Dedicated to my parents, Angela & Rino

Mamma, Papá, senza il vostro supporto, amore, dedizione e stima, la mia ambizione non avrebbe mai raggiunto livelli tali da volare in cieli così alti e sondare terre tanto lontane. Voi rappresentate per me quel modello esemplare di vita che sempre mi accompagnerà nel futuro, ovunque saró.

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 for printing this thesis is kindly acknowledged.

Cover photo: superb view of the Olympic Mountains from Vancouver Island, as a good memory of the sensational time I spent there during my PhD.

Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress

ISBN 978-90-8891-558-1

Optimizing hydrophilic interaction liquid chromatography for ultrasensitive proteome analysis

Optimaliseren van hydrofiele interactie vloeistofchromatografie voor ultragevoelige proteoomanalyse (met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op vrijdag 8 februari 2013 des middags te 2.30 uur

door

Serena Di Palma

geboren op 1 februari 1983, te Campobasso, Italië

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

Co-promotor: Dr. S. Mohammed

TABLE OF CONTENTS

Chapter 1

1

Introduction

1

TABLE OF CONTENTS

1 PROTEOMICS

Analytical protein chemistry, or "proteomics", refers to the branch of analytical science focusing on proteins.[1] The term proteome appeared in the literature in 1997 as a linguistic equivalent to the concept of genome and is used to describe the complete set of proteins expressed by the entire genome of a cell.[2] Moreover, proteomics expresses the ambition to obtain a global view at the protein level, in analogy to what is possible to obtain at the DNA and RNA levels, bridging the gap between our understanding of the genome sequence and the cellular response.[3, 4]

From a 'systems biology' point of view, proteomics delivers mainly three types of information. First, it embraces the study of interactions of proteins with other proteins, modified peptides, small molecules, and specific RNA and DNA sequences, a branch that is also defined as "interactomics".[5] A second type of information is the posttranslational modification (PTM) state of a protein.[6] PTMs can affect structure and stability of a protein, having potential effects on its biological function, as well as switching the protein into an active or deactivate state. Third, "expression proteomics" determines the relative and absolute amount of proteins in a sample.[7] This is analogous to "transcriptomics", which measures mRNAs by microarrays or deep sequencing methods. However, the mRNA levels have been found to not fully correlate with eventual protein expression due to differences in the extent of re-use of mRNA.[8] Thus, the main advantages of focusing on proteins is that it takes into account the regulation at the posttranscriptional and posttranslational level.[9] These three fields can be woven together and applied in a myriad of different formats to study biological and medical questions.

The technological basis of most current proteomics studies is biological mass spectrometry (MS). This field was first catapulted to mainstream prominence with the 'development of soft desorption ionization methods for mass spectrometric analyses of biological macromolecules', for which in 2002 John B. Fenn and Koichi Tanaka jointly received a Nobel Prize in Chemistry. Then, a number of decisive breakthroughs followed, notably optimized protocols to handle biological sample for MS analysis, powerful separation methods for the analysis of complex protein mixtures, automated peptide identification software and bioinformatics tools for data analysis, establishment of quantitative techniques, and improvements in mass spectrometric instrumentation.

During the past decade, most proteomics studies relied on tandem mass spectrometry (MS/MS) as the core technology, specifically on a method referred to as 'bottom-up' proteomics.[10] The key concept is fundamentally explained by the controlled decomposition of a proteome into peptides and their analysis by mass spectrometry. One idea behind this is that an ensemble of peptides has a narrower distribution of physico-chemical properties compared to a mixture of proteins, and peptides are analytically and preparatively less challenging than proteins. The other idea is that the fragmentation of peptides by tandem MS is well understood and occurs in a more or less predictable way, yielding fragment spectra that can be used for peptide identification.

1.1 Generic shotgun proteomics workflow

A 'standard' proteomics workflow does not exist and always depends on the specific research question and available resources. However, bottom-up proteomics can be summarized as following. Proteins are extracted from cells or tissues and digested into peptides. Peptides in the sample are separated, typically by liquid chromatography, and then ionized by electrospray ionization (ESI) or matrix-assisted laser desorption/ionization (MALDI). Ionization is required to transfer analytes into the mass spectrometer as gaseous ions, where they are subjected to fragmentation and their spectra are recorded. Fragment ion spectra are the currency of infor-

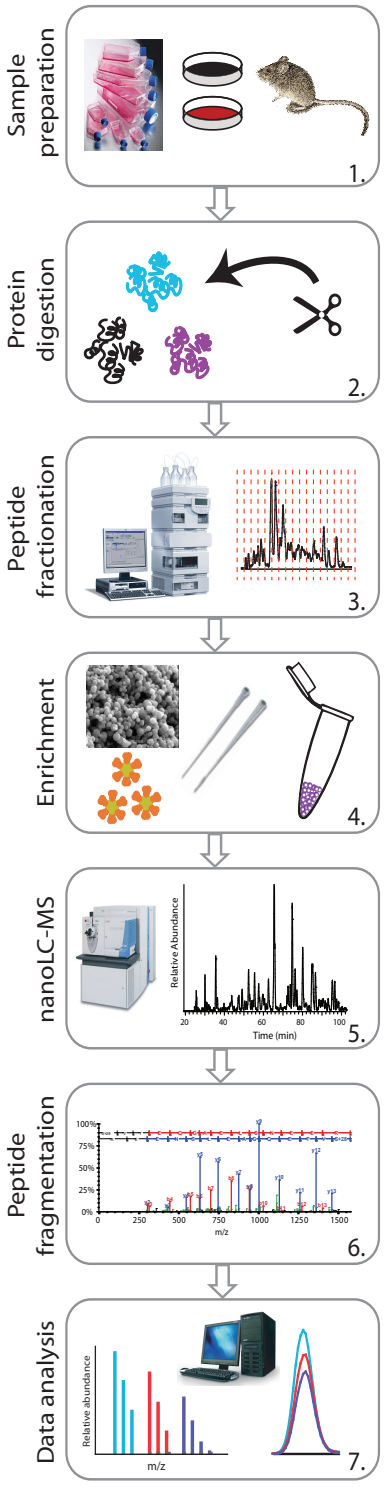

Figure 1: *General proteomics workflow.*

mation, as they can be assigned to peptide sequences from which the corresponding proteins are inferred. Fragment ion spectra also contain precious information to detect modified amino acid residues and to identify and locate modifications within the peptide sequence. Moreover, MS signals can be used to estimate the quantity of a peptide or protein.[11, 12] For every step of the process, including sample preparation and fractionation, MS data acquisition, quantification, and data analysis, multiple methods and tools have been developed.[13-15] This also applies to the MS instrumentation, which continuously enjoys increases in performance in regard to speed, mass accuracy, sensitivity and analytical robustness.[16, 17]

In Figure 1, I outline the general steps that would apply to most proteomics strategies and to the work described in this thesis. The sample preparation step largely depends on the type of sample and on the ultimate goal of the research, e.g. a comprehensive global proteome analysis, or a more targeted study on specific cell types, subcellular components or certain PTMs. A sample can include cells from laboratory culturing, isolated primary cells, tissue, biological fluid, plant material, etc. In general, the sample is lysed employing denaturing agents such as the surfactants sodium dodecyl sulfate (SDS) or sodium deoxycholate (SDC), or such as the chaotropic reagents urea or thiourea. The use of these reagents enhances cell and tissue solubilization and protein unfolding and denaturation, preventing their precipitation. However, surfactants, even at low concentrations, can preclude enzymatic digestion and dominate mass spectra due to their favorable ionizability and their great abundance compared to individual peptides. Therefore, depletion of detergents is a prerequisite for efficient mass-spectrometric analysis in proteomics.

The following step (Step 2) is the protein digestion with enzymes possessing specific substrate specificity to generate peptides that are in the molecular weight window suitable for MS analysis. The choice of enzyme is undoubtedly important as each enzyme generates a unique pool of peptides with differing sequence characteristics, length distribution, solubility and charge. The most common proteolytic enzyme is trypsin, which cleaves proteins specifically and efficiently at the C-terminal side of lysine and arginine residues.[18] These basic residues are adequately distributed over a protein sequence resulting in the gen-

eration of ideal peptide length for successful peptide sequencing by MS. Alternative proteases are also employed, sometimes in combination with trypsin, leading to significantly high protein sequence coverage. For example, Lys-C protease cuts at the C-terminal side of lysine and is also active at high concentration of denaturing agent. Thus, Lys-C can be combined with trypsin, which is more sensitive to reagent's concentration, to maximize the protein digestion. Recently, a complementary enzyme was introduced in proteomics, called Lys-N, which cleaves at the N-terminal side of lysine. The resulting peptides produce simplified fragmentation spectra when electron transfer dissociation (ETD) is used as a fragmentation technique.[19]

Digestion of protein mixtures results in a much larger number of proteolytic peptides than the initial number of proteins, hampering the chance for all those peptides to be sequenced by MS during a single LC-MS analysis. There are two main parameters limiting the analysis of complex peptide mixture and preventing full proteome identification. First, the instrument has a finite peptide sequencing speed. The other limitation is the dynamic range of measurement. Therefore, the complexity of the sample is often reduced in order to increase the possibility for the mass spectrometer to detect as many peptides as possible. The most common strategy to reduce complexity is the use of additional step(s) of peptide separation. In fact, proteomics is tightly linked and highly dependent on separation technologies. Though sample fractionation here is reported as step 3, the selection of an appropriate separation method is often the first step in designing a proteomic workflow. Often, peptide fractionation is based on liquid chromatography (LC) and several LC modes can be employed, which are mainly based on hydrophobicity, hydrophilicity, charge state, and size of peptides.[20] An extensive discussion of LC methods for peptide separation, their combination in multidimensional formats and relative applications in proteomics can be found in chapter 2.

The complexity of the proteome also includes protein modifications, such as phosphorylation. One of the major challenges in the analysis of PTMs, especially phosphorylation, is their relative low stoichiometry in comparison to 'regular' unmodified proteins. This will cause the majority of modified proteins (and peptides) to be masked by the more abundant unmodified analogues. To cope with this issue, a plethora of enrichment techniques are being developed. A number of methods that selectively target phosphorylated peptides also exist, which can be applied after (Step 4) or before sample fractionation. Several enrichment strategies take advantages of the chelating properties of some metals toward the phosphate group, such as immobilized Fe3+ ions,[21] or Metal Oxide Affinity Chromatography, such as TiO_2[22] or ZrO_2.[23] Some non-affinity based techniques, mainly chromatography based, offer additional strategies for enrichment. The most common are anion-exchange, mixed-bed, and hydrophilic interaction liquid chromatography. For a more detailed discussion of phosphoproteomic enrichment methods see sections 4.1 and 4.2.

In the last separation step (Step 5), which is often interfaced directly to MS via electrospray ionization (ESI), peptides are separated employing reversed-phase (RP) chromatography. RP has shown superior separation power, efficiency, and reproducibility compared to other LC modes, and has an excellent buffer compatibility with ESI. In fact, RP still remains the workhorse of separation methods in single-shot proteomics as well as last dimension in multidimensional strategies prior to MS analysis.

The mass spectrometer, then, determines the mass-to-charge (m/z) of peptides eluting from the LC column and electrosprayed into the MS. The instrument can isolate and select peptide ions for fragmentation in order to provide direct information on peptide sequences. The achievable proteome coverage depends on several parameters, including power, length and resolution of LC separation, dynamic range of detection, mass accuracy, scanning speed and fragmentation type of the mass spectrometer. Further discussion about fragmentation, MS

instrumentation and characteristics can be found in section 2.
The final step of the presented workflow is data analysis, referring to a set of bioinformatics tools required to compare the experimental MS/MS data to theoretical spectra, generated in silico from genomic databases, to obtain matches that fit the theoretical data with high probability. With this approach, candidate fragmentation spectra can be selected according to the peptide precursor mass, yielding a confident identification even when the MS/MS spectrum would not contain enough information to deduce the peptide sequence by direct interpretation. A drawback of this approach is that databases can be sometimes quite large and may result in false identifications. Therefore, robust statistical validations need to the applied to get an indication of inaccurate assignments or false positive matches in the datasets.

1.2 Proteomics strategies

From an extensive menu of available options for each single step, individual choices can be combined into different workflows and MS strategies, each addressing different types of biological inquiries.[12, 15] The most frequently used strategy is referred to as 'shotgun' or "discovery proteomics".[24] There, peptide precursor ions are detected in a survey scan and automatically selected using a process called data-dependent analysis. This strategy results in large datasets containing vast numbers of proteins and can enable quantitative comparisons between different samples either via stable isotope labeling or without labeling.[7, 14, 25] Shotgun proteomics does not require any prior knowledge of the composition of the sample, and thus each protein is newly discovered. In "directed proteomics", precursors are only selected for fragmentation if they are detectable in a survey scan and present on a list of predetermined precursor ions; an 'inclusion list'.[26, 27] This strategy results in datasets that identify and quantify specific, predetermined segments of a proteome at a higher level of reproducibility, compared to discovery proteomics. In "targeted proteomics", only predetermined peptides are selected for detection and quantification in a sample. The main mass spectrometric method that supports targeted proteomics is 'selected reaction monitoring' (SRM) also referred to as 'multiple reaction monitoring'.[28, 29] In SRM, specific MS assays are generated a priori for each targeted peptide, and then used to selectively detect and quantify analytes in multiple biological samples.[30] This method can generate highly reproducible and accurate datasets for small, preselected fractions of a proteome (typically one to a few hundreds peptides) at a wide dynamic range.[31] Finally, with recent advances in instrumentation, a forth strategy called "data-independent analysis"[32-35] is emerging in which no selection of the precursor ion occurs. Rather, all precursors are fragmented, either over the whole mass range or in swaths of smaller mass ranges.[36] Each of these strategies captures a different subset of the total proteome space, balancing trade-offs in comprehensiveness, reproducibility, selectivity, sensitivity, accuracy and dynamic range.[37]
In the following sections, several steps of the general workflow are described and discussed in more detail, as well as many aspects related to proteomics strategies, to help understanding the impact of the work presented in this thesis.

2 MASS SPECTROMETRY

MS has become the method of choice for the identification of peptides, and therefore proteins, in proteomics. A peptide can be identified based on the m/z of the precursor ion in combination with the m/z of product ions obtained by fragmenting the precursor in MS/MS mode. Following this approach, any protein can be theoretically identified by MS, making it an essential instrument in proteomics.
A mass spectrometer can be described as consisting of 4 different parts: ionization source,

mass analyzer, detector, and data processing system. A sample is introduced into the mass spectrometer through the ionization source. Ions are separated according to their m/z in the mass analyzer and captured by the detector. In a new round, ions can be selected and fragmented in a collision cell that can be tandem-in-space or tandem-in-time, after which the product ions are measured in the mass analyzer. The detector records the number of ions at each m/z value and, finally, a data processing system generates the mass spectrum in a suitable form. The following sections will give a brief overview of operational principles and key features of these technologies.

2.1 Ionization

Analysis in a mass spectrometer is performed under vacuum. To introduce a biological sample into a mass spectrometer, compounds must be converted to gas-phase ions without extensive degradation. In proteomics, this is typically performed by soft-ionization techniques, such as MALDI or ESI.

In MALDI, ions are generated from a solid state. The sample is mixed with a matrix and co-crystallized on a MALDI target plate by vaporization of the solvents. A rapid laser hits the plate and induces heating of the matrix 's crystals by accumulation of a large amount of energy. This excitation energy, which is thought to be partly transferred to the analytes, caused desorption of the matrix and desolvatation of the analytes. In this process, peptides are predominantly converted into singly protonated ion species.[38, 39]

Unlike MALDI, the ESI source produces ions from a solution. The analyte is dissolved in a solvent and dispersed into a fine aerosol through electrostatic charging.[40] 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.[41] Two models have been proposed to explain the final generation of gas-phase ions: the Ion Evaporation Model (IEM)[42] and the Charged Residue Model (CRM).[43] According to the IEM, gas-phase ions can be ejected directly from shrinking droplets before the Rayleigh limit is reached. This can occur during asymmetric droplet fission, when charge repulsions between similarly charged ions

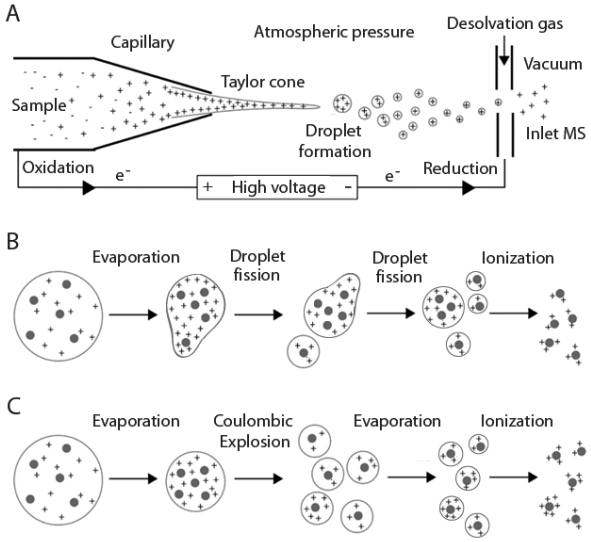

Figure 2: *Illustration of the instrumental setup to initiate electrospray (A), where a high voltage is applied between sample tip and counter electrode and schematic representation of the two models (B, C) explaining the formation of ions during electrospray ionization. B depicts the IEM model that holds mainly for smaller molecules such as peptides, where solvent evaporation leads to smaller and smaller droplets until ions will evaporates directly from the droplet surface. C depicts the CRM model associated to larger molecules such as intact proteins. The shrinking droplets will force the charged molecules closer together until repulsions of charges leads to Coulombic explosions and the formation of single ions in the gas phase.*

exceeds the surface tension of the droplet. In this way small molecules are capable of leaving larger droplets as charged ion species. In CRM, the initial droplet undergoes several cycles of solvent evaporation and asymmetric Coulombic explosions, until droplets contain on average only one charged molecule. The gas-phase ion is formed when all the solvent is evaporated and the ion will retain the residual charge carried by the ion in the solution before the final evaporation. ESI typically produces multiply charged ions, depending on the size of the ion and the amino acid sequence. These two theories are depicted in Figure 2.
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 conveniently interfaced to LC separation techniques.[44] Electrospray ionization is obtained applying a high voltage (1-6 kV) between the silica emitter at the end of the separation pipeline and the inlet of the MS, creating a continuous flow into the MS. An important development in ESI techniques includes nano-ESI,[45] in which the flow rate is lowered to a nanoliter-per-minute regimen. Reducing the flow rate at which peptides elute from the column, as well as the diameter of the emitter opening, increases the sensitivity of the electrospray process. This is caused by a higher efficiency in ion creation and better transfer into the MS. To suit nanoliter flow rates, RP-LC is performed with capillary columns of tens of micrometer inner diameter.

2.2 Mass analyzers

In the mass analyzer, ions are stored and separated according to their m/z. Several types of mass analyzers exist that determine m/z based on different principles. Ion trap, Orbitrap and ion cyclotron resonance separate ions based on their m/z resonance frequency; quadrupoles use m/z trajectory stability; time-of-flight analyzers use velocity or flight time.[46] Each mass analyzer has unique properties, such as mass range, analysis speed, resolution, ion transmission, sensitivity and dynamic range. Hybrid instruments are combinations of mass analyzers to address specific needs.[10, 11] Here, quadrupoles, ion traps, Fourier transform ion cyclotron resonance (FT-ICR) and Orbitrap mass analyzers will be discussed since these types of mass spectrometers have been used extensively in the works reported in following chapters.

2.2.1. Quadrupoles

Quadrupoles consist of 4 perfectly parallel, circular or ideally hyperbolic shaped rods. To each rod a combination of a static (DC) and oscillating (RF) electric field is applied.[47] Gas-phase ions are introduced along the axis into the middle of the radially positioned rods. The same potentials are applied to opposite rods, while adjacent rods have exactly the opposite potential being applied. This dynamic electric field creates a turbulent environment in which ions may or may not possess a stable trajectory. Ions of differing m/z values require different combinations of electric fields in order to stably traverse the quadrupole. If ions have no stable trajectories, they leave the ion path and collide with one of the rods, being discharged. Ions with a stable trajectory through the quadrupole move with a circular motion in the x- and y-axis, resulting in a corkscrew kind of motion through the rods. A schematic representation of a quadrupole MS and ions trajectory is reported in Figure 3.
Through manipulation of the electric fields, it is possible to isolate ions of a particular m/z value or obtain an entire spectrum by scanning through all m/z values. When the amplitude is set such that only ions with one m/z have stable trajectories and all the other ions are instable, the quadrupole acts as a mass filter. When the DC potential is set to ground and only the RF-potential is applied, the quadrupole functions as a wide band mass filter and it can be used as an ion guide to transport ions between different sections of an instrument. Furthermore, quadrupoles can be used as reaction chambers to induce ion fragmentation. Ions entering an RF-only

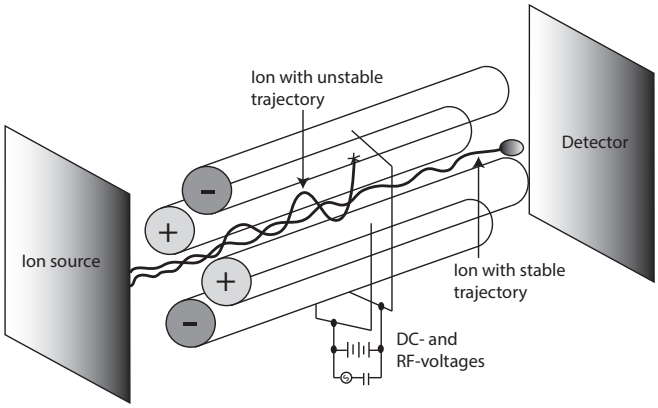

Figure 3: *Schematic illustration of a quadrupole mass spectrometer.*

quadrupole filled with an inert gas and possessing a sufficient kinetic energy, will collide with the gas molecules and will be fragmented. Mass filtering of the fragments is then achieved in another mass analyzer downstream.

Quadrupoles are often used in tandem MS studies and can be found in a triple quadrupole mass configurations (QqQ). Here, the second quadrupole is employed as a collision cell for collision induced dissociation (CID) and the third quadrupole as a mass-selective filter of the fragments.

If instead of four rods, six or eight rods are used, the instrument is called respectively hexapole or octopole, or in general multipole is used to define multiple rods.

2.2.2 Quadrupole ion trap and linear ion trap

The quadrupole ion trap (QIT),[48] also called 3D-trap, consists of 3 electrodes: a circular electrode (ring electrode) and two hyperbolic caps (endcaps), as depicted in Figure 4A. A three-dimensional quadrupolar field is established between an oscillating electric field to the ring electrode while the end-cap electrodes are kept at ground potential. Since there is no DC potential applied, the trap will allow ions of a broad m/z range to enter the QIT axially through a bore in one end-cap electrode. Ions are then excited into stable 3D-trajectories in the centre of the electrode assembly, which effectively means the ions are trapped. Discrimination of m/z is performed by applying a fundamental RF voltage to the ring: its frequency is constant but its amplitude (V) can be varied. Additional RF voltages of selected frequencies can be applied to the endcaps. Often, helium is used as a buffer gas to help focusing the ions toward the centre of the trap. Collisions with helium dampen the kinetic energy of the ions, trapping the ions more efficiently through collisional cooling. Furthermore, careful control of resonance excitation can be used to increase the kinetic energy of the ions to allow fragmentation by CID. Trapped ions are ejected by gradually changing the electrode voltages, resulting in unstable ajectories of ions at specific m/z values.

A different type of ion trap is the 2D-trap, also called linear ion trap (LIT),[48, 49] in which ions are trapped in two dimensions, i.e. along an axis, rather than at the focal point, i.e. at the centre of the trap, as for the 3D-trap. An LIT consists of a linear quadrupole with two end electrodes at the entrance and exit, which can be built of two plates or short quadrupoles, as depicted in Figure 4B. The ions are trapped in an axial motion by a static electric potential on end electrodes and can be ejected radially through slots in the rods. The design of 2D-traps

enables a larger ion storage volume compared to 3D-traps, circumventing the problem of overfilling the trap, referred to as "space-charge" effects, which limits the maximum load of an ion trap due to ion repulsions. Higher storage volumes lead to an increase in dynamic range and sensitivity.

2.2.3 FT-ICR

An FT-ICR instrument determines the m/z of ions based on their cyclotron frequency in a strong and constant magnetic field.[50] It is built from a superconductive magnet and a box-shaped cell with six electrodes aligned with the magnetic field. The cell consists of two pairs of opposite metal plates that are electrically coupled in pair, and two cylinders, at the end of the cell, functioning as trapping electrodes. The two pairs-electrodes are used as excitation and detection plates. Due to the magnetic field, ions inside the cells are trapped on stable orbital oscillations depending on their m/z and magnetic field strength. Next to the magnetic field, a pulsed RF-potential is applied to one pair of opposite metal plates to induce resonance excitation of oscillating ions. These cyclotron frequencies are then sensed by the detector plates and the image current is Fourier transformed to obtain a mass spectrum. With FT-ICR instruments, ions can be stored for a relatively long period of time (several seconds) in the cell with a very high vacuum system, achieving mass resolutions of more than 100.000 FWHM with sub-ppm mass accuracy. ICR mass analyzers can in principle be used for analyzing fragment ions; however, due to their great resolving power but low scan speed, they are commonly used for precursor spectrum acquisition. To address the slow acquisition rate in MS/MS mode, the FT-ICR analyzer is, nowadays, often combined with a linear ion trap analyzer, where ICR is used for MS and LIT for MS/MS.

2.2.4 Orbitrap

The Orbitrap is a relatively new mass analyzer which consists of a central spindle-shaped electrode and a surrounding barrel-shaped electrode that is split in half, separated by an insulating ring.[51] Ions are injected into the Orbitrap and trapped by an electrostatic field applying a DC voltage to the central electrode while the outer barrel is set to ground potential. Ions oscillate axially around the central electrode in a harmonic orbit, moving 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. The oscillating ions induce an image current that is detected with the help of a differential amplifier between the halves of the outer electrode of the orbitrap. The signals are converted into frequencies with the help of a Fourier transform and the frequencies are deconvoluted into m/z values to obtain a mass spectrum. In an Orbitrap, a mass resolution of over 100.000 FWHM [52] can be achieved. dependent on the time span the ions oscillate around the central electrode, with mass accuracies lower than 2 ppm or even in the ppb range when an internal calibration is applied.

2.2.5 Hybrid instruments

Nearly all modern mass spectrometers are hybrid instruments which are combinations of multiple mass analyzers to satisfy specific analytical requirements. For instance, FT-ICR and Orbitrap mass spectrometers are typically hybridized with a linear ion trap, such as the LTQ. MS measurements, where high mass resolution and accuracy is essential, are performed in the FT-ICR cell or Orbitrap, while MS/MS measurements are performed in the LTQ which has a faster scanning speed and where high mass accuracy is not essential. Figure 4C depicts an LTQ-Orbitrap instrument.

One example of a hybrid instrument is the Orbitrap Velos,[16] where quadrupoles and

multipoles, three linear ion traps and an Orbitrap are combined. The ions are transmitted into a dual linear ion trap with different gas pressures. The first trap accumulates ions and induces fragmentation at relative higher pressure (compared to the second trap). The scanning of the fragments is conducted in the second trap at a relative lower pressure. The ions are ejected through slits in the rods in order to reach the detectors. If fragmentation in the dual ion trap is not intended or if the fragments should be read out in the Orbitrap, the ions are guided by a multipole to another "linear" ion trap that is curved, the so called "C-trap". The C-trap is used to accumulate and focus ions in order to allow rapid and homogeneous injection of large ion populations into the Orbitrap. The Orbitrap analyzer is then utilized to read out the m/z with high resolution. The mass spectrometer has an additional multipole for fragmentation, which is a dedicated higher-energy collisional dissociation cell (HCD collision cell). In this case, the fragment analysis is performed in the Orbitrap.[53]

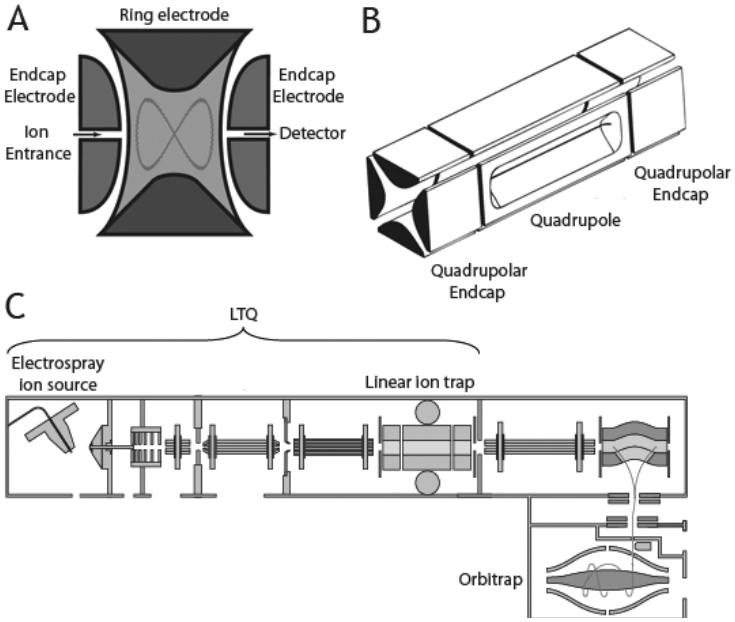

Figure 4: *Schematic pictures of: (A) quadrupole ion trap, (B) linear ion trap and (C) hybrid LTQ-Orbitrap mass spectrometer.*

2.3 Peptide fragmentation

Tandem mass spectrometers are capable of fragmenting peptide ions and to record the resulting product mass spectra. By measuring not only the m/z of intact peptides but also the m/z of their fragments, 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),[54] where 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.[55] CID can be performed in a collision cell that is tandem-in-space or in an ion trap that is tandem-in-time.[56] Fragmentation in a collision cell typically occurs on a short timescale (microseconds), usually involving multiple collisions, with the first one being the strongest, and at 'high' collision energy (from tens of eV to several keVs). In an ion trap, CID fragmentation occurs through slow heating. Here, the precursor undergoes multiple collisions with gas molecules, increasing the internal energy of the ion until dissociation occurs. The activation time is typically several milliseconds, but only a few eV of collision energy is required.

CID typically results in the disruption of single amide bonds with the formation of two types of fragments: b ions with the N-terminal end of the peptide and y ions with the C-termini, according to the Roepstorff-Fohlman nomenclature as reported in Figure 5A and B [57] 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.[58] This semi-random behavior leads to CID spectra containing ions corresponding to cleavage of most of the amide bonds. Considering that the mass difference between two adjacent amide bonds is essentially dictated by the amino acid and that 18 of the 20 amino acids have a unique mass, one can use CID spectra to obtain peptide sequence information.

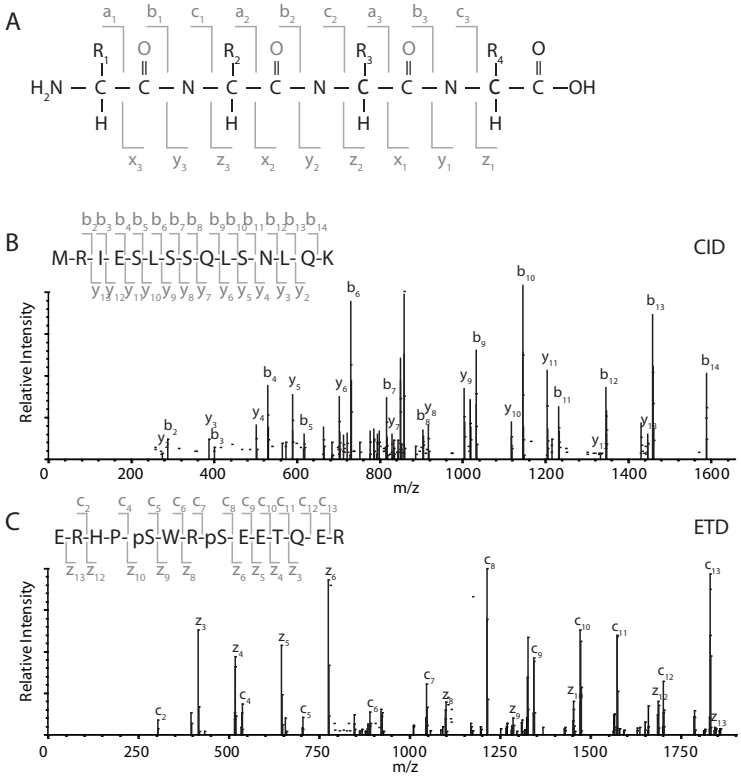

Figure 5: *(A) Roepstorff-Fohlman nomenclature of peptide fragments. (B) Example of a CID and (C) ETD spectra annotated according to the above nomenclature.*

An alternative electron-driven dissociation method, namely electron transfer dissociation (ETD),[59] is becoming more popular in proteomics since it is particularly suited for PTM analysis. In fact, acid labile modifications, such as phosphorylation and glycosylation, are left intact during ETD fragmentation. In ETD, electrons are transferred from a radical anion molecule to multiply protonated peptide cations causing backbone fragmentation at the N-Cα bond of amino acids. The predominant ions found in ETD spectra are c-type fragments at the N-termini and z-type fragments at C-termini, as exemplified in Figure 5C.[57] One disadvantage of ETD is the low conversation rate from precursor to fragment ions, resulting in charge reduced (but non-fragmented) precursor ions.

ETD and CID are complementary fragmentation techniques with respect to charge state. Several studies have demonstrated that ETD leads to a more efficient fragmentation for higher charged peptides and to the poor dissociation of doubly charged peptides.[60] In contrast, CID efficiently fragments doubly charged peptides and struggles more with increasing charge state. As modern high resolution mass spectrometers can determine the charge state of the precursor based on its isotopic distribution, and can also perform both fragmentation techniques, they can be programmed to choose the most appropriate technique automatically based on the charge state. Thus, combining both fragmentation techniques, more complete information on peptide sequences can be gained.[61]

3 PEPTIDE AND PROTEIN IDENTIFICATION

Tandem MS spectra provide direct information about the peptide sequence. However, the huge amount of spectra recorded in a standard LC-MS analysis causes manual interpretation to become a futile exercise. The predominant method to assign sequences to MS spectra is database searching.[13, 62] Database search algorithms, like Mascot,[63] first generate a candidate list of theoretical peptides that coincides with the mass of the precursor, within a defined mass window. The generation of the theoretical peptides is accomplished by in silico digestion of proteins from a specific protein database. Two main parameters affecting the theoretical number of candidate peptides, thus the size of the

database and specificity of the enzyme used for digestion, and possible modifications (both in vivo and in vitro). A very unspecific protease will lead to a long list of theoretical peptides. Moreover, peptides containing missed cleavages need to be taken into account, further increasing the search space. Then, when PTMs are expected, the search engine has to take multiple versions of the same peptide into account, where each contains one or more of the possible modifications.

The principles and assumptions of each search algorithm and their particular scoring functions can be slightly different. However, they all fundamentally exploit the same known rules regarding peptide fragments and generate lists of possible fragments and their masses. The matching between experimental and theoretical spectra is determined within the defined MS/MS mass window, which depends on the mass accuracy and mass precision of the mass analyzer employed for fragment read-out. For each peptide spectrum match (PSM) a score is reported that reflects the quality of the correlation. Again, each search algorithm scores matches and calculates penalties differently. When matching thousands of spectra against millions of possible candidate, the chance of obtaining an incorrect assignment exists. Thus, additional scoring models to assess the confidence of PSMs are generally applied, for example, by generating a statistical measure like the p-value or E value (expectation value). A way of calculating the p-value is to estimate a null distribution from a frequency histogram of the scores from all but the best candidate matched against the experimental spectrum. In the next step, the search engine references the best score against the null distribution. The E value corresponds to the number

of times the score is expected only by chance.
In addition to these scoring models, global approaches to compute optimized cut-off scores for individual datasets based on a false discovery rate (FDR) are often employed. In order to calculate the FDR, a second search is performed against a database with non-sense sequences, a decoy database. Such databases can be generated by reversing or randomizing the sequences of the original or 'target' database.[64] Any hit that is returned when searching against this decoy database (true negative) is equated to a false positive hit in the normal database. The FDR can then be calculated for a conservative threshold score by dividing the number of decoy PSMs by the total number of PSMs.[65] Typically, a peptide score threshold is chosen where the FDR is 1%. However, debate is still ongoing about how FDR should be exactly calculated, and whether these FDRs are really reliable.

4 PHOSPHOPROTEOMICS

Protein phosphorylation is a ubiquitous post-translational modification involved in several key intracellular processes including metabolism, excretion, homeostasis, transcriptional and translational regulation, cellular signaling and communication.[66] It is a reversible PTM that typically leads to changes in the conformation, activity and interactions of a protein within a very short timeframe. Phosphorylation often occurs at multiple residues of a protein (which can be serine, threonine and tyrosine residues) and in most cases is catalyzed by different protein kinases. Mass spectrometry (MS)-based phosphoproteomics has been gradually accepted as a powerful tool to analyze protein phosphorylation in a variety of biological contexts. However, significant analytical barriers still persist, hampering the routine application of phosphoproteomics. First, protein phosphorylations are present at substoichiometric levels. Second, the detection by MS is further impaired by low ionization efficiency and signal suppression in presence of non-phosphorylated species.[67]

4.1 Enrichment strategies

The success of phosphoproteomics by mass spectrometry greatly relies on the use of selective enrichment strategies, which decreases the background of unphosphorylated peptides, improving phosphopeptides identification by MS/MS sequencing. There are many phosphoproteomic enrichment strategies, depending on the type of sample and goals, and they are mainly applied after proteolytic digestion. Figure 6 gives an overview of some strategies currently available.
The most widely applied method is chemical coordination by affinity chromatography. Two common approaches in this area are immobilized metal ion affinity chromatography (IMAC) [68] or metal oxide affinity chromatography (MOAC) with materials such as porous titanium dioxide microspheres (TiO_2).[22] With IMAC, metal ions (Fe^{3+}, Al^{3+}, Co^{2+}, Ga^{3+}, Ti^{4+}) are chelated to coated beads, forming a stationary phase to which negatively charged phosphopeptides can selectively bind under acidic pH, leaving the vast majority of un-phosphorylated peptides in the flow through. Phosphopeptides are then eluted at alkaline pH.[21] One of the issues associated with this technique is the high level of unspecific binding with very complex peptide mixtures containing multiple acidic amino acid residues, reducing the selectivity of the method. One approach to circumvent this issue is the derivatization of the carboxyl groups on acidic residues by O-methyl esterification, though complete derivatization is challenging.[68] Another approach is based on the pH adjustment.[69] By acidification of the IMAC loading buffer (below pH 2), acidic residues are neutralized by protonation, while phosphopeptides (more acidic) will retain their negative charge and their binding affinity toward the IMAC resin.[70]
The specificity of TiO_2 for phosphopetides has been suggested to be based on Lewis acid-base interactions and is considered to be more specific than IMAC.[22] Unspecific binding of

non-phosphorylated peptides can be further reduced by using various organic acids including 2,5-dihydroxybenzoic acid (DHB), DMSO, phthalic acid or glycolic acid in the loading buffer.[71] TiO_2 and IMAC are often performed offline employing microcolumns. However, also on-line strategies have been described for TiO_2 in combination with RP, due to its good compatibility with RP. In an optimized design, a triple stage precolumn consisting of a RP part, a TiO_2 section and again a RP part is used.[72] Here, peptides are first trapped onto the RP. When eluting onto the TiO_2 by increasing the ACN concentration, non-phosphorylated peptides are directly moved toward the next RP part while phosphopeptides are retained in the TiO_2 section. The elution of phosphopeptides onto the second RP precolumn occurs by increasing the pH above 9, followed by an ACN gradient to achieve the final separation into the analytical column. This on-line approach allows the identification of phosphorylated and regular peptides in an automated manner, decreasing sample loss, and can also be applied in large scale experiments, in conjunction with ion-exchange chromatography.

Another strategy, SIMAC,[73] which stands for sequential elution from IMAC, combines the strengths of both IMAC and TiO_2, allowing the enrichment of mono- and multiply phosphorylated peptides from highly complex samples. In SIMAC, after an initial IMAC enrichment, the unbound fraction (flow through) and weekly bound singly phosphopeptides, which usually elute at acidic condition, are collected for another round of enrichment using TiO_2. Multiply phosphopeptides elute from the initial IMAC at basic condition (alkaline pH) and do not require another step of enrichment.

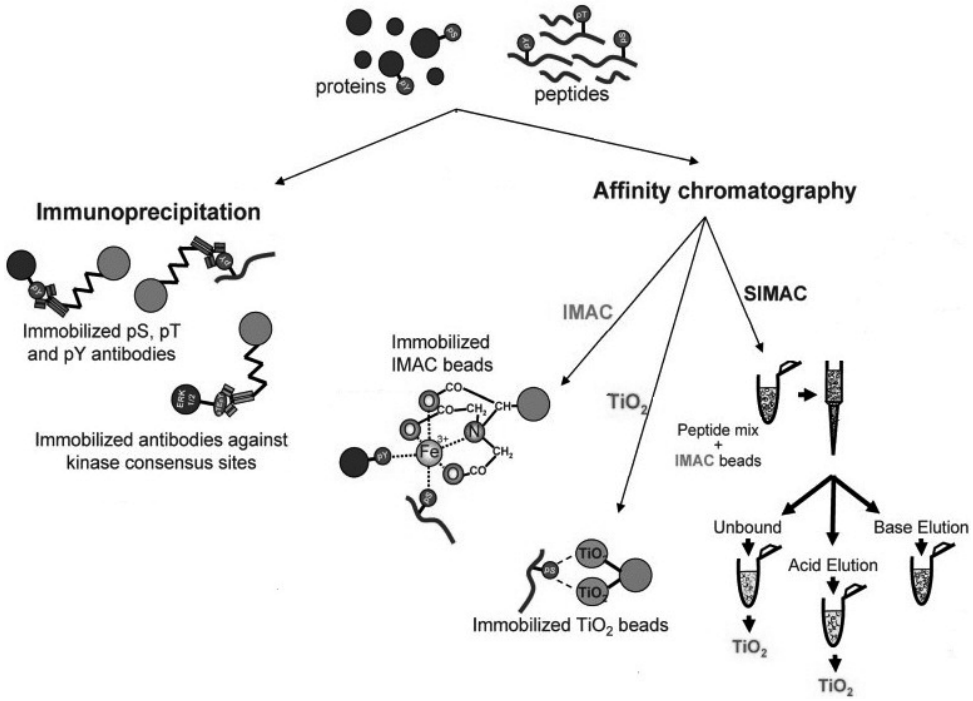

Figure 6: *Strategies for phospho-specific enrichment. Most commonly used strategies for phosphoprotein and phosphopeptide enrichment are based on immunoprecipitation and affinity chromatography.*

4.2 Fractionation strategies

Each enrichment technique is expected to be even more powerful when combined with pre-fractionation methods to decrease sample complexity prior to enrichment. Several LC methods are especially suited for this purpose, such as ion exchange chromatography or hydrophilic interaction liquid chromatography (HILIC). For large-scale phosphoproteomic studies, strong cation exchange (SCX) is the most efficient prefractionation technique prior to either IMAC[74] or TiO_2 chromatography[75], since SCX can also separate and isolate distinct groups of phosphopeptides from the bulk of non-phosphorylated peptides. HILIC can be combined with enrichment techniques prior or after any enrichment step.[76] In fact, in HILIC chromatography, polar compounds are highly retained, such as polar PTMs, providing a means to enrich those analytes and to decrease the background before a more specific enrichment strategy. However, HILIC is also employed after the affinity enrichment to dig deeper into the enriched pool and to achieve a better protein coverage. This topic will be further discussed in chapters 2 and 6.

Each fractionation method has specific drawbacks and advantages and, in general, it would be desirable to keep the workflow as simple as possible, with relatively few sample preparation steps in order to maximize the sensitivity. The more sample handling, the greater the risk of losing phosphopeptides onto, surfaces of tubes and tips and metallic LC components. However, this is less critical with a high amount of starting material and sometimes multiple steps are necessary to obtain a more comprehensive picture.

4.3 Phosphotyrosine enrichment

In nature, tyrosine phosphorylation occurs at low frequency compared to serine and threonine phosphorylation and its investigation relies almost exclusively on immunoaffinity precipitations. This approach is obviously limited by the specificity of the antibody. However, selective phosphotyrosine antibodies exist and can be utilized for the enrichment of phosphotyrosine at both protein and peptide levels.[77, 78] When performed at protein level, it can be difficult to determine the exact site of phosphorylation since Tyr phosphorylated and non-phosphorylated peptides resulting from protein digestion are both present in the mixture; also, proteins that bind to the Tyr phosphorylated proteins might be co-precipitated. Thus, immunoprecipitation is nowadays often performed at peptide level in order to obtain a sample that predominantly contain Tyr phosphopeptides.[79, 80] Moreover, (chemical) stable isotope labeling approaches can be implemented prior to immunoprecipitation for quantification so that potential variation introduced during sample handling can be decreased. Major disadvantages of using immunoprecipitation strategies at peptide level is the large amount of protein starting material required for an efficient enrichment and the lot-to-lot variability of the antibody.[81]

4.4 Phosphopeptide sequencing

In order to achieve site-specific phosphorylation information, it is important to be able to assign the phosphorylation to the correct residue. However, sequencing of phosphopeptides by tandem MS can induce the loss of the labile phosphate group. For instance, using CID typically results in the partial elimination of the phosphoric group (H_3PO_4, 98 Da, neutral loss) of phosphoserine and phosphothreonine. Most collision energy is then used for this fragmentation, resulting in a lower residual energy for peptide backbone fragmentation and in inadequate sequence information. This phenomenon is less common with phosphotyrosine and, generally, a characteristic immonium ion at m/z 216 in the fragment spectrum can be used as an indicator of its presence in the sequence.[77]

A few approaches can be used to overcome this issue. First, the ion originating from the loss

of phosphoric acid can be automatically selected and further fragmented by MS^3,[74] providing more phosphopeptide sequence information. Second, a multi-stage activation (MSA), also called 'pseudoMS^3', can be employed.[82] Herein, simultaneous fragmentation of the precursor ion isolated from the first MS and the theoretical neutral loss originated from the same ion occurs , combining MS^2 and MS^3 in a hybrid spectrum. Then, alternative fragmentation techniques to CID can also be employed, enhancing the dissociation of phosphopeptides at the peptide backbone rather than at the phosphoester group. One efficient fragmentation method for phosphopeptide analysis is ETD, where phosphogroups are left intact on the resulting c and z fragment ions, enabling the identification of specific phosphorylation sites.[83]

5 QUANTITATIVE PROTEOMICS

Quantification is of central importance in MS-based proteomics but also one of the most challenging tasks.[84, 85] Quantitative information comes in two forms: the absolute amount of protein in the sample, such as the determination of copy number of a protein in a cell, or the relative change in protein amount between two or more physiological states or in response to specific stimuli, drug treatments, etc. Mass spectrometry is not inherently a quantitative technique due to the fact that proteolytic peptides exhibit a wide range of physicochemical properties resulting in large differences in the MS response. However, in most proteomics workflows, quantitative information can be extrapolated in a number of ways. One major approach is based on the stable isotope dilution theory [86, 87] which states that a stable isotope-labeled peptide is chemically identical to its native counterpart during chromatographic and/or mass spectrometric analysis. Since a mass spectrometer can recognize the mass difference between labeled and unlabeled form, quantification is then performed comparing their respective signal intensities. More recently, alternative strategies have emerged, often referred to as label-free quantification.

Each quantification method has its particular strengths and weaknesses, though now they are starting to mature to an extent at which they can be meaningfully applied to biological studies on a proteomic scale. Regardless of the method chosen for labeling, the mass difference imparted by the addition of isotope labels has to be sufficiently large to distinguish the labeled and unlabeled form, or light and heavy isotope of the peptides, in the mass spectrum. Assuming that their MS response, such as ionization efficiency, is the same and that the introduction of different isotopes does not alter the chromatographic co-elution, quantification can be performed comparing their respective signal intensities.

5.1 Absolute quantification

Absolute quantification of protein amounts can be achieved by the addition of a known concentration of a stable isotope-labeled peptide (internal standard) to a protein digest and the subsequent comparison between the mass spectrometric signals of standard versus endogenous peptide in the sample.[88] To reduce interference from background ions, quantification can be performed on specific fragments of the peptide generated in the mass spectrometer using SRM or MRM.[89] In principle, absolute quantification is highly accurate and reproducible compared to label free and relative quantification. However, there are also certain limitations. First, the internal standard is introduced at a late stage in the sample preparation workflow and any sample manipulation prior to adding the synthetic peptides may bias the results. Second, the selection of appropriate internal standards requires previous identification of the endogenous peptide in the sample. Third, the synthesis of internal standards is rather expensive. In fact, this approach is attractive mostly for studies focusing on the determination of one or a few specific proteins of interest, for instance the validation of potential biomarkers in large

clinical studies.[90]

5.2 Relative quantification

For relative quantification, two main approaches are employed: stable isotope-based and label-free methods. Isotope-based methods incorporate heavy versions of specific molecules into peptides, either by chemical derivatization or by metabolic labeling.[7] The label can be introduced in the sample in different ways and at various points of the proteomics workflow, as showed in Figure 7. Since every preparative step is associated with a loss of analytes, the most accurate quantification can be expected if the label is introduced as early as possible in the workflow.

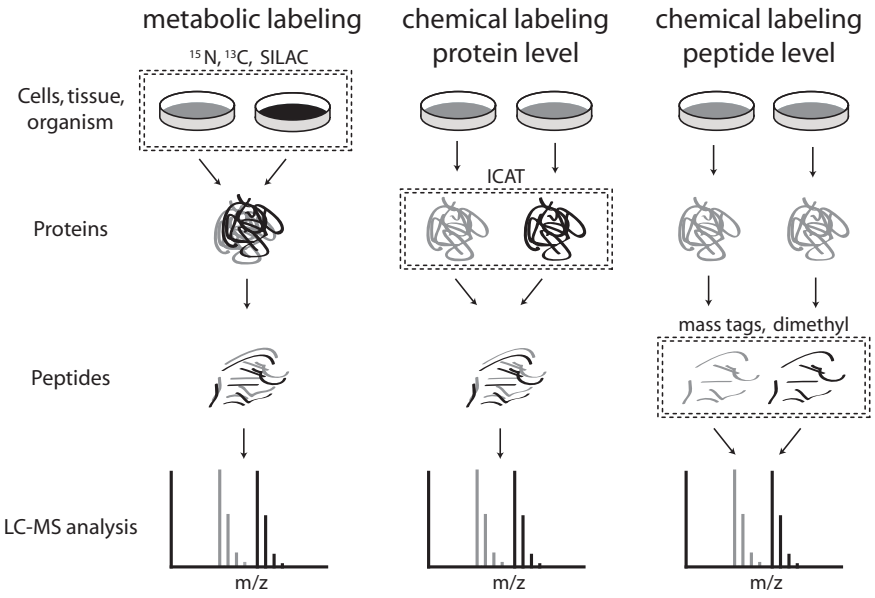

Figure 7: *Introduction of stable isotopes at different levels in the proteomics workflow, as indicated by the boxes.*

Label-free proteomics refers to the absence of an isotopic label and to the fact that quantification is performed by directly comparing MS measurements of different samples.[91] This can be achieved by comparing the signal intensity for any given peptide identified in separate runs to determine their relative abundance or employing the number of acquired spectra matching to a peptide/protein as an indication for its respective amount in a given sample. [92] An obvious advantage of label-free methods is the possibility of applying it to any type of sample. Frequently, a rough quantification is provided that will be sufficient to guide subsequent follow-up biological studies or more accurate quantitative experiments. A clear disadvantage of label-free methods is the need for several replicates and the probability of introducing errors during sample processing and analysis, especially in presence of interfering substances, such as abundant background proteins. The major difficulty is to minimize the fluctuation in LC-MS performance, above all variations of chromatographic profiles, in order to allow peak comparisons between different runs at the same retention time. Some of these limitations can be minimized by normalizing between runs with spiked-in calibrants, using abundant non-

changing peptides as landmarks between runs, and employing appropriate software for peaks matching inter-runs.[93]

5.2.1. Metabolic labeling

The earliest possible point for introducing a stable isotope signature into proteins is by metabolic labeling during cell growth and division. This is commonly realized by substituting natural ^{14}N or ^{12}C sources during cell culturing with their heavier isotopes, such as ^{15}N or ^{13}C,[86] or by substituting one or more essential amino acids with their heavy-labeled counterparts in order to obtain their full incorporation into newly synthetized proteins.[94] A principal advantage of metabolic incorporation is that the label is introduced in living cells. Thus, differentially treated samples can be already combined at the level of intact cells, before lysis, digestion and fractionation. In this way, the accuracy of quantification will not be altered by biochemical and MS procedures as these will affect both protein populations equally. Another advantage is the absence of side products or reactions, because the incorporation of isotopes is highly specific. However, the cost and time required for creating and maintaining these systems are often quite high, limiting their use in routine experiments.

Stable isotope labeling by amino acids in cell culture (SILAC)[94] is a metabolic labeling strategy that has proven to be a simple and powerful approach in quantitative proteomics. Cells are grown in minimal culture conditions that lack essential amino acids. Stable isotope containing versions of these amino acids are then provided in growth media in order to obtain in vivo incorporation. After several cell doublings, the complete cellular proteome will contain the supplied amino acid. Several amino acids can be used in the SILAC approach, however arginine and lysine are commonly used when combined with trypsin digestion since every peptide will be labeled at the carboxyl-terminus (except the one corresponding to C-terminal peptide).

Metabolic labeling is generally performed on cell systems that are grown in culture, although fruit flies,[95] plants[96] and rats[97] have, amongst other, also been isotopically labeled by supplying them with isotopically labeled sources.

5.2.2. Chemical labeling

When metabolic labeling cannot be applied, for example in case of human biopsies in biomedical research, a post-biosynthetic labeling approach, such as chemical labeling, is the method of choice. In chemical modification-based approaches, isotope-bearing chemical reagents are directed towards reactive sites on peptides or proteins. With the possibility to target multiple functional groups in the polypeptide chain, a huge number of chemical-labeling methods have been reported.[7] Notably, only a few have been applied in real experiments due to non-specific and incomplete labeling. In fact, even small amounts of side reactions would lead to an increase in sample complexity and a tremendous complication of the MS interpretation.

The prototypical example of chemical labeling is the isotope-coded affinity tag (ICAT), which is used for labeling at protein level.[87] The chemical reagent consists of a reactive group that is cysteine-directed, a polyether linker region with up to eight deuteriums, and a biotin group that allows recovery of labeled species. During a classical ICAT experiment, protein samples are denaturated, reduced and modified with the respective ICAT reagents. Samples can be pooled prior to proteolysis, avoiding biases introduced during sample digestions. However, some practical issues limited the use of ICAT, especially its applicability for only cysteine-containing proteins.

Other common labeling techniques target the primary amine groups (lysine and amino termini) of proteins/peptides, and can be specific and largely complete. One example is stable-isotope dimethyl labeling,[98] where triplex dimethyl labeling can be quickly generated by using a com-

bination of isotopomers of formaldehyde and cyanoborohydride, carrying deuterated and ^{13}C atoms. This approach is generally applied to tryptic peptides resulting in the introduction of dimethyl labels differing in at least 4 Da in their mass. To minimize sample loss, peptides can be labeled during solid phase extraction or directly online in the LC-MS setup, combining several sample handling steps.[99] This approach is particularly cost-friendly, with no limitations regarding the amount of sample that is to be labeled, and is applicable to any biological experiment. However, one of the major issues associated with deuterium containing tags (dimethyl labeling as well as other methods) is the chromatographic behavior in reversed phase or other LC modes. Deuterium displays subtle chemical differences from hydrogen, resulting in retention time shifts during LC that can affect the quantification.[100] This topic will be largely discussed in chapter 5.

Carboxylic acids in side chains of glutamic and aspartic acids, as well as C-termini of peptides, can be isotopically labeled by esterification employing deuterated alcohols.[101] This chemical approach is particularly interesting for quantification of phosphopeptides because esterification blocks the acidic groups preventing their unspecific binding to ion metals while purifying phosphopeptides on IMAC columns.[68] In this way, the specificity of phosphoenrichment can be improved.

Another method that has gained popularity is the use of isobaric tags, implemented into two main techniques called isobaric tag for relative and absolute quantification (iTRAQ)[102] and tandem mass tag (TMT).[103] These approaches use N-hydroxy-succinimmide chemistry to link the isobaric molecule to the primary amine of a peptide but add an innovative concept: a tag that generates a specific reporter ion in fragmentation spectra. In fact, these labels consist of a reporter moiety, a balance group to counterbalance the differential weight of the reporter, and an ester to link the peptide. In MS, differentially labeled peptides cannot be distinguished because they are isobaric and will not further increase the MS spectral complexity. But during MS/MS the reporter ions is cleaved off and those signals are then used to determine the relative abundance of each labeled sample. Moreover, this multiplexing strategy is available in 4- or 8-plex formats for iTRAQ,[104] and 6-plex for TMT,[105] which allows the analysis of multiple samples in a single analysis, increasing analytical throughput. However, an important requirement for isobaric tags approach is a good chromatographic separation, as co-eluting peptides of similar mass would contribute to the same reporter ion. Furthermore, specific MS instrumentation is required to observe reporter ions at low m/z, typically TOF instruments or current generation of Orbitrap mass spectrometers with HCD capabilities.[106]

6. OUTLINES OF THE THESIS

In the last decade, the field of proteomics has rapidly progressed with substantial advances in many aspects of the proteomics platform, particularly nanoLC separation, MS instrumentation and bioinformatics tools. However, significant improvements are still needed to generate proteome-wide data for limited number of cells. By enhancing the sensitivity of the LC-MS analysis, identification of lower abundant proteins can be achieved. The work described in this thesis highlights some relevant advances that can help accelerate proteomics towards a high level of depth in regard to proteome coverage and how these achievements can find ample application in several research lines.

