. *Typical CID spectra of doubly charged phosphorylated peptides. A) Serine phosphorylated FPTEGEpSDEEEDYER; B) Threonine phosphorylated TLAALEALDpTDGENEEEEYEAWK; C) Tyrosine phosphorylated VTSESDFoxMTEpYVVTR.*

The phosphate group of a phosphopeptide is relatively labile providing a low energy pathway that competes with backbone fragmentation. As a result, a CID spectrum of a phosphopeptide typically contains an intense neutral loss peak that resides at 98 Da or 80 Da lower than the precursor mass representing the loss of H_3PO_4 and HPO_3 respectively (Figure 2). Sometimes, the neutral loss peak dominates to such an extent that no other (sequence informative) ions can be observed. Loss of H_3PO_4 is dominant for serine phosphorylated peptides, for threonine it is present to a somewhat lesser extent. Phosphothreonine containing peptides might also exhibit a loss of HPO_3 [26]. Tyrosine phosphorylated peptides show a much lower extent of neutral loss, being typically exclusively HPO_3 [27].

Neutral loss of H_3PO_4 from phosphorylated serine and threonine residues

Several mechanisms have been suggested for the loss of H_3PO_4 in CID. Generally, neutral loss of H_3PO_4 is assumed to be a β-elimination reaction [26, 27]. However, alternative theories exist that describe a charge directed E2-elimination reaction, or S_N2-neighbouring group participation reactions, whereby an aziridine or an oxazoline ring is formed, respectively [28, 29]. The charge remote β-elimination reaction theory assumes a six-centered intermediate with cyclic rearrangement of three electron pairs [27]. Thereby, the hydrogen located at the α-carbon of the phosphorylated residue is transferred to the phosphate group, ultimately resulting in the formation of dehydroalanine or dehydroaminobutyric acid depending on whether the phosphorylation was on a serine or a threonine residue, respectively (Scheme 1A). In the charge directed pathways, protonation of the phosphate group itself



Scheme 1. *Proposed neutral loss pathways. A) charge remote β-elimination. B-D) charge directed pathways. B) SN2-neighbouring group participation; C) E2-elimination; D) SN2-neighbouring group participation. Indicated in grey is the hydrogen that is abstracted by the leaving phosphoric acid*

is associated with the neutral loss of H_3PO_4. The loss of H_3PO_4 from the single amino acid O-phosphoserine was explained by a nucleophilic attack of the β-carbon of the side chain by the N-terminus and was suggested to result in the formation of an aziridine structure (Scheme 1B) [28]. A dependence on the neighboring amine and leaving group was found, and this reaction might also occur for peptides rather than the studied single amino acid. In the E2-elimination reaction, the neighboring carbonyl oxygen abstracts the acidic hydrogen

on the α-carbon (Scheme 1C). Finally, the S_N2-neighbouring group participation reaction is initiated by the protonation of the phosphate group *via* a protonated neighboring amino acid followed by the nucleophilic attack of the β-carbon on the side chain of the phosphorylated amino acid by the amide carbonyl group. This results in the formation of a cyclic, five-membered oxazoline structure (Scheme 1D).
The charge-remote and charge-directed loss of H_3PO_4 are different in the origin of the hydrogen abstracted by the phosphate group. In the β-elimination reaction, the phosphate group abstracts the hydrogen from the α-carbon of the phosphorylated amino acid residue, while in the charge directed mechanism the hydrogen is the mobile proton. This origin of the hydrogen has been exploited to determine which mechanism dominates the neutral loss of H_3PO_4 by performing uniform and regioselective H/D exchange [27-29]. After H/D exchange of a phosphorylated peptide, all hydrogens except the backbone amide hydrogens are exchanged with deuterium atoms. If neutral loss of the phosphate group in CID results in a loss of 100 Da (HD_2PO_4) this would be indicative of a β-elimination reaction as the hydrogen then originates from the α-carbon. A loss of 101 Da (D_3PO_4) would be observed if a charge directed mechanism is responsible for the neutral loss. Interestingly, Tholey *et al.* [27], Reid *et al.* [28] and Palumbo *et al.* [29] performed similar H/D exchange experiments to determine the mechanism behind the neutral loss of the phosphate group, but drew opposite conclusions. Tholey *et al.* observed a loss of HD_2PO_4 and concluded the mechanism behind neutral loss to be β-elimination [27]. Reid *et al*. and Palumbo *et al.* repeated the experiment, but detected a loss of D_3PO_4 and concluded a charge directed mechanism to underlie the neutral loss of H_3PO_4 in CID, which may be a more justified conclusion considering the quality and detail of their spectra [28, 29]. Regio-selective side chain deuteration labeling of a phosphoserine containing peptide further confirmed their hypothesis and disproved β-elimination to be the dominant pathway in neutral loss [29]. Interestingly, MS^3 analysis of the singly and doubly charged $[M+nH-H_3PO_4]^{n+}$ and $[M+nH-H_2DPO_4]^{n+}$ product ions revealed rather different fragment ions in the tandem mass spectra. This would suggest that the loss of H_3PO_4 resulted in a peptide product exhibiting a different structure than the product formed after the loss of H_2DPO_4, implying that the charge directed and β-elimination reactions might be in competition, with the charge directed mechanism being the more abundant [29]. Additional confirmation of a charge directed neighboring group underlying the neutral loss of the phosphate group was obtained by comparing side chain losses in MS^3 spectra of neutral loss products of O-phosphoserine and several serine derivatives [28]. It appeared that in CID different neutral loss products arise from the $[M+H-H_3PO_4]^+$ ion of phosphoserine compared to the $[M+H]^+$ ion of dehydroalanine, while similar MS^3 spectra were obtained for a fixed charge derivative where only neighboring group participation reactions are expected to occur. This also indicates that other reaction mechanisms besides β-elimination are in play [28]. Additional neutral losses of CH_2O and CH_3CHO that were observed in the MS^3 spectra could only be explained by a S_N2-neighbouring group participation reaction and ruled out the charge directed E2-elimination reaction [29]. In summary, the neutral loss mechanism of H_3PO_4 from phosphorylated serine and threonine residues seems to be dominated by a charge-directed mechanism. However, this may happen concordantly with, to a lesser extent, a β-elimination mechanism.

Neutral loss of HPO_3 and H_3PO_4 from phosphorylated tyrosine residues

CID of tyrosine phosphorylated peptides on occasion results in the neutral loss of 80 Da which represents the loss of HPO_3 [26]. H_3PO_4 cannot be lost from the phosphorylated tyrosine residue because the aromatic ring does not allow an E2-elimination or S_N2-neighbour-

ing group participation reaction to occur. Furthermore, in the C-O-P structure of phosphorylated tyrosine, the C-O bond is stronger than the same bond of phosphorylated aliphatic amino acids due to stabilization by the aromatic ring [27]. The O-P bond is weaker and thus loss of HPO_3 is preferred. The exact mechanism of the neutral loss of HPO_3 from a phosphorylated tyrosine has yet to be fully established, but the involvement of a basic moiety has been suggested [30-32]. A protonated tyrosine phosphorylated peptide without basic residues did not produce any neutral loss [31]. However, the sodiated version of the same peptide did produce a neutral loss and it was suggested that a cation-π interaction between the sodium ion and the phenyl group of the phosphotyrosine residue would underlie the neutral loss of the sodiated phosphopeptide [31]. A similar interaction of the arginine or lysine side-chain with the π-cloud of the phosphotyrosine phenyl ring under proton depleted conditions was proposed to facilitate the gas phase phosphate fragmentation [30]. In some instances, a loss of 98 Da rather than 80 Da is observed for tyrosine phosphorylated peptides [30]. This can be attributed to a simultaneous or consecutive loss of HPO_3 from the phosphorylated tyrosine residue and H_2O from another residue [27, 32]. An alternative theory was formulated, wherein the HPO_3 group is transferred to an aspartic acid residue after which a cleavage of H_3PO_4 generates a succinimide [32].

Extent of neutral loss

The extent of neutral loss of H_3PO_4 and HPO_3 from phosphorylated residues is dependent on several aspects including the chemical structure of the phosphorylated amino acid residue, the amino acid sequence, charge state of the precursor ion, proton mobility, the type of mass spectrometer and the applied collision energy.

As discussed above, in CID, serine and threonine display mostly a loss of H_3PO_4. For phosphorylated threonine it has been suggested that it should exhibit slightly less neutral loss than serine [26, 27, 29]. This might be related to the steric hindrance of the β-methyl group in the side chain of threonine. Phosphorylated tyrosine shows the least neutral loss, but occasionally a neutral loss of HPO_3 can be observed as described above.

Precursor ion charge state and number of basic residues are interconnected characteristics in terms of the extent to which a neutral loss can be observed. Generally, with increasing charge, a phosphopeptide displays a reduced neutral loss [26, 27, 29, 33]. Furthermore, the extent of neutral loss seems to be dependent on the ratio of charge state *versus* number of basic amino acid residues (arginine, lysine, histidine)[29]. When the charge state is higher than the number of basic residues, less neutral loss is observed. Thus, when a 'mobile proton' is available, charge-directed backbone fragmentation can occur more easily and less fragments from the neutral loss fragmentation pathway are observed. In the case of no available mobile proton, the energy required for backbone cleavages is much higher and other pathways that do not require full deprotonation are energetically more favorable. One of these low-energy pathways is neutral loss, assisted by protonated arginine or lysine residues as participating S_N2 neighboring group donating entities [29]. For example, in Figure 3A, the fragmentation spectra of two peptides at charge 2+ and 3+ are compared. Where in the spectrum of the 2+ precursor ion (one 'mobile' proton) the neutral loss peak is the dominant peak, in the spectrum of 3+ (two 'mobile' protons present) this peak is significantly reduced relative to other (backbone) fragment ions.

Another parameter that determines the extent of neutral loss, is the amount of collision energy used in the CID experiment [34, 35]. Generally, ion trap mass spectrometers generate a higher level of neutral loss than 'tandem in space' collision cell mass spectrometers [36],

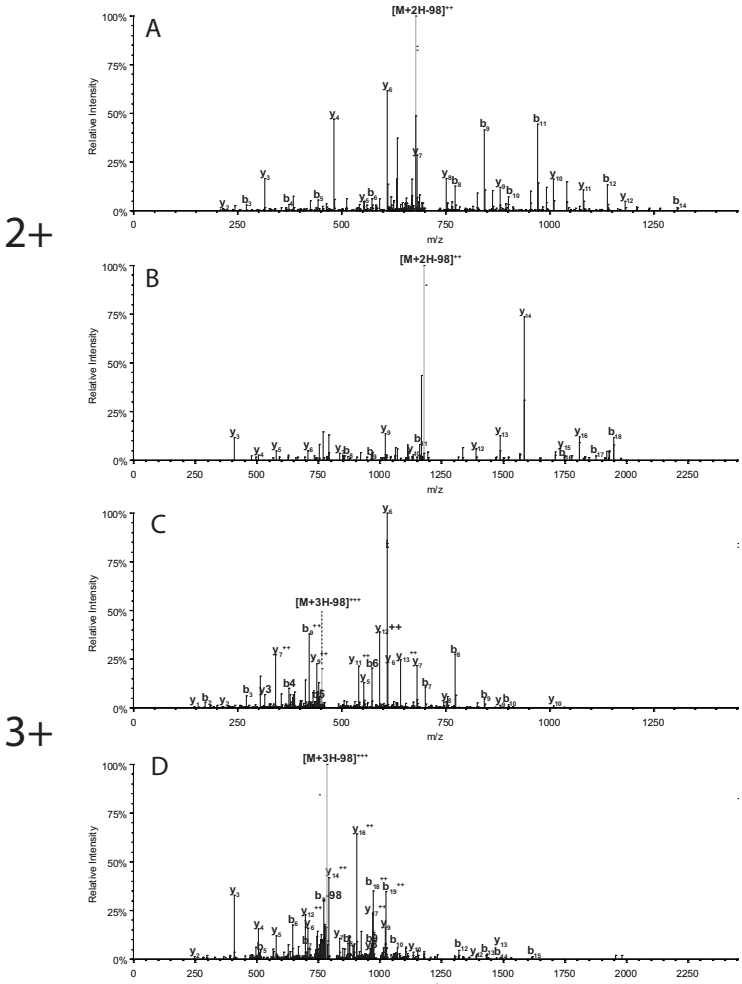


Figure 3. *The effect of precursor ion charge state on neutral loss. A higher extent of neutral loss can be observed in the CID spectra of doubly charged precursor ions (panel A and B) than in their triply charged counterparts (panel C and D). The neutral loss peak is highlighted in grey. A and C) GDVTAEEAAGApSPAK; B and D) EVSSLEGpSPPPCLGQEEAVCTK.*

which can be explained by the difference in collision energy used and the timeframe allowed for fragmentation. The slow heating approach of CID applied in ion traps through resonant excitation will favor low-energy pathways such as neutral loss fragmentation when compared to tandem in space instruments, where a higher level of internal energy is deposited much quicker. In Figure 4, fragmentation spectra are shown of two peptides analyzed by both an ion trap and a 'tandem in space' collision cell mass spectrometer. The neutral loss peak is less abundant in the collision cell MS spectra and more products of backbone cleavages can be observed.

Overall, differences in the extent of neutral loss can be explained by the above mentioned factors. However, it is still difficult to *a priori* determine the extent of neutral loss.

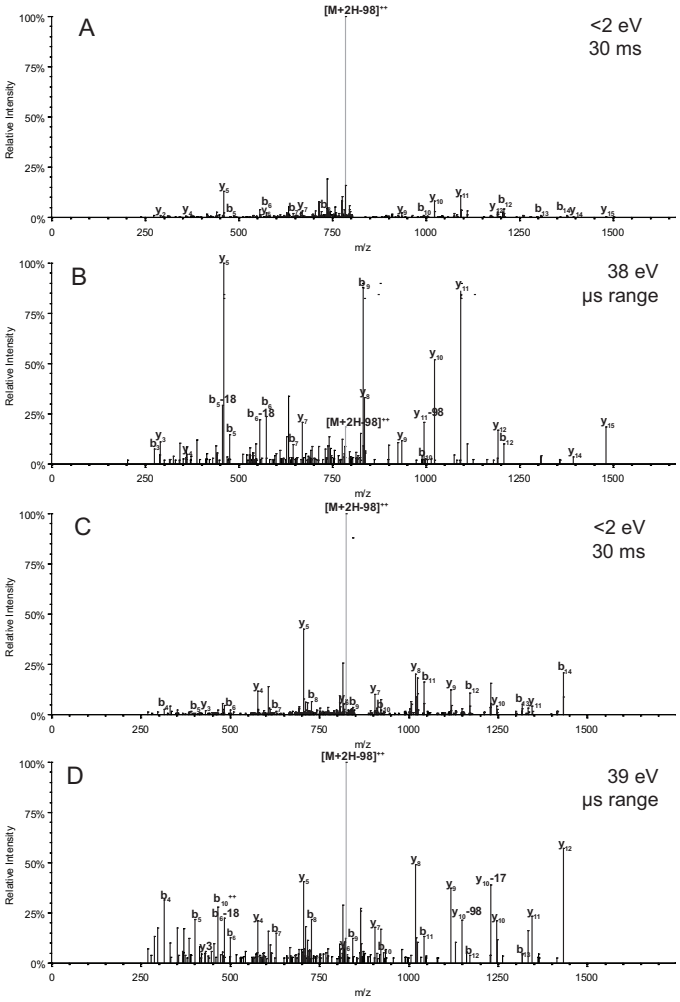


Figure 4. *The effect of collision energy and activation time on the extent of neutral loss. CID spectra of the same phosphopeptides were obtained at low collision energy (<2 eV) and relatively long activation times (30ms) in a linear ion trap (A and C) and higher collision energy (38-39 eV) and shorter activation time (microsecond range) in a Q-TOF (B and D) with the neutral loss peak highlighted in grey. A,B) SVSSNVASVpSPIPAGSK; C,D) GGSIS-VQVNSIKFDpSE.*

Phosphosite localization

Determining the exact site of phosphorylation can be even more demanding than identifying the phosphopeptide itself. Almost 16% of the amino acids in the current full NCBI database (20082710) are serine, threonine or tyrosine. Therefore, the chance of a given peptide to contain more than one possible phosphorylation is quite high. To confidently localize the site of phosphorylation requires specific diagnostic backbone fragments (also 'site determining ions') to be present [37]. To exemplify how to localize a phosphorylation site, in Figure 5A, a fragmentation spectrum of a tryptic peptide is shown and compared with the fragmentation spectrum of the phosphorylated analogue (Figure 5B). Seven potential phosphorylation sites are present in the sequence: Y3, S4, T6, S7, T10, T12 and T14. In figure 5B, fragment ions b_3 to

b_9 have masses corresponding to the fragment ions that are unmodified indicating that these peptide fragments are not phosphorylated. Therefore, Y3, S4, T6 and S7 can be excluded as phosphorylation sites. Examining the C-terminal fragments, y_3 to y_6 also correspond to non-modified peptide sequences. This excludes T12 and T14 as possible phosphorylation sites and leaves T10 as the only option. This is further confirmed by b_{10}-b_{15} and y_7-y_{14} which originate from phosphorylation containing peptide fragments. Finally, also some fragments are observed that have lost the phosphate group that further confirm the phosphorylation site to be T10. Note, backbone fragment ions are present n two forms, *i.e.* with and without the phosphate group, which corresponds to +80 Da or -18 Da respectively compared to the unmodified peptide. One should be aware that a loss of -18 Da is not a positive identification of a loss of the phosphate group, since it can also correspond to the loss of water from an unmodified residue. The fragment ion with an intact phosphate is required for direct validation. It is possible to use a fragment with a neutral loss if all other amino acid candidates can be eliminated. In Figure 5C, a fragmentation spectrum of another phosphopeptide is

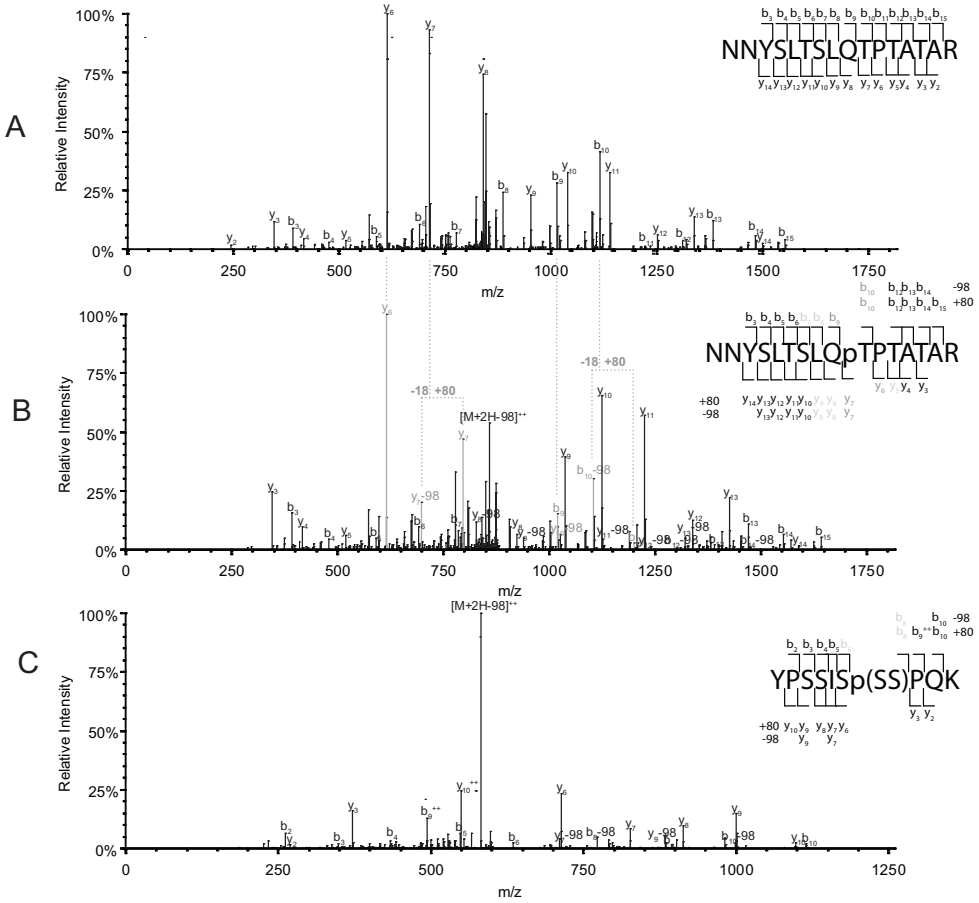


Figure 5. *Phosphorylation site localization. A) CID spectrum of a non-phosphorylated peptide compared with the CID spectrum of the phosphorylated counterpart (B). Indicated on the peptide sequence are the fragment ions that were found, including ions that lost 98 Da or were 80 Da heavier than in the non-phosphorylated peptide. Highlighted in dark grey are the site determining ions and the corresponding peaks in the spectrum. In light grey are indicated fragment ions that confirm the site localization C) A CID spectrum of a phosphopeptide of which the exact site of phosphorylation could not unambiguously be determined. In light grey are highlighted fragment ions that indicate that the phosphorylation is on either S7 or S8.*

depicted. Six possible phosphorylation sites are present on this peptide, one tyrosine and five serine residues. Fragment ions y_6-y_9 suggest that the phosphorylation site is on the S6, S7 or S8. Ions b_2-b_6 exclude Y1, S3, S4 and S6 to be phosphorylated. Yet, two options exist of residues that might be phosphorylated: S7 and S8. However, no site-determining ions exist that resolve the final ambiguity.

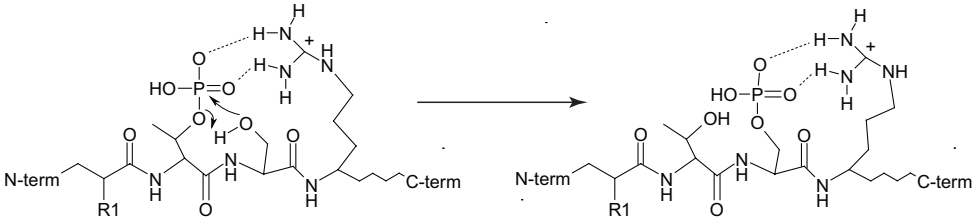
Recently, it was shown that under conditions of low proton mobility, *i.e.* when neutral loss is highest, phosphorylated serine and threonine-containing peptides that also contain an arginine might induce a charge-remote cleavage pathway that results in an increased intensity of the y-ion C-terminal of the phosphorylated residue [38]. This might be used as additional information to pinpoint the site of phosphorylation.

Several software tools and scoring schemes have been developed to automatically process fragmentation spectra of phosphorylated peptides and localize the site of modification. An often used and rather straightforward scoring system is the "delta score". A Mascot or Sequest score is calculated for all different possible localizations of the phosphorylation on the phosphopeptide. The localization with the highest score is the most likely candidate because the most fragment ions could be matched to this form of the peptide. The difference between the highest and the second highest gives an indication of the level of ambiguity in the phosphosite localization. The larger the difference, the more confident one can be in the localization of the highest score. In the example of Figure 5B, the peptide with phosphorylation at T10 obtained a Mascot score of 66, while the peptide with the second highest score, 54, had the phosphorylation at S7. In this example the difference –or delta score – would be 12. The Mascot score is given as S = -10* log(P) whereby P is the absolute probability that the returned hit is a random event. The phosphopeptide isoform with a score of 54 is therefore 16.6 times more likely a random hit than the one with the score of 66. In the second example, Figure 5C, the peptide with phosphorylation at S8 obtained a Mascot score of 35, while the peptide with phosphorylation at S7 scored 34. The delta score of only 1 is too low to confidently place the phosphorylation site on S8. In that case, the ambiguity of the phosphorylation site remains.

The post translational modification (PTM) scoring tool that is implemented in the MSQuant software [39, 40] is based on an algorithm that uses the four most intense ions per 100 *m/z* in an MS^2 or MS^3 spectrum. Theoretical spectra of all possible phosphorylation sites are then matched and the highest scoring option is reported. Ascore [37] is a variation on PTM scoring and is based on the presence and intensity of site-determining ions. A probability score is calculated of the phosphorylation localization by using these ions, whereby a score of 19 is considered highly confident. Several other scoring methods exist, some also incorporate MS^4 [41] or ECD/ETD spectra [42]. While these scoring tools are helpful in the processing of large scale phosphoproteomics datasets, they are probability-based algorithms that still have difficulties with peptides whereby multiple serine, threonine or tyrosine residues are close or next to each other. As in these instances only one b- and y-ion can determine the exact phosphorylation sites, localization scores are low and random matching chances may be high. Especially when spectra are noisy or when backbone fragments have to be found within the low intensity peaks, probabilities get low and erroneous assignment of site localization can occur.

Next to the fact that too little backbone fragmentation might make confident localization of phosphorylation difficult, a more serious effect might interfere with the correct localization. Palumbo *et al.* recently provided evidence that in the gas phase during tandem MS, the phosphate group might be rearranged and transferred prior to fragmentation to another serine, threonine or tyrosine residue present in the precursor peptide [43]. It has been sug-

gested before that rearrangement of a phosphomoiety might underlie the loss of H_3PO_4 from a tyrosine phosphorylated peptide [26, 32]. The phosphate group rearrangement was clearly displayed in a recent paper wherein synthetic phosphopeptides with known serine, threonine and tyrosine phosphorylation sites were subjected to and analyzed by tandem MS [43]. Of the spectra of the studied phosphopeptides, 45% exhibited ions that were the product of transfer reactions of the phosphate group. A protonated arginine was suggested to facilitate the transfer of the phosphate group from the original site of phosphorylation to another hydroxyl containing amino acid along the peptide backbone (Scheme 2) [43]. This was found to occur particularly under partial and non-mobile proton conditions, *i.e.* the conditions under which neutral loss of the phosphate group is most predominant. It was noted that this effect was dependent on the CID activation time. While at the shortest activation time (2 ms) in an ion trap mass spectrometer still rearrangement products ions were observed, in a triple quadrupole mass spectrometer -where the activation time is in the microsecond range- no evidence for these rearrangements were found. This effect can only be studied when the phosphorylation sites are known *a priori*. The extent of rearrangement cannot be established when studying larger phosphopeptide datasets of which the phosphorylation sites are still to be determined. Yet, these rearrangement effects might explain some of the contradictory phosphosite assignments.



Scheme 2. *Proposed mechanism for the gas phase phosphate group rearrangement from phosphothreonine to serine. Based on Palumbo et al. [43].*

MS^3 and multistage activation

A method to obtain more sequence information from peptides that show extensive neutral loss in their CID spectra, is the use of a second step of activation. By activating the ion produced by the neutral loss, peptide backbone fragments can be obtained that do reveal the amino acid sequence. In such an MS^3 experiment, which is generally performed in an ion trap mass spectrometer, first a survey scan is performed to determine the precursor masses and charge states (MS). The trap is then refilled and a precursor ion isolated for fragmentation (MS/MS, or MS^2) which may result in a spectrum with a dominant neutral loss peak (Figure 6A). In the third step, the trap is refilled, the precursor ion is isolated, activated and fragmentated followed by the isolation of the neutral loss fragment ion for a second step of activation and fragmentation (MS/MS/MS, or MS^3). As the low-energy pathway of neutral loss of the phosphate group is not available anymore, more backbone fragmentation occurs, which results in relatively more sequence-informative fragments (Figure 6B). The formerly phosphorylated serine residue can now be recognized as having lost 18 Da ($-H_2O$), although some speculation remains as to the exact structure of the amino acid [29]. MS^3 is normally performed in a data-dependent manner in which the third step of MS is trigged by the presence of an intense peak at the *m/z* of a neutral loss, which is calculated on the basis of the precursor ion mass and charge state (e.g. $[M+2H]^{2+}$-49, $[M+3H]^{3+}$-32.6, etc.). Data dependent neutral loss triggered MS^3 has been performed in several large-scale phosphoproteomics studies [7, 16, 39, 44-46] (and many more). Information from both MS^2 and MS^3 spectra can

be combined to determine the amino acid sequence of the analyzed peptide either as combined or separate spectra [47, 48].

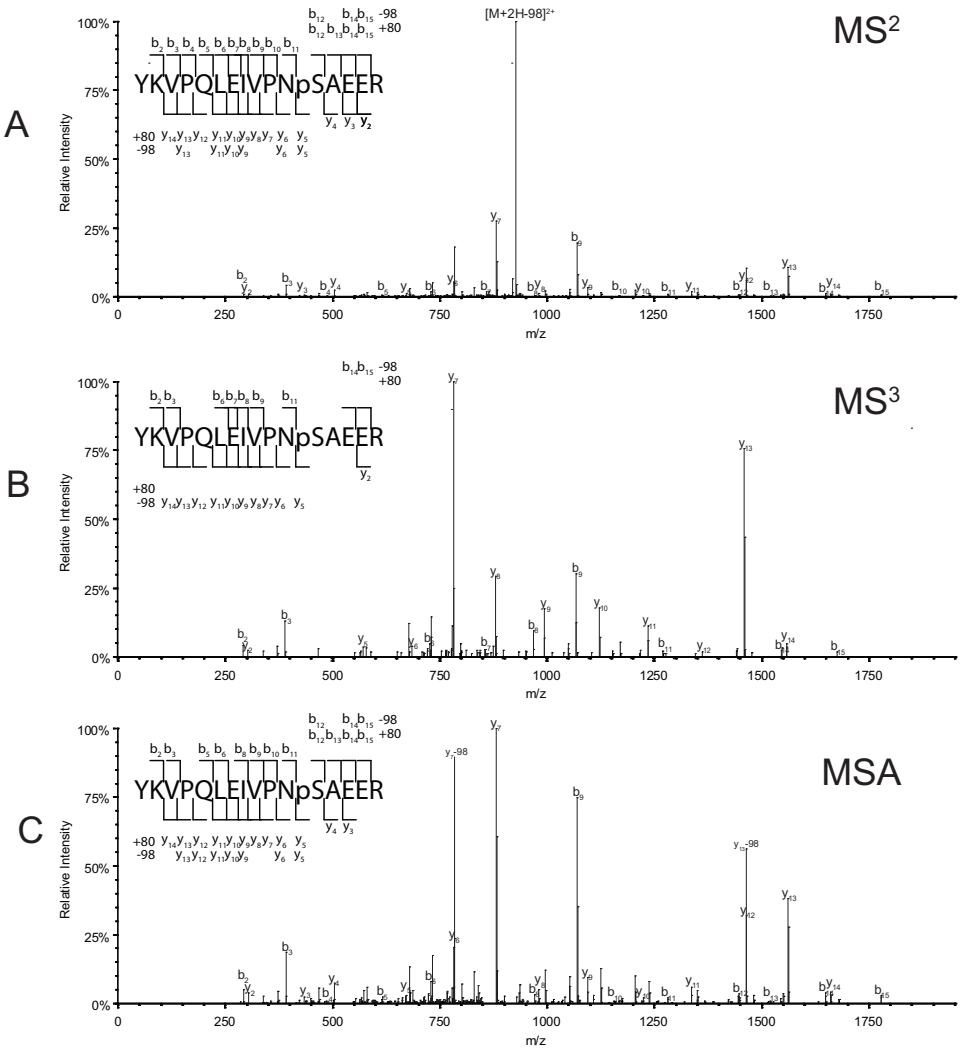


Figure 6. *Multiple stage fragmentation. A) MS/MS of doubly charged phosphopeptide YKVPQLEIVP-NpSAEER. B) MS3 of the $[M+2H-98]^{2+}$ from (A) (Note in the MSA spectrum all b-98 and y-98 are annotated as b and y). C) MSA of the same phosphopeptide with additional activation aimed at $[M+2H-98]^{2+}$.*

The repetition of MS^2 in order to perform an MS^3 experiment increases the analysis time per peptide considerably. To reduce the duty cycle and not to be confronted with contradictory sequences from MS^2 and MS^3 [47]. the MS^2 and MS^3 step can be combined in a 'pseudo-MS^3' or 'multistage activation' (MSA) type of analysis [49]. Herein, a precursor ion selected after the first MS analysis is activated at its *m*/*z*, followed by an additional activation at the *m*/*z* where theoretically the neutral loss ion should reside, while all the other fragments are still trapped in the ion trap. In Figure 6C, an MSA spectrum of a phosphopeptide is shown and

compared with the MS^2 and MS^3. The strong neutral loss peak from the MS^2 is missing in the MSA as the ion was further fragmented. Therefore, the (relative) intensity of sequence informative fragments is much increased. Normal b- and y-ions are formed, which are found at the same *m/z* as in the corresponding MS^2 spectrum. However, also b-98-ions and y-98-ions are present. These are the fragments of the neutral loss ion and are the same as the b- and y-ions found in the MS^3. Hence, an MSA spectrum can basically be regarded as a composite spectrum of the MS^2 and MS^3. Compared to an MSA spectrum, an MS^3 spectrum is generally of lower quality as the production of the neutral loss from the precursor ion is complemented by partial fragmentation, so the MS^3 will be constructed from a lower ion count [33]. MSA is most suited for samples that are rich in phosphopeptides. As no neutral loss triggering is performed, the additional time of activation might be wasted in the case of non-phosphorylated peptides where no neutral loss peak at $[M+nH-98]^{n+}$ is present. MSA has shown its usefulness in a number of large-scale studies [33, 39, 50].

