Proteomics enabled Vaccinology
Probing Antigen and Epitope Repertoires

Geert P.M. Mommen

ISBN 978-90-8891-814-8

The research described in this thesis was performed at the department of Formulation and Analytical Research of the Institute of Translational Vaccinology (Intravacc) (formerly the Netherlands Vaccine Institute), Bilthoven, The Netherlands and the Biomolecular Mass Spectrometry & Proteomics Group, Utrecht University, Utrecht, the Netherlands

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

Printing of this thesis was financially supported by the Institute of Translational Vaccinology (Intravacc)

Cover photo by Joost Uittenbogaard

Layout and artwork by Cas Houwen

Printed by Proefschriftmaken.nl || Uitgeverij BOXPress

Proteomics enabled Vaccinology
Probing Antigen and Epitope Repertoires

Vaccinologie mogelijk gemaakt door Proteomics
Onderzoeken van Antigen en Epitope Repertoires
(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 maandag 17 maart 2014 des ochtends te 10.30 uur

door

Geert Petrus Maria Mommen

geboren op 23 augustus 1981, te Venray

Promotor	Prof. dr. A.J.R. Heck
Co-promotor	Dr. ir. A.J.P.M. de Jong

Voor Cindy en Sanne

CHAPTER 01

General introduction

INTRODUCTION

Proteomics technologies have evolved rapidly over the last decade finding their applications in all kind of research areas. Therefore, it is not surprising that proteomics-based technologies also have been used to study the immune system. This particular application of proteomics, also referred to as immunoproteomics, is a relatively new concept of analytical science that links the research and scientific knowledge of immunology and proteomics. Although the term immunoproteomics first appeared in literature in 2001 [1], multidisciplinary studies have been applied for longer times and provided a better understanding of the cellular functioning of the immune system [2]. While the term proteomics describes the study of the complete set of proteins expressed by the entire genome [3], immunoproteomics focuses on the subset of proteins that is involved in immune responses (Figure 1) [4]. The immune system of higher organisms is composed of a variety of cellular sensors that have the ability to discriminate between self, altered self (tumor, autoimmunity) and nonself (antigenic substances from viruses, bacteria or eukaryotic parasites) species. These sensors serve as innate alarm signals, indicating the onset of infections or malignant transformations. After recognition, adaptive immune mechanisms are activated aiming at the clearance of the infected cells or pathogen and preventing further spread of the disease. In general, the substances (*e.g.* proteins, polysaccharides) that are recognized by immune receptors and elicit a specific immune response are called antigens.

In mammalian systems, there are two types of defense mechanisms, innate and adaptive, which differ in the receptors used to recognize pathogens [5]. The innate immune system is evolutionary conserved and provides an immediate, but nonspecific response to pathogens. A range of Pattern Recognition Receptors (PRRs) recognize conserved Pathogen-Associated Molecular Patterns (PAMP). PRRs trigger inflammatory responses and anti-pathogen and cell death mechanisms [6]. The innate immune system provides a first line of defense against microorganisms and facilitates the slower, but specific adaptive immune system. The adaptive immune system can be dived into a T cell-mediated cellular component and a B cell-mediated humoral component [5, 7]. T cells do not recognize antigen proteins directly, but rather the peptide fragments that are processed intracellular from self and nonself proteins. These peptides are presented by Major Histocompatibility Complex (MHC) molecules at the surface of all nucleated cells or special antigen processing cells (Figure 2). MHC-peptide complexes represent a comprehensive and balanced summary of the protein content and immune status of the cell. Recognition of MHC-peptide complexes on diseased cells by T cells results in a cytotoxic response and cell death. In contrast, B cells can recognize and bind to a specific pathogen directly by their surface-associated antibodies. After binding, the antigen-antibody complex is internalized, processed into peptides and displayed on MHC molecules for T cell response. Upon recognition, B cells start to produce antibodies that are able to recognize and neutralize pathogens, either by triggering an immune response like complement induced lysis of bacteria, or directly by inhibiting the function of particular proteins.