In chapter 2, recent advances in peptide separation by multidimensional liquid chromatography are reviewed. The most common LC-based techniques employed in proteomics are reversed-phase (RP), ion-exchange (IEX) and hydrophilic interaction liquid chromatography (HILIC). A detailed overview of these separations, as well their combination in multidimensional formats, are provided. Moreover, various instrumental setups and examples of their applicability in

proteomics research are thoroughly described.

In chapter 3, we introduce and evaluate the use of HILIC based strategies for the separation of complex peptide mixtures. Two zwitterionic stationary phases, ZIC-HILIC and ZIC-cHILIC, differing in the spatial orientation of their charged groups on the chromatographic material, are fully characterized with respect to separation efficiency and a number of peptide physico-chemical properties affecting peptide retention. We extensively tested the performances of these HILIC materials as first dimension in an off-line 2D-LC strategy, in combination with RP, for the analysis of low amounts (a few micrograms) of human cell digests. The protein identifications observed from such a level of material demonstrates that HILIC can rival traditional multidimensional strategies employed in proteomics.

In chapter 4, we apply the 2D ZIC-cHILIC-RP strategy to the analysis of a limited number of FACS-sorted colon stem cells. With an optimized sample preparation and workflow, we enabled for the first time the in-depth global proteome analysis of 10,000 colon stem cells, obtaining a high proteome coverage.

In chapter 5, we describe the feasibility of the dimethyl labeling method in combination with ZIC-cHILIC technology for quantitative proteomics. We address the potential issue of isotope effects perturbing the essential co-elution of differently labeled peptides under ZIC-cHILIC separation. The deuterium effect can be largely eliminated by choosing appropriate pH conditions which favors a mixed-mode ZIC-cHILIC separation based on combined hydrophilic and ionic interactions. This optimized approach is a suitable quantitative strategy, resulting in the quantification of thousands of proteins with negligible biases.

In chapter 6, we combine the HILIC separation with a phosphopeptide enrichment approach based on Ti^{4+}-IMAC to develop an efficient approach with the aim of maximizing the coverage of the cellular phosphoproteome. We design and systematically compare three strategies including: a sole Ti^{4+}-IMAC enrichment (1D); Ti^{4+}-IMAC enrichment followed by HILIC fractionation (2D); a pre-fractionation based on strong cation exchange, followed by Ti^{4+}-IMAC enrichment and a further step of fractionation by HILIC (3D). This work demonstrates the need to carry out extensive fractionation for deep mining of the phosphoproteome and shows that the choice of an appropriate analytical strategy mainly depends on sample amount and/or desirable analysis time.

In chapter 7, we summarize our expertise in the HILIC separation field into a protocol for sensitive shotgun proteomics. A few technical hurdles need to be addressed in order to implement HILIC in a highly sensitive proteomics workflow for the analysis of limited amounts of starting material. In the protocol, we report on our latest robust set-up for a ZIC-cHILIC chromatographic separation in a two-dimensional format. We give advice and propose practical solutions to issues that can be encountered during the development of a HILIC-based analytical strategy.

7. REFERENCES

[1] Pandey A, et al. Proteomics to study genes and genomes. Nature. 2000;405:837-46.
[2] James P. Protein identification in the post-genome era: the rapid rise of proteomics. Quarterly Reviews of Biophysics. 1997;30:279-331.
[3] Anderson NL, et al. Proteome and proteomics: New technologies, new concepts, and new words. Electrophoresis. 1998;19:1853-61.
[4] Wilkins MR, et al. From proteins to proteomes: Large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Bio-Technology. 1996;14:61-5.
[5] Collura V, et al. From protein-protein complexes to interactomics. Sub-cellular biochemistry. 2007;43:135-83.
[6] Mann M, et al. Proteomic analysis of post-translational modifications. Nat Biotech. 2003;21:255-61.
[7] Ong SE, et al. Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol. 2005;1:252-62.
[8] Gygi SP, et al. Correlation between protein and mRNA abundance in yeast. Molecular and Cellular Biology. 1999;19:1720-30.
[9] Hegde PS, et al. Interplay of transcriptomics and proteomics. Current Opinion in Biotechnology. 2003;14:647-51.
[10] Aebersold R, et al. Mass spectrometry-based proteomics. Nature. 2003;422:198-207.
[11] Yates JR, et al. Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. Annual Review of Biomedical Engineering2009. p. 49-79.
[12] Domon B, et al. Options and considerations when selecting a quantitative proteomics strategy. Nat Biotechnol. 2010;28:710-21.
[13] Nesvizhskii AI, et al. Analysis and validation of proteomic data generated by tandem mass spectrometry. Nature Methods. 2007;4:787-97.
[14] Cox J, et al. Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. Annual Review of Biochemistry. 2011;80:273-99.
[15] Mallick P, et al. Proteomics: a pragmatic perspective. Nat Biotechnol. 2010;28:695-709.
[16] Olsen JV, et al. A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. Molecular & Cellular Proteomics. 2009;8:2759-69.
[17] Makarov A, et al. Coupling liquid chromatography to Orbitrap mass spectrometry. Journal of Chromatography A. 2010;1217:3938-45.
[18] Olsen JV, et al. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Molecular & Cellular Proteomics. 2004;3:608-14.
[19] Taouatas N, et al. Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase. Nature Methods. 2008;5:405-7.
[20] Di Palma S, et al. Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. Journal of Proteomics. 2012;75:3791-813.
[21] Andersson L, et al. Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography. Analytical Biochemistry. 1986;154:250-4.
[22] Pinkse MWH, et al. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-nanoLC-ESI-MS/MS and titanium oxide precolumns. Analytical Chemistry. 2004;76:3935-43.
[23] Kweon HK, et al. Selective Zirconium Dioxide-Based Enrichment of Phosphorylated Peptides for Mass Spectrometric Analysis. Analytical Chemistry. 2006;78:1743-9.
[24] McDonald WH, et al. Shotgun proteomics: Integrating technologies to answer biological questions. Curr Opin Mol Ther. 2003;5:302-9.
[25] Neilson KA, et al. Less label, more free: Approaches in label-free quantitative mass spectrometry. Proteomics. 2011;11:535-53.
[26] Jaffe JD, et al. Accurate Inclusion Mass Screening. Molecular & Cellular Proteomics. 2008;7:1952-62.

[27] Schmidt A, et al. Directed mass spectrometry: towards hypothesis-driven proteomics. Current Opinion in Chemical Biology. 2009;13:510-7.
[28] Lange V, et al. Selected reaction monitoring for quantitative proteomics: a tutorial. Molecular Systems Biology. 2008;4.
[29] Gallien S, et al. Selected reaction monitoring applied to proteomics. J Mass Spectrom. 2011;46:298-312.
[30] Picotti P, et al. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. Nature Methods. 2010;7:43-U5.
[31] Addona TA, et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. Nat Biotechnol. 2009;27:633-U85.
[32] Geiger T, et al. Proteomics on an Orbitrap Benchtop Mass Spectrometer Using All-ion Fragmentation. Molecular & Cellular Proteomics. 2010;9:2252-61.
[33] Geromanos SJ, et al. The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. Proteomics. 2009;9:1683-95.
[34] Venable JD, et al. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. Nature Methods. 2004;1:39-45.
[35] Panchaud A, et al. Precursor Acquisition Independent From Ion Count: How to Dive Deeper into the Proteomics Ocean. Analytical Chemistry. 2009;81:6481-8.
[36] Gillet LC, et al. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. Molecular & Cellular Proteomics. 2012;11.
[37] Ahrens CH, et al. Generating and navigating proteome maps using mass spectrometry. Nature Reviews Molecular Cell Biology. 2010;11:789-801.
[38] Karas M, et al. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Analytical Chemistry. 1988;60:2299-301.
[39] Tanaka K. The origin of macromolecule ionization by laser irradiation (Nobel lecture). Angew Chem-Int Edit. 2003;42:3860-70.
[40] Fenn JB, et al. Electrospray ionization for mass spectrometry of large biomolecules. Science. 1989;246:64-71.
[41] Li K-Y, et al. Charge Limits on Droplets during Evaporation. Langmuir. 2005;21:3786-94.
[42] Iribarne JV, et al. On the evaporation of small ions from charged droplets. The Journal of Chemical Physics. 1976;64:2287-94.
[43] Dole M, et al. Molecular Beams of Macroions. The Journal of Chemical Physics. 1968;49:2240-9.
[44] Whitehouse CM, et al. Electrospray interface for liquid chromatographs and mass spectrometers. Analytical Chemistry. 1985;57:675-9.
[45] Wilm M, et al. Analytical properties of the nanoelectrospray ion source. Analytical Chemistry. 1996;68:1-8.
[46] de Hoffmann E, et al. Mass Spectrometry: Principle and Applications. Wiley. 2007;3rd edition.
[47] Paul W, et al. Ein neues Massenspektrometer ohne Magnetfeld. Journal of Physical Sciences. 1953;8:448-50.
[48] Hager JW, et al. High-performance liquid chromatography-tandem mass spectrometry with a new quadrupole/linear ion trap instrument. Journal of Chromatography A. 2003;1020:3-9.
[49] Douglas DJ, et al. Linear ion traps in mass spectrometry. Mass Spectrometry Reviews. 2005;24:1-29.
[50] Marshall AG, et al. Fourier transform ion cyclotron resonance detection: principles and experimental configurations. Int J Mass Spectrom. 2002;215:59-75.
[51] Makarov A. Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. Analytical Chemistry. 2000;72:1156-62.
[52] Michalski A, et al. Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap

Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes. Molecular & Cellular Proteomics. 2012;11.
[53] Makarov A, et al. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. Analytical Chemistry. 2006;78:2113-20.
[54] McLuckey SA. Principles of collisional activation in analytical mass spectrometry. J Am Soc Mass Spectrom. 1992;3:599-614.
[55] Paizs B, et al. Fragmentation pathways of protonated peptides. Mass Spectrometry Reviews. 2005;24:508-48.
[56] Sleno L, et al. Ion activation methods for tandem mass spectrometry. J Mass Spectrom. 2004;39:1091-112.
[57] Roepstorff P, et al. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomedical Mass Spectrometry 1984;11:601-.
[58] Frank AM. Predicting Intensity Ranks of Peptide Fragment Ions. Journal of Proteome Research. 2009;8:2226-40.
[59] Syka JEP, et al. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A. 2004;101:9528-33.
[60] Swaney DL, et al. Decision tree-driven tandem mass spectrometry for shotgun proteomics. Nat Meth. 2008;5:959-64.
[61] Frese CK, et al. Improved Peptide Identification by Targeted Fragmentation Using CID, HCD and ETD on an LTQ-Orbitrap Velos. Journal of Proteome Research. 2011;10:2377-88.
[62] Nesvizhskii AI. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. Journal of Proteomics. 2010;73:2092-123.
[63] Perkins DN, et al. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999;20:3551-67.
[64] Higgs RE, et al. Estimating the Statistical Significance of Peptide Identifications from Shotgun Proteomics Experiments. Journal of Proteome Research. 2007;6:1758-67.
[65] Weatherly DB, et al. A Heuristic Method for Assigning a False-discovery Rate for Protein Identifications from Mascot Database Search Results. Molecular & Cellular Proteomics. 2005;4:762-72.
[66] Graves JD, et al. Protein phosphorylation and signal transduction. Pharmacology & Therapeutics. 1999;82:111-21.
[67] Thingholm TE, et al. Analytical strategies for phosphoproteomics. Proteomics. 2009;9:1451-68.
[68] Ficarro SB, et al. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. Nat Biotech. 2002;20:301-5.
[69] Posewitz MC, et al. Immobilized gallium(III) affinity chromatography of phosphopeptides. Analytical Chemistry. 1999;71:2883-92.
[70] Kokubu M, et al. Specificity of Immobilized Metal Affinity-Based IMAC/C18 Tip Enrichment of Phosphopeptides for Protein Phosphorylation Analysis. Analytical Chemistry. 2005;77:5144-54.
[71] Larsen MR, et al. Highly Selective Enrichment of Phosphorylated Peptides from Peptide Mixtures Using Titanium Dioxide Microcolumns. Molecular & Cellular Proteomics. 2005;4:873-86.
[72] Pinkse MWH, et al. Highly Robust, Automated, and Sensitive Online TiO 2-Based Phosphoproteomics Applied To Study Endogenous Phosphorylation in Drosophila melanogaster. Journal of Proteome Research. 2007;7:687-97.
[73] Thingholm TE, et al. SIMAC (Sequential Elution from IMAC), a Phosphoproteomics Strategy for the Rapid Separation of Monophosphorylated from Multiply Phosphorylated Peptides. Molecular & Cellular Proteomics. 2008;7:661-71.
[74] Beausoleil SA, et al. Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci U S A. 2004;101:12130-5.
[75] Zarei M, et al. Comparison of ERLIC–TiO2, HILIC–TiO2, and SCX–TiO2 for Global Phospho-

proteomics Approaches. Journal of Proteome Research. 2011;10:3474-83.
[76] McNulty DE, et al. Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. Molecular & Cellular Proteomics. 2008;7:971-80.
[77] Steen H, et al. Tyrosine Phosphorylation Mapping of the Epidermal Growth Factor Receptor Signaling Pathway. J Biol Chem. 2002;277:1031-9.
[78] Pandey A, 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. Proceedings of the National Academy of Sciences. 2000;97:179-84.
[79] Zhang Y, et al. Time-resolved Mass Spectrometry of Tyrosine Phosphorylation Sites in the Epidermal Growth Factor Receptor Signaling Network Reveals Dynamic Modules. Molecular & Cellular Proteomics. 2005;4:1240-50.
[80] Rush J, et al. Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. Nat Biotech. 2005;23:94-101.
[81] Boersema PJ, et al. In-depth Qualitative and Quantitative Profiling of Tyrosine Phosphorylation Using a Combination of Phosphopeptide Immunoaffinity Purification and Stable Isotope Dimethyl Labeling. Molecular & Cellular Proteomics. 2010;9:84-99.
[82] Schroeder MJ, et al. A Neutral Loss Activation Method for Improved Phosphopeptide Sequence Analysis by Quadrupole Ion Trap Mass Spectrometry. Analytical Chemistry. 2004;76:3590-8.
[83] Boersema PJ, et al. Phosphopeptide fragmentation and analysis by mass spectrometry. J Mass Spectrom. 2009;44:861-78.
[84] Heck AJ, et al. Mass spectrometry-based quantitative proteomics. Expert Rev Proteomics. 2004;1:317-26.
[85] Bantscheff M, et al. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Analytical and Bioanalytical Chemistry. 2012;404:939-65.
[86] Oda Y, et al. Accurate quantitation of protein expression and site-specific phosphorylation. Proceedings of the National Academy of Sciences. 1999;96:6591-6.
[87] Gygi SP, et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol. 1999;17:994-9.
[88] Gerber SA, et al. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proceedings of the National Academy of Sciences. 2003;100:6940-5.
[89] Kirkpatrick DS, et al. The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. Methods. 2005;35:265-73.
[90] Pan S, et al. High Throughput Proteome Screening for Biomarker Detection. Molecular & Cellular Proteomics. 2005;4:182-90.
[91] Wang WX, et al. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. Analytical Chemistry. 2003;75:4818-26.
[92] Chelius D, et al. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. Journal of Proteome Research. 2002;1:317-23.
[93] Gillette MA, et al. Place of Pattern in Proteomic Biomarker Discovery†. Journal of Proteome Research. 2005;4:1143-54.
[94] Ong S-E, et al. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. Molecular & Cellular Proteomics. 2002;1:376-86.
[95] Krijgsveld J, et al. Metabolic labeling of C. elegans and D. melanogaster for quantitative proteomics. Nat Biotech. 2003;21:927-31.
[96] Gruhler A, et al. Stable Isotope Labeling of Arabidopsis thaliana Cells and Quantitative Proteomics by Mass Spectrometry. Molecular & Cellular Proteomics. 2005;4:1697-709.
[97] Wu CC, et al. Metabolic Labeling of Mammalian Organisms with Stable Isotopes for Quantitative

Proteomic Analysis. Analytical Chemistry. 2004;76:4951-9.
[98] Hsu JL, et al. Stable-isotope dimethyl labeling for quantitative proteomics. Anal Chem. 2003;75:6843-52.
[99] Boersema PJ, et al. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protoc. 2009;4:484-94.
[100] Zhang R, et al. Controlling deuterium isotope effects in comparative proteomics. Anal Chem. 2002;74:3662-9.
[101] Goodlett DR, et al. Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation. Rapid Communications in Mass Spectrometry. 2001;15:1214-21.
[102] Ross PL, et al. Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. Molecular & Cellular Proteomics. 2004;3:1154-69.
[103] Thompson A, et al. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. Analytical Chemistry. 2003;75:1895-904.
[104] Pierce A, et al. Eight-channel iTRAQ Enables Comparison of the Activity of Six Leukemogenic Tyrosine Kinases. Molecular & Cellular Proteomics. 2008;7:853-63.
[105] Dayon L, et al. Relative Quantification of Proteins in Human Cerebrospinal Fluids by MS/MS Using 6-Plex Isobaric Tags. Analytical Chemistry. 2008;80:2921-31.
[106] Bantscheff M, et al. Robust and Sensitive iTRAQ Quantification on an LTQ Orbitrap Mass Spectrometer. Molecular & Cellular Proteomics. 2008;7:1702-13.

Chapter 2

A review: Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis

Serena Di Palma*, Marco L. Hennrich*, Albert J. R. Heck, and Shabaz Mohammed
*These authors contributed equally to this work

J Proteomics. 2012 Jul 16;75(13):3791-813.

Abstract

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

TABLE OF CONTENTS

1. INTRODUCTION

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

However, the preferred handling of peptides over proteins faces challenges concerning the higher complexity of the proteome sample and the wider dynamic range, where thousands of peptides, with very similar mass to charge (m/z) ratios, are present at different levels of abundance.[13] Considering that the human proteome (containing approximately 20,000 genes) yields millions of peptides, it becomes clear that this task is not a trivial exercise.[14] Therefore, methods exhibiting high resolving power are required to maximize the separation of peptides before the mass spectrometric analysis. Improved resolving power can help decrease competition between peptides during ionization and reduce the likelihood of ion suppression, increasing the possibility for the detection of low-abundance peptides (and the protein of origin).

Not surprisingly, despite the rapid and massive advances in separation and instrument technology, current analytical strategies are still not sufficient to fully resolve complex biological samples.[15-18] Most approaches in proteomics end with peptide sequencing via nanolitre flow rate reversed phase chromatography and tandem mass spectrometry (nanoLC-MS). This step is relatively fast, reproducible and effective but its use in proteomics is hampered by the limited capability to resolve highly complex samples, such as cell lysate digests. In fact, to date, no separation method is capable of resolving all components in a single analytical dimension.[19, 20] Therefore, the combination of two or more orthogonal separation procedures is required to increase the peak capacity, and thus the overall resolving power.[21] In the last few years, several multidimensional chromatographic methods, coupled in either on-line or off-line mode, have been introduced to enhance in-depth proteome analysis and these approaches are now considered routine parts of the shotgun proteomics pipeline. The obvious need for separation/fractionation is to minimize the number of co-eluting peptides introduced into the MS at any time during the analysis, allowing the detection of a greater number of peptides.

The aim of this review is to highlight some of the advances and new developments that have been made in the area of liquid chromatographic-based separations. Electrophoretic based separations are covered in an accompanying review in this issue.[22] We will discuss, in-depth, each separation methodology, especially their roles in hyphenated multidimensional formats and what is required to maximize the resolution, reduce sample complexity, widen the overall

dynamic range and consequently increase the proteome coverage.

1.1 Orthogonality and peak capacity

Amino acids possess a number of physicochemical properties that can be used as the basis for a mode of separation. Peptides in a mixture greatly differ in their physicochemical properties and so separation of peptides can involve differences in polarity (for reversed phase, RP, and hydrophilic interaction liquid chromatography, HILIC), charge (ion exchange chromatography, IEX) or other peptide properties such as isoelectric point (isoelectric focusing). Some separations employ more than one property for selectivity, and a mixed-mode separation can be achieved when different mechanisms of interaction interplay. Electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) and HILIC with zwitterionic materials are two examples of chromatographic separations where the selectivity for peptides is based on the combination of charge and hydrophilic interactions. Figure 1 illustrates schematically the selectivity of six different types of stationary phases on the basis of polarity, hydrophobicity and charge (negative and positive) of peptides, anticipating their orthogonality in multidimensional combinations.[23] One can expect that ion exchange chromatography, which separates peptides based on charge and only partly on hydrophobicity, is more orthogonal to RP than the combination RP-RP in a 2D strategy which is mainly related to hydrophobicity in both dimensions. Along the same lines HILIC, employing a zwitterionic stationary phase, or ERLIC (a HILIC variant), using anion exchange stationary phase, can potentially display a higher orthogonality towards RP than anion or cation exchange since the peptide selectivity is driven by a combination of polar and ionic interactions.

Giddings first formalized the concept of multidimensional chromatographic separations in 1984.[24] When two separations systems based on different (non-correlative) retention mechanisms are coupled, the resulting 2D system has a higher resolving power than each single dimension. The resolving power of any chromatographic technique is measured by the peak capacity, which is defined as the maximum number of peaks that can be separated within a gradient time by a specific system.[25] The peak capacity in multidimensional strategies can be maximized by combining methods that separate peptides on the basis of different physicochemical properties; the final peak capacity for 2D separation results from the peak capacities in both separation dimensions[26]. Theoretically, the overall separation power becomes the product of each separation dimension, effectively creating a second opportunity for co-eluting analytes to be resolved in the orthogonal second dimension. In reality, several practical considerations limit the achievable peak capacity in 2D-LC. First of all, a situation of perfect orthogonality can rarely be achieved due to the fact that existing LC separation methods display a mixture of interactions with the stationary phase, increasing the overlap between different techniques.[23] Second, the inefficient transfer of separated zones from one dimension to the other can produce the remixing of separated peaks,[24, 26] resulting in a lower peak capacity than expected.

1.2 On-line and off-line set-ups

A multidimensional separation can be carried out in an on-line fashion, which involves a direct transfer of the eluent from the first dimension onto the next one, with no flow interruption;[27] or in an off-line way, based on fractions' collection in the first dimension and their analysis in the following dimension.[28] The collection of multiple fractions from an efficient separation/fractionation in the first dimension, in concert with a fast and comprehensive analysis in the second LC dimension, is a crucial step. When this step is carefully optimized, it can offer a significant advantage in terms of analysis time, reducing number of fractions and

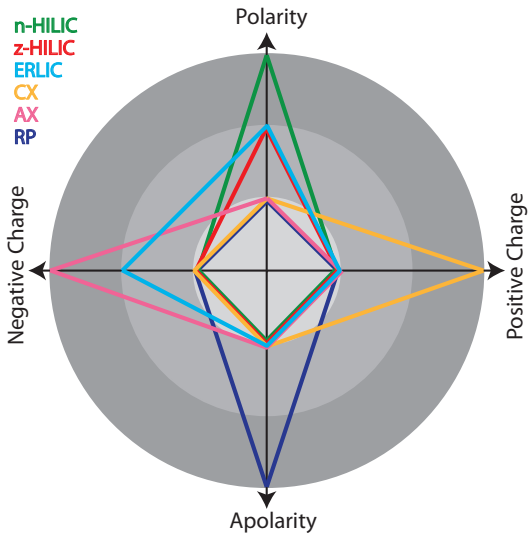

Figure 1: *Schematic of the selectivity of neutral HILIC (n-HILIC), zwitterionic HILIC (z-HILIC), ERLIC, cation exchange (CX), anion exchange (AX) and reversed phase (RP) chromatography. The light inner circle indicates low and the dark outer circle high selectivity towards positive charge, negative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

content of overlap between adjacent fractions.

Off-line set-ups are often more simple in design and operation, and each dimension can be fully and independently optimized.[29] Off-line approaches offer greater flexibility than on-line separations in the choice of LC modes, especially in the case of incompatibility of mobile phases between consecutive dimensions. In addition, fractions can be manipulated, i.e., diluted, concentrated or dissolved in a different solvent, chemically modified and, if necessary, reanalyzed. But the major limitations of off-line set-ups are related to higher risk of sample loss and contamination, due to sample handling, and lack of automation of the system. However, performing an experiment in an off-line fashion is not necessarily inferior, and, in real life, is the most common and favorable approach since it allows more combinations than on-line approaches with a uncompromised optimization of each single dimension.

When using on-line set-ups, sample handling is notably reduced since there is continuity in the

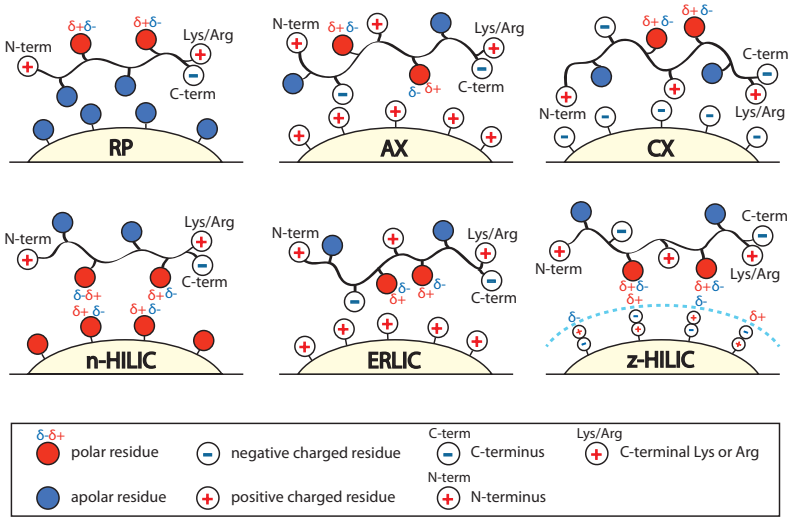

Figure 2: *Schematic illustration of interactions between polar, apolar, negatively charged and positively charged sites of a tryptic peptide and different stationary phases.*

sample transfer between consecutive dimensions realized via the use of switching valves, additional pumps and trapping columns. On-line configurations are usually designed to avoid dead volume and prevent sample loss, although the sample can still be exposed to additional column connections that may lead to carryover or loss.[30, 31] On-line hyphenation possesses certain advantages compared to off-line designs, for instance it can decrease the overall analysis time. However, it has also more stringent requirements, such as the solvent used for the elution in the first dimension must be a weak eluent in the second dimension, and the second dimension needs to be relatively fast in order not to lose the resolution achieved in the first dimension. In addition, whatever the on-line or off-line set-up, the column used in the second dimension must have a small internal diameter in order not to compromise the sensitivity of the mass spectrometric detection.

1.3 Multidimensional separation combined with mass spectrometry

The successful combination of multidimensional separations with mass spectrometry for proteins and peptides analysis was achieved with the advent of the soft ionization techniques MALDI[32, 33] and ESI,[34-36] which paved the way for modern bench-top MS proteomics, assisted by the continuous development of new powerful MS instrumentations.[37] While MALDI is usually combined with gel-based separations, high-pressure liquid chromatography(HPLC) separations are more frequently coupled to instruments with an ESI source requiring a continuous flow. This has allowed HPLC to become a standard front end for many proteomics applications, giving rise to several LC-MS set-ups.[5, 21, 38, 39]

Although MS technology has and will continue to rapidly improve with the advent of higher speed, higher resolution and more sensitive instrumentation,[40, 41] even the best dynamic range capabilities tolerated by MS are lower than the wide dynamic range of proteins present in samples. Thus, the deeper and comprehensive characterization of proteomes still largely relies on efficient separation technologies prior to MS. Orbitrap and QToF instrumentations, in ideal circumstances, can observe over approx. 4 orders of dynamic range. If in the same scan there are peptides exhibiting a higher dynamic range, the peptide that is of lower abundance, often, will not be observed. If the LC can separate the high and low abundance peptides, then there is more chance for the mass spectrometer to successfully observe and sequence the lower abundance peptide.

Several chromatographic techniques have been employed in 2D-LC configurations, whereby the second dimension is generally performed by RP[42, 43] due to its better compatibility with electrospray ionization (ESI) MS,[44] high resolving power and the advantageous sample desalting when the first dimension requires salt gradients. The vast majority of 2D-LC analyses implemented today utilize SCX coupled to RP in both on-line[45] and off-line[46] modes, mainly because of the good orthogonality of these two separations. Other 2D-LC strategies have been emerging in recent years as promising alternatives to this combination, including size exclusion chromatography (SEC),[47] affinity purification chromatography (AFC),[48] different types or combinations of ion exchangers,[49, 50] anion and cation mixed-bed exchange,[51] and HILIC. [17, 23, 52-54]

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

2. REVERSED-PHASE CHROMATOGRAPHY

Reversed-phase chromatography was introduced as a peptide separation method in 1976.[55] Though other types of chromatographic separations have undergone significant developments over the last decade, reversed-phase (RP) chromatography is still the most widely used method for sample preparation and, as well, separation of peptides prior to MS analysis.

The separation principle of RP is based on the partitioning of the analytes between a hydrophobic stationary phase (which can be visualized as a hydrophobic solvent) and a polar hydrophilic mobile phase. Peptides are loaded onto a RP column under low-organic-solvent conditions, which cause the peptides to partition into the RP material. Salts and the majority of components used in digestion protocols tend to remain in the low organic solvent and, thus, RP is often used to 'desalt' and concentrate the sample. Separation (or elution of the peptide) is then achieved by increasing the organic modifier content in the mobile phase and, if the mobile phase is of sufficient hydrophobicity, the peptide will start to partition between the two phases as it moves down the column. The order of the eluting peptides relates to the strength of their hydrophobic interactions with the stationary phase. The most common organic modifier employed in RP-LC-MS for peptides separation is acetonitrile, since it offers the best compromise in hydrophobicity, volatility, viscosity and MS compatibility.[56]

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

Nowadays, the RP stationary phases used most often are C18 resins, silica derivatized to possess an alkane chain containing 18 carbons, which can also be referred to as an octadecasilyl-group (ODS). Columns are capillary-scale (internal diameters are below 100 μm) and combined with nano-ESI-MS.[30, 36, 57, 58] The mobile phase employed in RP (a mixture of acetonitrile and water with the acidic ion-pairing reagent) has an excellent compatibility with ESI-MS.[44] The efficiency of the electrospray process is dictated by the size of the initial droplet size, which is dependent on the flow rate and the internal/external diameter of the electrospray needle.[36, 59] One of the most important parameters that affects separation efficiency is the linear velocity (cm/s), as a consequence of the van Deemter equation. Thus, there is a drive to reduce flow rates for improving the electrospray process and, in order to maintain separation efficiency (the linear velocity), the column i.d.s are reduced (see below for more details).

Significant efforts have been made to increase the resolving power,[60] sensitivity[61-64] and speed of analysis[65] in single dimension RP separation. Several parameters are of special interest in order to achieve this goal, such as length and inner diameter (i.d.) of the column,

velocity of the solvent, flow rate, particle size, temperature and pressure. Of these parameters, the choice of solvents and the range of flow rate are partially limited due to the hyphenation with ESI-MS. In contrast, the column length and i.d., and the particle size can be varied and have a strong influence on the resolving power. In theory, increasing the length of the column or reducing the particle size will lead to an increase in separation power, although both will be at the cost of an increase in pressure. Therefore, the pressure often represents the limiting factor.[15] The majority of experiments are still performed at room temperature, although it is well known that higher temperatures decrease the viscosity of the mobile phase, leading to a decrease in pressure. This can facilitate the use of longer columns and/or smaller particles, improving the separation power.[66]

A measure of the separation power is the peak capacity, described as the maximum number of peaks that can theoretically be separated in a defined system. It is calculated by dividing the gradient time by the peak width at 4σ (= peak width at 13.4% height). State-of-the-art RP columns provide a peak capacity of several hundred,[19, 42, 67] depending on their length and the gradient slope. A predicted maximum achievable peak capacity in single-dimensional (1D) RP-LC is within the range of 1400-1600.[68]

Various strategies have been proposed for LC peak capacity improvements in one-dimensional approaches: (i) decreasing the gradient slope (when column length is fixed), (ii) increasing the column length (L) with proportional increase in gradient time, and (iii) employing columns with smaller internal diameter and packed with smaller sorbent particles.[68] The first two strategies have diminishing returns, because the gains in peak capacity are not linear with the increase in analysis time and column length.[19] The third strategy could be more powerful since sufficiently small i.d. columns can provide higher separation efficiencies as well as optimized ESI sensitivity.[42] However, their use presents several technical issues that have to date effectively precluded the routine use of <50 μm i.d. LC columns. First of all, the fabrication of columns with i.d.'s less than 30 μm is extremely difficult. One reason is the agglomeration of particles, which can prevent successful packing.[42] Effective preparation of small i.d. columns requires smaller diameter particles (below 3 μm) that are highly uniform. Since nonporous particles have lower particle size deviation compared to porous particles, they represent the material of choice when very small i.d.s are chosen, at the cost of reduced sample loading capacity. The second major obstacle is the increased back pressure during the packing procedure. In general, the back pressure created by a column is inversely proportional to the square of the packing material diameter,[69] thus smaller-diameter columns can easily exceed the maximum operating pressure of conventional HPLC systems (400 bar). The prospects of increasing the separation power using smaller particles was the main driving force to develop high pressure systems capable of working in the range of 1000 bar. Special pumping equipment is required to operate such systems, called ultrahigh-pressure liquid chromatography (UHPLC). We refer to the excellent review of Jorgenson for more details.[70] Even though the use of columns with small i.d.s and small particles working at high pressures is not yet standard, several approaches have proven high potential in terms of separation power.[19, 71] Already, more than ten years ago, the group of Jorgenson used 30 μm i.d. columns of up to 66 cm length packed with 1.5 μm nonporous C18 silica based particles, resulting in highly efficient separations. In order to achieve an optimum flow rate for the separation, they chose to operate the system at approximately 1400 bar. Applying the maximum pressure achievable (up to 4100 bar) reduced the analysis time, but partly compromised the separation power.[72] In a follow-up study, they used 33 μm i.d. columns with a 27 cm length, packed with 1.0 μm nonporous C18 silica based particles, and demonstrated a peak capacity of 300 for a 30 min analysis of a tryptic ovalbumin digest.[73] A peak capacity of over 1000 has been shown by Shen et al. after 180 min analysis

of a tryptic yeast protein digest on a 87 cm x 19.8 μm i.d. column packed with 3 μm particles with a pore size of 300 Å.[42] A few years later the same group analyzed a tryptic digest of Shewanella oneidensis and reached a peak capacity of 1500 in 2000 min (33.3 h) by the use of a 200 cm x 50 μm i.d. column packed with 3 μm particles with a pore size of 300 Å.[74] All these promising applications show that this approach probably represents a bright future in peptide separation strategies.

An RP separation using 75 μm or 100- μm columns packed with 3-5 μm particles at nanoliter flow rate range (200-300 nL/min) provides, at present, the predominant approach for miniaturized systems. This is a set of parameters that can be relatively easily achieved packing in-house columns. Such columns are now also becoming available in chip form[75, 76] which allows a broader group of researchers to build and operate nanoLC systems.

An appealing alternative that potentially avoids the difficulties of packing traditional C18 columns is the use of silica-based monolithic columns thanks to their robustness, versatile surface chemistry, high column efficiency and fast separation. The permeability of monolithic columns is much higher than for packed C18 columns as a consequence of the presence of large macropores, which permit increased flows through the monolith and, thus, lower pressure requirements.[77] Their use has great advantages when increasing column length and decreasing inner diameter. Luo et al. reported the preparation and applicability of a long (70 cm) silica-based monolithic columns with i.d. of 20 μm[78] and, subsequently, a shorter 25 cm with 10 μm i.d. column,[79] providing high efficiency in separation at low back pressure. In contrast, the group of Tanaka used extremely long columns with larger i.d.s. They connected three C18 silica based monolithic columns of 100 μm i.d. (total length of 11-12 meters), reaching 1 million theoretical plates. They further reported a peak capacity of 380 in 215 min by the use of a 3 meter column.[80] Recently, Iwasaki et al. used a 3.5 meter-long silica based monolithic column (100 μm i.d.) at ambient temperature, with pressures less than 200 bar, resulting in a peak capacity over 600 during a 41h gradient.[81] However, monolithic columns suffer from some important limitations: first, the synthesis of each is unique, resulting in low column-to-column reproducibility; second, monolithic columns can easily be overloaded, which represents a serious obstacle when high sample amounts are being injected.

2.1 RP in multidimensional strategies

Although relatively complex mixtures can be handled by RP because of its excellent resolving power, reproducibility and efficiency, peptides analysis in shotgun experiments is still limited when using one-dimensional RP-based approaches. An efficient way to address the limited peak capacity is to integrate RP as part of a multidimensional separation strategy. Various combinations of orthogonal separation methods have been employed for more comprehensive analyses, as reported in other excellent reviews[21, 82]

The combination of RP as both first and second dimensions in 2D-LC strategies does not look promising at first glance, due to the similar hydrophobic-based retention in both dimensions. However, the selectivity for peptides under RP separation, based on hydrophobicity, can be modulated by varying other factors, such as type of stationary phase, mobile phase pH, choice of organic modifier and ion pairing reagent. Some of these combinations are helpful to achieve a higher degree of orthogonality in 2D-LC configurations. For instance, the effect of changing the type of stationary phase has been investigated using different functionalities, such as C18, phenyl, or pentafluorophenyl (PFP).[23] The selectivity showed a high correlation, thus the combination of two different types of RP columns provides only limited orthogonality and is not suitable for 2D-LC applications.[23] In contrast, the impact of the mobile phase pH has a more pronounced effect on altering peptide selectivity than the stationary phase.[83-86] Since

peptides are charged molecules comprised of ionizable basic and acidic functional groups, the change of mobile phase pH also affects their retention behavior. Neutralization of a charged residue decreases its hydrophilicity or increases its hydrophobicity and, consequently, leads to better retention. Thus, the pH is a potent tool to obtain more orthogonal separations and can be integrated in a 2D RP-RP approach based on the use of different pH values of the mobile phase in the first and second dimension. Gilar et al.[23] evaluated the impact of pH on peptide selectivity under RP separation using different pH conditions, ranging from 2.6 to 10. They showed that acidic peptides are better retained at low pH where the carboxylic acids are protonated/neutralized (increased hydrophobicity), while basic peptides are better retained at high pH, due to the deprotonation/neutralization of basic residues (Figure 3). In this respect the internal pKa values of the side chain functionalities play a major role, as they indicate at which pH value a moiety is charged or not (see Table 1).

The comparison, by Gilar et al., between LC systems operating at pH 2.6 versus 7.8, 8.5, or 10 showed that greater orthogonality was reached when a wider pH gap was used. This might be the result of the change in the protonation state of the primary amine of the N-terminus (pKa between 8 and 10) and histidine residues (pKa app. 6). The total deprotonation of histidine and the partial to total deprotonation of the N-terminus at a pH of 10 change the hydrophobicity, whereas at a pH of 7.8 or 8.5 the N-termini may still be protonated (see Figure 4). Taking into account these results, Gilar et al. designed a RP-RP bidimensional system using pH 10 in the first and pH 2.6 in the second dimension, employing columns with identical (C18) packing material, and confirming that the orthogonality was solely dictated by the pH effect.[23, 83] Nevertheless, the orthogonality of RP-RP in multidimensional systems needs further improvements. A strategy to address this problem was recently proposed by the group of Zou.[91] This mainly theoretical approach is based on 2D high pH-low pH RP-RP combined with a new fractionation plus pooling scheme where the early fractions are pooled with late fractions and then analyzed in the second RP dimension. In this way the separation space in the second dimension is used more

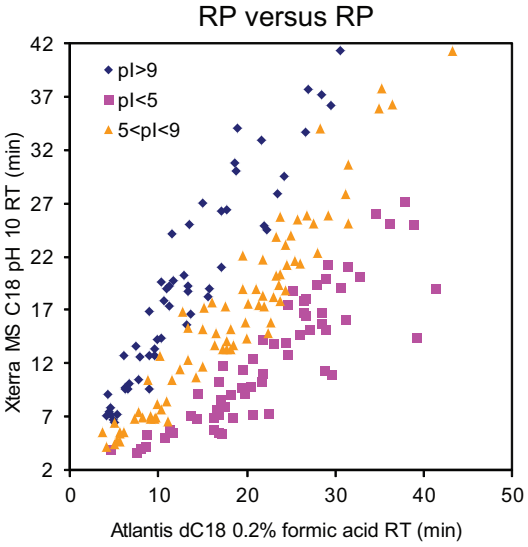

Figure 3: *Normalized retention time plot showing the impact of the pH on achieving orthogonality in RP-RP separation. Adapted from Gilar et al.[83]*

Figure 4: *Charge states of individual residues in the peptide XXHXXpSXXKPXXDXEXXR at pH3, pH7 and pH10. Red indicates a positive charge and blue indicates a negative charge. P in lowercase refers to phosphorylation. Letters in brackets indicate that the charge state is undefined. The charges were drawn based on the internal pKa values of the individual residues.*

	pKint (Nozaki&Tanford)[87]	pKint (Thurlkill)[88]	pKint (Gurd Lab)[89]	pKint (Wuthrich Lab) [90]
α-Carboxyl	3.8	3.7	3.3	x
Asp	4	3.7	3.9	3.9
Glu	4.4	4.25	4.2	4.2
His	6.3	6.5	6.8	6.9
α-Amino	7.5	8	8.1	x
Lys	10.4	10.4	10.5	11
Arg	12	x	x	x

Table 1: *Internal pKa values of the side chains of different amino acids and the N- and C-terminus as measured by different groups.[88]*

efficiently and, as a consequence, the orthogonality is increased.[91] A recent application of low-pH RP and high-PH RP with multiple concatenated fractions has shown some benefits compared to the conventional SCX-RP strategy, not solely in terms of peptide identifications, but also related to the decrease in sample loss and sample processing time since the desalting step is not needed for the 2D LC-MS analysis.[92]

Each RP-RP approach suffers from certain limitations, for instance the incompatibility of mobile phases due to the high organic content employed for the elution in the first dimension, which makes it difficult to design the 2D system in an on-line format. Therefore, off-line strategies are usually favorable, based on a partial evaporation of the collected fractions to decrease the organic content prior to the second dimension in order to minimize sample losses. A major concern related to the use of alkaline pH with silica-based stationary phase is the long-term stability of the column and the loss of performance during extended use, which may be tackled by using more pH-resistant polymer-based stationary phases, such as PS-DVB[85] or specifically end-capped silica particles.[23]

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

3. ION EXCHANGE CHROMATOGRAPHY

Ion exchange chromatography (IEX) has been used for peptide separation on HPLC systems for a number of decades.[93-97] The separation is mainly based on Coulombic interactions between charged groups of the analyte and the charged stationary phase. In order to elute the peptides salts are generally mixed into the mobile phase. The salt's cation or anion population competitively displaces the peptide. A higher concentration of salt is required for more strongly bound (more highly charged) peptides. Elution can also be conducted by a change of the mobile phase pH in order to neutralize or invert the charge of the analyte and/or the stationary phase. Principally, there are two main forms of IEX; cation exchange (CX) and anion exchange (AX).

3.1 Strong Cation Exchange-RP

Cation exchange (CX) materials were proven applicable for separating peptides in the 1980's[94-97] and the first attempts in two dimensional peptide separations containing a CX step were also conducted at that time.[98] Over the last decade, the combination of CX with RP has

been the most popular multidimensional approach for peptide separation. In CX the stationary phase has an anionic functional group that enables binding of molecules with cationic groups. In strong cation exchange (SCX) chromatography the functional groups are strong acids, which enables work over a wide pH range. The most common functional groups in SCX materials are strong acids, commonly sulfonic acid derivatives. Due to the low pKa value of the sulfonic acid group, these materials can be utilized at very low pH without losing their negative charge. Working at a low pH of approximately 3 is common in SCX of peptides. The application of weak cation exchange (WCX) materials on the contrary is restricted to a smaller pH range (above approx. pH 4), which explains the rarity of their use in shotgun proteomics. At a pH of 3, the carboxylic acid of the side chains of aspartic and glutamic acid, as well as the C-terminal carboxylic acid, are neutralized by protonation and the basic sites are positively charged, as visualized in Figure 4. This results in most of the tryptic peptides having a positive net charge, which enables binding to the anionic SCX material.

IEX, and thus SCX, separate predominantly based on charge, whereas RP-LC separation is based on hydrophobicity. The high orthogonality makes these two techniques good partners for two dimensional separation approaches.[23] When SCX is combined with RP, SCX is used as the first dimension. The two main reasons are the excellent compatibility of RP-LC with mass spectrometry and that the method with the highest separation power is applied as the last dimension.[68, 99]

Figure 5 illustrates that the separation power of RP is better than SCX and illustrates that peptides elute in clusters depending on their charge state in SCX whereas they are more evenly distributed in RP. Determination of the separation power of SCX and RP-LC has been conducted by Gilar et al.[23] For SCX they determined peak capacities of 63, 85 and 113 using 20, 40, and 80 min gradients, respectively. RP-LC reached a peak capacity over 350 during a 100 min gradient.[68] Even higher values can be reached when working with ultra-high pressures. For example, Köcher et al. managed a separation with a peak capacity of 700 when utilizing 2 µm material in combination with ultrahigh pressure and gradients of up to 10 hours.[19]

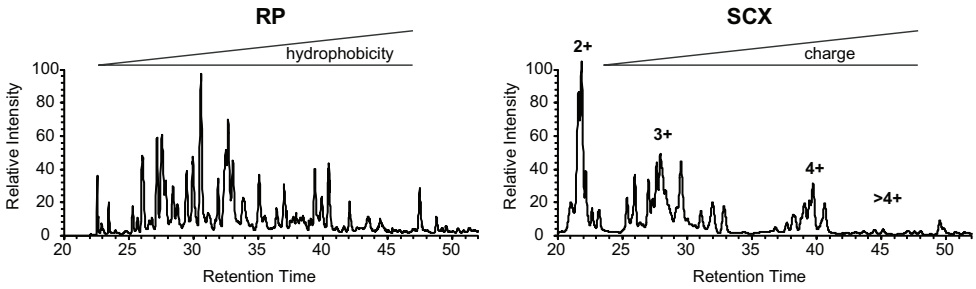

Figure 5: *Chromatograms of a mixture of tryptic peptides from bovine serum albumin and bovine casein separated by RP and SCX. The gradient time was approximately 30 minutes for each separation.*

SCX-RP has been applied to a great number of research topics. Numerous proteomes or subproteomes of bacteria,[100] fungi,[30, 45, 101] animals[75] and plants[102] have been analyzed by SCX-RP and the identification of thousands of proteins is now standard, as exemplified by the analysis of human embryonic stem cells, which resulted in the identification of over 5000 proteins even using a highly conservative Mascot score cutoff of greater than 35.[103]

The easiest way of combining SCX and RP is by constructing an off-line setup like the original from Takahashi et al.[98]. Here, the sample is first separated by SCX and fractions are col-

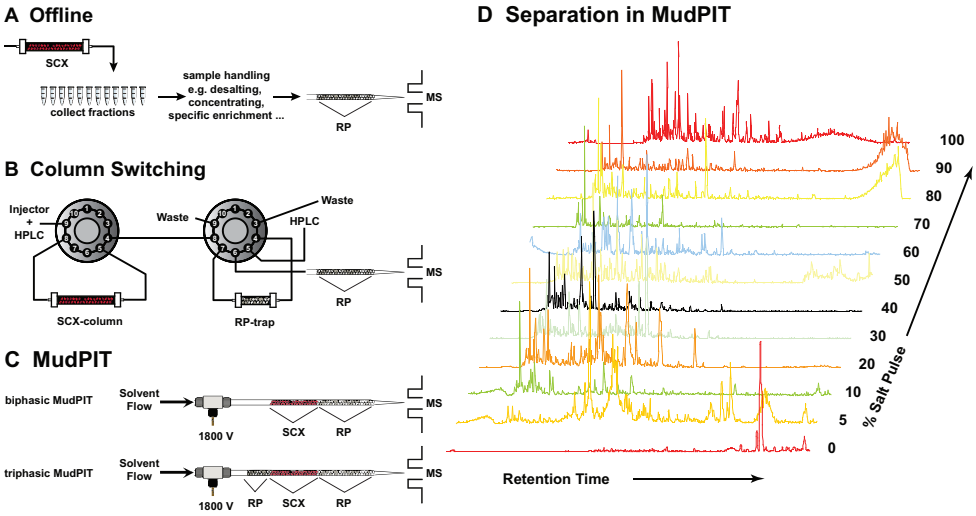

Figure 6: *Different instrumental setups for two dimensional separations applying SCX and RP. a) In an off-line setup the sample is first separated by SCX and fractions are collected. The fractions can be processed if needed and are subsequently separated by RP-LC and analyzed by MS. b) An example of an on-line column switching setup. The sample is first loaded onto the SCX column and eluted stepwise onto the trap column. The sample is then desalted and subsequently eluted onto the analytical RP column followed by MS analysis. c) In the MudPIT approach the SCX and RP material are in one capillary that also functions as spray tip for direct MS analysis. In the triphasic setup an additional RP phase is packed before the SCX and functions as a trap for desalting the sample prior to SCX-RP-MS. d) Chromatograms of a MudPIT analysis. Each color indicates the RP separation after one salt step. c) and d) are adapted from Yates et al.[45]*

lected, as depicted in Figure 6. As a consequence, there can be a big disparity in SCX and RP column size. Thus, there are no limitations in the SCX separation. The SCX column dimensions can be chosen depending on the sample amount to be analyzed and the optimum flow rate can be used. Also, all common buffer systems can be utilized for SCX, due to the possibility of removing unwanted constituents prior to RP-LC-MS analysis. Thus, non-MS-compatible phosphate buffers and salts such as sodium or potassium chloride are often applied for elution and the collected fractions are desalted prior to MS analysis. Furthermore, acetonitrile is commonly added to the mobile phases of the SCX separation. The addition of acetonitrile to the buffers reduces nonspecific binding to the support material. As a consequence the influence of the hydrophobicity of the peptide on the retention can be minimized,[104] and thus the separation is mainly based on the net charge of the analyte, which increases the orthogonality in SCX-RP. Thus, often, more than 20% acetonitrile is part of the SCX buffers in off-line setups. [23, 30, 50, 68] After the SCX separation the acetonitrile can be removed by concentrating the sample by vacuum centrifugation.

Furthermore, additional steps such as phosphopeptide enrichment can be conducted between SCX and RP.[75] An additional great advantage of an off-line approach is that selected fractions of interest can be concentrated, further processed, analyzed, reanalyzed or stored until further needed. The biggest disadvantage of the off-line approach are the sample losses, due to the sample handling between the different dimensions.

In addition to the off-line setup, various on-line setups have been used for SCX-RP of peptides. The on-line-approaches can be further divided into column-switching and multidimen-

sional protein identification technology (MudPIT) systems. In the late nineties the group of Yates provided the blueprint of two dimensional chromatography for proteomics, when they introduced MudPIT.[45, 101] Here, the SCX column and the subsequent RP column are combined in one fused silica capillary needle that is directly connected to the mass spectrometer and also acts as an ESI emitter (Figure 6c). In MudPIT the sample is first trapped on the SCX phase. The peptides are partially eluted from the SCX onto the RP phase by increasing salt steps. Prior to the next salt step, the peptides are eluted from the RP phase into the mass spectrometer by the use of an acetonitrile gradient (Figure 6c and d).

MudPIT has the major advantage that the whole procedure is automated and in general this reduces the time of the total analysis. One disadvantage of MudPIT is the limited sample amount that can be loaded. Furthermore, the loaded sample needs to be desalted prior to loading onto the system. In order to desalt on-line a triphasic column (RP-SCX-RP) can be utilized as depicted in Figure 6c. The additional RP material acts as a trap and is used to desalt the sample prior to the SCX-RP separation.[105] In the initial description of MudPIT, in which it was named integrated direct analysis of large protein complexes (integrated DALPC), nonvolatile salts (K2HPO4 and KCl) were used,[106] which turned out to be problematic in combination with MS.[45] Soon, they were substituted by the volatile salt ammonium acetate.[45, 71, 101, 105] Besides the advantage of being the easiest on-line setup, the major drawback is the limited lifetime of the MudPIT columns and the general risk of on-line approaches that if an error occurs during the long run a great part of the sample might be lost. Another development is the combination of ultrahigh-pressure liquid chromatography (UHPLC), operating at >1300 bar, with MudPIT (UHP-MudPIT).[71] Still, UHPLC is a relatively new technique and not easy to operate. Operating a triphasic UHP-MudPIT setup at elevated temperatures has proved to be beneficial compared to operating at room temperature. More identifications were achieved at a temperature of 45°C. By analyzing the cytosolic and membrane fractions using a 45 cm MudPIT column and a 20 hour 6-step run at 45°C, 46% of the proteome of Escherichia coli could be identified, which corresponds to almost 1900 proteins.[100]

In addition to pure SCX materials, mixed bed columns have been tested in MudPIT for their advantages over standard SCX-RP. Mixed bed columns combine at least two different modes of separation in one column, by mixing particles with different selectivity. Motoyama et al. investigated the use of weak anion exchange (WAX) and SCX material in a mixed bed.[51] They used a triphasic MudPIT setup and compared solely SCX (RP-SCX-RP) and WAX (RP-WAX-RP) with different blends of WAX and SCX (RP-WAX/SCX-RP). Blends of WAX and SCX material in a ratio of 1:2, 1:1, 2:1 and 5:1 were tested. In addition, WAX and SCX columns were connected in tandem and compared against the triphasic MudPIT setups. The 2:1 WAX:SCX blend afforded the most peptide identifications. In addition, they observed an improved orthogonality with the mixed bed compared to SCX alone. They stated that the increased orthogonality most likely contributes significantly to the improved identification rate.[51]

In column-switching based on-line approaches, the SCX and the C18 column are connected via switching valves, as depicted in the example in Figure 6b. Opiteck et al. were one of the first to build an SCX-RP on-line system to analyze whole proteins coupled to MS.[47] Since then, diverse setups have been constructed to connect SCX and RP-LC-MS. Often in these setups trap columns are integrated between the SCX and RP-LC. The trap columns enable desalting and in consequence the use of common non-volatile buffers and salts for elution in the initial SCX. In addition, the integration of trapping columns enables the concentration of the sample and thus affords more freedom in the flow rates applied for the initial SCX separation.[107] Furthermore, results can differ when a pump is used to deliver the buffers for gradient elution from the SCX[108] or step elution is conducted by injection of SCX elution buffer with de-

fined salt concentrations. In any case the two separation techniques need to be adapted to each other, which makes these setups very complicated and not appropriate for routine applications. Still, the advantages of reduced sample loss, relatively fast analysis due to automation and the applicability to minor amounts of sample, make these developments exciting.

3.2 SCX-RP for the Enrichment of Post-Translationally Modified Peptides

In addition to the separation of regular peptides, SCX has been optimized to enrich for certain classes of post-translationally modified peptides. The most prominent example is the enrichment of sub groups of phosphorylated tryptic peptides, which was first conducted with SCX by the group of Gygi.[75] Here, the pH of the SCX buffers is kept at approximately 3. In this pH range carboxylic acid residues, being predominantly protonated as their pKa value is higher than three (Table 1),[88] do not contribute to the net charge of the peptide, whereas the phospho-sites are still negatively charged and reduce the net charge by one in comparison to their unphosphorylated counterparts, as depicted in Figure 4. The difference in net charge state enables the separation of phosphopeptides from their non-phosphorylated counterparts. Thus, enrichment of pools of phosphopeptides that do not co-elute with less-charged regular peptides is possible.[50, 75, 108, 109] These pools are multiply phosphorylated peptides and singly phosphorylated peptides, that contain only a single basic amino acid. Singly phosphorylated peptides with more than one basic amino acid have the same net charge as non-phosphorylated peptides with one basic residue less. Thus, these classes of phosphopeptides co-elute with regular peptides and need further enrichment for which immobilized metal affinity chromatography (IMAC)[110, 111] or TiO_2[112-114] is most commonly used.

The enrichment of N-terminal acetylated peptides is also possible with SCX at a pH of 3.[109, 115-118] N-terminal acetylated peptides have a reduced net charge due to the neutralization of the basic N-terminus by acetylation. Thus, N-terminal acetylated peptides with only one basic amino acid can be separated in the same way, as described for the phosphorylated peptides. It is even possible to separate the N-terminal acetylated peptides from singly phosphorylated peptides, as depicted in Figure 7, even though they have the same net charge.[109]

One explanation is that the orientation of the peptide also has an influence in the SCX separation. Alpert et al. demonstrated that the carboxyl-group of the C-terminus interacts with the C-terminal basic residue of tryptic peptides and in consequence the N-terminal amino group has a higher exposed basicity. Thus, the N-terminus of both regular and phosphorylated peptides is oriented towards the stationary phase.[119] In the case of N-terminal acetylated peptides, it can be assumed that the orientation changes, due to the loss of basicity at the N-terminus. Now the less basic C-terminus is orientated towards the stationary phase, which accounts for N-terminal acetylated peptides being less well-retained compared to phosphorylated peptides (as depicted in Figure 7).