Several studies have described the pros and cons of performing MS^3 or MSA. Slightly different conclusions have been drawn regarding the effect of additional MS steps in the analysis of phosphopeptides. Initially, large benefits of MS^3 were reported whereby 43-56% of phosphopeptides could only be identified after performing MS^3 [44, 51]. Later, other work reported much smaller increases in the number of identified phosphopeptides by using MS^3 [37, 47]. A great redundancy in MS^2 and MS^3 data was shown wherein 82% of the phosphosites identified with MS^3 were already identified by MS^2 [37]. A few large scale phosphorylation studies have directly compared MS^2, MS^3 and MSA approaches [33, 50]. Large overlaps of phosphopeptides were found between MS^2, MS^3 and MSA in these high-throughput approaches. However, also some unique sets of peptides were identified for each of the approaches. For example, MS^3 and MSA identified more multiply phosphorylated peptides [50]. In this respect it is interesting to look at another study where a sample was reanalyzed by MSA with an inclusion list of all the precursor ions that showed a large neutral loss, but could not be identified by MS^3 [52]. An additional 107 out of a final total of 1628 phosphosites were identified in this way, showing that MSA is not exactly identical to MS^2 and MS^3 combined and a different subset of phosphopeptides may be identified. Ulintz *et al.* found MSA to perform slightly better than MS^3 and MS^2 [50], while Villén *et al.* found MS^2 to outperform MSA and MS^3 [33]. This discrepancy might be related to difference in MS set-up. Villén *et al.* used shorter LC-MS runs with relatively short ion accumulation times for MS^2 [33]. Any increase in the MS duty cycle because of the additional activation in MS^3 or MSA would have a stronger negative impact on the overall performance. Interestingly, in comparing the statistics of ion types found in MS^2, MS^3 and MSA spectra surprisingly few differences were found [50]. As expected, the MS^3 spectra contain more neutral loss fragment ions and fewer backbone cleavage fragment ions than MS^2 spectra. Based on ion statistics, MSA appears merely an MS^2 spectrum with the neutral loss peak removed rather than a combined MS^2 and MS^3 spectrum.

Also, the effect of MS^3 and MSA on phosphorylation localization was found to be limited [33, 50]. Basically, peptides that generate sequence-rich MS^3 or MSA spectra, *i.e.* spectra that can be used to localize the phosphosite, already generate sequence-rich MS^2 spectra, and in quite a few cases, the phosphosite could only be localized from the MS^2 spectrum and not from the MS^3 [33]. A perturbing observation concerning the phosphosite localization by multiple fragmentation stages, next to the above discussed rearrangement of the phosphate group, was that an apparent neutral loss of H_3PO_4 might actually be a simultaneous loss of HPO_3 from the phosphorylated residue and H_2O from another serine or threonine residue

[43]. In an MS^3 spectrum, normally the phosphorylation site is localized by determining the serine or threonine residue with -18 Da (*i.e.* $-H_2O$). If, however, H_2O is lost from another, non-phosphorylated serine or threonine residue, this might be assigned as the formerly phosphorylated site and while incorrect still get a high localization score [43].
Overall, it seems the additional effect of MS^3 and MSA is not very significant in large-scale phosphorylation analyses wherein high mass accuracy MS instrumentation is used in a high-throughput manner. Basically, although neutral loss ions may dominate the tandem mass spectra, the capacity of these ion traps is sufficient to still detect ion peaks diagnostic for peptide sequence determination as well as phosphosite localization. In these high-throughput approaches one normally strives to the highest achievable number of sequencing events. The extra time spent on additional fragmentation in MS^3 and MSA will decrease the number of sequencing attempts and this unconstructive effect might not be sufficiently compensated for by the additional information generated by MS^3 and MSA. Furthermore, the benefit of MS^3 and MSA is dependent on the experimental approach. For example, MS^3 and MSA can often still be useful when there is sufficient sequencing time, no limit in sample consisting solely of phosphopeptides. In such cases, MS^3 and MSA could be used to obtain sequence information of phosphopeptides that could not be identified in MS^2.

Phosphopeptide derivatization
To overcome the ionization issues of phosphopeptides in positive ion mode MS and the domination of neutral loss in CID, which are both associated with the nature of the phosphate group of a phosphorylated peptide, several derivatization and dephosphorylation approaches have been proposed and assessed to remove the negative charge of the phosphate group [5, 6, 53-58]. Dephosphorylation of the phosphopeptide by alkaline phosphatase can potentially solve both of the issues of neutral loss and poor ionization [57]. Recently, a number of groups have evaluated the efficacy of this approach where it was confirmed that after removal of the phosphorylation a higher signal intensity is observed in MS, particularly for multiply phosphorylated peptides [5, 6]. In contrast, Steen *et al.* saw no increased MS signal intensity of non-phosphorylated peptides compared to their phosphorylated counterpart [59]. However, the synthetic phosphopeptides used in this study were mostly rather atypical, non-tryptic peptides containing several basic residues. To identify phosphopeptides after dephosphorylation, a sample can be run in parallel with and without phosphatase treatment and the dephosphorylated peptide can be recognized by its appearance 80 Da lower than the disappearance of another peptide [5, 6]. The sequence of non-phosphorylated peptide would be relatively easy to determine since the fragmentation spectrum will not be dominated by a neutral loss peak. However, the greatest downside of this approach is that it will be impossible to obtain the specific site of phosphorylation when more than one possible phosphosite is present in the peptide sequence [6].
Another way to remove the phosphate group is by in-solution β-elimination [53, 54, 56, 60-63]. Exposing a phosphopeptide to highly alkaline conditions results in the β-elimination of the phosphate group which leaves a reactive dehydroamino acid that can react with a nucleophile in a Michael addition reaction [54, 56]. This, in turn, might be linked to an immobilizing agent for enrichment of the modified phosphopeptides [54, 60-63]. This has also been done in a one step process whereby the phosphorylated serine and threonine residues were chemically transformed into lysine analogs to allow phosphospecific proteolysis [62]. While neutral loss is abolished in these approaches, the site of original phosphorylation can still be recognized and localized. Using different isotopomers of the nucleophile in the Michael addition reaction step enables MS based quantification of the modified phosphopeptides [61].

However, the β-elimination reaction requires precise quantities of reactant to minimize side reactions [54]. Furthermore, the β-elimination reaction of phosphorylated threonine is much slower than for phosphorylated serine and phosphorylated tyrosine is left untouched [54]. Finally, it was demonstrated that also O-glycosylated serine and threonine residues [63, 64] and even the hydroxylgroups of non-phosphorylated serine and threonine can be modified in this β-elimination and Michael addition reaction, and therefore, impair the determination and localization of phosphorylation events [60, 65].

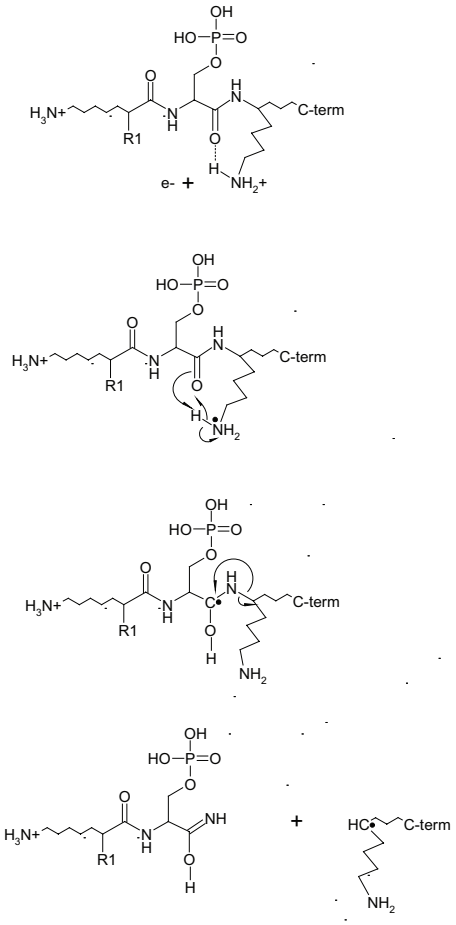
Phosphorylation analysis by targeted instrumental scan modes

Several other mass spectrometric approaches have been developed for a more targeted phosphorylation analysis using specific MS scanning modes. For example, neutral loss ion scanning on a triple quadrupole mass spectrometer can be used to detect the loss of the phosphate group from phosphopeptides [27, 34]. In such a set-up, Q1 is used to scan the whole mass range. Q2 is used as a collision cell, while Q3 is set to scan at $[M+nH-98]^{n+}$ to detect phosphopeptides that lose the phosphate group [66]. In another approach, precursor ion scanning was performed in the negative ion mode to detect the loss of *m/z* 97 ($H_2PO_4^-$) or the more phosphorylation specific loss of *m/z* 79 (PO_3^-) and/or 63 (PO_2^-) at higher collision energies [67-69]. By using the loss of PO_3^- tyrosine phosphorylated peptides can also be detected. The mass spectrometer can then be switched to positive ion mode to perform CID on only the peptide precursors that showed the loss that was scanned for [68, 70]. To analyze tyrosine phosphorylated peptides in positive ion mode a pseudo precursor ion scanning method can be used, where the time-of-flight (TOF) part of a quadrupole (Q)-TOF mass spectrometer can be set to 'scan' the immonium ion of phosphorylated tyrosine at *m/z* 216.043 [71, 72]. In general, however, the scan times for these ion scanning methods limit the use in online LC-MS.

The mass defect, *i.e.* the difference between the nominal and actual mass, of ^{31}P is greater than for most of the other atoms present in a peptide. The addition of HPO_3 to a peptide shifts the mass defect by -0.0337 amu, while the mass defects of ^{1}H, ^{14}N and ^{16}O are only 0.0078, 0.0031 and -0.0051 amu, respectively [73]. On the basis of purely the accurate mass, this mass defect can be used to predict whether a precursor ion is a phosphorylated or a non-phosphorylated peptide. With a measurement error of up to 20 ppm, 16.5-40.5% of known phosphopeptides could theoretically be identified as likely phosphorylated [73]. Implementation of knowledge on these mass defects in LC-MS approaches might involve targeted MS/MS sequencing of precursor ions that are likely phosphorylated.

ELECTRON-DRIVEN DISSOCIATION METHODS

As discussed above, fragmentation of phosphopeptides by CID is often dominated by neutral losses which can seriously hamper identification and phosphosite determination. Recently, novel electron based dissociation methods have been introduced that leave post translational modifications such as phosphorylation and glycosylation intact during fragmentation [74, 75]. In ECD, precursor ions are bombarded with near-thermal electrons (<0.2 eV). The capture of an electron by the peptide ion is an exothermic process which charge-reduces the peptide into a radical cation [76]. The basicity of the amide carbonyl oxygen is thereby increased such that it can abstract a proton from an amino acid residue in the peptide sequence. Very low energy barriers have been calculated for the dissociation of the N-Cα bond in the aminoketyl radical that is formed (Scheme 3)[76]. This results in the generation of c- and z-dot ions and, in much lower abundance, c-dot and z-ions, when a hydrogen is transferred. The presence of a radical makes this fragmentation very specific for the N-Ca



Scheme 3. *Proposed ECD/ETD fragmentation mechanism of phosphorylated peptides. Based on Syka et al. [75].*

bond and little secondary loss of termini and sidechains occur [74]. Only amine cleavages within proline residues are rarely observed, as this would require the cleavage of a second bond in the cyclic side chain [77]. It should be noted that ECD mechanisms are constantly being refined [74, 76, 78-80]. In Figure 7, an example of an ECD spectrum of a phosphopeptide is displayed. Compared to a CID spectrum of a phosphopeptide, it is predominantly made up of backbone fragments and no neutral loss can be observed. ECD is mostly performed on Fourier transform ion cyclotron resonance (FTICR) instruments, although some attempts have been made to implement ECD in modified ion traps [81-84]. An ion trap is required for ECD as the electron capture process by the precursor ion usually requires several milliseconds [85], which is longer than the residence time in a tandem in space collision cell. The means of trapping the ions for ECD are often based on a combination of magnetic and electric fields (FTICR) rather than radiofrequency (rf, other ion traps). The highest ECD efficiency is obtained with near-thermal electrons (<0.2 eV). In rf traps, however, electrons would maintain their thermal energy for only a few microseconds and, more importantly, will not be trapped [75, 86]. The electron capture cross section is proportional to the square of the ion charge [86], making ECD best suited for multiply charged ions generated by electrospray

ionization (ESI). However, only after electron production was changed from a conventional heated filament source, whereby the ion irradiation step took 3-30 secs [87], to an indirectly heated dispenser cathode, whereby the ion irradiation step takes a few milliseconds [85] could ECD be successfully coupled to online LC separation. Most studies have focused on digests of model proteins or simple protein mixtures and found better sequence coverage and more phosphopeptides compared to CID [88-92].

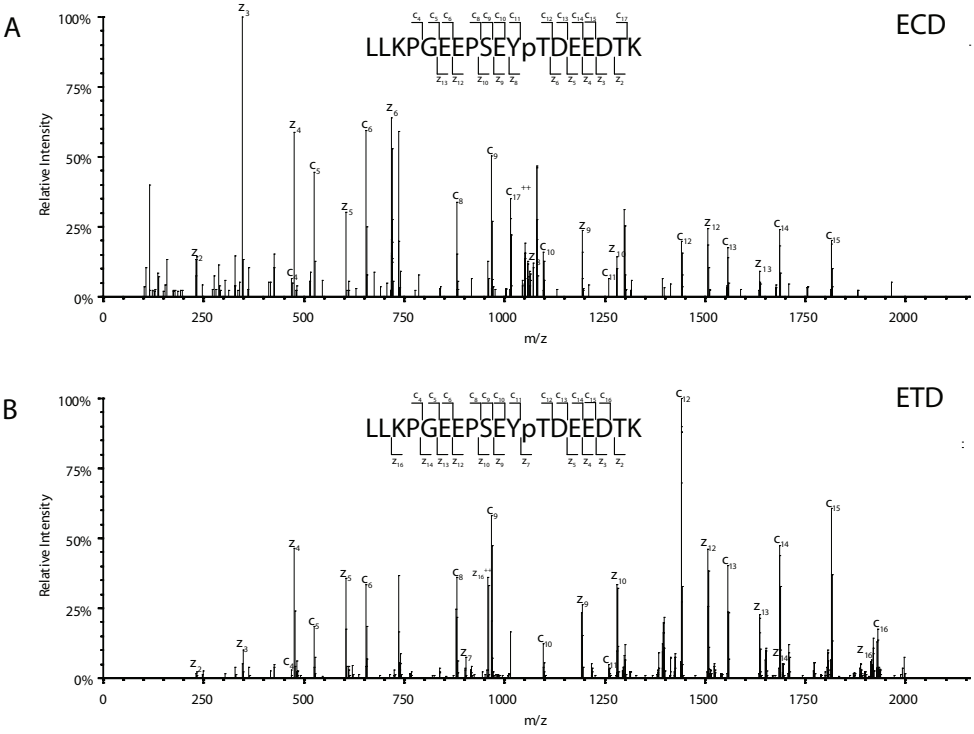


Figure 7. *A) ECD and B) ETD spectra of the same phosphopeptide. Indicated on the peptide sequence are the ions that were detected. The spectra are largely similar. The ECD spectrum was kindly provided by Dr. Helen J. Cooper and Dr. Steve M.M. Sweet.*

Up to now, the only large scale phosphoproteomics study using ECD acquired paired CID and ECD spectra by neutral loss dependent ECD [92]. Fifty percent of the phosphopeptides that were identified with the paired approach were identified with both CID and ECD, 34% was identified with CID alone and only 13% was identified with ECD alone, even though the ECD spectra were acquired in the high mass resolution FTICR cell and the CID spectra in a linear ion trap. The poorer performance of ECD might be explained by the analyzed peptides being trypsin products that are generally doubly charged after ESI. These peptides do fragment well in CID, but are suboptimal for ECD [86]. It has even been suggested that salt bridges might be formed between a protonated amide side chain and the deprotonated phosphogroup [93]. This will hold the two ECD fragments together at lower charge states and will only be broken by coulombic forces concurrent with higher charge states or at increased ECD electron energies. For phosphorylation-localization purposes ECD, however, outperformed CID thanks to the better fragmentation, higher mass accuracy of the FTICR and the lower background noise [92]. Next to the requirement of an (expensive) FTICR in-

strument, ECD still suffers from the low efficiency for peptide fragmentation of only 20-50% [74] and the need for 80 times more ions than CID which increases the duty cycle [92].

Electron based dissociation was made available for rf based ion trap mass spectrometers by using a different source of electrons, radical anions, and the fragmentation method is termed ETD [75]. ETD was first implemented on a modified quadrupole linear ion trap [75, 94], but has recently been made available for the high mass accuracy and resolution Orbitrap and Q-TOF instruments [95-97]. There is evidence that the fragmentation mechanism of ETD is similar to ECD, although thermodynamically an extra barrier of an electron leaving the radical anion is present [98, 99]. Cleavage also occurs at the N-Cα bond which generates c- and z-dot ions. In Figure 7B, an ETD spectrum is shown of the same phosphopeptides used to generate the ECD spectrum. The spectra are largely similar with c- an z- ions and no neutral loss. Radical anions generally used are anthracene [75], fluoranthene [100] and azobenzene [101] as they have a low enough electron affinity to pass the electron efficiently on to the precursor peptide ion. ETD has the same advantages as ECD for phosphopeptide sequencing: primarily backbone cleavages and no loss of labile modifications [75], but now with the additional advantage of being available on less cost prohibitive MS instruments. The ion trap is modified to trap both positively charged peptide ions and negatively charged radical anions. This is typically performed by confining the different ion species in separate segments on either side of the ion trap before ion-ion reactions are allowed to occur [75].
Similar to ECD, ETD is more efficient for higher-charged peptides, typically ≥3+ [102-106]. More precisely, the efficiency of ETD is improved by an increasing ratio of charge over number of amino acid residues, *i.e.* increasing peptide charge density [104]. At a lower charge density, which is the case in for example doubly charged tryptic peptides, non-covalent interactions within the peptide might result in electron transfer (ET) without dissociation. To overcome these interactions several approaches have been suggested. One way is to increase the charge state of peptide ions by charging them with for example m-nitrobenzyl alcohol that is added to the LC-MS eluents [99]. This solvent additive improves the electrospray ionization such that the charge density on the ionized peptides is increased. This improved the ETD fragmentation efficiency which was reflected in increased database scores. In another approach, using elevated bath gas temperatures, the sequence coverage of medium and large peptides was shown to increase [107]. It, however, requires modification of the mass spectrometer and increases the detector noise [107, 108]. One other way to improve fragmentation of doubly charged peptides and implemented in many current days ETD instruments is Electron Transfer with Collisionally Activated Dissociation (ETcaD)[103]. After ET, the product ion is dissociated via collisional activation at the *m*/*z* of the $[M+nH]^{n+\bullet}$. As the radical containing product ions are rather labile a lower energy is required for their dissociation than for typical CID. For fragmentation of doubly charged phosphopeptides this additional collision-based fragmentation might be useful. However, it might actually also result in the re-introduction of neutral loss of the phosphate group [103]. In a variation to ETcaD, in charge-reduced CID (CRCID), a charged reduced ET product is isolated and then fragmented to acquire cleaner and easier-to-interpret spectra [109].
Molina *et al*. [102] compared CID and ETD for the analysis of phosphopeptides and identified 70% more Lys-C phosphopeptides by ETD than CID with a 15% overlap. In a second study, however, 50% more non-modified tryptic peptides were identified with CID [110]. Both studies agreed in the advantage of ETD for sequence coverage, approximately 20% more peptide sequence was covered by ETD than CID, and the degree of complementarity, as the overlap of peptides identified with ETD and CID is relatively low [102, 110]. A simi-

larly small overlap was observed by Good *et al.* [104] which triggered them to compare the characteristics of the peptides that were identified with CID or ETD and use these characteristics in a decision-tree-based data-dependent MS approach whereby either CID or ETD was used for sequencing [105]. CID was shown to outperform ETD in the fragmentation of peptides with a charge state of 2+, and for peptides at higher *m/z* [105]. Less than 1% of the peptides identified with ETD were charge state 2+ [104] which is in sharp contrast with Molina *et al.* [110] where approximately half of the identified peptides with ETD were 2+. This is probably related to the different instrumentation and, more importantly, the use of ETcaD by Molina *et al.* [110].

Overall, ETD shows to be highly dependent on the charge state and the charge density [102-106]. As trypsin generally produces 2+ peptides, several alternative proteases have been used in ETD analysis including Lys-C, Glu-C, chymotrypsin and Lys-N [95, 100, 102, 104, 109, 111-113]. As lysine and arginine are similarly abundant, a large proportion of peptides generated by Lys-C would theoretically be 3+ and would therefore be excellent targets for ETD analysis. However, no increase in number of identified phosphopeptides was found between tryptic and Lys-C peptides [102]. This is probably related to the fact that next to the charge state, also the average length of a Lys-C peptide is increased compared to trypsin [100], so the peptide charge density is actually decreased. Furthermore, in practice, Lys-C products might not be so different from trypsin products. Trypsin was shown to cleave less specific close to phosphorylated residues as tryptic phosphopeptides turned out to have approximately four times more missed cleavages than normal peptides from the same digest [102]. Therefore, phosphopeptides already contain on average more than two charges. In the same study, Glu-C, which cleaves C-terminal of glutamic and, to a smaller extent, aspartic acid residues, was also tested for proteolysis as it would potentially generate peptides with a length in between tryptic and Lys-C peptides that might carry more charges [102]. However, the poor efficiency and specificity of the enzyme resulted in a dramatic decrease in the number of identifications. Finally, Lys-N, which is an enzyme that cleaves N-terminal of lysine residues [114], should produce peptides that are similar to Lys-C [112], only the position of the lysine residue would be different. This position of lysine showed to have a major impact on the fragmentation behavior in both CID [115] and ETcaD [112]. In Lys-N peptides, the basicity is concentrated at the N-terminus which affects the protonation after fragmentation. In the case of a doubly charged peptide, ET will charge-reduce the precursor ion such that after fragmentation one proton is available. This proton will be heavily drawn towards

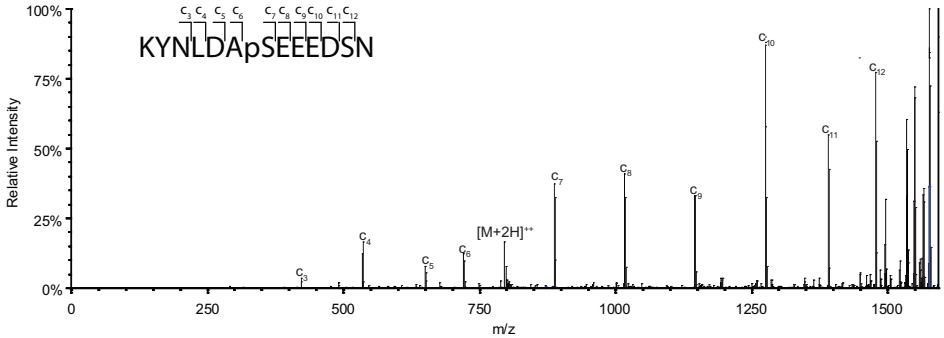


Figure 8 *ETcaD spectrum of a phosphorylated Lys-N peptide. A clear and straightforward fragmentation spectrum allows ladder sequencing of the peptide sequence and unambiguously localizes the phosphorylation at S7.*

the N-terminal fragment. ETcaD of doubly charged Lys-N therefore results in mass spectra that are dominated by N-terminal fragments and are largely simplified. The peptide sequence is thereby very easy to read from the spectrum and as generally almost full ion series are available, the exact site of phosphorylation can be readily determined (Figure 8) [112].

MS INSTRUMENTATION FOR LARGE SCALE PHOSPHORYLATION ANALYSIS

Special attention should be paid to the selection of an appropriate MS platform for large-scale phosphorylation analyses. Here, it is essential that phosphopeptides can be identified and the phosphorylation sites localized in an automated high-throughput, albeit confident, fashion as typically thousands of phosphopeptides are analyzed in parallel. An ideal mass spectrometer for large scale phosphopeptide analysis would (1) enable high resolution and mass accuracy analysis, (2) detect peptides over a wide dynamic range, (3) pose no restraints to the peptide sequence, (4) perform several tandem MS experiments per second, (5) generate little to no neutral loss, and (6) provide enough sequence-informative fragments to identify the phosphopeptides and localize sites of phosphorylation.

Matrix-assisted laser desorption/ionization (MALDI) based MS platforms have been successfully used for phosphopeptide analyses [13, 30, 116-119]. It has been shown that the response of phosphopeptides in MALDI-MS and the extent of neutral loss of the phosphate group during post source decay can be adjusted by changing the matrix and using additives such as phosphoric acid [120, 121]. An advantage of MALDI over ESI is the option to reanalyze samples. However, MALDI generates primarily singly charged peptide ions that thus have a low proton mobility in CID and therefore might show extensive neutral loss of the phosphate group. Furthermore, it takes relatively long to accumulate MALDI-CID spectra and therefore MALDI based MS instruments might not be well suited for high-throughput applications.

Therefore, nowadays most reported high-throughput phosphoproteomics experiments have used ESI-based platforms. ESI typically generates multiply charged peptide ions and especially its higher sensitivity when hyphenated to nano-LC has made ESI the routine ionization technique for phosphopeptide analysis. The drawback encountered by using an ion trap and CID, *i.e.* neutral loss-dominated spectra, might be somewhat overcome by implementing additional fragmentation in MS^3 or MSA approaches. However, as described above this comes at the expense of the number of peptides that can be sequenced as the duty cycle is increased (Figure 9). Tandem in space collision cell type of mass spectrometers such as Q-TOF instruments are nowadays capable of performing up to, theoretically, 10 MS/MS per second. Together with the adequate mass accuracy and resolution in both MS and MS/MS associated with TOF instruments and a reduced neutral loss and phosphate group rearrangement compared to ion trap mass spectrometers, the new generation Q-TOFs might become interesting tools for large-scale phosphoproteomics analyses. ECD and ETD of phosphopeptides do not result in neutral loss-dominated spectra and generate sequence-informative data. However, ECD still requires a long ion accumulation time (Figure 9) and both ETD and ECD are fastidious as to the charge state and charge density of the peptide. The phosphopeptides that are most efficiently fragmented in ECD and ETD have a high charge density. Ironically, these are phosphopeptides that in CID show less neutral loss. Nonetheless, the advent of Orbitrap and Q-TOF ETD instrumentation, allows high mass accuracy and resolution data to be generated. Using the m/z and precursor ion charge states the most efficient fragmentation method for that peptide -CID or ETD- can be selected to optimally sequence the contents of a sample. Thereby, ETD can be used as a complementary rather than a replacement MS/MS method for phosphopeptide analysis to increase the phosphoproteomic coverage.

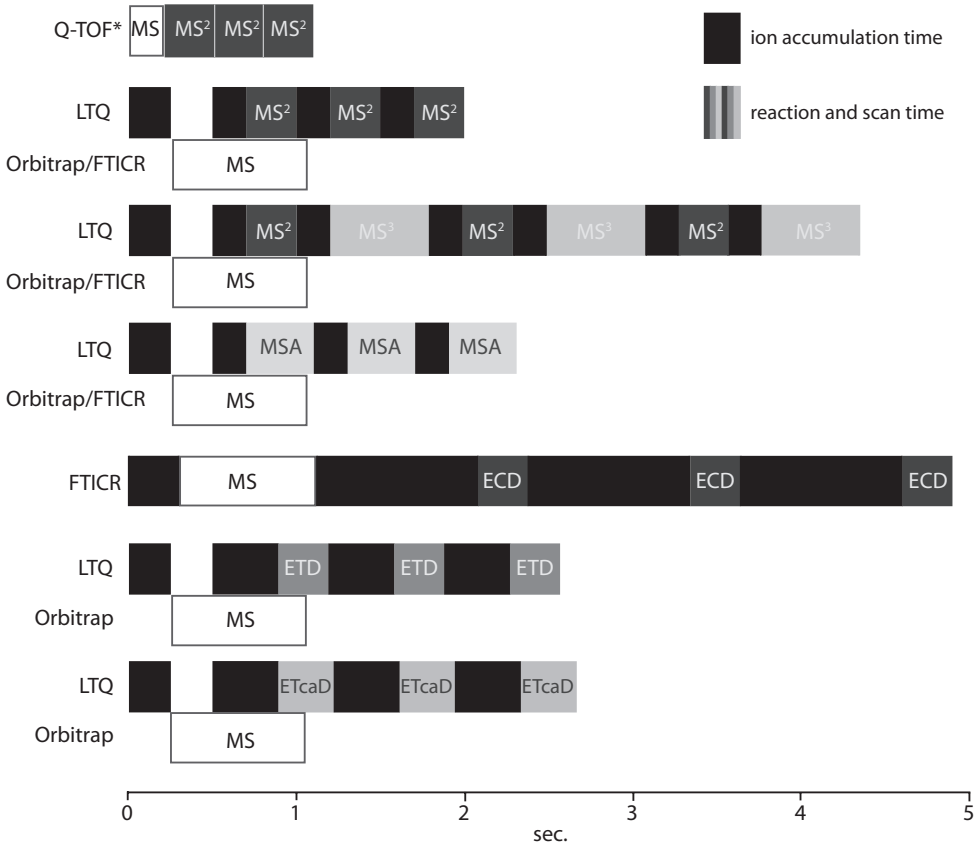


Figure 9. *Time window of ESI-MS approaches in phosphoproteomics. Depicted here are approximations of ion accumulation and reaction and scan times. In reality, these times depend on the amount of ions injected. * Current generation Q-TOFs.*

CONCLUSIONS

In this review, the fragmentation and analysis of phosphorylated peptide ions by mass spectrometry has been addressed. The instability of the phosphate group in MS warrants special attention and adapted MS approaches for the analysis of phosphopeptides. Although many attempts have been made to improve the analysis of phosphopeptides, clearly, a fully flawless MS technique for phosphorylation analysis does not exist (yet). Currently, the choice of mass spectrometer and MS strategy is dependent on whether quantity or quality of phosphopeptide identification is important; whether the ultimate aim is to identify as many phosphosites as possible, to localize phosphorylation sites with the highest confidence possible or to specifically target only a few phosphorylation events. This determines how the sample should be prepared, which protease should be selected, whether phosphopeptide enrichment should be performed or additional dimensions of LC separation should be implemented. And also, it determines which LC-MS approach should be chosen and how much time should be allowed for both LC separation and ion accumulation in the mass spectrometer, and whether additional fragmentation should be performed. Importantly, not only the analytical choices control the outcome of the study. Also, the bioinformatic processing of the MS data determines the results. For example, MS^2 and MS^3 spectra can be combined or searched separately [47]. Furthermore, some database search engines appear to be more

suited for ETD or CID and also for phosphorylation search engines seem to perform differently [48]. Further development of hybrid MS systems such as the Orbitrap-ETD and Q-TOF ETD that combine high-accuracy mass measurements with a choice of CID with or without additional fragmentation and ETD will certainly aid the progression of phosphorylation analysis by MS and may provide the researcher an 'all-in-one' system for both high-throughput and targeted analysis.