The *in vivo* recognition of antigens on diseased cells or pathogens by the immune system is highly specific and sensitive [4]. These antigens can be studied individually using well-established biochemical methods, such as immunoassays, gel-based methods or western blotting. The last decades, mass spectrometry (MS) has become an important analytical tool that allows for the high-throughput identification and quantification of proteins and peptides [8]. Figure 1 depicts

several examples where MS-based proteomics provides valuable insights in the immune system under healthy or stress conditions. These immunoproteomics applications include, for example, (*i*) the identification of T cell peptide antigens [9-12], (*ii*) structural analysis of B cell epitopes [13], (*iii*) quantitative profiling of the host-cell proteome, in response to pathogen invasion [14, 15], and (*iv*) the discovery of antigenic proteins that simulate the humoral immune system [2, 16].

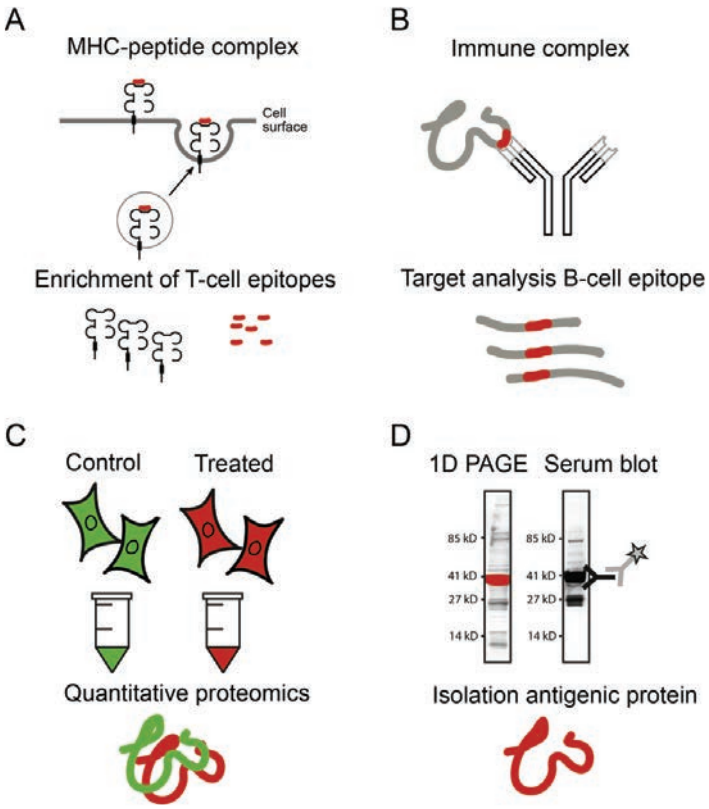


FIGURE 1. EXAMPLES OF IMMUNOPROTEOMICS APPLICATIONS. (A) Mass spectrometry (MS)-based sequencing of Major Histocompatibility Complex (MHC)-presented peptides provides a general insight into the antigen processing pathway and binding specificity of peptides to MHC molecules, as well as the identification of individual T-cell antigens that are associated with infectious diseases, cancer or transplant rejection. (B) Target MS approaches (*e.g.* cross-link chemistry, hydrogen-deuterium exchange or limited proteolysis) enable the structural elucidation of B-cell epitopes, *i.e.* the peptide fragment of a protein that interacts with an antibody. The identity of these B-cell epitopes is important for vaccine design, structural biology (protein-protein interactions) and monoclonal antibody therapy. (C) Quantitative MS-based analysis of the host-cell proteome in response to pathogenic invasion provides a global picture of the immune response, as early signatures of molecular pathways involved in human diseases. (D) The discovery of vaccine candidates or biomarkers by the identification of antigenic proteins which simulate the humoral immune system using the sera of recovering or vaccinated individuals.

The immune response is a resultant of complex biological processes that proceeds via specialized cells, involving different subcellular compartments or highly specific antigen-antibody interactions. The extreme heterogeneity and large dynamic range of protein abundances in a biological system may average out (sub)cellular responses and dilute the signal related to the biological question. Target enrichment methods (*e.g.* affinity chromatography, immunoprecipitation) reduce background contamination and enhance the dynamic range of the analysis, but such methods are often accompanied with significant loss of analytes [17]. To address these challenges, continuous improvements throughout all sections of the proteomics workflow are needed [18]. This chapter provides a general introduction in MS-based (quantitative) proteomics and highlights some relevant analytical approaches that are aimed at improving the selectivity and sensitivity in MS detection.