Recently, we published a three dimensional separation approach for the enrichment of basic phosphopeptides which combines SCX at a pH of 3 with a subsequent SCX separation at a pH of 1. In this tandem SCX approach, the phosphate switches from being negatively charged to neutral upon change of pH while all other standard amino acids remain the same. This change enables separation of phosphopeptides from regular peptides in the second SCX dimension. This approach allowed the identification of over 10,000 basic phosphopeptides.[120]

3.3 Anion Exchange-RP

The second form of IEX is anion exchange (AX). AX was proven to be applicable for separating peptides in the 1980s.[93] Anion exchange (AX)-RP has also been used for multidimensional approaches since that time[121, 122] and has been automated and miniaturized.[123]

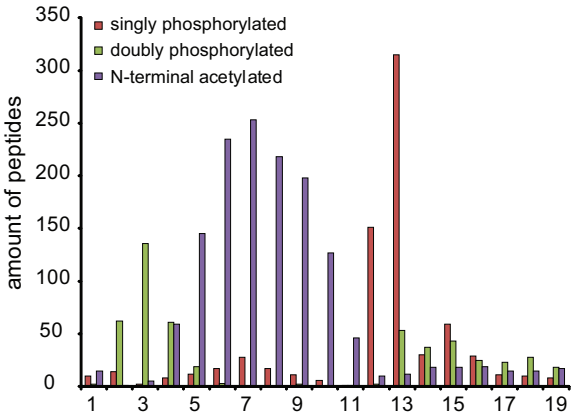

Figure 7. *Distribution of singly (red) and doubly (green) phosphorylated and N-terminal acetylated (violet) peptides in the early fractions of an SCX separation at pH 3.[109, 118]*

Inversely to CX, in AX the stationary phase contains positively charged residues that undergo Coulombic interactions with anionic sites of the analytes. Even though peptides have acidic residues, the separation of peptides by anion exchange is less common. This might be due to the fact that theoretically over 29% of human tryptic peptides will have a neutral or basic net charge at a pH lower than 8.5 and thus, these peptides are not expected to be retained by AX.[124] The reason for this is the high pKa values of the side chains of lysine and arginine, as given in Table 1.[88] Thus, when working at a pH of 8.5, lysine and arginine are still positively charged. If no additional acidic residues (besides the C-terminus) are present in the peptide the C-terminus is the only acidic residue present and the net charge of the peptide is zero or even positive when the peptide has missed cleavages. As a consequence these peptides do not bind in anion exchange.

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

AX is a very useful tool for specific separation and enrichment of acidic peptides, such as those from acidic proteins from brain[122] or phosphorylated peptides.[50, 124-128] The phosphogroup adds up to two negative charges to the peptide, depending on the pH. The additional negative charges causes the net charge of phosphopeptides (without missed cleavages) to be negative and as a consequence enables their retention on AX materials (Figure 8). Several phosphopeptide enrichment and separation methods have been developed that mainly apply SAX materials in combination with pH step[125] or gradient elution.[124, 126, 128] Due to the obvious bias of AX towards acidic solutes, AX has been combined with SCX in various ways in order to complement each other.[49, 50, 126, 129] Dai and coworkers[49] have used SCX as a first dimension in their off-line approach. The elution was conducted by different pH steps ranging from a pH of 3 up to a pH of 10. The SCX flow through (pH 2.5) was further separated with SAX followed by RP of all SCX and SAX fractions. The parameters for each step were carefully chosen, as the solvent conditions for the SCX separation were chosen in a way that the flow through of the SCX is dominated by acidic peptides including most of the identified phosphopeptides and SAX is well suited to separate this pool of peptides. In this way they improved their SCX dimension by further separating a pool of poorly resolved peptides with SAX. Recently, Hennrich et al. [50] showed that weak anion exchange (WAX) at a low pH is especially suited to further separate phosphopeptides from early SCX fractions. In contrast to the approach from Dai et al. Hennrich et al. used a refined SCX which was conducted at a pH of 3 and applied a salt gradient for elution. This setup leads to some fractions enriched for singly phosphorylated peptides (Figure 7). These two to three fractions contain thousands

of phosphopeptides. One of these fractions was further separated by WAX in order to overcome issues of dynamic range and complexity. The WAX was conducted at a constant pH of approximately 3 and a salt gradient was applied for elution. Using this off-line SCX-WAX-RP approach over 11,000 unique phosphopeptides could be identified from just this one SCX fraction.[50]

A complete on-line 3 dimensional system has been developed by Zhou et al. (Figure 8).[130] They combined RP at high pH with SAX and RP at low pH and compared it against RP-RP high-low pH. The use of SAX as intermediary step also allowed the solvent incompatibilities of the two RP steps to be circumvented. They observed an increased identification rate with RP-SAX-RP compared to RP-RP when analyzing complex proteomes of Escherichia coli and Saccharomyces cerevisiae. In addition, they found that sampling 101 instead of 51 fractions in the final RP dimension added less than 5% protein and 6% unique peptide identifications.[130] This is valuable information and should be tested when a multidimensional separation is first constructed, in order to identify the optimum compromise between experiment time and proteome coverage. In a follow up study the same group applied the RP-SAX-RP approach to phosphopeptide analysis. They achieved a peak capacity of over 3500 in 126.5 hours, leading to approximately 12,000 phosphopeptide identifications.[131]

In addition AX-RP mixed mode columns have been applied for peptide separation. Mixed mode columns combine at least two different modes of separation in one column. In mixed mode columns two or more chemical groups with different binding specificities are combined in the uniform stationary phase.[132] Phillips and co-workers[132] used a commercially available mixed mode column in a two dimensional setup. They applied the mixed mode column as first dimension and compared it to a standard SCX column. The mixed mode column combined an anion exchange (AX) and RP functionality and was used in an off-line setup. As the separation with the RP/AX stationary phase is based on hydrophobic and electrostatic interac-

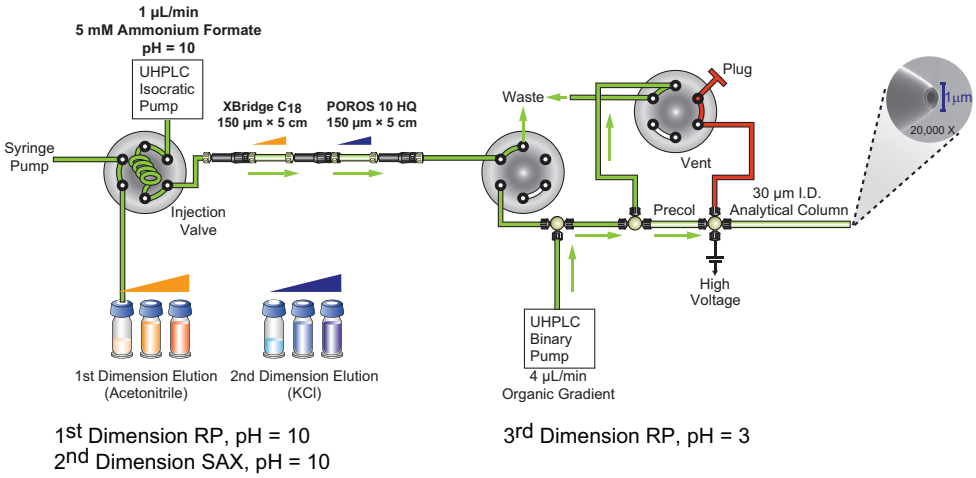

Figure 8: *Schematic diagram of an automated, on-line nanoflow RP-SAX-RP system. The autosampler first loads the sample and is then used to inject the elution buffers for the first- (acetonitrile, orange) or the second- (KCl, blue) dimension, respectively. Second dimension fractions are diluted (4:1) and acidified with reversed phase solvent A (0.1% formic acid, 3% acetonitrile) introduced by an ultrahigh-pressure (UHP) binary pump. An additional six-port valve provides a vented third dimension column configuration and allows concentration of the peptides on the precolumn. An organic gradient is delivered by the UHP binary pump to elute peptides from the third dimension column for MS/MS analysis. Adapted from Zhou et al.[130]*

tions, a gradient with an increase in acetonitrile (3%-80%) in conjunction with a decrease in buffer concentration (20 mM ammonium formate to 2 mM ammonium formate) and pH (6.5-2.5) was applied. They observed an increase in peptide separation efficiency and a more homogeneous distribution of tryptic peptides with the RP/AX column when compared to SCX. In addition, the RP/AX setup resulted in at least 50% more unique peptide identifications.

4. HYDROPHILIC INTERACTIONS CHROMATOGRAPHY

The term HILIC was coined by Alpert in 1990[133] to describe normal-phase chromatography with water miscible mobile phases to separate hydrophilic compounds, such as proteins and peptides, but separations using these conditions were originally described by Linden et al. already in 1975, where similar conditions were employed for a carbohydrate separation.[134]
The exact separation mechanism in HILIC is still not fully understood but the hydrophilic partitioning model can be used to adequately rationalize many trends observed in analyte retention.[133] A sufficiently polar stationary phase and a typically low aqueous mobile phase (5-20 % water in ACN)[133] are employed creating a water-enriched liquid layer around the polar stationary phase. The separation is achieved by partitioning of analytes from the dynamic mobile phase into this stagnant aqueous layer around the stationary phase.[133, 135] Elution is then obtained by increasing the water content of the mobile phase. Factors governing the retention are both hydrogen bonding[136], the extent of which depends on the acidity and basicity of the peptides, and dipole-dipole interactions, which depend on the dipole moments and the polarizability of the analytes.[133] In HILIC, the retention of peptides increases with increasing polarity or hydrophilicity of peptides and stationary phases, opposite to the trend observed in RP. However, the order of elution upon HILIC separation is not simply the reverse of RP, showing elution trends related to a set of physicochemical properties rather than simply hydrophilicity. Several papers and reviews have attempted to examine the variables involved in peptide retention under HILIC conditions,[135-138] but the exact retention mechanisms of HILIC are largely determined by the individual stationary and mobile phases in use, the type and concentration of salts and the pH.[139, 140]
The most common stationary phases applied for peptide separation in HILIC mode include underivatized silica that contain functional groups such as siloxane, silanols and a small quantity of metals, [141, 142] and derivatized silica, which can be neutral stationary phases, such as TSKgel Amide-80;[138] ionic stationary phases, such as the weak cation-exchanger PolyCAT A,[143] the cation exchanger polysulphoethyl A,[133] the weak anion exchanger PolyWAX LP;[127] zwitterionic stationary phases, such as ZIC-HILIC[144, 145] and ZIC-cHILIC.[54] Each of these materials is capable of generating a semi-immobilized aqueous layer on their polar surface.[133] However, different stationary and mobile phases display different retention characteristics, and the type of salt employed can also influence the retention behavior.[146]
When peptide separation is performed using a neutral hydrophilic TSKgel Amide-80 stationary phase, the primary interaction is claimed to be hydrogen bonding. The retention is mostly based on the overall hydrophilicity of the analytes,[137, 147] although some electrostatic effects might play a role under specific buffer's conditions, especially when employing silica-based materials.[127, 140] The main advantage of this kind of HILIC stationary phase is the use of 'salt-free' buffers, when not considering as salts the use of volatile electrolytes in the mobile phase such as TFA and formic acid. The elution is predominantly obtained by increasing the polarity of the mobile phase (higher ratio water/organic solvent). Thus, this separation mode is directly compatibility with MS detection or any further step of separation in multidimensional strategies when HILIC is used as first step.
With ionic stationary phases, besides polarity, electrostatic interactions can also play an impor-

tant role in increasing or decreasing the selectivity for peptides,[133, 135] but the downside of those ionic interactions is the need of higher salt concentration in the buffer to disrupt the attractive interactions and to obtain the elution,[127, 141] which may cause dramatic ionization suppression and negative effects with MS detection. The choice of salts is, however, limited by their solubility in highly organic buffers,[140] and ammonium formate and acetate are typically chosen for their good compatibility with MS. A novel method utilizing 'saltless' pH gradient with a WCX-HILIC chromatography was described for the automated on-line separation of hypermodified histone peptides, directly introduced via nanoelectrospray into the mass spectrometer, improving the ionization and sensitivity during the MS analysis.[148]

When zwitterionic materials with <20 mM salt concentrations are employed, a good compromise between selectivity and compatibility can be achieved since peptide separation involves both hydrophilic and electrostatic interactions,[149] but the latter interactions are weaker compared to normal ionic exchangers and do not require high salt concentration for the elution.[135] An advantageous property of zwitterionic stationary phases relies on the amphoteric character of their functional moieties, containing two oppositely charged groups in close proximity at a stoichiometric ratio, the charges of which are permanent and unaffected by pH changes. Hence, assuming the same charge distribution on the stationary phase, the optimization of the mobile phase pH is mainly dictated by the analytes charge and represents a potent factor to modulate the interactions between charged peptides and the zwitterionic stationary phase. Changing pH within a specific range influences the ionization of peptides due to protonation or deprotonation of certain amino acids, affecting their hydrophilicity and, consequently, their retention in HILIC mode.[52] Recently, two pH conditions, one neutral and one acidic, have been extensively evaluated with respect to their influence on peptides retention using a zwitterionic stationary phase (ZIC-cHILIC).[54] It was shown that ZIC-cHILIC selectivity at pH 6.8 was mostly driven by the peptide hydrophilicity, consistently with the hydrophilic partitioning model, whereas at pH 3.5 the retention was governed by a stronger mixed-mode mechanism based on balanced hydrophilic and electrostatic interactions (Figure 9).

4.1 HILIC in multidimensional strategies

HILIC is a relatively new player in the proteomics field, providing particular advantages in comparison to more established techniques. First, the high organic content of the buffer, together with the use of low salt concentration or volatile salts, makes HILIC highly compatible with ESI-MS, potentially increasing its sensitivity[146, 150, 151] and suggesting that HILIC can be an interesting final dimension.[152] Second, the ideal applicability for polar compounds allows the analysis of highly hydrophilic species that would be otherwise lost under RP analysis.[153] Third, the high orthogonality or complementarity to RP makes HILIC a suitable candidate in multidimensional approaches for the analysis of complex samples.[23, 52] All these reasons have contributed to an exponential increase in the application of HILIC chromatography in proteomics, and especially in multidimensional strategies. Furthermore, HILIC can represent an excellent alternative to SCX as a first dimension, providing higher resolution[54] and potentially allowing the enrichment of polar protein post translation modifications (PTMs),[53] such as glycosylation,[52, 154] phosphorylation[147, 155] and N-acetylation.[52]

The successful combination of HILIC with RP in two-dimensional systems relies on their opposite selection for polarity on peptide retention. HILIC chromatography is strongly affected by other factors, such as SCX, RP at high pH and SEC. Boersema et al.[52] further confirmed the competitive level of orthogonality for HILIC when compared to SCX, though a different stationary phase was employed, showing less clustering of similarly charged peptides, as highlighted in Figure 10. This behavior was explained on the basis of the mixed-mode interaction

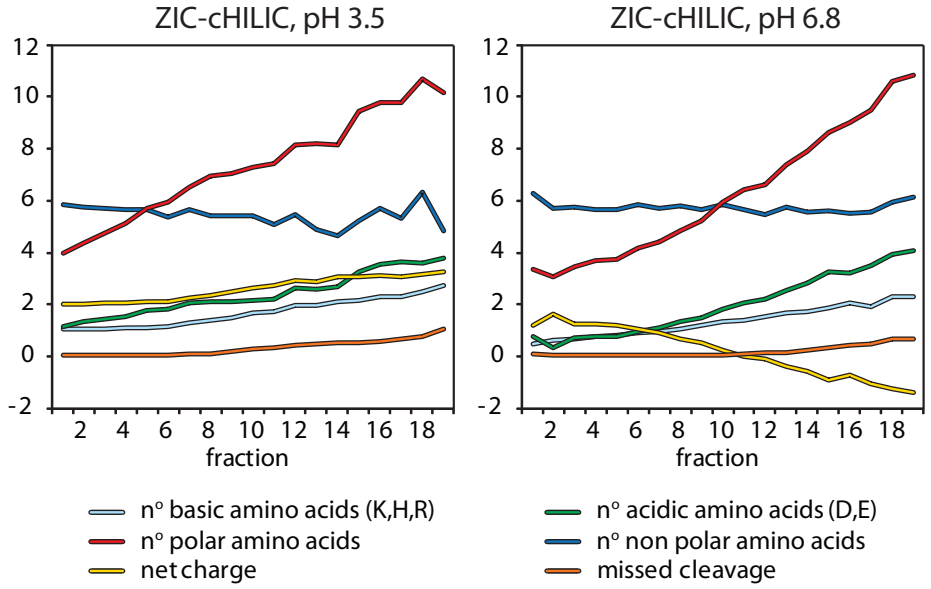

Figure 9: *Analyses of peptide characteristics observed in ZIC-cHILIC fractions at a pH of 3.5 and 6.8 analyzing the effect of basicity, acidity, net charge, polarity, and number of missed cleavages on HILIC retention time. Adapted from Di Palma et al.[54]*

with the zwitterionic stationary phase, where peptides are on one hand retained by charge, due to electrostatic interactions, and on the other hand by hydrophilic partitioning, allowing a larger spread in the distribution of peptides with similar charge.

Although aqueous-organic mobile phases of similar composition (a mixture of ACN and water) are employed for HILIC and RP, compatibility issues in 2D approaches would arise from differences in the elution strengths in the two modes. Mobile phases with high concentration of acetonitrile (ACN) employed with HILIC are strong eluents for RP; thus the hyphenation of

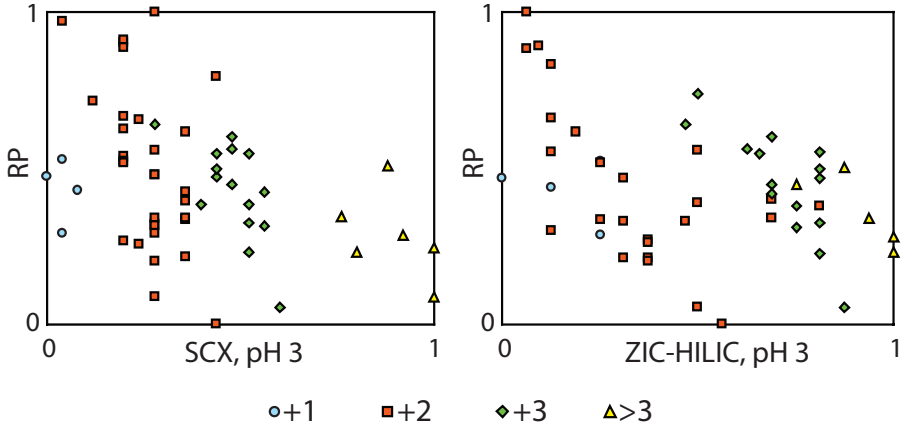

Figure 10: *Normalized retention time plots for SCX-RP and ZIC-HILIC-RP. The net charge of the peptides are indicated with (circle) 1+, (square)2+, (rhomb) 3+, and (triangle) >3+. (Left panel) SCX pH 3 versus RP; (right panel) ZIC-HILIC pH 3 versus RP. Adapted from Boersema et al.[52]*

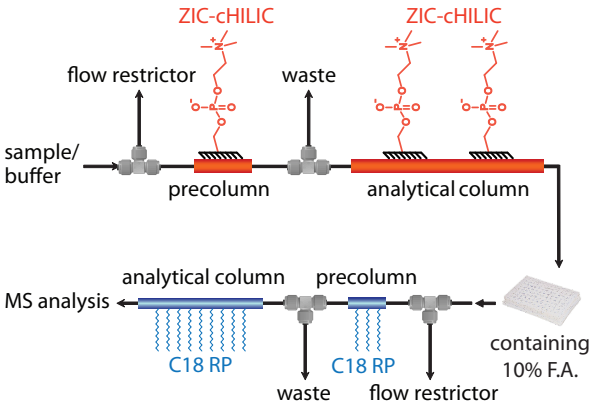

Figure 11: *Schematic design of a 2 dimensional HILIC-RP approach.[17] The sample is concentrated on a ZIC-cHILIC precolumn and subsequently separated at nanoliter flowrates. One-minute fractions are directly collected in a 96 well plate already containing an acidified water solvent. Subsequently, the fractions are directly analyzed by RP-LC-MS with a RP precolumn for desalting and enrichment. Adapted from Di Palma et al.[54]*

HILIC with RP in on-line formats is difficult as it would demand the use of larger second dimension columns in order to dilute the eluent from the first dimension. To overcome this problem, we[54] recently designed an off-line 2D system based on the combination of zwitterionic HILIC and RP in nanoscale format, which represents a refinement from an earlier set-up developed by Boersema et al.[52] In the new design, as schematically shown in Figure 11, the HILIC eluent is directly collected during the separation as 1-min fractions into a 96 well plate already containing an acidified water solvent. Since the HILIC separation is performed at nanoliters per minute, the volume necessary to dilute the HILIC eluent is on the microliter scale and the fractions are sufficiently aqueous and large to avoid evaporation. This strategy could be also defined as in-line, as it is optimized to reduce sample handling, which causes sample loss and contamination, and to allow a direct analysis of the fractions with the subsequent RP-LC-MS/MS. This 2D-ZIC-cHILIC-RP approach combines excellent resolution in both first and second dimension with minimal sample loss, and allows for a more sensitive proteome analysis in comparison to current methods that require more input material. In fact, the analysis can be performed on a mere few micrograms of starting material, achieving proteome coverage comparable to 'large scale' strategies. To probe the sensitivity of this strategy with an actual small-scale sample, we[17] analyzed 10,000 FACS-sorted colon stem cells, directly after the extraction from the mouse intestine, allowing the identification of 15,775 unique peptides originating from 3,775 proteins. A further comparison between the obtained proteome data and previous microarray experiments from the same colon stem cells confirmed the quality of the method, showing that 95% of the proteins detected in this study were also found to be expressed at the mRNA level.

4.2 HILIC for the analysis of PTMs

Given the suitability for the separation of highly polar molecules, HILIC would seem to be ideal in phosphoproteomics and glycoproteomics approaches, or in the targeted analysis of other PTMs such as acetylation, where it can represent an alternative enrichment technique or a chromatographic step in order to selectively reduce the sample complexity. For instance, Young et al. [148] developed an effective on-line HILIC-MS method for the analysis of complex mixture of modified histone forms enabling to chromatographically distinguish isobaric modifications such as trimethylation and acetylation and allowing the thorough characterization of the human histones H3.2 and H4. Since a general overview of HILIC applications for PTMs analysis was provided in an earlier review,[53] here we report solely the latest endeavours in implementing HILIC in the phosphoproteomics and glycoproteomics fields.

Phosphopeptides are, in theory, good candidates for analysis by HILIC because of the increased overall hydrophilicity and their higher negative charge in comparison to regular peptides. McNulty et al.[147] showed that HILIC could be a good first dimension in multidimensional systems for the targeted analysis of phosphopeptides. They compared the combinations of HILIC (using the TSKgel Amide-80 stationary phase) and IMAC in alternate sequences (HILIC-IMAC versus IMAC-HILIC) to establish which approach would be more effective for the selective enrichment of phosphopeptides. Though approximately the same number of peptides were identified with the two approaches, in the IMAC-HILIC experiment only 60% of the total peptides obtained were phosphorylated, while the HILIC-IMAC strategy showed a very selective phospho-enrichment (higher than 99%). However, phosphopeptides were evenly distributed throughout the HILIC fractions, indicating that HILIC chromatography alone, employing TSKgel Amide-80 stationary phase, cannot be used as an efficient enrichment strategy, but represents a good fractionation prior to any more specific phospho-enrichment approach to improve the selectivity. We[54] have investigated the use of two zwitterionic materials, ZIC-HILIC and ZIC-cHILIC, for phosphoproteomics. These two stationary phases exhibit an opposite spatial arrangement of their charged groups, with ZIC-HILIC possessing a negative charge as a distal moiety on the surface, and ZIC-cHILIC a positive charge. In theory, the more exposed positive charge on the surface of ZIC-cHILIC should electrostatically interact with the negatively charged phosphogroup of peptides, enhancing their binding and their retention. Since the phosphopeptides were evenly spread throughout the fractions in both cases, no obvious differences were found between the two zwitterionic materials, demonstrating that in this case the inverted charge arrangement on these stationary phases did not affect the separation. In fact, the surface charge properties of these two materials have been studied through zeta-potential measurements showing that they both exhibit a negative surface charge at a wide range of pH, although their spatial arrangement should, in principle, favor generating a positive charge.[145] This result confirmed that HILIC itself is not an efficient enrichment approach for phosphoproteomics, requiring the combination with more specific strategy.[147] We also believe that it may be possible to increase the phospho-selectivity by the addition in the mobile phase of hydrophobic ion-pairing reagents, such as TFA, in order to preferentially reduce the interactions of regular peptides with the stationary phase, as has been done with glycopeptide enrichment.[156]

Due to the high complexity and heterogeneity of protein glycosylation, the analysis of this PTM is at the present a challenging task. The investigation of complex glycopeptide pools and insightful glycan structural elucidation require good chromatographic separations in combination with glycoenrichment strategies.[157] It is also common to enzymatically remove the glycans attached to the peptides or proteins prior to the MS analysis, thereby reducing the complexity and facilitating glycosylation site determination. Hägglund et al.[158] described an analytical method based on glycopeptide enrichment by HILIC (using microcolumns packed with ZIC-HILIC materials), followed by two different enzymatic deglycosylation strategies for a global proteome analysis of N- and O-glycosylation sites from human plasma proteins. A modified protocol was reported by Thaysen-Andersen et al.[159] to extend this technique for site-specific characterization of low amounts of N-glycosylated proteins after immunoaffinity chromatography, followed by gel electrophoresis separation and in-gel digestion. The peptides obtained were further enriched for glycosylation using HILIC microcolumns, allowing a comprehensive glycoprofiling of the tissue inhibitor of metalloproteinases-1.

For glycopeptides separation, Takegawa et al. [160, 161] already highlighted the potential of HILIC. They showed that ZIC-HILIC columns allowed a high resolution separation of isomeric glycoforms based on structural recognition. Different glycopeptides containing neutral

and sialylated N-glycans as well as released N-glycans were well separated.[160, 161] Wohlgemuth et al.[152] recently described a scheme for glycopeptides profiling based on a ZIC-HILIC pre-enrichment followed by a 2D RP-ZIC-cHILIC separation employing monolithic capillary columns in both dimensions, showing that these two chromatographic techniques are complementary for glycopeptide separation and their combination considerably improved the site-specific elucidation of glycans.

Mysling et al.[162] described an efficient glycoproteomic approach based on ZIC-HILIC separation on a microcolumn format. They explored the use of different mobile phases in order to improve the selectivity of the method and highlighted that the use of an ion-pairing reagent such as TFA in the mobile phase notably increased the efficiency of glycopeptide enrichment (see Figure 12). The use of TFA in the mobile phase acts as an ion pairing agent for the analytes as well as for the stationary phase. The effect on the zwitterionic stationary phase is a higher contribution of hydrophilic partitioning in the HILIC retention since the electrostatic interactions are suppressed by the ion pairing reagent. The effect on the analytes will be a larger hydrophilicity difference between glycosylated and non-glycosylated peptides,[156] with a preferential decrease in hydrophilicity of 'regular' peptides. In spite of that, this ion-pairing-HILIC approach is solely applicable in off-line format due to the presence of TFA that reduces the MS sensitivity and significantly lowers the pH of the mobile phase, affecting the long term stability of ZIC-HILIC silica-based resins.

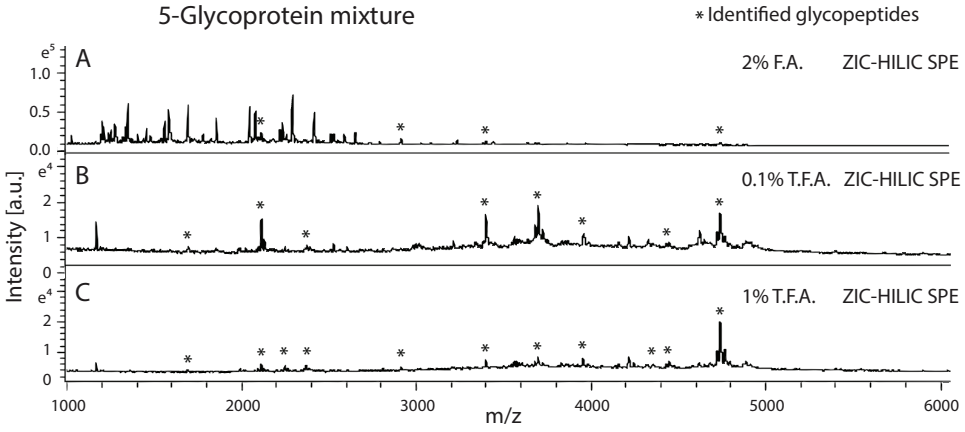

Figure 12: *Glycopeptide enrichment from a peptide mixture derived from a tryptic digest of five glycoproteins (RNase B, ovalbumin, serotransferrin, fetuin, and R-1-acid glycoprotein). The digest was divided in three equal fractions and applied to 2% FA, 0.1% TFA, and 1% TFA ZIC-HILIC SPE and the retained fractions analyzed using MALDI-TOF MS (A-C). Numerous signals in the low m/z region (m/z 1000-2500) were observed for the 2% FA ZIC-HILIC SPE retained fraction, which most likely corresponded to nonglycosylated peptides. These analytes were depleted in the 0.1% TFA and 1% TFA ZIC-HILIC SPE retained fraction, and a number of signals in the higher m/z region were observed (m/z 3000-5000) corresponding to glycopeptides (assigned with asterisks (*)). Adapted from Mysling et al.[162]*

4.3 Alternative separation based on HILIC: ERLIC

A special variant of HILIC that departs from merely partitioning and exploits additionally superimposed electrostatic interactions is called ERLIC (electrostatic repulsion-hydrophilic interaction chromatography).[127] An ion-exchange stationary phase is used with a highly or-

ganic mobile phase, similar to HILIC, in order to generate hydrophilic interactions with polar analytes. In this way, all peptides in a mixture would be retained through hydrophilic partitioning, despite charged peptides being repelled to some extent by similar charges present on the stationary phase. Since hydrophilic interactions, on one hand, and electrostatic repulsions, on the other hand, have opposite effects on peptide retention, an isocratic resolution of heterogeneous peptides can be achieved.

The ERLIC separation, based on anion exchange chromatography in combination with hydrophilic interactions, has shown good resolving power and high orthogonality to RP; thus, it is especially suitable for first dimensional fractionation of peptides in 2D-LC settings. Hao et al[163]. introduced an ERLIC-RP method for fractionating peptides based on both pI and polarity using a salt-free pH gradient of increasing water content. They compared this strategy to the traditional SCX-RP for the analysis of rat kidney tissue and showed that, in their hands, ERLIC outperformed SCX, identifying 4821 proteins and 30,659 unique peptides. The dataset seemed to have a better representation of highly hydrophobic and basic peptides. They subsequently confirmed their results extending the strategy in a quantitative setting with iTRAQ labeled peptides.[164]

The ERLIC technique has been recently introduced as a potential phosphopeptide enrichment method using a WAX column[127] as an alternative to immobilized metal affinity chromatography (IMAC) and Lewis acids, such as titania,[165] and zirconia.[166] Peptides containing phosphate groups maintain their negative charge even at pH values low enough to deprotonate acidic amino acids. This characteristic can be used to enhance the electrostatic interaction of negatively charged phosphopeptides with the positively charged ERLIC stationary phase, allowing their selective isolation from a peptide mixture. In addition, the retention is further improved by the use of a high concentration of organic solvents (ACN), promoting hydrophilic interactions of the polar phosphate group with the column. A salt and aqueous gradient is then used to gradually elute phosphopeptides.

A comparative phosphoproteomic study of ERLIC versus SCX-IMAC has been reported for the evaluation of isolation/enrichment of phosphopeptides showing that, unlike SCX-IMAC, the ERLIC approach achieved both sufficient enrichment of phosphopeptides and fractionation in one step, though SCX-IMAC outperformed ERLIC in the enrichment efficiency. The results revealed that only 12% of phosphopeptides were commonly identified in both strategies. The ERLIC approach enriched more selectively for multiply phosphorylated peptides with acidophilic motifs, while SCX-IMAC covered a wider variety of motifs, including acidophilic and basophilic. This study highlighted that the two methods are complementary to each other and should be used in combination to obtain a more comprehensive phosphoproteome coverage.[167] A recent publication has demonstrated that when ERLIC is combined with an efficient phospho-enrichment strategy, such as IMAC, prior to the separation, its performance becomes comparable to other common strategies (in this case SCX and HILIC) in terms of phosphopeptide identifications. In this work ERLIC also produced (relatively) the highest proportion of multiply phosphorylated peptides.[168] This work, once again, confirms that, in order to achieve optimal phosphopeptide coverage, it is necessary to decrease the background of nonphosphorylated species which suppress the identification by MS of lower abundant components.[169]

Zarei et al.[170] also systematically compared 3 different chromatographic techniques, ERLIC, HILIC and SCX, coupled to TiO_2 as a specific phosphopeptide enrichment, evaluating the fractionation performance of each workflow based on the number of detected phosphopeptides, percentage of singly and multiply phosphorylated peptides, and their distribution over the applied gradients. The majority of phosphopeptides in SCX and HILIC were singly

phosphorylated, 63% and 68%, respectively, whereas in ERLIC 69% were multiphosphorylated, confirming that ERLIC has a higher efficiency to bind and fractionate multiphosphorylated peptides versus SCX and HILIC. However, the use of TFA as electrolyte in the mobile phase is not ideal in ERLIC separation since it forms ion pairs with basic residues present on peptides, rendering them more hydrophobic. More importantly, the TFA can also bind to the positively charged surface of the ERLIC stationary phase reducing the available surface charge for phosphopeptide binding, which can explain the under-representation of singly phosphorylated peptides when compared to multiply phosphorylated peptides.

Theoretically ERLIC can be applied for the enrichment of other types of protein post-translational modifications, as long as they display sufficient charges, either negative or positive. Zhang et al.[171] reported an optimized ERLIC-based protocol for the simultaneous enrichment of glyco- and phosphopeptides from mouse brain membrane, providing the opportunity to study the interaction of two different PTMs. Phosphopeptides mostly contained one phosphogroup, thus one negative charge, while N-linked glycopeptides possessed, presumably, extra negative charges originating from negatively charged sialic acid groups present on the sugars. As phospho- and glycopeptides differ in their charge and hydrophilicity, their retention and elution profile would be different. Indeed, they found that phosphopeptides mainly eluted with an organic amount of approximately 70-60%, while the majority of glycopeptides did not elute until reaching a lower amount of organic solvent (30%). Thus, increasing the gradient slope with a wider range of organic solvent (from 70 to 25 % of acetonitrile) and optimizing the elution gradient allowed not only the enrichment of both phospho- and glycopeptides, but also a differential distribution of these two modified peptides in different fractions, with the phosphopeptides eluting earlier and the glycopeptides eluting mainly in later fractions (see Figure 13). This approach resulted in the identification of 519 glycoproteins and 337 phosphorylation sites from a relatively small amount of sample (1 mg in replicate analyses). In addition, the

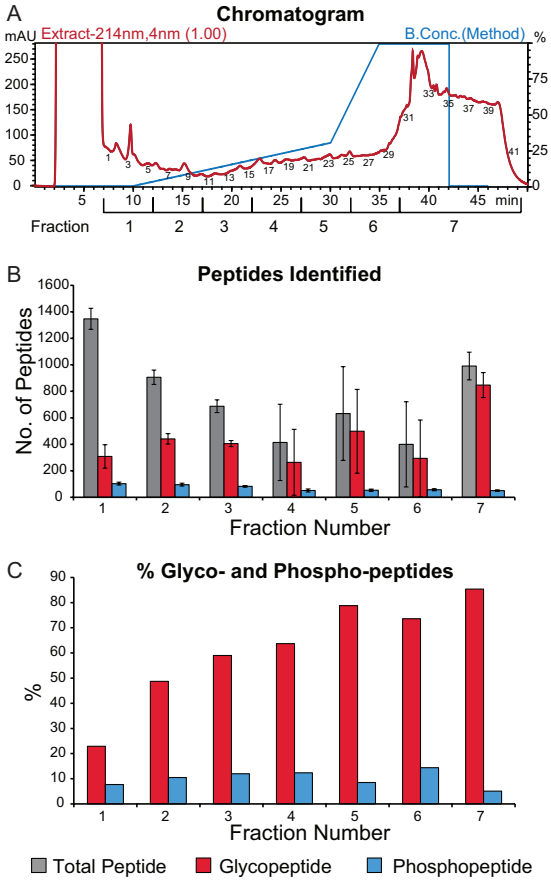

Figure 13: *A, ERLIC chromatogram of 1mg of mouse brain tryptic digest. Forty fractions from 7 to 50 min were collected and then combined to seven final fractions as shown for b and c. Distribution of peptides in the seven collected fractions collected. B, average number of peptides, glycopeptides, and phosphopeptides identified in each fraction from six replicates. Error bars are the S.D. C, percentage of glyco- and phosphopeptides identified in each fraction. Adapted from Zhang et al.[171]*

optimization of the sample preparation in combination with the efficient fractionation allow the identification of phosphorylated membrane proteins, which are particularly difficult to identify due to their intrinsic higher hydrophobicity and low abundance of phosphopeptides. An interesting finding was that more than 40% of these membrane proteins were simultaneously phosphorylated and glycosylated, suggesting that these two PTMs could modulate protein function in a cooperative way.

5. CONCLUSIONS AND OUTLOOK

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

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

We believe that the multidimensional setups of the future will focus on optimization of these points in order to further increase the total peak capacity of coming methods. The simplest multidimensional system to construct is an RP-RP, with SCX-RP being not far behind. The major reason for this situation is that these two separations are well established with vendors providing excellent choices. However, in recent years the promise shown by ERLIC, HILIC and mixed bed material have raised new possibilities and so the choice of multidimensional strategies will expand and become less straightforward.

It should be mentioned that the majority of configurations discussed in this review will provide excellent results regarding proteome coverage and it is essentially the question of what sub proteome populations one wants to observe that will guide one to a preferred strategy. Finally, massive improvements in the next few years are expected especially in the field of peptide separation because of availability of LC systems capable of operating at UHPLC pressures and columns exploiting such equipment. We expect that the power of separations and the proteomics results from their use will continue to improve at an exponential rate.

6. ACKNOWLEDGEMENTS

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

7. REFERENCES

[1] Chait BT. Mass Spectrometry in the Postgenomic Era. Annual Review of Biochemistry. 2011;80:239-46.

[2] Yarmush ML, Jayaraman A. Advances in proteomic technologies. Annual Review of Biomedical Engineering. 2002;4:349-73.

[3] Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003;422:198-207.

[4] Yates JR, Ruse CI, Nakorchevsky A. Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. Annual Review of Biomedical Engineering. 2009. p. 49-79.

[5] Yates JR. Mass spectral analysis in proteomics. Annu Rev Biophys Biomolec Struct. 2004;33:297-316.

[6] Shteynberg D, Deutsch EW, Lam H, Eng JK, Sun Z, Tasman N, et al. iProphet: Multi-level Integrative Analysis of Shotgun Proteomic Data Improves Peptide and Protein Identification Rates and Error Estimates. Molecular & Cellular Proteomics. 2011;10:doi: 10.1074/mcp.M111.007690.

[7] McDonald WH, Yates JR. Shotgun proteomics: Integrating technologies to answer biological questions. Curr Opin Mol Ther. 2003;5:302-9.

[8] McDonald WH, Yates JR. Shotgun proteomics and biomarker discovery. Dis Markers. 2002;18:99-105.

[9] Reid GE, McLuckey SA. 'Top down' protein characterization via tandem mass spectrometry. J Mass Spectrom. 2002;37:663-75.

[10] Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, et al. Mapping intact protein isoforms in discovery mode using top-down proteomics. Nature. 2011;480:254-8.

[11] Kellie JF, Tran JC, Lee JE, Ahlf DR, Thomas HM, Ntai I, et al. The emerging process of Top Down mass spectrometry for protein analysis: biomarkers, protein-therapeutics, and achieving high throughput. Molecular BioSystems. 2010;6:1532-9.

[12] Cox J, Mann M. Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. In: Kornberg, RD, Raetz CRH, Rothman JE, Thorner JW,, editor. Annual Review of Biochemistry. 2011. 80:p. 273-99.

[13] Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Molecular & cellular proteomics : MCP. 2002;1:845-67.

[14] Tyers M, Mann M. From genomics to proteomics. Nature. 2003;422:193-7.

[15] Tolley L, Jorgensen JW, Moseley MA. Very high pressure gradient LC/MS/MS. Analytical Chemistry. 2001;73:2985-91.

[16] Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, et al. Deep proteome and transcriptome mapping of a human cancer cell line. Mol Syst Biol. 2011;7:548.

[17] Di Palma S, Stange D, van de Wetering M, Clevers H, Heck AJR, Mohammed S. Highly Sensitive Proteome Analysis of FACS-Sorted Adult Colon Stem Cells. Journal of Proteome Research. 2011;10:3814-9.

[18] Beck M, Schmidt A, Malmstroem J, Claassen M, Ori A, Szymborska A, et al. The quantitative proteome of a human cell line. Mol Syst Biol. 2011;7:549.

[19] Köcher T, Swart R, Mechtler K. Ultra-High-Pressure RPLC Hyphenated to an LTQ-Orbitrap Velos Reveals a Linear Relation between Peak Capacity and Number of Identified Peptides. Analytical Chemistry. 2011;83:2699-704.

[20] Nagaraj N, Kulak NA, Cox J, Neuhaus N, Mayr K, Hoerning O, et al. Systems-wide perturbation analysis with near complete coverage of the yeast proteome by single-shot UHPLC runs on a bench-top Orbitrap. Molecular & Cellular Proteomics. 2011. doi:10.1074/mcp.M111.013722

[21] Fournier ML, Gilmore JM, Martin-Brown SA, Washburn MP. Multidimensional separations-based shotgun proteomics. Chemical Reviews. 2007;107:3654-86.

[22] Ramautar R, Heemskerk AAM, Hensbergen PJ, Deelder AM, Busnel J-M, Mayboroda OA. CE-MS for proteomics: recent developments and applications. Journal of Proteomics. 2012.

[23] Gilar M, Olivova P, Daly AE, Gebler JC. Orthogonality of separation in two-dimensional liquid chromatography. Anal Chem. 2005;77:6426-34.

[24] Giddings JC. Two-dimensional separations - concept and promise. Analytical Chemistry. 1984;56:1258A-1270A.
[25] Giddings JC. Maximum number of components resolvable by gel filtration and other elution chromatographic methods. Analytical Chemistry. 1967;39:1027-1028.
[26] Giddings JC. Concepts and comparisons in multidimensional separation. Journal of High Resolution Chromatography. 1987;10:319-23.
[27] Dixon SP, Pitfield ID, Perrett D. Comprehensive multi-dimensional liquid chromatographic separation in biomedical and pharmaceutical analysis: a review. Biomedical Chromatography. 2006;20:508-29.
[28] Majors RE. High Performance Liquid Chromatography Columns and Column Technology: A State-of-the-Art Review (Parts I and II). Journal of Chromatographic Science. 1980;18:393&487.
[29] Vollmer M, Horth P, Nagele E. Optimization of two-dimensional off-line LC/MS separations to improve resolution of complex proteomic samples. Analytical Chemistry. 2004;76:5180-5.
[30] Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. Journal of Proteome Research. 2003;2:43-50.
[31] Wagner K, Miliotis T, Marko-Varga G, Bischoff R, Unger KK. An automated on-line multidimensional HPLC system for protein and peptide mapping with integrated sample preparation. Analytical Chemistry. 2002;74:809-20.
[32] Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Analytical Chemistry. 1988;60:2299-301.
[33] Tanaka K. The origin of macromolecule ionization by laser irradiation (Nobel lecture). Angew Chem-Int Edit. 2003;42:3860-70.
[34] Whitehouse CM, Dreyer RN, Yamashita M, Fenn JB. Electrospray interface for liquid chromatographs and mass spectrometers. Analytical Chemistry. 1985;57:675-9.
[35] Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. Science. 1989;246:64-71.
[36] Wilm M, Mann M. Analytical properties of the nanoelectrospray ion source. Analytical Chemistry. 1996;68:1-8.
[37] Makarov A, Scigelova M. Coupling liquid chromatography to Orbitrap mass spectrometry. Journal of Chromatography A. 2010;1217:3938-45.
[38] Domon B, Aebersold R. Review - Mass spectrometry and protein analysis. Science. 2006;312:212-7.
[39] Romijn EP, Krijgsveld J, Heck AJR. Recent liquid chromatographic-(tandem) mass spectrometric applications in proteomics. Journal of Chromatography A. 2003;1000:589-608.
[40] Olsen JV, Schwartz JC, Griep-Raming J, Nielsen ML, Damoc E, Denisov E, et al. A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. Molecular & Cellular Proteomics. 2009;8:2759-69.
[41] Andrews GL, Simons BL, Young JB, Hawkridge AM, Muddiman DC. Performance Characteristics of a New Hybrid Quadrupole Time-of-Flight Tandem Mass Spectrometer (TripleTOF 5600). Analytical Chemistry. 2011;83:5442-6.
[42] Shen Y, Smith RD. Proteomics based on high-efficiency capillary separations. Electrophoresis. 2002;23:3106-24.
[43] Emmett MR, Caprioli RM. Micro-electrospray mass-spectrometry - ultra-high-sensitivity analysis of peptides and proteins. J Am Soc Mass Spectrom. 1994;5:605-13.
[44] Shen Y, Smith RD, Unger KK, Kumar D, Lubda D. Ultrahigh-Throughput Proteomics Using Fast RPLC Separations with ESI-MS/MS. Analytical Chemistry. 2005;77:6692-701.
[45] Wolters DA, Washburn MP, Yates JR, 3rd. An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem. 2001;73:5683-90.
[46] Wang N, Xie C, Young JB, Li L. Off-Line Two-Dimensional Liquid Chromatography with Maxi-

mized Sample Loading to Reversed-Phase Liquid Chromatography-Electrospray Ionization Tandem Mass Spectrometry for Shotgun Proteome Analysis. Analytical Chemistry. 2009;81:1049-60.
[47] Opiteck GJ, Jorgenson JW, Anderegg RJ. Two-Dimensional SEC/RPLC Coupled to Mass Spectrometry for the Analysis of Peptides. Analytical Chemistry. 1997;69:2283-91.
[48] Geng M, Ji J, Regnier FE. Signature-peptide approach to detecting proteins in complex mixtures. Journal of chromatography A. 2000;870:295-313.
[49] Dai J, Jin W-H, Sheng Q-H, Shieh C-H, Wu J-R, Zeng R. Protein Phosphorylation and Expression Profiling by Yin-Yang Multidimensional Liquid Chromatography (Yin-Yang MDLC) Mass Spectrometry. Journal of Proteome Research. 2006;6:250-62.
[50] Hennrich ML, Groenewold V, Kops GJPL, Heck AJR, Mohammed S. Improving Depth in Phosphoproteomics by Using a Strong Cation Exchange-Weak Anion Exchange-Reversed Phase Multidimensional Separation Approach. Analytical Chemistry. 2011;83:7137-43.
[51] Motoyama A, Xu T, Ruse CI, Wohlschlegel JA, Yates JR, 3rd. Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. Analytical Chemistry. 2007;79:3623-34.
[52] Boersema PJ, Divecha N, Heck AJ, Mohammed S. Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. Journal of Proteome Research. 2007;6:937-46.
[53] Boersema PJ, Mohammed S, Heck AJ. Hydrophilic interaction liquid chromatography (HILIC) in proteomics. Anal Bioanal Chem. 2008;391:151-9.
[54] Di Palma S, Boersema PJ, Heck AJR, Mohammed S. Zwitterionic Hydrophilic Interaction Liquid Chromatography (ZIC-HILIC and ZIC-cHILIC) Provide High Resolution Separation and Increase Sensitivity in Proteome Analysis. Analytical Chemistry. 2011;83:3440-7.
[55] Gruber KA, Stein S, Brink L, Radhakrishnan A, Udenfriend S. Fluorometric assay of vasopressin and oxytocin: a general approach to the assay of peptides in tissues. Proceedings of the National Academy of Sciences. 1976;73:1314-8.
[56] Guo DC, Mant CT, Taneja AK, Parker JMR, Hodges RS. Prediction of peptide retention times in reversed-phase high-performance liquid-chromatography .1. Determination of retention coefficients of amino-acid-residues of model synthetic peptides. Journal of Chromatography. 1986;359:499-517.
[57] Chervet JP, Ursem M, Salzmann JP. Instrumental Requirements for Nanoscale Liquid Chromatography. Analytical Chemistry. 1996;68:1507-12.
[58] Meiring HD, van der Heeft E, ten Hove GJ, de Jong APJM. Nanoscale LC–MS(n): technical design and applications to peptide and protein analysis. Journal of Separation Science. 2002;25:557-68.
[59] Schmidt A, Karas M, Dulcks T. Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: When does ESI turn into nano-ESI? J Am Soc Mass Spectrom. 2003;14:492-500.
[60] Shen Y, Zhang R, Moore RJ, Kim J, Metz TO, Hixson KK, et al. Automated 20 kpsi RPLC-MS and MS/MS with Chromatographic Peak Capacities of 1000-1500 and Capabilities in Proteomics and Metabolomics. Analytical Chemistry. 2005;77:3090-100.
[61] Wilkins JA, Xiang R, Horvath C. Selective enrichment of low-abundance peptides in complex mixtures by elution-modified displacement chromatography and their identification by electrospray ionization mass spectrometry. Analytical Chemistry. 2002;74:3933-41.
[62] Xiang R, Horvath C, Wilkins JA. Elution-modified displacement chromatography coupled with electrospray ionization-MS: On-line detection of trace peptides at femtomole level in peptide digests. Analytical Chemistry. 2003;75:1819-27.
[63] Ficarro SB, Zhang Y, Lu Y, Moghimi AR, Askenazi M, Hyatt E, et al. Improved Electrospray Ionization Efficiency Compensates for Diminished Chromatographic Resolution and Enables Proteomics Analysis of Tyrosine Signaling in Embryonic Stem Cells. Analytical Chemistry. 2009;81:3440-7.
[64] Masuda T, Sugiyama N, Tomita M, Ishihama Y. Microscale Phosphoproteome Analysis of 10.000 Cells from Human Cancer Cell Lines. Analytical Chemistry. 2011;83:7698-703.

[65] Wang XL, Stoll DR, Carr PW, Schoenmakers PJ. A graphical method for understanding the kinetics of peak capacity production in gradient elution liquid chromatography. Journal of Chromatography A. 2006;1125:177-81.
[66] Dolan JW. Temperature selectivity in reversed-phase high performance liquid chromatography. Journal of chromatography A. 2002;965:195-205.
[67] Thakur SS, Geiger T, Chatterjee B, Bandilla P, Frohlich F, Cox J, et al. Deep and Highly Sensitive Proteome Coverage by LC-MS/MS Without Prefractionation. Molecular & Cellular Proteomics. 2011;10.
[68] Gilar M, Daly AE, Kele M, Neue UD, Gebler JC. Implications of column peak capacity on the separation of complex peptide mixtures in single- and two-dimensional high-performance liquid chromatography. Journal of Chromatography A. 2004;1061:183-92.
[69] Gehrke CW, Wixom RL. Chromatography : a science of discovery: Hoboken, N.J. : Wiley; 2010.
[70] Jorgenson JW. Capillary Liquid Chromatography at Ultrahigh Pressures. Annual Review of Analytical Chemistry. 2010;3:129-50.
[71] Motoyama A, Venable JD, Ruse CI, Yates JR. Automated ultra-high-pressure multidimensional protein identification technology (UHP-MudPIT) for improved peptide identification of proteomic samples. Analytical Chemistry. 2006;78:5109-18.
[72] MacNair JE, Lewis KC, Jorgenson JW. Ultrahigh pressure reversed-phase liquid chromatography in packed capillary columns. Analytical Chemistry. 1997;69:983-9.
[73] MacNair JE, Patel KD, Jorgenson JW. Ultrahigh-Pressure Reversed-Phase Capillary Liquid Chromatography: Isocratic and Gradient Elution Using Columns Packed with 1.0-μm Particles. Analytical Chemistry. 1999;71:700-8.
[74] Hsu JL, Huang SY, Chen SH. Dimethyl multiplexed labeling combined with microcolumn separation and MS analysis for time course study in proteomics. Electrophoresis. 2006;27:3652-60.
[75] Ballif BA, Villen J, Beausoleil SA, Schwartz D, Gygi SP. Phosphoproteomic analysis of the developing mouse brain. Molecular & Cellular Proteomics. 2004;3:1093-101.
[76] Raijmakers R, Berkers CR, de Jong A, Ovaa H, Heck AJ, Mohammed S. Automated online sequential isotope labeling for protein quantitation applied to proteasome tissue-specific diversity. Molecular & Cellular Proteomics. 2008;7:1755-62.
[77] Tanaka N, Kobayashi H, Nakanishi K, Minakuchi H, Ishizuka N. Peer Reviewed: Monolithic LC Columns. Analytical Chemistry. 2001;73:420 A-429 A.
[78] Luo Q, Shen Y, Hixson KK, Zhao R, Yang F, Moore RJ, et al. Preparation of 20-μm-i.d. Silica-Based Monolithic Columns and Their Performance for Proteomics Analyses. Analytical Chemistry. 2005;77:5028-35.
[79] Luo Q, Tang K, Yang F, Elias A, Shen Y, Moore RJ, et al. More Sensitive and Quantitative Proteomic Measurements Using Very Low Flow Rate Porous Silica Monolithic LC Columns with Electrospray Ionization-Mass Spectrometry. Journal of Proteome Research. 2006;5:1091-7.
[80] Miyamoto K, Hara T, Kobayashi H, Morisaka H, Tokuda D, Horie K, et al. High-Efficiency Liquid Chromatographic Separation Utilizing Long Monolithic Silica Capillary Columns. Analytical Chemistry. 2008;80:8741-50.
[81] Iwasaki M, Miwa S, Ikegami T, Tomita M, Tanaka N, Ishihama Y. One-Dimensional Capillary Liquid Chromatographic Separation Coupled with Tandem Mass Spectrometry Unveils the Escherichia coil Proteome on a Microarray Scale. Analytical Chemistry. 2010;82:2616-20.
[82] Sandra K, Moshir M, D'hondt F, Tuytten R, Verleysen K, Kas K, et al. Highly efficient peptide separations in proteomics: Part 2: Bi- and multidimensional liquid-based separation techniques. Journal of Chromatography B. 2009;877:1019-39.
[83] Gilar M, Olivova P, Daly AE, Gebler JC. Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. Journal of separation science. 2005;28:1694-703.