ACKNOWLEDGEMENTS

This work was supported by the Netherlands Proteomics Centre (http://www.netherlandsproteomicscentre.nl/), a program embedded in the Netherlands Genomics Initiative. We thank Dr. Helen J. Cooper and Dr. Steve M. M. Sweet from the School of Biosciences, University of Birmingham, Birmingham, UK, for providing ECD spectra. Furthermore, we thank Dr. A.F. Maarten Altelaar, Dr. Javier Munoz and Nadia Taouatas for providing additional fragmentation spectra.

REFERENCES

[1] Manning, G., Whyte, D. B., Martinez, R., Hunter, T., Sudarsanam, S., The Protein Kinase Complement of the Human Genome. *Science* 2002, *298*, 1912-1934.

[2] Hoffert, J. D., Knepper, M. A., Taking aim at shotgun phosphoproteomics. *Anal. Biochem.* 2008, *375*, 1-10.

[3] Paradela, A., Albar, J. P., Advances in the Analysis of Protein Phosphorylation. *J. Proteome Res.* 2008, *7*, 1809-1818.

[4] Reinders, J., Sickmann, A., State-of-the-art in phosphoproteomics. *Proteomics* 2005, *5*, 4052-4061.

[5] Ishihama, Y., Wei, F. Y., Aoshima, K., Sato, T., *et al.*, Enhancement of the efficiency of phosphoproteomic identification by removing phosphates after phosphopeptide enrichment. *J. Proteome Res.* 2007, *6*, 1139-1144.

[6] Marcantonio, M., Trost, M., Courcelles, M., Desjardins, M., Thibault, P., Combined enzymatic and data mining approaches for comprehensive phosphoproteome analyses. *Mol. Cell. Proteomics* 2008, *7*, 645-660.

[7] Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., *et al.*, Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U. S. A.* 2004, *101*, 12130-12135.

[8] Pinkse, M. W. H., Mohammed, S., Gouw, L. W., van Breukelen, B., *et al.*, Highly robust, automated, and sensitive on line TiO2-based phosphoproteomics applied to study endogenous phosphorylation in Drosophila melanogaster. *J. Proteome Res.* 2008, *7*, 687-697.

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

[10] Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., *et al.*, Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat. Biotechnol.* 2002, *20*, 301-305.

[11] Ruse, C. I., McClatchy, D. B., Lu, B., Cociorva, D., *et al.*, Motif-Specific Sampling of Phosphoproteomes. *J. Proteome Res.* 2008, *7*, 2140-2150.

[12] Zhang, X., Ye, J., Jensen, O. N., Roepstorff, P., Highly Efficient Phosphopeptide Enrichment by Calcium Phosphate Precipitation Combined with Subsequent IMAC Enrichment. *Mol. Cell. Proteomics* 2007, *6*, 2032-2042.

[13] Pandey, A., Podtelejnikov, A. V., Blagoev, B., Bustelo, X. R., *et al.*, Analysis of receptor signaling pathways by mass spectrometry: Identification of Vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors. *Proc. Natl. Acad. Sci. U. S. A.* 2000, *97*, 179-184.

[14] Rush, J., Moritz, A., Lee, K. A., Guo, A., *et al.*, Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat. Biotechnol.* 2005, *23*, 94-101.

[15] Villen, J., Gygi, S. P., The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat. Protoc.* 2008, *3*, 1630-1638.

[16] Lemeer, S., Pinkse, M. W. H., Mohammed, S., van Breukelen, B., *et al.*, Online Automated in Vivo Zebrafish Phosphoproteomics: From Large-Scale Analysis Down to a Single Embryo. *J. Proteome Res.* 2008, *7*, 1555-1564.

[17] McNulty, D. E., Annan, R. S., Hydrophilic Interaction Chromatography Reduces the Complexity of the Phosphoproteome and Improves Global Phosphopeptide Isolation and Detection. *Mol. Cell. Proteomics* 2008, *7*, 971-980.

[18] Roepstorff, P., Fohlman, J., Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* 1984, *11*, 601.

[19] Biemann, K., Contributions of mass spectrometry to peptide and protein structure. *Biomed. Environ. Mass Spectrom.* 1988, *16*, 99-111.

[20] Harrison, A. G., Yalcin, T., Proton mobility in protonated amino acids and peptides. *Int. J. Mass Spectrom.* 1997, *165*, 339-347.

[21] Burlet, O., Yang, C. Y., Gaskell, S. J., Influence of Cysteine to Cysteic Acid Oxidation on the Collision-Activated Decomposition of Protonated Peptides - Evidence for Intraionic Interactions. *J. Am. Soc. Mass Spectrom.* 1992, *3*, 337-344.

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

[23] Paizs, B., Suhai, S., Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.* 2005, *24*, 508-548.

[24] Huang, Y. Y., Triscari, J. M., Tseng, G. C., Pasa-Tolic, L., *et al.*, Statistical characterization of the charge state and residue dependence of low-energy CID peptide dissociation patterns. *Anal. Chem.* 2005,

77, 5800-5813.
[25] Huang, Y. Y., Tseng, G. C., Yuan, S. S., Pasa-Tolic, L., *et al.*, A data-mining scheme for identifying peptide structural motifs responsible for different MS/MS fragmentation intensity patterns. *J. Proteome Res.* 2008, *7*, 70-79.
[26] DeGnore, J. P., Qin, J., Fragmentation of phosphopeptides in an ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.* 1998, *9*, 1175-1188.
[27] Tholey, A., Reed, J., Lehmann, W. D., Electrospray tandem mass spectrometric studies of phosphopeptides and phosphopeptide analogues. *J. Mass Spectrom.* 1999, *34*, 117-123.
[28] Reid, G. E., Simpson, R. J., O'Hair, R. A. J., Leaving group and gas phase neighboring group effects in the side chain losses from protonated serine and its derivatives. *J. Am. Soc. Mass Spectrom.* 2000, *11*, 1047-1060.
[29] Palumbo, A. M., Tepe, J. J., Reid, G. E., Mechanistic insights into the multistage gas-phase fragmentation behavior of phosphoserine- and phosphothreonine-containing peptides. *J. Proteome Res.* 2008, *7*, 771-779.
[30] Moyer, S. C., Cotter, R. J., Woods, A. S., Fragmentation of phosphopeptides by atmospheric pressure MALDI and ESI/ion trap mass spectrometry. *J. Am. Soc. Mass Spectrom.* 2002, *13*, 274-283.
[31] Moyer, S. C., VonSeggern, C. E., Cotter, R. J., Fragmentation of cationized phosphotyrosine containing peptides by atmospheric pressure MALDI/ion trap mass spectrometry. *J. Am. Soc. Mass Spectrom.* 2003, *14*, 581-592.
[32] Metzger, S., Hoffmann, R., Studies on the dephosphorylation of phosphotyrosine-containing peptides during post-source decay in matrix-assisted laser desorption/ionization. *J. Mass Spectrom.* 2000, *35*, 1165-1177.
[33] Villén, J., Beausoleil, S. A., Gygi, S. P., Evaluation of the utility of neutral-loss-dependent MS3 strategies in large-scale phosphorylation analysis. *Proteomics* 2008, *8*, 4444-4452.
[34] Schlosser, A., Pipkorn, R., Bossemeyer, D., Lehmann, W. D., Analysis of protein phosphorylation by a combination of elastase digestion and neutral loss tandem mass spectrometry. *Anal. Chem.* 2001, *73*, 170-176.
[35] Hoffman, M. D., Sniatynski, M. J., Rogalski, J. C., Le Blanc, J. C. Y., Kast, J., Multiple neutral loss monitoring (MNM): A multiplexed method for post-translational modification screening. *J. Am. Soc. Mass Spectrom.* 2006, *17*, 307-317.
[36] Lehmann, W. D., Kruger, R., Salek, M., Hung, C. W., *et al.*, Neutral loss-based phosphopeptide recognition: A collection of caveats. *J. Proteome Res.* 2007, *6*, 2866-2873.
[37] Beausoleil, S. A., Villen, J., Gerber, S. A., Rush, J., Gygi, S. P., A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* 2006, *24*, 1285-1292.
[38] Gehrig, P. M., Roschitzki, B., Rutishauser, D., Reiland, S., Schlapbach, R., Phosphorylated serine and threonine residues promote site-specific fragmentation of singly charged, arginine-containing peptide ions. *Rapid Commun. Mass Spectrom.* 2009, *23*, 1435-1445.
[39] Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., *et al.*, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 2006, *127*, 635-648.
[40] Olsen, J. V., Mann, M., Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc. Natl. Acad. Sci. U. S. A.* 2004, *101*, 13417-13422.
[41] Ruttenberg, B. E., Pisitkun, T., Knepper, M. A., Hoffert, J. D., PhosphoScore: An open-source phosphorylation site assignment tool for MSn data. *J. Proteome Res.* 2008, *7*, 3054-3059.
[42] Bailey, C. M., Sweet, S. M. M., Cunningham, D. L., Zeller, M., *et al.*, SLoMo: Automated Site Localization of Modifications from ETD/ECD Mass Spectra. *J. Proteome Res.* 2009, *8*, 1965-1971.
[43] Palumbo, A. M., Reid, G. E., Evaluation of Gas-Phase Rearrangement and Competing Fragmentation Reactions on Protein Phosphorylation Site Assignment Using Collision Induced Dissociation-MS/MS and MS3. *Anal. Chem.* 2008, *80*, 9735-9747.
[44] Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., *et al.*, Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol. Cell. Proteomics* 2005, *4*, 310-327.
[45] Benschop, J. J., Mohammed, S., O'Flaherty, M., Heck, A. J. R., *et al.*, Quantitative Phosphoproteomics of Early Elicitor Signaling in Arabidopsis. *Mol. Cell. Proteomics* 2007, *6*, 1198-1214.
[46] Lemeer, S., Jopling, C., Gouw, J., Mohammed, S., *et al.*, Comparative Phosphoproteomics of Zebrafish Fyn/Yes Morpholino Knockdown Embryos. *Mol. Cell. Proteomics* 2008, *7*, 2176-2187.
[47] Ulintz, P. J., Bodenmiller, B., Andrews, P. C., Aebersold, R., Nesvizhskii, A. I., Investigating MS2/MS3 matching statistics. *Mol. Cell. Proteomics* 2008, *7*, 71-87.
[48] Wan, Y., Cripps, D., Thomas, S., Campbell, P., *et al.*, PhosphoScan: A Probability-Based Method for Phosphorylation Site Prediction Using MS2/MS3 Pair Information. *J. Proteome Res.* 2008, *7*, 2803-2811.

[49] Schroeder, M. J., Shabanowitz, J., Schwartz, J. C., Hunt, D. F., Coon, J. J., A Neutral Loss Activation Method for Improved Phosphopeptide Sequence Analysis by Quadrupole Ion Trap Mass Spectrometry. *Anal. Chem.* 2004, *76*, 3590-3598.
[50] Ulintz, P. J., Yocum, A. K., Bodenmiller, B., Aebersold, R., *et al.*, Comparison of MS2-Only, MSA, and MS2/MS3 Methodologies for Phosphopeptide Identification. *J. Proteome Res.* 2009, *8*, 887-899.
[51] Hoffert, J. D., Wang, G., Pisitkun, T., Shen, R. F., Knepper, M. A., An Automated Platform for Analysis of Phosphoproteomic Datasets: Application to Kidney Collecting Duct Phosphoproteins. *J. Proteome Res.* 2007, *6*, 3501-3508.
[52] Schmidt, A., Gehlenborg, N., Bodenmiller, B., Mueller, L. N., *et al.*, An integrated, directed mass spectrometric approach for in-depth characterization of complex peptide mixtures. *Mol. Cell. Proteomics* 2008, *7*, 2138-2150.
[53] Molloy, M. P., Andrews, P. C., Phosphopeptide derivatization signatures to identify serine and threonine phosphorylated peptides by mass spectrometry. *Anal. Chem.* 2001, *73*, 5387-5394.
[54] Oda, Y., Nagasu, T., Chait, B. T., Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotechnol.* 2001, *19*, 379-382.
[55] Zhou, H., Watts, J. D., Aebersold, R., A systematic approach to the analysis of protein phosphorylation. *Nat. Biotechnol.* 2001, *19*, 375-378.
[56] Thompson, A. J., Hart, S. R., Franz, C., Barnouin, K., *et al.*, Characterization of protein phosphorylation by mass spectrometry using immobilized metal ion affinity chromatography with on-resin beta-elimination and Michael addition. *Anal. Chem.* 2003, *75*, 3232-3243.
[57] Stensballe, A., Andersen, S., Jensen, O. N., Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis. *Proteomics* 2001, *1*, 207-222.
[58] Liao, P. C., Leykam, J., Andrews, P. C., Gage, D. A., Allison, J., An approach to locate phosphorylation sites in a phosphoprotein - mass mapping by combining specific enzymatic degradation with matrix-assisted laser-desorption ionization mass-spectrometry *Anal. Biochem.* 1994, *219*, 9-20.
[59] Steen, H., Jebanathirajah, J. A., Rush, J., Morrice, N., Kirschner, M. W., Phosphorylation analysis by mass spectrometry - Myths, facts, and the consequences for qualitative and quantitative measurements. *Mol. Cell. Proteomics* 2006, *5*, 172-181.
[60] McLachlin, D. T., Chait, B. T., Improved beta-elimination-based affinity purification strategy for enrichment of phosphopeptides. *Anal. Chem.* 2003, *75*, 6826-6836.
[61] Goshe, M. B., Conrads, T. P., Panisko, E. A., Angell, N. H., *et al.*, Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analyses. *Anal. Chem.* 2001, *73*, 2578-2586.
[62] Knight, Z. A., Schilling, B., Row, R. H., Kenski, D. M., *et al.*, Phosphospecific proteolysis for mapping sites of protein phosphorylation. *Nat. Biotechnol.* 2003, *21*, 1047-1054.
[63] Poot, A. J., Ruijter, E., Nuijens, T., Dirksen, E. H. C., *et al.*, Selective enrichment of Ser-/Thr-phosphorylated peptides in the presence of Ser-/Thr-glycosylated peptides. *Proteomics* 2006, *6*, 6394-6399.
[64] Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., *et al.*, Mapping Sites of O-GlcNAc Modification Using Affinity Tags for Serine and Threonine Post-translational Modifications. *Mol. Cell. Proteomics* 2002, *1*, 791-804.
[65] Li, W., Backlund, P. S., Boykins, R. A., Wang, G. Y., Chen, H. C., Susceptibility of the hydroxyl groups in serine and threonine to beta-elimination/Michael addition under commonly used moderately high-temperature conditions. *Anal. Biochem.* 2003, *323*, 94-102.
[66] McLachlin, D. T., Chait, B. T., Analysis of phosphorylated proteins and peptides by mass spectrometry. *Curr. Opin. Chem. Biol.* 2001, *5*, 591-602.
[67] Neubauer, G., Mann, M., Parent ion scans of large molecules. *J. Mass Spectrom.* 1997, *32*, 94-98.
[68] Carr, S. A., Huddleston, M. J., Annan, R. S., Selective Detection and Sequencing of Phosphopeptides at the Femtomole Level by Mass Spectrometry. *Anal. Biochem.* 1996, *239*, 180-192.
[69] Huddleston, M. J., Annan, R. S., Bean, M. F., Carr, S. A., Selective detection of phosphopeptides in complex-mixtures by electrospray liquid-chromatography mass-spectrometry. *J. Am. Soc. Mass Spectrom.* 1993, *4*, 710-717.
[70] Williamson, B. L., Marchese, J., Morrice, N. A., Automated Identification and Quantification of Protein Phosphorylation Sites by LC/MS on a Hybrid Triple Quadrupole Linear Ion Trap Mass Spectrometer. *Mol. Cell. Proteomics* 2006, *5*, 337-346.
[71] Steen, H., Kuster, B., Fernandez, M., Pandey, A., Mann, M., Detection of Tyrosine Phosphorylated Peptides by Precursor Ion Scanning Quadrupole TOF Mass Spectrometry in Positive Ion Mode. *Anal. Chem.* 2001, *73*, 1440-1448.

[72] Salek, M., Alonso, A., Pipkorn, R., Lehmann, W. D., Analysis of Protein Tyrosine Phosphorylation by Nanoelectrospray Ionization High-Resolution Tandem Mass Spectrometry and Tyrosine-Targeted Product Ion Scanning. *Anal. Chem.* 2003, *75*, 2724-2729.
[73] Bruce, C., Shifman, M. A., Miller, P., Gulcicek, E. E., Probabilistic enrichment of phosphopeptides by their mass defect. *Anal. Chem.* 2006, *78*, 4374-4382.
[74] Zubarev, R. A., Electron-capture dissociation tandem mass spectrometry. *Curr. Opin. Biotechnol.* 2004, *15*, 12-16.
[75] Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 2004, *101*, 9528-9533.
[76] Syrstad, E. A., Turecek, F., Toward a general mechanism of electron capture dissociation. *J. Am. Soc. Mass Spectrom.* 2005, *16*, 208-224.
[77] Cooper, H. J., Hudgins, R. R., Håkansson, K., Marshall, A. G., Secondary fragmentation of linear peptides in electron capture dissociation. *Int. J. Mass Spectrom.* 2003, *228*, 723-728.
[78] Turecek, F., N-C-alpha bond dissociation energies and kinetics in amide and peptide radicals. Is the dissociation a non-ergodic process? *J. Am. Chem. Soc.* 2003, *125*, 5954-5963.
[79] Leymarie, N., Costello, C. E., O'Connor, P. B., Electron capture dissociation initiates a free radical reaction cascade. *J. Am. Chem. Soc.* 2003, *125*, 8949-8958.
[80] Cooper, H. J., Håkansson, K., Marshall, A. G., The role of electron capture dissociation in biomolecular analysis. *Mass Spectrom. Rev.* 2005, *24*, 201-222.
[81] Baba, T., Hashimoto, Y., Hasegawa, H., Hirabayashi, A., Waki, I., Electron capture dissociation in a radio frequency ion trap. *Anal. Chem.* 2004, *76*, 4263-4266.
[82] Satake, H., Hasegawa, H., Hirabayashi, A., Hashimoto, Y., Baba, T., Fast multiple electron capture dissociation in a linear radio frequency quadrupole ion trap. *Anal. Chem.* 2007, *79*, 8755-8761.
[83] Ding, L., Brancia, F. L., Electron capture dissociation in a digital ion trap mass spectrometer. *Anal. Chem.* 2006, *78*, 1995-2000.
[84] Silivra, O. A., Kjeldsen, F., Ivonin, I. A., Zubarev, R. A., Electron capture dissociation of polypeptides in a three-dimensional quadrupole ion trap: Implementation and first results. *J. Am. Soc. Mass Spectrom.* 2005, *16*, 22-27.
[85] Tsybin, Y. O., Hakansson, P., Budnik, B. A., Haselmann, K. F., *et al.*, Improved low-energy electron injection systems for high rate electron capture dissociation in Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun. Mass Spectrom.* 2001, *15*, 1849-1854.
[86] Zubarev, R. A., Horn, D. M., Fridriksson, E. K., Kelleher, N. L., *et al.*, Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal. Chem.* 2000, *72*, 563-573.
[87] Zubarev, R. A., Kelleher, N. L., McLafferty, F. W., Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* 1998, *120*, 3265-3266.
[88] Stensballe, A., Jensen, O. N., Olsen, J. V., Haselmann, K. F., Zubarev, R. A., Electron capture dissociation of singly and multiply phosphorylated peptides. *Rapid Commun. Mass Spectrom.* 2000, *14*, 1793-1800.
[89] Shi, S. D. H., Hemling, M. E., Carr, S. A., Horn, D. M., *et al.*, Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry. *Anal. Chem.* 2001, *73*, 19-22.
[90] Sweet, S. M. M., Creese, A. J., Cooper, H. J., Strategy for the identification of sites of phosphorylation in proteins: Neutral loss triggered electron capture dissociation. *Anal. Chem.* 2006, *78*, 7563-7569.
[91] Kweon, H. K., Hakansson, K., Metal oxide-based enrichment combined with gas-phase ion-electron reactions for improved mass spectrometric characterization of protein phosphorylation. *J. Proteome Res.* 2008, *7*, 749-755.
[92] Sweet, S. M. M., Bailey, C. M., Cunningham, D. L., Heath, J. K., Cooper, H. J., Large-scale localization of protein phosphorylation by use of electron capture dissociation mass spectrometry. *Mol. Cell. Proteomics* 2009, *8*, 904-912.
[93] Creese, A. J., Cooper, H. J., The Effect of Phosphorylation on the Electron Capture Dissociation of Peptide Ions. *J. Am. Soc. Mass Spectrom.* 2008, *19*, 1263-1274.
[94] Reid, G. E., Wells, J. M., Badman, E. R., McLuckey, S. A., Performance of a quadrupole ion trap mass spectrometer adapted for ion/ion reaction studies. *Int. J. Mass Spectrom.* 2003, *222*, 243-258.
[95] McAlister, G. C., Berggren, W. T., Griep-Raming, J., Horning, S., *et al.*, A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer. *J. Proteome Res.* 2008, *7*, 3127-3136.
[96] McAlister, G. C., Phanstiel, D., Good, D. M., Berggren, W. T., Coon, J. J., Implementation of electron-transfer dissociation on a hybrid linear ion trap-orbitrap mass spectrometer. *Anal. Chem.* 2007, *79*, 3525-3534.

[97] Xia, Y., Chrisman, P. A., Erickson, D. E., Liu, J., *et al.*, Implementation of Ion/Ion Reactions in a Quadrupole/Time-of-Flight Tandem Mass Spectrometer. *Anal. Chem.* 2006, *78*, 4146-4154.
[98] Gunawardena, H. P., He, M., Chrisman, P. A., Pitteri, S. J., *et al.*, Electron transfer versus proton transfer in gas-phase ion/ion reactions of polyprotonated peptides. *J. Am. Chem. Soc.* 2005, *127*, 12627-12639.
[99] Kjeldsen, F., Giessing, A. M. B., Ingrell, C. R., Jensen, O. N., Peptide sequencing and characterization of post-translational modifications by enhanced ion-charging and liquid chromatography electron-transfer dissociation tandem mass spectrometry. *Anal. Chem.* 2007, *79*, 9243-9252.
[100] Chi, A., Huttenhower, C., Geer, L. Y., Coon, J. J., *et al.*, Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 2007, *104*, 2193-2198.
[101] Xia, Y., Gunawardena, H. P., Erickson, D. E., McLuckey, S. A., Effects of cation charge-site identity and position on electron-transfer dissociation of polypeptide cations. *J. Am. Chem. Soc.* 2007, *129*, 12232-12243.
[102] Molina, H., Horn, D. M., Tang, N., Mathivanan, S., Pandey, A., Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 2007, *104*, 2199-2204.
[103] Swaney, D. L., McAlister, G. C., Wirtala, M., Schwartz, J. C., *et al.*, Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal. Chem.* 2007, *79*, 477-485.
[104] Good, D. M., Wirtala, M., McAlister, G. C., Coon, J. J., Performance characteristics of electron transfer dissociation mass spectrometry. *Mol. Cell. Proteomics* 2007, *6*, 1942-1951.
[105] Swaney, D. L., McAlister, G. C., Coon, J. J., Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat. Methods* 2008, *5*, 959-964.
[106] Van den Toorn, H. W. P., Mohammed, S., Gouw, J. W., Van Breukelen, B., Heck, A. J. R., Targeted SCX based peptide fractionation for optimal sequencing by collision induced, and electron transfer dissociation. *J. Proteomics Bioinform.* 2008, *1*, 379-388.
[107] Pitteri, S. J., Chrisman, P. A., McLuckey, S. A., Electron-transfer ion/ion reactions of doubly protonated peptides: Effect of elevated bath gas temperature. *Anal. Chem.* 2005, *77*, 5662-5669.
[108] Wiesner, J., Premsler, T., Sickmann, A., Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. *Proteomics* 2008, *8*, 4466-4483.
[109] Wu, S. L., Huehmer, A. F. R., Hao, Z. Q., Karger, B. L., On-line LC-MS approach combining collision-induced dissociation (CID), electron-transfer dissociation (ETD), and CID of an isolated charge-reduced species for the trace-level characterization of proteins with post-translational modifications. *J. Proteome Res.* 2007, *6*, 4230-4244.
[110] Molina, H., Matthiesen, R., Kandasamy, K., Pandey, A., Comprehensive comparison of collision induced dissociation and electron transfer dissociation. *Anal. Chem.* 2008, *80*, 4825-4835.
[111] Mohammed, S., Lorenzen, K., Kerkhoven, R., Breukelen, B. v., *et al.*, Multiplexed Proteomics Mapping of Yeast RNA Polymerase II and III Allows Near-Complete Sequence Coverage and Reveals Several Novel Phosphorylation Sites. *Anal. Chem.* 2008, *80*, 3584-3592.
[112] Taouatas, N., Drugan, M. M., Heck, A. J. R., Mohammed, S., Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase. *Nat. Methods* 2008, *5*, 405-407.
[113] Taouatas, N., Altelaar, A. F. M., Drugan, M. M., Helbig, A. O., *et al.*, Strong Cation Exchange-based Fractionation of Lys-N-generated Peptides Facilitates the Targeted Analysis of Post-translational Modifications. *Mol. Cell. Proteomics* 2009, *8*, 190-200.
[114] Nonaka, T., Hashimoto, Y., Takio, K., Kinetic characterization of lysine-specific metalloendopeptidases from Grifola frondosa and Pleurotus ostreatus fruiting bodies. *J. Biochem.* 1998, *124*, 157-162.
[115] Boersema, P. J., Taouatas, N., Altelaar, A. F. M., Gouw, J. W., *et al.*, Straightforward and de novo peptide sequencing by MALDI-MS/MS using a Lys-N metalloendopeptidase. *Mol. Cell. Proteomics* 2009, *8*, 650-660.
[116] Qin, J., Chait, B. T., Identification and characterization of posttranslational modifications of proteins by MALDI ion trap mass spectrometry. *Anal. Chem.* 1997, *69*, 4002-4009.
[117] Annan, R. S., Carr, S. A., Phosphopeptide analysis by matrix-assisted laser desorption time-of-flight mass spectrometry. *Anal. Chem.* 1996, *68*, 3413-3421.
[118] Hoffman, M. D., Rogalski, J. C., Sniatynski, M. J., Locke, J., Kast, J., A multiplexed post-translational modification monitoring approach on a matrix-assisted laser desorption/ionization time-of-flight/time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* 2007, *21*, 2147-2156.
[119] Zhang, X. L., Herring, C. J., Romano, P. R., Szczepanowska, J., *et al.*, Identification of phosphoryla-

tion sites in proteins separated by polyacrylamide gel electrophoresis. *Anal. Chem.* 1998, *70*, 2050-2059.
[120] Kjellstrom, S., Jensen, O. N., Phosphoric acid as a matrix additive for MALDI MS analysis of phosphopeptides and phosphoproteins. *Anal. Chem.* 2004, *76*, 5109-5117.
[121] Stensballe, A., Jensen, O. N., Phosphoric acid enhances the performance of Fe(III) affinity chromatography and matrix-assisted laser desorption/ionization tandem mass spectrometry for recovery, detection and sequencing of phosphopeptides. *Rapid Commun. Mass Spectrom.* 2004, *18*, 1721-1730.

Chapter 7:
In depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immuno-affinity purification and stable isotope dimethyl labeling

Paul J. Boersema[1,2,*], Leong Yan Foong[3,*], Vanessa M.Y. Ding[3,*], Simone Lemeer[1,2,*,†], Bas van Breukelen[1,2], Robin Philp[3], Jos Boekhorst[4], Berend Snel[4], Jeroen den Hertog[5,6], Andre B. H. Choo[3,7#] and Albert J. R. Heck[1,2,8#]

[1] 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; [2] Netherlands Proteomics Centre; [3] Bioprocessing Technology Institute, A*STAR (Agency for Science, Technology and Research), 20 Biopolis Way #06 - 01 Centros, Singapore, 138668, Singapore; [4] Bioinformatics, Department of Biology, Faculty of Science, Utecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands; [5] Hubrecht Institute – KNAW & University Medical Center Utrecht, the Netherlands; [6] Institute of Biology Leiden, Leiden University, the Netherlands; [7] Division of Bioengineering, Faculty of Engineering, National University of Singapore, Singapore; [8] Centre for Biomedical Genetics, the Netherlands

* These authors contributed equally to this work and should be considered as joint first authors

Based on *Mol. Cell. Proteomics* 2009, accepted

SUMMARY

Several mass spectrometry based assays have emerged for the quantitative profiling of cellular tyrosine phosphorylation. Ideally, these methods should reveal the exact sites of tyrosine phosphorylation, be quantitative and not cost-prohibitive. The latter is often an issue as typically several milligrams of (stable isotope labeled) starting protein material is required to enable the detection of low abundant phosphotyrosine peptides. Here, we adopted and refined a peptide centric immuno-affinity purification approach for the quantitative analysis of tyrosine phosphorylation by combining it with a cost-effective stable isotope dimethyl labeling method.