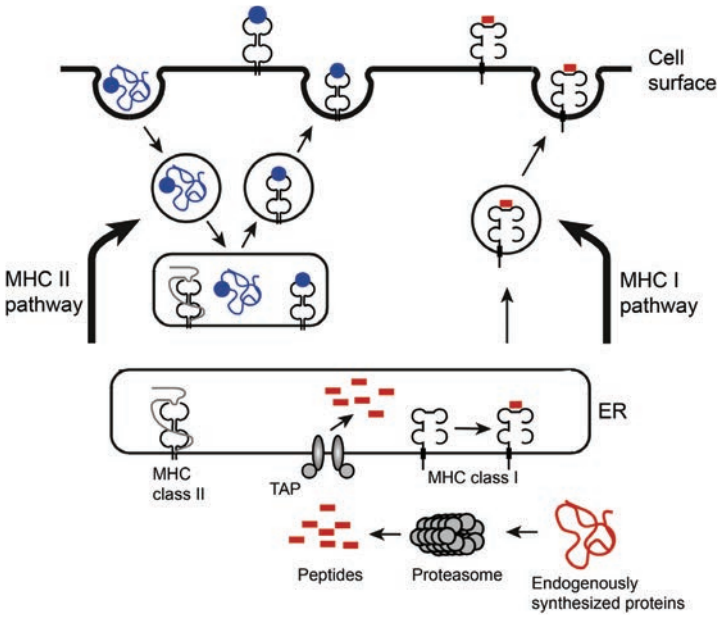


FIGURE 2. MAJOR HISTOCOMPATIBILITY COMPLEX (MHC) PATHWAY. The two different classes of MHC molecules, class I (right) and class II (left), differ in intracellular peptide-presenting pathway. MHC class I molecules are expressed by all nucleated cells. The MHC-class I-associated peptides originate from endogenously synthesized self-proteins and foreign proteins in case of infection. These proteins are cleaved by a number of different proteases and the resulting peptides are translocated into the endoplasmic reticulum (ER) by transporter associated with antigen processing (TAP) protein. There they are subjected to N-terminal trimming by aminopeptidases to a size that is suitable for loading onto MHC class I molecules (typically 8-11 amino acid residues) and transported to the cell surface for immune surveillance. The MHC class II pathway operates in antigen presenting cells (APCs), such as dendritic cells, macrophages and B lymphocytes. MHC class II molecules display peptide fragments (typically 8-25 amino acid residues) from extracellular self- or foreign proteins that have been captured by cell surface receptors, internalized and degraded in early or late endosomal compartments. Figure adapted from [99].

MASS SPECTROMETRY-BASED PROTEOMICS

Hunt *et al.* [19] were among the first who recognized the need for sensitivity in MS-based (immuno) proteomics. Already decades ago, they reported the first application of MS in the analysis of *in vivo*-generated peptides presented by major histocompatibility complex (MHC) molecules. To identify the femtomolar amounts of peptides isolated from MHC molecules, they developed the revolutionary sensitive microcapillary-column liquid chromatography-mass spectrometry (LC-MS) method. This so-called nanoflow LC technology is nowadays one of the most prevalent techniques in proteomics.

Figure 3A depicts the general workflow of MS-based proteomics. The first step of any proteomics experiment is the extraction of proteins from cells, tissues or body fluids. Although a wide variety of protocols is available [20, 21], samples are typically treated with denaturing agents (*e.g.* sodium dodecyl sulfate (SDS), urea, guanidine hydrochloride) to enhance the solubilization of cells, tissue and proteins [22, 23]. A clear disadvantage of the use of irreversible denaturing agents like SDS is their inhibiting effects on enzymatic digestion or the loss of binding to monoclonal antibodies. In addition, detergents are generally not compatible with MS analysis because they often dominate the MS spectrum due to their high abundances, resulting in a dramatic suppression of the analyte signal.