[84] Delmotte N, Lasaosa M, Tholey A, Heinzle E, Huber CG. Two-Dimensional Reversed-Phase × Ion-Pair Reversed-Phase HPLC: An Alternative Approach to High-Resolution Peptide Separation for Shotgun Proteome Analysis. Journal of Proteome Research. 2007;6:4363-73.
[85] Toll H, Oberacher H, Swart R, Huber CG. Separation, detection, and identification of peptides by ion-pair reversed-phase high-performance liquid chromatography-electrospray ionization mass spectrometry at high and low pH. Journal of chromatography A. 2005;1079:274-86.
[86] Manadas B, English JA, Wynne KJ, Cotter DR, Dunn MJ. Comparative analysis of OFFGel, strong cation exchange with pH gradient, and RP at high pH for first-dimensional separation of peptides from a membrane-enriched protein fraction. Proteomics. 2009;9:5194-8.
[87] Nozaki Y, Tanford C. [84] Examination of titration behavior. In: Hirs CHW, editor. Methods in Enzymology: Academic Press; 1967. p. 715-34.
[88] Thurlkill RL, Grimsley GR, Scholtz JM, Pace CN. pK values of the ionizable groups of proteins. Protein Sci. 2006;15:1214-8.
[89] Keim P, Vigna RA, Morrow JS, Marshall RC, Gurd FRN. C-13 nuclear magnetic-resonance of pentapeptides of glycine containing central residues of serine, threonine, aspartic and glutamic acids, asparagine, and glutamine .53. J Biol Chem. 1973;248:7811-8.
[90] Richarz R, Wuthrich K. C-13 NMR chemical-shifts of common amino-acid residues measured in aqueous-solutions of linear tetrapeptides H-GLY-GLY-X-L-ALA-OH. Biopolymers. 1978;17:2133-41.
[91] Song C, Ye M, Han G, Jiang X, Wang F, Yu Z, et al. Reversed-phase-reversed-phase liquid chromatography approach with high orthogonality for multidimensional separation of phosphopeptides. Analytical Chemistry. 2010;82:53-6.
[92] Wang Y, Yang F, Gritsenko MA, Wang Y, Clauss T, Liu T, et al. Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. Proteomics. 2011;11:2019-26.
[93] Dizdaroglu M. Weak anion-exchange high-performance liquid-chromatography of peptides. Journal of Chromatography. 1985;334:49-69.
[94] Isobe T, Takayasu T, Takai N, Okuyama T. High-performance liquid-chromatography of peptides on a macroreticular cation-exchange resin - application to peptide-mapping of bence-jones proteins. Analytical Biochemistry. 1982;122:417-25.
[95] Cachia PJ, Vaneyk J, Chong PCS, Taneja A, Hodges RS. Separation of basic peptides by cation-exchange high-performance liquid-chromatography. Journal of Chromatography. 1983;266:651-9.
[96] Mant CT, Hodges RS. Separation of peptides by strong cation-exchange high-performance liquid-chromatography. Journal of Chromatography. 1985;327:147-55.
[97] Alpert AJ, Andrews PC. Cation-exchange chromatography of peptides on poly(2-sulfoethyl aspartamide)-silica. Journal of Chromatography. 1988;443:85-96.
[98] Takahashi N, Takahashi Y, Putnam FW. Two-dimensional high-performance liquid-chromatography and chemical modification in the strategy of sequence-analysis - complete amino-acid-sequence of the lambda light chain of human immunoglobulin-d. Journal of Chromatography. 1983;266:511-22.
[99] Issaq HJ, Chan KC, Janini GM, Conrads TP, Veenstra TD. Multidimensional separation of peptides for effective proteomic analysis. J Chromatogr B. 2005;817:35-47.
[100] Fränzel B, Wolters DA. Advanced MudPIT as a next step toward high proteome coverage. Proteomics. 2011;11:3651-6.
[101] Washburn MP, Wolters D, Yates JR, 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001;19:242-7.
[102] Bayer EM, Bottrill AR, Walshaw J, Vigouroux M, Naldrett MJ, Thomas CL, et al. Arabidopsis cell wall proteome defined using multidimensional protein identification technology. Proteomics. 2006;6:301-11.
[103] Van Hoof D, Munoz J, Braam SR, Pinkse MW, Linding R, Heck AJ, et al. Phosphorylation dynamics

during early differentiation of human embryonic stem cells. Cell Stem Cell. 2009;5:214-26.
[104] Burke TWL, Mant CT, Black JA, Hodges RS. Strong cation-exchange high-performance liquid-chromatography of peptides - effect of non-specific hydrophobic interactions and linearization of peptide retention behavior. Journal of Chromatography. 1989;476:377-89.
[105] McDonald WH, Ohi R, Miyamoto DT, Mitchison TJ, Yates JR. Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. Int J Mass Spectrom. 2002;219:245-51.
[106] Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, et al. Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol. 1999;17:676-82.
[107] Davis MT, Beierle J, Bures ET, McGinley MD, Mort J, Robinson JH, et al. Automated LC-LC-MS-MS platform using binary ion-exchange and gradient reversed-phase chromatography for improved proteomic analyses. Journal of Chromatography B. 2001;752:281-91.
[108] Lim KB, Kassel DB. Phosphopeptides enrichment using on-line two-dimensional strong cation exchange followed by reversed-phase liquid chromatography/mass spectrometry. Analytical Biochemistry. 2006;354:213-9.
[109] Gauci S, Helbig AO, Slijper M, Krijgsveld J, Heck AJ, Mohammed S. Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. Analytical Chemistry. 2009;81:4493-501.
[110] Gruhler A, Olsen JV, Mohammed S, Mortensen P, Faergeman NJ, Mann M, et al. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. Molecular & Cellular Proteomics. 2005;4:310-27.
[111] Villen J, Gygi SP. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nature Protocols. 2008;3:1630-8.
[112] Lemeer S, Pinkse MWH, Mohammed S, van Breukelen B, den Hertog J, Slijper M, et al. Online automated in vivo zebrafish phosphoproteomics: From large-scale analysis down to a single embryo. Journal of Proteome Research. 2008;7:1555-64.
[113] Benschop JJ, Mohammed S, O'Flaherty M, Heck AJR, Slijper M, Menke FLH. Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis. Molecular & Cellular Proteomics. 2007;6:1198-214.
[114] Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell. 2006;127:635-48.
[115] Mischerikow N, Heck AJR. Targeted large-scale analysis of protein acetylation. Proteomics. 2011;11:571-89.
[116] Aivaliotis M, Gevaert K, Falb M, Tebbe A, Konstantinidis K, Bisle B, et al. Large-scale identification of N-terminal peptides in the halophilic archaea Halobacterium salinarum and Natronomonas pharaonis. Journal of Proteome Research. 2007;6:2195-204.
[117] Dormeyer W, Mohammed S, van Breukelen B, Krijgsveld J, Heck AJR. Targeted analysis of protein termini. Journal of Proteome Research. 2007;6:4634-45.
[118] Helbig AO, Gauci S, Raijmakers R, van Breukelen B, Slijper M, Mohammed S, et al. Profiling of N-Acetylated Protein Termini Provides In-depth Insights into the N-terminal Nature of the Proteome. Molecular & Cellular Proteomics. 2010;9:928-39.
[119] Alpert AJ, Petritis K, Kangas L, Smith RD, Mechtler K, Mitulovic G, et al. Peptide Orientation Affects Selectivity in Ion-Exchange Chromatography. Analytical Chemistry. 2010;82:5253-9.
[120] Hennrich ML, van den Toorn HWP, Groenewold V, Heck AJR, Mohammed S. Ultra Acidic Strong Cation Exchange Enabling the Efficient Enrichment of Basic Phosphopeptides. Analytical Chemistry. 2012;84:1804-8.
[121] Takahashi N, Ishioka N, Takahashi Y, Putnam FW. Automated tandem high-performance liquid-chromatographic system for separation of extremely complex peptide mixtures. Journal of Chromatog-

raphy. 1985;326:407-18.
[122] Matsuoka K, Taoka M, Isobe T, Okuyama T, Kato Y. Automated high-resolution 2-dimensional liquid-chromatographic system for the rapid and sensitive separation of complex peptide mixtures. Journal of Chromatography. 1990;515:313-20.
[123] Holland HA, Jorgenson JW. Separation of nanoliter samples of biological amines by a comprehensive 2-dimensional microcolumn liquid-chromatography system. Analytical Chemistry. 1995;67:3275-83.
[124] Dai J, Wang LS, Wu YB, Sheng QH, Wu JR, Shieh CH, et al. Fully Automatic Separation and Identification of Phosphopeptides by Continuous pH-Gradient Anion Exchange Online Coupled with Reversed-Phase Liquid Chromatography Mass Spectrometry. Journal of Proteome Research. 2009;8:133-41.
[125] Nuhse TS, Stensballe A, Jensen ON, Peck SC. Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. Molecular & Cellular Proteomics. 2003;2:1234-43.
[126] Nie S, Dai J, Ning ZB, Cao XJ, Sheng QH, Zeng R. Comprehensive Profiling of Phosphopeptides Based on Anion Exchange Followed by Flow-Through Enrichment with Titanium Dioxide (AFET). Journal of Proteome Research. 2010;9:4585-94.
[127] Alpert AJ. Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. Anal Chem. 2008;80:62-76.
[128] Han GH, Ye ML, Zhou HJ, Jiang XN, Feng S, Jiang XG, et al. Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography. Proteomics. 2008;8:1346-61.
[129] Wagner Y, Sickmann A, Meyer H, Daum G. Multidimensional nano-HPLC for analysis of protein complexes. J Am Soc Mass Spectrom. 2003;14:1003-11.
[130] Zhou F, Sikorski TW, Ficarro SB, Webber JT, Marto JA. Online Nanoflow Reversed Phase-Strong Anion Exchange-Reversed Phase Liquid Chromatography–Tandem Mass Spectrometry Platform for Efficient and In-Depth Proteome Sequence Analysis of Complex Organisms. Analytical Chemistry. 2011;83:6996-7005.
[131] Ficarro SB, Zhang Y, Carrasco-Alfonso MJ, Garg B, Adelmant G, Webber JT, et al. Online Nanoflow Multidimensional Fractionation for High Efficiency Phosphopeptide Analysis. Molecular & Cellular Proteomics. 2011;10.
[132] Phillips HL, Williamson JC, van Elburg KA, Snijders APL, Wright PC, Dickman MJ. Shotgun proteome analysis utilising mixed mode (reversed phase-anion exchange chromatography) in conjunction with reversed phase liquid chromatography mass spectrometry analysis. Proteomics. 2010;10:2950-60.
[133] Alpert AJ. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. J Chromatogr. 1990;499:177-96.
[134] Linden JC, Lawhead CL. Liquid-chromatography of saccharides. Journal of Chromatography. 1975;105:125-33.
[135] Hemstrom P, Irgum K. Hydrophilic interaction chromatography. J Sep Sci. 2006;29:1784-821.
[136] Yoshida T. Peptide separation by Hydrophilic-Interaction Chromatography: a review. J Biochem Biophys Methods. 2004;60:265-80.
[137] Yoshida T. Peptide separation in normal phase liquid chromatography. Analytical Chemistry. 1997;69:3038-43.
[138] Yoshida T. Calculation of peptide retention coefficients in normal-phase liquid chromatography. Journal of Chromatography A. 1998;808:105-12.
[139] Guo Y, Gaiki S. Retention and selectivity of stationary phases for hydrophilic interaction chromatography. Journal of Chromatography A. 2011;1218:5920-38.
[140] Guo Y, Gaiki S. Retention behavior of small polar compounds on polar stationary phases in hydrophilic interaction chromatography. Journal of Chromatography A. 2005;1074:71-80.
[141] Jandera P. Stationary and mobile phases in hydrophilic interaction chromatography: a review. Ana-

lytica Chimica Acta. 2011;692:1-25.
[142] Jandera P. Stationary phases for hydrophilic interaction chromatography, their characterization and implementation into multidimensional chromatography concepts. Journal of Separation Science. 2008;31:1421-37.
[143] Lindner H, Sarg B, Meraner C, Helliger W. Separation of acetylated core histones by hydrophilic-interaction liquid chromatography. Journal of Chromatography A. 1996;743:137-44.
[144] Intoh A, Kurisaki A, Fukuda H, Asashima M. Separation with zwitterionic hydrophilic interaction liquid chromatography improves protein identification by matrix-assisted laser desorption/ionization-based proteomic analysis. Biomedical Chromatography. 2009;23:607-14.
[145] Jiang W, Irgum K. Tentacle-type Zwitterionic stationary phase, prepared by surface-initiated graft polymerization of 3- N,N-dimethyl-N-(methacryloyloxyethyl)-ammonium propanesulfonate through peroxide groups tethered on porous silica. Analytical Chemistry. 2002;74:4682-7.
[146] Fountain KJ, Xu J, Diehl DM, Morrison D. Influence of stationary phase chemistry and mobile-phase composition on retention, selectivity, and MS response in hydrophilic interaction chromatography. Journal of Separation Science. 2010;33:740-51.
[147] McNulty DE, Annan RS. Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. Molecular & Cellular Proteomics. 2008;7:971-80.
[148] Young NL, DiMaggio PA, Plazas-Mayorca MD, Baliban RC, Floudas CA, Garcia BA. High Throughput Characterization of Combinatorial Histone Codes. Molecular & Cellular Proteomics. 2009;8:2266-84.
[149] Di Palma S, Raijmakers R, Heck AJR, Mohammed S. Evaluation of the Deuterium Isotope Effect in Zwitterionic Hydrophilic Interaction Liquid Chromatography Separations for Implementation in a Quantitative Proteomic Approach. Analytical Chemistry. 2011;83:8352-6.
[150] Weng N. Bioanalytical liquid chromatography tandem mass spectrometry methods on underivatized silica columns with aqueous/organic mobile phases. Journal of Chromatography B. 2003;796:209-24.
[151] Grumbach ES, Diehl DM, Neue UD. The application of novel 1.7 μm ethylene bridged hybrid particles for hydrophilic interaction chromatography. Journal of Separation Science. 2008;31:1511-8.
[152] Wohlgemuth J, Karas M, Jiang W, Hendriks R, Andrecht S. Enhanced glyco-profiling by specific glycopeptide enrichment and complementary monolithic nano-LC (ZIC-HILIC/RP18e)/ESI-MS analysis. Journal of Separation Science. 2010;33:880-90.
[153] Xu RN, Boyd B, Rieser MJ, Ei-Shourbagy TA. Simultaneous LC-MS/MS quantitation of a highly hydrophobic pharmaceutical compound and its metabolite in urine using online monolithic phase-based extraction. Journal of Separation Science. 2007;30:2943-9.
[154] Palmisano G, Lendal SE, Engholm-Keller K, Leth-Larsen R, Parker BL, Larsen MR. Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. Nature Protocols. 2010;5:1974-82.
[155] Wu C-J, Chen Y-W, Tai J-H, Chen S-H. Quantitative Phosphoproteomics Studies Using Stable Isotope Dimethyl Labeling Coupled with IMAC-HILIC-nanoLC-MS/MS for Estrogen-Induced Transcriptional Regulation. Journal of Proteome Research. 2011;10:1088-97.
[156] Ding W, Nothaft H, Szymanski CM, Kelly J. Identification and quantification of glycoproteins using ion-pairing normal-phase liquid chromatography and mass spectrometry. Molecular & Cellular Proteomics. 2009;8:2170-85.
[157] Scott NE, Parker BL, Connolly AM, Paulech J, Edwards AVG, Crossett B, et al. Simultaneous Glycan-Peptide Characterization Using Hydrophilic Interaction Chromatography and Parallel Fragmentation by CID, Higher Energy Collisional Dissociation, and Electron Transfer Dissociation MS Applied to the N-Linked Glycoproteome of Campylobacter jejuni. Molecular & Cellular Proteomics. 2011;10:doi: 10.1074/mcp.M000031-MCP201.
[158] Hägglund P, Matthiesen R, Elortza F, Højrup P, Roepstorff P, Jensen ON, et al. An Enzymatic De-

glycosylation Scheme Enabling Identification of Core Fucosylated N-Glycans and O-Glycosylation Site Mapping of Human Plasma Proteins. Journal of Proteome Research. 2007;6:3021-31.
[159] Thaysen-Andersen M, Thøgersen IB, Nielsen HJ, Lademann U, Brünner N, Enghild JJ, et al. Rapid and Individual-specific Glycoprofiling of the Low Abundance N-Glycosylated Protein Tissue Inhibitor of Metalloproteinases-1. Molecular & Cellular Proteomics. 2007;6:638-47.
[160] Takegawa Y, Deguchi K, Keira T, Ito H, Nakagawa H, Nishimura S-I. Separation of isomeric 2-aminopyridine derivatized N-glycans and N-glycopeptides of human serum immunoglobulin G by using a zwitterionic type of hydrophilic-interaction chromatography. Journal of Chromatography A. 2006;1113:177-81.
[161] Takegawa Y, Ito H, Keira T, Deguchi K, Nakagawa H, Nishimura S-I. Profiling of N- and O-glycopeptides of erythropoietin by capillary zwitterionic type of hydrophilic interaction chromatography/electrospray ionization mass spectrometry. Journal of Separation Science. 2008;31:1585-93.
[162] Mysling S, Palmisano G, Hojrup P, Thaysen-Andersen M. Utilizing Ion-Pairing Hydrophilic Interaction Chromatography Solid Phase Extraction for Efficient Glycopeptide Enrichment in Glycoproteomics. Analytical Chemistry. 2010;82:5598-609.
[163] Hao P, Guo T, Li X, Adav SS, Yang J, Wei M, et al. Novel Application of Electrostatic Repulsion-Hydrophilic Interaction Chromatography (ERLIC) in Shotgun Proteomics: Comprehensive Profiling of Rat Kidney Proteome. Journal of Proteome Research. 2010;9:3520-6.
[164] Hao P, Qian J, Ren Y, Sze SK. Electrostatic Repulsion-Hydrophilic Interaction Chromatography (ERLIC) versus Strong Cation Exchange (SCX) for Fractionation of iTRAQ-Labeled Peptides. Journal of Proteome Research. 2011;10:5568-74.
[165] Zhou H, Low TY, Hennrich ML, van der Toorn H, Schwend T, Zou H, et al. Enhancing the Identification of Phosphopeptides from Putative Basophilic Kinase Substrates Using Ti (IV) Based IMAC Enrichment. Molecular & Cellular Proteomics. 2011;10:doi: 10.1074/mcp.M110.006452.
[166] Kweon HK, Håkansson K. Selective Zirconium Dioxide-Based Enrichment of Phosphorylated Peptides for Mass Spectrometric Analysis. Analytical Chemistry. 2006;78:1743-9.
[167] Gan CS, Guo T, Zhang H, Lim SK, Sze SK. A Comparative Study of Electrostatic Repulsion-Hydrophilic Interaction Chromatography (ERLIC) versus SCX-IMAC-Based Methods for Phosphopeptide Isolation/Enrichment. Journal of Proteome Research. 2008;7:4869-77.
[168] Chen X, Wu D, Zhao Y, Wong BHC, Guo L. Increasing phosphoproteome coverage and identification of phosphorylation motifs through combination of different HPLC fractionation methods. Journal of Chromatography B. 2011;879:25-34.
[169] Chien K-y, Liu H-C, Goshe MB. Development and Application of a Phosphoproteomic Method Using Electrostatic Repulsion-Hydrophilic Interaction Chromatography (ERLIC), IMAC, and LC–MS/MS Analysis to Study Marek's Disease Virus Infection. Journal of Proteome Research. 2011;10:4041-53.
[170] Zarei M, Sprenger A, Metzger F, Gretzmeier C, Dengjel J. Comparison of ERLIC–TiO2, HILIC–TiO2, and SCX–TiO2 for Global Phosphoproteomics Approaches. Journal of Proteome Research. 2011;10:3474-83.
[171] Zhang H, Guo T, Li X, Datta A, Park JE, Yang J, et al. Simultaneous Characterization of Glyco- and Phosphoproteomes of Mouse Brain Membrane Proteome with Electrostatic Repulsion Hydrophilic Interaction Chromatography. Molecular & Cellular Proteomics. 2010;9:635-47.

2

Chapter 3

ZIC-HILIC and ZIC-cHILIC provide high resolution separation and increase sensitivity in proteome analysis

Serena Di Palma, Paul J. Boersema, Albert J. R. Heck, and Shabaz Mohammed

Anal Chem. 2011 May 1;83(9):3440-7.

Abstract

The complexity of peptide mixtures that are analyzed in proteomics necessitates fractionation by multidimensional separation approaches prior to mass spectrometric analysis. In this work, we introduce and evaluate Hydrophilic Interaction Liquid Chromatography (HILIC) based strategies for the separation of complex peptide mixtures. The two zwitterionic HILIC materials (ZIC-HILIC and ZIC-cHILIC) chosen for this work differ in the spatial orientation of the positive and negative charged groups. On-line experiments revealed a pH-independent resolving power for the ZIC-cHILIC resin while ZIC-HILIC showed a decrease in resolving power at an acidic pH. Subsequently, we extensively evaluated the performances of ZIC-HILIC and ZIC-cHILIC as first dimension in an off-line 2D-LC strategy in combination with reversed phase (RP), with respect to peptide separation efficiency and how the retention time correlates with a number of peptide physicochemical properties. Both resins allowed the identification of more than 20,000 unique peptides corresponding to over 3500 proteins in each experimental condition from a remarkably low (1.5 μg) amount of starting material of HeLa lysate digestion. The resulting data allows the drawing of a comprehensive picture regarding ZIC- and ZIC-cHILIC peptide separation characteristics. Furthermore, the extent of protein identifications observed from such a level of material demonstrates that HILIC can rival or surpass traditional multidimensional strategies employed in proteomics.

1. INTRODUCTION

A primary driving force in the development of enabling technologies in mass spectrometry (MS) based proteomics is the ability to identify ever smaller amounts of proteins originating from samples of increased complexity[1]. Currently, the best strategy for the analysis of protein mixtures is proteolytic digestion prior to LC-MS/MS to convert hard-to-handle proteins by MS into chemically well-behaved peptides. Such an approach overcomes many issues associated with protein mixture analysis[2]. The greatest challenges reside in cumulative effects, including limited sample material[3-5], fast and non-reproducible sample degradation, the vast dynamic range in protein abundances and the presence of a plethora of post-translational modifications[6]. While proteomics is expected to yield direct biological insights, all of these difficulties render any comprehensive project an exhausting and often humbling exercise[7].

The analysis of peptides in the background of highly complex proteomic samples demands analytical techniques with very high resolving power[8]. To tackle this challenge, different separation strategies have been introduced for the analysis of complex mixtures, highlighting the nearly indispensable role of LC coupled to MS in proteomics research[9]. It is generally accepted that no single stage of chromatography is capable of fully resolving the complex mixture of peptides that results from a complex proteolytic digest. Therefore, combining two or even more orthogonal (multimodal) separation procedures have been introduced, which dramatically improve the overall separation power and result in a much larger number of identified peptides [2, 8, 10-12]. In theory, the peak capacity of the system is the product of the peak capacity of each orthogonal dimension[13]. While it has been demonstrated that increasing the number of dimensions of separation prior to MS analysis increases the number of peptides that may be identified, a balance between the time invested and the final results obtained must be carefully considered[1].

Several different types of chromatography may be used in tandem but the last separation step is typically reversed-phase (RP) LC at nanoliter flow rates since it can provide a high resolving power, effective desalting of the samples and good compatibility with ESI and MS detection[14]. Several first dimensions, not all of them LC-based, have been demonstrated to be very powerful in tandem with RP, including strong-cation exchange (SCX)[2, 15-17], strong-anion exchange (SAX)[18], anion and cation mixed-bed ion exchange (ACE)[19], in-gel [20-22] or in-solution isoelectric focusing techniques (OFFGEL)[23].

Relatively new in the field of proteomics is the use of hydrophilic interaction liquid chromatography (HILIC [24]) as a first dimension for 2D LC proteomic strategies[13] with promising results[25, 26]. Compared to all commonly used peptide separation modes, HILIC has one of the highest degrees of orthogonality to RPLC[13]. Although the retention in HILIC increases with increasing polarity or hydrophilicity, opposite to the trend observed in RP, HILIC is not a variation of NPLC as the HILIC technique employs water-miscible solvents compatible with mass spectrometry and the elution is achieved by a water gradient [24, 27-29]. Samples are loaded at high organic solvent concentration and eluted by increasing the polarity of the mobile phase.

Interestingly, several quite different HILIC stationary phases have been introduced[30] including derivatized silica materials, such as the cation exchanger polysulphoethyl A [24], the weak cation exchanger Polycat A[31], the weak anion exchanger PolyWAX[32], TSKgel amide-80[26, 33], and zwitterionic ZIC-HILIC[25, 34-36]. Although their exact mechanisms of chromatographic action are quite different, they all generate a hydrophilic layer around their functional groups. Zwitterionic separation materials, which we employ in this study, are uniquely characterized by carrying both positive and negative charges on the surface. The electrostatic interactions between those two oppositely charged groups in close proximity at a stoichiometric ratio, relatively weaken the interactions of this stationary phase with the charged analytes when

compared to normal ionic exchangers such as SCX.
In the present study we explore the capabilities of two different HILIC zwitterionic stationary phases in a MuDPIT proteomics workflow, evaluating them in terms of peptide separation efficiency and total peptide/protein identification when used in combination with RP-LC MS/MS. Analysis of cell lysate digest at amounts equating to approximately 1.5 micrograms were performed and trends in the identified peptide populations evaluated. Under such conditions, each configuration was capable of identifying approximately 3500 proteins and 20,000 unique peptides demonstrating that HILIC can represent a powerful foundation for a sensitive multidimensional strategy.

2. EXPERIMENTAL SECTION

Chemicals and Materials. Bovine serum albumin, α- and β-casein, iodoacetamide and ammonium acetate were supplied by Sigma-Aldrich (Steinheim, Germany). Lys-C was obtained from Roche Diagnostics (Mannheim, Germany). Ammonium bicarbonate and dithiothreitol (DTT) were purchased from Fluka (Buchs, Switzerland). HPLC-S gradient grade acetonitrile was purchased from Biosolve (Valkenwaard, The Netherlands). Acetic acid was obtained by Merck KGaA (Darmstadt, Germany) and high purity water obtained from a Milli-Q purification system (Millipore, Bedford, MA). Formic acid was obtained from Merck (Darmstadt, Germany). Sep-Pak Vac tC18 1 cm3 cartridges were obtained from Waters Corporation (Milford, MA).

Preparation of a Protein Standard Mixture. For the optimization and evaluation of the on-line ZIC and ZIC-cHILIC system, we prepared a model standard peptide mixture consisting of combined protein digests of bovine serum albumin (BSA) and α- and β-casein. Each protein was digested separately and subsequently mixed in a 1:1:1 ratio. 2.5 μL of 45 mM DTT were added to 10 μl of 4 μg/μL dissolved protein and incubated at 37 °C for 30 min. Subsequently, 2.5 μL of 110 mM iodoacetamide were added to the mixture and kept at room temperature in the dark for 30 min. The mixture was diluted 6 times with 50 mM ammonium bicarbonate and 1 μg of trypsin was added for overnight digestion at 37 °C. The mixture was diluted in buffer A for subsequent LC separation.

Sample preparation. A total of 5.6 μL of 45 mM DTT was added to 40 μL of 1 μg/μL dissolved HeLa cell lysate and incubated at 50 °C for 30 min. After cooling down, 12.5 μL of 100 mM iodoacetamide were added to the mixture and kept at room temperature in the dark for 30 min. Subsequently, 5 μL of a 0.1 μg/μL Lys-C solution were added to the mixture and kept at 37 °C for 4 h. The mixture was diluted 4 times with a 50 mM ammonium bicarbonate solution (2M urea final concentration). A total of 1.2 μg of trypsin was added for overnight digestion at 37 °C. The sample was desalted using Sep-pak C18 and the eluent was dried in vacuo, reconstituted in 10% formic acid and stored at -20 °C until required. For subsequent LC separation the sample was diluted in buffer A.

HILIC-LC Buffers. For HILIC at pH 6.8 buffer A was 95% acetonitrile, 0.5% acetic acid and 5 mM ammonium acetate; buffer B was 5 mM ammonium acetate. For pH 3.5 buffer A was 95% acetonitrile, 2% formic acid and 5 mM ammonium acetate; buffer B 0.07% formic acid and 5 mM ammonium acetate.

RP-LC Buffers. Buffer A for nano-LTQ-Orbitrap analysis was 0.1 M acetic acid; buffer B was 80% acetonitrile and 0.1 M acetic acid.

On-line system: HILIC-MS. ZIC/ZIC-cHILIC–MS was performed on an "inert" Dionex "Ulitimate" LC system using a vented column setup. The trapping column was a ZIC-HILIC, 100 μm × 20 mm, 3.5 μm, 200 Å or a ZIC-cHILIC 100 μm × 20 mm, 5 μm, 200 Å; the analytical column was ZIC-HILIC, 75 μm × 250 mm, 3.5 μm, 200 Å or a ZIC-cHILIC 75 μm × 250

mm, 5 μm, 200 Å. All columns were packed in-house. Trapping was performed at 5 μL/min for 10 min at 100% buffer A; elution was achieved with a gradient of 0-55 % buffer B in 40 min at a flow rate of 0.35 mL/min passively split to 300 nL/min. Column output was coupled with an LTQ-FTICR mass spectrometer. Nanospray was achieved using a distally coated fused silica emitter (New Objective, Cambridge, MA) biased to 2.2 kV. The LTQ-FT-ICR was operated in positive ion mode, from 350 to 1500 m/z in MS mode and with an AGC value of 5.00e+05. The two most intense parent ions (with a threshold above 500) were isolated and fragmented by CID in data-dependent mode with an AGC value of 1.00e+04. Ions were fragmented in the linear ion trap using CID with normalized collision energy of 35 and 30 ms activation time.

Off-line 2D-LC system: HILIC-RP-MS. First dimension analysis as described above. One-min fractions were collected during the elution in a 96 plate with each well containing 40 μL of 10% formic acid per fraction. No additional sample modification was performed before the fractions' analysis. A volume of 20 μL of collected fractions was used for subsequent nanoLC-LTQ-Orbitrap-MS (Orbitrap Discovery, Thermo). An Agilent 1100 series LC system was equipped with a Reprosil, 100 μm × 20 mm, 5 μm, 120 Å trapping column and a Reprosil, 50 μm × 250 mm, 3 μm, 120 Å analytical column. Trapping was performed at 5 μL/min for 10 min with solvent A (0.1M acetic acid); elution was achieved with a gradient from 13 to 28 % in 57 min and up to 50 % in 25 min of solvent B, with a flow rate of 0.35 mL/min passively split to 100 nL/min. The column effluent was directly introduced into the ESI source of the MS. Nanospray was achieved using a distally coated fused silica emitter biased to 1.7 kV. The mass spectrometer was operated in data dependent mode to automatically switch between MS and MS/MS.

Protein identification. Raw MS data were converted to peak lists using Bioworks Browser software, version 3.1.1. These peak lists were searched using the Mascot search engine (version 2.2.0) against a concatenated target-decoy database containing the IPI Human database (version 3.52, 148380 sequences). The settings were: trypsin with maximum 2 missed cleavages, carbamidomethyl (C) as fixed modification, oxidation (M) and phosphorylation (S, T) as variable modifications. In order to circumvent issues related to MS calibration and to generate a significant level of false positives, peptides tolerance was initially set to 50 ppm for the MS level mass accuracy and 0.9 Da for MS/MS. Subsequently, the peptide identifications were filtered for true mass accuracy (precursor mass should be between >-5ppm to <15ppm) and to a minimum score of 20 using an in-house developed tool, called Rockerbox [37]. The result corresponded to an FDR below 1%. Distributions for the mass accuracy of the filtered dataset can be found in supplementary Figure 5 and the peptide score distribution can be found in supplementary Figure 6. Supplementary data 1 contains zipped files of the direct Mascot HTML output of the filtered files. For the purposes of identifying peptide populations for the HILIC fractionation, a further level of filtering related to unique peptides was performed and if a peptide was found in more than one fraction, it was assigned only to the first fraction. Final peptide lists can be found in Supplementary Table 3.

3. RESULTS AND DISCUSSION

The aim of this study was to explore the potential of two different HILIC zwitterionic stationary phases in a MuDPIt proteomics workflow. Therefore, we first optimized and evaluated, using standard proteins mixtures, two on-line nanoflow zwitterionic (ZIC-HILIC and ZIC-cHILIC) MS/MS configurations (Figure 1).

3.1 On-line system: HILIC-MS

Separation of peptides using zwitterionic materials in HILIC mode involves both hydrophilic

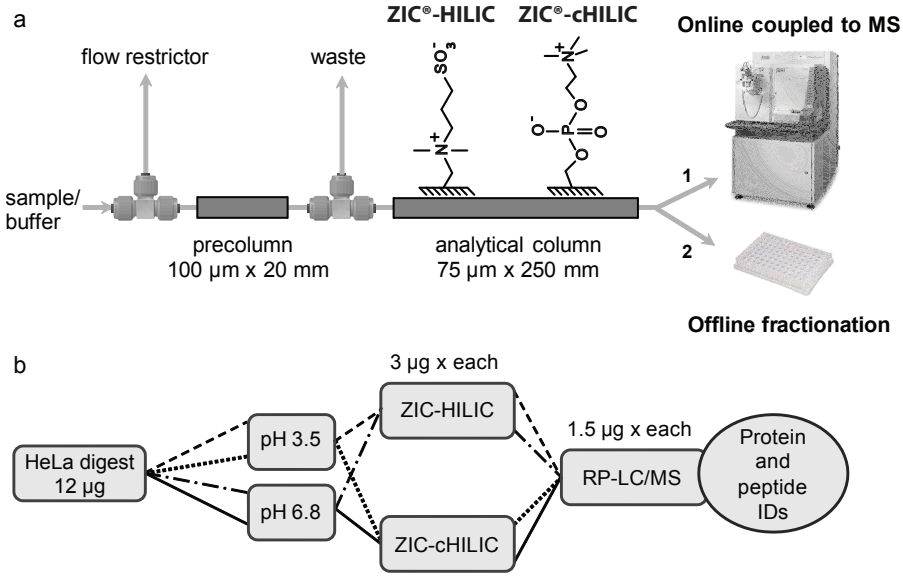

Figure 1: *a) Schematic design of the nano on-line HILIC-MS systems (1) and first-dimensional separation configurations (2) used in this study. After fractionation, the eluent was collected as one-minute fractions in a 96 well plate containing 10 % formic acid and subsequently analysed by nano-RP-LC-MS. b) Schematic workflow of different settings employed with the above offline configuration (2) for all experiments performed: pH 3.5, ZIC-HILIC (- -); pH 6.8, ZIC-HILIC (— ·); pH 3.5, ZIC-cHILIC (···); pH 6.8, ZIC-cHILIC (—). For each fractionation, the same amount of sample, corresponding to 3 µg of lysate, were used and 1.5 µg of each were analysed in the RP second dimension.*

and ionic interactions. ZIC-HILIC is a sulfobetaine type of zwitterionic stationary phase that possesses a slight negative surface charge due to the spatial orientation of the sulfonic group at the distal end of the zwitterionic moiety[38]. The inverted zwitterionic stationary phase, here termed ZIC-cHILIC, has a cholinate group as zwitterionic material and exhibits an opposite charge arrangement, with the quaternary ammonium as a distal moiety and the negatively charged phosphoric group in the proximal location to the bead surface (Figure 1). We hypothesized that this opposite spatial orientation would affect peptide separation due to the enhanced interactions of the distal ammonium group with negatively charged acid peptides and/or phosphopeptides. To test such a hypothesis, we examined first the chromatographic behavior and retention time of several model peptides from standard mixtures using either ZIC-HILIC or ZIC-cHILIC at two different pHs. We showed previously[25] that HILIC can be performed at different pHs, and pH 3.5 and 6.8 are optimal for orthogonality (with RP) and separation, respectively. A benefit of zwitterionic stationary phases relies on the fact that the charge of the chromatographic material does not change at different pH values. However, the charge state of the peptide in the buffer is affected by the pH, which in turn affects the peptide's hydrophilicity and, thus, its retention.

A first step in our evaluation was to develop an effective on-line nano-LC system based on either ZIC-HILIC or ZIC-cHILIC stationary phases. The initial efforts involved the optimization of mobile phases (in terms of organic content and salt concentration), loading buffers, flow rate and gradient. The choice of the optimal mobile phases was related not only to

electrospray ionization and MS compatibility[39], but also to separation efficiency and peptide solubility. According to the hydrophilic partitioning model[24], it is crucial to keep the water concentration in the eluent within certain limits (> 3%) to maintain sufficient hydration of the stationary phase particles and to ensure minimal peptide solubility.
To test these configurations, equal amounts of peptide standard mixtures were injected into the ZIC-HILIC or ZIC-cHILIC LC system. Comparing the two ZIC-HILIC chromatograms, the resolving power at pH 3.5 was slightly lower than at pH 6.8 (Figure 2 a, b), while using ZIC-cHILIC the separation was equally good at both pH values: peaks were found to have a FWHM < 0.3 min and, particularly at higher pH, were sharper and almost baseline-separated (Figure 2 c, d).

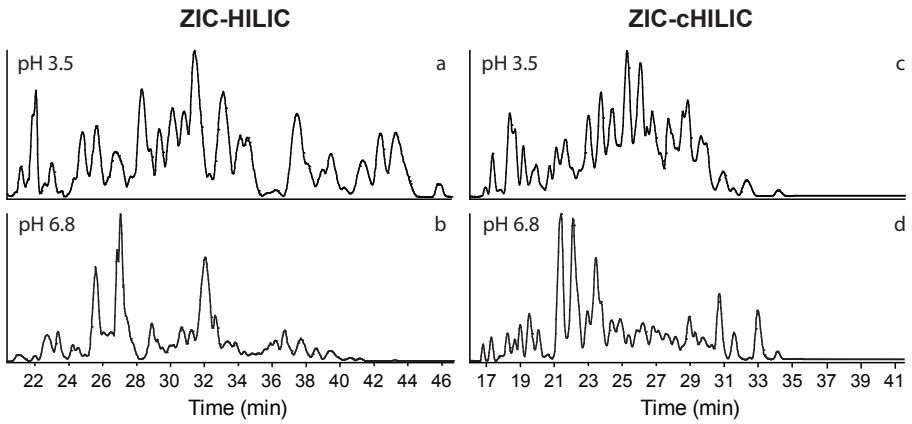

Figure 2: *LC-MS chromatograms of a 2 pmol standard peptide mixture using the ZIC-HILIC and ZIC-cHILIC online configurations (trap column: 0.1 x 20 mm, analytical column: 0.075 x 20mm) connected to an LTQ-FT-ICR mass spectrometer. (a) ZIC-HILIC base peak chromatograms with a running buffer at a pH of 3.5, (b) ZIC-HILIC base peak chromatograms with a running buffer at a pH of 6.8, (a) ZIC-cHILIC base peak chromatograms with a running buffer at a pH of 3.5, (a) ZIC-cHILIC base peak chromatograms with a running buffer at a pH of 6.8.*

In general, the observed patterns of elution appeared to be consistent with a HILIC-based separation mode in which polar peptides are more retained in comparison to less polar peptides. To further investigate trends correlated to the peptide net charge, we compared the retention times of 'model peptides'chosen from the standard mixture at the two pH values. Their extracted ion chromatograms for the ZIC-cHILIC configuration are shown in Figure 3, while their sequences, amino acid composition in terms of basic, acidic, polar and non-polar residues and relative net charge at the two pH conditions are listed in Supplementary Table 1. Examination of the retention times of these peptides showed a trend of longer retention time at higher pH, related to the increasing number of ionized acidic residues and, thus, increased polarity. For example, the presence of six acidic residues in the peptides 6 and 7 (see Supplementary Table 1) adds six extra negative charges at a pH of 6.8 and increases the overall polarity of these peptides, enhancing the interaction with the zwitterionic material and, thus, extending the retention time. A relevant shift in the retention times of those two peptides was indeed observed at a pH of 3.5, most likely due to the change in ionization of the acidic residues and net charge variations. To confirm our hypothesis, we also analyzed the behavior of peptides with less acidic residues whose net charged is less influenced by the change in pH. Retention of peptides 3

and 5, which have only three acidic residues, were only partly affected by the pH of the buffer. Similar behavior was observed for the ZIC-HILIC material (data not shown and described earlier[25]) showing the same dependency on the pH of the buffer. We also observed that most of the phosphopeptides eluted at the end of the gradient, which is not surprising since phosphopeptides are generally more polar than their non-phosphorylated counterpart[26]. However, in addition to phosphorylated peptides, those peptides containing multiple hydrophilic amino acids also eluted later, providing thus no obvious enrichment of phosphopeptides.

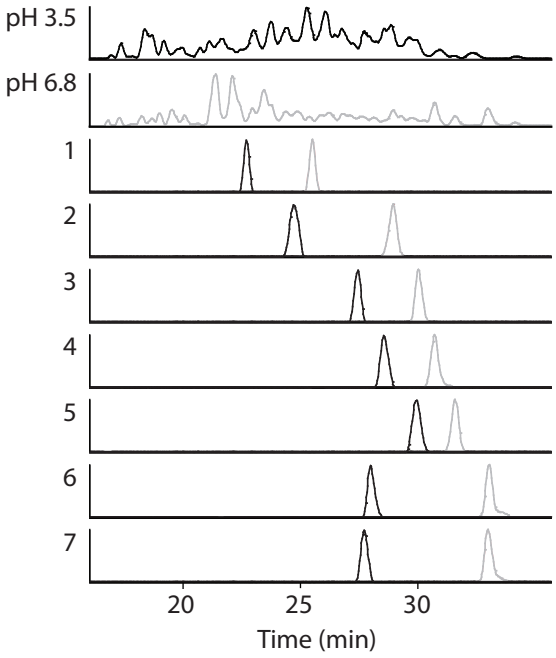

Figure 3: *Extracted ion chromatograms from 7 selected peptides separated by ZIC-cHILIC online using a running buffer pH of 3.5 (top) and 6.8 (bottom). The peptide numbers correspond with those used in Supplementary Table 1.*
(1) LFTFHADICTLPDTEK (2) VPQLEIVPNpSAEER (3) YKVPQLEIVPNpSAEER (4) ECCHGDLLECADDR (5) TCVADESHAGCEK (6) DIGpSEpSTEDQAMEDIK (7) FQpSEEQQQTEDELQDK.

3.2 Gradient optimization

We next optimized the gradient in terms of separation for an off-line configuration. Concurrently, the aim was to find a good compromise in terms of optimum resolution in the first dimension and acceptable time invested for the analysis of all the fractions in the second dimension in an offline configuration. Gradients of 45, 60, 75, 90 and 120 minutes were evaluated for ZIC-cHILIC at pH 3.5 (Supplementary Table 2) and 6.8 (data not shown). The separation obtained with the short gradients of 45 and 60 min provided the best peak capacity and a resolution comparable to RP (Supplementary Figure 1). No further improvements were achieved through shallower, longer, gradients. Similar results and behavior were observed also for the ZIC-HILIC system. The optimal resolution at shorter analysis times prompted us to choose the 60-min run time as optimal for an off-line fractionation in our 2D-LC strategy.

3.3 Off-line 2D-LC system: HILIC-RP-MS

To prepare our configuration for adaptation into a high-throughput proteomics workflow, we designed an off-line system combining ZIC-HILIC or ZIC-cHILIC separation, as a fractionation step, with a RP second dimension (Figure 1b). Since the high organic buffer used in first dimension does not allow a direct hyphenation with RP, the eluent of the HILIC separation was collected during the run as 1-minute fractions in a 96 well plate already containing 40 µL of 10% FA in each well. Thus, sample handing was reduced and the fractions were sufficiently large and aqueous, decreasing evaporation, allowing direct compatibility with the subsequent RP nano-LC-MS/MS analysis.
We examined the performance of the 2D-HILIC-RP systems with complex proteome samples

consisting of a human cell lysate digestion. Separations for 3 µg of HeLa cell lysate digest were performed by ZIC-HILIC and ZIC-cHILIC as first dimension at a pH of 3.5 or 6.8 using the same optimized 60-minute analysis. In the 20-minute range, wherein most peptides were expected to elute, 20 single-minute fractions were collected. Half of each collected HILIC fraction was subsequently injected and further separated and analyzed by a nano RP-LC run using a 120-min analysis time. Overall we thus analyzed by LC MS/MS maximally 1.5 microgram of a cell lysate digestion.

The use of ZIC-HILIC as a first dimension allowed the identification of 28316 unique peptides assigned to 4180 proteins at pH 6.8 and 14482 peptides assigned to 2816 proteins at pH 3.5. The better separation of ZIC-HILIC at pH 6.8 led to more peptides being identified after fractionation. In contrast, near-identical numbers of peptides (and proteins) were identified using ZIC-cHILIC at both pH 6.8 and pH 3.5. (Figure 4), which is in line with the similar chromatographic resolution observed for the simple mixtures analyzed (see Figure 2 c, d). Note, these experiments do not indicate which set of conditions are superior since no replicates are performed but the results do demonstrate that the overall separation efficiency of both ZIC-HILIC and ZIC-cHILIC is comparable (Supplementary Table 3). For comparison, analyzing 1.5 µg of the same sample solely on the second RP dimension led to less than 800 proteins and approximately 2200 unique peptides being identified, highlighting the high-quality separation power of our 2D-HILIC-RP-LC-MS system.

The distribution of unique peptides identified in each of the HILIC fractions was investigated for both the ZIC-HILIC and ZIC-cHILIC configurations running at either neutral or acidic pH (Figure 5). Peptides were nearly equally distributed throughout the 20-minute elution window, more so when operating the HILIC systems at a pH of 6.8. We found that shorter peptides eluted in general in the early HILIC fractions, whereas larger peptides were found in the later fractions, possibly due to the fact that the polar backbone provides a major contribution to the retention profile. A detailed comparative evaluation in terms of basicity, acidity, hydrophilicity, hydrophobicity, net charge and miss cleavages is provided in Figure 6 for ZIC-

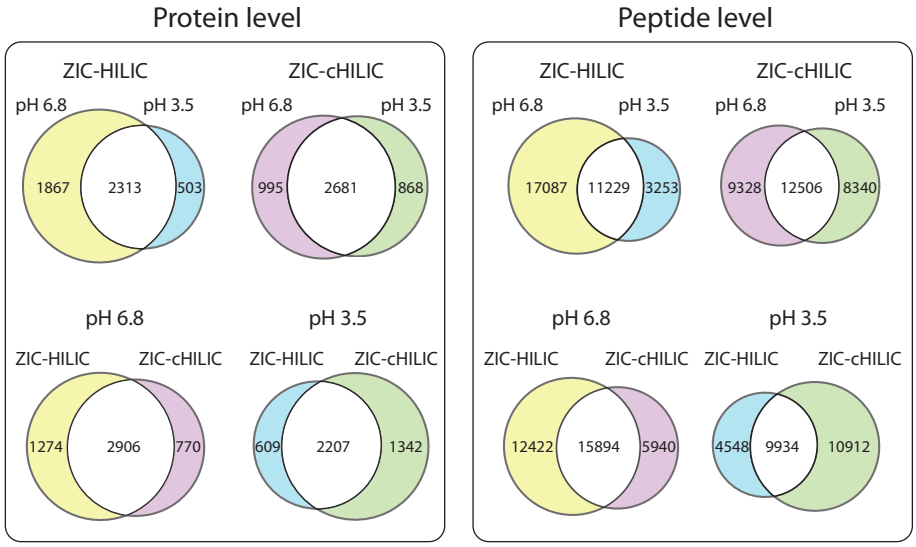

Figure 4: *Total number of unique protein and peptide identifications detected from 1.5 microgram of cell lysate digest, using the different evaluated zwitterionic HILIC RP LC MS/MS configurations. The Venn diagrams reveal the numbers and overlap of the a) proteins and b) peptides.*

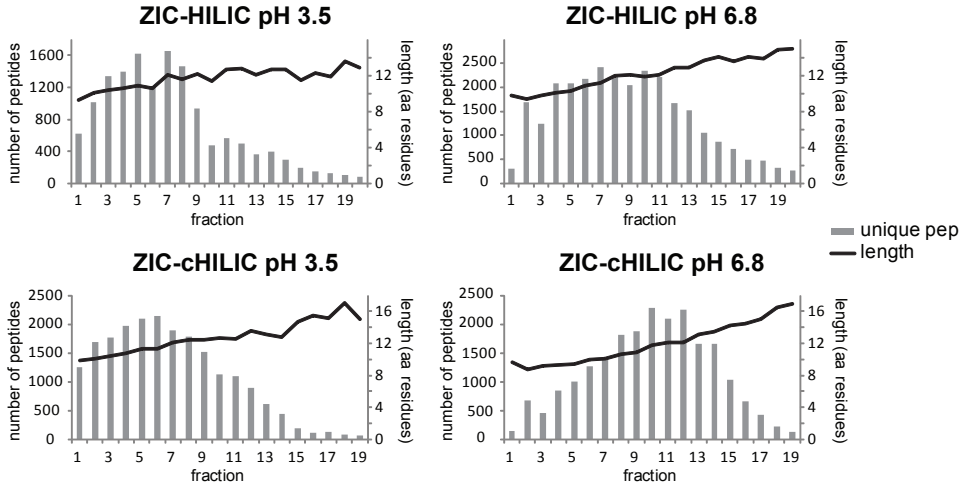

Figure 5: *Fractional distributions of unique peptides (left y-axis) from a Hela lysate digest after ZIC-HILIC and ZIC-cHILIC fractionation at different buffer pHs. Additionally the solid line shows the average length of peptides per fraction (right y-axis).*

cHILIC (ZIC-HILIC analysis available as Supplementary Figure 2).
Analysis of the effect of the peptides net charge on the retention time in HILIC showed an opposite trend at the two applied pH values (green lines), in line with the above described results (Figure 3 and Supplementary Table 1). A strong negative charge trend was observed at pH 6.8, likely due to deprotonation of acidic residues and the increased number of negatively charged phosphopeptides, while a positive charge trend at pH 3.5 could be observed and is most likely the consequence of peptides containing histidine residues, which are deprotonated above pH 6. Similar net charge trend lines were emerging from the ZIC-HILIC fraction analysis, underlying that the change of HILIC material only partially affects this mechanism.
The trend lines also clearly revealed that the HILIC selectivity is mostly driven by the peptide hydrophilicity, consistently with the hydrophilic partitioning model. As expected, early fractions were largely comprised of hydrophobic peptides (violet line) while an obvious increase in

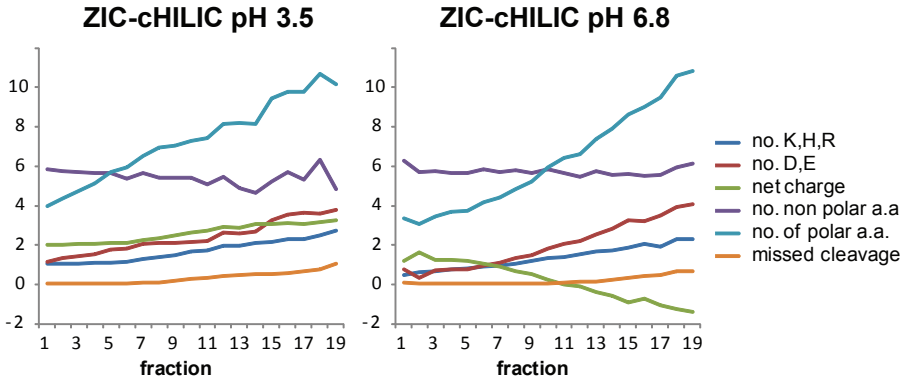

Figure 6: *Analyses of peptide characteristics observed in ZIC-cHILIC fractions at a pH of a) 3.5 and b) 6.8 analyzing the effect of basicity, acidity, net charge, polarity and number of missed cleavages on HILIC retention time.*

polarity (light blue line) was observed in the late fractions at either pHs, related to the increased number of acidic (red line) and basic (blue line) residues in the peptides. In summary, the observed trends confirm that HILIC separation is governed by a mixed-mode mechanism based on both hydrophilic and ionic interactions.
In our earlier work[25] we established that the orthogonality of ZIC-HILIC to RP is dependent on the pH of the running buffer. At a pH of 3 a higher orthogonality was seen than at higher pH. In this work, we observe that this pH dependent orthogonality is again apparent both for the ZIC-HILIC and ZIC-cHILIC configurations (Supplementary Figure 3). On average larger differences in retention times were observed in the late HILIC fractions compared to the earlier fractions, confirming the effect of changing the net charge state of the peptides. Nevertheless, in general our data highlight that high separation power (obtained at a pH of 6.8) is a more relevant aspect than increased orthogonality (obtained at a pH of 3.5) when the aim is to detect as many peptides as possible from complex lysate digests.
Since the observed trend lines are very similar for ZIC-HILIC and ZIC-cHILIC, we conclude that the separation efficiency is not significantly affected by the different charge arrangement in the functional groups of these stationary phases. The zeta-potential measurements of a phosphorylcholine type of zwitterionic stationary phase [36], which should behave as our ZIC-cHILIC, have been previously reported. The authors studied the surface charge properties of these materials (alongside sulfobetaine derivatives) in a wide range of pH (pH 3-7) and showed that all these materials have a negative surface charge, although their spatial charge arrangement 'should' favor generating a positive charge. We believe that our data is indicating similar results for ZIC-cHILIC and ZIC-HILIC stationary phases; however, further investigation is required.

3.4 Enrichment of phosphorylated peptides

Prompted by the work of McNulty et al., who used a HILIC stationary phase coupled with IMAC for phosphoproteomic studies[26], we next investigated the suitability of our zwitterionic HILIC configurations for targeted analysis of phosphopeptides. Phosphopeptides are in theory ideal candidates for enrichment because of their increased overall hydrophilicity and their higher negative charge.
The distribution of phosphorylated peptides under ZIC-HILIC and ZIC-cHILIC separations was evaluated at pH 6.8 and pH 3.5 (Figure 7). Although a higher amount of phophopeptides was found with ZIC-HILIC at pH 6.8, no obvious enrichment in specific HILIC fractions was observed at either pHs. The spread distribution of these phosphopeptides through all fractions convinced us that zwitterionic HILIC alone as used here cannot be used as an efficient strategy for phosphoproteomics.
It is noteworthy that we detected nearly 900 unique phosphopeptides in the analysis of only 1.5 μg of lysate using the ZIC-HILIC configuration at a pH of 6.8 and almost 500 phosphopeptides using ZIC-cHILIC. To our knowledge, this is a remarkably high number of phosphopeptides for such a small amount of starting material demonstrating, once again, the separation power of the configuration. Interestingly, this phosphopeptides population was obtained purely through sample complexity reduction and no enrichment.
We believe that the enrichment efficiency of phosphopeptides by HILIC may be possible by the addition of an ion pairing reagent, such as TFA, in the mobile phase minimizing the overlap of phosphopeptides and non-phosphorylated peptides, as the addition of a hydrophobic ion pairing reagents, will preferentially reduce polar/ electrostatic interactions of 'regular' peptides[40] [41] and thus increase the hydrophilicity difference between 'regular' and phosphopeptides. Unfortunately, TFA in the buffer will significantly lower the pH of the mobile phase, which will affect the long-term stability of the stationary phase.

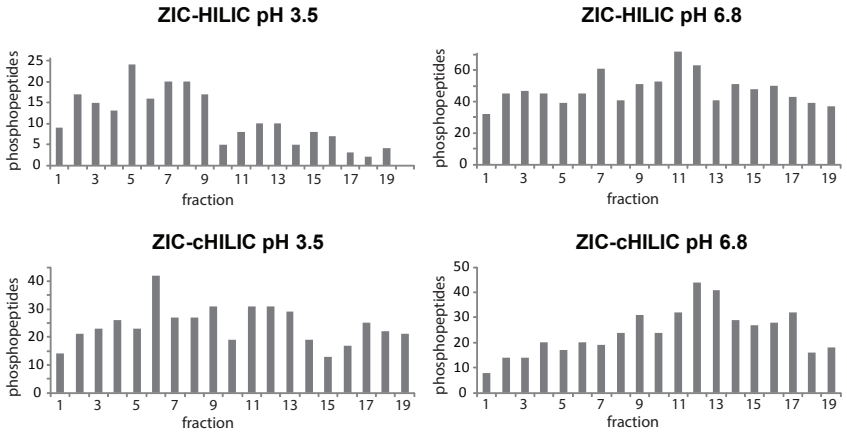

Figure 7: *Distribution of phosphorylated peptides over ZIC-HILIC and ZIC-cHILIC fractions at a pH of 3.5 and 6.8.*

4. CONCLUSIONS

One of the holy grails of proteomics, still far out of reach, is the single cell single proteome concept. Analyzing the proteome of small amount of cells is a challenging task at present, and a suite of special analytical tools is required to deal with these types of sample. Many techniques commonly used to handle large number of cells with good outcomes cannot be directly applied to a small number of cells. The zwitterionic HILIC based approaches demonstrated here describe a major step forward towards more sensitive proteome analysis. We discovered that the zwitterionic materials which have opposite orientations generated similar separations although the separation power of ZIC-cHILIC was more independent of pH. The use of two orthogonal separation techniques such as zwitterionic-HILIC and RP, the first capable to retain peptides based on the hydrophilicity and weak ionic interactions, the second one based on the hydrophobicity, in combination with nanoflow rate LC system and high resolution mass spectrometry, has allowed the identification of thousands of proteins from a very modest amount (1.5 μg) of starting material.

5. ACKNOWLEDGEMENTS

This work was supported by the Netherlands Proteomics Centre, embedded in the Netherlands Genomics Initiative. We kindly acknowledge Dr. Wen Jiang from Merck SeQuant for providing the zwitterionic HILIC materials.

6. SUPPORTING INFORMATION AVAILABLE

Tables listing peptides and proteins identified over ZIC-HILIC and ZIC-cHILIC fractionations at pH 3.5 and 6.8. Supplementary Table 1 lists the number of basic, acidic, non-polar and polar residues, and net charges at pH 3.5 and 6.8 for the analysis of peptides used in the Figure 3. Supplementary Table 2 contains the 5 LC methods used for the gradient optimization. Supplementary Table 3 summarizes the results for proteins and peptides identifications obtained using the 4 different fractionation methods for the analysis of the same amount of HeLa digest. Supplementary Figure 1 shows LC-MS chromatograms obtained by the ZIC-cHILIC configuration with a running buffer of pH 3.5 using different gradient times, ranging from 45 to 120 min. Supplementary Figure 2 shows the analysis of peptide characteristics observed

in ZIC-HILIC fractions at a pH of 3.5 and 6.8 in terms of basicity, acidity, net charge, polarity and number of missed cleavages. Supplementary Figure 3 shows the normalized peptide retention times of all peptides detected using ZIC-cHILIC and ZIC-HILIC at the two applied pH conditions. Supplementary Figure 4 shows the distribution of phosphorylated peptides over ZIC-HILIC and ZIC-cHILIC fractions at a pH of 3.5 and 6.8. Supplementary Figure 5 shows the mass accuracy of the filtered datasets for all the four different conditions and supplementary Figure 6 the peptide score distribution at selected ion score ranges. This material is available free of charge at http://pubs.acs.org.