We were able to identify by mass spectrometry, using just two LC-MS/MS runs, more than 1100 unique non-redundant phosphopeptides in HeLa cells from about 4 mg of starting material without requiring any further affinity enrichment as close to 80% of the identified peptides were tyrosine phosphorylated peptides. Stable isotope dimethyl labeling could be incorporated prior to the immuno-affinity purification, even for the used large quantities (mg) of peptide material, enabling the quantification of differences in tyrosine phosphorylation upon pervanadate treatment or EGF stimulation. Analysis of the EGF-stimulated HeLa cells, a frequently used model system for tyrosine phosphorylation, resulted in the quantification of 73 regulated unique phosphotyrosine peptides. The quantitative data was found to be exceptionally consistent with the literature, evidencing that such a targeted quantitative phosphoproteomics approach can provide reproducible results. In general, the combination of immuno-affinity purification of tyrosine phosphorylated peptides with large scale stable isotope dimethyl labeling provides a cost-effective approach that can alleviate variation in sample preparation and analysis as samples can be combined early on. Using this approach a rather complete qualitative and quantitative picture of tyrosine phosphorylation signaling events can be generated.

INTRODUCTION

Reversible tyrosine phosphorylation plays an important role in numerous cellular processes like growth, differentiation and migration. Phosphotyrosine signaling is tightly controlled by the balanced action of protein tyrosine kinases (PTKs) and protein tyrosine phosphatases. Aberrant tyrosine phosphorylation has been suggested to be an underlying course in multiple cancers [1]. Therefore, the identification of tyrosine phosphorylated proteins and the investigation into their involvement in signaling pathways is important. Several groups have attempted to comprehensively study tyrosine phosphorylation by proteomics means [2-5]. However, large scale identification of tyrosine phosphorylation sites by mass spectrometry (MS) can be hindered by the low abundance of tyrosine phosphorylated proteins. Especially, signaling intermediates are usually low abundant proteins that show substoichiometric phosphorylation levels. In addition, the identification by mass spectrometry of phosphopeptides from a complex cellular lysate digest is often complicated by ion suppression effects due to a high background of non-phosphorylated peptides. Enrichment of tyrosine phosphorylated proteins or peptides prior to mass spectrometric detection is therefore essential. Traditionally, antibodies against phosphorylated tyrosine have been used to immunoprecipitate tyrosine phosphorylated proteins from cultured cells [2-4, 6-8]. This phosphoprotein immuno-affinity purification method has for example been used to study the global dynamics of phosphotyrosine signaling events after EGF stimulation using stable isotope labeling with amino acids in cell culture (SILAC)[2]. This approach led to the identification of known and previously unidentified signaling proteins in the EGF receptor (EGFR) pathway, including their temporal activation profile after stimulation of the EGFR, providing crucial information for modeling signaling events in the cell. However, as the identification and quantification of these phosphorylated proteins in these studies was not necessarily based on tyrosine phosphorylated peptides, but largely on non-phosphorylated peptides, little information is derived on the exact site(s) of tyrosine phosphorylation. Also, binding partners of tyrosine phosphorylated proteins, that itself are not tyrosine phosphorylated, might be co-precipitated and impair the tyrosine phosphorylation quantification. Immuno-affinity purification of phosphotyrosine peptides -rather than proteins- using anti-phosphotyrosine antibodies [5, 9-16] significantly facilitates the identification of the site(s) of phosphorylation as it greatly alleviates most of the above mentioned problems since the tyrosine phosphorylated site can be directly identified and quantified.

Accurate MS-based quantification is typically performed by stable isotope labeling. The isotopes can be incorporated metabolically during cell culture as in stable isotope labeling by amino acids in cell culture (SILAC)[17] or chemically as in isobaric tag for relative and absolute quantitation (iTRAQ)[18] or stable isotope dimethyl labeling [19-21]. Typically, most precise quantification can be obtained by metabolic labeling as the different samples can be combined at the level of intact cells [22]. However, metabolic labeling is somewhat limited to biological systems that can be grown in culture and the medium may have an effect on the growth and development of the cells. iTRAQ has been used in conjunction with phosphotyrosine peptide immunoprecipitation [5]. As the chemical labeling is performed before immunoprecipitation, the differentially labeled samples can be precipitated together, thereby neutralizing the potentially largest source of variation. However, as this phosphotyrosine peptide immunoprecipitation is typically performed on several hundreds of micrograms to milligrams of protein sample, iTRAQ provides in these cases a rather cost prohibitive means.

Here, we present an optimized immuno-affinity purification approach for the analysis of tyrosine phosphorylation and combine it with stable isotope dimethyl labeling [19-21, 23].

We efficiently enriched and identified by MS 1112 unique phosphopeptides derived from 4 mg of starting protein material without any further affinity chromatographic enrichment, whereby up to 80% of the peptides analyzed in the final LC run were phosphotyrosine peptides. We further advanced the method by introducing triplex stable isotope dimethyl labeling prior to immunoprecipitation. We quantified differences in tyrosine phosphorylation upon pervanadate treatment or EGF stimulation to detect site-specific changes in tyrosine phosphorylation. 128 unique phosphotyrosine peptides were identified and quantified upon pervanadate treatment. By employing an internal standard comprising of both mock and pervanadate treated sample, we could more confidently identify and quantify phosphorylation sites that are strongly regulated and on-off situations. Analysis of EGF-stimulated HeLa cells resulted in the quantification of 73 unique phosphotyrosine peptides. Most of the upregulated phosphotyrosine peptides that were identified have previously been reported to be involved in the EGFR signaling pathway, validating our approach. However, for the first time we found TFG to become also highly tyrosine phosphorylated upon EGF stimulation together with some tyrosine phosphorylation sites on for example IRS2, SgK269 and DLG3 that have not been firmly established earlier to be involved in EGFR signaling.
In general, we show that the combination of immuno-affinity purification of tyrosine phosphorylated peptides with large scale chemical stable isotope dimethyl labeling provides a cost-effective approach that can alleviate variation in immunoprecipitation and LC-MS as samples can be combined before immunoprecipitation and the necessity of performing additional enrichment is removed by an optimization of the protocol. With only a single LC-MS run, already a rather complete qualitative and quantitative picture of a signaling event can be generated.

MATERIALS AND METHODS

Cell culture, stimulation, digest preparation

HeLa cells were grown to confluence in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum (Invitrogen) and 0.05 mg/ml penicillin/streptomycin (Invitrogen). Cells were placed in serum-free medium 16 h before EGF stimulation. Cells were stimulated with 150 ng/ml EGF for 10 or 30 min, 1mM Pervanadate (prepared by incubating 1 mM orthovanadate with 1 mM hydrogen peroxide for several min) for 10 min or left untreated. Cells were washed with cold phosphate buffered saline and lysed. Before labeling and immunoprecipitation, cells were lysed in 8 M Urea/ 50 mM ammoniumbicarbonate, 5 mM sodium phosphate, 1 mM potassium fluoride, 1 mM sodium orthovanadate and EDTA-free protease inhibitor cocktail (Sigma). Samples were reduced with DTT at a final concentration of 10 mM at 56 °C, subsequently samples were alkylated with iodoacetamide at a final concentration of 55 mM at RT. The samples were diluted to 2 M Urea/50 mM ammoniumbicarbonate and trypsin (1:100; Promega) was added. Digestion was performed overnight at 37 °C.

Stable isotope labeling by reductive amination of tryptic peptides

Tryptic peptides were desalted, dried *in-vacuo* and re-suspended in 100 µL of triethylammonium bicarbonate (100 mM). Subsequently, formaldehyde-H_2 (573 µmol) was added, vortexed for 2 min followed by the addition of freshly prepared sodium cyanoborohydride (278 µmol). The resultant mixture was vortexed for 60 min at RT. A total of 60 µl ammonia (25%) was added to consume the excess formaldehyde. Finally, 50 µl formic acid (100%) was added to acidify the solution. For intermediate labels, formaldehyde-D_2 (573 µmol) was used. For the heavy labeling, ^{13}C-D_2-formaldehyde (573 µmol) and freshly prepared cyanoborodeuteride (278 µmol) was used [20, 21]. The light, intermediate and heavy dimethyl labeled samples were mixed in 1:1:1 ratio based on total peptide amount, determined by running an aliquot of the labeled samples on a regular LC-MS run and comparing overall peptide signal intensities.

Immunoprecipitation

Labeled peptides were mixed, desalted, dried down and re-dissolved in immunoprecipitation (IP) buffer (50mM Tris pH 7.4, 150 mM NaCl, 1% n-octyl-beta-d-glucopyranoside and 1x Complete mini (Roche diagnostics)). Prior to IP, pY99 agarose beads (Santa Cruz) were washed in IP buffer. The labeled peptide mixture was added to the pY99 agarose beads and incubation was performed overnight at 4°C. Beads were washed 3 times with IP buffer and 2 times with water. Peptides were eluted by adding 0.15% TFA for 20 min at RT. Eluted peptides were desalted and concentrated on STAGE-tips.

On-line nanoflow liquid chromatography

Nanoflow LC-MS/MS was performed by coupling an Agilent 1100 HPLC system (Agilent Technologies, Waldbronn, Germany) to a LTQ-Orbitrap mass spectrometer (Thermo Electron, Bremen, Germany) as described previously [24]. Dried fractions were reconstituted in 10 µL 0.1 M acetic acid and delivered to a trap column (Aqua^{tm} C18, 5 µm, (Phenomenex, Torrance, CA, USA); 20 mm × 100 µm ID, packed in-house) at 5 µL/min in 100% solvent A (0.1 M acetic acid in water). Subsequently, peptides were transferred to an analytical column (ReproSil-Pur C18-AQ, 3 µm, Dr. Maisch GmbH, Ammerbuch, Germany; 40 cm × 50 µm ID, packed in-house) at ~100 nL/min in a 2 or 3 hour gradient from 0 to 40% solvent B (0.1 M acetic acid in 8/2 (v/v) acetonitrile/water). The eluent was sprayed via distal coated emitter

tips (New Objective), butt-connected to the analytical column. The mass spectrometer was operated in data dependent mode, automatically switching between MS and MS/MS. Full scan MS spectra (from *m/z* 300-1500) were acquired in the Orbitrap with a resolution of 60,000 at *m/z* 400 after accumulation to target value of 500,000. The three most intense ions at a threshold above 5000 were selected for collision-induced fragmentation in the linear ion trap at a normalized collision energy of 35% after accumulation to a target value of 10,000.

Data analysis

All MS^2 spectra were converted to single DTA files using Bioworks 3.3 using default settings. Runs were searched using an in-house licensed MASCOT search engine (Mascot (version 2.1.0) software platform (Matrix Science, London, UK)) against the Human IPI database version 3.36 (63012 sequences) with carbamidomethyl cysteine as a fixed modification. Light, intermediate and heavy dimethylation of peptide N-termini and lysine residues, oxidized methionine and phosphorylation of tyrosine, serine and threonine were set as variable modifications. Trypsin was specified as the proteolytic enzyme and up to two missed cleavages were allowed. The mass tolerance of the precursor ion was set to 5 ppm and for fragment ions 0.6 Da. Peptides were assigned to the first protein hit reported by Mascot. The assignment of phosphorylation sites of identified phosphopeptides was performed by the PTM scoring algorithm implemented in MSQuant as described previously [25]. Individual MS/MS spectra from phosphopeptides were accepted for a Mascot score ≥20. This threshold was experimentally tested (see Results section) and the FDR at this score for phosphotyrosine peptides only was estimated to be 2% by performing a concatenated decoy database search. All identified phosphopeptides that were found to be differentially phosphorylated were manually validated.

Quantification

Quantification of peptide triplets of which at least one has obtained a Mascot peptide score of 20 was performed using an in-house dimethyl-adapted version of MSQuant (http://msquant.sourceforge.net/), as described previously [20]. Briefly, peptide ratios were obtained by calculating the extracted ion chromatograms (XIC) of the "light", "intermediate" and "heavy" forms of the peptide using the monoisotopic peaks only. The total XIC for each of the peptide forms was obtained by summing the XIC in consecutive MS cycles for the duration of their respective LC-MS peaks in the total ion chromatogram using FT-MS scans. This total XIC was then used to compute the peptide ratio. Heavy and light labeled peptides were found to largely co-elute. Quantified proteins were normalized against the Log_2 of the median of all peptides quantified. StatQuant, an in-house developed program [26], was used for normalization, outlier detection and determination of standard deviation. Ratios of phosphotyrosine levels were normalized to the ratios of (non-specifically binding) non-phosphorylated peptides. Ratios derived from different charge states of the peptide and/or missed cleavages with the same phosphorylation sites were log averaged. Clustering of the phosphotyrosine profiles upon EGF stimulation was performed by K-means clustering (Eucledian distance).

Datasets

The HeLa phosphoproteome was compared to other datasets by mapping protein identifiers and phosphosite locations from Phospho.ELM (version 8.2) and Rikova *et al*. [10] to IPI human version 3.36. Only sites that could be mapped unambiguously to a single IPI identifier were included, resulting in 1405 and 3955 mapped phosphotyrosines, respectively. Overlap

was determined by counting the number of identical sites in the different combinations of datasets.

RESULTS

To first evaluate our protocols and to chart the potential arsenal of phosphorylation sites in HeLa cells, we performed immuno-affinity enrichment of tyrosine phosphorylated peptides from HeLa cell lysate digests in which hyper-phosphorylation was induced via inhibition of phosphatases by pervanadate. To allow for a comprehensive coverage of tyrosine phosphorylation, HeLa cells that were stimulated with pervanadate or mock treated were lysed and digested by trypsin and mixed 1:1, followed by peptide immuno-affinity purification of 4 mg total starting protein material by an antiphosphotyrosine antibody immobilized on agarose beads. The peptides that eluted from the immuno-affinity resin were then analyzed without any further enrichment in two separate LC-MS runs using a two and a three hour LC gradient, respectively. Peptides were identified by matching the fragmentation spectra against the IPI human database (3.36). In Figure 1A, a part of the base peak chromatogram of the 3 hour run is shown to illustrate the efficiency and specificity of the immunoprecipitation. The resolved peaks nearly all represent tyrosine phosphorylated peptides, while the few non-phosphorylated peptides of abundant proteins that were detected are largely masked by these phosphopeptides. In Figure 1B, the apparent specificity of the immunoprecipitation and the number of identified tyrosine phosphorylated peptides with decreasing Mascot threshold score is plotted. By decreasing the Mascot threshold score the number of identified redundant tyrosine phosphorylated peptides increases considerably. However, by decreasing the threshold the chance of including false positives is also increased substantially [27]. This is reflected in the number of identified non-tyrosine phosphorylated peptides which increases exponentially below a score of 20. Consequently, the apparent immunoprecipitation specificity remains above 75% from threshold score 45 to 20 and then suddenly drops. A Mascot score of 20 was therefore taken as a threshold score as this lies above the apparent inflection point. At this score the false discovery rate (FDR) as calculated by performing a concatenated decoy database search is seemingly large, approximately 9%. However, when only tyrosine phosphorylated peptides are taken into account the FDR is only 2%. Tandem mass spectra are available in the PRIDE (Proteomics Identifications) database under accession number 9779. To create a unique, non-redundant phosphotyrosine peptide library, we filtered the identified peptides. Phosphorylation sites that were also identified with a methionine oxidation or a miss-cleavage were considered redundant, but different states of additional phosphorylation on serine, threonine and/or tyrosine were considered unique. This led to the identification of 729 unique phosphotyrosine peptides in the 2 hour run and 970 unique phosphotyrosine peptides in the 3 hour run. The overlap between these two runs was very large (Supplemental Figure 1A[1]), leading to the identification of 1112 unique phosphotyrosine peptides with a total of 983 unique tyrosine phosphorylation sites. A list of these identified phosphopeptides including the sites of phosphorylation and Mascot scores are listed in Supplementary Table 2. This is one of the largest experimental datasets of phosphotyrosine peptides detected in a single experiment without any further TiO_2 or IMAC enrichment. We compared our dataset with phosphotyrosine peptides and sites reported in the recent update of the Phospho.ELM [28] database and with a recently reported large phosphotyrosine peptide dataset of lung cancer cell lines consisting of the phosphotyrosine peptides cumulatively identified from 20 different cell lines and 30 different tissues samples [10]. Approximately 28% of the phosphotyrosine sites that were identified in the current study overlapped with Rikova *et al.*[10] (Supplemental Figure 1B). The Phospho.

1 Supplementary data for this chapter is available at http://www.mcponline.org/cgi/content/full/M900291-MCP200/DC1

ELM database reports whether a phosphorylation site is identified by small scale analysis (low throughput, LTP) or large scale, typically LC-MS based, analysis (high throughput, HTP). The overlap of our dataset with the HTP database is small with 14.3% (Supplemental Figure 1B), but still more than three times larger than the overlap with the LTP database (Supplemental Figure 1C).

The size of the dataset allowed us to do a statistically relevant motif analysis of the residues neighboring the sites of phosphorylation in our dataset to identify conserved sequence motifs that can be recognized by protein tyrosine kinases. Using motif-X [29], the occurrence of motifs from our phosphotyrosine dataset is compared with the occurrence of these motifs in the total human proteome. We found 6 highly significantly enriched motifs that could

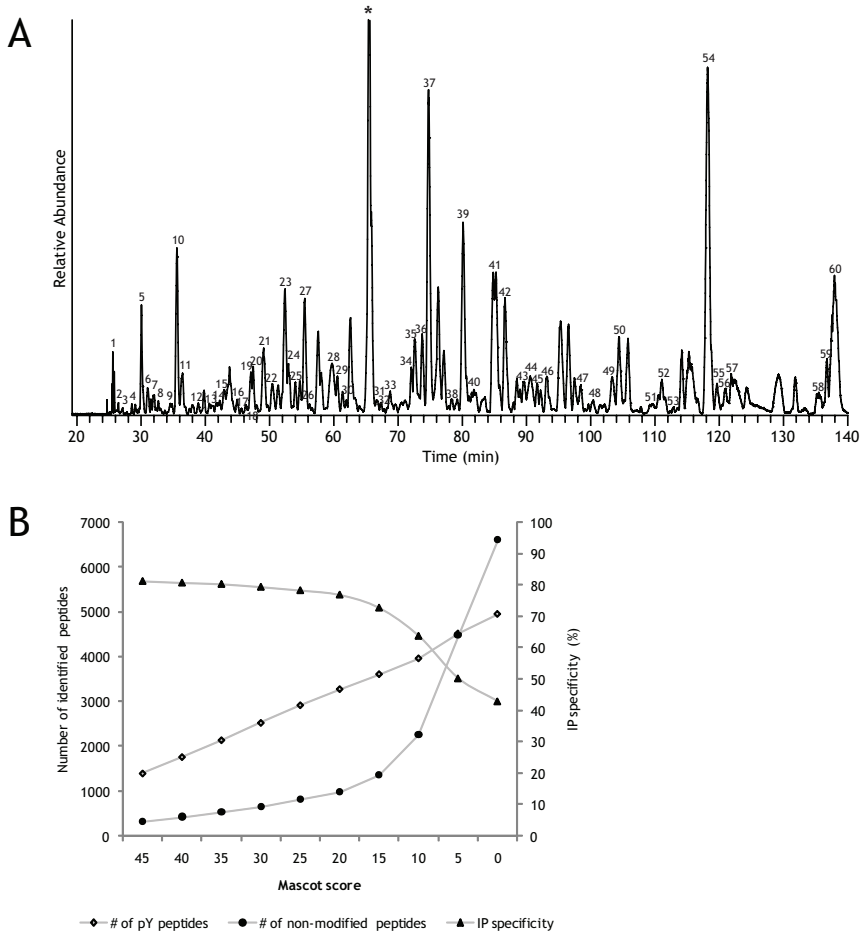


Figure 1. *A) LC-MS base peak chromatogram of the eluate from the phosphotyrosine immunoprecipitated pervanadate treated HeLa digest. The numbers indicate intense peaks representing phosphotyrosine containing peptides. Only a few non-phosphorylated peptides were detected in this eluate. A numbered list of these abundant phosphopeptides is available as Supplemental Table 1. * represents a peak of an ion that could not be identified and most likely is a non-peptide species. Bl) Number of identified phosphotyrosine and non-phosphorylated peptides as a function of the applied Mascot threshold score revealing the apparent specificity of the immunopurification. With decreasing Mascot score threshold the number of non-modified peptides in the dataset increases exponentially below 20. The inflection point at score 20 indicates that at an even lower score, relatively more false positive identifications are introduced in the dataset.*

classify two-thirds of our observed phosphotyrosine sites. These motifs are displayed as WebLogos [30] in Figure 2. The motif with the highest occurrence has a serine residue at the P+1 position. Surprisingly, this motif, has not been previously reported and only one rather promiscuous kinase/phosphatase motif and three also promiscuous binding motifs are known in PhosphoMotif finder[31] with a serine at that position. The second motif contains a proline at the P+3 position. Two motifs show a negatively charged amino acid (aspartic or glutamic acid) at the P+1 position. And two already previously identified motifs as obtained from analysis in mouse brain [11], show a negatively charged amino acid at the P-3 position. To obtain an indication of the potential kinases responsible for phosphorylation of the identified phosphotyrosine peptides we used NetworKIN [32, 33]. For 925 phosphosites (pY, pS and pT) we could identify the potential upstream kinase. The Insulin receptor group kinases was shown to be the upstream kinase for more than 50% of the identified phosphotyrosine peptides that were identified in our dataset, whereby the insulin receptor and insulin-like growth factor 1 receptor contribute equally. Also the EphA receptors, and in particular EphA4, seem to have a high number of substrates represented in our dataset (see Supplemental Figure 2).

Quantitative phosphoproteomics

To allow for quantification of tyrosine phosphorylation, we next set out to incorporate stable isotope dimethyl labeling [19-21] into the phosphotyrosine peptide immunoprecipitation workflow. Therefore, HeLa cells were mock treated or treated with 1 mM pervanadate for 10 min. Cells were lysed and proteins digested. Peptides were labeled with light, intermediate or heavy dimethyl, essentially as described previously [20], but with adopted protocols to allow for the labeling of a few milligrams of sample [21]. Peptides derived from the untreated cells were labeled with light dimethyl, peptides derived from a 1:1 mixture of untreated and pervanadate treated cells were labeled with intermediate dimethyl (as an internal standard) and peptides derived from the pervanadate treated cells were labeled with heavy dimethyl. Labeled peptides originating from 2 mg of protein sample for each of the three samples were mixed in a 1:1:1 ratio, and from this complex peptide mixture tyrosine phosphorylated peptides were enriched by immunoprecipitation (Figure 3). Using this approach we were able to identify and quantify from a single LC-MS run 128 unique tyrosine phosphorylated triplet peptides originating from 99 phosphoproteins. As expected, most of the detected peptides show an abundance profile whereby the ion signal from the non-stimulated phosphopeptide is very low and the signal from the intermediate labeled phosphopeptide is equal to half of the sum of the untreated peptide signal (light) and treated phosphopeptide signal (heavy) (Figure 4A). On average the heavy labeled phosphopeptides were 2.22 (±0.49) times more intense than the intermediate. A list of all quantified phosphopeptides is available as Supplementary Table 3 and tandem mass spectra are available in the PRIDE database under accession number 9780. Interestingly, 15 tyrosine phosphorylated peptides seemed not to be affected by pervanadate treatment as the abundance profile did not significantly change (Table 1 and Figure 4B). A GO term [34] and motif-X [29] analysis was performed to find within this set of unaffected phosphopeptides, and their corresponding proteins, an enrichment of certain biological processes, molecular function or conserved sequences, but the set was probably too small to detect any significant enrichment. However, an enrichment was found of phosphotyrosine sites that fall within a protein family (Pfam) domain [35] (p=2.7e-6, Fisher's exact test). 12 of the 15 unaffected phosphotyrosine sites (80%) fall within a Pfam domain, while only 18% of the remainder of phosphotyrosine sites in the quantified database fall within such globular domains which is close to a previously reported 15-17%

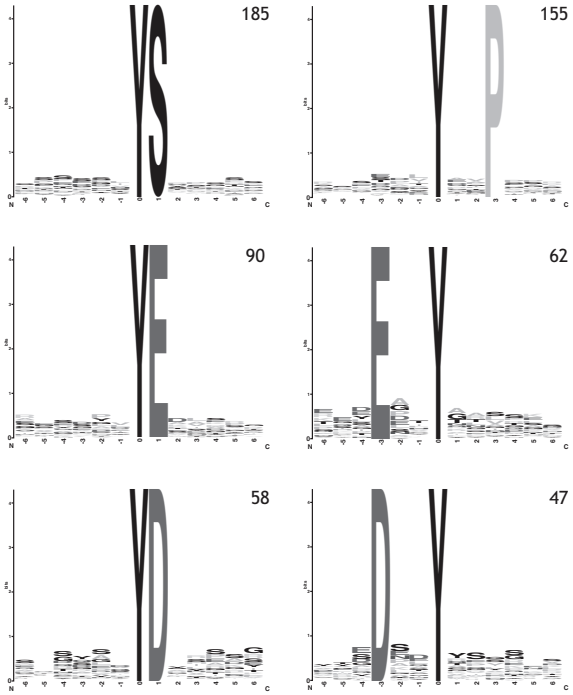


Figure 2. *Motif-X [29] analysis of the here acquired HeLa phosphotyrosine dataset identified six conserved motifs, here displayed as WebLogos [30]. The numbers indicate the number of peptides exhibiting this motif in the full dataset.*

[36]. It remains therefore elusive what is the exact cause of the apparent resistance to phosphatases at these particular phosphotyrosine sites.

With the quantitative method established, we next set out to investigate tyrosine phosphorylation mediated EGFR signaling pathways using stable isotope dimethyl labeling followed by phosphotyrosine immunoprecipitation. In our study, HeLa cells were stimulated with EGF for 0, 10 or 30 min to study the temporal phosphotyrosine signaling pathways after stimulation of the EGF receptor. We used essentially similar methods as described above for the pervanadate treated cells (Figure 3). Briefly, HeLa cells were stimulated with 150 ng/ml EGF for 0, 10 or 30 min. Cells were lysed and proteins were digested. The derived peptide mixtures were stable isotope labeled with light, intermediate or heavy dimethyl labels and mixed in a 1:1:1 ratio after which tyrosine phosphorylated peptides were immunoprecipitated using the antiphosphotyrosine antibody. 73 unique phosphotyrosine peptides, originating from 52 phosphoproteins could be differentially quantified over all three time-points after EGF stimulation in a single LC-MS analysis without any additional affinity enrichment. These associated phosphoproteins including their sites of phosphorylation are listed in Table 2 (For a full table with Mascot and PTM scores, see Supplemental Figure 4) and tandem mass spectra are available in the PRIDE database under accession number 9777. All except one of the phosphotyrosine peptides that were identified, have been reported before and can be found in the PhosphoSitePlus database (www.phosphosite.org)[37]. The observed tem-

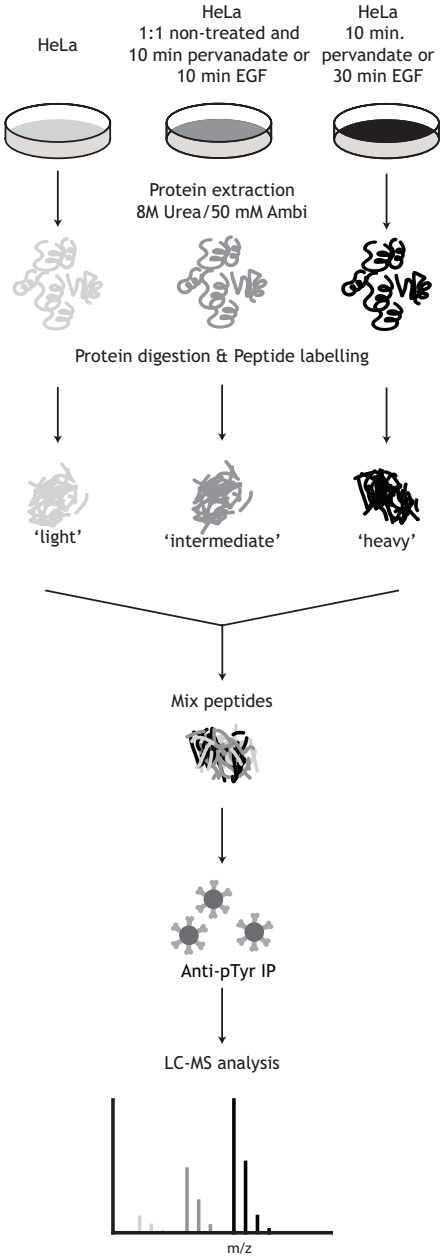


Figure 3. *Experimental scheme for the quantitative phosphotyrosine proteome studies. Cell cultures were mock treated or stimulated with pervanadate or EGF. After lysis and enzymatic digestion, peptides were differentially stable isotope dimethyl labeled and combined before immunoprecipitation with a phosphotyrosine specific antibody. The precipitate was analyzed by LC-MS followed by quantification using the triplet peaks originating from the different isotopes.*

poral ratio profiles largely clustered into three groups (Figure 5A). Cluster 1 consists of sites that show no or only a small change in phosphorylation levels. Cluster 2 consists of tyrosine phosphorylation sites that show an immediate upregulation upon EGF stimulation that remains after 30 min. Cluster 3 comprises of phosphorylation sites that show a large and quick (10 min) increase in tyrosine phosphorylation that already diminishes at 30 min. In the latter cluster autophosphorylation sites of EGFR are found along with established direct EGFR interactors such as Gab1, Cbl and SHC1. In cluster 2, are downstream targets like p38a, stat3 and GSK3-β. Some phosphotyrosine sites were found to be regulated that have not been conclusively established to be in involved in EGFR signaling such as sites on SgK269, IRS2, HNRNPA1 and ATP1A1. The tyrosine phosphorylation on TFG has not been reported before. Finally, in cluster 1, phosphotyrosine peptides were detected from CTTN, ENO1 and WASL that showed an apparent downregulation upon EGF stimulation. Representative examples of tyrosine phosphorylated peptides from the three clusters and their temporal profiles are shown in Figure 5B-D.