The second step in a typical proteomics workflow is the enzymatic digestion of proteins into peptides, preferably using a sequence-specific enzyme that cleaves proteins in a controlled and predictable way. The analysis of peptides is less challenging than proteins, because they are easy to separate by LC and the amino acid sequence information can easily be retrieved by MS analyses. Trypsin is most commonly used for proteolytic digestion, which cleaves proteins efficiently and specifically at the C-terminal side of the basic amino acids lysine and arginine [24]. Trypsin digestion yields primarily peptides with at least two basic groups, one being the C-terminal amino acid side chain and one being the N-terminal amino group. Both sites are protonated under acidic LC conditions, which is ideally for the successful identification of the peptide sequence by tandem MS.

Prior to MS analysis, peptide mixtures are usually separated by reversed phase LC using a gradient of a water-miscible organic solvent. ElectroSpray Ionization (ESI) and Matrix-Assisted Laser Desorption Ionization (MALDI) are the most suitable techniques to generate gaseous ions from peptides. ESI enables the direct coupling of the column effluent with the mass spectrometer and is therefore more straightforward, quicker and nowadays most frequently used [25, 26]. Peptide mixtures are often of such complexity that many peptides co-elute from the column and simultaneously reach the MS. Additional dimensions in peptide separation are optional to identify as many peptides as possible by MS. Today, a wide variety of column materials is available, allowing the researcher to design the appropriate multidimensional LC mode for advanced peptide separations [27]. The most common method couples Strong Cation eXchange (SCX) as a first dimensional separation with reversed phase (RP) chromatography as the second dimensional separation, a method referred to as multidimensional protein identification technology (MudPIT) [28]. The second dimensional separation of peptides is almost exclusively performed by RP chromatography because of the excellent resolving power, high efficiency and reproducibility, and because the elution solvents are highly compatible with ESI-MS [29].

After transferring the peptide ions into the gas phase by ESI, they are separated based on their mass-to-charge (*m/z*) ratio by the mass spectrometer in an MS scan. The instrument software automatically selects, with user-specified criteria, a predefined number of these *m/z* values as precursor ions for subsequent tandem mass spectrometry (MS/MS). Tandem MS provides direct information of the peptide sequence and involves the fragmentation of peptides into smaller, sequence-informative ions, which are recorded in an MS/MS spectrum. The MS and MS/MS data are used as input for bioinformatics tools to elucidate the corresponding peptide sequence. Database search engines, such as SEQUEST and Mascot, compare the experimental MS/MS spectra to theoretical MS/MS spectra, generated *in silico* from proteome databases [30, 31]. To control the number of incorrect identifications (False Discovery Rate (FDR)), the obtained data set is statistically validated through decoy search strategies in which the MS/MS spectra are searched against a random database containing nonexisting protein sequences [32].