7. REFERENCES

[1] de Godoy LMF, et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature. 2008;455:1251-U60.
[2] Washburn MP, et al. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001;19:242-7.
[3] Umar A, et al. NanoLC-FT-ICR MS improves proteome coverage attainable for approximately 3000 laser-microdissected breast carcinoma cells. Proteomics. 2007;7:323-9.
[4] Wang N, et al. Development of Mass Spectrometry-Based Shotgun Method for Proteome Analysis of 500 to 5000 Cancer Cells. Analytical Chemistry. 2010;82:2262-71.
[5] Waanders LF, et al. Quantitative proteomic analysis of single pancreatic islets. Proc Natl Acad Sci U S A. 2009;106:18902-7.
[6] Rabilloud T. Two-dimensional gel electrophoresis in proteomics: Old, old fashioned, but it still climbs up the mountains. Proteomics. 2002;2:3-10.
[7] Tyers M, et al. From genomics to proteomics. Nature. 2003;422:193-7.
[8] Vollmer M, et al. Optimization of two-dimensional off-line LC/MS separations to improve resolution of complex proteomic samples. Analytical Chemistry. 2004;76:5180-5.
[9] Issaq HJ, et al. Multidimensional separation of peptides for effective proteomic analysis. J Chromatogr B. 2005;817:35-47.
[10] Romijn EP, et al. Recent liquid chromatographic-(tandem) mass spectrometric applications in proteomics. Journal of Chromatography A. 2003;1000:589-608.
[11] Wu CC, et al. The application of mass spectrometry to membrane proteomics. Nat Biotechnol. 2003;21:262-7.
[12] Zhou H, et al. A fully automated 2-D LC-MS method utilizing online continuous pH and RP gradients for global proteome analysis. Electrophoresis. 2007;28:4311-9.
[13] Gilar M, et al. Orthogonality of separation in two-dimensional liquid chromatography. Anal Chem. 2005;77:6426-34.
[14] Wang WX, et al. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. Analytical Chemistry. 2003;75:4818-26.
[15] Dong MQ, et al. Quantitative mass spectrometry identifies insulin signaling targets in C-elegans. Science. 2007;317:660-3.
[16] Dormeyer W, et al. Targeted analysis of protein termini. Journal of Proteome Research. 2007;6:4634-45.
[17] Taouatas N, et al. Strong Cation Exchange-based Fractionation of Lys-N-generated Peptides Facilitates the Targeted Analysis of Post-translational Modifications. Molecular & Cellular Proteomics. 2009;8:190-200.
[18] Dai J, et al. Fully Automatic Separation and Identification of Phosphopeptides by Continuous pH-Gradient Anion Exchange Online Coupled with Reversed-Phase Liquid Chromatography Mass Spectrometry. Journal of Proteome Research. 2009;8:133-41.
[19] Motoyama A, et al. Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. Analytical Chemistry. 2007;79:3623-34.
[20] Krijgsveld J, et al. In-gel isoelectric focusing of peptides as a tool for improved protein identification. Journal of Proteome Research. 2006;5:1721-30.
[21] Essader AS, et al. A comparison of immobilized pH gradient isoelectric focusing and strong-cation-exchange chromatography as a first dimension in shotgun proteomics. Proteomics. 2005;5:24-34.
[22] Cargile BJ, et al. Immobilized pH gradients as a first dimension in shotgun proteomics and analysis of the accuracy of pI predictability of peptides. Electrophoresis. 2004;25:936-45.
[23] Hubner NC, et al. Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. Proteomics. 2008;8:4862-72.

[24] Alpert AJ. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. J Chromatogr. 1990;499:177-96.
[25] Boersema PJ, et al. Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. Journal of Proteome Research. 2007;6:937-46.
[26] McNulty DE, et al. Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. Molecular & Cellular Proteomics. 2008;7:971-80.
[27] Naidong W. Bioanalytical liquid chromatography tandem mass spectrometry methods on underivatized silica columns with aqueous/organic mobile phases. Journal of chromatography B, Analytical technologies in the biomedical and life sciences. 2003;796:209-24.
[28] Hagglund P, et al. A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. Journal of Proteome Research. 2004;3:556-66.
[29] Hemstrom P, et al. Hydrophilic interaction chromatography. J Sep Sci. 2006;29:1784-821.
[30] Jandera P. Stationary phases for hydrophilic interaction chromatography, their characterization and implementation into multidimensional chromatography concepts. Journal of Separation Science. 2008;31:1421-37.
[31] Lindner H, et al. Application of hydrophilic-interaction liquid chromatography to the separation of phosphorylated H1 histones. Journal of Chromatography A. 1997;782:55-62.
[32] Alpert AJ. Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. Anal Chem. 2008;80:62-76.
[33] Yoshida T. Peptide separation in normal phase liquid chromatography. Analytical Chemistry. 1997;69:3038-43.
[34] Intoh A, et al. Separation with zwitterionic hydrophilic interaction liquid chromatography improves protein identification by matrix-assisted laser desorption/ionization-based proteomic analysis. Biomedical Chromatography. 2009;23:607-14.
[35] Jiang W, et al. Tentacle-type Zwitterionic stationary phase, prepared by surface-initiated graft polymerization of 3- N,N-dimethyl-N-(methacryloyloxyethyl)-ammonium propanesulfonate through peroxide groups tethered on porous silica. Analytical Chemistry. 2002;74:4682-7.
[36] Jiang W, et al. Zwitterionic stationary phase with covalently bonded phosphorylcholine type polymer grafts and its applicability to separation of peptides in the hydrophilic interaction liquid chromatography mode. Journal of chromatography A. 2006;1127:82-91.
[37] van den Toorn HWP, et al. RockerBox: Analysis and Filtering of Massive Proteomics Search Results. Journal of Proteome Research. 2011;10:1420-4.
[38] Cook HA, et al. A mechanism of separation in electrostatic ion chromatography. Analytical Chemistry. 2001;73:3022-7.
[39] Young NL, et al. High Throughput Characterization of Combinatorial Histone Codes. Molecular & Cellular Proteomics. 2009;8:2266-84.
[40] Ding W, et al. Identification and quantification of glycoproteins using ion-pairing normal-phase liquid chromatography and mass spectrometry. Molecular & Cellular Proteomics. 2009;8:2170-85.
[41] Mysling S, et al. Utilizing Ion-Pairing Hydrophilic Interaction Chromatography Solid Phase Extraction for Efficient Glycopeptide Enrichment in Glycoproteomics. Analytical Chemistry. 2010;82:5598-609.

Chapter 4

Highly sensitive proteome analysis of FACS-sorted adult colon stem cells

Serena Di Palma, Daniel Stange, Marc van de Wetering, Hans Clevers, Albert J. R. Heck, and Shabaz Mohammed

J Proteome Res. 2011 Aug 5;10(8):3814-9.

Abstract

In proteomics, multidimensional liquid chromatography combined with mass spectrometry has become a standard technique to reduce sample complexity and tackle the vast dynamic range. Such fractionation is necessary to obtain a comprehensive analysis of biological samples such as tissues and cell lines. However, extensive fractionation comes at the expense of sample losses, hampering the analysis of limited material. We previously described a highly sensitive multidimensional chromatographic strategy based on a combination of hydrophilic interaction liquid chromatography and reversed phase chromatography, which allows proteomic analysis with minimal sample losses. Here we apply this strategy to the analysis of a limited number of FACS-sorted colon stem cells extracted from mouse intestine, obtaining a proteome coverage comparable to current methods that generally require 100-fold more starting material. We propose that this alternative multidimensional chromatographic technology will find ample application such as in the analysis of distinct cellular populations obtained by laser microdissection.

1. INTRODUCTION

In the last few years, characterisation of proteomes has accelerated dramatically, driven by the completion of many genomes and the rapid evolution of protein sample preparation and mass spectrometry technologies.[1, 2] Improvements in reversed phase chromatography have now made it possible to identify several thousand proteins in a single analysis using a few micrograms of material.[3] Still, comprehensive proteomic experiments often require several millions of cells corresponding to milligrams of protein.

The current generation of mass spectrometers has exquisite sensitivity and mass accuracy and can successfully sequence and quantify proteins using nanograms of material.[4, 5] The need for larger amounts of material originates from the requirement to fractionate the tens of thousands of proteins present in a cell which all vary in abundance.[6] Each fractionation step involves sample loss and, when planning an experiment with limited material, it is necessary to create a balance between minimal sample preparation for sensitivity [4] and adequate fractionation for comprehensivity.[2, 7, 8]

The most common strategy for global proteome screens involves the use of two peptide level separations where the second separation is predominantly nanoliter flow reversed phase (RP) chromatography.[8] Strong cation exchange (SCX) has proven to be very popular as the initial separation step since it also has advantages for phosphopeptide enrichment.[7, 9] To improve sensitivity in the SCX-RP approach, 'online' configurations have been developed, even though they suffer from necessary compromises in terms of separation quality.[8] Although SCX is by far the most dominant player as a primary stage in multidimensional peptide separation technologies, alternative approaches are constantly being explored. For instance, our laboratory (and others) has been exploring the potential of hydrophilic interaction liquid chromatography (HILIC) as the primary stage. [10, 11] During this journey we have discovered that quite a few technical hurdles had to be overcome before it could be efficiently implemented in a highly sensitive proteomics workflow.

Adult stem cells, present in the majority of multi-cellular organisms albeit at typically very low numbers, are characterized by their ability to renew themselves and/or differentiate into a diverse range of specialized cell types. It is unsurprising that large efforts are being placed in the proteomic characterization of the stem cell function. [12-14] Most of this work has been performed on cultured (embryonic) stem cells, where the sample amount typically does not pose a serious limitation. In contrast, adult stem cells have to be extracted from the organism, putting a severe constraint on the amount of starting material. Therefore, more sensitive proteomic strategies, such as the one described here, are essential.

In the intestine, identification and characterisation of downstream Wnt target genes led to the discovery of Lgr5,[15] a gene that is uniquely expressed in the stem cells of several adult tissues such as intestine, hair follicles and stomach.[16, 17] A breakthrough in these studies has been the generation of a mouse strain in which GFP has been knocked into the Lgr5 locus.[16] A cell sorting method has been established that allows a very high enrichment (>95%) of these GFP+ cells, albeit in rather small quantities for proteomics.[18] Here we report on our latest endeavours in implementing HILIC-RP in a proteomic workflow, improving significantly the sensitivity, combining excellent multidimensional separation power with minimal sample losses, and enabling for the first time the in-depth global proteome analysis of 10,000 colon stem cells, directly after the extraction from the mouse intestine.

2. EXPERIMENTAL SECTION

Materials. Protease inhibitor cocktail and Lys-C were obtained from Roche Diagnostics (Mannheim, Germany). Iodoacetamide (IAA) and ammonium acetate were supplied by Sigma-

Aldrich (Steinheim, Germany). Ammonium bicarbonate and dithiothreitol (DTT) were purchased from Fluka (Buchs, Switzerland). HPLC-S gradient grade acetonitrile was purchased from Biosolve (Valkenwaard, The Netherlands). Acetic acid was obtained by Merck KGaA (Darmstadt, Germany) and high purity water obtained from a Milli-Q purification system (Millipore, Bedford, MA). Formic acid was obtained from Merck (Darmstadt, Germany). Oasis and Sep-Pak Vac tC18 1 cm3 SPE cartridges were obtained from Waters Corporation (Milford, MA).

HeLa cell digestion. HeLa cells were grown in free suspension until confluence was reached. After harvesting the cells, lysis was carried out by resuspending them in lysis buffer containing 50 mM ammonium bicarbonate, 8M urea, EDTA free protease inhibitor cocktail, 1mM potassium fluoride, 1mM sodium orthovanadate and 5mM potassium phosphate. The suspension was vortexed and incubated for 20 min on ice. After spinning down unbroken cells and debris at 1000g for 10 min at 4 °C the protein concentration was determined by the 2DQuant Kit (GE Healthcare, Diegem, Belgium). A total of 5.6 µL of 45 mM DTT was added to 40 µL of 1 µg/µL Hela cell lysate and incubated at 50 °C for 30 min. After cooling down, 12.5 µL of 100 mM iodoacetamide was added to the mixture and kept at room temperature in the dark for 30 min. Subsequently, 5 µL of a 0.1 µg/µL Lys-C solution were added to the mixture and kept at 37 °C for 4 h. The mixture was diluted 4 times with a 50 mM ammonium bicarbonate solution (2M urea final concentration). A total of 1.2 µg of trypsin was added for overnight digestion at 37 °C. The sample was desalted using Sep-pak C18 and the eluent was dried in vacuo, reconstituted in 10% formic acid and stored at -20 °C until required.

GFP+ Cell Sorting. Freshly isolated colons from Lgr5-EGFP-IRES-CreERT2 mice were incised along their length. The intestinal tissue was washed for 5 minutes in PBS/EDTA (5mM) and subsequently incubated in fresh PBS/EDTA for 30 minutes at 4ºC. Vigorous shaking yielded free crypts which were incubated in PBS supplemented with 10 mg/ml trypsin and DNAse (0.8 µg/µl) for 30 minutes at 37ºC. After incubation, cells were spun down, resuspended in SMEM (Invitrogen, Carlsbad, CA) and filtered through a 40µm mesh. Single GFP-expressing cells were isolated using a MoFlo cell sorter (DAKO, Heverlee, BE). Propidium Iodide was used to exclude dead cells. After sorting, the cells were collected in SMEM, spun down and snap frozen.

GFP+ cells lysate and digestion. The 30,000 GFP+ colon stem cells were lysed in 8M Urea/50 mM ammonium bicarbonate, EDTA-free and phosphatase inhibitor cocktail. The sample was reduced with 2µl of 200 mM DTT for 25 min at 56 °C and alkylated with 4 µl of 200 mM IAA for 30 min in the dark at RT. Subsequently, Lys-C (1:75) was added to the mixture and kept at 37 °C for 4 h. The sample was diluted to 2M Urea/50 mM ammonium bicarbonate and trypsin was added. Digestion was performed overnight at 37 °C. The sample was desalted using OASIS resin C18 before subsequent analysis.

GFP+ cells first dimensional (HILIC) separation and fractionation. First dimensional ZIC-cHILIC separation was performed on a Famos/Ultimate LC system, using a vented column set-up. The trapping column was ZIC-cHILIC (Merck Sequant, Umea, SE), 100 µm × 20 mm, 5 µm, 200 Å; the analytical column was ZIC-cHILIC, 75 µm × 270 mm, 5 µm, 200 Å. All columns were packed in-house. The sample (corresponding to 10,000 GFP+ cells) was injected on the trap column just after the desalting. Trapping was performed at 15 µL/min for 10 min at 100% buffer A (95% acetonitrile, 0.5% acetic acid and 5mM ammonium acetate); elution was achieved with a gradient of 0-55 % buffer B (5mM ammonium acetate) in 40 min at a flow rate of 0.35 mL/min passively split to 300 nL/min. One-min fractions were collected during the elution in a 96 well plate, with each well containing 40 µL of 10% formic acid per fraction. No additional sample modification was performed before the fractions' analysis.

GFP+ cells second dimensional separation. A volume of 20 µL of collected fractions (corresponding to 5,000 cells) was used for subsequent nanoLC-LTQ-Orbitrap-MS. An Agilent 1100 series LC system was equipped with a Reprosil, 100 µm × 20 mm, 5 µm, 120 Å trapping column and a Reprosil , 50 µm × 400 mm, 3 µm, 120 Å analytical column. Trapping was performed at 5 µL/min for 10 min with solvent A (0.1M acetic acid); elution was achieved with a gradient from 13 to 28 % in 110 min and up to 50 % in 30 min of solvent B (80/20 acetonitrile/water (v/v) containing 0.1M acetic acid), with a flow rate of 0.35 mL/min passively split to 100 nL/min. The column effluent was directly introduced into the ESI source of the MS (Orbitrap Velos). Nanospray was achieved using a distally coated fused silica emitter biased to 1.7 kV. The mass spectrometer was operated in data dependent mode to automatically switch between MS and MS/MS. Full-scan MS spectra (300–1500 m/z) were acquired with a resolution of 60,000 at 400 m/z and accumulation to a target value of 500,000. After the survey scan, the 20 most intense precursors were selected for collision-induced dissociation in the linear ion trap at normalized collision energy of 35%, a q value of 0.25, an activation time of 10 ms and a target value of 5000. The signal threshold for triggering an MS/MS event was set to 500 counts. For internal mass calibration the 445.120025 and 519.138600 ions were used as lock mass with a target lock mass abundance of 0%. Charge state screening was enabled, and precursors with unknown charge state or a charge state of 1 were excluded. Dynamic exclusion was enabled (exclusion size list 180, exclusion duration 90 s).

1D RP-LC analysis. The samples (corresponding to 5000 GFP+ colon stem cells and 100 ng of HeLa cells digests) were diluted in 10 % formic acid and injected on the trap column for subsequent nanoLC-LTQ-Orbitrap-MS. Details for the LC instrument, solvents and MS setting reported in the 'GFP+ cells second dimensional separation' section. Trapping was performed at 5 µL/min for 10 min with solvent A; elution was achieved with a gradient from 10 to 23 % in 230 min and up to 50 % in 40 min of solvent B, with a flow rate of 0.35 mL/min passively split to 100 nL/min. The column effluent was directly introduced into the ESI source of the MS (Orbitrap Velos). Mass spectrometric analysis was performed as described above.

Protein identification. After MS measurement, data was analysed with the Mascot software (version 2.2.04). LTQ Orbitrap spectra from GFP+ and HeLa samples were searched against SwissProt 56.2 Fasta. The database search parameters were set to consider a peptide tolerance of 50 ppm and MS/MS tolerance was 0.9 Da. Further settings: trypsin with 2 missed cleavages; carbamidomethyl (C) as fixed modifications; oxidation (M) and phosphorylation (S, T) as variable modifications. Subsequently, the results were filtered using an in-house program that incorporates Percolator.[19] However, the minimum peptide score was set to 20 (expect value ≤0.01) which corresponded to an FDR < 1%.

3. RESULTS AND DISCUSSION

A population of 30,000 colon stem cells was FACS sorted and subsequently digested using Lys-C and Trypsin. We started with a total of 30,000 cells to have sufficient material for a systematic evaluation using one-dimensional (1D) and two-dimensional (2D) separations. A portion of the sample corresponding to 5,000 cells was subjected to a standard 1D RP nanoLC-MS analysis using a rather long 5 hour gradient elution. The inspection of the resulting LC-MS data revealed the presence of a number of abundant non-peptide (singly charged) contaminants, most likely introduced into the sample during either isolation of crypts or in the FACS procedure for the isolation of the GFP marked cells. When we performed a standard protein (bovine serum albumin) digestion in the presence of the FACS solution, no contaminant peaks were observed, thus eliminating this solution as the source of the contamination. Additionally, a number of very dominant highly charged peptide signals dominated the chromatogram (Sup-

plementary Figure 1a). A standard database search analysis identified only 380 proteins and 759 unique peptides (Table 1 and Dataset S1). The total number of protein identifications was lower than anticipated, most likely caused by suppression of the peptide signals by the non-peptide contaminants. To test this hypothesis, 100 nanograms of a HeLa lysate digest (which was estimated to be approximately equivalent to the amount of protein extracted from 5000 GFP+ cells) was analyzed in an identical fashion. Even though the signal observed in the 1D RP LC-MS analysis was somewhat higher for the GFP+ stem cell sample, the HeLa sample showed fewer dominant peaks (Supplementary Figure 1b). A standard database search analysis of the HeLa data led to the identification of 1085 proteins and 3643 unique peptides (Table 1 and Dataset S2), which is in line with expectations for such an amount of material analyzed using this LC-MS configuration and comparable to recent results reported by, for instance, Mann and co-workers.[4, 20] This higher number of identifications in the HeLa sample confirms our hypothesis that the analysis of GFP+ stem cells is hampered by the presence of non-peptide contaminants and a high dynamic range caused by certain dominant proteins.

Building on the knowledge gained from our previous work with HILIC,[10, 11, 21] we applied a refined highly sensitive ZIC-cHILIC-based system, combining excellent multidimensional separation power with minimal sample losses, and optimizing sample preparation in order to analyze minute amounts of FACS-sorted stem cells. The new column material is a resin possessing a cholinate functional group (Figure 1) which we found to be more pH independent whereby the separation power is retained when the pH conditions became more acidic.[11] We evaluated a range of gradients (Supplementary Figure 2) and found that an analysis time of 60 minutes (containing a gradient of 40 minutes) provided the optimal balance between peptide resolving power and number of fractions. Peptide peak widths at the base were less than a minute allowing the generation of 20-30 one minute peptide fractions in which the majority of peptides were distributed.

Then, we applied our ZIC-cHILIC 2D strategy to the analysis of 10,000 colon stem cells. These cells were subjected to ZIC-cHILIC fractionation, leading to 27 fractions. Half of each fraction was, subsequently, separated and analyzed by the same nano RP-LC system used earlier but applying shorter 180-minute analyses. The workflow is summarized in Figure 2. Effectively,

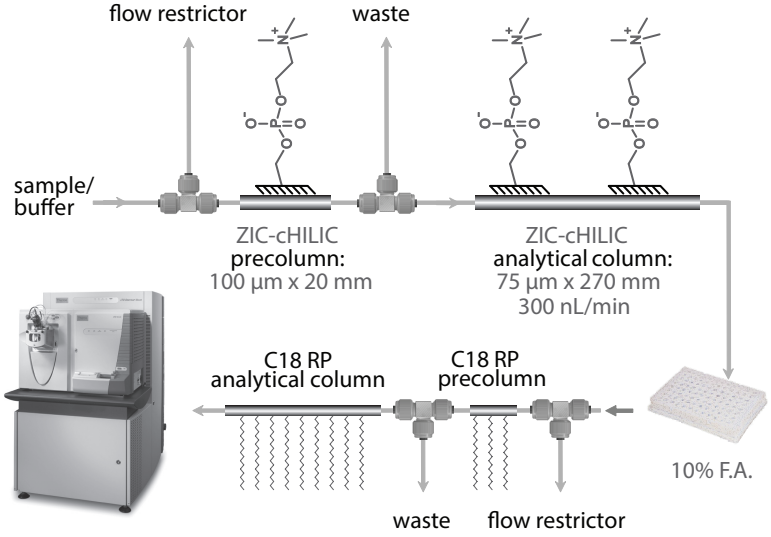

Figure 1: *Schematic of the ZIC-cHILIC-RP design.*

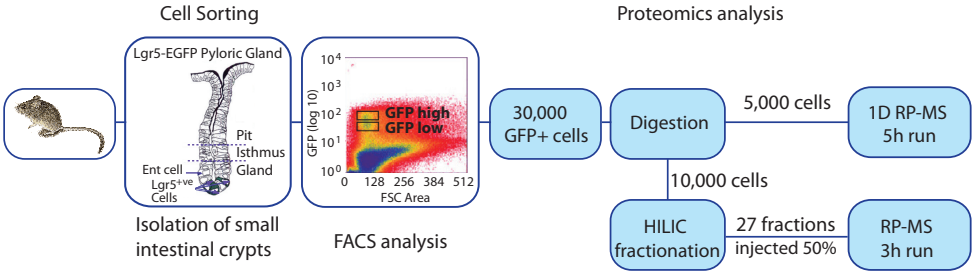

Figure 2: *Schematic representation of the employed workflow for the isolation of adult colon stem cells from mouse intestine and the subsequent proteomic analysis performed by the HILIC multidimensional strategy.*

considering the overall 2D-LC-MS/MS workflow, we analyzed 5,000 stem cells, i.e. the same amount used for the single-stage RP analysis, having a second batch available as control. Combining the data of all ZIC-cHILIC fractions allowed the cumulative identification of 15,775 unique peptides, originating from 3,775 proteins (Table 1, Dataset S3 and Figure 3a). The multidimensional ZIC-cHILIC-RP strategy led to a 20-fold and 10-fold improvement in peptide and protein identifications, respectively, when compared to the 1D RP analysis. The proteome coverage we achieved rivals that reached by current "state-of-the-art" proteome analyses that often require several millions of cells.[2, 8, 20] A detailed comparison between the two strategies revealed that only 5 (out of 380) proteins present in the 1D analysis were not observed in the multidimensional ZIC-cHILIC-RP experiment, illustrating that very little sample loss occurred incorporating the additional separation step (Figure 3c).

A GO analysis for molecular function and biological processes of the 3,775 proteins detected by our multidimensional ZIC-cHILIC-RP strategy revealed that we cover all functional categories, including, for instance, 151 proteins involved in transcription, 72 receptor related proteins and 41 proteins in translation regulation activity (Supplementary Figure 3 and 4). Comparing the number of peptide identifications, an increase from 759 to 15,775 unique peptides was achieved (Table 1, Figure 3b). Only 101 peptides observed in the 1D analysis were not identified in the 2D experiment, which potentially may be attributed to the incompatibility of certain peptides with the HILIC solvents. This low number corresponds to approximately 0.6% of the total of 15,775 unique, confidently (i.e. FDR < 1%) identified peptides in our ZIC-cHILIC-RP experiment.

Evidently, multidimensional separation strategies introduce an undesired extension of the analysis time, in our case from 5 hours to 80 hours. However, resolving the identifications to a fraction by fraction basis, it is evident that the majority of peptides are observed in a core of 20 fractions (Figure 3a). Optimizing the choice of fractions and customizing further the used LC gradients may reduce the analysis time

Method	Cell type	Amount	Peptides	Proteins
1D RP-LC	HeLa	100 ng	3643	1085
1D RP-LC	Stem cells	5000 cells	759	380
2D HILIC-RP-LC	Stem cells	5000 cells	15775	3775

Table 1: *Direct comparison between the two single-stage RP analyses and the ZIC-cHILIC-RP strategy in terms of total unique peptides and proteins identified from HeLa and GFP+ stem cell digests at the reported amount of starting material. The ZIC-cHILIC-RP fractionation method results in the identification of 20 times more unique peptides assigned to almost 10 times more proteins.*

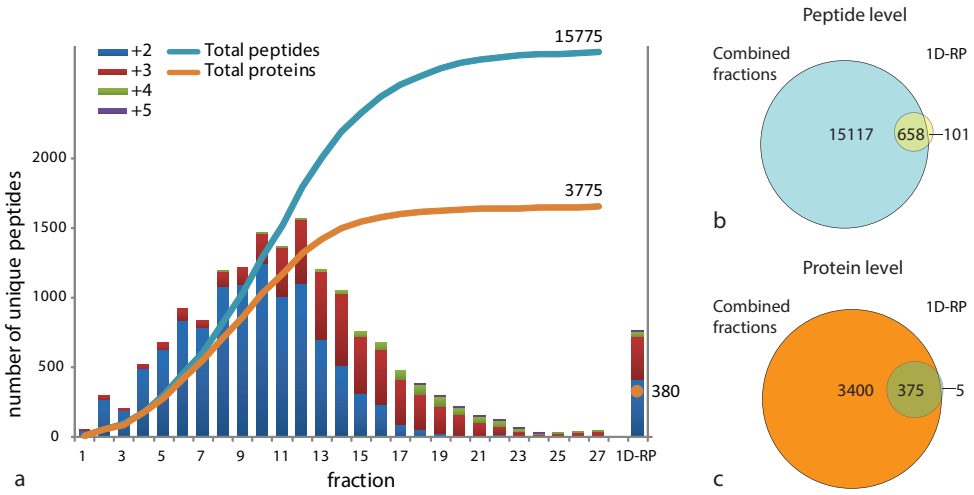

Figure 3: *Analysis of FACS-sorted colon stem cells with a 2D ZIC-cHILIC approach and a single-stage RP strategy. a: Cumulative number of identified unique peptides and proteins and distribution of unique peptides across the ZIC-cHILIC fractions and the single-stage RP run, obtained by analyzing 5,000 GFP+ colon stem cells from an initial 30,000 cells population. Within each column the distribution of peptide charge (+2, +3, +4 and +5) is indicated. The cumulated total number of peptides and proteins is indicated with a blue and orange line, respectively. Although 2+, 3+ and 4+ peptides elute over quite a number of fractions, a trend of an increase in peptide charge in the later fractions can be observed. b: Venn diagrams showing the overlap of identified unique peptides between the ZIC-cHILIC-RP approach (reported as combined fractions) and the single-stage RP analysis (reported as single run). c: Venn diagrams showing the overlap of identified proteins between the HILIC-RP approach (reported as combined fractions) and the single-stage RP analysis (reported as single run). A large overlap is observed suggesting the peptides identified in the single-stage LC run are merely higher abundant peptides that are all also detected in the ZIC-cHILIC-RP strategy.*

for our ZIC-cHILIC-RP strategy to 40 hours, without any significant loss of data.
To further confirm the quality and reliability of our ZIC-cHILIC-based strategy, we compared the obtained proteome data to previous microarray data from the same colon GFP+ cells and observed that 95% of the proteins detected in our study were also found to be expressed at the mRNA level. Furthermore, 21% of the GFP+ colon stem cell specific genes could be detected in our proteomics screen and two of these proteins, CD44 and EphB3, have been reported to be expressed highest at the bottom of the crypt, where the colon stem cells reside.[22, 23]

4. CONCLUSIONS

Notwithstanding the huge advances made in proteomics in the last decade, significant improvements are still needed to access proteome-wide data for limited number of cells, for instance originating from laser microdissections, adult stem cells and other scenarios where sample material will be a limiting factor. We demonstrate that the multidimensional ZIC-cHILIC based strategy developed in our laboratory,[11] when compared to current methods, leads to a significant reduction of sample complexity with nearly negligible sample losses. In combination with isotope labeled or label-free based quantification, our method can now probe with high-sensitivity specific cell proteomes, requiring just thousands of cells.

5. ACKNOWLEDGEMENTS
We kindly acknowledge Paul Boersema for fruitful discussion and critical reading of the manuscript. This work was in part supported by the PRIME-XS project, grant agreement number 262067, funded by the European Union 7th Framework Programme; The Netherlands Proteomics Centre, embedded in the Netherlands Genomics Initiative, and the Centre for Biomedical Genetics.

6. SUPPORTING INFORMATION AVAILABLE
The data associated with this manuscript may be downloaded from the ProteomeCommons.org Tranche.
Supplementary Figure 1: Single-stage RP LC-MS base peak intensity chromatograms (LTQ-Orbitrap Velos) from (a) sample corresponding to the 30,000 GFP+ colon cell digest and loading equivalent to 5000 cells and (b) 100 ng HeLa digests.
Supplementary Figure 2: Gradient optimization of ZIC-cHILIC at pH 6.8. The sample for the analyses was a peptide mixture consisting of BSA, alpha- and beta-casein digests. The peak width at half maximum (FWHM) is indicated alongside total analysis time. Gradients utilized for each analysis are specified in Supplementary Table 1.
Supplementary Figure 3: GO biological process and molecular function analysis of proteins (n=3775) identified in the HILIC-RP LC-MS analysis of 5000 GFP+ colon stem cells. Mostly proteins involved in metabolic processes and with a catalytic and binding activity were identified.
Supplementary Figure 4: A protein classification analysis of what was identified in the HILIC-RP LC-MS analysis of 5000 GFP+ colon stem cells.
Supplementary Table 1: Gradients utilized for each analysis as reported in Supplementary Figure 2.
Dataset S1: Excel Table containing peptides and protein results for the 1D RP-LC analysis of GFP+ cells.
Dataset S2: Excel Table containing all peptides and protein results for the 1D RP-LC analysis of HeLa cells.
Dataset S3: Excel Table containing all peptides and protein results for the 2D HILIC-RP-LC analysis of GFP+ cells.
This material is available free of charge at http://pubs.acs.org.

7. REFERENCES

[1] de Godoy LM, et al. Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. Genome Biol. 2006;7:R50.

[2] de Godoy LMF, et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature. 2008;455:1251-U60.

[3] Köcher T, et al. Ultra-High-Pressure RPLC Hyphenated to an LTQ-Orbitrap Velos Reveals a Linear Relation between Peak Capacity and Number of Identified Peptides. Analytical Chemistry. 2011;83:2699-704.

[4] Waanders LF, et al. Quantitative proteomic analysis of single pancreatic islets. Proc Natl Acad Sci U S A. 2009;106:18902-7.

[5] Valaskovic GA, et al. Attomole protein characterization by capillary electrophoresis mass spectrometry. Science. 1996;273:1199-202.

[6] Anderson NL, et al. A Human Proteome Detection and Quantitation Project. Molecular & Cellular Proteomics. 2009;8:883-6.

[7] Gauci S, et al. Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. Analytical Chemistry. 2009;81:4493-501.

[8] Washburn MP, et al. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001;19:242-7.

[9] Beausoleil SA, et al. Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci U S A. 2004;101:12130-5.

[10] Boersema PJ, et al. Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. Journal of Proteome Research. 2007;6:937-46.

[11] Di Palma S, et al. Zwitterionic Hydrophilic Interaction Liquid Chromatography (ZIC-HILIC and ZIC-cHILIC) Provide High Resolution Separation and Increase Sensitivity in Proteome Analysis. Analytical Chemistry. 2011;83:3440-7.

[12] Van Hoof D, et al. Phosphorylation dynamics during early differentiation of human embryonic stem cells. Cell Stem Cell. 2009;5:214-26.

[13] Swaney DL, et al. Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. Proc Natl Acad Sci U S A. 2009;106:995-1000.

[14] Rigbolt KT, et al. System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. Sci Signal. 2011;4:rs3.

[15] van de Wetering M, et al. The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell. 2002;111:241-50.

[16] Barker N, et al. Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature. 2007;449:1003-U1.

[17] Barker N, et al. Lgr5(+ve) Stem Cells Drive Self-Renewal in the Stomach and Build Long-Lived Gastric Units In Vitro. Cell Stem Cell. 2010;6:25-36.

[18] van der Flier LG, et al. Transcription Factor Achaete Scute-Like 2 Controls Intestinal Stem Cell Fate. Cell. 2009;136:903-12.

[19] van den Toorn HWP, et al. RockerBox: Analysis and Filtering of Massive Proteomics Search Results. Journal of Proteome Research. 2011;10:1420-4.

[20] Olsen JV, et al. A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. Molecular & Cellular Proteomics. 2009;8:2759-69.

[21] Boersema PJ, et al. Hydrophilic interaction liquid chromatography (HILIC) in proteomics. Anal Bioanal Chem. 2008;391:151-9.

[22] Wielenga VJM, et al. Expression of CD44 in Apc and Tcf mutant mice implies regulation by the WNT pathway. Am J Pathol. 1999;154:515-23.

[23] Batlle E, et al. beta-catenin and TCF mediate cell positioning in the intestinal epithelium

by controlling the expression of EphB/EphrinB. Cell. 2002;111:251-63.

Chapter 5

Evaluation of the deuterium isotope effect in ZIC-cHILIC separations for implementation in a quantitative proteomic approach

Serena Di Palma, Reinout Raijmakers, Albert J. R. Heck, and Shabaz Mohammed

Anal Chem. 2011 Nov 1;83(21):8352-6.

Abstract

Quantitative methodologies for the global in-depth comparison of proteomes are frequently based on chemical derivatization of peptides with isotopically distinguishable labeling agents. In the present work, we set out to study the feasibility of the dimethyl labeling method in combination with ZIC-cHILIC (zwitterionic hydrophilic interaction liquid chromatography) technology for quantitative proteomics. We first addressed the potential issue of isotope effects perturbing the essential co-elution of differently labeled peptides under ZIC-cHILIC separation. The deuterium incorporation-induced effect can be largely eliminated by favoring the mixed-mode ZIC-cHILIC separation based on combined hydrophilic and ionic interactions. Then, we evaluated the performance and applicability of this strategy using a sample consisting of human cell lysate. We demonstrate that our approach is suitable to perform unbiased quantitative proteome analysis, still quantifying more than 2,500 proteins when analyzing only a few micrograms of sample.

1. INTRODUCTION

The recent advances in mass spectrometry-based proteomics have allowed the systematic interrogation of complex proteomes and the identification of differentially expressed proteins in cells, tissues and body fluids, opening new horizons for a broad range of applications[1]. Multidimensional liquid chromatography (LC) in combination with mass spectrometry still remains the main workflow for proteomic strategies[2]. Hydrophilic interaction liquid chromatography (HILIC) has found widespread use in the analytical field since the pioneering work of Alpert[3]. HILIC has a high degree of orthogonality to reversed phase (RP) chromatography[4-6], which typically is the last separation step prior to MS. Hence, the combination of HILIC and RP provides a powerful alternative in multidimensional separation strategies[5, 7, 8].

To explore the dynamics of a whole proteome requires a systematic assessment of quantitative differences in protein abundance. One of the most powerful enabling strategies involves the use of stable isotope labeled proteins or peptides introduced by either chemical or metabolic labeling[9, 10]. It is well known that the introduction of stable isotopes into peptides may affect their retention time under RP-LC conditions[11]. While little or no resolution of peptide isoforms has been observed during RP when 13C and 15N were used as coding agents[12, 13], deuterated versions (of coding agents) can show distinctive chromatographic behavior, the so-called deuterium effect[14].

Recently, we refined and optimized a multiplex stable isotope labeling approach[15-18], based on reductive amination ('dimethyl labeling')[19]. The use of deuterated isotopomers of formaldehyde and sodium cyanoborohydride to incorporate dimethyl labels at the N-terminus or the ε-amino group of lysine residues induces a mass difference of at least 4 Da and allows the comparison of three samples in parallel[16]. Although dimethyl labeling is inexpensive and efficiently incorporated into biomolecules, deuterated species are slightly more hydrophilic than their non-deuterated counterpart and, when using a HILIC-based separation, the different isoforms often fail to co-elute[15]. Consequently, the isotope ratio will vary across the elution of isotopologous peptides[14] and a biased quantification might arise.

The aim of the present study is to evaluate and eliminate, if necessary, the deuterium isotope effect in ZIC-cHILIC separations, which would allow us to introduce a highly sensitive and powerful quantitative proteomic workflow based on dimethyl labeling[16] and two-dimensional ZIC-cHILIC-RP-MS/MS[7].

2. EXPERIMENTAL SECTION

Chemicals and Materials. Bovine serum albumin, iodoacetamide, formaldehyde (37%) and ammonium acetate were supplied by Sigma-Aldrich (Steinheim, Germany). Ammonium bicarbonate, sodium cyanoborohydride and dithiothreitol (DTT) were purchased from Fluka (Buchs, Switzerland). HPLC-S gradient grade acetonitrile was purchased from Biosolve (Valkenswaard, The Netherlands). Acetic acid was obtained by Merck KGaA (Darmstadt, Germany) and high purity water obtained from a Milli-Q purification system (Millipore, Bedford, MA). Formic acid was obtained from Merck (Darmstadt, Germany). Sep-Pak Vac tC18 1 cm3 cartridges were obtained from Waters Corporation (Milford, MA).

Sample preparation. For the evaluation of the deuterium effect under on-line ZIC-cHILIC separation, we prepared a standard peptide mixture from bovine serum albumin (BSA), whereas for the ZIC-cHILIC fractionation we employed a human cell lysate (HeLa). Each sample was digested, as described previously[7], and subsequently split in 3 equal aliquots for triple dimethyl labeling with the light, medium and heavy isotopes.

Dimethyl Labeling Reagents. Each aliquot was desalted and isotope labeled directly onto Sep-pak C18 according to the original protocol[16] with slight modifications. Per sample/label, 5

ml of labeling solution was prepared: 4.5 ml of 50 mM sodium phosphate buffer pH 7.5 were mixed with 250 µl of 4% (vol/vol) formaldehyde in water (CH_2O, CD_2O or $^{13}CD_2O$) and 250 µl of 0.6 M cyanoborohydride in water ($NaBH_3CN$ or $NaBD_3CN$). For optimal labeling efficiency, each of the SepPak columns (50 mg, 1cc, C18) was flushed 5 times with 1 ml of the respective labeling reagent (light, intermediate or heavy), for at least 10 minutes. After labeling, each channel was washed with RP solvent A to remove the excess labeling solution. The labeled sample was subsequently eluted using the appropriate HILIC loading buffer (see below).

HILIC-LC and RP-LC Buffers. For HILIC at pH 6.8, buffer A was 95% acetonitrile, 0.5% acetic acid and 5 mM ammonium acetate; buffer B was 5 mM ammonium acetate. For pH 3.5, buffer A was 95% acetonitrile, 2% formic acid and 5 mM ammonium acetate; buffer B 0.07% formic acid and 5 mM ammonium acetate. For RP, buffer A was 0.1 M acetic acid; buffer B was 80% acetonitrile and 0.1 M acetic acid.

On-line (ZIC-cHILIC-MS) and off-line (ZIC-cHILIC-RP-MS) systems. ZIC-cHILIC–MS was performed on an "inert" Dionex "Ultimate" LC system using a vented column setup. The trapping column was a ZIC-cHILIC 100 µm × 20 mm, 5 µm, 200 Å (Merck SeQuant, Umeå, Sweden); the analytical column was a ZIC-cHILIC 75 µm × 250 mm, 5 µm, 200 Å. All columns were packed in-house. Trapping was performed at 10 µL/min for 10 min at 100% buffer A; elution was achieved with a gradient of 0-55 % buffer B in 40 min at a flow rate of 0.35 mL/min passively split to 300 nL/min. Column output was coupled with an LTQ-Orbitrap mass spectrometer (Orbitrap Discovery, Thermo). In the off-line approach, the first dimensional analysis was performed as described above. During ZIC-cHILIC elution, one-min fractions were collected in a 96-well plate with each well containing 40 µL of 10% formic acid per fraction. A volume of 20 µL of collected fractions was used for subsequent nanoLC-LTQ-Orbitrap-MS. An Agilent 1100 series LC system was equipped with a Reprosil, 100 µm × 20 mm, 5 µm, 120 Å trapping column and a Reprosil, 50 µm × 400 mm, 3 µm, 120 Å (Dr. Maisch GmbH, Ammerbuch, Germany) analytical column. Trapping was performed at 5 µL/min for 10 min with solvent A (0.1M acetic acid); elution was achieved with a gradient from 13 to 28 % in 57 min and up to 50 % in 25 min of solvent B, with a flow rate of 0.35 mL/min passively split to 100 nL/min. The column effluent was directly introduced into the ESI source of the MS. Further details for MS analysis can be found in our previous work[7].

Data Analysis. For protein identification and quantification, raw MS data were converted to peak lists using Proteome Discoverer software, version 1.2 (Thermo Fisher Scientific, Bremen). Peak lists were searched using MASCOT software (version 2.3.04) against a concatenated database containing the Uniprot human database (version 56.2) and an equivalent decoy database which was created using MaxQuant [20]. Further settings included the enzyme trypsin with 2 missed cleavages, carbamidomethyl (C) as fixed modification; oxidation (M), phosphorylation (S, T), dimethyl (K) and (N-terminal), dimethyl 2H(4) (K) and (N-terminal), dimethyl 2H(6) 13C(2) (K) and (N-terminal) as variable modifications. Peptide tolerance was initially set to 50 ppm for +2, +3 charged peptides, and MS/MS tolerance was 0.6 Da. The mgf files were subsequently filtered by mass-accuracy (final mass accuracy was ±7 ppm) using a Proteome Discoverer mass filter. A minimum peptide score was set to 20 and globally the data corresponded to an FDR < 1%. Other filters used were a peptide length between 7 and 36 amino acids, peptide rank 1. Proteome Discoverer was used in a quantitative mode; extracted ion chromatograms (XIC) were integrated for all peptides to calculate the relative peptide ratios of 'light', 'medium' and 'heavy' forms. To evaluate the quantitation quality, the $\log_2$ Heavy/Light, Medium/Light and Heavy/Medium ratios were calculated, and each dataset (pH 3.5 and 6.8) was normalized using the median of all values. For the evaluation of the deuterium isotope effect upon fractionation, the datasets were filtered for unique peptides per fraction. For each peptide identified

in two consecutive fractions, the average $\log_2$ Heavy/Light, Medium/Light or Heavy/Medium ratios between the two fractions were calculated, and the absolute difference between the observed ratios and the average ratios in those fractions was used as a measure of deviation.

3. RESULTS AND DISCUSSION

3.1 Peptide separation by ZIC-cHILIC

When using multidimensional chromatography, quantitative analysis based on isotope labeling strategies requires that each isotopic peptide analogue co-elutes during every chromatographic separation or fractionation step. When partial resolution of the isotopic forms of a peptide occurs, a substantial error in the isotope ratio determination may be introduced in quantification measurements.

We started with the evaluation of a potential deuterium isotope effect on the chromatographic resolution of differentially labeled peptides using a simple one dimensional ZIC-cHILIC separation under two different conditions, i.e. at acidic and nearly neutral pH. Separation of peptides using zwitterionic materials, such as ZIC-c, in HILIC mode involves both hydrophilic and ionic interactions, even though the latter interactions are weaker when compared to normal ion exchangers[21]. Factors governing the retention include: hydrogen bonding[22], the extent of which depends on the acidity and basicity of the peptides, and dipole-dipole interactions, which depend on the dipole moments and the polarizability of the analytes[3].

One advantageous property of zwitterionic materials relies on the fact that the charge of their functional groups is permanent and unaffected by pH changes. Recently, two pH conditions, one nearly neutral and one acidic, have been extensively evaluated in our group with respect to their influence on peptide retention in ZIC-cHILIC[7]. We showed that ZIC-cHILIC selectivity at pH 6.8 was mostly driven by the peptide hydrophilicity, consistent with the hydrophilic partitioning model, whereas at pH 3.5 the retention was governed by a stronger mixed-mode mechanism based on hydrophilic and ionic repulsive ionic interactions (Figure 1).

As normal tryptic peptides are mainly positively charged at pH 3.5, this mechanism may favor the electrostatic repulsion with the more exposed quaternary ammonium on the ZIC-cHILIC surface, creating an ERLIC-like mode[23] (electrostatic repulsion-hydrophilic interaction chromatography) where analytes are on one side repelled by similar charges, on the other side retained on the aqueous layer by hydrophilic partitioning. Proposed factors governing the analytes' retention at the two pH conditions are summarized in Table 1.

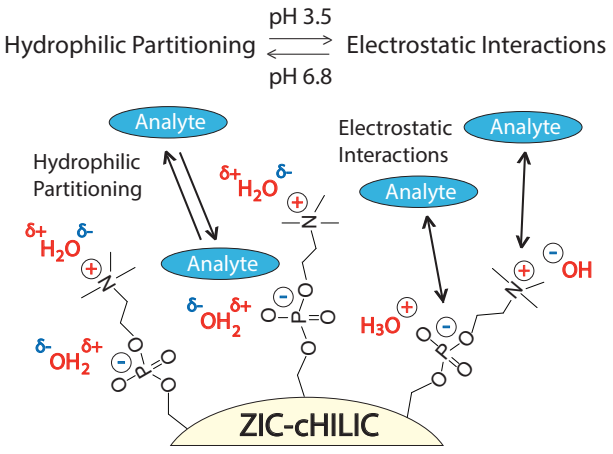

Figure 1: *Schematic representation of the mechanism of retention in HILIC mode using ZIC-cHILIC stationary phase at an acidic and neutral pH.*

ZIC-cHILIC	Hydrophilic Partitioning	Hydrogen Bondings	Electrostatic Interactions	Dipole-Dipole Interactions
pH 3.5	++	++	++	++
pH 6.8	+++	++	+	++

Table 1: *Proposed factors governing the analytes' retention on ZIC-cHILIC stationary phases at a pH of 3.5 and 6.8. At neutral pH values, the hydrophilicity is a more dominant factor when compared to the ionic interactions, while at acidic pH the separation is largely governed by a stronger mixed-mode mechanism based on more balanced hydrophilic and ionic interactions.*

3.2 Deuterium isotope effect in ZIC-cHILIC retention at acidic and neutral pH

Boersema et al. reported previously a deuterium isotope effect within zwitterionic HILIC based separations when fractionating triple dimethyl labeled samples[15]. In contrast, when Guo et al. investigated the deuterium isotope effect on a non-ionic TSK-gel Amide 80 HILIC stationary phase, no isotope effect was found[24]. It seems that these different findings are related to the specific HILIC-mode involved in the separation mechanisms when using these two different HILIC materials.

In general, deuterated peptides are slightly more hydrophilic than their non-deuterated counterparts and this increased hydrophilicity might lead to partial or even complete resolution of differentially labeled peptides under HILIC separation conditions[15]. Taking into account

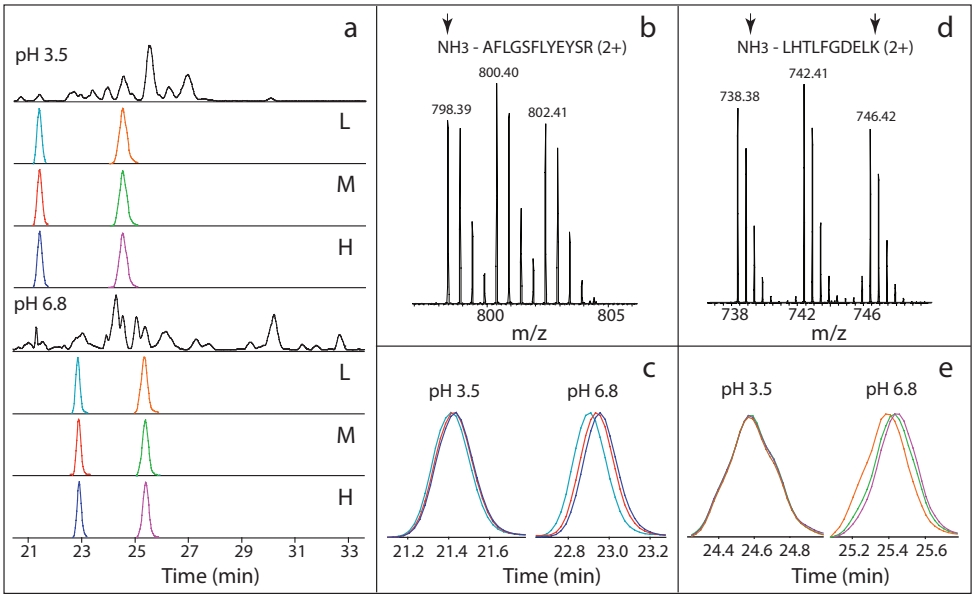

Figure 2: *Comparison of the deuterium isotope effect when the analysis of a triple dimethyl sample is performed on-line at a pH of 3.5 and 6.8 using ZIC-cHILIC separation. a) LC-MS chromatograms (black lines) of a triplex dimethyl labeled BSA (1 pmol) at the two indicated pH values and extracted ion chromatograms (XIC) of the light (L) (light blue and orange), medium (M) (red and green) and heavy (H) (blue and purple) isotopes of two peptides (AFLGSFLYEYSR and LHTLFGDELK). (b, d) MS spectra and (c, e) overlaid XICs of L, M and H isotopomers of the two peptides reported in panel a with the same colors assigned to each isotope and relative retention times at indicated buffer pHs.*

the ZIC-cHILIC separation mechanism proposed above, we hypothesized that the deuterium isotope effect would hamper the co-elution of isotopologous peptides under neutral pH ZIC-cHILIC separation, due to the hydrophilicity-driven selectivity. However, this effect may be partially corrected at acidic pH, where the mixed-mode ERLIC-like HILIC mechanism provides a more orthogonal separation, minimizing the hydrophilicity difference between deuterated and non-deuterated isotopes potentially improving the co-elution of different isotopomers.

To test this hypothesis, we used an on-line ZIC-cHILIC strategy at two pH conditions using a simple BSA tryptic digest after triplex dimethyl labeling, wherein each peptide exhibits three labels varying in the number of deuterium atoms; the 'light' form (L) has no deuterium (D) atom, the 'medium' (M) has 4 D atoms, the 'heavy' (H) has 6 D atoms and 2 13C atoms[16]. As the number of deuterium atoms increases, the resolution of structurally similar species might also increase[12, 14]. We observed that peptide isotopologues varying by at least 4 (M) or 6 (H) deuterium atoms exhibited shifts in their retention times in comparison to their non-deuterated analogues (L) at pH 6.8. In contrast, such peptide triplets co-eluted at pH 3.5 (Figure 2). Illustrative triplet peaks for two BSA peptides were chosen as examples to show the expected 1:1:1 ratio between light, medium and heavy labels (Figure 2 b, d). The overlaid extracted ion chromatograms of L, M and H isotopes at pH 6.8 and 3.5 demonstrate the extent of the deuterium effect. At pH 6.8 the deuterated peptides are somewhat resolved upon elution, while at pH 3.5 the three peaks corresponding to the three labeled species are indistinguishable (Figure 2 c, e). Although the overall resolving power is similar for ZIC-cHILIC at either pH value, it seems that the separation mechanism in operation at neutral pH resolves peptides containing several deuteriums from their non-heavy isotopic counterparts. The observed retention times for the more hydrophilic heavy and medium labels were higher than for light labels. In addition, the isotope shift was larger between the heavy and the light labels (6 D's difference) than between the medium and the light labels (4 D's difference). This difference suggests that the ZIC-cHILIC separation at pH 6.8 is mainly driven by hydrophilicity. In contrast, the co-elution of the isotopologue triplets observed at pH 3.5 might be attributed to the mixed-mode separation mechanism of both hydrophilic and electrostatic interactions that is in operation, consistent with our hypothesis.

3.3 Deuterium isotope effect in ZIC-cHILIC fractionation in a 2D-ZIC-cHILIC-RP strategy

To further validate our initial results and to demonstrate the feasibility of larger scale dimethyl labeling in combination with ZIC-cHILIC fractionation at pH 3.5, we extended our study to an off-line multidimensional ZIC-cHILIC–RP strategy. In principle, in any off-line 2D chromatography approach based on the collection of fractions across a chromatographic peak, there is the possibility that a peak is separated over two fractions. Assuming a situation of imperfect co-elution of the isotopic forms, the isotopic abundances would be different in the two consecutive fractions, leading to different calculated ratios for the same analyte and introducing a deviation from the 'true' ratio in the sample.

A lysate consisting of HeLa cells was digested using trypsin and three equal aliquots were derivatised with light, medium and heavy labels. A fraction of the combined sample, corresponding to 2 μg for each isotopic form, was fractionated in the first dimension employing ZIC-cHILIC at the two different pHs. Subsequently, half of each fraction was analysed by RP LC MS/MS. A schematic representation of the employed proteomic workflow is reported in Figure 3.

The combination of the data generated from all ZIC-cHILIC fractions allowed the cumulative identification and quantification of more than 15,000 unique peptides (FDR < 0.5%) originat-

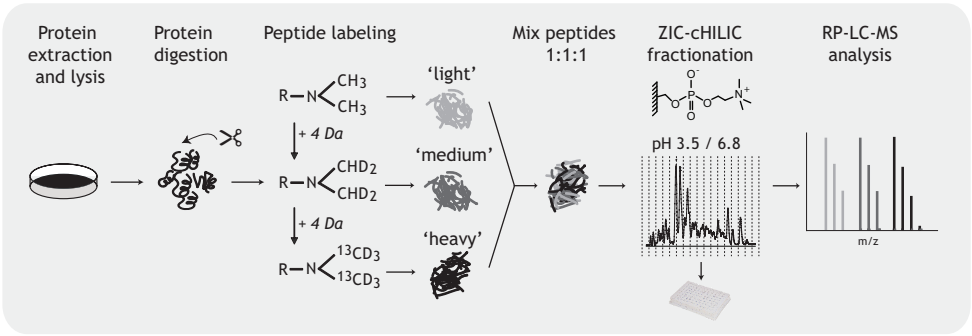

Figure 3: *Experimental scheme for the quantitative proteomic strategy. After protein lysis and enzymatic digestion, peptides were differentially stable isotope dimethyl labeled. By using three differential stable isotopes, peptide triplets can be obtained that differ in mass by a minimum of 4 Da, when the digested peptides do not contain lysines. In the latter case, the mass shift would be increased proportionally to the number of labeled lysine residues present in the sequence. Then, the 3 labeled samples (light, medium and heavy) were combined in an equimolar ratio. The 6 μg mixture (containing 2 μg per each label) was fractionated by ZIC-cHILIC at a pH of 3.5 and 6.8, followed by RP-LC-MS analysis (half the amount of 20 collected fractions, overall corresponding to 3 μg) for quantification using the triplet peaks originating from the different isotopomers.*

ing from more than 2,200 proteins, at each pH condition. Next, we filtered our dataset to keep only those peptides found in consecutive fractions and calculated their isotopic ratios (heavy/light, medium/light and heavy/medium) in both fractions. The percentage of peptides identified in consecutive fractions was rather low, corresponding to approximately 15%, demonstrating that the ZIC-cHILIC separation power was minimally affected when introducing triple dimethyl labeled isotopes[7].

Theoretically, since differentially labeled peptides were originally present at equimolar amounts, the $\log_2$ ratios of H/L, M/L and H/M in consecutive fractions should be approximately the same and close to zero. In practice, when a partial resolution of the labeled forms of a peptide occurs, the $\log_2$ ratios should vary in consecutive fractions. In order to evaluate this effect, we calculated the average of the $\log_2$ ratio (H/L, M/L and H/M) of the same peptides found in consecutive fractions and reported the absolute $\log_2$ difference as a parameter to estimate the deviation from the experimental average value (Figure 4a). In agreement with the 1:1:1 ratio of the pooled labeled samples, the measured average of the $\log_2$ ratios was in general very close to zero at both pH values. However, the observed trends of deviation showed that substantial differences in the ratios were introduced when the fractionation was performed at pH 6.8.