Comparing the identified phosphotyrosine peptides from the three separate experiments, evidently, a large overlap can be observed between the more comprehensive dataset of 1112 phosphopeptides in the pervanadate blocked cells and the two other quantitative experiments (Supplemental Figure 3). 86% of the phosphotyrosine peptides identified in the quantitative pervanadate study overlapped with the library, while 74% of the EGF study overlapped with the library set. Also, 51% of the phosphotyrosine peptides identified in the EGF experiment overlapped with those identified in the quantitative pervanadate experiment. Ratios of these phosphopeptides were plotted against each other (Supplemental Figure 4). Some of the sites that showed an increase in tyrosine phosphorylation upon phosphatase inhibition showed no increase upon EGF stimulation. However, tyrosine phospho-

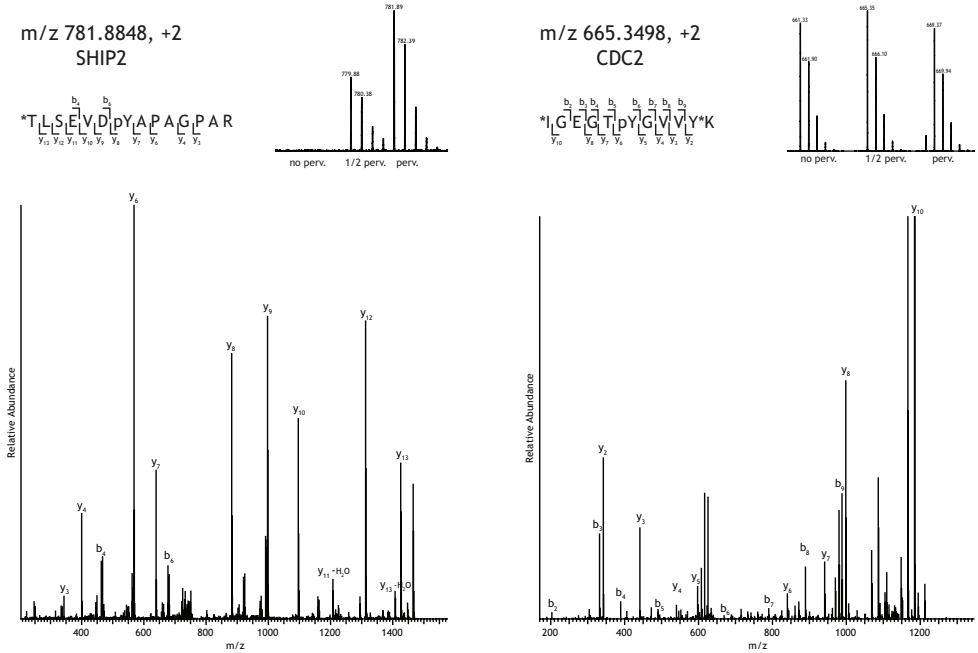


Figure 4. *Representative examples of mass spectra and fragmentation spectra of peptides identified and quantified after pervanadate treatment. A) The abundance of SHIP-2 peptide TLSEVDpYAPAGPAR (fragmentation spectrum shown of m/z 781.8848, +2, heavy dimethyl labeled) is dramatically increased by pervanadate treatment. B) The abundance of peptide IGEGTpYGVVYK (fragmentation spectrum shown of m/z 665.3498, +2, intermediate dimethyl labeled) is not affected by pervanadate treatment. * indicates the site of stable isotope dimethyl labeling.*

IPI acc. nr.	Gene	Peptide	Site 1	Site 2	ratio I/L	STDEV I/L	ratio H/L	STDEV H/L
IPI00016932	SHIP-2	NSFNNPApYpYVLEGVPHQLLPPEPPSPAR	Y986	Y987	1.04	-	2.38	-
IPI00008530	RPLP0	IIQLLDDpYPK	Y24		0.81	-	0.65	-
IPI00000352	DYRK1B	IYQpYIQSR	Y273		0.92	-	1.03	-
IPI00291175	VCL	SFLDSGpYR	Y822		0.93	-	0.73	-
IPI00396485	EEF1A1	EHALLApYTLGVK	Y141		1.02	-	0.93	-
IPI00026689	CDC2	IEKIGEGTpYGVVYK	Y15		1.06	0.05	1.12	0.10
IPI00028570	GSK3B	QLVRGEPNVSpYICSR	Y216		1.10	0.01	1.68	0.15
IPI00654623	TNS3	KLSLGQpYDNDAGGQLPFSK	Y540		1.15	-	1.19	-
IPI00023503	CDK3	VEKIGEGTpYGVVYK	Y43		1.16	-	1.41	-
IPI00026689	CDC2	IGEGpTpYGVVYK	T14	Y15	1.27	0.04	1.49	0.08
IPI00013721	PRPF4B	LCDFGSASHVADNDITPpYLVSR	Y849		1.29	0.08	1.33	0.08
IPI00012885	PTK2	Y(ox)MEDSTpYYKASK	Y576		1.30	-	1.42	-
IPI00008438	RPS10	IAIpYELLFK	Y12		1.41	-	1.70	-
IPI00013981	YES1	LIEDNEpYTAR	Y426		1.47	-	2.22	-
IPI00166578	PRAGMIN	EATQPEPIpYAESTKR	Y411		1.66	0.03	2.44	0.23

Table 1. *Phosphotyrosine peptides whose abundance did not change significantly upon pervanadate treatment of HeLa cells. L, light labeled peptide (mock treated), I, intermediate labeled peptide (1:1 mock and pervanadate treated), H, heavy labeled (pervanadate treated).*

rylation events that showed a large increase upon EGF stimulation also showed a strong increase upon phosphatase inhibition by pervanadate. Interestingly, some of the phosphotyrosine sites that were not affected by pervanadate treatment, such as CDC2 (pY15) and GSK3β (pY216) did increase in abundance upon EGF stimulation.

DISCUSSION

Analysis of tyrosine phosphorylation in cells or tissue is extremely important to understand critical signaling processes involved in processes like development and human disease. In this work, we explored and further optimized an enrichment method for phosphotyrosine peptides based on immuno-affinity purification using phosphotyrosine specific antibodies. Our protocol allowed a very high level of enrichment efficiency providing an eluted fraction dominated by phosphotyrosine peptides, requiring no further IMAC or TiO_2 enrichment prior to LC-MS/MS analysis. From just two of these eluates we were able to identify 1112 unique, non-redundant phosphotyrosine peptides derived from only 4 mg of starting material. The overlap of our HeLa cells phosphotyrosine sites with the dataset reported in Rikova *et al.* [10] is 28%. Notably, the dataset was of Rikova *et al.* was obtained by analyzing several dozens of non-small cell lung cancer (NSCLC) cell lines and NSCLC tumors. The different constitution of their and our cells, together with a different antibody used for phosphotyrosine peptide enrichment, might explain the relative small overlap in tyrosine phosphopeptides, while the greater protein amount and larger number of LC-MS analyses might have allowed them for an even deeper penetration into the tyrosine phosphoproteome. The overlap between LTP and HTP phosphosites in the Phospho.ELM database is surprisingly low indicating that the tyrosine phosphoproteome has not yet been fully mapped [28]. Not unexpectedly, the overlap between our dataset and the HTP dataset is larger than with the smaller LTP dataset. Apparently, a different subset of phosphosites is detected by classical biology means than by LC-MS even though datasets derived by the latter method typically contain hundreds of phosphosites. It is estimated that of the peptides that are identified in a vertebrate cell to be phosphorylated approximately 0.05-2% is phosphorylated on tyrosine residues [25, 38]. If these numbers are taken as true, the number of phosphotyrosine sites identified in this study (over 1000) would suggest that there may be concurrently even up to 1,000,000 serine and threonine phosphorylation sites present as well. Therefore, even with ever improving analytical tools it still remains a great challenge to comprehensively analyze the complete cellular phosphoproteome.

Several stable isotope labeling based quantification methods have been used in combination with phosphoproteomic approaches, including chemical labeling such as iTRAQ [4, 5, 13] and metabolic labeling such as SILAC [2, 39]. The advantage of a chemical modification approach over metabolic labeling is the ability to label samples after cell lysis and digestion. This makes the approach more generically applicable as it also allows the quantitative analysis of biological samples that cannot be grown in culture such as human body fluids or tissue. The benefit of peptide level immunoprecipitation combined with stable isotope labeling is that differently labeled samples can be combined prior to immunoprecipitation, thereby neutralizing the potentially largest source of variation. Furthermore, quantification at the peptide level allows the separate analysis of phosphorylation events on the same protein. Here, we introduced stable isotope dimethyl labeling for quantification of immunoprecipitated phosphotyrosine peptides. As the starting material for these immunoprecipitation typically is several milligrams of protein, the use of iTRAQ labeling can be cost prohibitive, while stable isotope dimethyl labeling is performed with inexpensive generic reagents and

IPI acc. nr.	Gene	Peptide	Site 1	Site 2	log2 ratio 10/0min	stdev	log2 ratio 30/0min	stdev
IPI00011676	N-WASP	VIpYDFIEK	Y256		-1.49	0.66	-1.86	0.74
IPI00465248	ENO1	AAVPSGASTGIpYEALELR	Y44		-0.90	0.25	-0.08	0.25
IPI00029601	CTTN	TQTPPVSPAPQPTEERLPSSPVpYEDAASFK	Y421		-0.73	-	-1.47	-
IPI00220644	PKM2	TATESFASDPILpYRPVAVALDTK	Y105		-0.68	-	-0.13	-
IPI00012885	PTK2	GSIDREDGSLQGPIGNQHIpYQPVGKPDPAAPPK	Y861		-0.52	0.45	0.43	0.09
IPI00216423	ITSN2	REEPEALpYAAVNK	Y967		-0.41	-	-0.79	-
IPI00021439	ACTB	IWHHTFpYNELR	Y91		-0.32	-	0.57	-
IPI00008530	RPLP0	IIQLLDDpYPK	Y24		-0.14	-	-0.16	-
IPI00215949	HIPK2	AVCSTpYLQSR	Y361		-0.13	-	-0.34	-
IPI00013721	PRPF4B	LCDFGSASHVADNDITPpYLVSR	Y849		-0.10	-	0.09	-
IPI00000352	DYRK1B	IYQpYIQSR	Y273		-0.06	-	-0.04	-
IPI00217966	LDHA	QVVESApYEVIK	Y239		0.02	-	0.30	-
IPI00641339	BCAR1	GLPPSNHHAVpYDVPPSVSK	Y306		0.05	-	0.58	-
IPI00552750	TNK2	KPTpYDPVSEDQDPLSSDFKR	Y574		0.08	0.12	-1.30	0.25
IPI00641339	BCAR1	AQQGLpYQVPGPSPQFQSPPAK	Y128		0.16	-	0.77	-
IPI00641339	BCAR1	HLLAPGPQDIpYDVPPVR	Y249		0.20	0.15	0.73	0.22
IPI00301058	VASP	VQIpYHNPTANSFR	Y39		0.21	0.12	0.26	0.06
IPI00654623	TNS3	LSLGQpYDNDAGGQLPFSK	Y540		0.24	-	0.40	-
IPI00021076	PKP4	NNYALNTTATpYAEPYRPIQYR	Y478		0.32	-	0.77	-
IPI00306959	KRT7	LSSARPGGLGSSSLpYGLGASRPR	Y40		0.32	0.10	0.82	0.01
IPI00641339	BCAR1	RPGPGTLpYDVPRER	Y387		0.33	0.15	1.00	0.07
IPI00021267	EPHA2	TYVDPHTpYEDPNQAVLK	Y594		0.35	-	0.69	-
IPI00166578	PRAGMIN	EATQPEPIpYAESTK	Y411		0.39	0.14	0.20	0.23
IPI00291175	VCL	SFLDSGpYR	Y822		0.40	-	0.35	-
IPI00220030	PXN	VGEEEHVpYSFPNK	Y118		0.43	0.19	0.50	0.11
IPI00013981	Src	LIEDNEpYTAR	Y426		0.44	-	0.72	-
IPI00418471	VIM	SLYASSPGGVpYATR	Y61		0.45	-	0.65	-
IPI00220030	PXN	FIHQQPQSSSPVpYGSSAK	Y88		0.47	-	0.50	-
IPI00396485	EEF1A1	STTTGHLIpYK	Y29		0.50	-	0.32	-
IPI00298347	PTPN11	IQNTGDpYYDLYGGEK	Y62		0.50	-	0.75	-
IPI00026689	CDC2	IGEGpTpYGVVYK	T14	Y15	0.58	0.00	1.22	0.00
IPI00026689	CDC2	IGEGTYGVVpYKGR	Y19		0.59	0.15	0.49	0.03
IPI00016932	SHIP-2	ERLpYEWISIDKDEAGAK	Y886		0.63	-	1.55	-
IPI00021267	EPHA2	VLEDDPEATpYTTSGGKIPIR	Y772		0.65	-	1.04	-
IPI00220030	PXN	FIHQQPQSSpSPVpYGSSAK	S85	Y88	0.74	0.00	0.91	0.00
IPI00396485	EEF1A1	EHALLApYTLGVK	Y141		0.80	-	0.61	-
IPI00182469	CTNND1	SLDNNpYSTPNER	Y898		0.90	-	0.33	-
IPI00012885	PTK2	YMEDSTpYYK	Y576		0.91	-	0.71	-
IPI00182469	CTNND1	LNGPQDHSHLLpYSTIPR	Y96		0.98	-	1.03	-
IPI00737545	SGK269	NAIKVPIVINPNApYDNLAIYK	Y635		1.17	-	0.98	-
IPI00022353	TYK2	LLAQAEGEPCpYIR	Y292		1.19	-	1.01	-
IPI00215965	HNRNPA1	SSGPpYGGGGQYFAKPR	Y341		1.22	-	0.45	-
IPI00028570	GSK3B	GEPNVSpYICSR	Y216		1.34	0.74	0.95	0.39
IPI00176903	PTRF	SFTPDHVVpYAR	Y308		1.44	-	1.15	-
IPI00014454	RIN1	EKPAQDPLpYDVPNASGGQAGGPQRPGR	Y36		1.53	0.05	1.81	0.11
IPI00412752	STAT3	YCRPESQEHPEADPGSAAPpYLK	Y705		1.56	-	0.96	-
IPI00029601	CTTN	GPVSGTEPEPVpYSMEAADYR	Y446		1.57	-	1.21	-
IPI00026689	CDC2	IEKIGEGTpYGVVYK	Y15		1.66	-	0.71	-
IPI00334715	GRLF1	NEEENIpYSVPHDSTQGK	Y1105		1.69	0.00	1.64	0.04
IPI00552750	TNK2	VSSTHpYYLLPERPSYLER	Y913		1.87	-	0.89	-
IPI00216423	ITSN2	LIpYLVPEK	Y552		1.93	-	1.26	-
IPI00022462	TFRC	SAFSNLFGGEPLSpYTR	Y20		2.04	-	1.41	-
IPI00737545	SGK269	SSAIRpYQEVWTSSTSPR	Y531		2.04	-	1.63	-
IPI00464978	IRS2	SYKAPYTCGGDSDQpYVLMSSPVGR	Y825		2.16	-	1.61	-

Table 2. *Identified and quantified tyrosine phosphopeptides after EGF stimulation.*

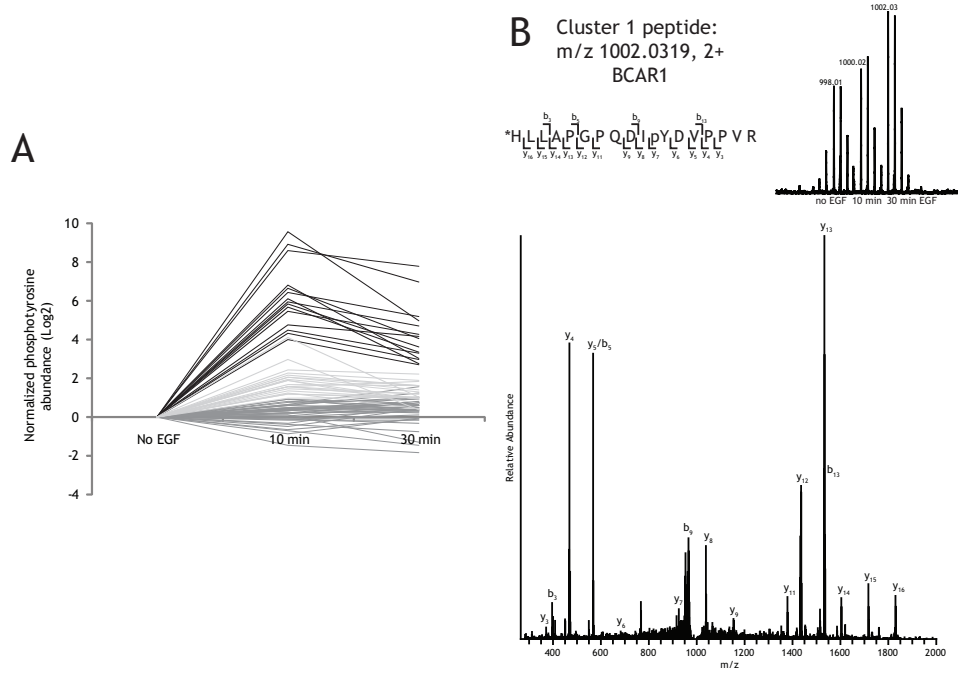

A
Normalized phosphotyrosine abundance (Log2)
No EGF
10 min
30 min
B
Cluster 1 peptide:
m/z 1002.0319, 2+
BCAR1
*HLLLAPGLPQDIpYLDVPLPVR
1002.03
1000.02
998.01
no EGF 10 min 30 min EGF
Relative Abundance
m/z
C
Cluster 2 peptide:
m/z 689.6682, 3+
SGK269
*SSALRpYLQEVLWLTSSLTSPR
689.67
691.01
688.33
no EGF 10 min 30 min EGF
Relative Abundance
m/z
D
Cluster 3 peptide:
m/z 1273.0631, 2+
RBCK1
*N S Q EAEVSCCP F IDN LIpY S C S G*K
1268.85
1273.06
no EGF 10 min EGF 30 min EGF
Relative Abundance
m/z

Figure 5 (previous page). *A) Clustering of tyrosine phosphorylation profiles. Cluster 1 (dark grey): no or only a small change in phosphorylation levels; cluster 2 (light grey): an immediate upregulation that remains after 30 min; cluster 3 (black): strong and quick increase in phosphorylation. B, C, D) Representative examples of mass spectra and fragmentation spectra of peptides identified and quantified after EGF stimulation from each of the clusters. B) BCAR1 peptide HLLAPGPQDIpYDVPPVR (fragmentation spectrum shown of m/z 1002.0319, +2, heavy dimethyl labeled) is from cluster 1 with only a slight increase in phosphorylation upon EGF stimulation. C) SGK269 peptide SSAIRpYQEVWTSSTSPR (fragmentation spectrum shown of m/z 689.6682, +3, intermediate dimethyl labeled) is from cluster 2 with a larger increase after EGF stimulation. D) RBCK1 peptide NSQEAEVSCPFIDNTpYSCSGK (fragmentation spectrum shown of m/z 1273.0631, +2, heavy dimethyl labeled) is from cluster 3 showing an extensive increase in abundance after EGF stimulation. * indicates the site of stable isotope dimethyl labeling.*

thereby does not pose financial restrictions to the amount of sample to be labeled [21]. Importantly, the chemical labeling does not seem to significantly affect the immunoprecipitation process itself.

We were able to identify and quantify 128 unique phosphotyrosine peptides upon pervanadate induction, all detected by their characteristic peptide triplets that can be easily distinguished in the LC-MS spectra. As expected most of the peptide ion signals of the internal standard intensity were equal to half of the sum of the untreated and treated signal intensities, confirming that stable isotope dimethyl labeling does not alter phosphorylation states or impair immunoprecipitation. For most phosphotyrosine peptides, the detected ratios were close to the expected 0:1:2, indicating that tyrosine phosphorylation was largely enhanced upon pervanadate treatment. In Figure 4A, an example of such a phosphotyrosine peptide can be seen. No signal is detected at the m/z where the light labeled, non-pervanadate treated, phosphotyrosine peptide would reside. The intermediate labeled phosphotyrosine peptide from the internal standard is half the intensity of the heavy labeled phosphotyrosine peptide. The internal standard facilitates the detection and quantification of such an on-off situation. Without the internal standard only a single isotope envelope would be visible. The internal standard confirms the on-off situation and validates the identified sequence by the number of lysine residues that can be easily determined based on the mass shift between the intermediate and heavy labeled peptides. Surprisingly, about 15 out of 128 of the quantified peptides showed a close to 1:1:1 ratio in the three samples, indicating that these phosphorylation sites were not affected by pervanadate treatment. Pervanadate is thought to inhibit phosphatases by irreversibly oxidizing the catalytic cysteine [40]. The fact that all of these phosphopeptides were selected by the mass spectrometer for sequencing indicates that these phosphopeptides are relatively abundant. These sites might therefore be constitutively phosphorylated. However, some of the sites showed an increase in phosphorylation upon EGF stimulation and, assuming that the time scale of 10 min is too low for protein synthesis, this suggests that some non-phosphorylated tyrosine residues are phosphorylated upon EGF stimulation. Explanations for the unaltered phosphorylation levels upon pervanadate treatment might involve specific phosphatases that are not inhibited by pervanadate or phosphorylation sites that are not easily accessible for these phosphates, so that inhibition does not lead to increased phosphorylation under the condition used here. The here observed enrichment of phosphotyrosine sites residing in Pfam domains, which are protein domains with a known function and tertiary structure, for this class of peptides could be in agreement with the latter hypothesis. It has been suggested that proteins become typically phosphorylated on sites outside these functional domains [36, 41] as phosphorylation sites on the interspersing variable regions might be more exposed for kinases and phosphatases. We hypothesize that the unaffected phosphorylation sites might therefore be structurally essential or concealed in a non-accessible tertiary structural element.

The EGF receptor plays an important role in a variety of cellular processes and has been studied intensively over the years. The EGFR signaling pathway has been investigated by a variety of (phospho-)proteomic approaches [2, 4, 5, 9, 13, 25, 42, 43]. Upon activation by a growth factor, the PTK activity of EGFR is enhanced [44]. Our quantitative phosphotyrosine peptide immunoprecipitation, therefore, provides an excellent way to study EGFR signaling. For this reason, and as the pathway is very well documented in the literature, we took it here as a model system to further test our new approach. We analyzed HeLa cells that were stimulated for 0, 10 or 30 min by EGF. 73 unique tyrosine phosphorylated triplet peptides were identified and quantified over all time points. Most of the tyrosine phosphopeptides we identified have been reported before also in the context of EGF stimulation. For example, more than 30% of the here identified phosphosites overlap with those identified in Zhang *et al.*[5]. In that report, similar EGF stimulation conditions were used and the effect was also measured after 0, 10 and 30 min. Reassuringly, the observed ratios for these overlapping phosphopeptides for 10 min and 30 min after EGF stimulation correlated very well with those found in our study (correlation coefficients 0.87 and 0.82, respectively, see Figure 6). This is especially remarkable considering different cell lines were used, HeLa *versus* 184A1 parental human mammary epithelial cells, suggesting a high conservation of the EGFR signaling pathway between these two cell lines. Moreover, it reveals that quantitative tyrosine phosphoproteome studies can provide highly reproducible results, even between different laboratories and cell lines used, which is a status not yet readily achieved in global phosphoproteome studies. Most of the identified phosphotyrosine peptides that are increased upon EGF stimulation are known members of the EGF signaling pathway. The EGFR autophosphorylation sites pY1172 and pY1197 showed an increase in phosphorylation after 10 min EGF stimulation whereby this phosphorylation decreased again after 30 min, consistent with previously published data [5, 45]. Furthermore, phosphopeptides of Gab1 (pY659), STAM2 (pY192 and pY371), RBCK1 (pY288), Cbl (pY674), SHIP-2 (pY986 and pY1135) and Epsin (pY17) showed an extensive increase in phosphorylation and have shown to be involved in EGF signaling and internalization [5, 12, 15, 46, 47].

Some of the tyrosine phosphorylated peptides that showed an increase in abundance after EGF stimulation have not been reported before or the phosphosite has not been described profoundly to be involved in EGFR signaling. For example, here, we observed IRS2 pY260 to become more phosporylated upon EGF stimulation [48, 49]. Also, TFG (pY392) shows high tyrosine phosphorylation in response to EGF treatment, but the phosphosite has not been reported before. Recently, this protein has been described to be a substrate of Src [50] like SgK269, a pseudokinase for which we found two phosphosites (pY531 and pY635) to be upregulated [51, 52]. Upregulation of both these phosphotyrosine sites has not been associated with EGF stimulation before, but pY635 was suggested to be involved in Src interaction [52] and was shown to be downregulated upon drug induced EGFR inhibition [9]. With these two proteins we might shed new light on a less well established Src-based part of the EGFR signaling pathway. Next, DLG3 (pY673), which was shown to be downregulated upon drug induced EGFR inhibition [9] was here, in agreement, found to be upregulated upon EGF stimulation.

Finally, only a few tyrosine phosphorylation sites were observed to be significantly downregulated upon EGF stimulation: N-WASP (pY256), ENO1 (pY44) and CTTN (pY421). Interestingly, CTTN has been shown to bind N-WASP. pY421 is one of three residues known to be

phosphorylated by Src while Erk can phosphorylate CTTN at two other serine residues [53, 54]. This Erk and Src phosphorylation of CTTN was suggested to act as a switch on-switch off mechanism respectively for the activation of N-WASP leading to actin polymerization which is important in for example cytokinesis [55]. What exactly the outcome of the down-regulation of these sites is regarding actin polymerization remains to be elucidated.

Our quantitative phosphotyrosine profiling in HeLa cells enables a direct comparison of the response of individual phosphorylation sites upon EGF stimulation and pervanadate treatment. Some sites show a strong increase after pervanadate treatment, but are seemingly unaffected by EGF stimulation, while others are increased in abundance after both EGF stimulation and pervanadate treatment. The existence of the latter category would suggest that in the pervanadate treatment conditions EGFR stimuli exist or kinases are active that target substrates that are also in the EGFR signaling pathway. This is not surprising as the fetal calf serum on which these cells are grown contain hormones and growth factors. This is reflected in the overrepresentation of insulin receptor, insulin-like growth factor receptor and ephrin receptors in the upstream kinases as predicted by NetworKIN in the larger library phosphotyrosine dataset.

In conclusion, immunoprecipitation of tyrosine phosphorylated stable isotope dimethyl labeled peptides allows the quantitative analysis of tyrosine phosphorylation. As isotope labeling is performed after cell lysis and enzymatic digestion, the method is applicable to virtually any sample type, including human tissue. As demonstrated for the EGFR signaling pathway, by these means protein tyrosine kinase signaling pathways can be studied in a relatively quick and inexpensive manner. Several phosphotyrosine sites could be newly identified or further substantiated to be involved in EGFR signaling. By performing quantification on the peptide, rather than protein levels, different phosphorylation events on the same protein can be readily monitored.

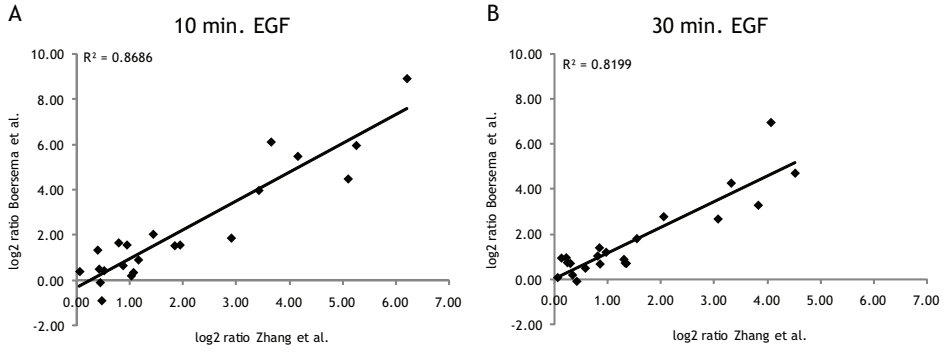


Figure 6. *Ratios of changes in tyrosine phosphorylation A) 10 min and B) 30 min after EGF stimulation detected in the current study plotted against ratios of overlapping phosphotyrosine sites found in Zhang et al.[5].*

Peptide Sequence/phosphosite	pY-site	Zhang et al. (2005) log2 ratio, 10/0 min	Zhang et al. (2005) log2 ratio, 30/0 min	Boersema et al. (2009) log2 ratio, 10/0 min	Boersema et al. (2009) log2 ratio, 30/0 min
AAVPSGASTGIpYEALELR	Y44	0.47	0.42	-0.90	-0.08
LCDFGSASHVADNDITPpYLVSR	Y849	0.44	0.07	-0.10	0.09
EATQPEPIpYAESTKR	Y411	0.06	0.34	0.39	0.20
VGEEEHVpYSFPNK	Y118	0.52	0.59	0.43	0.50
TYVDPHTpYEDPNQAVLK	Y594	1.07	0.86	0.35	0.69
YMEDSTpYYK	Y576	1.16	1.35	0.91	0.71
IGEGTpYGVVYK	Y15	0.79	0.30	1.66	0.71
HLLAPGPQDIpYDVPPVR	Y249	1.03	1.33	0.20	0.73
IQNTGDpYYDLYGGEK	Y62	0.42	0.25	0.50	0.75
VSSTHpYYLLPERPSYLER	Y857	2.91	1.31	1.87	0.89
YCRPESQEHPEADPGSAAPpYLK	Y705	1.94	0.23	1.56	0.96
GEPNVSpYICSR	Y279	0.39	0.14	1.34	0.95
VLEDDPEATpYTTSGGKIPIR	Y772	0.87	0.82	0.65	1.04
GPVSGTEPEPVpYSMEAADYR	Y446	0.95	0.97	1.57	1.21
SAFSNLFGGEPLSpYTR	Y20	1.44	0.85	2.04	1.41
EKPAQDPLpYDVPNASGGQAGGPQRPGR	Y36	1.84	1.55	1.53	1.81
VADPDHDHTGFLTEpYVATR	Y204	3.42	3.07	3.98	2.68
LVNEAPVYSVpYSK	Y374	3.66	2.06	6.12	2.78
NSFNNPApYYVLEGVPHQLLPPEPPpSPARA PVPSATK	Y986/ S1003	5.10	3.83	4.49	3.29
GSHQISLDNPDpYQQDFFPK	Y1172	4.16	3.32	5.49	4.27
ELFDDPSpYVNVQNLDK	Y317	5.26	4.52	5.97	4.71
NSQEAEVSCPFIDNTpYSCSGK	Y288	6.21	4.07	8.93	6.95

Table 3. *Ratios of changes in tyrosine phosphorylation upon EGF stimulation of phosphorylation sites detected in the current study and Zhang et al. [5].*

ACKNOWLEDGEMENTS

This work was supported by the Netherlands Proteomics Centre (http://www.netherlandsproteomicscentre.nl), a program embedded in the Netherlands Genomics Initiative. We thank Dr. Rune Linding for his help with the NetworKIN analysis.