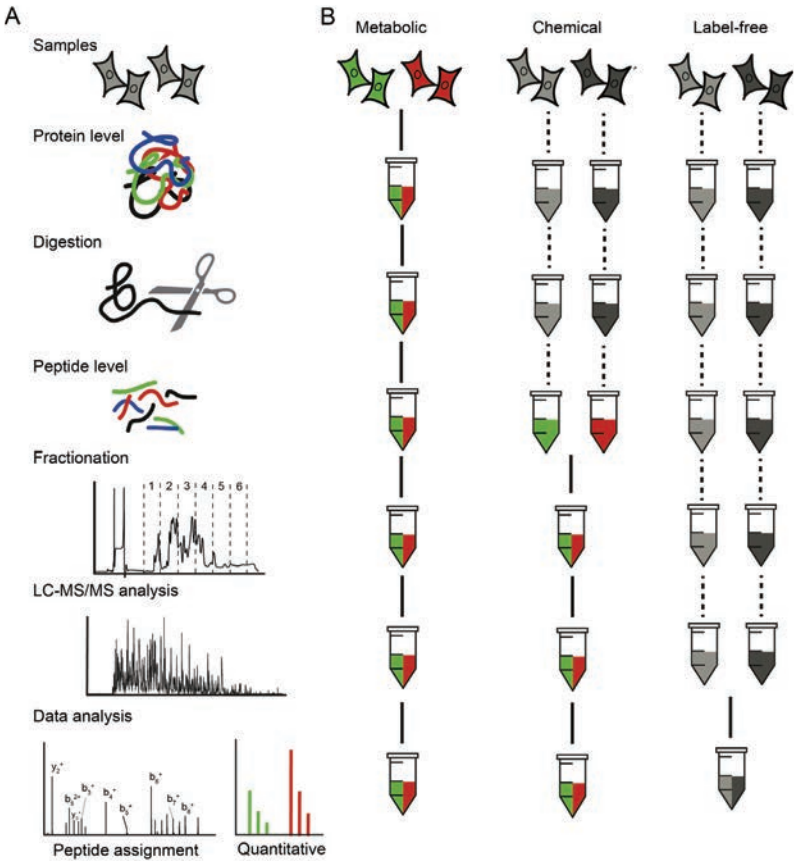


FIGURE 3. WORKFLOWS IN BOTTOM-UP (A) AND QUANTITATIVE PROTEOMICS (B). For metabolic and chemical labeling, samples are labeled with stable isotopes at different levels in the proteomic workflow, as indicated by the red and green color coding. Upon isotope labeling, samples are combined as highlighted by the horizontal lines. The dashed line represents the steps where samples are treated separately. In label-free quantification, samples are processed individually and combined at the data analysis level. Figure adapted from [100].

MULTIDIMENSIONAL PEPTIDE SEPARATION

Proteome samples may contain thousands of different proteins in concentration abundances ranging five orders of magnitude [33]. The complexity of proteome samples will further increase upon digestion since each protein yields multiple peptide fragments. To reduce sample complexity and increase the depth of proteome coverage, samples are usually fractionated, at the cellular level (membrane, cytosol, subcellular compartments), at the protein level (*e.g.* after SDS PAGE separation) or the peptide level [34, 35]. Typically, two or more subsequent LC separation methods are used to increases the overall peptide separation power, such that fewer peptides reach the MS simultaneously. In most 2D approaches, reversed-phase (RP) LC is used as second dimensional separation in combination with various first-dimensional separation methods, such as strong cation exchange (SCX), hydrophilic interaction chromatography (HILIC) and reversed phase LC with subsequent high pH and low pH elution [36-38].

Although multidimensional peptide separation methods are already standard in most proteomics workflows, these methods are not always applicable to small sample quantities [17]. One of the major limitations of multidimensional separation approaches is the high risk of sample loss, sample contamination and sample dilution as a result of the additional sample handling steps and exposure to LC materials. Significant advances have been made in the miniaturization of chromatographic systems to reduce nonspecific adsorption, improve chromatographic resolving power and increase MS sensitivity [39]. Sample handling can be further reduced by online multidimensional setups because the sample transfer between consecutives separation dimensions is realized in a fully automated fashion using switching valves and trapping columns. However, the disadvantages of online formats are the inability to re-analyze individual sample fractions and the limited flexibility in optimizing the chromatography in each dimension due to the more stringent solvent requirements [40]. Moreover, chromatographic systems allow for sample preconcentration or peptide enrichment, which is often a prerequisite for genuine biological samples that contain high concentrations of detergents and salts.

The most frequently used 2D separation method today combines SCX in the first separation dimension and RP chromatography in the second dimension. These two techniques are moderate orthogonal since the separation principle of SCX is primarily based on charge, whereas RP separation is based on hydrophobicity. The sample is first fractionated by SCX and individual fractions are collected for subsequent RP analysis. Since the displacement of peptides in SCX is accomplished using an MS incompatible salt gradient, fractions need to be desalted prior to MS analysis. SCX is profound in the removal of detergent. These compounds, typically used in upstream cell lysis or protein denaturing protocols, co-elute with peptides in reversed phase LC and seriously interfere with MS detection. Since detergents lack retention on SCX chromatography, due to the absence of positively charged functional groups, they end up in the void volume, a region that is usually low in peptide abundances.

The peptides obtained from enzymatic digestion (*e.g.* with trypsin) have rather similar properties under SCX chromatography because the majority of these peptides have two basic sites, one at lysine or arginine side chain (C-terminus) and one at the free amine group at the N-terminus. At a low pH, these peptides will possess two positive charges, meaning that they largely co-elute

under SCX [37]. The use of reversed phase (RP) chromatography in the first separation dimension is an attractive alternative because of the superior resolving power of RP over SCX. However, the combination of RP as both first and second dimension lacks orthogonality because of the similar hydrophobic-based retention mechanisms in both dimensions. There are several factors which can modulate the selectivity in peptide retention in RP chromatography, such as the type of stationary phase or ion paring reagents, but the most pronounced effect can be achieved by changing the pH of the mobile phase. Gilar *et al.* [38] designed an RP-RP system for complex peptide mixtures using pH 10 in the first dimension and pH 2.6 in the second dimension, a 2D approach that combines the high separation power of RP with moderate orthogonality.