In addition, the reported average deviation in the case of H/L ratio (0.32) was somewhat higher in comparison to the average deviation for M/L (0.22) and H/M (0.15) (Figure 4a), confirming that the deuterium effect was proportional to the increased number of deuteriums present in the species. In contrast, when we fractionated the sample at pH 3.5, the calculated average deviations were more than a factor of 2 lower for both H/L and M/L ratios (0.13 and 0.10) and slightly lower for H/M ratio (0.11) when compared to the data obtained at pH 6.8, highlighting the limited extent of deuterium isotope effect on peptide retention time at this acidic pH. In line with these observations, we expected that peptides containing more deuterium atoms after labeling would show a more pronounced deuterium effect. Thus, we investigated the deuterium effect in two distinct classes of peptides, differing in the number of labeling sites: N-terminally labeled (no lysine containing tryptic) peptides and N-terminally plus lysine labeled (lysine tryp-

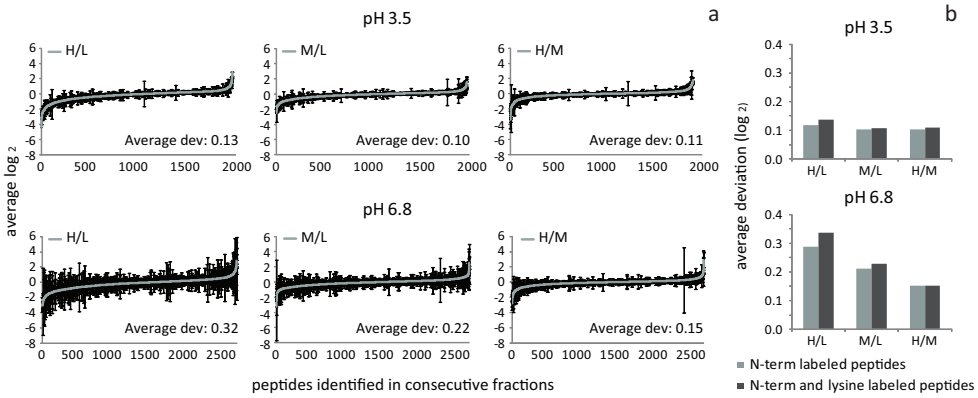

Figure 4: *a) Average log_2 ratios (y axis) of heavy divided by light (H/L), medium divided by light (M/L) and heavy divided by medium (H/M) reported only for those peptides that were identified and quantified in two consecutive fractions (x axis) after ZIC-cHILIC fractionation at pH 3.5 and 6.8 from a HeLa lysate sample. The error bars represent the deviation (reported as absolute value) from the average log_2 value. For instance, if we call i the first fraction where a peptide was found and i+1 the second fraction, the average log_2 ratio and deviation for H/L were calculated as: Average = (log_2 H/L i + log_2 H/L i+1)/2; Deviation = |log_2 (H/L) i – log_2 H/Laverage|. The average deviations with relative standard deviation were also calculated: at pH 3.5, 0.13 for H/L (STDEV 0.17), 0.10 for M/L (STDEV 0.15), 0.11 for H/M (STDEV 0.17); at pH 6.8, 0.32 for H/L (STDEV 0.45), 0.22 for M/L (STDEV 0.32), 0.15 for H/M (STDEV 0.21). b) Comparison between the average deviations (reported for each ratio) for peptides labeled solely at the N-termini (no lysine containing tryptic cleaved) versus peptides labeled at both N-terminus and lysine residues (lysine tryptic and miss cleaved peptides) at the two pH conditions. In the case of N-terminally labeled peptides, the difference in number of deuterium atoms is 6 between H and L species, 4 between M and L, 2 between H and M. For N-terminally plus lysine labeled peptides, the difference in deuterium atoms increases proportionally to the number of lysine residues. The analysis is restricted to peptides found in two consecutive fractions*

tic and miss cleaved) peptides. At pH 6.8, we could observe higher average deviations in the case of N-terminally plus lysine labeled peptides (for both H versus L and M versus L labels) as a consequence of the higher number of deuterium atoms (Figure 4b). The trend of deviation at acidic pH was in general lower than at pH 6.8 and similar for both types of labeled peptides, confirming that, even in presence of additional labeling sites, the deuterium effect is minimal. To further show the separation occurring at pH 6.8 between light and heavier species and the coelution occurring at pH 3.5, we calculated for peptides found in two consecutive fractions the difference of log_2 ratios (H/L and M/L) between the earlier and later fraction and used this value to visualize the direction of the deviation. Using this approach, negative values indicated that the light species (less hydrophilic) would be the predominant form in the earlier fraction in comparison to heavier species, while a positive value would describe the opposite phenomenon. This behavior can be observed in Figure 5: at pH 6.8 the curve of deviation was moved towards negative values (especially in the case of H/L), indicating a higher bias as a consequence of the light species being resolved from the medium and heavy forms. In contrast, at acidic pH the curve was more flat and closer to zero (baseline), underlining that, albeit peptides may elute in consecutive fractions, the differently labeled isoforms coeluted during the fractionation procedure.

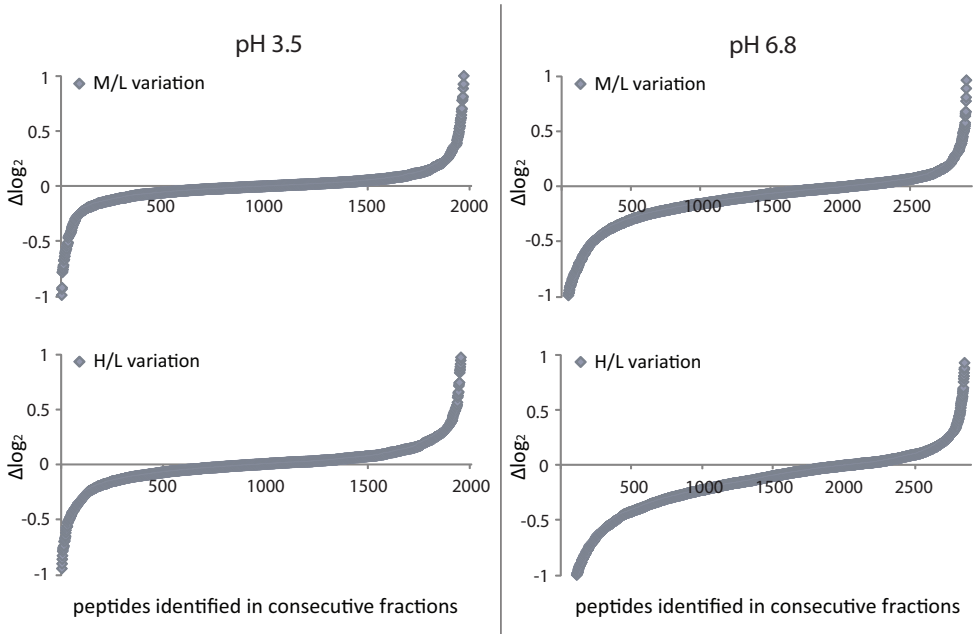

Figure 5: *Relative difference between M/L and H/L log_2 ratios for peptides identified and quantified in consecutive fractions. Shown is half of the difference between the log_2 ratio of a quantified peptide in two consecutive fractions. If the light labeled peptide (hydrogen containing, less hydrophilic) is more abundant in the first fraction, then the value will be negative. The graphs show that at pH 6.8 more separation of the light labeled peptides in the first of the two fractions occurs compared to pH 3.5; i.e. more combinations of fractions show negative values.*

Overall, using the proposed quantitative strategy based on dimethyl labeling in combination with ZIC-cHILIC-RP-MS/MS technology at pH 3.5, it is possible to minimize the deuterium effect without compromising the separation power. Starting with a few micrograms of material, we could confidently quantify 17,583 unique peptides (FDR < 0.5%), assigned to 2,630 proteins using pH 3.5, and 16,434 unique peptides assigned to almost 2,300 proteins at pH 6.8, although the former pH condition possessed better precision due to the elimination of the deuterium isotope effect.

4. CONCLUSIONS

We demonstrate that ZIC-cHILIC separations at an acidic pH of 3.5, unlike at a neutral pH of 6.8, exhibits no deuterium isotope effect. We hypothesize that the negligible effect is due to a stronger mixed-mode separation, decreasing the hydrophilicity difference of deuterated and non-deuterated species by enhancing ion-exchange interactions with the zwitterionic stationary phase. At a pH of 6.8, charge effects become smaller and hydrophilicity effect enhances, leading to the polarity differences of labels containing hydrogen and deuterium resolvable. Furthermore, we show that our two-dimensional approach, at acidic conditions, in combination with dimethyl labeling represents a highly sensitive proteomic workflow that allows comprehensive quantitative analysis of very modest amount of starting material.

5. ACKNOWLEDGEMENTS

We thank Dr. Paul Boersema and Dr. Natalie Thompson for critically reading the manuscript. We kindly acknowledge Dr. Wen Jiang, from Merck SeQuant, for providing the ZIC-cHILIC materials. This work was in parts supported by the PRIME-XS project, grant agreement number 262067, funded by the European Union 7th Framework Programme; The Netherlands Proteomics Centre, embedded in the Netherlands Genomics Initiative, and the Centre for Biomedical Genetics; the Netherlands Organization for Scientific Research (NWO) with the VIDI grant (700.10.429).

6. REFERENCES

[1] Cox J, et al. Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. Annual Review of Biochemistry. 2011;80:273-99.

[2] Washburn MP, et al. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001;19:242-7.

[3] Alpert AJ. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. J Chromatogr. 1990;499:177-96.

[4] Gilar M, et al. Orthogonality of separation in two-dimensional liquid chromatography. Anal Chem. 2005;77:6426-34.

[5] Boersema PJ, et al. Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. Journal of Proteome Research. 2007;6:937-46.

[6] McNulty DE, et al. Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. Molecular & Cellular Proteomics. 2008;7:971-80.

[7] Di Palma S, et al. Zwitterionic Hydrophilic Interaction Liquid Chromatography (ZIC-HILIC and ZIC-cHILIC) Provide High Resolution Separation and Increase Sensitivity in Proteome Analysis. Analytical Chemistry. 2011;83:3440-7.

[8] Di Palma S, et al. Highly Sensitive Proteome Analysis of FACS-Sorted Adult Colon Stem Cells. Journal of Proteome Research. 2011;10:3814-9.

[9] Heck AJ, et al. Mass spectrometry-based quantitative proteomics. Expert Rev Proteomics. 2004;1:317-26.

[10] Julka S, et al. Quantification in proteomics through stable isotope coding: a review. J Proteome Res. 2004;3:350-63.

[11] Hsu JL, et al. Dimethyl multiplexed labeling combined with microcolumn separation and MS analysis for time course study in proteomics. Electrophoresis. 2006;27:3652-60.

[12] Zhang R, et al. Controlling deuterium isotope effects in comparative proteomics. Anal Chem. 2002;74:3662-9.

[13] Gouw JW, et al. Optimizing identification and quantitation of 15N-labeled proteins in comparative proteomics. Anal Chem. 2008;80:7796-803.

[14] Zhang R, et al. Fractionation of isotopically labeled peptides in quantitative proteomics. Anal Chem. 2001;73:5142-9.

[15] Boersema PJ, et al. Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. Proteomics. 2008;8:4624-32.

[16] Boersema PJ, et al. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protoc. 2009;4:484-94.

[17] Raijmakers R, et al. Assessing biological variation and protein processing in primary human leukocytes by automated multiplex stable isotope labeling coupled to 2 dimensional peptide separation. Mol Biosyst. 2009;5:992-1003.

[18] Raijmakers R, et al. Automated online sequential isotope labeling for protein quantitation applied to proteasome tissue-specific diversity. Molecular & Cellular Proteomics. 2008;7:1755-62.

[19] Hsu JL, et al. Stable-isotope dimethyl labeling for quantitative proteomics. Anal Chem. 2003;75:6843-52.

[20] Cox J, et al. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26:1367-72.

[21] Hemstrom P, et al. Hydrophilic interaction chromatography. J Sep Sci. 2006;29:1784-821.

[22] Yoshida T. Peptide separation by Hydrophilic-Interaction Chromatography: a review. J Biochem Biophys Methods. 2004;60:265-80.

[23] Alpert AJ. Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of

charged solutes and selective isolation of phosphopeptides. Anal Chem. 2008;80:62-76.
[24] Guo K, et al. Stable-isotope dimethylation labeling combined with LC-ESI MS for quantification of amine-containing metabolites in biological samples. Anal Chem. 2007;79:8631-8.

Chapter 6

Towards a comprehensive characterization of a human cancer cell phosphoproteome

Houjiang Zhou*, Serena Di Palma*, Christian Preisinger, Mao Peng, Ayse Nur Polat, Albert J. R. Heck, and Shabaz Mohammed
*These authors contributed equally to this work

J Proteome Res. 2012 Nov 28

Abstract

Mass spectrometry (MS)-based phosphoproteomics has achieved extraordinary successes in qualitative and quantitative analysis of cellular protein phosphorylation. Considering that an estimated level of phosphorylation in a cell is placed at above 100,000 sites, there is still much room for improvement. Here, we attempt to extend the depth of phosphoproteome coverage while maintaining realistic aspirations in terms of available material, robustness and instrument running time. We developed three strategies, where each provided a different balance between these three key parameters. The first strategy simply used enrichment by Ti^{4+}-IMAC followed by reversed chromatography LC-MS (termed 1D). The second strategy incorporated an additional fractionation step through the use of HILIC (2D). Finally, a third strategy was designed employing first an SCX fractionation, followed by Ti^{4+}-IMAC enrichment and an additional fractionation by HILIC (3D). A preliminary evaluation was performed on the HeLa cell line. Detecting 3700 phosphopeptides in about 2 hours, the 1D strategy was found to be the most sensitive but limited in comprehensivity, mainly due to issues with complexity and dynamic range. Overall, the best balance was achieved using the 2D based strategy, identifying close to 17,000 phosphopeptides with less than 1 mg of material in about 48 hours. Subsequently, we confirmed our finding with the K562 cell sample. When sufficient material was available, the 3D strategy increased the phosphoproteome coverage allowing over 22,000 unique phosphopeptides to be identified. Unfortunately, the 3D strategy required more time and over 1 mg of material before it started to outperform 2D. Ultimately, combining all strategies, we were able to identify over 16,000 and nearly 24,000 unique phosphorylation sites from the cancer cell lines HeLa and K562, respectively. In summary, we demonstrate the need to carry out extensive fractionation for deep mining of the phosphoproteome and provide a guide for appropriate strategies depending on sample amount and/or analysis time.

1. INTRODUCTION

Protein phosphorylation is a key component of cellular signal transduction and plays a critical role in many biological processes [1, 2]. Aberrant protein phosphorylation can often be correlated with diseases [3, 4]. Hence, comprehensive characterization of protein phosphorylation is crucial [5, 6]. One of the aims of phosphoproteomics is to globally identify all phosphorylation sites present in a cell [7-9] at a given/defined timepoint. However, the task is not simple since protein phosphorylation is often of low stoichiometry and covers a high dynamic range, placing considerable strains on its detection by mass spectrometry.

In the past decade, a multitude of technological developments have allowed substantial advances in large-scale phosphoproteome profiling of tissues [5, 10-12], cancer cells [13-15],stem cells [16-19], etc. A common step in most methods involves enrichment of phosphopeptides. Currently, the most commonly used chemical chelation affinity-based methods are immobilized metal ion affinity chromatography (IMAC) using a transition metal (Fe^{3+} [20], Ga^{3+}[21], Zr^{4+}[22] or Ti^{4+}[14, 23]) or metal oxide affinity chromatography (MOAC) with a material such as titanium dioxide (TiO_2) [24, 25]. However, in the face of extreme complexity of cellular protein digests, the single step enrichment is not effective enough to identify all phosphopeptides. Therefore, extensive fractionation using liquid chromatography-based techniques are required in conjunction with enrichments in order to reduce the complexity of the proteome. Among fractionation-based techniques, strong cation exchange (SCX)[12, 26-28], strong anion exchange (SAX)[29, 30], electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) [31] and hydrophilic interaction liquid chromatography (HILIC) [32, 33] represent some popular possibilities. However, SCX is by far the most common fractionation technique, where peptides are separated depending on their solution-phase charge state at an acidic pH (2.7). In fact, SCX combined with IMAC/MOAC has proven effective at identifying/quantifying thousands of phosphorylation sites in many biological systems [6, 15, 18, 34-36]. However, SCX fractionation can hinder the analysis of phosphopeptides rich in basic residues [14]. Interestingly, contrary to the well-established strategy of SCX followed by TiO_2, Gerber and co-workers have recently reported an alternate strategy wherein TiO_2 was followed by SCX [37] and demonstrated good reproducibility and a nearly comparable phosphopeptide identification level. Although great gains have been made through enrichment and fractionation, the field is still scratching the surface of the phosphoproteome, since it has been suggested that over 100,000 sites are present at any time in a cell[38]. In order to dig deeper, it appears necessary to go above and beyond two dimensional chromatography of an enriched phosphopeptide sample. Ficarro et al. reported an on-line multidimensional RP-SAX-RP-LC separation system and demonstrated an impressively highly efficient separation and phosphopeptide identification rate [39]. We recently demonstrated that an additional WAX separation/fractionation between SCX and RP-LC can lead to a significant increase in the number of identified phosphopeptides [40].

Apart from effective phosphopeptide enrichment and separation, efficient MS analysis of phosphopeptides is also crucial for profiling a phosphoproteome. For instance, peptide fragmentation technologies such as CID, 'HCD' and ETD have shown distinct advantages toward phosphopeptides with differing physicochemical characteristics [41], where some reports have exemplified that CID possesses sensitivity and speed [42], 'HCD' has more comprehensive fragmentations and lower neutral losses [43] and ETD is more efficient for sequencing highly charged phosphopeptides [14, 44, 45]. Mass spectrometers such as the LTQ-Orbitrap Velos or Elite [46] provide a powerful platform for deep phosphoproteome analysis due to their sensitivity, speed and possession of all three fragmentation technologies.

Recently, we developed a high-resolution and highly sensitive HILIC separation [47, 48], as

well as a less biased phosphopeptide enrichment approach based on Ti^{4+}-IMAC [14, 23]. To leverage the strength of both these technologies, here we combined them to develop an efficient approach with the aim of maximizing the coverage of the cellular phosphoproteome. We therefore designed and systematically assessed three strategies including a sole Ti^{4+}-IMAC enrichment (1D), Ti^{4+}-IMAC enrichment followed by HILIC fractionation (2D), and SCX-Ti^{4+}-IMAC followed by HILIC fractionation (3D). The evaluation was carried on two human cancer cell lines (i.e. HeLa and K562, the latter system is often used as model for chronic myeloid leukemia (CML)), generating phosphoproteome resources of respectively 17,000 and nearly 24,000 unique phosphorylation sites.

2. EXPERIMENTAL SECTION

All experimental procedures are depicted in Figure 1.

Cell culture and sample preparation. HeLa cells were cultured in DMEM and K562 were cultured in RPMI 1640 media, both supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin at 37 °C in 5% CO2. Cells were lysed in buffer containing 50 mM ammonium bicarbonate (pH 8.0), 8 M urea, 1 mM sodium orthovanadate, complete EDTA-free protease inhibitor mixture (Roche) and phosSTOP phosphatase inhibitor mixture (Roche) with ultrasonication for 1 min at 60 W using 30 cycles with 50% duty. Cell debris was then removed by centrifugation at 20000 × g for 15 min at 4 °C. Protein concentration was measured using a Bradford assay (BioRad, Veenendaal, The Netherlands) and then split into 2 mg aliquots for enzymatic digestion. Proteins were reduced with 10 mM dithiothreitol for 30 mins at 56 °C and alkylated by addition of 20 mM iodoacetamide for 30 min and incubation in the dark at room temperature. Lysates were then digested using Lys-C at an enzyme/protein ratio of 1/50 and incubation for 4 h at 37 °C. Then the solution was diluted to a final urea concentration of 2 M with 50 mM ammonium bicarbonate, and trypsin was added with at an enzyme/protein ratio of 1/50 and incubated at 37 °C overnight. The digestion was quenched by acidification to 2% formic acid. The digests were desalted using Sep-Pak C18 cartridges, dried in vacuo and stored at -80 °C for further use.

SCX for peptide fractionation (only for 3D strategy). The desalted digests were reconstituted in 10% formic acid and then loaded onto a C18 cartridge (50 mm × 4.6 mm) (Aqua, Phenomenex) using an Agilent 1100 HPLC system and SCX fractionation was carried out as described before [14, 27]. After combining a few neighbor 'blank' fractions, a total of 22 fractions were collected and distinguished into 3 main populations corresponding to +1, +2, and >+2 charged state peptides, according to the separation obtained in SCX-mode and recorded by the UV chromatogram. Each fraction was desalted using 50cc Sep-Pak C18 cartridges (Waters Corporation). The eluted peptides were then dried in vacuo and reconstituted in the Ti^{4+}-IMAC loading buffer for phosphopeptide enrichment.

Phosphopeptide enrichment (for 1D, 2D and 3D strategies). The Ti^{4+}-IMAC beads were prepared as previous reports [14, 23, 49]. The Ti^{4+}-IMAC Gel-loader tips were built using a C8 plug and the Ti^{4+}-IMAC beads (500 μg of beads/200 μl pipette tip). To reduce variations of enrichment processes, in parallel spin tip enrichment was used in this study. The Ti^{4+}-IMAC spin tip was constructed by placing the gel loader tip into an eppendorf tube with a centrifuge column adaptor. The spin tip enrichment is clean, fast and easy to use and each tip can enrich phosphopeptides up to 250 μg of cellular protein digest. The prepared spin tip was conditioned by using 30 μl of loading buffer (80% ACN/6% TFA) and centrifugation at 2500 × g for 2 min. Non-fractionated protein digests or the desalted SCX fractionations in 80% ACN/6% TFA were applied to the spin tip and centrifuged at 1000 × g for 30 min. Then, the spin tip was sequentially washed with 30 μl of washing buffer 1 (50% ACN, 0.5% TFA

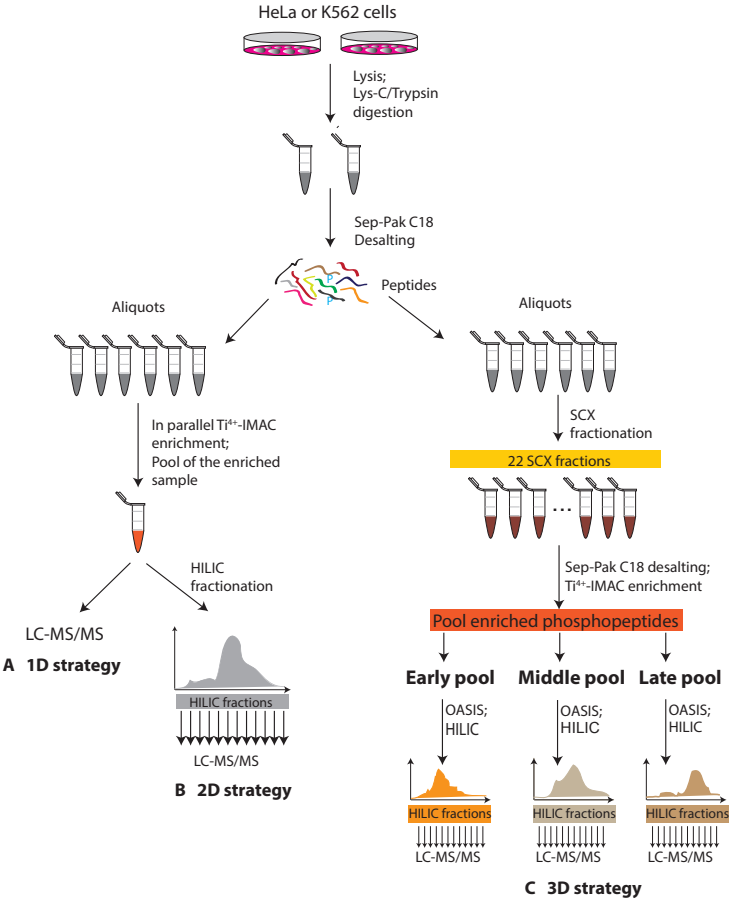

Figure 1: *Schematic overview of the evaluated 1D, 2D and 3D phosphoproteome strategies. HeLa or K562 cells lysates were digested into peptides using a double digestion with Lys-C/trypsin. The resulting peptides were acidified and then desalted using a Sep-Pak C18 column. Aliquots were prepared for Ti^{4+}-IMAC enrichment or for the SCX pre-fractionation. A, In the 1D strategy, Ti^{4+}-IMAC enriched samples were subjected to a direct 2 hours LC-MS analysis. B, In the 2D strategy, Ti^{4+}-IMAC enriched samples were first fractionated by HILIC. Subsequently, HILIC fractions were subjected to 2 hour LC-MS analyses. C, In the 3D strategy, SCX was first used to fractionate the digest. The SCX fractions were desalted and pooled into three discrete populations based on the expected charge states of "1+", "2+", and "3+ or higher". Subsequently, these 3 populations were enriched using Ti^{4+}-IMAC. Afterwards, each of the enriched phosphopeptides population was fractionated by HILIC, and the fractions subjected to 2 hour LC-MS analyses.*

containing 200 mM NaCl) and additional washing buffer 2 of 30 µl of 50% ACN/0.1% TFA, each centrifuge at 1500 × g for 15 min, respectively. The bound peptides were eluted into a new collection tube (already containing 35 µl of 10% formic acid) with 20 µl of 10% ammonia by centrifugation at 1500 × g for 15 min. A final elution was performed with 2 µl of 80% ACN/2% formic acid for 2 min. The collected eluate was further acidified by adding 3 µl of 100% formic acid. For 1D strategy, the enriched phosphopeptides was directly injected to RP for LC-MS/MS analysis. For 2D and 3D strategy, the resulting phosphopeptides were then desalted using the OASIS system (Waters Corporation, Milford, MA) and eluted directly with the HILIC loading buffer containing 95% ACN, 2% formic acid and 5 mM ammonium acetate, ready for the HILIC fractionation.

HILIC for the enriched phosphopeptides fractionation (for 2D and 3D strategies). The fractionation of the enriched phosphopeptides by HILIC was performed on an "inert" Dionex Ultimate LC systems using a vented column setup and operated as reported previously [47].

The enriched phosphopeptides from un-fractionated HeLa or K562 digests (for 2D) or from SCX fractions were loaded on a home-packed HILIC column (ZIC-cHILIC, 100 μm × 20 mm, 5 μm, 200 Å) at 100% loading buffer (95% ACN, 2% formic acid, and 5 mM ammonium acetate). The analytical HILIC column (ZIC-cHILIC, 75 μm × 200 mm, 5 μm, 200 Å) was in line with the HILIC trap column, connected via a T-piece (see the detailed schematic in our previous work [47]. The flow rate during loading step was set at 10 μl/min. The flow though the trap column was collected during the 10-min loading step and dried in vacuum for further RP-LC-MS/MS analysis. Eluting buffer consisted of 0.07% formic acid, containing 5 mM ammonium acetate. Elution was achieved with a gradient of 0-55% in 40 min at a flow rate of 0.35 ml/min passively split to 300 nl/min. During HILIC elution, 1-min fractions were collected in a 96-well plate, with each well containing 40 μl of 10% formic acid per fractions, from 24 min to 38 min and 2-min fractions were taken from both early (20-23 min) and late (39-50 min) gradient. Half of each of HILIC fractions was subjected to RP-LC-MS/MS analysis.

Mass spectrometric analysis. The peptides were subjected to a reversed phase nano-LC-MS/MS analysis consisting of an Agilent 1200 series HPLC system (as described before [50]) connected to an LTQ-Orbitrap Velos Mass Spectrometer equipped with an electron transfer dissociation (ETD) source (ThermoFisher Scientific, Bremen, Germany). The HPLC system was equipped with a 20 mm ReproSil-Pur 120 C18-AQ (Dr. Maisch GmbH, Germany) trapping column (home packed; 100 μm, 5 μm) and a 400-mm ReproSil-Pur 120 C18-AQ (Dr. Maisch GmbH, Germany) analytical column (home packed; 50 μm, 3 μm). Trapping was performed at 5 μl/min in solvent A (0.1 M acetic acid in ddH2O) for 10 min. Elution was performed with a gradient from 10% to 26% (v/v) solvent B (0.1 M acetic acid in 80%/20%: ACN/ ddH2O) in 65 min followed by a gradient of 26%–50% (v/v) in 25 min and up to 100% in 3 min, 100% for 2 min and finally back to 100% solvent A for 15 min with a total analysis of 120 min. The flow rate was passively split from 0.6 ml/min to 100 nl/min. During the gradient, the column effluent was directly introduced into the ESI source of the MS. Nanoelectrospray was achieved using a distally coated fused silica emitter (360 μm outer diameter, 20 μm inner diameter, 10 μm tip inner diameter; constructed in house) biased to 1.7 kV. The LTQ-Orbitrap Velos instrument was operated in the data-dependent mode to automatically switch between MS and MS/MS. MS data were acquired with a data-dependent decision tree method as previously described [51]. Briefly, survey full-scan MS spectra were acquired after accumulation to a target value of 5e5 in the linear ion trap from m/z 350 to m/z 1500 in the Orbitrap with a resolution of 60,000 at m/z 400. For internal mass calibration, the 445.120025 ion was used for lock mass with a target lock mass abundance of 0%. Charge state screening was enabled and precursors with unknown charge state or a charge state of 1 were excluded. Dynamic exclusion was enabled (exclusion size list 500, exclusion duration 40 s). After the survey scans, the ten most intense precursors were subjected to HCD, ETD-IT or ETD-FT fragmentation. A programmed data-dependent decision tree determined the choice of the most appropriate technique for a selected precursor [51]. Essentially, doubly charged peptides were subjected to HCD fragmentation and more highly charged peptides were fragmented using ETD. The normalized collision energy for HCD was set to 40% and the target value was set to 3e4. The resulting fragments were detected in the Orbitrap with resolution 7500. Supplemental activation was enabled for ETD. The ETD reagent target value was set to 2e5 and the reaction time to 50 ms. The ten most intense precursor ions were sequentially isolated to a target value of 5000 and fragmented in the high-pressure linear ion trap collision cell with ETD. The resulting fragments were detected in either the LTQ or Orbitrap.

Data analysis. All MS data were processed with Proteome Discoverer (version 1.3, Thermo Scientific) with standardized workflows. In details, peak lists containing HCD and ETD frag-

mentation were generated with Proteome Discoverer with a signal-to-noise threshold of 3. The ETD non-fragment filter was also taken into account as precursor peak removal within a 4 Da, charge-reduced precursor removal within a 2 Da, and removal of known neutral losses from charge-reduced precursor within a 2 Da (the maximum neutral loss mass was set to 120 Da). All generated peak lists were searched against a Swissprot database (version 56.2, taxonomy filter was set to Homo sapiens) using Mascot software (version 2.3.02 Matrix Science, UK). The database search was performed with the following parameters: a mass tolerance of ±50 ppm for precursor masses; ±0.6 Da for ETD-ion trap fragment ions; ±0.05 Da for HCD and ETD-orbi trap fragment ions, allowing two missed cleavages, cysteine carbamidomethylation as fixed modification and methionine oxidation, phosphorylation on serine, threonine, and tyrosine and protein N-terminal acetylation as variable modifications. The enzyme was specified as trypsin and the fragment ion type was specified as electrospray ionization FTMS-ECD, ETD-TRAP, and ESI-QUAD-TOF for the corresponding mass spectra. A decoy search was enabled. The resulting results were filtered for a 1% false discovery rate (FDR) at the PSM level utilizing the percolator-based algorithm [52]. In addition, only PSMs with Mascot score ≥ 20 were accepted. We grouped potential phosphorylation sites into three classes depending on their site localization probabilities by phosphoRS, which was developed by Mechtler group [53] and is implemented in Proteome Discoverer. In the category with highest confidence in localization (class I), the site has a localization probability for the phospho-group of at least 0.7. In class II, the localization probability is between 0.7 and 0.3, Class III sites had the localization probabilities below 0.3. All classified sites were listed in supplementary material. All results in terms of all PSM, phospho-PSMs, unique phosphopeptides, unique phosphorylation sites, phosphoproteins, phosphopeptide specificity are summarized in Table 1 and 2 and can be found in details within supplementary information.

Functional annotation. For functional annotation analysis of phosphoprotein data sets, we used the annotation tool PANTHER [54] and Ingenuity Pathway Analysis (IPA) [55]. Canonical pathways analyses were performed with p value of 0.05 (Fisher's exact test). Phosphorylation motifs enriched in our phosphoproteome data sets were obtained with the Motif-X algorithm at a significance of $p<10^{-6}$ [56].To predict which protein kinase phosphorylated the identified phosphorylation site, we used the algorithm of group-based prediction system (GPS)[57]. GPS provides an high-throughput approach for prediction of phosphorylation site-specific kinase.

3. RESULTS

3.1 Quality of enrichment

The major aim of this study was to develop and evaluate efficient methods to gain further depth in the profiling of cellular phosphoproteomes, whereby we use a relatively new affinity material, a polymer-based Ti^{4+}-IMAC spin tip. First, we evaluated the optimal loading of the Ti^{4+}-IMAC spin tips by applying differing amounts of K562 lysate digests (0.1, 0.25, 0.5 and 1 mg). We found that the number of identified phosphopeptides reached a plateau when 250 μg of starting material was brought onto the Ti^{4+}-IMAC spin tips, whereby 90% of the ~3000 enriched peptides were phosphopeptides. At higher loading amounts the specificity became lower. This result convinced us to perform the enrichment steps on sample amount corresponding to maximally 250 μg of digest and to pool together multiple enriched aliquots to obtain higher amount of materials (when necessary).

3.2 Evaluation of 1D, 2D and 3D strategies employing HeLa cells: 2D provides the best balance when working with limited materials

The challenge of combining multiple dimensions of chromatography is to create a balance be-

tween the gain in resolving power versus the time required for the fractions' analysis, also taking into account the additional sample loss occurring with any extra separation step. An initial thorough evaluation of the performance of the 1D (Ti^{4+}-IMAC-RP), 2D (Ti^{4+}-IMAC-HILIC-RP) and 3D (SCX-Ti^{4+}-IMAC-HILIC-RP) strategies was performed using a HeLa digest (see Figure 1 for a schematic workflow). After HeLa cells digestions, we made 250 μg aliquots. For the 1D and 2D strategies, we next performed a Ti^{4+}-IMAC based enrichment on a number of these aliquots and the enriched peptides were further pooled. First, enriched phosphopeptides corresponding to 125 μg of HeLa digests were analyzed by a single 2-hour RP-LC-MS/MS analysis. From this 1D experiment we could already identify 3,726 unique phosphopeptides (FDR < 1%) originating from 1,570 phosphoproteins. We achieved a specificity level of 85%, and a phosphopeptide identification frequency of > 1600 per hour MS time (see Table 1).

We next evaluated our 2D strategy on an equivalent amount of material. Specifically, enriched phosphopeptides were subjected to HILIC fractionation and a total of 20 fractions were collected and analyzed with a 2-hour RP-LC-MS/MS analysis. In this experiment, we identified 9,066 unique phosphopeptides. The overall phosphopeptide selectivity (80%) was quite similar when compared to the 1D strategy (85%); however, due to the longer analysis time, the phosphopeptide identification frequency dropped to ~ 160 per hour MS time (see Table 1). To see where we gained mostly from the pre-fractionation, we calculated the number of unique phosphopeptides observed in each HILIC fraction and then plotted the results for each fraction as a heat map (Figure 2A). We observed a Gaussian distribution where the main phosphopeptide fractions were found spanning the middle of the 20-min separation window (Figure 2A, (column 2D-125 μg-RP)). We analyzed the separation power of the HILIC and discovered that most of phosphopeptides (90%) appeared solely in one of the HILIC fractions.

Next, we wanted to investigate if the sample amount was limiting the outcome of the 2D separation strategy. We repeated the 2D experiment with more input material, corresponding to 500 μg. This led to an almost double number of unique identified phosphopeptides, equivalent to 16,637. Approximately 86% of the phosphopeptides identified from the 2D experiment using 125 μg were also observed in the 500 μg scale experiment (Figure 3B). Such a high overlap suggests a good reproducibility with limited issues related to stochastic sampling.

Afterwards, we evaluated whether dynamic range (and complexity) was playing a restrictive role in the effective identification of phosphopeptides. In fact, when increased the analysis time by 20 folds with the 2D approach versus the 1D, we only improved results by a factor of 5. Thus, we performed a 3D experiment (Figure 1) on a 500 μg scale using an aliquot of the same HeLa digest. This 3D strategy was based on an initial step of pre-fractionation, using a well-established SCX system [27], prior to the 2D approach. Thus, the resulting fractions from SCX were first desalted, then subjected to phosphopeptide enrichment using the Ti^{4+}-IMAC spin-tips system, and finally fractionated by HILIC. Our SCX roughly separates tryptic peptides in three categories based on their charge at pH 2.7, namely '1+' (here referred as 'early pool'), '2+'(referred as ' middle pool'), and '3+ or higher' (referred to as the 'late pool') [14, 27, 58]. Each of these three charge-based pools was treated as an individual population. The three sets of HILIC fractions deriving from the 3 SCX populations (in total 3 x 21 fractions) were analyzed by 2 hours of RP-LC-MS/MS. A total of 4,738, 4,167 and 5,018 unique phosphopeptides were identified from the early, middle and late pool, respectively (Table 1). Notably, the enrichment selectivity in the three pools showed differences, in which the early pool had the highest level of specificity (96%), the middle and the late pool resulted in lower specificities of ~70% (Table 1). The high specificity observed for the early pool underlies the benefit of using a dual (phospho)enrichment by SCX and Ti^{4+}-IMAC (or any chelation strategy) toward low charge phosphopeptides [14]. In fact, the early pool from SCX mainly contains singly phosphorylated

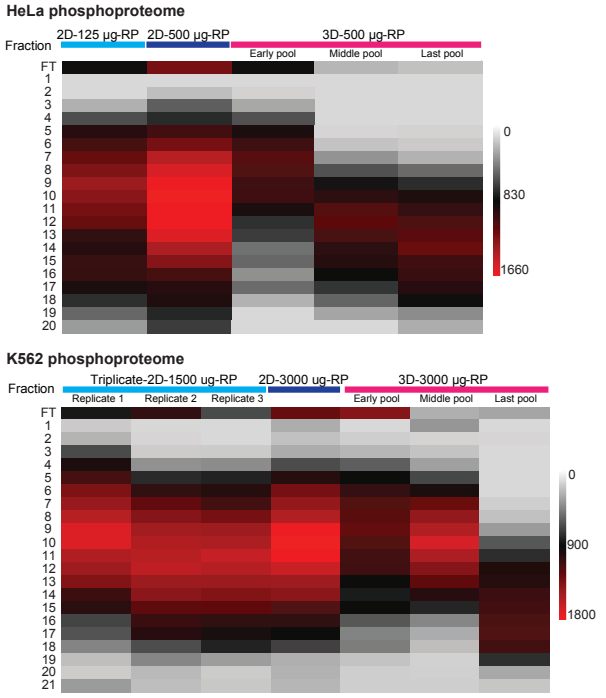

Figure 2. *Heat maps displaying the number of unique phosphopeptides detected in each HILIC fraction in 2D or 3D strategies. In general, a Gaussian distribution was observed with respect to the number of peptides versus the fraction number. In the 2D approach, the fractions containing most phosphopeptides were observed in the middle of the 20-min separation window. In the 3D HeLa experiment, the 'early' pool had its maximum at ~ fractions 7-10; the 'middle' at fractions 10-12 and 'late' at fractions 11-13. The K562 3D dataset showed similar trends, although with higher identification rates. A displays data for the HeLa phosphoproteome; B for the K562 phosphoproteome.*

peptides. The reason for the lower specificity in higher charge SCX fractions was expected and most likely due to higher sample complexity and higher difficulty in binding phosphopeptides with multiple basic residues to the Ti^{4+}-IMAC spin-tips [14].

We further evaluated the benefit of performing an SCX pre-fractionation, followed by an additional HILIC fractionation. We calculated again the number of unique phosphopeptides observed in each HILIC fraction as a heat map. We could observed, within the HILIC fractions, three distinct profiles for the three different SCX pools (Figure 2A): for the early pool, a maximum peptide identification was observed for fractions 7-10; for the middle pool, the maximum identification was centered around fractions 10-12, while for the late pool around fractions 11-13. Finally, when merging results of all three pools together, we could achieve 11,980 unique phosphopeptides with an overall specificity of 81%. The overlap of identified phosphopeptides between the 2D and 3D strategies was 70%. Surprisingly, the cumulative identification in the 3D strategy was lower than within the 2D strategy (11,980 versus 16,637), at least when using the same amount of material (Table 1).

Upon the level of peptides observed in the 1D, 2D and 3D strategies, it became quite apparent that extending the number of dimensions could induce sample loss, and compromise the analysis. In fact, additional sample handing processes are required to desalt the SCX fractions before IMAC enrichment, and the IMAC samples need to be further handled for subsequent HILIC fractionation. Unsurprisingly, the total loss was greater for the 3D strategy than the 2D strategy, justifying the lower identification (for 3D).

3.3 Evaluation of 1D, 2D and 3D strategies on K562 cells: 3D provides the highest phosphoproteome coverage when sample is not limited

To further pinpoint the strengths/weaknesses of the 2D and 3D strategy, we chose to evaluate their performance with significantly more sample amount and employing a different cell

line, K562. After digestion, the sample was analyzed with 1D, 2D and 3D strategies, as described for the HeLa cells (Figure 1). As a starting point, we evaluated how much material was required employing K562 samples to generate maximal results in 1D analyses. We found that the amount of material required was higher than for HeLa; in fact, we needed 750 µg, while only 125 µg for HeLa. A total of 4,104 unique phosphopeptides on 1,763 proteins could be identified, with nearly 100% specificity (Table 2), analyzing 750 µg of K562 with the 1D strategy. The number of sites obderved for K562 was comparable to that for HeLa. Next, we performed triplicate 2D experiments, analyzing double amount of material as for 1D (1.5 mg). Each 2D analysis generated ~14,000 unique phosphopeptides and, when combining together, a total of 19,800 unique phosphopeptides could be identified with 50% overlap among the 3 experiments (Figure 3C). We also performed only one time our 2D strategy with an increased amount of material corresponding to 3 mg, allowing the identification of nearly 17,000 unique phosphopeptides (Table 2). Significantly, up to 70% of phosphopeptides were identified in common between any two 2D experiments, confirming the good reproducibility of the 2D strategy. Although the 3 mg experiment provided a 20% increase over the 1.5 mg experiment, it was not a dramatic achievement. The result highlighted that, increasing the sample amount up to 3 mg,

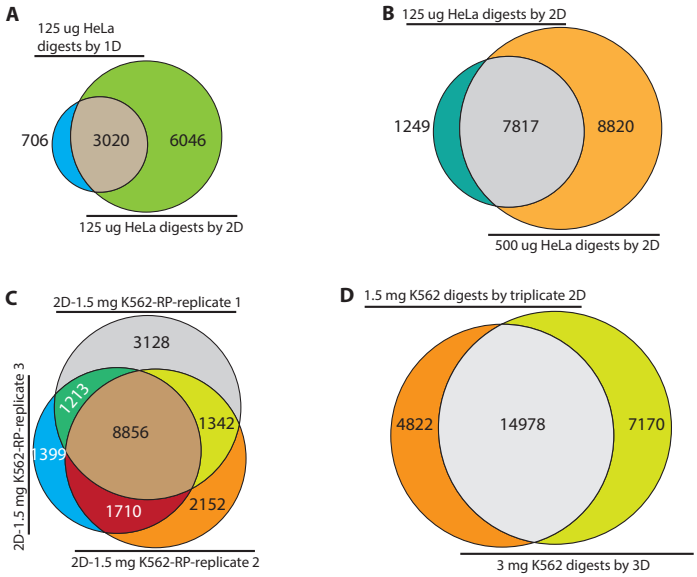

Figure 3: *Venn diagrams displaying the overlap of identified phosphopeptides between several 1D, 2D and 3D strategies. A and B display overlaps for HeLa experiments employing 1D and 2D strategies. C shows overlaps for triplicate analyses with 2D approaches for K562 cells. D shows the overlap between the cumulative data of 1.5 mg of K562 cells in triplicate using 2D strategies versus the 3D experiment starting with 3 mg of K562 lysate.*

Analytical strategy	Sample amount[a]	All PSMs	Phospho PSMs	Selectivity	No. of unique phos-peptides	No. of unique phos-sites	No. of unique phos-proteins	MS analysis time (h)	Identification efficiency	
									sites per hour	sites per µg material
1D	125 µg	6602	5582	85%	3726	3351	1570	2	1676	27
2D	125 µg	27133	21745	80%	9066	7746	2825	48	161	62
2D	500 µg	55176	48175	87%	16637	13681	3891	48	285	27
3D[b]	Early pool	11677	11165	96%	4738	4819	2190	48	100	10
	Middle pool	13699	9931	73%	4167	4041	1971	48	84	8
	Last pool	14466	11134	77%	5018	4466	2127	48	93	9
Summary of 3D	500 µg	39842	32230	81%	11980	10051	3424	144	70	20
Summary of 1D, 2D and 3D	1.25 mg	128572	107732	84%	20788	16740	4284	290	58	30

Table 1. *Summary for the HeLa analysis of all PSMs, Phospho-PSMs, unique identified phosphopeptides, unique phosphorylation sites, phosphoproteins, phosphopeptide specificity, MS analysis time and identification efficiency. [a] represents the total amount of sample that was subjected to LC-MS/MS analysis. [b] the total amount in the 3D experiment was 500 µg.*

the final LC-MS step was partly saturated and incapable of handling the high complexity and dynamic range within each HILIC fraction.
To confirm the finding, we analyzed this high amount of material (3 mg K562 digests) employing the 3D strategy, since an extra dimension should reduce further sample complexity. In total, from 66 LC-MS analyses we identified 9,155, 10,655 and 6,131 unique phosphopeptides from the early, middle and late SCX pools, respectively (Table 2). Merging the three SCX data sets, we detected in total 22,148 unique phosphopeptides on 4,708 phosphoproteins, corresponding to a 32% increase compared to the 2D experiment (22,148 versus 16,722) on the same amount of sample material. In fact, the additional orthogonal SCX step decreased sample complexity and increased the phosphoproteome depth, although requiring about three times more LC-MS time. Notably, this 3D strategy generated more unique phosphopeptides than the sum of the triplicate 2D strategy, though both experiments used similar levels of material (3 mg versus 1.5 mg X 3) and the same LC-MS analysis time. Furthermore, overlapping these two results, we found 14,978 unique phosphopeptides identified in common, representing nearly 60% (Figure 3D).

Analytical strategy	Sample amount[a]	All PSMs	Phospho PSMs	Selectivity	No. of unique phos-peptides	No. of unique phos-sites	No. of unique phos-proteins	MS analysis time (h)	Identification efficiency	
									sites per hour	sites per ug material
1D	750 µg	5783	5718	99%	4104	3776	1763	2	1888	5
Triplicate-2D	1.5 mg-replicate 1	58952	42952	73%	14539	12127	3763	48	253	8
	1.5 mg-replicate 2	44214	39301	89%	14060	11877	3753	48	247	8
	1.5 mg-replicate 3	48532	37846	78%	13178	11041	3609	48	230	7
2D	3 mg	62724	56499	90%	16722	13963	4132	48	291	5
3D[b]	Early pool	43037	35028	81%	9155	8861	3167	48	185	9
	Middle pool	48570	38172	79%	10655	10093	3511	48	210	10
	Last pool	21999	15404	70%	6131	5347	2408	48	111	5
Summary of 3D	3 mg	113682	88634	78%	22148	18055	4708	144	125	6
Summary of 1D, 2D and 3D	10.25 mg	328104	270950	83%	28993	23196	5116	386	60	7

Table 2: *Summary for the K562 analysis of all PSMs, Phospho-PSMs, unique identified phosphopeptides, unique phosphorylation sites, phosphoproteins, phosphopeptide specificity, MS analysis time and identification efficiency.* [a] *represents the total amount of sample that was subjected to LC-MS/MS analysis.* [b] *the total amount in the 3D experiment was 3 mg.*

3.3 The HeLa and K562 cancer cell phosphoproteomes at a glance

Integrating all data led to a total of 16,740 unique phosphorylation sites on 4,284 phosphoproteins for the HeLa cell line and 23,196 unique phosphorylation sites on 5,116 phosphoproteins for the K562 cell line (Tables 1 and 2). Over 80% of the sites were Class I, i.e. unambiguous (see data analysis in experimental section for definition). Over 40% of phosphorylation sites in both the HeLa and K562 data sets were previously unreported, based on comparison with the PhosphositePlus [59] and the Phospho.ELM database (Phospho.ELM 9.0, Sep 2010) [60]. Of note, we could identify a total of 339 (for HeLa) and 363 (for K562) phosphorylated membrane proteins and 276 (for HeLa) and 331 (for K562) phosphorylated protein kinases, indicative of the depth of our phosphoproteome analysis. To the best of our knowledge, this data constitutes the most comprehensive phosphorylation maps of HeLa and K562 cell lines to date. We assessed the HeLa and K562 phosphoproteome separately in terms of canonical pathways, enriched motifs and dominant (major kinases pathways) kinases involved. Extensive studies have shown that the most cellular signaling pathways rely on sequential and coordinated phosphorylation of constituent pathway proteins to relay and propagate the initial signal [5]. A canonical pathway analysis was performed by fitting the identified phosphoproteins from HeLa or K562 into known pathways using IPA. A total of 257 (for HeLa) and 263 (for K562) pathways were identified by Ingenuity in each phosphoproteome data set, respectively. Clearly,

phosphoproteins significantly influence a broad range of cellular pathways within both HeLa and K562. The analysis indicated that most of the phosphoproteins are globally involved in common cellular pathways such as the protein ubiquitination pathway and the ERK/MAPK signaling pathway. For instance, the highest frequency of protein phosphorylation was found in DNA methylation and transcriptional repression signaling pathways in both HeLa (78%) and K562 (83%) cells.

Another example is the ERK/MAPK pathway, which is a key ubiquitous pathway that transduces cellular information on growth, differentiation and carcinogensesis within a cell [61]. We found that 77 phosphoproteins (for HeLa) and 92 phosphoproteins (for K562) were connected with the ERK/MAPK pathway. Although the coverage of canonical signaling pathways in HeLa and K562 was quite similar, we did observe the involved proteins with different phosphorylation patterns. One prominent expected example is the BCR/ABL fusion gene that has constitutive tyrosine kinase activity in K562 cells. A well-known hallmark in the K562 cell line is CML (chronic myeloid leukemia) signaling pathway, which is initialized by tyrosine phosphorylated BCR/ABL [62]. We mapped 42 phosphoproteins from HeLa and 50 phosphoproteins from K562 involved in this pathway, respectively. An illustrative example of how the current dataset and strategies may be used to sample and map the CML signaling pathway is shown in Figure 4.

3.4 Overview of the identified phosphorylation site in HeLa and K562 cells

The distribution of serine, threonine and tyrosine phosphorylation sites of HeLa and K562 are similar to previous large-scale phosphoproteomic studies [5, 8], i.e. for serine phosphorylation ~83%, for threonine ~15% and for tyrosine ~2%. We also analyzed the distribution of phosphosites per protein (Figure 5A) and found that 60% of total HeLa phosphoproteins and 65% of total K562 phosphoproteins contain more than one site, in agreement with the findings reported on mouse tissue phosphoproteomes [5]. For a global view, we compared all identified phosphorylation sites from HeLa and K562. A total of 9,858 phosphorylation sites were identified to be in common (Figure 5B). The high overlap may be indicative of a (more than likely) biased approach. A second (less likely) reason is that these technologies are catching the majority of the abundant and most common phosphorylations present in cells.

Motif-X was used to see if different kinase families are more or less active in the two cell lines [56]. We obtained 143 and 154 motifs from HeLa and K562 with high significance ($P<10^{-6}$), respectively. The enriched motifs broadly contained proline-directed, acidic, basic, tyrosine and non-typical motifs.

4. CONCLUSIONS

In the last decade, extensive efforts in technology development for MS-based phosphoproteomics have allowed remarkable progress in large-scale phosphoproteomics. However, remaining limitations of current analytical strategies lie in inefficient separation, low specificity and inherent bias of phosphopeptide enrichment methods, which still hamper optimal profiling of cellular protein phosphorylation. In this study, we combined analytical strategies by leveraging the strength of Ti^{4+}-IMAC and HILIC, as well as SCX, with an aim of maximal in-depth characterization of the cellular phosphoproteome. For the 1D strategy, we could achieve a high specificity (nearly 90%) of phosphopeptide enrichment without any fractionation, limiting LC-MS analysis time to 2-3 hours. When applying HILIC to separate the enriched phosphopeptides (2D), we could double the number of phosphopeptide identifications. Importantly, the additional HILIC separation shows highly orthogonal characteristic with RP-LC and, thus, provides a feasibility of increasing the coverage of the phosphoproteome. Besides, our 2D

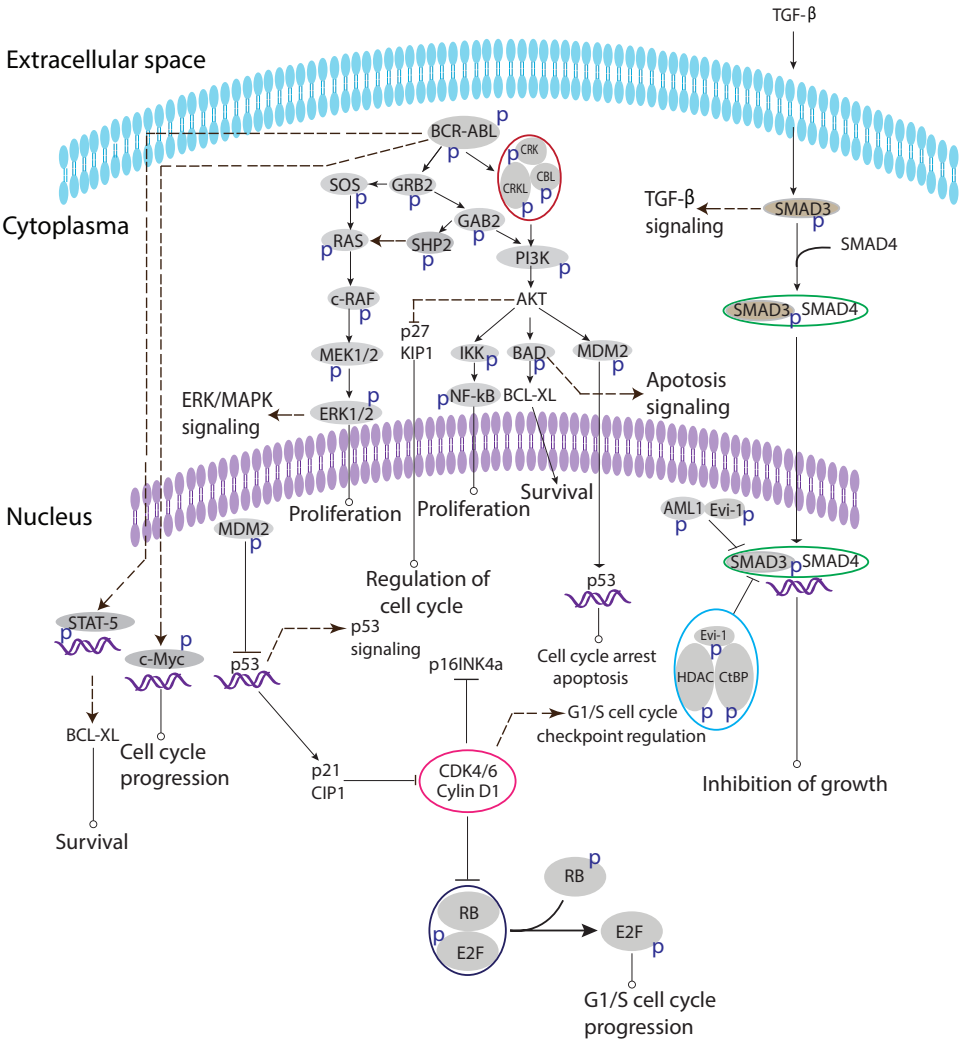

Figure 4. *Chronic Myeloid Leukemia signaling pathway [63, 64] and phosphoproteins observed in our K562 analyses. The identified phosphoproteins are shaded in grey and annotated with a purple "p".*

experiment could cover over 80% phosphopeptides detected in the 1D experiment, emphasizing that there is little bias introduced by HILIC. Here, the superior performance of the 2D strategy can be attributed to a few facts. First, Ti^{4+}-IMAC enrichment enables us to efficiently recover all types of phosphorylated peptides. Second, the low abundant phosphopeptides are better separated from the high abundant phosphopeptides by the additional HILIC fractionation and, therefore, more easily detected by MS. Third, multiple MS analyses could alleviate MS under-sampling, which is often present in shotgun proteomics. The 2D strategy proved to be the method of choice when sample material was limited and only showed levels of saturation when analyzing up to 3 mg. Above this level of material, dynamic range represented the bigger issue and the benefits of a 3D strategy by SCX-Ti^{4+}-IMAC-HILIC became apparent. Considering that the level of phosphorylation present in a cell may be above 100,000 sites at

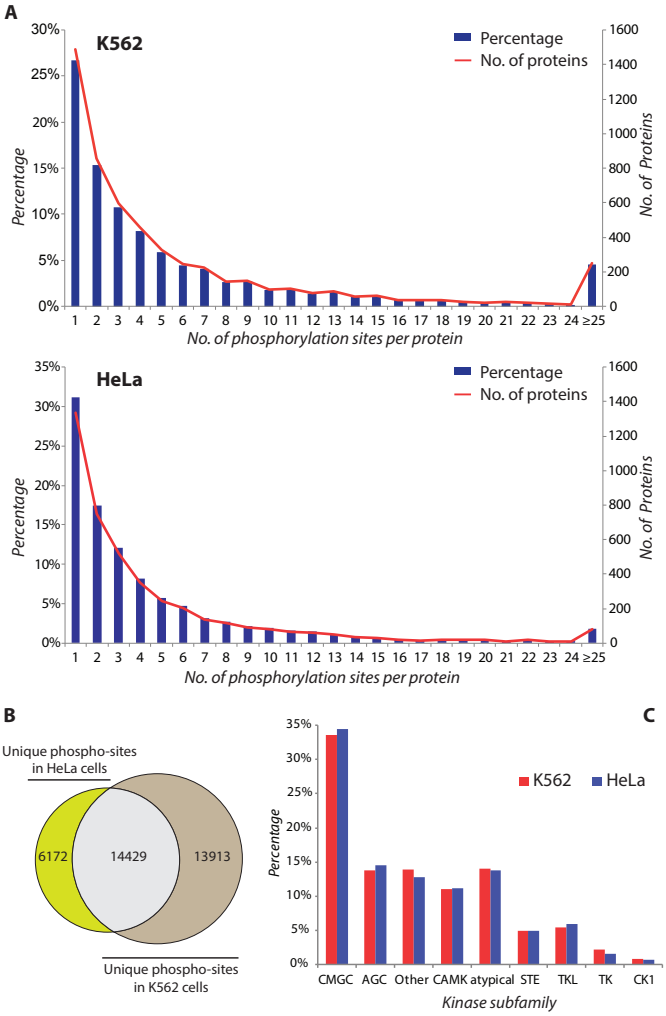

Figure 5. *Overview of the identified phosphorylation sites in HeLa and K562 cells. A, Histogram depicting the number of sites detected per phosphoprotein. B, Global overlap of identified phosphorylation sites between HeLa and K562 experiments. C, Relative contribution of different kinase subfamilies as predicted by the algorithm "group-based prediction system" [57].*

any given timepoint, we have to conclude that we likely still do not reach comprehensivity. Our data, however, reveal strong overlap in-between replicate analysis, indicating that robust phosphoproteomics datasets can nowadays be generated.