Supplementary data is available free at http://www.mcponline.org/cgi/content/full/M900291-MCP200/DC1

REFERENCES

[1] Blume-Jensen, P., Hunter, T., Oncogenic kinase signalling. *Nature* 2001, *411*, 355-365.
[2] Blagoev, B., Ong, S. E., Kratchmarova, I., Mann, M., Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat Biotechnol* 2004, *22*, 1139-1145.
[3] Steen, H., Kuster, B., Fernandez, M., Pandey, A., Mann, M., Tyrosine phosphorylation mapping of the epidermal growth factor receptor signaling pathway. *J Biol Chem* 2002, *277*, 1031-1039.
[4] Thelemann, A., Petti, F., Griffin, G., Iwata, K., *et al.*, Phosphotyrosine signaling networks in epidermal growth factor receptor overexpressing squamous carcinoma cells. *Mol Cell Proteomics* 2005, *4*, 356-376.
[5] Zhang, Y., Wolf-Yadlin, A., Ross, P. L., Pappin, D. J., *et al.*, Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics* 2005, *4*, 1240-1250.
[6] Pandey, A., Podtelejnikov, A. V., Blagoev, B., Bustelo, X. R., *et al.*, Analysis of receptor signaling pathways by mass spectrometry: Identification of Vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors. *Proc Natl Acad Sci U S A* 2000, *97*, 179-184.
[7] Amanchy, R., Kalume, D. E., Iwahori, A., Zhong, J., Pandey, A., Phosphoproteome analysis of HeLa cells using stable isotope labeling with amino acids in cell culture (SILAC). *J Proteome Res* 2005, *4*, 1661-1671.
[8] Hinsby, A. M., Olsen, J. V., Bennettt, K. L., Mann, M., Signaling initiated by overexpression of the fibroblast growth factor receptor-1 investigated by mass spectrometry. *Mol Cell Proteomics* 2003, *2*, 29-36.
[9] Guo, A., Villen, J., Kornhauser, J., Lee, K. A., *et al.*, Signaling networks assembled by oncogenic EGFR and c-Met. *Proc Natl Acad Sci U S A* 2008, *105*, 692-697.
[10] Rikova, K., Guo, A., Zeng, Q., Possemato, A., *et al.*, Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* 2007, *131*, 1190-1203.
[11] Ballif, B. A., Carey, G. R., Sunyaev, S. R., Gygi, S. P., Large-scale identification and evolution indexing of tyrosine phosphorylation sites from murine brain. *J Proteome Res* 2008, *7*, 311-318.
[12] Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A., White, F. M., Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc Natl Acad Sci U S A* 2007, *104*, 5860-5865.
[13] Wolf-Yadlin, A., Kumar, N., Zhang, Y., Hautaniemi, S., *et al.*, Effects of HER2 overexpression on cell signaling networks governing proliferation and migration. *Mol Syst Biol* 2006, *2*, 15.
[14] Zheng, H., Hu, P., Quinn, D. F., Wang, Y. K., Phosphotyrosine proteomic study of interferon alpha signaling pathway using a combination of immunoprecipitation and immobilized metal affinity chromatography. *Mol Cell Proteomics* 2005, *4*, 721-730.
[15] Tong, J., Taylor, P., Jovceva, E., St-Germain, J. R., *et al.*, Tandem immunoprecipitation of phosphotyrosine-mass spectrometry (TIPY-MS) indicates C19ORF19 becomes tyrosine-phosphorylated and associated with activated epidermal growth factor receptor. *J Proteome Res* 2008, *7*, 1067-1077.
[16] Rush, J., Moritz, A., Lee, K. A., Guo, A., *et al.*, Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat Biotechnol* 2005, *23*, 94-101.
[17] Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., *et al.*, Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol Cell Proteomics* 2002, *1*, 376-386.
[18] Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., *et al.*, Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. *Mol Cell Proteomics* 2004, *3*, 1154-1169.
[19] Hsu, J. L., Huang, S. Y., Chow, N. H., Chen, S. H., Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem* 2003, *75*, 6843-6852.
[20] Boersema, P. J., Aye, T. T., van Veen, T. A. B., Heck, A. J. R., Mohammed, S., Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. *Proteomics* 2008, *8*, 4624-4632.
[21] Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S., Heck, A. J. R., Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protocols* 2009, *4*, 484-494.
[22] Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* 2007, *389*, 1017-1031.
[23] Lemeer, S., Jopling, C., Gouw, J., Mohammed, S., *et al.*, Comparative Phosphoproteomics of Zebrafish Fyn/Yes Morpholino Knockdown Embryos. *Mol Cell Proteomics* 2008, *7*, 2176-2187.
[24] Raijmakers, R., Berkers, C. R., de Jong, A., Ovaa, H., *et al.*, Automated Online Sequential Isotope Labeling for Protein Quantitation Applied to Proteasome Tissue-specific Diversity. *Mol Cell Proteomics* 2008, *7*, 1755-1762.

[25] Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., *et al.*, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 2006, *127*, 635-648.
[26] van Breukelen, B., van den Toorn, H. W. P., Drugan, M. M., Heck, A. J. R., StatQuant: A post quantification analysis toolbox for improving quantitative mass spectrometry. *Bioinformatics* 2009, *25*, 1472-1473.
[27] Käll, L., Storey, J. D., MacCoss, M. J., Noble, W. S., Assigning Significance to Peptides Identified by Tandem Mass Spectrometry Using Decoy Databases. *J Proteome Res* 2008, *7*, 29-34.
[28] Diella, F., Gould, C. M., Chica, C., Via, A., Gibson, T. J., Phospho.ELM: a database of phosphorylation sites update 2008. *Nucl Acids Res* 2008, *36*, D240-244.
[29] Schwartz, D., Gygi, S. P., An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol* 2005, *23*, 1391-1398.
[30] Crooks, G. E., Hon, G., Chandonia, J.-M., Brenner, S. E., WebLogo: A Sequence Logo Generator. *Genome Res* 2004, *14*, 1188-1190.
[31] Amanchy, R., Periaswamy, B., Mathivanan, S., Reddy, R., *et al.*, A curated compendium of phosphorylation motifs. *Nat Biotechnol* 2007, *25*, 285-286.
[32] Linding, R., Jensen, L. J., Ostheimer, G. J., van Vugt, M., *et al.*, Systematic discovery of in vivo phosphorylation networks. *Cell* 2007, *129*, 1415-1426.
[33] Linding, R., Jensen, L. J., Pasculescu, A., Olhovsky, M., *et al.*, NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucl Acids Res* 2008, *36*, D695-D699.
[34] Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., *et al.*, Gene Ontology: tool for the unification of biology. *Nat Genet* 2000, *25*, 25-29.
[35] Sonnhammer, E. L. L., Eddy, S. R., Durbin, R., Pfam: A comprehensive database of protein domain families based on seed alignments. *Proteins* 1997, *28*, 405-420.
[36] Boekhorst, J., van Breukelen, B., Heck, A., Snel, B., Comparative phosphoproteomics reveals evolutionary and functional conservation of phosphorylation across eukaryotes. *Genome Biol* 2008, *9*, R144.
[37] Hornbeck, P. V., Chabra, I., Kornhauser, J. M., Skrzypek, E., Zhang, B., Phosphosite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* 2004, *4*, 1551-1561.
[38] Mann, M., Ong, S.-E., Grønborg, M., Steen, H., *et al.*, Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol* 2002, *20*, 261-268.
[39] Kratchmarova, I., Blagoev, B., Haack-Sorensen, M., Kassem, M., Mann, M., Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation. *Science* 2005, *308*, 1472-1477.
[40] Huyer, G., Liu, S., Kelly, J., Moffat, J., *et al.*, Mechanism of inhibition of protein-tyrosine phosphatases by vanadate and pervanadate. *J Biol Chem* 1997, *272*, 843-851.
[41] Nuhse, T. S., Stensballe, A., Jensen, O. N., Peck, S. C., Phosphoproteomics of the Arabidopsis plasma membrane and a new phosphorylation site database. *Plant Cell* 2004, *16*, 2394 - 2405.
[42] Morandell, S., Stasyk, T., Skvortsov, S., Ascher, S., Huber, L. A., Quantitative proteomics and phosphoproteomics reveal novel insights into complexity and dynamics of the EGFR signaling network. *Proteomics* 2008, *8*, 4383-4401.
[43] Huang, P. H., Mukasa, A., Bonavia, R., Flynn, R. A., *et al.*, Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. *Proc Natl Acad Sci U S A* 2007, *104*, 12867-12872.
[44] Schlessinger, J., Ligand-Induced, Receptor-Mediated Dimerization and Activation of EGF Receptor. *Cell* 2002, *110*, 669-672.
[45] Schulze, W. X., Deng, L., Mann, M., Phosphotyrosine interactome of the ErbB-receptor kinase family. *Mol Syst Biol* 2005, *1*, 13.
[46] Pesesse, X., Dewaste, V., De Smedt, F., Laffargue, M., *et al.*, The Src Homology 2 Domain Containing Inositol 5-Phosphatase SHIP2 Is Recruited to the Epidermal Growth Factor (EGF) Receptor and Dephosphorylates Phosphatidylinositol 3,4,5-Trisphosphate in EGF-stimulated COS-7 Cells. *J Biol Chem* 2001, *276*, 28348-28355.
[47] Prasad, N. K., Decker, S. J., SH2-containing 5'-inositol phosphatase, SHIP2, regulates cytoskeleton organization and ligand-dependent down-regulation of the epidermal growth factor receptor. *J Biol Chem* 2005, *280*, 13129-13136.
[48] Yamauchi, T., Ueki, K., Tobe, K., Tamemoto, H., *et al.*, Tyrosine phosphorylation of the EGF receptor by the kinase Jak2 is induced by growth hormone. *Nature* 1997, *390*, 91-96.
[49] Gogg, S., Smith, U., Epidermal Growth Factor and Transforming Growth Factor alpha Mimic the Effects of Insulin in Human Fat Cells and Augment Downstream Signaling in Insulin Resistance. *J Biol Chem* 2002, *277*, 36045-36051.
[50] Amanchy, R., Zhong, J., Molina, H., Chaerkady, R., *et al.*, Identification of c-Src Tyrosine Kinase Substrates Using Mass Spectrometry and Peptide Microarrays. *J Proteome Res* 2008, *7*, 3900-3910.

[51] Leroy, C., Fialin, C., Sirvent, A., Simon, V., *et al.*, Quantitative Phosphoproteomics Reveals a Cluster of Tyrosine Kinases That Mediates Src Invasive Activity in Advanced Colon Carcinoma Cells. *Cancer Res* 2009, *69*, 2279-2286.
[52] Luo, W., Slebos, R. J., Hill, S., Li, M., *et al.*, Global Impact of Oncogenic Src on a Phosphotyrosine Proteome. *J Proteome Res* 2008, *7*, 3447-3460.
[53] Head, J. A., Jiang, D. Y., Li, M., Zorn, L. J., *et al.*, Cortactin tyrosine phosphorylation requires Rac1 activity and association with the cortical actin cytoskeleton. *Mol Biol Cell* 2003, *14*, 3216-3229.
[54] Nieto-Pelegrin, E., Martinez-Quiles, N., Distinct phosphorylation requirements regulate cortactin activation by TirEPEC and its binding to N-WASP. *Cell Commun Signal* 2009, *7*, 11.
[55] Martinez-Quiles, N., Ho, H.-Y. H., Kirschner, M. W., Ramesh, N., Geha, R. S., Erk/Src Phosphorylation of Cortactin Acts as a Switch On-Switch Off Mechanism That Controls Its Ability To Activate N-WASP. *Mol Cell Biol* 2004, *24*, 5269-5280.

Chapter 8: Quantitative tyrosine phosphorylation profiling in FGF-2 stimulated human embryonic stem cells

Paul J. Boersema[1,2,*], Vanessa M.Y. Ding[3,*], Leong Yan Foong[3,*], Simone Lemeer[1,2,*], Christian Preisinger[1,2], Subaashini Natarajan[3], Lee Dong Yup[3], Geoffrey Koh[3], Jos Boekhorst[4], Berend Snel[4], Andre B.H. Choo[3,5] and Albert J.R. Heck[1,2]

[1] 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; [2] Netherlands Proteomics Centre; [3] Bioprocessing Technology Institute, A*STAR (Agency for Science, Technology and Research), 20 Biopolis Way #06 - 01 Centros, Singapore, 138668, Singapore; [4] Bioinformatics, Department of Biology, Faculty of Science, Utecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands; [5] Division of Bioengineering, Faculty of Engineering, National University of Singapore, Singapore

*These authors contributed equally to this work and should be considered as joint first authors

This is a preliminary report, a manuscript is in preparation for publication

ABSTRACT

FGF signaling appears to be essential in the process of self-renewal of human embryonic stem cells (hESCs). FGF-2 is used for long-term culture of hESC, but the exact role of exogenous FGF-2 is unclear. To elucidate the role of FGF-2 in maintenance of the pluripotency state of hESC, a large-scale targeted phosphoproteomics approach was undertaken to investigate specifically tyrosine phosphorylation events following FGF-2 stimulation. Using an optimized immuno-affinity purification strategy, a cumulative total of 735 unique tyrosine phosphorylation sites were identified. This is the largest inventory to date for hESC. Combining the enrichment approach with a stable isotope dimethyl labeling strategy, a quantitative picture of the early signaling events in FGF-2 stimulated hESC could be generated. A temporal tyrosine phosphorylation profile was generated for 316 unique peptides. Activation of all four FGFRs was found upon FGF-2 stimulation. Furthermore, transactivation of several other receptor tyrosine kinases (RTKs) was observed. Increased tyrosine phosphorylation was confirmed for most of these RTKs using a human phospho-RTK array. Downstream proteins with increased tyrosine phosphorylation include Src family members and their substrate targets. These downstream Src substrates are involved in the regulation of cytoskeletal/actin depending processes that are important in the structural maintenance of the pluripotency state of hESC. Therefore we hypothesize that FGF-2 induces self-renewal partly through the regulation of cytoskeletal and actin depending processes.

INTRODUCTION

Human embryonic stem cells (hESC) are derived from the inner cell mass of the blastocyst [2, 3]. They have the ability to self-renew and are capable of differentiating into the three embryonic germ layers (ectoderm, endoderm, and mesoderm) both *in vitro* and *in vivo* [4, 5]. Importantly, these cells are a powerful tool for drug screening, studying early lineage differentiation *in vitro*, and generating specific cell phenotypes for therapeutic applications. Several pathways have been implicated in hESC self-renewal, such as transforming growth factor-β (TGF-β)/Activin-A/Nodal, Sphingosine-1-phosphate/platelet-dervied growth factor (S1P/PDGF), fibroblast growth factor-2 (FGF-2) (reviewed in [6]) and insulin growth factor (IGF)/insulin [7]. The process of self-renewal appears to be regulated synergistically through various pathways via growth factor or cytokine supplementation. FGF signaling is of central importance to hESC self-renewal. FGF-2 is widely used for long-term culture of hESC and human induced pluripotent stem (iPS) cells in feeder and feeder–free systems [3, 8-10]. Despite its importance in maintaining the undifferentiated phenotype, the exact role of exogenous FGF-2 is still unclear.

FGFs execute their biological actions by binding, dimerizing and activating cell surface fibroblast growth factor receptors (FGFRs) [11, 12]. There are four known human FGFRs with tyrosine kinase activity, namely FGFR1, 2, 3 and 4 which are members of the receptor tyrosine kinase (RTK) family that govern a wide variety of cellular processes from cell motility and differentiation to proliferation. hESC expresses all four FGF-receptors [13-17]. Blocking FGFR signaling in hESC leads to rapid differentiation [13, 17]. This suggests that FGF-mediated signaling is important for embryonic stem cell self-renewal. Following FGF-2 stimulation, activation of the FGF/FGFRs in hESC typically results in signal transduction of the FGF canonical pathways namely, mitogen activated protein kinase (MAPK) and phosphoinositide 3- kinase (PI3-K) pathways [13, 17, 18]. However, the downstream signaling events following FGF-2 stimulation, its link to hESC self-renewal and the maintenance of pluripotency remain to be determined.

Phosphorylation of RTKs, such as FGFRs, elicits a plethora of downstream signaling events upon stimulation which is a highly dynamic process. Reversible protein phosphorylation plays a critical role in regulating signaling pathways and cellular processes. Conventional techniques used to identify these phosphorylation events are limited, for example, by the availability and efficiency of the detection antibodies. Therefore, more recently, large scale proteomics methods have been used to profile phosphorylation in hESC [19-21]. The large scale analysis of phosphorylation, and particularly tyrosine phosphorylation, by LC-MS is challenging because of the low stoichiometry of phosphorylation. Therefore, various techniques have been developed to enrich for phosphopeptides prior to MS analysis, for example using TiO_2 [22] or IMAC [23] for total phosphopeptides. Further specific enrichment of tyrosine phosphorylated peptides by phosphotyrosine-specific antibodies has been used as well [24-26].

To understand the broader implications of FGF signaling in hESC, we have undertaken a large-scale phosphoproteomics approach to investigate the global tyrosine phosphorylation events following FGF-2 stimulation. In this study, using an optimized immuno-affinity purification strategy, a cumulative total of 735 unique tyrosine phosphorylation sites were detected using two biological replicates. Combining this enrichment approach with a stable isotope dimethyl labeling strategy [27-29], a quantitative picture of the early signaling events in FGF-2 stimulated hESC could be generated.

MATERIALS AND METHODS

Cell culture and cell lysis

Human embryonic stem cell line, HES-3 (46 X,X) from ES Cell International (ESI, Singapore) was cultured on Matrigel-coated (Becton, Dickinson and Company, Franklin Lakes, NJ) tissue culture dishes supplemented with conditioned medium (CM) from mitomycin-C treated immortalised mouse embryonic fibroblast (ΔE-MEF)[30]. Medium used for culturing hESC was knockout (KO) medium, which contained 85% KO-DMEM supplemented with 15% KO serum replacer, 1 mM L-glutamine, 1% non-essential amino acids and 0.1 mM 2-mercaptoethanol (All from Invitrogen, Carlsbad, CA). For routine culture, the medium was supplemented with 10 ng/ml of FGF-2 (Invitrogen), and medium was changed daily. The cultures were passaged weekly following enzymatic treatment as previously described [31]. For FGF-2 starved cultures, cells were maintained as mentioned above in the absence of FGF-2 for 5-7 population doublings (PD, 1 PD = ~1 day). Cells were then stimulated with 10 ng/ml of FGF-2 at the indicated time points.

Cells were lysed on ice in 7M urea, 2M thiourea, 4% CHAPS, 40mM Tris, 50 μg/ml DNase, 50 μg/ml RNase, 1mM sodium orthovanadate and 1X *PhosSTOP* (Roche Diagnostics, Switzerland) in the presence of protease inhibitors. Protein concentration was determined using a Bradford Assay. Total protein lysate of 6 mg per time point were reduced with DTT at a final concentration of 10 mM at 56°C. Subsequently, lysates were alkylated with 55 mM iodoacetamide. Lysates were diluted 6-fold with 100 mM ammonium bicarbonate and digested overnight with trypsin.

Stable isotope labeling by reductive amination of tryptic peptides

Tryptic peptides were desalted with Sep-Pak C18 column (Waters, Milford, MA), eluted peptides were lyophilized, and re-suspended in 100 μL of triethylammonium bicarbonate (100 mM). Subsequently, stable isotope dimethyl labeling was performed as described before [26, 27, 29] using formaldehyde-H_2 and cyanoborohydride, formaldehyde-D_2 and cyanoborohydride and formaldehyde-^{13}C-D_2 and cyanoborodeuteride to generate light, intermediate and heavy dimethyl labels, respectively. The light, intermediate and heavy dimethyl labeled samples were mixed in 1:1:1 ratio based on total peptide amount, determined by running an aliquot of the labeled samples on a regular LC-MS run and comparing overall peptide signal intensities. Labeled peptides were mixed, desalted with Sep-Pak C18 column and lypholized.

Immunoprecipitation (IP) of tyrosine phosphorylated peptides

IP was performed as described in Chapter 7 [26]. Peptide mixtures (both non-labeled and stable isotope dimethyl labeled) were dissolved in IP buffer containing 50mM Tris (pH7.4), 150mM NaCl, 1% NOG, and 1x complete mini (Roche Diagnostics). pY-99 antibody beads (Santa Cruz Biotechnology Inc., Santa Cruz, CA) (prewashed three times with IP buffer) were added into each peptide mixture and incubated overnight at 4°C with gentle rotation. After incubation, the beads were then washed three times with 1 ml of IP buffer followed by two times with 1 ml of water, all at 4°C. Peptides were eluted with 0.15% TFA and centrifuged at 1500g for 1 min to separate the antibody beads from the eluate. Eluted peptides were desalted and concentrated on STAGE-tips.

On-line nanoflow LC-MS

Nanoflow LC-MS/MS was performed by coupling an Agilent 1100 HPLC system (Agilent Technologies, Waldbronn, Germany) to an LTQ-Orbitrap mass spectrometer (Thermo Elec-

tron, Bremen, Germany) as described previously. Dried fractions were reconstituted in 10 μL 0.1 M acetic acid and delivered to a trap column (Aqua™ C18, 5 μm, (Phenomenex, Torrance, CA, USA); 20 mm × 100 μm ID, packed in-house) at 5 μL/min in 100% solvent A (0.1 M acetic acid in water). Subsequently, peptides were transferred to an analytical column (ReproSil-Pur C18-AQ, 3 μm, Dr. Maisch GmbH, Ammerbuch, Germany; 40 cm × 50 μm ID, packed in-house) at ~100 nL/min in a 2 hour (non-labeled) or 3 hour (stable isotope dimethyl labeled) gradient from 0 to 40% solvent B (0.1 M acetic acid in 8/2 (v/v) acetonitrile/water). The eluent was sprayed via distal coated emitter tips (New Objective), butt-connected to the analytical column. The mass spectrometer was operated in data dependent mode, automatically switching between MS and MS/MS. Full scan MS spectra (from *m/z* 300-1500) were acquired in the Orbitrap with a resolution of 60,000 at *m/z* 400 after accumulation to target value of 500,000. The three most intense ions at a threshold above 5000 were selected for collision-induced fragmentation in the linear ion trap at normalized collision energy of 35% after accumulation to a target value of 10,000.

Data analysis

All MS^2 spectra were converted to single DTA files using Bioworks 3.3 using default settings. Runs were searched using an in-house licensed MASCOT search engine (Mascot version 2.1.0) software platform (Matrix Science, London, UK) against the Human IPI database version 3.36 (labeled sample; 63012 sequences) or version 3.37 (non-labeled sample; 69164 sequences) with carbamidomethyl cysteine as a fixed modification. Light, intermediate and heavy dimethylation of peptide N-termini and lysine residues (for labeled samples only), oxidized methionine and phosphorylation of tyrosine, serine and threonine were set as variable modifications. Trypsin was specified as the proteolytic enzyme and up to two missed cleavages were allowed. The mass tolerance of the precursor ion was set to 5 ppm and for fragment ions 0.6 Da. Peptides were assigned to the first protein hit reported by Mascot. The assignment of phosphorylation sites of identified phosphopeptides was performed by the PTM scoring algorithm implemented in MSQuant as described previously [32]. Individual MS/MS spectra from phosphopeptides were accepted for a Mascot score ≥ 20. The FDR at this score for pY peptides was found to be less than 1% by performing a concatenated decoy database search. All identified phosphopeptides that were found to be differentially phosphorylated were manually validated.

Quantification

Quantification of peptide triplets of which at least one had obtained a Mascot peptide score of 20 was performed using an in-house stable isotope dimethyl labeling-adapted version of MSQuant (http://msquant.sourceforge.net/), as described previously [26, 28]. Briefly, peptide ratios were obtained by calculating the extracted ion chromatograms (XIC) of the "light", "intermediate" and "heavy" forms of the peptide using the monoisotopic peaks only. The total XIC for each of the peptide forms was obtained by summing the XIC in consecutive MS cycles for the duration of their respective LC-MS peaks in the total ion chromatogram using FT-MS scans. This total XIC was then used to compute the peptide ratio. Quantified proteins were normalized against the Log_2 of the median of all peptides quantified. StatQuant, an in-house developed program [33], was used for normalization, outlier detection and determination of standard deviation. Ratios of phosphotyrosine levels were normalized to the ratios of several of the most abundant (non-specifically binding) non-phosphorylated peptides.

Western Blot and IP of phosphoproteins

Cells were lysed in Cell Lysis Buffer (Cell Signaling Technology, Beverly, MA) supplemented with 1 mM PMSF. Protein concentrations were determined using the DC protein assay (Bio-Rad laboratories Inc., Hercules, CA). Twenty micrograms of each sample was mixed with Laemmli buffer and boiled for 5 minat 95°C. All samples were subjected to SDS-PAGE and electro-transferred onto PDVF membranes (0.2 μm, Bio-Rad). Membranes were probed using the corresponding primary antibodies at the indicated dilutions. After incubation with the primary antibodies, appropriate peroxidase-conjugated secondary antibodies (Dako, Glostrup, Denmark) or fluorescence secondary antibodies (LI-COR Biosciences, Lincoln, NE) were used to detect the bound antibodies. Protein bands were visualized either using a chemiluminescence detection reagent ECL Plus (Amersham, GE Healthcare, UK) or LI-COR ODYSSY imaging system (LI-COR Biosciences). All antibodies used are shown in Supplementary Table 1[1].
For IP of phosphotyrosine (pY) proteins, cells were harvested and precleared with Protein A beads. The precleared lysates were subjected to agarose-conjugated 4G10 antibody (Cell Signaling Technology). After incubating with 4G10 antibody for 2 h at 4°C, anti-pY antibody pY99 (Santa Cruz) was added to the immunoprecipitation for an additional 4 h at 4°C . The beads were then washed three times with lysis buffer, followed by elution of the immunoprecipitated proteins in Laemmli buffer after boiling for 5 min at 95°C. Western blotting was performed as mentioned above.

Human phospho-RTK Array
Analysis of protein expression using a human phospho-RTK antibody array (R&D Systems Inc., Minneapolis, MN) was performed according to manufacturer's instructions. Briefly, capture and control antibodies were spotted in duplicates on nitrocellulose membranes and incubated overnight with 300 μg protein lysate. The membrane was then washed extensively with buffer provided and, further incubated with pan anti-phospho-tyrosine antibody conjugated to horseradish peroxidase (HRP). After incubation, arrays were washed and visualized using chemiluminescence ECL Plus (Amersham) and developed using chemiluminescence.

Dataset Comparison
The hESC phosphoproteome was compared to other datasets by mapping known pY site locations obtained from Phospho.ELM (version 8.2) and Rikova *et. al*. (2007)[1]. Using the IPI human database, 1382 and 4117 unique pY sites were mapped out from phospho.ELM and Rikova *et. al.* (2007) datasets respectively. Overlap was determined by counting the number of identical sites between t he datasets.

1 Supplementary data to this chapter is available at http://tinyurl.com/thesispjb

RESULTS

Effect of FGF-2 stimulation on hESC phosphoproteome

To study the effect of FGF-2 in the maintenance of hESC pluripotency, hESCs were FGF-2 starved for 5-7 doublings, followed by stimulation with 10 ng/mL FGF-2 for 0, 1, 5, 15 and 30 min and 1, 2, 4, 8 and 24 hours. A rapid increase in phosphorylation of both ERK/MAPK (Thr185 and Tyr187) and AKT (Ser473) was observed after FGF-2 stimulation (Figure 1A) suggesting the expected activation of the FGF receptors as ERK/MAPK and AKT are downstream targets of FGFRs. To profile the early effect of the activation of FGFR on the tyrosine phosphorylation of hESC an anti-pY specific antibody was used to selectively enrich for tyrosine phosphorylated peptides from 6 mg of tryptic lysate digest per cell state (0, 1, 5, 15 and 60 min)[26]. The precipitated peptides were then analyzed by LC-MS. Respectively 153, 376, 273, 285 and 287 unique tyrosine phosphorylation sites could be identified from the 0, 1, 5, 15 and 60 min by FGF-2 stimulated cells with cumulatively 597 unique tyrosine phosphorylation sites. In a biological replicate, a total of 574 tyrosine phosphorylation sites could be identified leading to an overall cumulative 735 unique tyrosine phosphorylation sites in the two biological replicates (see Supplemental Table 2). The more than double increase in tyrosine phosphorylation after 1 min of FGF-2 stimulation suggests a prompt activation of several tyrosine phosphorylated signaling pathways. The overlap of identified pY sites between each sample time point is relatively large, typically between 60% and 80%. Also, the 76% overlap between the biological replicates is larger than is observed in typical shotgun LC-MS phosphoproteomics approaches, suggesting that significant reproducibility in phosphoproteomics can be achieved by enrichment of tyrosine phosphorylated peptides using IP, and thus allowing comprehensive profiling of tyrosine phosphorylation in complex cell lysates. Many of the pY sites were classified by Panther [34] to be on proteins that are involved in growth factor signaling, such as the FGF, EGF, VEGF and PDGF signaling pathways.

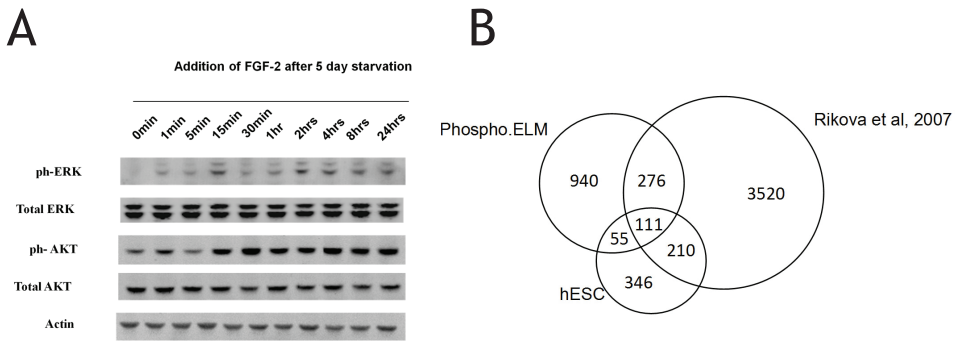


Figure 1. *A) Western blot showing an increase in phosphorylation of both ERK and AKT upon FGF-2 stimulation. B) Overlap of identified phosphotyrosine sites between our dataset (hESC), those deposited in Phospho. ELM (v 8.2), and those identified in Rikova et al. [1].*

Comparison of our dataset with previously reported large datasets of pY site locations obtained in Phospho.ELM (version 8.2) and Rikova et al. [1] yielded an overlap of 44% and 23% respectively (Figure 1B). A motif analysis was performed on the residues of the adjourning sites of phosphorylation in our cumulative hESC dataset. Using Motif-X [35], overrepresented sequence motifs were extracted which might be related to consensus sequences that are recognized by kinases. Employing human IPI database as the background sequences, 7 motif patterns were found to be significantly overrepresented, covering over 53% of our dataset

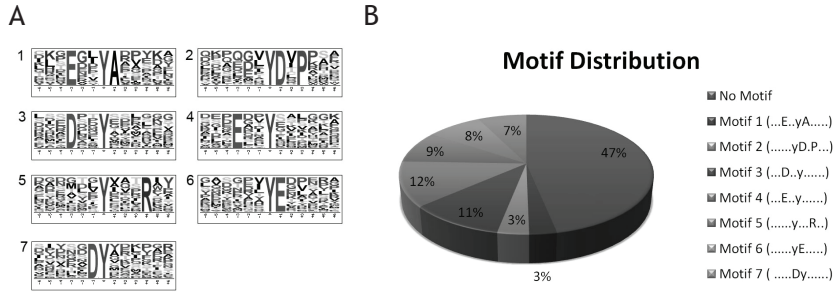


Figure 2. *Tyrosine phosphorylation motifs overrepresented in the current dataset. A) Seven motifs were found to be significantly overrepresented. B) Distribution of different motifs in the dataset.*

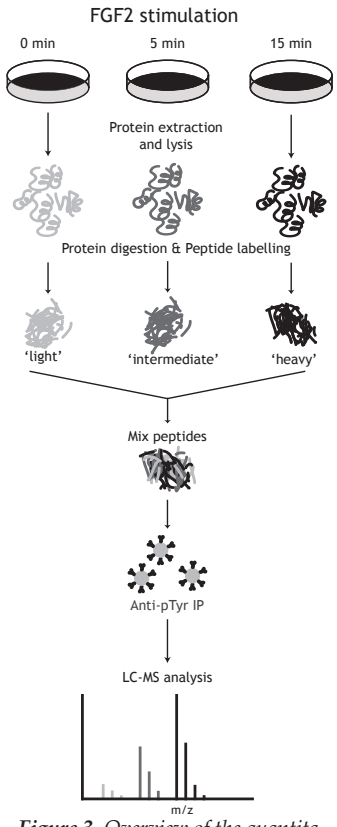


Figure 3. *Overview of the quantitative proteomics workflow. Samples are stimulated for different times with FGF-2 followed by lysis and digestion. The peptides are differentially labeled by stable isotope dimethyl labeling and combined followed by simultaneous enrichment of tyrosine phosphorylated peptides using a phosphotyrosine specific antibody. The enriched fraction is analyzed by LC-MS.*

(Figure 2). In particular, the motifs where Aspartic acid (D) or Glutamic acid (E) occupy position P-3 (Motifs 3 and 4), covered 23% of the phosphopeptide sequences. Although motifs 1 and 2 are only present in approximately 6% of the cumulative databases, they occur more frequently in our dataset as compared to the background sequences. Further comparison of the 7 motif patterns identified in our cumulative dataset to both phospho.ELM and Rikova databases, revealed that motif 1 and motif 5 are overrepresented only in our hESC dataset.