A relative new 2D separation approach in proteomics combines Hydrophilic Interaction Chromatography (HILIC) and RP chromatography. HILIC provides several advantages in comparison to more established techniques, such as high resolving power and excellent orthogonality with RP due to the opposite selectivity in polarity [41]. Recently, di Palma *et al.* [17] designed and optimized a highly sensitive 2D strategy, combining first dimensional HILIC with second dimensional RP chromatography, and identified over 1500 proteins from a limited number of Fluorescence-activated Cell Sorting (FASC)-sorted colon stem cell.

In addition to the wide variety of stationary phases and online or off-line configurations available, column (internal) dimensions are also essential parameters that influence the separation efficiency [27, 42, 43]. Moreover, the second dimension RP column should have a small internal diameter (i.d.) to ensure sensitivity in MS detection. Typically, fused silica capillary columns of 50, 75 or 100 µm i.d. packed with RP stationary phase are used. Small i.d. columns operate at 100-500 nL/min flow-rates and provide high separation efficiency and sensitivity in ESI-MS. The analytical requirements of nanoflow LC include dedicated pumps or switching valves and zero dead-volume connections. These strict requirements have precluded the routine use of microcapillary columns with < 50 µm i.d.'s, although several reports describe the design and successful use of ultra-small formats (10 – 25 µm i.d.) in ultrasensitive proteomics applications [44-47].

MASS SPECTROMETRY

Developments in mass spectrometry have enabled a technical breakthrough in the high throughput identification of proteins and peptides. To generate detectable ions from proteins and peptides, the most suitable strategies are electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) because they are able to transfer the analytes into the gas phase without extensive degradation [25, 26]. ESI is nowadays the most predominant ionization technique used in MS-based analysis of peptides or proteins. In principle, a high voltage is applied to an electrically conductive needle placed at the end of the column, which causes the creation of electrically charged spray of fine droplets, followed by subsequent desolvation to form gas-phase analyte ions [48]. The produced ions are transferred into the mass spectrometer where they are separated based on their *m/z*. There are quite distinct types of mass analyzers used in proteomics, each different in design and performance [33, 49-52]. Orbitrap and Fourier transform ion cyclotron mass analyzers separate ions based on their *m/z* resonance frequency, quadrupoles and ion traps use *m/z* stability and time-of-flight (TOF) analyzers use flight time. More recently, hybrid instrument have been designed that combine, the strengths of different mass analyzers. Table 1 summarizes the comparative performance and applications of the currently available instruments most commonly used in proteomics.

TABLE 1. Brief description of the principles and applications of the mass filters generally used in the mass spectrometers employed in proteomics research. Table adapted from [33, 49-52].