5. ACKNOWLEDGEMENTS

This work was supported in part by the PRIME-XS project with the grant agreement number 262067, funded by the European Union 7th Framework Program; The Netherlands Proteomics Centre, embedded in the Netherlands Genomics Initiative; the Netherlands Organization for Scientific Research (NWO) with the VIDI grant (700.10.429).

6. REFERENCES

[1] Pawson T, et al. Protein phosphorylation in signaling--50 years and counting. Trends Biochem Sci. 2005;30:286-90.

[2] Hunter T. Tyrosine phosphorylation: thirty years and counting. Curr Opin Cell Biol. 2009;21:140-6.

[3] Brognard J, et al. Protein kinase signaling networks in cancer. Curr Opin Genet Dev. 2011;21:4-11.

[4] Harsha HC, et al. Phosphoproteomics in cancer. Mol Oncol. 2010;4:482-95.

[5] Huttlin EL, et al. A tissue-specific atlas of mouse protein phosphorylation and expression. Cell. 2010;143:1174-89.

[6] Olsen JV, et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell. 2006;127:635-48.

[7] Aebersold R, et al. Mass spectrometry-based proteomics. Nature. 2003;422:198-207.

[8] Cox J, et al. Quantitative, high-resolution proteomics for data-driven systems biology. Annu Rev Biochem. 2011;80:273-99.

[9] Olsen JV, et al. Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci Signal. 2010;3:ra3.

[10] Ballif BA, et al. Large-scale identification and evolution indexing of tyrosine phosphorylation sites from murine brain. J Proteome Res. 2008;7:311-8.

[11] Villen J, et al. Large-scale phosphorylation analysis of mouse liver. Proc Natl Acad Sci U S A. 2007;104:1488-93.

[12] Ballif BA, et al. Phosphoproteomic analysis of the developing mouse brain. Mol Cell Proteomics. 2004;3:1093-101.

[13] Leitner A, et al. Probing the phosphoproteome of HeLa cells using nanocast metal oxide microspheres for phosphopeptide enrichment. Anal Chem. 2010;82:2726-33.

[14] Zhou H, et al. Enhancing the identification of phosphopeptides from putative basophilic kinase substrates using Ti (IV) based IMAC enrichment. Mol Cell Proteomics. 2011;10:M110 006452.

[15] Dephoure N, et al. A quantitative atlas of mitotic phosphorylation. Proc Natl Acad Sci U S A. 2008;105:10762-7.

[16] Van Hoof D, et al. Phosphorylation dynamics during early differentiation of human embryonic stem cells. Cell Stem Cell. 2009;5:214-26.

[17] Brill LM, et al. Phosphoproteomic analysis of human embryonic stem cells. Cell Stem Cell. 2009;5:204-13.

[18] Rigbolt KT, et al. System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. Sci Signal. 2011;4:rs3.

[19] Phanstiel DH, et al. Proteomic and phosphoproteomic comparison of human ES and iPS cells. Nat Methods. 2011;8:821-7.

[20] Andersson L, et al. Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography. Anal Biochem. 1986;154:250-4.

[21] Posewitz MC, et al. Immobilized gallium(III) affinity chromatography of phosphopeptides. Anal Chem. 1999;71:2883-92.

[22] Zhou H, et al. Zirconium phosphonate-modified porous silicon for highly specific capture of phosphopeptides and MALDI-TOF MS analysis. J Proteome Res. 2006;5:2431-7.

[23] Zhou H, et al. Specific phosphopeptide enrichment with immobilized titanium ion affinity chromatography adsorbent for phosphoproteome analysis. J Proteome Res. 2008;7:3957-67.

[24] Pinkse MW, et al. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. Anal Chem. 2004;76:3935-43.

[25] Pinkse MWH, et al. Highly Robust, Automated, and Sensitive Online TiO 2-Based Phosphoproteomics Applied To Study Endogenous Phosphorylation in Drosophila melanogaster. Journal of Proteome Research. 2007;7:687-97.

[26] Beausoleil SA, et al. Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci U S A. 2004;101:12130-5.
[27] Gauci S, et al. Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. Anal Chem. 2009;81:4493-501.
[28] Alpert AJ, et al. Peptide orientation affects selectivity in ion-exchange chromatography. Anal Chem. 2010;82:5253-9.
[29] Dai J, et al. Protein phosphorylation and expression profiling by Yin-yang multidimensional liquid chromatography (Yin-yang MDLC) mass spectrometry. J Proteome Res. 2007;6:250-62.
[30] Han G, et al. Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography. Proteomics. 2008;8:1346-61.
[31] Alpert AJ. Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. Anal Chem. 2008;80:62-76.
[32] McNulty DE, et al. Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. Mol Cell Proteomics. 2008;7:971-80.
[33] Albuquerque CP, et al. A multidimensional chromatography technology for in-depth phosphoproteome analysis. Mol Cell Proteomics. 2008;7:1389-96.
[34] Benschop JJ, et al. Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis. Mol Cell Proteomics. 2007;6:1198-214.
[35] Kettenbach AN, et al. Quantitative phosphoproteomics identifies substrates and functional modules of Aurora and Polo-like kinase activities in mitotic cells. Sci Signal. 2011;4:rs5.
[36] Koch A, et al. Mitotic substrates of the kinase aurora with roles in chromatin regulation identified through quantitative phosphoproteomics of fission yeast. Sci Signal. 2011;4:rs6.
[37] Kettenbach AN, et al. Rapid and reproducible single-stage phosphopeptide enrichment of complex peptide mixtures: application to general and phosphotyrosine-specific phosphoproteomics experiments. Anal Chem. 2011;83:7635-44.
[38] Lemeer S, et al. The phosphoproteomics data explosion. Current Opinion in Chemical Biology. 2009;13:414-20.
[39] Ficarro SB, et al. Online nanoflow multidimensional fractionation for high efficiency phosphopeptide analysis. Mol Cell Proteomics. 2011;10:O111 011064.
[40] Hennrich ML, et al. Improving depth in phosphoproteomics by using a strong cation exchange-weak anion exchange-reversed phase multidimensional separation approach. Anal Chem. 2011;83:7137-43.
[41] Boersema PJ, et al. Phosphopeptide fragmentation and analysis by mass spectrometry. J Mass Spectrom. 2009;44:861-78.
[42] Jedrychowski MP, et al. Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. Mol Cell Proteomics. 2011;10:M111 009910.
[43] Nagaraj N, et al. Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. J Proteome Res. 2010;9:6786-94.
[44] Chi A, 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;104:2193-8.
[45] Swaney DL, et al. Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. Proc Natl Acad Sci U S A. 2009;106:995-1000.
[46] Michalski A, et al. Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. Mol Cell Proteomics. 2011:doi: 10.1074/mcp.O111.013698.
[47] Di Palma S, et al. Zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC and ZIC-cHILIC) provide high resolution separation and increase sensitivity in proteome analysis. Anal Chem. 2011;83:3440-7.

[48] Di Palma S, et al. Highly sensitive proteome analysis of FACS-sorted adult colon stem cells. J Proteome Res. 2011;10:3814-9.
[49] Mohammed S, et al. Strong cation exchange (SCX) based analytical methods for the targeted analysis of protein post-translational modifications. Curr Opin Biotechnol. 2010;22:9-16.
[50] Hornbeck PV, et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res. 2011;40:D261-70.
[51] Diella F, et al. Phospho.ELM: a database of phosphorylation sites--update 2008. Nucleic Acids Res. 2008;36:D240-4.
[52] Data were analyzed through the use of IPA (Ingenuity® Systems, www.ingenuity.com).
[53] Dhillon AS, et al. Untying the regulation of the Raf-1 kinase. Arch Biochem Biophys. 2002;404:3-9.
[54] Horita M, et al. Blockade of the Bcr-Abl kinase activity induces apoptosis of chronic myelogenous leukemia cells by suppressing signal transducer and activator of transcription 5-dependent expression of Bcl-xL. J Exp Med. 2000;191:977-84.
[55] Faderl S, et al. Mechanisms of disease - The biology of chronic myeloid leukemia. N Engl J Med. 1999;341:164-72.
[56] Fabbro D. BCR-ABL SIGNALING A new STATus in CML. Nat Chem Biol. 2012;8:228-9.
[57] Schwartz D, et al. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. Nat Biotechnol. 2005;23:1391-8.
[58] Xue Y, et al. GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. Mol Cell Proteomics. 2008;7:1598-608.
[59] Yu Z, et al. Preparation of monodisperse immobilized Ti(4+) affinity chromatography microspheres for specific enrichment of phosphopeptides. Anal Chim Acta. 2009;636:34-41.
[60] Raijmakers R, et al. Automated online sequential isotope labeling for protein quantitation applied to proteasome tissue-specific diversity. Mol Cell Proteomics. 2008;7:1755-62.
[61] Frese CK, et al. Improved Peptide Identification by Targeted Fragmentation Using CID, HCD and ETD on an LTQ-Orbitrap Velos. Journal of Proteome Research. 2011;10:2377-88.
[62] Kall L, et al. Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods. 2007;4:923-5.
[63] Taus T, et al. Universal and confident phosphorylation site localization using phosphoRS. J Proteome Res. 2011;10:5354-62.
[64] Thomas PD, et al. PANTHER: A library of protein families and subfamilies indexed by function. Genome Res. 2003;13:2129-41.

Chapter 7

A protocol: ZIC-cHILIC as a fractionation method for sensitive and powerful shotgun proteomics

Serena Di Palma, Albert J. R. Heck, and Shabaz Mohammed

Nat Protocols. 2012 Oct 25;7(11):2041-55.

Abstract

Multidimensional liquid chromatography (LC) combined with mass spectrometry (MS) has become a standard technique in proteomics to reduce sample complexity and tackle the dynamic range in protein abundance. Fractionation is necessary to obtain a comprehensive analysis of complex biological samples such as tissue and mammalian cell lines. However, extensive fractionation comes at the expense of sample loss, presenting a bottle-neck in the analysis of limited amounts of material. In this protocol, we describe a two-dimensional chromatographic strategy based on a combination of hydrophilic interaction liquid chromatography (HILIC; with a zwitterionic packing material, ZIC-cHILIC) and reversed phase (RP) chromatography, which allows proteomic analyses with minimal sample loss. Experimental aspects related to obtaining maximum recovery will be discussed, including how to optimally prepare samples for this system. Examples involving protein lysates originating from cultured cell lines and cells sorted by flow cytometry will be used to show the power, sensitivity and versatility of the technique. Once the ZIC-cHILIC fractionation system has been optimized and standardized, this protocol requires approximately 5-6 days, including sample preparation and fraction analysis.

1. INTRODUCTION

In the past decade, mass spectrometry-based proteomics has allowed the systematic characterization of complex proteomes, thereby becoming an essential tool for biologists and biochemists in their efforts to understand the molecular mechanisms regulating cellular systems.[1, 2] Most proteomics research still requires large amounts of starting material, with several tens of millions of cells- corresponding to milligrams of protein- often being used to access low abundance components.[3]

However, one of the ultimate aspirations of proteomics is the ability to perform single-cell single-proteome analyses and to identify all proteins in a particular biological system, such as a specific cell type or subcellular components/subfractions.[4-9] Rapid innovations in mass spectrometric instrumentation,[10] including improvements in speed and sensitivity, have allowed the analysis of cell populations numbering as low as the thousands.[6, 7, 11] To achieve such a level of detection, each section of the proteomics pipeline needs to be optimized. Focus on sample preparation [12, 13] and peptide separations[14-20] are essential and can represent the key to success for ultra-sensitive proteomics analyses.[4, 6, 7, 21-23]

The most common strategy for global proteome screenings involves the use of two-dimensional separations at the peptide level, where the second separation is predominantly nanoliter-flow reversed phase (RP) chromatography and the first dimension is one of many different techniques that display an orthogonal mechanism of separation towards RP, increasing the overall resolving power.[24-31] Strong cation exchange (SCX) has proven to be the most dominant player in two-dimensional strategies because of its good orthogonality and additional advantages in the enrichment of specific post-translational modifications.[25, 32-35] However, SCX suffers certain limitations regarding the quality of separation, in that clusters of similarly charged tryptic peptides co-elute;[36] in addition, the amount of salt needed for peptide elution can compromise the second step of separation, requiring sample clean-up before the final LC-MS analysis.

Alternative approaches to SCX-RP are constantly being explored, in particular by varying the first stage of chromatography.[37] HILIC represents such an alternative for SCX, and has found widespread use in the analytical field since the pioneering work of Alpert.[38-40] Several different HILIC stationary phases have been used so far,[41] including derivatized silica materials, which can be neutral,[42, 43] ionic,[38, 44, 45] and zwitterionic, such as ZIC-HILIC[46, 47] and ZIC-cHILIC.[48]. Although the exact mechanisms of chromatographic action are different,[37] they share common advantages. HILIC separations are ideal for the analysis of polar and highly hydrophilic compounds, and they often require low salt concentration (or volatile salts) in the buffers, allowing direct coupling to ESI-MS.[49-51] Moreover, HILIC has been shown to have one of the highest degrees of orthogonality to RP of all commonly used peptide separation modes, including SCX, RP at high pH and size exclusion chromatography[36, 52] and it therefore represents an appealing and excellent alternative as a first dimension in two-dimensional strategies.

The application of HILIC with zwitterionic materials is gaining momentum in the proteomics field.[36, 40, 46, 48, 53] Its successful combination with RP in two-dimensional systems relies on their opposite selection for polarity on peptide retention. For example, when compared to SCX, zwitterionic HILIC results in less clustering of similarly charged peptides.[36] Zwitterionic separation materials carry both positive and negative charges on their surface, and separation using them involves both hydrophilic and electrostatic interactions,[54]. However, the electrostatic interactions are weaker than those associated with ionic exchangers, such as SCX.[55] One of the major issues associated with the combination of HILIC and RP in two-dimensional approaches arises from the incompatibility of the mobile phases when employing an

on-line setup. To overcome this problem, we recently designed an off-line 2D system (ZIC-cHILIC-RP) in a nanoscale format,[48] which represents a refinement from an earlier setup developed in our laboratory.[36] To improve sensitivity and to handle small sample volumes, in the new design the column dimensions have been reduced (to 75 μm ID). Furthermore, the HILIC eluent is directly collected during the separation and diluted in an acidified water solvent, compatible with the second dimension, without further manipulation. As the HILIC separation is performed at a flow rate of nanoliters per minute, the volume necessary to dilute the HILIC eluent is on the microliter scale, creating idyllic samples for the consecutive RP-LC-MS step. This 2D method has shown excellent resolution in both first and second dimension with minimal sample loss, and has enabled highly sensitive proteome analyses when compared to current methods that often require more input material.[5, 48] The sample preparation and experimental setup have been optimized for small-scale proteome analysis, equal to a few thousand cells or a few micrograms of proteins (1-10 μg), and has been successfully applied to the in-depth proteome analysis of lysates originating from cultured cell lines[48, 54] or sorted stem cells extracted from mice intestinal tissue.[5]

In addition, this strategy is compatible with most of the stable isotopic quantitative methods used in modern proteomics, as far as the differentially labeled peptides are undistinguishable species during the chromatographic separation.[54, 56] For instance, we have demonstrated[54] the applicability of our two-dimensional HILIC-RP strategy in combination with dimethyl labeling for the comparison of 3 samples in parallel.

The protocol describes how to construct a HILIC system for a 2 dimensional analysis of a sample. The protocol includes column packing, LC setup and gradient optimization. The time required for an optimal set-up depends on the level of expertise of the reader and assumes some experience with nanoLC and column packing. Once the HILIC apparatus has been constructed, the protocol requires maximally 5-6 days, including the MS analysis, when starting at the sample preparation stage. Naturally, the timeline can vary as a function of the number of fractions one wants to analyze and length of LC-MS gradient that is chosen.

2. EXPERIMENTAL DESIGN

During our optimization, we discovered that a few technical hurdles had to be addressed before HILIC could be efficiently employed in a highly sensitive proteomics workflow. In this protocol we report our latest robust set-up for implementing a ZIC-cHILIC chromatographic separation in a 2D ZIC-cHILIC-RP strategy.

The following experimental design has been optimized in our laboratory and can be modified or adapted according to the sample under study and the user's need:

1. Each step of sample preparation prior to the HILIC fractionation, which includes cell lysis, protein digestion and sample clean-up, has to be optimized for low microgram-scale amounts of starting material. However, sample preparation can be adjusted for higher amount of starting material (Supplementary Method 1). When performing the protocol for the first time, it is advised to test any single step on a small sample amount (< 10 μg), such as a simple protein mixture consisting of BSA, α- and β-casein. The optimization of sample preparation is crucial to maximizing sample recovery and minimizing sample loss occurring when handling small and low-volume samples.

2. The sample desalting and clean up before the HILIC fractionation is essential to remove undesirable salts, reagents and buffers that may compromise the separation, thus decreasing the resolving power. Peptides must be concentrated in a high-organic content buffer, preferably the HILIC loading buffer, to enhance the binding efficiency toward the HILIC material.

3. The gradient elution for HILIC fractionation is chosen according to sample complexity and

number of fractions one wants to collect. The goal of selecting a suitable gradient is to find a good compromise in terms of optimum resolution during the HILIC separation in the first dimension, and in terms of acceptable time invested for the analysis of all the fractions in the second dimension.
4. Before using the off-line setup for sample fractionation, the HILIC separation should be monitored in an on-line manner, preferably with an MS read-out, or, alternatively, a UV detector. The on-line screening can be performed using a standard tryptic peptide mixture (for instance with a mixture of BSA, α- and β-casein digests) to easily allow the evaluation of the retention time window appropriate for fractionation. When this step is carefully optimized, it gives the advantage of selecting the right number of fractions to be analyzed in the subsequent second dimension, reducing the MS analysis time.
5. The protocol is illustrated mainly for the qualitative proteome analysis of small-scale samples obtained by cultured cells and from tissue extraction combined with FACS-sorting, but it can be adapted or modified for sample coming from different origins, such as cells isolated by laser microdissection, subcellular components or subfractions, etc. When a specific sample type is analyzed for the first time, it is strongly recommended to test and optimize each step of the sample preparation, in particular lysis, protein extraction, and digestion. The procedure reported here, is, in theory, generally applicable to any source of sample. However, the volume of lysis buffer, reduction and alkylation reagents have to be adjusted according to sample volume/amount (or size, if for instance it is a dissected sample). Moreover, the digestion efficiency should be evaluated and optimized where attention should be paid to the concentration of the solution and the ratio between enzyme and substrate.
6. The HILIC system has been tested for sample amounts ranging from 1 to 10 μg. However, the size of the trap and analytical columns can be increased to accommodate the need of analyzing a higher amount of sample.
7. The method can also be combined with most of the quantitative strategies used in proteomics, such as metabolic and chemical labeling. As an optional step, here we report on how to implement the dimethyl labeling, which is itself described in detail in a previous protocol,[57] in this two dimensional HILIC-RP pipeline. The dimethyl labeling is introduced at the peptide level and can allow the relative quantification and comparison of 3 samples in parallel, without compromising the sensitivity of the analysis. It can be easily performed during the step of on-column sample clean-up, solely adding an extra step to the qualitative 'standard' protocol. When considering quantitative experiments, keep in mind the variation observed at the biological level is higher than the variation observed at the stable isotope labeling and proteomics analysis level (i.e. biological replicates are far more important than technical replicates).
8. The RP-LC-MS set-up description is omitted in this protocol. However, the procedure related to column packing is applicable also for RP trap and analytical columns, and the vented column design has been extensively described in previous work.[58, 59]

3. MATERIALS

REAGENTS

* ZIC-cHILIC, Zwitterionic Bonded Phase on Silica 5 μm, 200 Å (Merck SeQuant, cat. no. 2612-000)
* Fused silica with 75 μm ID and 360 μm OD (Polymicro, cat. no. TSP075375)
* Fused silica with 100 μm ID 360 μm OD (Polymicro, cat. no. TSP100375)
* Formamide, (Merck, cat. no. 104008)
* Kasil potassium silicate solution (PQ Europe, cat. no. 1624)
* Acetone (Merck cat. no. 100014)

* Ethanol 96% (Anthonides, cat. no. I-A5.102.46A)
* Acetic acid (Merck, cat. no. 1.00063)
* Acetonitrile (ACN) (Biosolve, cat. no. 012007)
* Ammonium acetate (Sigma-Aldrich, cat. no. A1542)
* Formic acid (Fluka, cat. no. 94318)
* High purity water obtained from a Milli-Q purification system (Millipore)
* Urea (Merck, cat. no. 66612)
* Ammonium bicarbonate (NH_4HCO_3; Fluka, cat. no. 09830)
* Complete Mini EDTA-free cocktail (Roche, cat. no. 11.836.170.001)
* PhosphoSTOP Phosphatase Inhibitor cocktail (Roche, cat. no. 04.906.845.001)
* DL-dithothreitol (DTT) (Sigma-Aldrich, cat. no. 43815)
* Iodoacetamide (IAA) (Sigma-Aldrich, cat. no. I6125)
* Trypsin (Promega, cat. no. V528A)
* Lysyl endopeptidase (Lys-C) MS grade (Wako Chemicals, cat. no. 129-02541)
* HeLa cell pellet (Cil Biotech, cat. no. cc-01-10-50)
* Bovine serum albumin (BSA; Sigma-Aldrich, cat. no. A2153)
* Alpha-casein (Sigma-Aldrich, cat. no. C6780)
* Beta-casein (Sigma-Aldrich, cat. no. C6905)

Optional reagents for dimethyl labeling:
* Formaldehyde (CH_2O) (37% (vol/vol), Sigma, cat. no. 252549)
* Formaldehyde (CD_2O) (20%, 98% D, Isotec, cat. no. 492620)
* Formaldehyde ($^{13}CD_2O$) (20%, 99% 13C, 98% D, Isotec, cat. no. 596388)

! CAUTION Formaldehyde solutions and formaldehyde vapors are toxic; prepare solutions in a fume hood.

* Sodium cyanoborohydride ($NaBH_3CN$) (Fluka, cat. no. 71435)
* Sodium cyanoborodeuteride ($NaBD_3CN$) (96% D, Isotec, cat. no. 190020)
* Sodium dihydrogen phosphate (NaH_2PO_4) (Merck, cat. no. 1.06346)
* Di-sodium hydrogen phosphate (Na_2HPO_4) (Merck, cat. no. 1.06580)

Oasis C18 solvents
* Conditioning solvent: 100% acetonitrile.
* Equilibrating/washing solvent: 10% (vol/vol) formic acid in water.
* Eluting solvent: ZIC-cHILIC solvent A (see below).

HPLC solvents
* ZIC-cHILIC: solvent A, 95% (vol/vol) acetonitrile, 2% (vol/vol) formic acid and 5 mM ammonium acetate; solvent B, 0.07% (vol/vol) formic acid and 5 mM ammonium acetate; final pH is 3.5 for both solvents.
* RP: solvent A, 0.6% (vol/vol) acetic acid, pH 2.9; solvent B, 0.6% (vol/vol) acetic acid and 80% (vol/vol) ACN, pH 3.5.

4. EQUIPMENT
* Silica cutter (CRS)
* Oven, 100°C (Tamson)
* Magnetic stirrer (Ikamag)
* Pressure bomb
* Tank of compressed Helium
* Microscope (Cole Farmer, 1477.53.031)
* TFE Teflon tubing 0.3 mm (Sigma, cat. no. 58698-U)

* Eppendorf Centrifuge 5417R (Eppendorf)
* LTQ Orbitrap mass spectrometer (Thermo Fisher) and/or OrbitrapVelos (Thermo)
* NanoHPLC system (Dionex 'Ultimate')
* NanoHPLC system (1100 Agilent)
* Milli-Q purification system (Millipore)
* Oasis Vac tC18 3cc cartridges (Waters Corporation)
* Visiprep DL Vacuum manifold system (Supelco)
* Flow control valve liners for the Visiprep-DL (Supelco)
* Vacuum centrifuge (Thermo Scientific)
* Vortex mixer (VWR)
* Vacuum manifold (Waters Corporation)
* Oasis HLB μElution plate 30 μm (Waters Corporation, cat. no. 186001828BA)
* Low-binding Eppendorf tubes (Eppendorf, cat. no. 0030.108.116)
* PicoTip emitter Silica tip (New Objective, cat. no. FS360-20-10)

5. REAGENT SETUP

Lysis buffer

The lysis buffer is prepared in 50 mM NH_4HCO_3(40 mg/10 mL). For 10 mL, the lysis buffer contains: 4.8 g urea (for a final concentration of 8M); 1 tablet of *Complete Mini EDTA-free cocktail*; 1 tablet of *phosphoSTOP Phosphatase Inhibitor cocktail*. Δ CRITICAL It is recommended to first dissolve urea with a lower volume of NH_4HCO_3 solution (i.e. 7 mL), and then to bring it to a final volume of 10 mL. Δ CRITICAL It is strongly suggested to prepare all the reagents fresh and to add the two tablets just before use. Keep the lysis buffer on ice.

Protein reduction and alkylation reagents

Prepare 25 mM DTT (3.8 mg/mL) and 50 mM IAA (9.2 mg/mL), both in 50 mM NH_4HCO_3. Δ CRITICAL It is recommended to prepare these reagents just before use.

Standard peptide mixture

A standard peptide mixture consisting of combined protein digests of bovine serum albumin (BSA), α- and β-casein is prepared. Each protein is digested separately and subsequently mixed in a 1:1:1 ratio. Per protein, a 5 μg/μL of solution is prepared in water. Per 10 μL of solution, add 2 μL of 25 mM DDT and incubate at 30 °C for 40 min. Subsequently, add 4 μL of 50 mM IAA and incubate at room temperature (20-22 °C) in the dark for 30 min. Then, add another aliquot of 2 μL of 25 mM DDT, and incubate at 30 °C for 40 min. The mixture is diluted 5 times with 50 mM ammonium bicarbonate, and 1 μg of trypsin is added for overnight digestion at 37 °C. The digest is stored at -20 °C and thaw before use. The sample is diluted in HILIC solvent A prior to injection.

Optional for quantitative strategy: on-column dimethyl labeling reagents

* Per sample/label: 900 μl of 50 mM sodium phosphate buffer pH 7.5 (prepared by mixing 200 μl of 50 mM NaH_2PO_4 with 700 μl of 50 mM Na_2HPO_4) is mixed with 50 μl of 4% (vol/vol) formaldehyde in water (CH_2O, CD_2O or $^{13}CD_2O$) and 50 μl of 0.6 M cyanoborohydride in water ($NaBH_3CN$ or $NaBD_3CN$).

Δ CRITICAL It is recommended to prepare these reagents just before use and keep them on ice. See previous protocol for details.[57]

* For the correct isotope combinations of formaldehyde and cyanoborohydride, follow this scheme:

Light label	CH_2O combined with $NaBH_3CN$
Intermediate label	CD_2O combined with $NaBH_3CN$
Heavy label	$^{13}CD_2O$ combined with $NaBD_3CN$

Δ CRITICAL Labeling reagent mixtures should be kept at 4 °C and not stored longer than 24h to ensure labeling efficiency.

6. EQUIPMENT SETUP

On-line ZIC-cHILIC-MS system.

The on-line ZIC-cHILIC-MS is performed on a HPLC system directly coupled to an MS. Our setup employs a Dionex "Ultimate" LC system, where the column effluent is introduced into the ESI source of an LTQ-Orbitrap (or other) mass spectrometer via a distally coated fused silica emitter biased to 2.2 kV. The voltage is applied directly to the metal coated tip using the standard Thermo nanospray source. The HPLC is equipped with a 100 μm × 20 mm, 5 μm, 200 Å, ZIC-cHILIC trap column, and a ZIC-cHILIC 75 μm × 200 mm, 5 μm, 200 Å, analytical column, according to a vented column setup, as shown in Figure 1.

The trap column is employed for rapid loading and concentration of the sample. When the system is designed without trap column, the sample loading might require longer time, thus pre-concentration of the sample is needed. All columns are packed in-house (see details in the procedure). The trap column is designed with a double frit (Figure 2). The regular frit at the front end of the column allows for the column packing; the second frit at the back side of the column prevents back-flushing of the packed material during large pressure changes or unforeseen pressure shocks (e.g. system depressurizing). The analytical column has a single frit since it is less likely to unpack. The tubing connections from the LC pump to the column system are constructed with 100 μm fused silica capillary. The six-port valve is used as a switching valve to control both the flow direction and the on/off switching of the flow restrictor.

Trapping is performed at 100% solvent A with an unsplit flow rate of 10 μL min^{-1} for 10 min or more, depending on the sample injection volume. During the trapping, the flow restrictor is shut off by a plug, while the high-flow restriction of the analytical column is used to direct the flow to the waste outlet (a fused silica capillary of 200 μm × 300 cm which possesses no resistance to the applied flow rate). Hence, there is no flow through the analytical column in this valve position (defined in Figure 1 as trap mode). During the analytical separation, a linear gradient of solvent B passes through both trap and analytical columns. This is achieved by switching the valve to open the flow restrictor and close the waste outlet (defined in Figure 1 as analysis mode), and increasing the pump flow rate at the predetermined value of 0.30 mL min^{-1}, passively split via the restrictor to nanoliters per minute flow rates. The actual column flow rate depends on several parameters, including the resistance offered by the restrictor and the column backpressure, which relates to the packing material size, the column length and inner diameter of the capillary. With our current column setup and with a restrictor capillary of 50 μm × 350 cm, the column flow is expected to be ~300 nl min^{-1} at 100% solvent A (see Table 1 for more details).

Valve position	Pump flow rate μL min^{-1}	Restrictor	Column flow rate	Solvent	Expected pressure (bar)
Trap mode	10	closed	10 μL min^{-1}	100 % A	~ 40-50
Analysis mode	100	open	~ 100 nL min^{-1}	100 % A	~ 20
	200	open	~ 200 nL min^{-1}	100 % A	~ 35-40
	300	open	~ 300 nL min^{-1}	100 % A	~ 55-60
	300	open	~ 250 nL min^{-1}	50% A	~ 105-115

Table 1: *Expected pressure value at the specified solvent composition and flow rate.*

Off-line ZIC-cHILIC-RP system: fractionation. In the offline ZIC-cHILIC-RP-MS, the first dimensional fractionation is performed using the settings described above, but the column output is not connected with the ESI-MS. During the ZIC-cHILIC elution, the column effluent is directly collected as 1-min or 2-min fractions (depending on the gradient used and the number of fractions one wants to obtain) in a 96-well plate, with each well containing 40 µL of 10% formic acid per fraction. The operator can choose the amount and volume appropriate for subsequent nanoLC-RP-MS analysis, although we recommend approximately half the volume (20 µL).

Off-line ZIC-cHILIC-RP-MS system: analysis of fractions. For the fractions analysis most of the reversed phase nanoLC-MS configurations can be used. It is preferable to employ a system that contains a trap column so as to accommodate the sample volumes generated by the HILIC fractionation. In our case an Agilent 1100 HPLC system is connected to a mass spectrometer and is equipped with a 100 µm × 20 mm C18 trapping column and a 50 µm × 400 mm C18 analytical column, using a vented column configuration. Trapping is performed at 5 µL min^{-1} for 10 min with RP solvent A; whereas gradient elution is performed at a column flow rate of ~100 nl min^{-1} (see schematic in Figure 3). Further details for MS analysis can be found in our previous work.[48, 60]

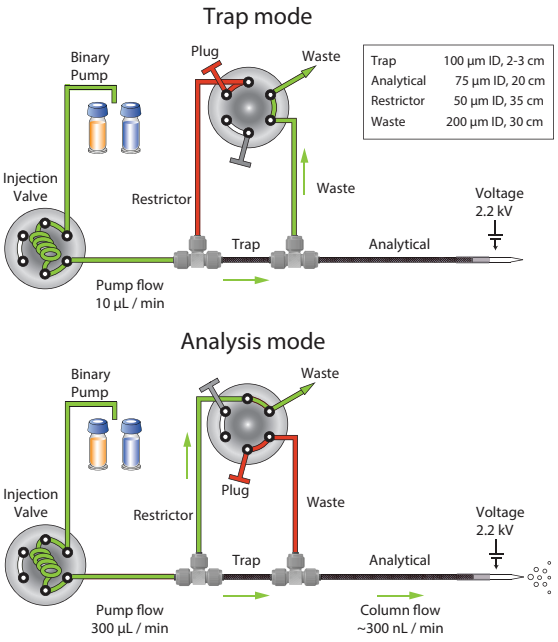

Figure 1: *Schematic representation of the nanoflow ZIC-cHILIC system. Valve positions are given for the loading step (trap mode) and analytical separation (analysis mode). At the moment of sample loading, the flow rate is 10 µL min^{-1} through the trap column only. During this step, no flow passes through the analytical column. After 10 min, the valve is switched to analysis mode, allowing the flow to pass through both trap and analytical column. The pump flow rate is 300 µL min^{-1}, passively split via the restrictor to ~300 nL min^{-1}. At this stage the ESI-MS spectra are acquired.*

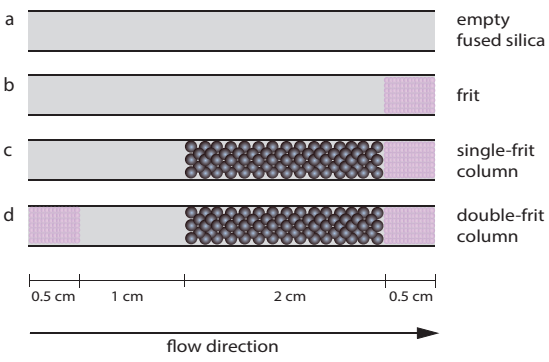

Figure 2: *Schematic representation of the fused silica capillaries during different steps of the packing procedure. The procedure starts with an empty fused silica (a), a silica with frit (b), a single-frit column (c), and, at the end, a double-frit column (d).*

7. PROCEDURE

Packing the HILIC nanoLC columns **O TIMING ~ 7-8 h**

1 | Porous ceramic frits for the trap and analytical columns (steps 1-5) Create a 20 cm-fused silica capillary with a 100 µm ID for the trap

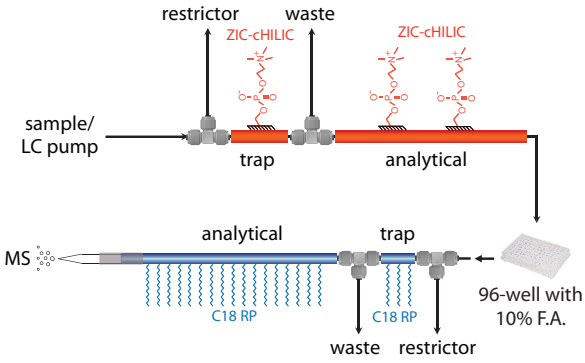

Figure 3: *Schematic design of a two-dimensional HILIC-RP approach, adapted from Di Palma et al.[5] The sample is first concentrated on a ZIC-cHILIC trap column and afterwards separated at nanoliter flow rates in the analytical column. One-minute fractions are directly collected in a 96-well plate already containing an acidified aqueous solvent. Subsequently, the fractions are directly analyzed by RP-LC-MS with a RP trap column for sample desalting and enrichment, followed by the separation and MS detection.*

column and a 50 cm-fused silica capillary with a 75 µm ID for the analytical column.

2 | Prepare the frit solution in a glass vial by mixing 300 µL of potassium silicate and 100 µL of formamide. Vortex immediately upon the addition of formamide.[61] The amount of reagents can be adjusted as long as the final ratio of potassium silicate/formamide is 3:1.

3 | Dip one extremity of the 75 and 100 µm ID fused silica capillaries into the glass vial for a few seconds (2-3 sec).

Δ CRITICAL STEP Check that the frit mixture has entered the fused silica capillaries, reaching a height of 1-2 cm.

4 | Place the tubing into an appropriate vessel and place in a pre-heated oven, which is at 100°C, for one hour in order for the frit solution to polymerize.

5 | Clean the outside of the frit-containing silica capillaries with EtOH.

6 | Trap column with double frit (step 6-28) Cut the frit-ending of the 100 µm silica with a high-precision silica cutter to obtain a frit of 5 mm (see Figure 2, b)

Δ CRITICAL STEP Perform this operation under a source of light or using a microscope.

7 | Place a glass vial containing 1 mL of acetone into the pressure bomb.

8 | Insert the silica, with the frit inside the glass vial, through the high-pressure vessel until almost reaching the bottom of the vial.

9 | Connect the pressure bomb to the compressed helium tank and set the pressure to 50 bar.

! CAUTION Wear protective glasses.

Δ CRITICAL STEP Check that there is flow through the silica.

10 | Rinse the capillary for 2 min with acetone and switch off the bomb.

11 | Prepare the ZIC-cHILIC slurry putting a mini spoon (corresponding to approximately 1 mg) of the material into a glass vial and adding 2 mL of fresh acetone and a mini magnetic stirrer.

12 | Place the vial with slurry in the pressure bomb and switch on the magnetic stirrer.

13 | Insert the silica, with the open-end (without the frit) facing the glass vial, through the high-pressure vessel until almost reaching the bottom of the vial.

14 | Connect the bomb to compressed helium and set the pressure to 50 bar.

15 | Fill the 100 µm silica with ~2 cm of ZIC-cHILIC slurry (see Figure 2, c)

16 | Release the pressure slowly to avoid back-flow of the slurry.

17 | Prepare a fresh 1 mL of acetone in a glass vial and place it in the pressure bomb.

18 | Rinse the packed capillary (trap column) with 1 mL of acetone.

19 | Turn off the pressure when the acetone vial is dried and remove the column from the

bomb.
Δ CRITICAL STEP Use a microscope to check if the column is properly packed, without any gaps between the frit and the material.
? TROUBLESHOOTING
20 | Dry the column with air.
21 | Cut the column at the open-end at 2 cm away from the packed material.
22 | Prepare a fresh frit solution in a glass vial, as reported in the step above.
23 | Dip the open-end of the column in the frit solution for 1-2 seconds.
Δ CRITICAL STEP This step should be performed under a source of light at an angle that allows the operator to monitor the frit solution migrating through the capillary for less than 1 cm, without reaching the ZIC-cHILIC material.
24 | Place the trap into the oven at 100°C for one hour to let the frit polymerize.
25 | Clean the outside of the silica column with EtOH
26 | Cut the second frit of the column at a height of approximately 5 mm (see Figure 2,d)
Δ CRITICAL STEP Perform this operation under a source of light or using a microscope.
27 | Check the ability of the second frit to prevent back-flush. This operation can be performed connecting the trap column (at the end of the second frit) with an empty fused silica (100 μm ID x 20 cm) via a 2 cm-Teflon tubing. Afterwards, the 20 cm-fused silica is flushed with 50% ACN/H2O via the pressure bomb at 50 bar. When quickly releasing the pressure, the packing material should not back-flush through the empty fused silica.
! CAUTION Wear protective glasses.
Δ CRITICAL STEP Do not reverse the flow direction to avoid unpacking of the material.
Δ CRITICAL STEP If the resistance offered by the second frit is too high, and no solvent passes through the column when flushing it, the length of the second frit can be shortened to less than 5 mm.
28 | Label the trap-column indicating the direction of the flow toward the first frit.

29 | Analytical column (step 29-35) Cut the frit-ending of the 75 μm silica with a high-precision silica cutter to obtain a length of 5 mm.
30 | Repeat steps 7-13 (as described for the trap column).
31 | Connect the bomb to compressed helium and set the pressure to 50 bar.
32 | Fill the 75 μm silica with approximately 20 cm of ZIC-cHILIC slurry.
33 | Repeat steps 16-19 (as described for the trap column).
Δ CRITICAL STEP Use a microscope or a source of light to check if the column is properly packed, without any gaps between the frit and the material (see Figure 4a to visualize the apparatus employed during the column packing; and Figure 4b to monitor how the column looks while packing)
? TROUBLESHOOTING
34 | Cut the column at the open-end at 1 cm away from the packed material.
35 | Label the analytical column indicating the direction of the flow toward the frit.

ZIC-cHILIC-LC design **O TIMING ~ 2-3 weeks**
36 | On-line HILIC setup (step 36-44) Connect the trap column to the fused capillary coming from the LC pump and the flow restrictor via the first T-piece (see Figure 1 and 4c).
Δ CRITICAL STEP The trap has to be connected to obtain a flow direction towards the first frit.
37 | Flush the trap with 100% HILIC solvent A in analysis mode at 100 μL min^{-1} for 15 min and then increase the pump flow rate to 300 μL min^{-1} for 15 min.

Δ CRITICAL STEP Before increasing the pump flow rate to 300 μL min^{-1}, monitor if the pressure is at the expected value of ~20 bar when flushing the trap with 100 μL min^{-1} (see Table 1). If the pressure is higher than expected, try to shorten the length of the flow restrictor to decrease the flow through the trap column.
? TROUBLESHOOTING
38 | Decrease the flow rate to 10 μL min^{-1} and, after 1 sec, switch the valve to trap mode (blocking of restrictor line).
Δ CRITICAL STEP Check if the pressure stabilizes, after 3-4 min, at ~40-50 bar (see Table 1).
? TROUBLESHOOTING
39 | Connect the analytical column to the trap column and the waste line via the second T-piece (see Figure 1 and 4c).
40 | Switch the valve to analysis mode (restrictor line open).
41 | Initially, apply a flow rate of 100 μL min^{-1} for 15 min and, then, increase the flow rate to 300 μL min^{-1} for 15 min.
Δ CRITICAL STEP Check if the pressure is as expected at both flow rate values, as reported in Table 1.
? TROUBLESHOOTING
42 | Monitor the flow rate through the analytical column at 300 μL min^{-1}.
Δ CRITICAL STEP The column flow rate can be measured via a 0.3 mm ID Teflon sleeve butt-connected to the end of the analytical column. Checking the distance that the flow needs to travel through the Teflon sleeve in one minute, it is possible to calculate the column flow. If the distance is between 4-5 mm, the flow through the column is ~300 nl min^{-1}.
43 | Connect the column output to a coated silica emitter via the Teflon tube, and into the ESI-MS (see Figure 4d).
44 | Set the capillary voltage at 2.2 kV and monitor the electrospray and MS signal.
Δ CRITICAL STEP The capillary voltage can be adjusted in order to obtain a stable electrospray. Usually, for flow rates of 200-300 nl min^{-1}, an ideal voltage is between 1.7 and 2.2 kV.
? TROUBLESHOOTING
45 | Gradient optimization (step 45-47) Set up different gradients, as reported in Table 2

Analysis time (min)	Trapping condition	Analysis flow rate	Gradient	Equilibration
45	10 min 0% B 10 μl min^{-1}	300 μl min^{-1}	10-62% B in 16 min, 62-70% in 5 min, 70% for 2 min	10 min 0% B 300 μl min^{-1}
60			10-55% B in 31 min, 55-70% in 5 min, 70% for 2 min	
75			10-45% B in 46 min, 45-70% in 5 min, 70% for 2 min	

Table 2: *Gradients of 45, 60 and 75 minutes used for fractionation by HILIC*

for 45-, 60- and 75-minute analysis runs, or choose a more appropriate gradient, according to your need.
46 | Test the gradients injecting into the on-line HILIC setup a standard peptide mixtures consisting of BSA, α- and β-casein digests. See Reagent Setup for preparation.
Δ CRITICAL STEP It is recommended to inject first a small amount of peptide mixture (~ 1 picomole) to avoid column overloading. After checking the chromatogram and MS signal (TIC), the sample concentration can be increased according to the observed results. Sample volume depends on the available sample loop. We suggest to employ a loop of 20-25 μL.
? TROUBLESHOOTING

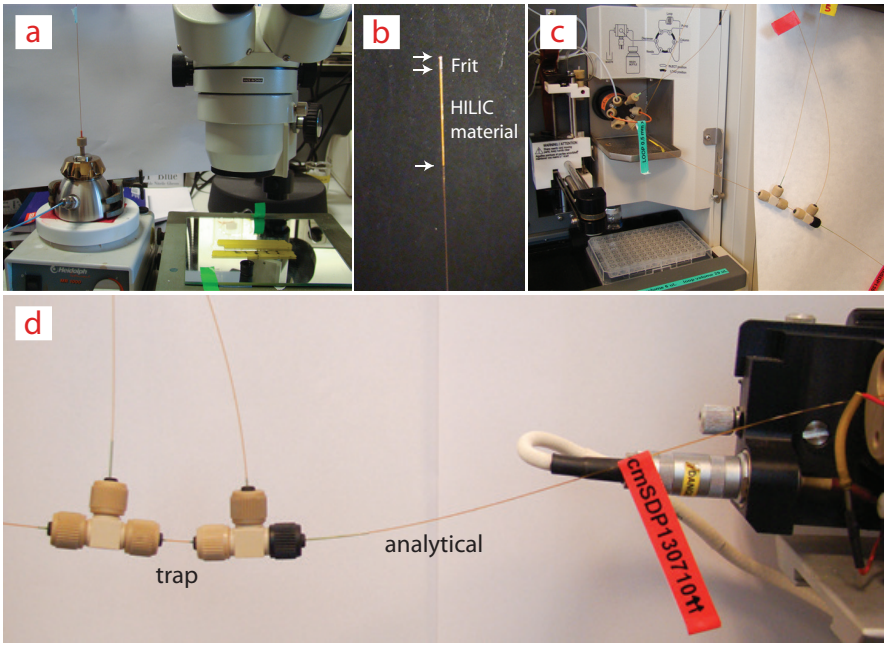

Figure 4: *Experimental setups including a) column packing employing a vessel connected with the pressure bomb and a microscope for monitoring the packing efficiency; b) column during the packing procedure where the frit (between two top arrows) and the packed material (between two bottom arrows) are visible under a source of light; c) vented column setup connected to the LC system; d) analytical column output connected to the ESI-MS source for on-line HILIC configuration.*

47 | Monitor the chromatographic separation, in terms of resolution (peak width at FWHM), retention time and the MS signal.

Δ CRITICAL STEP If the observed peaks are broad (FWHM > 1 minute), the separation can be optimized by testing more/less shallow gradients and increased/decreased flow rates to improve the resolution (narrower peaks) and to change the elution window. Be aware that each LC system can have different lag time between the pump gradient and the actual column gradient, which can cause shifts in the elution profile.

? TROUBLESHOOTING

Sample preparation O **TIMING ~ 2 days**

48 | Cell lysis Lyse the cells as described in options A, B and C respectively for cultured cells, FACS-sorted cells and other sources of cells. See Supplementary Method 1 for an alternative sample preparation approach that would be appropriate for larger amounts/volumes of starting material.

A. For cultured cells (1-10 µg or ~10,000-100,000 cells of expected starting material);

(i) Resuspend the cells with at least 20 µL of lysis buffer and vortex for 10 min at 4 °C

B. For FACS sorted cells (10,000-100,000 cells)

(i) Resuspend the sorted cells in at least 20 µL of lysis buffer and vortex for 10 min at 4 °C

Δ CRITICAL STEP If the sorting procedure is optimized to obtain the sorted cells in a volume below 50 µL, it is recommended to lyse the cells soon after the sorting procedure. In this

case, we suggest to add urea or a concentrated urea solution directly to the sorted cells in order to obtain a urea final concentration between 6M and 8M.
Δ CRITICAL STEP If the cells are sorted in a volume higher than 50 µL, the sample can be dried down and resuspended in 20 µL of lysis buffer.
C. For other sources of sample corresponding to <10 µg of starting material
(i) If the samples have a different origin (distinct cellular population isolated by laser microdissection, subcellular components, embryos, etc.), add enough lysis buffer to cover the pellet, starting with 20 µL, and vortex for 10 min at 4 °C.
Δ CRITICAL STEP It is suggested not to sonicate the sample, independent from its origin. Sample recovery might be lower.

49 | Protein reduction and alkylation (step 49-54) Per 20 µL of solution, add 2 µL of DTT from 25 mM stock to obtain a final concentration of approximately 2 mM.
50 | Vortex, spin down and incubate for 40 min at 30 °C or at room temperature (20-22 °C) for 1h with gentle agitation
51 | Add to the solution 2 µL of IAA from 50 mM stock to obtain a final concentration of 4 mM.
52 | Vortex, spin down and incubate for 30 min in the dark at room temperature
53 | Add to the solution 2 µL of DTT from 25 mM stock to obtain a final concentration of 4 mM together with the previous amount
Δ CRITICAL STEP This step is recommended to reduce over-alkylation
54 | Vortex, spin down and incubate at room temperature for 1h with gentle agitation.

55 | Enzymatic digestion (step 55-58) Add Lys-C endopeptidase from a diluted stock (e.g. 0.01 µg/uL) at an enzyme/protein ratio 1:50 to 1:100 (suggested volume 2-10 µL) and incubate for 4 hr at 37° C.
Δ CRITICAL STEP This step is recommended to improve the efficiency of the digestion. Lys-C operates successfully at high concentration of urea converting insoluble proteins into soluble peptides, which are more easily accessible to trypsin.
56 | Dilute the solution at least 5 times with 50 mM NH4HCO3 (pH 7.8), to obtain a urea concentration below 2M.
57 | Add trypsin at an enzyme/substrate ratio of 1:50 to 1:100 from a diluted stock (e.g. 0.01 µg/uL) (suggested volume 2-10 µL) and incubate overnight at 37° C. An excess of enzyme usually does not represent an issue and the steps with Lys-C and/or trypsin can be repeated, when a ratio 1:100 is used to increase the efficiency of the digestion.
58 | Quench the trypsin digestion by acidifying the solution with formic acid to a final concentration of 1-5% to obtain a final sample volume between 100 and 200 µL
Δ CRITICAL STEP Do not add directly 100% formic acid but preferably use a solution diluted in water (e.g. 40-50 %)

59 | OASIS sample clean-up - and optional dimethyl labeling (steps 59-69) Condition one column of the OASIS cartridge with 200 µL of ACN.
60 | Equilibrate with 200 µL of 10% formic acid (FA) twice.
61 | Load the tryptic peptide digest.
62 | Wash with 200 µL of 10% FA twice.
63 | Optional step for dimethyl labeling (steps 63-64) Flush each of the OASIS columns (maximally three columns containing the three samples to be differentially labeled) five times with 200 µL of the respective labeling reagent (light, intermediate or heavy).

64 | Wash with 200 μL of 10% formic acid (FA) twice to remove the excess of labeling reagents.
65 | Elute the tryptic peptides (after step 62) or the labeled tryptic peptides (after step 64) according to the steps 66-67.
66 | Elute with 25 μL of HILIC solvent A (this volume can be increased if the HILIC system is designed to allow the injection of a higher sample volume)
Δ CRITICAL STEP Collect the 1st eluate in one low-binding Eppendorf tube and store it in the refrigerator at 4 °C.
67 | Elute a second time with 100 μL of HILIC solvent A and dry down the eluate.
Δ CRITICAL STEP Collect the 2nd eluate in another low-binding Eppendorf tube and dry it for 10-15 min by vacuum centrifugation
Δ CRITICAL STEP Due to the high amount of organic solvent in which the eluate is dissolved, the sample can be dried relatively quickly. Monitor the speed of the drying process. It is preferable to reduce the sample to maximally 1-2 μL and to avoid complete dryness.
68 | Reconstitute the second eluate with the 25 μL of the 1st eluate, vortex and spin down.
69 | Optional for dimethyl labeled samples Mix the 3 differentially labeled samples at a 1:1:1 ratio and bring to the desirable final volume for the HILIC injection by vacuum centrifugation.

Two dimensional ZIC-cHILIC-RP-MS O **TIMING ~ 3 days**
! CAUTION: Wear protective glasses.
70 | First dimension separation: Condition column prior to HILIC fractionation (steps 70-72) Before the fractionation, condition the LC system, including the HILIC columns. Start by equilibrating both the trap and analytical column with 100% solvent A for 10-15 min at 300 μL min^{-1} with the switching valve in analysis mode (flow restrictor open).
Δ CRITICAL STEP Monitor the pressure. After full equilibration, it should be in the range 50-60 bar, as reported in Table 1.
? TROUBLESHOOTING
71 | Decrease the pump flow rate to 10 μL min^{-1} and, after one second, set the switching valve in trap mode (flow restrictor closed).
Δ CRITICAL STEP The pump flow rate has to be set to the actual loading flow rate just before starting the analysis to keep both columns fully equilibrated and to ensure a correct pressure value during the injection and loading of the sample.
? TROUBLESHOOTING
Δ CRITICAL STEP An expected pressure value in trapping mode is 40-50 bar, as reported in Table 1. This value can fluctuate depending on the column length and backpressure offered by the packing material and double frit. It is suggested to adjust the flow rate between 5 and 20 μL to obtain a pressure value below 100 bar.
? TROUBLESHOOTING
72 | From a 96-well plate, inject a standard sample into the ZIC-cHILIC–MS configuration in order to compare the acquired chromatogram with your standard reference.
Δ CRITICAL STEP It is recommended to run a standard peptide mixture, prior to the sample analysis, choosing the gradient and settings that will be employed with the real sample to monitor the system equilibration, chromatographic separation, and elution window.
? TROUBLESHOOTING
73 | First dimension separation: HILIC fractionation (steps 73-78) Place the sample (25 μL) in a 96-well plate.
74 | Inject it onto the off-line ZIC-cHILIC-LC system for sample fractionation. See EQUIPMENT SETUP for details about LC configuration and settings.

75 | Collect the flow through (FT) of the trapping column via the waste line in an Eppendorf tube during the loading step (10 min).
76 | Dry down the FT by vacuum centrifugation.
77 | Fill 18-26 wells of a 96-well plate with 40 µL of 10% F.A, according to the chosen gradient time, and keep the plate in a cooled autosampler in order to prevent evaporation
78 | Collect one-minute (or two-minute fractions) into subsequent wells, following the fractionation scheme timeline reported in Table 3 for analysis runs of 45, 60 and 75 minutes.

Analysis run (min)	2-min fractions from/to (min)	1-min fractions from/to (min)	2-min fractions from/to (min)	Total fractions n°
45	18-23	24-34	35-42	18
60	20-25	26-38	39-46	20
75	20-31	32-47	48-55	26

Table 3: *Suggested fraction collection timeline*

79 | Second dimension separation: fractions' analysis by RP-MS (steps 79-81) Reconstitute the FT in 40 µL of 10% F.A.
80 | Analyze the fractions and FT directly by RP-LC-MS/MS injecting half volume (20 µL).
81 | Store the second half at -80 °C for an optional re-analysis.

Data Searching O **TIMING ~ 1-2 days**
82 | Parameters setting for data searching. Search the tandem mass spectra against an appropriate database, e.g., Swiss-prot, using an appropriate search algorithm, e.g., Mascot (http://www.matrixscience.com/). Include the enzyme trypsin in your search and set carbamidomethyl (C) as fixed modification and oxidation (M) as variable modification. With a quantitative analysis employing the 3 differential isotopic labelings, add dimethyl (K) and (N-terminal), dimethyl $^{2}H(4)$ (K) and (N-terminal), dimethyl $^{2}H(6)$ $^{13}C(2)$ (K) and (N-terminal) as variable modifications. For peptide identification, a minimum mascot score of 20 and an FDR below 1% are chosen. In this protocol, the Proteome Discoverer software package (Thermo Scientific) was used to process the data but other software packages, which also support the dimethyl labeling quantitation, are available, e.g. MaxQuant, [62] etc.