Quantitative phosphoproteomics

For a more precise quantitative profiling of tyrosine phosphorylation, enrichment of tyrosine phosphorylated peptides by immunoprecipitation was combined with stable isotope dimethyl labeling [26-29]. The activation profile of ERK and AKT detected by Western blot (Figure 1A), together with the number of tyrosine phosphorylation sites detected in our tyrosine phosphorylation profiling dataset, indicated that most of the tyrosine phosphorylation events happen within 15 min after FGF-2 stimulation. Therefore, the 0, 5 and 15 min post FGF-2 stimulation time points were chosen for the quantitative profiling study. Figure 3 displays an overview of the quantitative phosphoproteomics strategy used in this study.

Lysate (6 mg) for each time point was digested followed by stable isotope dimethyl labeling whereby the non-stimulated sample (0 min) was labeled with light dimethyl labels, the 5 min time point with intermediate labels, and finally the 15 min time point with heavy dimethyl labels. The differentially labeled samples were mixed 1:1:1 and enriched simultaneously for tyrosine phosphorylated peptides by immunoprecipitation. The enriched tyrosine phosphorylated peptides were then analyzed by LC-MS, using a 3 hour elution gradient. By comparing their respective signal intensities the different FGF-2 stimulated time

points were relatively quantified.

From our analysis, 316 unique tyrosine phosphorylated peptides (light, intermediate, heavy) with 300 unique tyrosine phosphorylation sites could be identified and quantified (see Supplemental Table 3 and 4). The quantitative analysis demonstrated the increase upon FGF-2 stimulation in tyrosine phosphorylation of FGFRs, which includes the autophosphorylation tyrosine phosphorylation sites on FGFR 1, 2, 3, and 4 and some of their canonical downstream effectors, (see Figure 4 for the abundance profiles of several selected tyrosine phosphorylated peptides). Activation of FGF receptors is known to induce MAPK and PI3-K pathways [17, 18]. In the quantitative dataset we do indeed find an increase in tyrosine phosphorylation in the activation loop of MAPK1 and 3, which confirms the initial western blot from Figure 1A as the same phosphorylation sites were detected that increased in intensity. Furthermore, two members of the PI3-K family were found with increases in tyrosine phosphorylation upon FGF-2 stimulation. Additionally, a large number of tyrosine phosphorylated peptides of proteins not directly involved in the canonical FGF pathway were identified with an increase in tyrosine phosphorylation post FGF-2 stimulation.

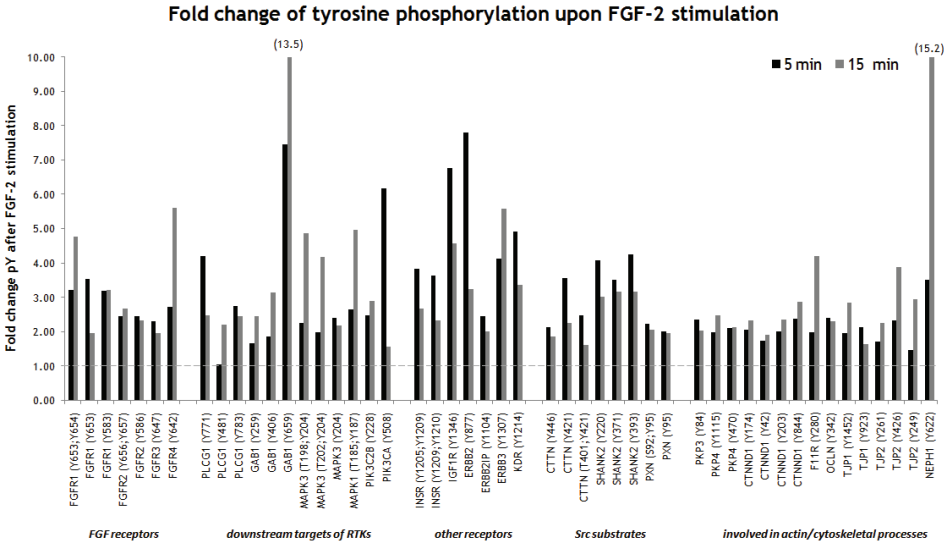
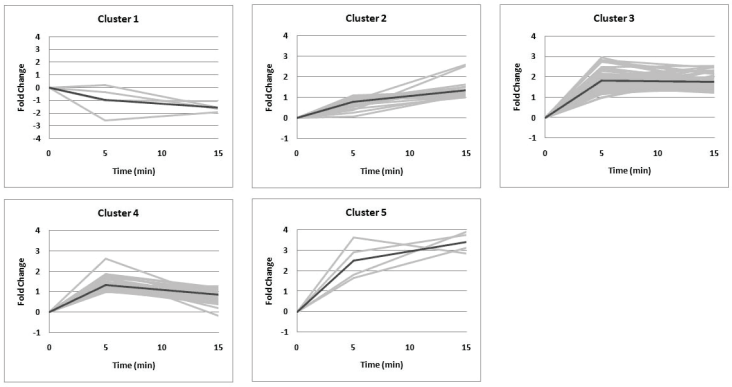


Figure 4. *Quantitative profiles of tyrosine phosphorylation abundance upon FGF-2 stimulation for a selection of proteins.*

To classify the response of these proteins to FGF-2, a cluster analysis was performed based on the temporal tyrosine phosphorylation profiles. Phosphopeptides that showed and increase in tyrosine phosphorylation of at least two times at either of the time points were clustered into 5 different groups, each describing a particular profile pattern (Figure 5). From the clusters, most of the phosphopeptides (approximately 34.5%) showed sustained or transient activity (Clusters 2 and 3) upon stimulation. No significant enrichment within the different clusters was found for the motifs that were found to be overrepresented in hESC as described in Figure 1.

Human phospho-receptor tyrosine kinase array

In the quantitative tyrosine phosphorylation profiling study of the effect of FGF-2 on hESC, several receptors other than FGFR were shown to increase in tyrosine phosphorylation upon



Cluster	Number of Peptides	Description
1	4	Down regulated
2	25	Slow/delayed activation
3	56	Sustained activation
4	53	Transient activation
5	4	Fast activation

Figure 5. *Clustering of tyrosine phosphorylation profiles in response to FGF-2 stimulation.*

FGF-2 stimulation. To confirm the activation of these receptors a human phospho-receptor tyrosine kinase (phospho-RTK) array was used. With the human phospho-RTK array, relative tyrosine phosphorylation of 42 different RTKs can be simultaneously monitored. This includes receptors from the FGF, EGF, VEGF, and insulin receptor families. In the human phospho-RTK array, RTKs are captured by antibodies spotted on a nitrocellulose membrane. The levels of total tyrosine phosphorylation of the RTKs are then determined using a pan phosphotyrosine antibody that is conjugated to horseradish peroxidase which can be detected by chemiluminescence. In the array dataset, an additional time point (60 min) was included to understand the pattern of phosphorylation sustenance of these receptors in the stimulated hESC. A rapid increase in phosphorylation of all FGF receptors was observed (Figure 6). Furthermore, phosphorylation levels increased for Insulin receptor (INSR), Insulin growth factor 1 receptor (IGF1-R), Ephrin type A receptors (EphA) 1 and 2, Vascular endothelial growth factor receptor 2 (VEGF2/KDR).

DISCUSSION

Many cytokines and growth factors have been shown to play a role in maintaining hESC self-renewal. For example, almost all hESC culture platforms require FGF-2 supplementation. In our previous study [17], we demonstrated that the withdrawal of exogenous FGF-2 from hESC results in loss of pluripotent marker expression (Oct-4, Tra-1-60, and Podocalyxin). The presence of all four FGFRs on Oct-4 positive cells was also illustrated. Furthermore, blocking FGFR signaling in hESC was shown to lead to rapid differentiation [13, 17]. Taken together, these data suggest that FGF-mediated signaling is important for the maintenance of the undifferentiated hESC phenotype. To understand FGF-2 mediated signaling in hESC, a large-scale phosphoproteomics approach was embarked upon to investigate the global tyrosine phosphorylation events following FGF-2 stimulation.

The main challenge in identification of tyrosine phosphoproteomics is the low abundance of tyrosine phosphorylated proteins. In this study, we used a pY-specific antibody to enrich for these low abundant pY-peptides from large quantities of complex samples. Using this approach, a cumulative 735 unique tyrosine phosphorylation sites in two biological replicates

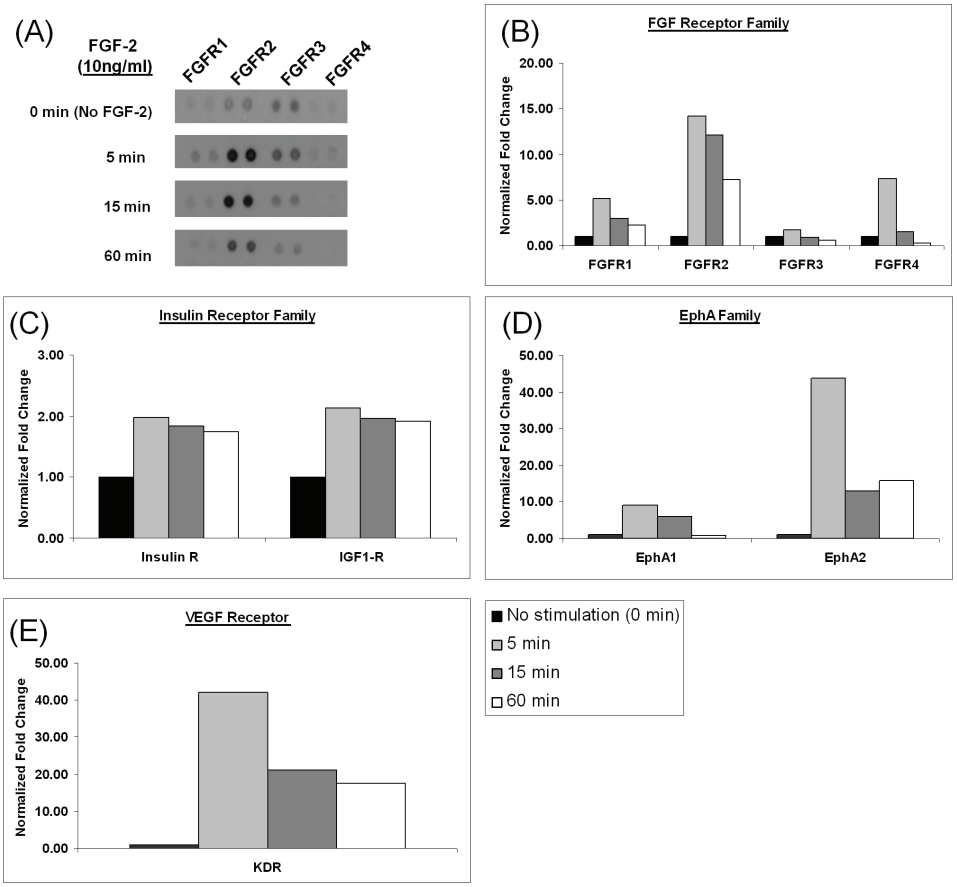


Figure 6. *Human phospho-RTK array to detect proteins activated upon FGF-2 stimulation. A, B) Phosphorylation levels of FGFR 1, 2, 3, and 4 are increased upon FGF-2 stimulation. C, D, E) Proteins in the insulin receptor, ephrin A, and VEGF receptor family show an increase in phosphorylation upon FGF-2.*

were identified. To our knowledge, this is to date the largest hESC tyrosine phosphorylation dataset reported. The levels of tyrosine phosphorylation seem surprisingly large. Often the inhibition of phosphatases with, for example, pervanadate is necessary to detect tyrosine phosphorylation sites [26], while here, FGF-2 starved cells were only stimulated with exogenous FGF-2. Seemingly, ample tyrosine phosphorylation network pathways are active in hESC. Previous large scale hESC phosphorylation profiling studies reported thousands of phosphorylation sites but were largely biased to serine and threonine phosphorylation with only 2-4% of the identified phosphorylation sites were on tyrosine residues [19-21].

Quantitative profiling of tyrosine phosphorylation

In order to study key early signaling events of FGF-2 stimulated hESC, it is important to understand the temporal involvement of the FGFRs and its subsequent substrates post FGF-2 stimulation. Results from Western blotting of FGF signaling downstream effectors (ph-ERK and ph-AKT) demonstrated rapid responses (5-15 min) of these downstream effectors to FGF-2 stimulation, hence we selected 0, 5, and 15 min post-FGF-2 for our quantitative study. pY-peptide immunoprecipitation using a pY specific antibody was combined with stable isotope dimethyl labeling to quantitatively profile tyrosine phosphorylation. Thus, 316 unique

tyrosine phosphorylated peptides were identified and quantified. By performing quantitative phosphoproteomics at the peptide level, temporal dynamics of specific phosphorylated sites rather than total tyrosine phosphorylation of the whole protein can be followed.

FGF receptors and other receptor tyrosine kinases

Expectedly, increases in tyrosine phosphorylation at multiple sites of FGFR 1, 2, 3, and 4 were detected upon FGF-2 stimulation which was confirmed by the human phospho-RTK array. The phosphorylation of autophosphorylation sites suggests indeed activation of these receptors which is further corroborated by the detection of increased tyrosine phosphorylation of downstream targets of canonical FGF signaling, such as PLC-γ, GAB1, MAPK1/3, and PI3-K.

Interestingly, several other RTKs were identified in our quantitative proteomics dataset to show a temporal tyrosine phosphorylation response upon FGF-2 stimulation. Amongst these receptors were the IR, IGF1-R, ERBB2, ERBB3, EphA1 and 2, EphB3 and 4, and KDR/VEGFR2. IGF1-R and IR have been shown to play a role in the self-renewal of hESC [7, 36]. Furthermore, tyrosine phosphorylation of ERBB2, ERBB3, KDR/VEGFR2, FGFR3, and FGFR4 has been demonstrated before upon stimulation with both conditioned medium and FGF2 [36]. The activation of most of these receptors was confirmed by the human phospho-RTK array. With the array, changes in total tyrosine phosphorylation can be robustly and sensitively measured and it has therefore been used to determine the activation state of the captured RTKs {}. Next to activation of the FGF receptors, in our study, the phospho-RTK array showed increased total tyrosine phosphorylation upon stimulation with FGF-2 for the VEGF receptor, insulin receptor family and ephrin family members. The increased phosphorylation is in agreement with the increase in tyrosine phosphorylation that was observed using quantitative LC-MS. The latter method however has the advantage that it can determine exactly which tyrosine sites are differentially phosphorylated and therefore provide higher resolution data. Transactivation (*i.e.* the activation of a given receptor activates a heterologous receptor [37]) has been suggested for FGFR and other RTKs [38-40]. For example, FGFR1 was shown to be capable of tyrosine phosphorylating EphA4 in an adult kinase-negative mutant cell line [38]. Also, it has been demonstrated that a transactivation mechanism might even be required to establish certain physiological processes [40]. The elucidation of these transactivations and cross-talk between different pathways, once again, illustrates that a linear view of signal transduction is a large simplification of the actual pathway {}. Some RTKs share several of the same downstream targets, including the above mentioned PLC-γ, GAB1, MAPK1/3, and PI3-K which might make it difficult to determine whether phosphorylation of any of these proteins was the result of upstream activation of FGFR or one of the other RTKs. A more comprehensive view is therefore requisite for the determination of signaling pathways underlying the effect of certain stimuli. Here, we show the potential involvement of several RTKs in the FGF-2 induced self-renewal of hESCs which could aid in the elucidation of the triggered signaling pathway.

Src family kinases and substrates

Src family kinases (SFKs) represent a group of tyrosine kinases that are strongly activated by, amongst others, RTKs [41]. Several tyrosine phosphorylation sites were identified on members of SFKs, but the extensive homology in particularly the activation loop of different members of the family complicates the determination of the exact family member. Interestingly, although only relatively small increases in tyrosine phosphorylation were observed for the SFK members, larger increases in tyrosine phosphorylation were observed for Src

substrates and, for example, focal adhesion kinase (FAK). Src and FAK are known to form a tight complex after activation by RTKs or Integrins [42, 43]. Src binds to phosphorylated FAK and promotes further phosphorylation of tyrosine residues on FAK.
Cortactin has long been known to be a very efficient substrate for SFKs and was found in our screen to exhibit increased tyrosine phosphorylation levels. These phosphorylation events are supposed to create docking sites for proteins containing the SH2 phosphotyrosine binding domain, such as Src family kinases themselves or adaptor proteins [44, 45]. These interactions are then thought to influence actin polymerization and thus the turnover of actin networks. SHANK2 was also found to be heavily phosphorylated upon induction with FGF. It is a large scaffold protein associated with actin, and also a binding partner of Cortactin [46]. Another protein that shows increased tyrosine phosphorylation is the scaffold protein Paxillin, which localizes at the sites of cell adhesions. Together with a large number of interaction partners, Paxillin is heavily involved in cytoskeletal reorganization and cell adhesion [47-50].
Plakophilin-3, Plakophilin-4, Catenin delta-1, F11R, Occludin, KIRREL/Neph1, TJP-1 and TJP-2 are found in tight junctions and an increased tyrosine phosphorylation was detected in our screen. Of these proteins Catenin delta-1, Occludin, TJP-1 and TJP-2 have been described as genuine Src kinase substrates; the phosphorylation of the aforementioned sites has been shown to modulate the regulation of cell-cell adhesions as well as the formation of tight junctions [47, 51, 52].
Another activated kinase to be found upregulated upon FGF-treatment was CDK5. CDK5 is an atypical member of the cyclin dependent kinase family, which has rather been found to be involved in the regulation of CNS development [53]. Many of its substrates are involved in cytoskeletal regulation and neuronal migration such as c-Abl, Src, PAK1 or beta-catenin.

CONCLUSIONS

Taken together these findings, FGF-2 treatment of hESC results in the activation of all four FGFR family members as well as transactivation of members of several other receptor families such as the insulin receptor family, the Ephrin type A family and the VEGF receptor. Increased tyrosine phosphorylation was found on downstream substrate targets of these RTKs including Src family kinases and their substrates. Seemingly, major and very fast changes in the regulation of cytoskeletal/actin depending processes are induced, including modulation of cell-adhesion, cell-cell adhesion, migration and the formation of tight junctions. Many of the proteins found in our screen that show a distinct increase in tyrosine phosphorylation induced by FGF have been described as integral components and also key players in the aforementioned processes in adult cells. Here, we show activation of several of these pathways in embryonic stem cells upon FGF-2 stimulation. For the differentiation of embryonic cells, cytoskeletal and actin depending processes are majorly involved. Therefore, the role of FGF-2 in the maintenance of the pluripotency state of hESCs might be partly found in the regulation of these processes.

ACKNOWLEDGEMENTS

This work was supported by the Netherlands Proteomics Centre (http://www.netherlandsproteomicscentre.nl), a program embedded in the Netherlands Genomics Initiative.
Supplementary data is available free at http://tinyurl.com/thesispjb

REFERENCES

[1] Rikova, K., Guo, A., Zeng, Q., Possemato, A., *et al.*, Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* 2007, *131*, 1190-1203.

[2] Reubinoff, B. E., Pera, M. F., Fong, C. Y., Trounson, A., Bongso, A., Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. *Nat Biotechnol* 2000, *18*, 399-404.

[3] Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., *et al.*, Embryonic stem cell lines derived from human blastocysts. *Science* 1998, *282*, 1145-1147.

[4] Phillips, B. W., Hentze, H., Rust, W. L., Chen, Q. P., *et al.*, Directed differentiation of human embryonic stem cells into the pancreatic endocrine lineage. *Stem Cells Dev* 2007, *16*, 561-578.

[5] Xu, X. Q., Graichen, R., Soo, S. Y., Balakrishnan, T., *et al.*, Chemically defined medium supporting cardiomyocyte differentiation of human embryonic stem cells. *Differentiation* 2008, *76*, 958-970.

[6] Avery, S., Inniss, K., Moore, H., The regulation of self-renewal in human embryonic stem cells. *Stem Cells Dev* 2006, *15*, 729-740.

[7] Bendall, S. C., Stewart, M. H., Menendez, P., George, D., *et al.*, IGF and FGF cooperatively establish the regulatory stem cell niche of pluripotent human cells in vitro. *Nature* 2007, *448*, 1015-U1013.

[8] Ludwig, T. E., Bergendahl, V., Levenstein, M. E., Yu, J. Y., *et al.*, Feeder-independent culture of human embryonic stem cells. *Nat Methods* 2006, *3*, 637-646.

[9] Xu, C. H., Rosler, E., Jiang, J. J., Lebkowski, J. S., *et al.*, Basic fibroblast growth factor supports undifferentiated human embryonic stem cell growth without conditioned medium. *Stem Cells* 2005, *23*, 315-323.

[10] Xu, R. H., Peck, R. M., Li, D. S., Feng, X. Z., *et al.*, Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nat Methods* 2005, *2*, 185-190.

[11] Ibrahimi, O. A., Zhang, F. M., Hrstka, S. C. L., Mohammadi, M., Linhardt, R. J., Kinetic model for FGF, FGFR, and proteoglycan signal transduction complex assembly. *Biochemistry* 2004, *43*, 4724-4730.

[12] Mohammadi, M., Olsen, S. K., Ibrahimi, O. A., Structural basis for fibroblast growth factor receptor activation. *Cytokine Growth Factor Rev* 2005, *16*, 107-137.

[13] Dvorak, P., Dvorakova, D., Koskova, S., Vodinska, M., *et al.*, Expression and potential role of fibroblast growth factor 2 and its receptors in human embryonic stem cells. *Stem Cells* 2005, *23*, 1200-1211.

[14] Brandenberger, R., Wei, H., Zhang, S., Lei, S., *et al.*, Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. *Nat Biotechnol* 2004, *22*, 707-716.

[15] Bhattacharya, B., Miura, T., Brandenberger, R., Mejido, J., *et al.*, Gene expression in human embryonic stem cell lines: unique molecular signature. *Blood* 2004, *103*, 2956-2964.

[16] Sperger, J. M., Chen, X., Draper, J. S., Antosiewicz, J. E., *et al.*, Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc Natl Acad Sci U S A* 2003, *100*, 13350-13355.

[17] Ding, V. M. Y., Ling, L., Natarajan, S., Yap, M., *et al.*, FGF-2 activates PI-3K and WNT signaling pathways to maintain undifferentiated hESC and iPS cell phenotype. *Stem Cells* 2009, *submitted*.

[18] Li, J., Wang, G. W., Wang, C. Y., Zhao, Y., *et al.*, MEK/ERK signaling contributes to the maintenance of human embryonic stem cell self-renewal. *Differentiation* 2007, *75*, 299-307.

[19] Swaney, D. L., Wenger, C. D., Thomson, J. A., Coon, J. J., Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A* 2009, *106*, 995-1000.

[20] Van Hoof, D., Munoz, J., Braam, S. R., Pinkse, M. W. H., *et al.*, Phosphorylation Dynamics during Early Differentiation of Human Embryonic Stem Cells. *Cell Stem Cell* 2009, *5*, 214-226.

[21] Brill, L. M., Xiong, W., Lee, K. B., Ficarro, S. B., *et al.*, Phosphoproteomic Analysis of Human Embryonic Stem Cells. *Cell Stem Cell* 2009, *5*, 204-213.

[22] Pinkse, M. W. H., Uitto, P. M., Hilhorst, M. J., Ooms, B., Heck, A. J. R., Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-nanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem* 2004, *76*, 3935-3943.

[23] Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., *et al.*, Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat Biotechnol* 2002, *20*, 301-305.

[24] Rush, J., Moritz, A., Lee, K. A., Guo, A., *et al.*, Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat Biotechnol* 2005, *23*, 94-101.

[25] Zhang, Y., Wolf-Yadlin, A., Ross, P. L., Pappin, D. J., *et al.*, Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics* 2005, *4*, 1240-1250.

[26] Boersema, P. J., Foong, L. Y., Ding, V. M. Y., Lemeer, S., *et al.*, In depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immuno-affinity purifi-

cation and stable isotope dimethyl labeling. *Mol Cell Proteomics* 2009, *in press*.
[27] Hsu, J. L., Huang, S. Y., Chow, N. H., Chen, S. H., Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem* 2003, *75*, 6843-6852.
[28] Boersema, P. J., Aye, T. T., van Veen, T. A. B., Heck, A. J. R., Mohammed, S., Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. *Proteomics* 2008, *8*, 4624-4632.
[29] Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S., Heck, A. J. R., Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protocols* 2009, *4*, 484-494.
[30] Choo, A., Padmanabhan, J., Chin, A., Fong, W. J., Oh, S. K. W., Immortalized feeders for the scale-up of human embryonic stem cells in feeder and feeder-free conditions. *J Biotechnol* 2006, *122*, 130-141.
[31] Ding, V., Choo, A. B. H., Oh, S. K. W., Deciphering the importance of three key media components in human embryonic stem cell cultures. *Biotechnol Lett* 2006, *28*, 491-495.
[32] Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., *et al.*, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 2006, *127*, 635-648.
[33] van Breukelen, B., van den Toorn, H. W. P., Drugan, M. M., Heck, A. J. R., StatQuant: A post quantification analysis toolbox for improving quantitative mass spectrometry. *Bioinformatics* 2009, *25*, 1472-1473.
[34] Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H. Y., *et al.*, PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res* 2003, *13*, 2129-2141.
[35] Schwartz, D., Gygi, S. P., An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol* 2005, *23*, 1391-1398.
[36] Wang, L., Schuiz, T. C., Sherrer, E. S., Dauphin, D. S., *et al.*, Self-renewal of human embryonic stem cells requires insuhn-like growth factor-1 receptor and ERBB2 receptor signaling. *Blood* 2007, *110*, 4111-4119.
[37] Wetzker, R., Bohmer, F. D., Transactivation joins multiple tracks to the ERK/MAPK cascade. *Nat Rev Mol Cell Biol* 2003, *4*, 651-657.
[38] Yokote, H., Fujita, K., Jing, X. F., Sawada, T., *et al.*, Trans-activation of EphA4 and FGF receptors mediated by direct interactions between their cytoplasmic domains. *Proc Natl Acad Sci U S A* 2005, *102*, 18866-18871.
[39] Polanska, U. M., Fernig, D. G., Kinnunen, T., Extracellular Interactome of the FGF Receptor-Ligand System: Complexities and the Relative Simplicity of the Worm. *Dev Dyn* 2009, *238*, 277-293.
[40] Esposito, C. L., D'Alessio, A., de Franciscis, V., Cerchia, L., A Cross-Talk between TrkB and Ret Tyrosine Kinases Receptors Mediates Neuroblastoma Cells Differentiation. *PLoS ONE* 2008, *3*, e1643.
[41] Brunton, V. G., Frame, M. C., Src and focal adhesion kinase as therapeutic targets in cancer. *Curr Opin Pharmacol* 2008, *8*, 427-432.
[42] Parsons, J. T., Focal adhesion kinase: the first ten years. *J Cell Sci* 2003, *116*, 1409-1416.
[43] van Nimwegen, M. J., van de Water, B., Focal adhesion kinase: a potential target in cancer therapy. *Biochem Pharmacol* 2007, *73*, 597-609.
[44] Daly, R. J., Cortactin signalling and dynamic actin networks. *Biochem J* 2004, *382*, 13-25.
[45] Buday, L., Downward, J., Roles of cortactin in tumor pathogenesis. *Biochim Biophys Acta* 2007, *1775*, 263-273.
[46] Boeckers, T. M., Bockmann, J., Kreutz, M. R., Gundelfinger, E. D., ProSAP/Shank proteins - a family of higher order organizing molecules of the postsynaptic density with an emerging role in human neurological disease. *J Neurochem* 2002, *81*, 903-910.
[47] Mitra, S. K., Schlaepfer, D. D., Integrin-regulated FAK-Src signaling in normal and cancer cells. *Curr Opin Cell Biol* 2006, *18*, 516-523.
[48] Schaller, M. D., Paxillin: a focal adhesion-associated adaptor protein. *Oncogene* 2001, *20*, 6459-6472.
[49] Playford, M. P., Schaller, M. D., The interplay between Src and integrins in normal and tumor biology. *Oncogene* 2004, *23*, 7928-7946.
[50] Deakin, N. O., Turner, C. E., Paxillin comes of age. *J Cell Sci* 2008, *121*, 2435-2444.
[51] Rao, R., Occludin phosphorylation in regulation of epithelial tight junctions. *Ann N Y Acad Sci* 2009, *1165*, 62-68.
[52] Saito, K., Enya, K., Oneyama, C., Hikita, T., Okada, M., Proteomic identification of ZO-1/2 as a novel scaffold for Src/Csk regulatory circuit. *Biochem Biophys Res Commun* 2008, *366*, 969-975.
[53] Dhariwala, F. A., Rajadhyaksha, M. S., An unusual member of the Cdk family: Cdk5. *Cell Mol Neurobiol* 2008, *28*, 351-369.

Summary

In proteomics, high-tech nano-LC and MS instrumentation is used to routinely sequence proteins at a large scale. The protein status of (part of) an organism can thus be determined, which can be dynamic in, for example, protein expression and post translational modifications (PTMs). The field of proteomics has evolved rapidly over the last decade with substantial technical advances. However, improvements are still necessary to increase the sensitivity and dynamic range of LC-MS in order to dig deeper into the proteome to identify lower abundant proteins, to advance MS based quantification and to enhance throughput. In this thesis, several technological developments are described to advance proteomics and their applicability is demonstrated in several different research lines.

In chapter 1, the concept of proteomics is introduced and several practical aspects of a proteomics workflow are highlighted. A universal proteomics workflow does not exist as different research questions and the availability of instruments require different approaches. However, several components come back in many proteomics experimental procedures. Generally, one of the first steps is the lysis of a biological sample followed by enzymatic digestion. As the complexity of the sample is increased tremendously by the digestion, several ways of fractionation can be applied, such as strong cation exchange (SCX) and hydrophilic interaction liquid chromatography (HILIC). To isolate proteins or peptides containing certain PTMs, enrichment techniques can be applied before or after sample fractionation such as immobilized metal affinity chromatography, TiO_2 or antibody based immunoprecipitation. The most important step in a proteomics experiment is the sequencing of peptides by LC-MS to identify the protein contents of the sample. This is performed by matching the masses of intact peptides and their fragments with the theoretical masses as derived from genomic databases.