Timing
Packing the HILIC nanoLC columns: ~ 7-8 h
Steps 1-5, Porous ceramic frits for the trap and analytical columns: ~1.5 h
Steps 6-28, Trap column with double frit: ~ 2-3 h
Steps 29-35, Analytical column: ~ 2-3 h
ZIC-cHILIC-LC design: ~ 2-3 weeks
Steps 36-44, On-line HILIC setup: ~ 5-6 days
Steps 45-47, Gradient optimization: ~ 5-6 days
Sample preparation: ~ 2 days
Step 48, Cell lysis: ~ 1h
Steps 49-54, Protein reduction and alkylation: ~ 2 h
Steps 55-58, Enzymatic digestion: ~ 4h and overnight
Steps 59-69, OASIS sample clean-up - and optional dimethyl labeling: ~ 2-3 h
Two dimensional ZIC-cHILIC-RP-MS: ~ 3 days

Steps 70-78, First dimension separation: HILIC fractionation: ~ 2-3 h
Steps 79-81, Second dimension separation: fractions' analysis by RP-MS.: ~ 2-3 days
Data searching: ~ 1-2 days
Step 82 Parameters setting for data searching.

Troubleshooting
Troubleshooting advice can be found in the following Table 4.

8. ANTICIPATED RESULTS

In this protocol, we describe a setup for enabling a powerful and highly sensitive proteome analysis on minute amounts of sample, ranging from 1 μg to a maximum of 10 μg, regardless of its biological origin. For the implementation of the HILIC separation in the 2D off-line configuration, we optimized the gradient profile to find a balance between optimal resolution in the first dimension and desirable time for fractions' analysis in the second dimension. We evaluated three analysis times (45, 60, and 75 min), as illustrated in Figure 5, and we could observe a good resolution independent of the chosen analysis time. We found the 60-min run to be a good balance between separation power and analysis time, which corresponds to the collection of approximately 20 fractions. A suggested timeline for fraction collection using these 3 gradients is reported in Table 3.
To provide an example of the performance of our two-dimensional method on a low-scale experiment, we have applied it to the proteome analysis of human cell lines [48, 54]. The HILIC fractionation was performed on 3 μg of HeLa cells digest scale with the 60-min analysis time. In the range wherein most peptides were expected to elute, single-minute fractions were collected to obtain a maximum of 20 fractions. Half of each fraction was subsequently injected and further analyzed by RP-LC run with an analysis time of 120-min. Overall, we analyzed approximately 1.5 μg of the digest by LC-MS/MS with a total analysis time of 40 hours. This strategy allowed the identification of approximately 20,000 unique peptides assigned to almost 3,600 proteins. The distribution of unique peptides in each HILIC fractions showed a Gaussian profile, with maximal identifications observed between fractions 5-9, corresponding to the range of elution between 25 and 30 minute during the 60-min run. Then, we observed that shorter and more hydrophobic peptides eluted on average earlier than larger and more hydrophilic peptides, consistently with the HILIC retention

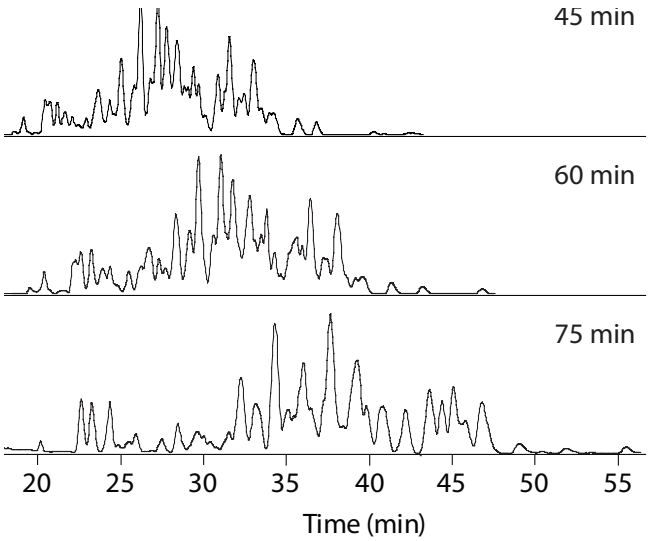

Figure 5: *LC-MS chromatograms for the analysis of 1 picomole of a standard peptide mixture (BSA, α- and β-casein) using the ZIC-cHILIC configuration with different gradient times, as reported in Table 2, ranging from 45 to 75 min. Adapted from Di Palma et al.[48]*

mode (see Figure 6).
To probe the sensitivity of this strategy with an actual small-scale sample, we also analyzed 10,000 adult colon stem cells (from initial 30,000 cells sorted by flow cytometry) directly after the extraction from the mouse.[5] The results showed trends similar to those obtained by HeLa cells, enabling the identification of 15,775 unique peptides originating from 3,775 proteins

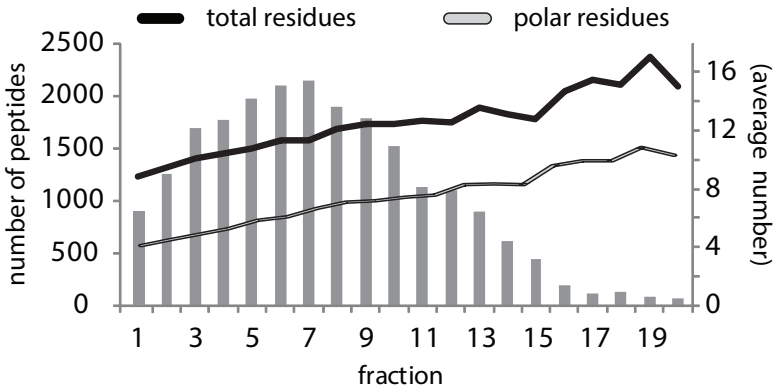

Figure 6: *Fractional distributions of unique peptides (left y-axis) from a HeLa digest after ZIC-cHILIC fractionation. Additionally, the solid lines show the average total number of amino acids (length of peptides) and number of polar amino acids per fraction (right y-axis).*

and providing in-depth proteome coverage that usually requires at least 10 times more starting material.

9. ACKNOWLEDGEMENTS
We want to thank Paul J. Boersema and Reinout Raijmakers for suggestions and critical reading of the manuscript. We thank all members of the Biomolecular Mass Spectrometry and Proteomics Group, particularly Harm Post, Arjan Barendregt and Soenita Goerdayal for their help. This work was in part supported by the PRIME-XS project, Grant Agreement Number 262067, funded by the European Union Seventh Framework Program. The Netherlands Proteomics Centre, embedded in The Netherlands Genomics Initiative and The Netherlands Organization for Scientific Research (NWO) with the VIDI Grant 700.10.429 (for SM) is acknowledged for funding.

Author contribution
S.D.P. designed the experiments, performed experiments, analyzed data and wrote the paper. S.M. and A.J.R.H. designed the experiments, supervised the project and wrote the paper.

Table 4: *Troubleshooting advice.*

Step(s)	Problem	Possible reason	Solution
19, 33	Gap in the packing of the column	There has not been a constant flow of packing material or the material got trapped somewhere	Agitate the column close to the gap area with your fingers or tap the column with a ruler to help removing the obstruction Flush the column with 20% ACN/H2O employing the pressure bomb set at 50 bar for at least 1h Wash the column overnight with the LC in analysis mode setting 50% solvent A at 300 uL min-1
37, 41, 70	High back-pressure in analysis mode	Restrictor is blocked	Replace the restrictor line
		Restrictor is too long	Shorten the restrictor's length or choose a larger ID restrictor line
		T-piece connections or switching valves are blocked	Wash the component and, if the issue persists, replace the part
		Trap or analytical column is blocked (least likely)	Replace the part
37, 38, 71	High back-pressure in trap mode	Waste line, trap column or switching valve is blocked	Replace the part
37, 38, 71	Low pressure or no pressure in trap mode	Trap column is overused or one of the frits is damaged	Replace the trap
		Leakage at the T-piece connections	Reconnect the column to the T-piece
44	No MS signal	No electrospray	Replace the silica emitter if used for long time Check if there is flow through the column Set the capillary voltage at appropriate values according to the column flow rate: for 100-200 nL min-1, ~ 1.5-1.7 kV for 200-300 nL min-1, ~ 1.7-2.2 kV
		Obstruction and/or leakage at Teflon sleeve	Replace/reposition sleeve
44	Unstable MS signal	Unstable electrospray	See suggestions as for no MS signal
		Split electrospray or droplets at the tip of the emitter	Set the silica emitter at an optimal position into the ESI source
46, 47, 72	Broad peaks	Dead volumes in the flowpath	Check the sleeve after the analytical column
		Column is not properly packed or contains a gap as can be observed under a microscope	Use the suggestions as for gap in the packing of the column or replace it
		Sample overloading	Try to reduce sample amount injection
47, 72	Peak tailing	Sub-optimal gradient	Try to adjust the gradient using a steeper slope
		A gap is present somewhere after the pre-column first connection	Check all appropriate connections
47, 72	Fluctuation in retention time	Analytical column is not fully equilibrated	Wash the column with 100% solvent A before starting a new run Check if the equilibration step at the end of the gradient is enough to allow the re-conditioning

		Solvent A/B mixer is unstable or LC pumps are not working properly	Check the components and eventually contact supplier to replace the part
		LC pumps are not working properly	Check it and eventually replace it
		Air bubbles in the system	Purge the lines at high flow rate
		Temperature instability	Monitor the temperature in the room with a proper airco system Use a column oven
47, 72	Lower peak intensity than expected	The sample is not completely injected	Avoid bubbles or particles in the sample with a centrifugation step before injection Check if there is a leakage in the injection system Check if sample volume is appropriate for sample loop size Check if all the sample volume is correctly delivered into the sample loop Check if the flow during the trapping condition allows the sample to reach and completely pass through the trap column
		Ionization is compromised	Replace the silica emitter and apply appropriate voltage.
47, 72	Incomplete chromatogram or missing peaks	Trap column overloaded	Collect the flow through the trap via the waste line and test it for the missing peaks Reduce sample amount injected
		Inefficient trap column	Replace the trap column
		Adsorption to metallic component of the injection and LC system	Wash the system with a chelating agent such as EDTA
		Sample is dissolved in a high concentration of aqueous solvent to be efficiently trapped	Dry down sample in vacuum centrifuge and redissolve in 100% buffer A

10. REFERENCES

[1] Bensimon A, Heck AJR, Aebersold R. Mass Spectrometry–Based Proteomics and Network Biology. Annual Review of Biochemistry. 2012;81:379-405.

[2] Cox J, Mann M. Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. Annual Review of Biochemistry. 2011;80:273-99.

[3] Anderson NL, Anderson NG, Pearson TW, Borchers CH, Paulovich AG, Patterson SD, et al. A Human Proteome Detection and Quantitation Project. Molecular & Cellular Proteomics. 2009;8:883-6.

[4] Waanders LF, Chwalek K, Monetti M, Kumar C, Lammert E, Mann M. Quantitative proteomic analysis of single pancreatic islets. Proc Natl Acad Sci U S A. 2009;106:18902-7.

[5] Di Palma S, Stange D, van de Wetering M, Clevers H, Heck AJR, Mohammed S. Highly Sensitive Proteome Analysis of FACS-Sorted Adult Colon Stem Cells. Journal of Proteome Research. 2011;10:3814-9.

[6] Thakur D, Rejtar T, Wang D, Bones J, Cha S, Clodfelder-Miller B, et al. Microproteomic analysis of 10,000 laser captured microdissected breast tumor cells using short-range sodium dodecyl sulfate-polyacrylamide gel electrophoresis and porous layer open tubular liquid chromatography tandem mass spectrometry. Journal of Chromatography A. 2011;1218:8168-74.

[7] Masuda T, Sugiyama N, Tomita M, Ishihama Y. Microscale Phosphoproteome Analysis of 10.000 Cells from Human Cancer Cell Lines. Analytical Chemistry. 2011;83:7698-703.

[8] Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, et al. Deep proteome and transcriptome mapping of a human cancer cell line. Mol Syst Biol. 2011;7.

[9] Koroleva OA, Cramer R. Single-cell proteomic analysis of glucosinolate-rich S-cells in Arabidopsis thaliana. Methods. 2011;54:413-23.

[10] Makarov A, Scigelova M. Coupling liquid chromatography to Orbitrap mass spectrometry. Journal of Chromatography A. 2010;1217:3938-45.

[11] Nagaraj N, Alexander Kulak N, Cox J, Neuhauser N, Mayr K, Hoerning O, et al. System-wide Perturbation Analysis with Nearly Complete Coverage of the Yeast Proteome by Single-shot Ultra HPLC Runs on a Bench Top Orbitrap. Molecular & Cellular Proteomics. 2012;11.

[12] Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Meth. 2009;6:359-62.

[13] Liu W, Hou Y, Chen H, Wei H, Lin W, Li J, et al. Sample preparation method for isolation of single-cell types from mouse liver for proteomic studies. Proteomics. 2011;11:3556-64.

[14] Thakur SS, Geiger T, Chatterjee B, Bandilla P, Fröhlich F, Cox J, et al. Deep and Highly Sensitive Proteome Coverage by LC-MS/MS Without Prefractionation. Molecular & Cellular Proteomics. 2011;10.

[15] Kocher T, Pichler P, Swart R, Mechtler K. Analysis of protein mixtures from whole-cell extracts by single-run nanoLC-MS/MS using ultralong gradients. Nat Protocols. 2012;7:882-90.

[16] Waanders LF, Almeida R, Prosser S, Cox J, Eikel D, Allen MH, et al. A Novel Chromatographic Method Allows On-line Reanalysis of the Proteome. Molecular & Cellular Proteomics. 2008;7:1452-9.

[17] Unwin RD, Griffiths JR, Whetton AD. Simultaneous analysis of relative protein expression levels across multiple samples using iTRAQ isobaric tags with 2D nano LC-MS/MS. Nature Protocols. 2010;5:1574-82.

[18] Shen Y, Smith RD, Unger KK, Kumar D, Lubda D. Ultrahigh-Throughput Proteomics Using Fast RPLC Separations with ESI-MS/MS. Analytical Chemistry. 2005;77:6692-701.

[19] Dai J, Jin W-H, Sheng Q-H, Shieh C-H, Wu J-R, Zeng R. Protein Phosphorylation and Expression Profiling by Yin-Yang Multidimensional Liquid Chromatography (Yin-Yang MDLC) Mass Spectrometry. Journal of Proteome Research. 2006;6:250-62.

[20] Motoyama A, Venable JD, Ruse CI, Yates JR. Automated ultra-high-pressure multidimensional protein identification technology (UHP-MudPIT) for improved peptide identification of proteomic samples. Analytical Chemistry. 2006;78:5109-18.

[21] Altelaar AFM, Heck AJR. Trends in ultrasensitive proteomics. Current Opinion in Chemical Biology.

2012;16:206-13.
[22] Bandura DR, Baranov VI, Ornatsky OI, Antonov A, Kinach R, Lou X, et al. Mass Cytometry: Technique for Real Time Single Cell Multitarget Immunoassay Based on Inductively Coupled Plasma Time-of-Flight Mass Spectrometry. Analytical Chemistry. 2009;81:6813-22.
[23] Salehi-Reyhani A, Kaplinsky J, Burgin E, Novakova M, deMello AJ, Templer RH, et al. A first step towards practical single cell proteomics: a microfluidic antibody capture chip with TIRF detection. Lab on a Chip. 2011;11:1256-61.
[24] Washburn MP, Wolters D, Yates JR, 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001;19:242-7.
[25] Wolters DA, Washburn MP, Yates JR, 3rd. An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem. 2001;73:5683-90.
[26] Motoyama A, Xu T, Ruse CI, Wohlschlegel JA, Yates JR, 3rd. Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. Analytical Chemistry. 2007;79:3623-34.
[27] Opiteck GJ, Jorgenson JW, Anderegg RJ. Two-Dimensional SEC/RPLC Coupled to Mass Spectrometry for the Analysis of Peptides. Analytical Chemistry. 1997;69:2283-91.
[28] Dai J, Wang LS, Wu YB, Sheng QH, Wu JR, Shieh CH, et al. Fully Automatic Separation and Identification of Phosphopeptides by Continuous pH-Gradient Anion Exchange Online Coupled with Reversed-Phase Liquid Chromatography Mass Spectrometry. Journal of Proteome Research. 2009;8:133-41.
[29] Geng M, Ji J, Regnier FE. Signature-peptide approach to detecting proteins in complex mixtures. Journal of chromatography A. 2000;870:295-313.
[30] Gilar M, Olivova P, Daly AE, Gebler JC. Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. Journal of separation science. 2005;28:1694-703.
[31] Hao P, Guo T, Li X, Adav SS, Yang J, Wei M, et al. Novel Application of Electrostatic Repulsion-Hydrophilic Interaction Chromatography (ERLIC) in Shotgun Proteomics: Comprehensive Profiling of Rat Kidney Proteome. Journal of Proteome Research. 2010;9:3520-6.
[32] Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li JX, et al. Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci U S A. 2004;101:12130-5.
[33] Gauci S, Helbig AO, Slijper M, Krijgsveld J, Heck AJ, Mohammed S. Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. Analytical Chemistry. 2009;81:4493-501.
[34] Ballif BA, Villen J, Beausoleil SA, Schwartz D, Gygi SP. Phosphoproteomic analysis of the developing mouse brain. Molecular & Cellular Proteomics. 2004;3:1093-101.
[35] Lewandrowski U, Zahedi RP, Moebius J, Walter U, Sickmann A. Enhanced N-Glycosylation Site Analysis of Sialoglycopeptides by Strong Cation Exchange Prefractionation Applied to Platelet Plasma Membranes. Molecular & Cellular Proteomics. 2007;6:1933-41.
[36] Boersema PJ, Divecha N, Heck AJ, Mohammed S. Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. Journal of Proteome Research. 2007;6:937-46.
[37] Di Palma S, Hennrich ML, Heck AJR, Mohammed S. Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. Journal of Proteomics. 2012;75:3791-813.
[38] Alpert AJ. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. J Chromatogr. 1990;499:177-96.
[39] Want EJ, Wilson ID, Gika H, Theodoridis G, Plumb RS, Shockcor J, et al. Global metabolic profiling procedures for urine using UPLC-MS. Nat Protocols. 2010;5:1005-18.
[40] Palmisano G, Lendal SE, Engholm-Keller K, Leth-Larsen R, Parker BL, Larsen MR. Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. Nature Protocols. 2010;5:1974-82.

[41] Jandera P. Stationary phases for hydrophilic interaction chromatography, their characterization and implementation into multidimensional chromatography concepts. Journal of Separation Science. 2008;31:1421-37.
[42] McNulty DE, Annan RS. Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. Molecular & Cellular Proteomics. 2008;7:971-80.
[43] Kolarich D, Jensen PH, Altmann F, Packer NH. Determination of site-specific glycan heterogeneity on glycoproteins. Nat Protocols. 2012;7:1285-98.
[44] Lindner H, Sarg B, Meraner C, Helliger W. Separation of acetylated core histones by hydrophilic-interaction liquid chromatography. Journal of Chromatography A. 1996;743:137-44.
[45] Alpert AJ. Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. Anal Chem. 2008;80:62-76.
[46] Intoh A, Kurisaki A, Fukuda H, Asashima M. Separation with zwitterionic hydrophilic interaction liquid chromatography improves protein identification by matrix-assisted laser desorption/ionization-based proteomic analysis. Biomedical Chromatography. 2009;23:607-14.
[47] Jiang W, Irgum K. Tentacle-type Zwitterionic stationary phase, prepared by surface-initiated graft polymerization of 3- N,N-dimethyl-N-(methacryloyloxyethyl)-ammonium propanesulfonate through peroxide groups tethered on porous silica. Analytical Chemistry. 2002;74:4682-7.
[48] Di Palma S, Boersema PJ, Heck AJR, Mohammed S. Zwitterionic Hydrophilic Interaction Liquid Chromatography (ZIC-HILIC and ZIC-cHILIC) Provide High Resolution Separation and Increase Sensitivity in Proteome Analysis. Analytical Chemistry. 2011;83:3440-7.
[49] Fountain KJ, Xu J, Diehl DM, Morrison D. Influence of stationary phase chemistry and mobile-phase composition on retention, selectivity, and MS response in hydrophilic interaction chromatography. Journal of Separation Science. 2010;33:740-51.
[50] Boersema PJ, Mohammed S, Heck AJ. Hydrophilic interaction liquid chromatography (HILIC) in proteomics. Anal Bioanal Chem. 2008;391:151-9.
[51] Grumbach ES, Diehl DM, Neue UD. The application of novel 1.7 μm ethylene bridged hybrid particles for hydrophilic interaction chromatography. Journal of Separation Science. 2008;31:1511-8.
[52] Gilar M, Olivova P, Daly AE, Gebler JC. Orthogonality of separation in two-dimensional liquid chromatography. Anal Chem. 2005;77:6426-34.
[53] Jiang W, Fischer G, Girmay Y, Irgum K. Zwitterionic stationary phase with covalently bonded phosphorylcholine type polymer grafts and its applicability to separation of peptides in the hydrophilic interaction liquid chromatography mode. Journal of chromatography A. 2006;1127:82-91.
[54] Di Palma S, Raijmakers R, Heck AJR, Mohammed S. Evaluation of the Deuterium Isotope Effect in Zwitterionic Hydrophilic Interaction Liquid Chromatography Separations for Implementation in a Quantitative Proteomic Approach. Analytical Chemistry. 2011;83:8352-6.
[55] Hemstrom P, Irgum K. Hydrophilic interaction chromatography. J Sep Sci. 2006;29:1784-821.
[56] Boersema PJ, Aye TT, van Veen TA, Heck AJ, Mohammed S. Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. Proteomics. 2008;8:4624-32.
[57] Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJ. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protoc. 2009;4:484-94.
[58] Raijmakers R, Berkers CR, de Jong A, Ovaa H, Heck AJ, Mohammed S. Automated online sequential isotope labeling for protein quantitation applied to proteasome tissue-specific diversity. Molecular & Cellular Proteomics. 2008;7:1755-62.
[59] Meiring HD, van der Heeft E, ten Hove GJ, de Jong APJM. Nanoscale LC–MS(n): technical design and applications to peptide and protein analysis. Journal of Separation Science. 2002;25:557-68.
[60] Frese CK, Altelaar AFM, Hennrich ML, Nolting D, Zeller M, Griep-Raming J, et al. Improved Pep-

tide Identification by Targeted Fragmentation Using CID, HCD and ETD on an LTQ-Orbitrap Velos. Journal of Proteome Research. 2011;10:2377-88.
[61] Cortes HJ, Pfeiffer CD, Richter BE, Stevens TS. Porous ceramic bed supports for fused silica packed capillary columns used in liquid chromatography. Journal of High Resolution Chromatography. 1987;10:446-8.
[62] Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26:1367-72.

Chapter 8

Summary
Nederlandse Samenvattig
Curriculum vitae
List of publications

SUMMARY
Peptide mass spectrometry (MS) is an invaluable analytical method for biological and biomedical research. This technique, when integrated with liquid chromatography (LC) and database search tools, allows highly sensitive qualitative characterization and highly accurate quantitative comparison of proteomes. Continuous technological advances in MS instrumentation and peptide fractionation strategies lead to an in depth knowledge of proteomes, although a full proteome coverage is still far out of reach due to extreme differences in protein abundances and post-translational modifications (PTMs). Moreover, the targeted analysis of subsets of proteomes, for instance phosphorylated proteins, requires specialized enrichment methods to gain insights into cellular processes that would be inadequately covered by analyzing the whole proteome.
Though the field of proteomics has rapidly evolved in recent years, improvements are still required to increase the sensitivity and dynamic range of LC-MS. In the framework of this thesis, several technological developments are described to advance proteomics towards sensitivity and comprehensivity, and their applicability is demonstrated in 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 increases 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.

In chapter 2, we aimed to highlight some of the advances and new developments that have been made in the area of liquid chromatographic-based separations. We choose to focus on specific LC methods which represent some of the major breakthroughs in peptide separation science and whose applicability will rapidly increase in the proteomic field; namely, reversed-phase (RP), ion exchange (IEX), and hydrophilic interaction liquid chromatography (HILIC). Each separation methodology is discussed in depth, especially their roles in hyphenated multidimensional formats. Special emphasis is given to aspects such as maximizing the resolution, reducing sample complexity, widening the overall dynamic range and, consequently, increasing the proteome coverage.

In chapter 3, two specific zwitterionic-type HILIC stationary phases are evaluated and further optimized for two dimensional-LC (2D-LC) configurations. Exploring the capabilities of both ZIC-HILIC and ZIC-cHILIC, we observed a mixed mode separation consisting of (strong) polar and (weak) electrostatic interactions between 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. We discovered that, although the two zwitterionic materials have opposite charge orientations, they generated similar separations. However, it became apparent that the separation power of ZIC-cHILIC was more pH-independent, while ZIC-

HILIC separation was negatively affected at an acidic pH.
We then implemented these zwitterionic HILIC stationary phases in a multidimensional proteomics workflow, in combination with RP chromatography, and evaluated them in terms of peptide separation efficiency and sensitivity. The use of two orthogonal separation techniques, such as zwitterionic HILIC and RP, in combination with nanoflow rate LC systems and high resolution mass spectrometry, allowed the identification of thousands of proteins from a very modest amount (μg) of starting material.

In chapter 4, we applied the ZIC-cHILIC-RP multidimensional strategy to the analysis of a limited number of (FACS) sorted colon stem cells extracted from mouse intestine. We demonstrated that this robust set-up leads to a significant reduction of sample complexity with nearly negligible sample losses. In fact, our method required just thousands of cells to probe with high sensitivity a specific population of adult colon stem cells. We obtained a proteome coverage comparable to current methods that generally requires 100-fold more starting material.

In chapter 5, we studied the feasibility of combining a quantitative approach based on dimethyl labeling with the ZIC-cHILIC separation for quantitative proteomics. We addressed the potential issue of deuterium isotope effect that would introduce an error during the quantification. In fact, the incorporation of deuterium atoms via dimethyl labeling could alter the fundamental co-elution of differently labeled peptides under ZIC-cHILIC separations. We demonstrated the influence of choosing a specific pH of the solvent to eliminate the deuterium isotope effect. At pH 6.8, hydrophilicity is the major factor governing peptide retention, hampering the co-elution of hydrogen- and deuterium-containing labeled peptides due to differences in their polarity. In contrast, at acidic pH values there is a stronger mixed-mode separation, which decreases the hydrophilicity difference of deuterated and non-deuterated species. We evaluated our findings in the multidimensional ZIC-cHILIC-RP strategy and showed that our approach is suitable to perform unbiased quantitative proteome analysis, resulting in the quantification of thousands of proteins.

In chapter 6, we combined our high-resolution and highly sensitive ZIC-cHILIC separation with a specific phosphopeptide enrichment method based on Ti^{4+}-IMAC with the aim of maximizing the coverage of cellular phosphoproteomes. We designed and systematically compared three analytical strategies including: i) a single Ti^{4+}-IMAC enrichment; ii) Ti^{4+}-IMAC enrichment followed by HILIC fractionation; iii) an SCX-based pre-fractionation, followed by Ti^{4+}-IMAC enrichment and a further step of fractionation by HILIC. The evaluation was carried on two human cancer cell lines (i.e. HeLa and K562). The comparison between the 3 different workflows showed that an extensive fractionation, as reported for the last strategy, is necessary to achieve comprehensivity and in-depth phosphoproteome coverage. However, this comes at the expense of higher sample amounts and longer MS analysis time. Ultimately, we were able to identify over 20,000 and nearly 29,000 unique phosphorylation sites from HeLa and K562 cell lines, respectively, generating useful phosphoproteome resources for the scientific community.

In chapter 7, we described a detailed protocol to implement HILIC fractionation into a sensitive shotgun proteomics strategy. Analyzing the proteome of small amount of cells is a challenging task at present, and a suite of special analytical tools is required to deal with these types of sample. Many techniques commonly used to handle large number of cells with good outcomes cannot be directly applied to a small number of cells. The zwitterionic HILIC based approaches described in this protocol represents a major step forward towards more sensitive

proteome analysis. Experimental aspects related to obtaining maximum recovery from sample preparation, and how to optimally prepare samples for this system, are discussed. Examples involving protein lysates originating from cultured cell lines and cells sorted by flow cytometry are used to show the power, sensitivity and versatility of the technique.
We proposed ZIC-cHILIC as an alternative fractionation method for ultrasensitive proteomics that will find ample application in the analysis of distinct cellular populations, for instance obtained by laser microdissection.

NEDERLANDSE SAMENVATTIG

Peptide massa spectrometrie (MS) is een bruikbare analytische methode voor biologisch en biomedisch onderzoek. Wanneer deze techniek geïntegreerd wordt met vloeistofchromatografie (liquid chomatography; LC) en databank zoekmachines zijn uiterst gevoelige kwalitatieve profileringen en zeer nauwkeurige kwantitatieve vergelijkingen van proteomen mogelijk. Voortdurende technologische ontwikkelingen in MS instrumentatie en peptide scheidingstechnieken leiden tot een diepgaande kennis van proteomen. De analyse van een volledig proteoom is evenwel nog buiten bereik vanwege extreme verschillen in eiwithoeveelheden en post-translationele modificaties (PTMs). Verder zijn specifieke verrijkingsmethoden nodig voor de gerichte analyse van delen van een proteoom, zoals bijvoorbeeld gefosforyleerde eiwitten. Door deze verrijkingsmethoden kan inzicht worden verkregen in cellulaire processen die niet volledig gedetecteerd kunnen worden wanneer het complete proteoom wordt onderzocht.
Hoewel het onderzoeksveld van proteomics zich snel heeft ontwikkeld in de afgelopen jaren zijn verbeteringen nog steeds nodig om de gevoeligheid en het dynamische bereik van LC-MS te verhogen. In het kader van dit proefschrift zijn verscheidene technologische ontwikkelingen beschreven die proteomics vooruit brengen in gevoeligheid en omvattendheid en de toepasbaarheid is aangetoond in verscheidene onderzoekslijnen.

In hoofdstuk 1 is het concept van proteomics geïntroduceerd en verscheidene praktische aspecten van een proteomics werkstroom zijn uitgelicht. Een universeel proteomics draaiboek bestaat niet, omdat verschillende onderzoeksvragen en beschikbaarheid van instrumenten verschillende benaderingen vergen. Toch komen verscheidene onderdelen terug in veel experimentele proteomics procedures. In het algemeen is de lysis van een biologisch monster een van de eerste stappen, gevolgd door enzymatische digestie. Omdat de complexiteit van het monster sterk verhoogd wordt door de digestie, kunnen verscheidene manieren van fractionering worden toegepast, zoals kation uitwisselingschromatografie (strong cation exchange; SCX) en hydrofiele interactie vloeistofchromatografie (hydrophilic interaction liquid chromatography; HILIC). Om eiwitten of peptiden te isoleren die bepaalde PTMs bevatten kunnen verrijkingstechnieken worden toegepast vóór of na de monsterfractionering, zoals geïmmobiliseerde metaalaffiniteitschromatografie (immobilized metal affinity chromatography; IMAC), TiO_2 of op antilichaam gebaseerde immunoprecipitatie. De belangrijkste stap in een proteomics experiment is de aminozuurvolgordebepaling van peptiden met LC-MS om de eiwitten in een monster te bepalen. Dit wordt gedaan door middel van het vergelijken van de massa's van intacte peptiden en hun fragmenten met de theoretische massa's die afgeleid kunnen worden uit genoom databanken.

In hoofdstuk 2 lichtten we enkele nieuwe ontwikkelingen uit die gedaan zijn in het veld van op LC gebaseerde scheidingen. We richtten ons daarbij op specifieke LC methoden die beschouwd kunnen worden als belangrijke doorbraken in peptide scheidingsonderzoek en waarvoor snel vele toepassingen in proteomics gevonden zullen worden; namelijk omgekeerde fase (reversed-phase; RP), ionenuitwisseling en HILIC. Deze scheidingsmethoden werden in detail besproken, daarbij richtend op hun toepassingen in multidimensionaal gekoppelde opstellingen. Speciale aandacht werd gegeven aan aspecten als het maximaliseren van resolutie, vermindering van monstercomplexiteit, vergroten van het dynamisch bereik en de als gevolg daarvan meer omvattende proteoomanalyse.

In hoofdstuk 3 zijn twee specifieke zwitterion-type HILIC stationaire fases geëvalueerd en verder geoptimaliseerd voor gebruik in twee dimensionale LC (2D-LC) opstellingen. Terwijl we

de eigenschappen van zowel ZIC-HILIC als ZIC-cHILIC bestudeerden namen we een tweeledig scheidingsmechanisme waar, bestaand uit (sterke) polaire en (zwakkere) electrostatische interacties tussen de peptiden en de stationaire fase. Het uiteindelijke scheidingsmechanisme kan worden gewijzigd door aanpassing van de pH van de buffers, omdat dit de hydrofiliciteit en lading van peptiden verandert. We hebben ontdekt dat de twee zwitterionische materialen resulteren in vergelijkbare scheidingsprofielen, ondanks de tegengestelde ladingsoriëntaties. De scheidingskracht van ZIC-HILIC bleek negatief beïnvloed te worden door een lage pH, terwijl ZIC-cHILIC minder pH gevoelig bleek te zijn.
We implementeerden daarna de zwitterionische HILIC stationaire fases in een multidimensionale proteomics werkstroom in combinatie met RP chromatografie en evalueerden de peptidescheidingsefficiëntie en gevoeligheid. Het gebruik van twee orthogonale scheidingstechnieken, zoals zwitterionisch HILIC en RP, in combinatie met nano-LC systemen en MS met hoge resolutie, maakte de identificatie van duizenden eiwitten in een kleine hoeveelheid (µg) van uitgangsmateriaal mogelijk.

In hoofdstuk 4 hebben we de multidimensionale ZIC-cHILIC-RP methode toegepast in de analyse van een slechts beperkt aantal met flowcytometrie gesorteerde stamcellen die geëxtraheerd waren uit de dikke darm van een muis. We hebben aangetoond dat deze solide opstelling tot een verlaging van de monstercomplexiteit leidt met verwaarloosbaar verlies van materiaal. Inderdaad hadden we met onze methode slechts enkele duizenden cellen nodig om met hoge gevoeligheid een populatie van volwassen darmstamcellen te karaktariseren. We bereikten een diepgaande proteoomanalyse die vergelijkbaar is met huidige methoden die meer dan 100 keer zoveel uitgangsmateriaal vergen.

In hoofdstuk 5 hebben we de haalbaarheid getest van de combinatie van een kwantitatieve aanpak gebaseerd op dimethyl labelen met ZIC-cHILIC scheiding voor kwantitatieve proteomics. We hebben het potentiële probleem van een deuterium isotoopeffect besproken dat de kwantifikatie negatief zou kunnen beïnvloeden. Inderdaad kan de incorporatie van deuteriumatomen door middel van dimethyl labelen er voor zorgen dat verschillend gelabelde peptiden niet meer tegelijkertijd elueren tijdens ZIC-cHILIC scheidingen. We hebben aangetoond dat de keuze van een specifieke buffer pH het deuterium effect kan oplossen. Met pH 6.8 is hydrofiliciteit de belangrijkste bepaler voor peptideretentie wat er toe kan leiden dat waterstof en deuterium bevattende gelabelde peptiden met hun licht verschillende polariteit niet meer gelijktijdig elueren. Daarentegen kan een zure pH het hydrofiliciteitsverschil tussen gedeutereerde en niet-gedeutereerde peptiden verlagen vanwege een sterker effect van electrostatische interacties in de peptideretentie. De resultaten van de multidimensionale ZIC-cHILIC-RP aanpak tonen aan dat onze methode geschikt is voor nauwkeurige kwantitatieve proteoomanalyse en resulteert in de kwantificatie van duizenden eiwitten.

In hoofdstuk 6 hebben we onze hoge resolutie en erg gevoelige ZIC-cHILIC scheidingsmethode gecombineerd met een specifieke fosfopeptide verrijkingsmethode die gebaseerd is op Ti^{4+}-IMAC met het maximaliseren van de diepgang van analyse van cellulaire fosfoproteomen als doel. We hebben drie analytische strategieën ontworpen en systematisch geanalyseerd: i) enkel verrijking met Ti^{4+}-IMAC; ii) Ti^{4+}-IMAC verrijking gevolgd door HILIC fractionering; iii) een op SCX gebaseerd pre-fractionering, gevolgd door Ti^{4+}-IMAC verrijking en een volgende HILIC factioneringsstap. De evaluatie vond plaats gebruik makend van twee human kankercellijnen (HeLa en K562). De vergelijking van de drie verschillende werkstromen toonde aan dat een uitgebreide fractionering, zoals in strategie iii, nodig is voor een diepgaande analyse van het

fosfoproteoom. Daar staat dan weer wel een hoger vereisde hoeveelheid monster en langere MS analysetijd tegenover. Alles bij elkaar genomen konden we meer dan 20,000 en ongeveer 29,000 unieke fosfopeptiden identificeren van respectievelijk HeLa en K562 cellen, waarmee we een bruikbare fosfoproteoom informatiebron hebben gegenereerd voor de wetenschappelijke gemeenschap.

In hoofdstuk 7 hebben we in een gedetailleerd protocol beschreven hoe HILIC fractionering in een gevoelige "shotgun" proteomics strategie geïntegreerd kan worden. Het is tegenwoordig nog een uitdagende taak om het proteoom van kleine hoeveelheden cellen te analyseren. Een reeks aan speciale analytische methoden is vereist voor het behandelen van dit soort monsters. Vele van de technieken die normaal gesproken voor de analyse van grote hoeveelheden cellen worden gebruikt kunnen niet direct gebruikt worden voor kleine hoeveelheden cellen. De op zwitterionische HILIC gebaseerde procedures die in dit protocol worden beschreven betekenen een belangrijke stap voorwaarts voor gevoeligere proteoomanalyse. We bespraken enkele experimentele aspecten die belangrijk zijn voor een optimale monstervoorbereiding om een maximale opbrengst te genereren. We gebruikten voorbeelden van eiwitlysaten verkregen uit gecultiveerde cellijnen en cellen gesorteerd met flowcytometrie om de scheidingskracht, gevoeligheid en veelzijdigheid van de techniek aan te tonen.
We stelden ZIC-cHILIC voor als een alternatieve fractioneringsmethode voor ultragevoelige proteomics dat een ruime toepassing zal vinden in de analyse van specifieke cel populaties zoals bijvoorbeeld verkregen met laser capture microdissectie.

CURRICULUM VITAE

Scientific High School: "Principessa Elena", Campobasso, Italy
Date: July 2001, mark: 100/100

Master's degree: "La Sapienza", Rome University, Italy
Faculty: Pharmaceutical and Technological Chemistry
Date: 19th July 2007, *cum laude*

Doctorate: Utrecht University, the Netherlands
Faculty: Chemistry
Date: 8th February 2013

Work experience:

Master research project
Rome University, Italy
September 2006-July 2007
Subject: Technological Chemistry - Drug delivery systems

Junior research chemist
Merck Research Laboratories (IRBM), Pomezia, Italy
October 2007- January 2009
Subject: Preclinical Drug Metabolism and Pharmacokinetics - Analytical Chemistry

PhD fellowship
Biomolecular Mass Spectrometry and Proteomics Group, Utrecht University, the Netherlands
February 2009- November 2012
Subject: peptide separation strategies by multidimensional approaches; enabling technologies in proteomics workflows

Visiting PhD, Genome British Columbia-Proteomics Centre, Victoria University (BC), Canada
April-June 2012
Subject: absolute quantification by Multiple Reaction Monitoring

Brief summary of research experience:
During my undergraduate internship, I worked at the Technological Chemistry department focusing on drug delivery systems where I developed hydrogels-based systems used for the controlled release of anti-inflammatory drugs in oral and topical formulations. There, I gained a good practical experience in many analytical methodologies, including UV, IR and HPLC.
After M.Sc. graduation, I joined Merck working in the Preclinical Drug Metabolism and Pharmacokinetic group as young research chemist. My main research included the development of bioanalytical assays for qualitative/quantitative analyses of small molecules and peptides drug candidates and the evaluation of pharmacokinetic profiles after drug administration. There, I reached a good knowledge on liquid chromatography coupled to mass spectrometry (LC-MS), acquiring expertize in multiple-reaction-monitoring (MRM) methodologies.
During my PhD in the Biomolecular Mass Spectrometry and Proteomics Group, I continued my interest for mass spectrometry and technology, applying it into proteomics. My research

is focused on peptide separation strategies, phospho-enrichment tools and quantitative approaches implemented in proteomics workflows. This thesis described in detail the research performed under the supervision of Shabaz Mohammed and Albert Heck.
As a visiting PhD at the University of Victoria (Canada), I was focused on absolute quantification methods for viral protein analysis employing MRM.

Presentation at international conferences/symposia:

- 'HILIC day' symposium, Darmstadt, Germany, September 2012 (invited speaker)
- 60th ASMS conference on Mass Spectrometry, Vancouver, BC, May 2012 (poster)
- 59th ASMS conference on Mass Spectrometry, Denver, CO, June 2011 (oral)
- Proteomic Forum, Berlin, Germany, April 2011 (poster)
- Nordic Proteomic Network, Odense, Denmark, June 2010 (oral)

Annual meetings in the Netherlands:

- progress meeting of Netherlands Proteomics Centre (poster in 2011 and 2012)
- Bijvoet Simposium (poster in 2010, oral in 2011)
- NWO, Chemische Wetenschappen studiegroep, Analytische Chemie (oral in 2010)
- Nederlandse Vereniging voor Massa Spectrometrie (NVMS) (oral in 2012)

Academic activities:

Supervisor of a practical course in Analytical Chemistry for bachelor students at Utrecht University (February-May 2009, 2010)

Lecturer in the yearly course 'Advanced Biomolecular Mass Spectrometry and Proteomics', with the subject 'Peptide separation strategies'

7: Zhou H*, **Di Palma S***, Preisinger C, Peng M, Polat AN, Heck AJ, Mohammed S. Towards a comprehensive characterization of a human cancer cell phosphoproteome. Journal of Proteome Research. 2012 Nov 28.

6: **Di Palma S**, Mohammed S, Heck AJ. ZIC-cHILIC as a fractionation method for sensitive and powerful shotgun proteomics. Nature Protocols. 2012 Oct 25;7(11):2041-55.

5: **Di Palma S***, Hennrich ML*, Heck AJ, Mohammed S. Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. Journal of Proteomics. 2012 Jul 16;75(13):3791-813.

4: **Di Palma S**, Raijmakers R, Heck AJ, Mohammed S. Evaluation of the deuterium isotope effect in zwitterionic hydrophilic interaction liquid chromatography separations for implementation in a quantitative proteomic approach. Analytical Chemistry. 2011 Nov 1;83(21):8352-6.

3: **Di Palma S**, Stange D, van de Wetering M, Clevers H, Heck AJ, Mohammed S. Highly sensitive proteome analysis of FACS-sorted adult colon stem cells. Journal of Proteome Research. 2011 Aug 5;10(8):3814-9.

2: **Di Palma S**, Boersema PJ, Heck AJ, Mohammed S. Zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC and ZIC-cHILIC) provide high resolution separation and increase sensitivity in proteome analysis. Analytical Chemistry. 2011 May 1;83(9):3440-7.

1: Pucci V, **Di Palma S**, Alfieri A, Bonelli F, Monteagudo E. A novel strategy for reducing phospholipids-based matrix effect in LC-ESI-MS bioanalysis by means of HybridSPE. Journal of Pharmaceutical and Biomedical Analysis. 2009 Dec 5;50(5):867-71.

IN PREPARATION

Zoumaro-Djayoon A*, Di Palma S*, Peng M*, Post H, Preisinger C, Munoz J, Heck AJ. Sequence analysis of phosphotyrosine peptides enriched by means of immuno-affinity precipitation and metal-based affinity chromatography. Manuscript in preparation.

Simigdala N, Di Palma S, Goetze S, Reim G, Heck AJ, Mohammed S, Hafen E, Basler K, Wollscheid B, Brunner E. The *Drosophila* wing disc proteotranscriptome at spatial resolution reveals novel compartment-specific markers. Manuscript in preparation.

Di Palma S, Benevento M, Heck AJ. et al. MRM as an accurate methodology to quantitatively elucidate the viral protein composition in presence of endogenous proteases. Manuscript in preparation.

* Joined first author

ACKNOWLEDGEMENTS

If I think back to the starting point, I realize that time has passed too fast. I still remember that day when I packed my 'life' in a few suitcases and I said goodbye to my friends and my beautiful Rome. I left Italy on a sunny and windy day, and I landed in a misty and mysterious country. I also remember that weird mixture of feelings: excitement and fear, serenity and turmoil, curiosity and nostalgia, altogether. That was the beginning of a new 'adventure' in the Netherlands, in a little city, where I knew nobody. And what an adventure! I spent the most incredible 4 years of my life.

I have to admit that it couldn't have been as wonderful without some special people that I met in Utrecht and supported me during this journey. And I like to end this thesis thanking all those people, my friends, colleagues and family, for encouraging and giving me the energy and enthusiasm to reach the 'final destination'.

Dear Albert, I want to thank you for believing in my capabilities, trusting me, giving me the freedom to deal with my projects and making me every day more independent and self-confident. I am impressed about your multi-tasking busy life. You are capable to come to work after travelling the whole night long from the other side of the world and, still, have the strength and curiosity to discuss with me about progresses in my research. I really appreciated your supportive and compassionate attitude, which always gave me enthusiasm and stimuli to do 'more' and 'better'. And I thank you immensely for allowing me to visit Christoph's labs in Canada and explore the stunning Vancouver Island. I will never forget that experience, such as that it even deserved the cover of my thesis! Albert, I am happy that our professional relationship has grown to such a mature extent that we can easily converse about everything in a direct and constructive way. It has been a truly pleasure to work with you. I am sure we will meet often in the future!

Shabaz, our similar 'analytical mind' perfectly matched and helped us to understand each other at a glance, to design experiments in 3 minutes and to make every project successful in short time. Thanks for everything, I got to learn a lot from you! I wish you a big good luck with your future career in Oxford. And try not to hire crazy female italian Ph.D. students, it could be quite demanding!

Corine, I've never met such an organized and efficient person as you… and you are always willing to help everyone with a big smile. This is the best attitude in life. I am really grateful to you for being so available with me. I actually can't remember a single day when I didn't come to your office or wrote you an e-mail asking for something! Please, come with me wherever I will end up or I will be lost!!! Thanks!

Thanks to my 2 beautiful paranymphs, Adja and Soenita! Adja, you are a good friend of mine, not only a colleague. We went through everything together, always trying to smile… though sometimes there was nothing to laugh about! But we made it… I enjoyed a lot our time together, chatting during lunch, in the lab, in the office, during symposia, everywhere! Just a favor: could I borrow your Timeo for a few days, pleeeeeease!? Soenita, from the first day I joined the group, I was so glad to have you. You were the first one showing me in the lab how to troubleshoot with the LC-MS and many other things, meanwhile feeding me with a lot of candies and delicious cakes. You are such a caring and sweet person, I will miss you a lot and I will order your cakes even from Zurich!

Thanks to the 'real strength' of the lab: Arjan and Harm. How could I manage without you, guys? Arjan, I can't express my feelings here, in front of everyone… in my hearth you will always be my 'primo' secondo! Harmito, you are so funny… I can't stop laughing when I work and talk to you… but you can't stop talking at all!!! Please, once in a while: mond dicht!

My dear Natalie, my list of thanking for you could be very long. Shall I just mention all the time

you spent proofreading my drafts? What about watering all my plants in our office? I enjoyed very much talking to you, asking your opinion and advices… and you know how much I admire you! I will definitely miss you.
Reinout, I could recognize your steps in the corridor within hundreds of people. Thanks for helping me in a myriad of occasions. Though you were super-busy, you always said 'yes' to my requests of help… I shouldn't say, but you are one of the few people I totally trust, personally and professionally.
Marco, my German friend, you can be so quite and fast, as precise and omnipresent! We can have such nice conversations, about science or other topics. I missed you this year running all around the lab at the speed of the light. I am happy we are still in touch!
Javi, why did you leave? I am still looking for you in your office. I needed your help to design in an 'extraordinary fashion' my thesis… but you were not around! I will never forgive you for that… My crazy Spanish friend, I still laugh when I think about you cycling home half-drunk after long parties, though I don't laugh when thinking of the consequences!
A big THANKS to all my wonderful colleagues/friends of the Biomolecular Mass Spectrometry and Proteomics Group, current and previous. You are so many and I am terrified to forget somebody. I already apologize if it will happen.
First of all, thanks to my favorite (obviously) Italian 'clan': Ele, Sara, Salvo, Marco, Fabio, Gianluca and Piero. It was so fun to speak Italian while the others were trying to catch our topic! Probably at the end they will understand something, we should be more careful!
Then, the German group. Christian F., thanks for your help with the MS and for sharing 'epical hikes' and glorious moments outside the lab. Christian Sr., for me you are still half-Austrian, half-German. Why did I put you with the German people, sorry! Kristina and Andreas, and the lovely dinners at your place! My 'brother' Tommasino, it was weird without you in the office. I missed you, your desk and our long chat! BTW, my plants survived… I think Natalie and I got a good training from you. Then, Charlotte, Hanka, Nikolai, Nadine, etc.
Of course, the Canadian crazy couple, Sarah and Glen. I was so enthusiastic to come visit you in Edmonton that I couldn't keep myself driving slowly… I hate the Canadian cops! We still drink some last Pils to remember how awesome it was having you in Utrecht!
Then, the British girls, Becky and Jess. Becky, it is not the same here without you! The French people, Nikolas and Lucrèce. My darling, why didn't you join the group earlier? I like you, your macarons and our coffee break so much! The Iberian girls, Alba and Ana. The 'Chinese-Asian mafia': Houjiang (ciao bello! You are so social, especially with pretty girls!), Teck, Hongtao, Vincent, Donna and Mao. The Turkish girls, sweet Ayse and Basak. My favorite Kenyan 'Latin dancer', Benji.
And obviously, thanks to all the Dutchies. Mirjam, you are always available and kind to me, thanks for all your favors! Monique, you have such a passion and enthusiasm when you explain things to your students and talk about your work! Keep it, it's such a good example! Maarten, I like your non-talkative way of talking, it is all quite clear to me! Bas and Henk, you are both becoming Italian, too much espresso! Erik, I will organize your next trip since you enjoyed so much the one in Victoria; Martje, you are the NPC heart; Patrick, thanks for the Dutch translations; Geert, luckily I had you fixing my troubles with computers! And still, Arjen, Pepijn, Renske vG. and P., Esther, Joost, Kees, Onno, Jeffrey, Peter, etc.
A special thanks to some 'international' previous colleagues and friends: Nadia, Thin Thin, Joana, Sharon, Poupak, Martina. I have to say, girls, you are all real 'talkative' people, even more than me! It was too quiet at the end without you… and no gossip around! I missed you!
I have to say that the most exciting time I spent during my PhD was in Victoria. No offence for the Netherlands… but Canada is just a breathtaking place to live in, offering wonderful views

of mountains, rocks, oceans, lakes…! Thank you, Christoph! I enjoyed so much working in your labs with your group. I felt so comfortable and at home since the first moment, perhaps because I was sitting next to lovely Ute! Thanks to all the great people that helped me there: Melissa, Angela, Derek, Darryl, etc. I was also so lucky to meet in Victoria some special friends, super 'crazy' and hyperactive, that were travelling all around with me. Especially Nikita, Addi and Karlito (of course their real names don't matter!). Nikita, I even chose one of the pictures of our adventurous 'whale watching' for my cover! How proud can you be!

Now, it is the moment to thank the wonderful people that did not 'scientifically' contribute to this thesis, but without their support, friendship, love, patience and happiness I couldn't go through my 'dark moments'. I am talking about my best girl-friends with whom I share my daily life! Claudette, I surely spent most of my time in Utrecht with you. I got to know you deeply and to appreciate your 'Porsche' taste, which luckily corresponds most of the time with mine! You are adorable and capricious… you missed me so much while I was in Canada that you were calling me during the night 'cause you couldn't sleep! I did miss you too, and I will miss you even more when I will move to Switzerland… how can I be sharp without you?! Paolina, chiquitita rabiosa, I can confirm that you are the most 'légère' of my friends, and I can finally explain what I mean! It is so easy and nice to chat, have fun, talk about serious stuff and enjoy time with you in a very spontaneous way, without complications! Clearly, I love the way you are! Francis, luckily I had you in Utrecht organizing amazing Italian dinners and reminding me how beautiful Italy is (especially Tuscany)… otherwise I will probably be too Dutch now! I will come to visit you at your future agriturismo toscano, since I know you will not come to Switzerland… it is not in Europe, right? Aninha, my super dancy Brasilian friend! I have missed so much you, the picnic at Wilhelmina Park with Prosecco and your cool parties with the rhythm of samba! Come back! Jo, we spent great time together, dancing, chatting, sometimes crying, and then laughing again! When you will come to visit me, please, bring your amazing bacalhau! A special thanks also to my boy-friends, Peng, Shusil and RK, that I had the pleasure to meet at the very beginning of my time in Utrecht. We had amazing tours in the Netherlands, it was always so fun!

I can't forget to thank all the people from the Merck-IRBM group, where all started! Fabio, grande Fabio, I am immensely grateful to you for teaching me the fundamentals of mass spectrometry, for infusing your passion for research into me and, more importantly, for showing me how respect and professionality are pillars in a working environment. I consider you as my father in science! Enzino, you had the capability to teach me how to be quick, efficient and precise, altogether, like if it was a game or a bet! And I am so thankful to you for inspiring me with your determination and stubborness! I started my PhD just because I wanted to be like you! My pretty women, Marinella, Annarella and Maria, it couldn't have been so extraordinary without you. For me every day was amazing 'cause we were working together and laughing! I miss so much those days and our long long chats... Edith and Enzo S., you are such examples of 'good' and 'nice' leaders, a mixture of authority and compassion. I should learn from you! Finally, an enormous thanks to the 'crazy car pooling' clan: Simona, Michele, Francesca, Caterina and (kind of) Bufali! How to have a better early start and late end of the day, guys?! I was a 'young sister' for you, and it was fantastic to have your protection and fraternal advices... However, as all young sisters do, I didn't listen to you and also here my colleagues think that I am out of head!!!!!!

I want to thank enormously my Italian friends, Iole, Fede, Daniela, Ale & Ale, Vale & Andrea, Dani & Gianluca, Elly & Stefano, Chiara, Miri, and many others... always very present, overenthusiastic, supportive and extremely curious about my life abroad. And it was funny when, at the end of my stories, I always heard the fatidical question: "Serenella, when are you coming

back to Italy? You will come back one day, right?". My dear friends, I will always come back to Italy, because I want to see you and because I love my beautiful country... and wherever I will end up living in the future, you will always be the pillars of my life.
Now I have two big families to thank: the Italian and the Dutch one!
Il gatto e la volpe, il sordo, il muto e la chef! Questo é il quadretto perfetto della mia famiglia meravigliosa... non oso dare alcun attributo, ma a ciascuno il suo! Mamma, Papá, Danny, Paoletto e nonna Tittina, vi porterei con me in capo al mondo per avervi sempre vicini ed a portata di mano... Peró, vi chiedo tre favori: non gridate, imparate l'inglese ed evitate di chiamare il mio Pauli POOOL! Terry, per fortuna sei arrivata tu a mitigare il caos! Vi adoro!
Lieve allemaal, lieve Emma en Peter, Dick, Eduard, Karin en Otto, Jaap and Cilia, Anthoinet en Arnold (en mijn 5 mooie neefjes en nichtjes: Henk-Jan, Peter, Kirsten, Rob en Inge). Ik ben blij deel uit te maken van jullie grote familie, ook ben ik blij dat jullie familie een stuk rustiger is dan de mijne... Dit zorgt voor een goede balans voor mijn oren!! Ik vind het erg fijn om tijd met jullie te spenderen en dan met name tijdens het familieweekend.
My dear Pauli, I am still not sure if I should thank you for being a special colleague, a great officemate, my best friend, or a wondeful life partner. What I know for sure is that without you, and YOU with each definition, these 4 years of PhD wouldn't have the same meaning, value and sparkle. Paultje, you are the only one capable to understand, share and cope with my acme of happiness and deepness of frustration. You took care of me daily in the lab from my first day till you moved to Munich, explaining me every single detail with passion, calmness and devotion. I still wonder why you did all that...??!! Then, even far apart, you were my everywhere-, every day-, every hour- and always available-supervisor, discussing with me doubts, disappointments, encouraging me to overcome hurdles, and enjoying with me success and satisfaction. I couldn't be luckier to meet a person as you in my life. After sharing incredible and extraordinary moments, so much that I don't feel to mention those magic memories here, I look forward to move with you to Zurich to start another adventure together, in close, close contact to you. My Pauli, my sweetie, my joy...

Serena

"There are and have been and will be an infinite number of things on earth. Individuals all different, all wanting different things, all knowing different things, all loving different things, all looking different. Everything that has been on earth has been different from any other thing. That is what I love: the differentness, the uniqueness of all things and the importance of life... I see something that seems wonderful; I see the divineness in ordinary things."

Diane Arbus. November 28, 1939