Chapter 2 focuses specifically on the use of HILIC in proteomics. This relatively new and little explored LC phase exhibits some features that can be utilized in proteomics. The orthogonality of separation with reversed phase (RP)-LC and the more even dispersion of peptides over the LC run, makes HILIC an adequate first dimension for the fractionation of complex peptide mixtures. Furthermore, HILIC was found to be useful in the enrichment of phosphorylated, N-acetylated and glycosylated peptides and has been used for the separation of differently modified histones.
In chapter 3, a specific variety of HILIC is evaluated and further optimized for two dimensional-LC (2D-LC); zwitterionic HILIC (ZIC-HILIC). A mixed mode separation was observed for ZIC-HILIC consisting of both electrostatic and polar interactions between the peptides and stationary phase. The final separation mechanism can be altered by adjusting the pH of the solvent as this affects the hydrophilicity and charge of peptides. Compared to the often used SCX, less clustering of the typically ubiquitous +2 and +3 charged peptides was observed, rendering ZIC-HILIC an attractive alternative to SCX in multidimensional peptide separations. This was confirmed by the analysis of a complex biological sample consisting of several thousands of proteins whereby a semi-automatic ZIC-HILIC fractionation was setup that mixes the elution with enough aqueous buffer to allow direct consecutive analysis by RP LC-MS. More than a thousand proteins could be identified using the 2D-LC system.

Several approaches of stable isotope labeling have been developed to allow for MS based quantification in proteomics. Some incorporate the isotopes at metabolic level during, for example, cell culturing. In other methods proteins or peptides are chemically labeled with isotopomeric labels. In chapter 4, the development of a triplex stable isotope dimethyl labe-

ling approach is reported. By using different isotopomers of formaldehyde and cyanoborohydride, three different dimethyl labels can be generated in order to simultaneously analyze peptides from three different samples by LC-MS. The advantage of this approach is the utilization of cheap and readily available chemicals and therefore, the method does not pose limits to the amount of sample that has to be labeled. Furthermore, the labeling reaction goes to close to completion and virtually any sample can be labeled as the labeling is performed post-lysis and -digestion. It was found that the first dimension of a 2D-LC system has to be chosen carefully when working with stable isotope dimethyl labeled samples. For example, in HILIC a small retention shift is observed for the differently labeled peptides due to the different numbers of deuterium the labels contain. This results in a slight differential elution of isotopomeric peptides over consecutive fractions. As SCX separation is not based on hydrophilicity or hydrophobicity, this isotope effect was not observed and was therefore used in the application of stable isotope dimethyl labeling in a targeted study wherein the method was used to distinguish specific from aspecific cAMP binding proteins when performing pull-down experiments using immobilized cAMP. In total, 142 proteins could be identified and quantified, including several cAMP binding proteins and several of their A-kinase anchoring protein binding partners.

Trypsin is the most often used enzyme for proteolysis in proteomics approaches. The resultant peptides have a basic N-terminus and C-terminus (Lys or Arg residue). After collision induced dissociation (CID) a mass spectrum typically consists of both b- and y-ions. In chapter 5, the metalloendopeptidase Lys-N is investigated for MALDI-MS/MS proteomics applications. Lys-N produces peptides with an N-terminal Lys residue and therefore the basicity is concentrated at the N-terminus. Fragmentation in MALDI of peptides where this N-terminal Lys is the only basic group results in the generation of primarily N-terminal fragments. The CID spectra are straightforward and the sequence can be easily read off since often complete sequence ladders of b-ions are present. The extent of this effect was statistically supported by performing Lys-N digestion on a complex cellular lysate. It was calculated that b-ions were responsible for 94% of the total intensity of b- and y-ions in Pro-free peptides. Furthermore, it was found that these straightforward spectra facilitate *de novo* sequencing. Several peptides could be identified from an ostrich muscle sample without bioinformatic database searching tools.

Phosphorylated peptides typically fragment differently compared to non-modified peptides. In CID for example, fragmentation of phosphorylated peptides can result in spectra that are dominated by a single peak that represents the neutral loss of the phosphate group. In chapter 6, this neutral loss effect and ways to accommodate challenges of MS based analysis of phosphopeptides are reviewed. The neutral loss of the phosphate group of a phosphate has often been described as the result of a β-elimination reaction during CID. However, multiple lines of evidence are discussed that suggest that charge-directed S_N2-neighbouring group participation reactions might have a more dominant effect. The extent of neutral loss is dependent on several aspects including the chemical structure of the phosphorylated residue, the amino acid sequence of the peptide, the charge state of the precursor ion, proton mobility, the type of mass spectrometer and the applied collision energy. The neutral loss can hinder identification of the peptide sequence and the determination of the exact site of phosphorylation. Several alternative fragmentation approaches such as MS^3, multistage activation, ECD and ETD have been developed to accommodate some of the challenges of MS analysis of phosphopeptides. Although all these methods show certain

advantages over CID, they also seem to identify complementary and not always larger sets of phosphorylated peptides. Therefore, not one ideal fragmentation methods exists that is capable of fully and comprehensively determining a phosphoproteome. Often the faster sequencing speed of CID MS/MS allows a larger set of phosphopeptides to be identified.

The analysis of Tyr phosphorylation is rather challenging due to the low levels of Tyr phosphorylation. Approaches have been developed whereby Tyr phosphorylated peptides are immunoprecipitated using antibodies that are specific for Tyr phosphorylation. In chapter 7, the enrichment efficiency of such phosphopeptide immuno-affinity purification was shown to be rather high and more than 1000 phospho-Tyr peptides could be identified after two separate LC-MS runs. In the optimized immunoprecipitation protocol the input is several milligrams of sample. To introduce stable isotope labeling in the protocol, triplex stable isotope dimethyl labeling was combined with phosphoTyr immunoprecipitation. Tyr phosphorylation was profiled after pervanadate and EGF stimulation, respectively. Using the quantitative immunoprecipitation approach and just a single LC-MS run, a rather complete qualitative and quantitative image was generated of Tyr phosphorylation signaling events. Many known Tyr phosphorylation sites and several sites that have not or only poorly been described to be involved in EGF receptor signaling were detected.
In chapter 8, the quantitative immunoprecipitation method was applied to study the role of FGF-2 stimulation in human embryonic stem cells (hESCs). FGF-2 is important in the maintenance of the pluripotency state of hESCs. FGF-2 stimulation was shown to activate MAPK and AKT pathways. In our study, several hundreds of Tyr phosphorylation sites could be identified and quantified of which some 140 showed a differential regulation upon FGF-2 stimulation. These regulated sites were found on proteins that include FGF receptors, members of the Src family, proteins involved in actin polymerization and cyclin dependent kinases.

Nederlandse Samenvatting

In proteomics worden high-tech nano vloeistofchromatografie (liquid chromatography; LC) en massa spectrometrie (MS) instrumenten gebruikt om op grote schaal eiwitten te kunnen identificeren. Op deze manier kan bepaald worden welke eiwitten in (een gedeelte van) een organisme voorkomen. Deze eiwit status kan erg dynamisch zijn door veranderingen in eiwit expressie en post translationele modificaties (PTMs). Het proteomics veld heeft zich snel geëvolueerd over de laatste decennia met belangrijke technische ontwikkelingen. Toch zijn technische verbeteringen nog steeds noodzakelijk om de gevoeligheid en het dynamische bereik van LC-MS te verhogen om zo dieper in het proteome te graven en minder voorkomende eiwitten te kunnen identificeren, om op MS gebaseerde kwantificatie methoden te verbeteren en om de analyse snelheid te verhogen. In dit proefschrift worden enkele technologische proteomics ontwikkeling beschreven en hun toepassing in een aantal verschillende onderzoekslijnen.

In hoofdstuk 1 wordt het concept proteomics geïntroduceerd en passeren verscheiden praktische aspecten van een proteomics experiment de revue. Een universeel draaiboek voor een proteomics experiment bestaat niet, omdat verschillende onderzoeksvragen en de beschikbaarheid van bepaalde apparatuur verschillende werkwijzen noodzaken. Over het algemeen is een van de eerste stappen van een proteomics experiment de lysis van een biologische monster gevolgd door enzymatische digestie. Verscheidene manieren van fractioneren kunnen vervolgens toegepast worden zoals strong cation exchange (SCX) en hydrophilic interaction liquid chromatography (HILIC). Om eiwitten of peptiden met bepaalde PTMs te kunnen isoleren kunnen verrijkingsmethode gebruikt worden voor of na monster fractionering zoals 'immobilized metal affinity chromatography', TiO_2 of op antilichaam gebaseerde immunoprecipitatie. De belangrijkste stap in een proteomics experiment is het sequencen van peptiden met LC-MS om te bepalen welke eiwitten zich in een monster bevinden. Dit kan gedaan worden door de massa's van intacte peptiden en hun fragmenten te vergelijken met de theoretische massa's uit een genomische databank.

Hoofdstuk 2 richt zich specifiek op het gebruik van HILIC in proteomics. Deze relatief nieuwe en weinig gekarakteriseerde LC fase bezit enkele eigenschappen die interessant zijn voor proteomics. De orthogonaliteit van scheiding in vergelijking met 'reversed phase' (RP)-LC en de betere verspreiding van peptiden over een LC scheiding maken van HILIC een geschikte eerste dimensie voor de fractionering van complexe peptiden mengsels. HILIC blijkt ook bruikbaar voor de verrijking van gefosforyleerde, N-geacetyleerde en geglycosyleerde peptiden en het is toegepast voor de scheiding van verschillend gemodificeerde histonen.
In hoofstuk 3 is een specifieke versie van HILIC geëvalueerd en geoptimaliseerd voor gebruik in twee dimensionale-LC (2D-LC); zwitterionic-HILIC (ZIC-HILIC). Aan de scheiding met ZIC-HILIC blijkt een tweeledig mechanisme ten grondslag te liggen met zowel electrostatische als polaire interacties tussen de peptiden en de stationaire fase. Het uiteindelijke scheidingsmechanisme kan aangepast worden door de pH van de oplosmiddelen te veranderen waardoor de hydrofiliciteit en lading van peptiden verandert. In vergelijking met het veel gebruikte SCX vindt minder clusteren van +2 en +3 geladen peptiden plaats wat ZIC-HILIC een aantrekkelijk alternatief voor SCX maakt voor multidimensionale peptide scheidingen. Dit werd bevestigd door de analyse van een complex biologisch monster dat bestond uit enkele duizenden eiwitten. Een semiautomatische ZIC-HILIC fractionering was ontworpen waarbij het eluaat direct met genoeg waterige buffer wordt gemengd waardoor analyse met RP LC-MS direct mogelijk is. Meer dan duizend eiwitten konden geïdentificeerd worden door gebruik te maken van dit 2D-LC systeem.

Verscheidene methoden van stabiele isotoop labelen zijn ontwikkeld om op MS gebaseerd kwantificatie mogelijk te maken in proteomics. Sommige van die methoden introduceren de isotopen op metabool niveau, bijvoorbeeld tijdens celkweek. In een andere benadering worden eiwitten of peptiden chemische gelabeld met isotopomere labels. In hoofdstuk 4 wordt de ontwikkeling van een triplex stabiele isotoop dimethyl labeling methode besproken. Door verschillende isotopomeren van formaldehyde en cyanoborohydride te gebruiken kunnen drie verschillende dimethyl labels gegenereerd worden waardoor tegelijkertijd drie verschillende monster met LC-MS geanalyseerd kunnen worden. Het voordeel van deze methode is dat goedkope en algemeen verkrijgbare chemicaliën gebruikt worden waardoor de hoeveelheid monster dat gelabeld moet worden geen beperkende factor is. Verder is de reactie zo goed als volledig en kan vrijwel elk monster gelabeld worden, omdat de labeling plaats vindt na lysis en digestie. De juiste keuze van de eerste dimensie van een 2D-LC systeem bleek van belang te zijn wanneer er gewerkt wordt met stabiele isotoop dimethyl gelabelde monsters. Met HILIC kan een kleine retentie verschuiving optreden van de verschillend gelabelde peptide door het verschillende aantal deuteriums dat de labels bevatten. Dit kan resulteren in een niet evenredige elutie van verschillende isotopomere peptiden over opeenvolgende fracties. SCX scheiding is niet gebaseerd op hydrophiliciteit of hydrophobiciteit en een isotoop effect is daardoor verwaarloosbaar. SCX was daarom gebruikt in een toepassing van stabiele isotoop dimethyl labeling waarbij de methode werd gebruikt om onderscheid te maken tussen specifiek en aspecifiek aan cAMP bindende eiwitten in pull-down experimenten waarbij gebruik wordt gemaakt van geïmmobiliseerd cAMP. In totaal konden 142 eiwitten geïdentificeerd en gekwantificeerd worden inclusief verscheidene cAMP bindende eiwitten en enkele van hun AKAPs.

Trypsine is het meest gebruikte enzym voor proteolysis in proteomics. De peptiden die op deze manier geproduceerd worden hebben een basische N-terminus en C-terminus (Lys of Arg residu). De spectra die worden gegenereerd met collision induced dissociation (CID) bestaan over het algemeen uit b- en y-ionen. In hoofdstuk 5 is de metalloendopeptidase Lys-N getest voor toepassing in MALDI-MS/MS. Lys-N produceert peptiden met een N-terminale Lys residu en daardoor is de basiciteit van het peptide geconcentreerd op de N-terminus. Fragmentatie met MALDI-MS/MS van peptiden waarbij Lys de enige basische aminozuur is resulteert in de productie van voornamelijk N-terminale fragmenten. De CID spectra zien er eenvoudig uit en de aminozuur volgorde van het peptide kan makkelijk afgelezen worden, omdat vaak volledige sequentie ladders van b-ionen aanwezig zijn. De mate van het effect werd statistisch ondersteund door een Lys-N digestie van een complex cellulair lysaat. B-ionen bleken verantwoordelijk voor 94% van de totale intensiteit van b- en y-ionen van niet-Pro-bevattende peptiden. Verder blijken de vereenvoudigde spectra *de novo* sequencing te vergemakkelijken. Verscheidene peptiden konden geïdentificeerd worden in een spiermonster van een struisvogel zonder bioinformatische programma's.

Gefosforyleerde peptiden fragmenteren over het algemeen anders dan hun niet gefosforyleerde equivalenten. In CID kunnen gefosforyleerde peptiden bijvoorbeeld spectra genereren die gedomineerd worden door een enkele piek veroorzaakt door de 'neutral loss' van de fosfaat groep. In hoofdstuk 6 wordt deze neutral loss besproken samen met manieren om om te gaan met de hindernissen die opgeworpen kunnen worden in de analyse van fosfopeptiden met MS. Vaak wordt de neutral loss van de fosfaat groep omschreven als een β-eliminatie reactie tijdens CID. Er komt echter meer en meer data beschikbaar die doet ver-

moeden dat een 'charge-directed S_N2 neighbouring group participation reaction' ten grondslag ligt. De mate van neutral loss hangt af van verscheidene aspecten zoals de chemische structuur van het gefosforyleerde residu, de aminozuurvolgorde van het peptide, de lading van het precursor ion, 'proton mobility', het type massaspectrometer en de hoeveelheid energie die toegepast wordt om fragmentatie te induceren. De neutral loss kan de identificatie van de peptide sequentie en de lokalisatie van de fosforylering bemoeilijken. Verscheidene alternatieve fragmentatie technieken zijn daarom ontwikkeld zoals MS^3, 'multistage activation', ECD en ETD. Hoewel deze technieken enkele voordelen hebben ten opzichte van CID, blijken ze ook andere subgroepen van peptiden te identificeren. Er bestaat daarom niet één ideale fragmentatie techniek die een phosphoproteome volledig kan bepalen. Vaak blijken meer fosfopeptiden geïdentificeerd te kunnen worden met CID MS/MS door de hogere sequentie snelheid.

De analyse van Tyr fosforylering wordt behoorlijk bemoeilijkt door de lage hoeveelheden van Tyr fosforyleringen. Er zijn methodes ontwikkeld waarbij Tyr gefosforyleerde peptiden geïmmunoprecipiteerd kunnen worden door gebruik te maken van antilichamen met specificiteit voor Tyr fosforylering. In hoofdstuk 7 wordt aangetoond dat de efficiëntie van zo'n fosfopeptide immunoprecipitatie erg hoog kan zijn en dat meer dan 1000 fosfoTyr peptiden geïdentificeerd konden worden in twee LC-MS analyses. De input in het geoptimaliseerde protocol is enkele milligrammen. Om stabiele isotoop labeling toe te kunnen passen werd daarom gekozen om de fosfoTyr immunoprecipitatie te combineren met stabiele isotoop dimethyl labeling. Tyr fosforylering na pervanadaat of EGF stimulatie kon op deze manier gekarakteriseerd worden. Door gebruik te maken van deze kwantitatieve immunoprecipitatie benadering kon in een enkele LC-MS analyse een behoorlijk compleet kwalitatief en kwantitatief beeld gegenereerd worden van Tyr fosforyleringen. Vele Tyr fosfosites die bekend zijn in EGF receptor signaal transductie konden geïdentificeerd worden, maar ook enkele onbekende of weinig beschreven sites werden geïdentificeerd.
In hoofdstuk 8 is de kwantitatieve immunoprecipitatie methode toegepast op de analyse van de rol van FGF-2 stimulatie in humane embryonale stamcellen (hESC). FGF-2 is belangrijk voor de instandhouding van de pluripotente toestand van hESC. In onze studie konden enkele honderden Tyr fosforyleringen geïdentificeerd en gekwantificeerd worden. Ongeveer 140 daarvan lieten een differentiële regulatie door FGF-2 stimulatie zien. Deze fosforyleringen konden gedetecteerd worden op eiwitten zoals FGF receptoren, leden van de Src familie en eiwitten die betrokken zijn bij actine polymerisatie en 'cyclin dependent kinases'.

Curriculum Vitae

Paul Boersema is op 22 mei 1982 in Assen geboren. Na het doorlopen van het VWO aan het Gomarus College te Groningen, begon hij de studie biologie aan de Rijksuniversiteit Groningen met een specialisatie in medische biologie. Onderdeel van de afstudeerfase waren onderzoeksstages in de afdeling pathologie en medische biologie aan de Rijksuniversiteit Groningen onder leiding van Dr. Aalzen de Haan en het Conway Institute Proteome Research Centre aan het University College Dublin in Ierland onder leiding van Prof. Dr. Michael Dunn en Dr. Melanie Föcking. Tijdens deze laatste stage werd hij geïntroduceerd met het concept proteomics en raakte zo geïnteresseerd in de potentie van dit nieuwe vakgebied dat hij na het afstuderen als assistent in opleiding in dienst trad bij de biomoleculaire massa spectrometrie en proteomics groep aan de Universiteit van Utrecht. De resultaten van het onderzoek dat onder leiding van Prof. Dr. Albert Heck en Dr. Shabaz Mohammed plaats vond zijn beschreven in dit proefschrift. Enkele maanden van het promotietraject werden besteed aan het Research Center for Proteome Analysis van Prof. Dr. Rong Zeng in het Shanghai Institutes for Biological Sciences and Institute of Biochemistry and Cell Biology in Shanghai in China. Onderzoeksresultaten werd verder gepresenteerd op nationale en internationale congressen in onder meer Indianapolis, Seoul, Amsterdam en San Francisco.

List of publications

- **P.J. Boersema**, L.Y. Foong, V.M.Y. Ding, S. Lemeer, B. van Breukelen, R. Philp, J. Boekhorst, B. Snel, J. den Hertog, A.B.H. Choo, A.J.R. Heck (2009) In depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immuno-affinity purification and stable isotope dimethyl labeling. *Molecular & Cellular Proteomics* (*in press)*

- M.L. Hennrich, **P.J. Boersema**, N. Mischerikow, A.J.R. Heck, S. Mohammed (2009) The effect of chemical modifications on the peptide fragmentation behavior upon electron transfer induced dissociation. *Analytical Chemistry*, Sep. 81(18):7814-7822

- **P.J. Boersema**, S. Mohammed, A.J.R. Heck (2009) Phosphopeptide fragmentation and analysis by mass spectrometry. *Journal of Mass Spectrometry*, Jun. 44(6):861-878 (review)

- **P.J. Boersema**, N. Taouatas, A.F.M. Altelaar, J.W. Gouw, P.L. Ross, D.J. Pappin, A.J.R. Heck, S. Mohammed (2009) Straightforward and de novo peptide sequencing by MALDI-MS/MS using a Lys-N metallopeptidase. *Molecular & Cellular Proteomics*, Apr. 8(4):650-660

- **P.J. Boersema**, R. Raijmakers, S. Lemeer, S. Mohammed, A.J.R. Heck (2009) Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nature Protocols*, Mar. 4(4):484-494

- **P.J. Boersema**, T.T. Aye, T.A. van Veen, A.J.R. Heck, S. Mohammed (2008) Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. *Proteomics*, Nov. 8(22):4624-32

- **P.J. Boersema**, S. Mohammed, A.J.R. Heck (2008) Hydrophilic interaction liquid chromatography (HILIC) in proteomics. *Analytical and Bioanalytical Chemistry*, May 391(1):151-159 (review)

- **P.J. Boersema**, N. Divecha, A.J.R. Heck, S. Mohammed (2007) Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. *Journal of Proteome Research*, Mar. 6(3):937-946.

- M. Föcking, **P.J. Boersema**, N. O´Donoghue, G. Lubec, S.R. Pennington, D.R. Cotter, M.J. Dunn (2006) 2-D DIGE as a titative tool for investigating the HUPO Brain Proteome Project mouse series. *Proteomics*, Sep. 6(18):4914-4931

Dankwoord

It is wonderful to see my PhD work materialize in the shape of this booklet. But this result of four years of work could not have been produced without the help, support and company of colleagues, friends and family. Running the risk of forgetting someone, I will to try to list the people that I am greatly indebted to.

Als eerste moet ik Albert natuurlijk bedanken, want zonder jou zou ik dit proefschrift niet hebben kunnen schrijven. Bedankt voor het vertrouwen dat je in me hebt gehad door me een AIO-plaats aan te bieden na slechts een korte sollicitatieronde. Bedankt voor alle ondersteuning, ideeën en supersnelle correcties van manuscripten. En natuurlijk voor de uitzonderlijke kans om een tijdje in China te kunnen vertoeven.
Shabaz, it has been great to have you as my day-to-day supervisor. It was funny to see you struggle to remain into some kind of boss-student relation while trying to be involved in all the social events and being always on the lookout for some juicy gossip. Anyway, thanks for teaching me all the ins and outs of LC-MS, proteomics and science in general and for checking and correcting manuscripts and other output very quickly and thoroughly.

Of course I should also mention the people from the group where it all started: the Conway Institute of Proteome Research Center in Dublin, Ireland. It was here that I got enthusiastic about proteomics and its potential. Thanks Mike, Melanie (and Peter and kids), Steve, Jane, Jennifer, Caitriona, Niaobh, Ashling and others for the wonderful time I had in Dublin.

Then there are a lot of colleagues from the Biomolecular Mass Spectrometry and Proteomics Group that I am grateful to. Over the years the group has grown and changed quite a lot. And so did the group dynamics and atmosphere. I am happy for the many people that I met and could socialize with. Spending long days at work also means that a large part of the social life actually takes place in the lab. It has been wonderful to work together with people for which this was the same and to interact with people of so many different nationalities.
In het begin van mijn tijd in Utrecht was er de (Nederlandse) bierclub en koffieclub om mee te socializen. Annemieke, het is toch verwonderlijk hoe snel acht jaar (jaren?) voorbij kunnen gaan. Nog bedankt trouwens voor de uitnodiging voor je promotiefeestje, maar ik zit dan helaas nog in Dublin. Eef bedankt voor je zuidelijke gezelligheid, punctuele aankondigingen van lunch en koffie en je grootse levenswijsheden per e-mail. Jeroen Ko., nog nooit heb ik zo'n atypische wetenschapper gezien, niet alleen door je humor, maar ook je botte opmerkingen en je muzieksmaak (ik weet nog steeds niet of ik er trots op moet zijn dat ik de Dikke Lul Band live heb gezien). Toch doe je het wetenschappelijk gezien goed en ben je prima gezelschap om bijvoorbeeld een aantal dagen in Cincinnati en New York te verkeren. Sjemig Llléon, jij was toch wel de beste kantoorgenoot. Goed gezelschap, altijd in voor een geintje maar op de juiste momenten serieus en met een prima muzieksmaak. De dagelijkse Gummbah moet ik nu missen: "Wat zeggen we dan tegen de…". Ook bedankt voor je gastvrijheid in Oslo. Joost bedankt voor je eeuwige bereidheid om bier te drinken, zelfs als dat betekende dat je 's nachts per fiets terug naar Amersfoort moest. Jij hebt me ook warm gemaakt voor snowboarden waar ik de rest van mijn leven blij mee kan zijn. Verder waren Jeroen J., Martijn en Harm-Jan een goed gezelschap voor uitjes naar Gulpen of gewoon de Vooghel. Simone, het was altijd het kwartiertje wachten waard om naar huis te fietsen en eens heerlijk cynisch tegen ons werk aan te kijken. Wanneer gaan we ook al weer naar Singapore? Arjen, jij was er, ging weg en kwam net op tijd weer terug om de A in SJAP van Annemieke over te nemen. Bedankt voor je gezelschap tijdens de vele lunches.
I realize that I was quite privileged in that I did not have a mass spectrometer to take care of.

Nadia and Sharon G., you are two of the people that I did not envy for the amount of time you spent at the instruments. Thanks for your dedication, the least I could do is to listen to your complaints after you said 'yes, yes' to the boss. Next to Nadia -btw thanks for all the sweets and the complementary chat-, also Thin Thin, Donna (in the UMC for a year), Erwin and Marco were great office mates in Z305, although some plants needed to be grown to get some privacy to be able to think with eyes closed.
Also my current office mates are really enjoyable. Chhiti ("You changed, since Poupaks party!"), thanks for your Indian influence. Thomas, you complemented the office with your interesting, often very honest, humor. Serena, why did you not join our group earlier! Thanks to your Italian passion, energy, high fun-factor, craziness, hospitability, great cooking and many more qualities, last year flew by like a breeze. It was always good fun to teach you LC-MS and proteomics or sit behind your computer to try and get you to understand how Mascot and all the different servers in the biomass network work. Let's plan our next holidays in a bit more synchronized manner. Tvucdb!
Arjan, het was altijd leuk om met jou te geinen in het lab. Corine, bedankt voor alle administratie en declaraties die je voor me hebt geregeld. Kristina, thanks for your lovely dinners and interesting experiences afterwards. Christian sr., thanks for the pub quiz nights and upgrading the pTyr stories. Sarah and Glen (leugenaar!) thanks for your hospitability and endless energy. It is wonderful to have you around during drinking sessions and wintersport trips. Rebecca, for some reason I cannot drink gin-tonic anymore. Tom, not really a colleague, but thanks for taking me to the Nieuwegein circuit for cycling and for the other bicycle runs. Charlotte, thanks to you I could confirm that death metal is not really my thing ;-). Soenita, bedankt voor alle taarten, cakes, koekjes en snoep waar je ons altijd mee verwende. Javier, thanks for teaching us to curse in Spanish and the wonderful parties at your place. Poupak, bedankt voor je hartelijkheid en de introductie tot de Iraanse keuken. Tieneke, bedankt dat ik voor mijn promotie nog even bij jou kan afkijken hoe het ook al weer werkt.
Danny, Pieter, Henk, Robert, Bas, Madalina, Shahram and Wim -the 'computer guys'- I am grateful for all the programming and solving issues with especially MSQuant and StatQuant. Furthermore I want to thank Adja, Andreas, Ayse, Bas N., Bas T., Benjamin, Christian F., Erik, Esther, Gideon, Harm, Hongtao, Houjiang, Ioana, Jeff, Jeffrey, Jeroen Kr., Kees, Lars, Luigi, Maarten, Martina, Martje, Mirjam, Monique, Nasrin, Natasja M., Natasja V., Nikiana, Nikolai, Onno, Pepijn, Reinout, Renske, Salvatore, Sara, Silvia, Silvie, Teck, Vincent G., Vincent H., and Werner for their help and company.

Also some collaborators I would like to thank for introducing some biology into my thesis: Sue and Tokameh at the Hubrecht and Sharon M. and Frank at Molecular Genetics. It is always interesting to talk with non-mass spectrometrists and broaden your horizon. Thanks to you our LC-MS results were more than just lists of proteins and expression ratios. Vanessa and Yan, it was wonderful to host your visit in Holland. Thanks for introducing our lab to pTyr IPs. Jos en Berend, bedankt voor jullie hulp met de bioinformatica kant.

Zhibin Ning, Yangshen Liu, Qingrun Li, Zhou Hu, Fu Ning, Yibo Wu, Rong Zeng, Paul Shieh and the others in Shanghai I have to acknowledge for a wonderful time in China. Thanks for helping me out in the lab and, more importantly, also outside the lab with getting train tickets, tips on where to go, finding good restaurants and more generally, thanks for introducing me to the Chinese culture.

After work it is good to be also hanging out with people outside the lab. Angela, Maria, RK

and others, thanks for the wonderful evenings and diners. Albert-Jan en Harry, Saffier 130 is nog steeds mijn beste woonervaring. Leuk om dat nog een beetje door te kunnen zetten met weekendjes weg, omdat we wat verder van elkaar wonen (Albert-Jan dan). Het is helaas nog een beetje onduidelijk waar ik de komende voorjaarstrip kan hosten. Edwin, ik moest toch wel lachen toen jij ook met een AIO-onderzoek begon. Ik ben er nu mee klaar, jij moet nog even vooruit. Dat betekent dat ik in ieder geval de eerst komende tijd nog een prima plek in Groningen heb om een biertje te komen drinken en over cultuur met een hoofdletter C te praten. Ook de jongens uit Winsum kan ik natuurlijk niet vergeten, hoewel een gedeelte Winsum inmiddels heeft verlaten en in een redelijk andere levensfase verkeert van trouwen, koopwoningen en kinderen.

En dan is er natuurlijk nog altijd de back-up van de familie. Anthoinet, Arnold en de kids, Jaap en Cilia, Dick, Karin en Otto en Eduard, het was altijd leuk om met zijn allen of een aantal een weekend door te brengen in Winsum of een van de andere plaatsen waar we verbleven. Grappig om te zien dat we ouder worden, veranderen, relaties aangaan, gaan werken, maar dat alles toch voelt als vanouds wanneer we bij elkaar zijn. Papa en mama bedankt voor jullie ondersteuning. We zijn anders gaan denken over bepaalde dingen, maar naar Winsum gaan voelt nog steeds als thuis komen.

After finishing primary school, high school and university, another phase in my life is over. I am excited about starting something new, but at the same time sad to finish what I have now. Anyway, thanks to all of you I can always look back on the wonderful time I had during my PhD.

Paul