Mass spectrometers	Applications and Characteristics
ION TRAP. The (three-dimensional) ion trap captures ions, fragments ions of a particular *m/z* and analyses the subsequent fragment ions to generate tandem mass spectra.	**ION TRAPS.** High throughput peptide identification, multistage MS (MSn), PTM identification. <100 ppm accuracy, femtomole sensitivity, fast, 10^4 dynamic range, CID/ETD.
TRIPLE QUADRUPOLE. Quadrupole mass spectrometers select ions by time-varying electric fields between four rods, which permit a stable trajectory only for ions of a particular *m/z*. The first section of the triple quadrupole selects ions of a particular *m/z*, which are fragmented in the collision cell (second quadruple), while fragment ions are analyzed by the third quadrupole, creating a tandem mass spectrum.	**TRIPLE QUADRUPOLES.** Quantification in selective reaction monitoring (SRM), PTM characterization in neutral loss scanning mode. <100 ppm accuracy, femtomole sensitivity, fast, 10^4 dynamic range, CID.
QUADRUPOLE TIME-OF-FLIGHT (Q-TOF). The Q-TOF instrument combines a triple quadrupole with a reflector time-of-flight section for measuring the *m/z* values of the ions. In the reflector TOF section, ions are accelerated to high kinetic energy and are separated along a flight tube as a result of their different velocities.	**QUADRUPOLE TIME-OF-FLIGHT (Q-TOF).** Peptide identification, protein characterization, PTM identification. <5 ppm accuracy, femtomole sensitivity, fast scanning, 10^4 dynamic range, CID.
LINEAR ION TRAP-ORBITRAP. A hybrid analyzer consisting of a combination of a linear ion trap and an Orbitrap mass analyzer. The Orbitrap is a type of Fourier Transform (FT) ion trap mass spectrometer where ions oscillate along and around a central spindle. The *m/z* of the trapped ions are proportional to the frequency of the oscillations, which are Fourier transformed to yield a high precision mass spectrum.	**LINEAR ION TRAP-ORBITRAP.** Peptide identification, protein characterization, multistage MS (MSn), PTM identification. <2 ppm accuracy, femtomole sensitivity, slow scanning, 10^3 dynamic range, CID/ETD.
FOURIER TRANSFORM ION CYCLOTRON RESONANCE MASS SPECTROMETER (FT-ICR). The FT-ICR also traps the ions, but does so with help of strong magnetic fields. The FT-ICR is often combined with a linear ion trap for efficient isolation.	**FOURIER TRANSFORM ION CYCLOTRON RESONANCE MASS SPECTROMETER (FT-ICR).** Peptide identification, protein characterization, multistage MS (MSn), PTM identification. <2 ppm accuracy., femtomole, sensitivity, slow scanning, 10^3 dynamic range, CID/ETD

In bottom-up proteomics, a mass spectrometer ideally provides sufficient mass accuracy to assign the unique elemental composition, *i.e.* the total mass of the amino acid sequence composition, to the measured peptide ion. High mass accuracy greatly reduces the possible number of amino acid sequences responsible for an observed signal. The amino acid composition of relative small peptides can be determined with a 1 part-per-million (ppm) mass accuracy, a level of accuracy that is almost routine for current mass analyzers using internal or external calibration [53-55]. Tandem mass spectrometry (MS/MS) provides the additional information needed to determine the peptide amino acid sequence and to locate possible modifications (*e.g.* post-translational modifications). In an MS/MS scan, a precursor ion is isolated and fragmented into smaller sequence-informative fragments, which are subsequently recorded in a MS/MS spectrum. In the most ideal situation, one MS/MS spectrum is acquired for each peptide within a complex peptide mixture. However, this is not yet reality because the sequencing of all peptides is hampered by the finite sequencing speed of the MS and the dynamic range of measurement. Nonetheless, recent instruments which provide high mass accuracy, fast scanning speed and large dynamic range of detection are the Orbitrap series (Thermo) and triple TOF 5600 (AB Sciex) [56, 57].

The Obritrap analyzer was mainly used for the proteomics research reported in this thesis. In the Orbitrap mass analyzer, ions orbit around a central electrode while simultaneously moving back and forward along its axis [58]. The frequency of this axial motion is *m/z* dependent. The Orbitrap measures the image current of all ions present, each with a characteristic axial frequency, and uses Fourier transformation algorithms to convert the time-domain signals into *m/z* values. When coupled to a linear ion trap (LTQ-Orbitrap), the hybrid instrument has the advantage of both high resolution and mass accuracy of the Orbitrap and the speed and sensitivity of the LTQ. The LTQ-Orbitrap can operate in parallel: while the Orbitrap acquires a full scan to determine the accurate *m/z* of peptide ions, the LTQ performs Collision-Induced Dissociation (CID) or Electron-Transfer Dissociation (ETD) and records the *m/z* values of the of resulting fragment ions. The low precision (~100 ppm) of the LTQ, in combination with the high mass accuracy precursor ion *m/z*-value, is sufficient to obtain peptide sequence information because the fragmentation pattern of peptides is determined by the amino acid composition, and 18 of the 20 amino acids have a unique mass. However, the fragment ions can also be analyzed with high mass accuracy and resolution using the Orbitrap analyzer, which improves the identification of peptides and enhances the specificity in the database search analysis [59, 60]. In this case, the fragments ions generated upon CID or ETD are accumulated and focused by the so-called C-trap and subsequently analyzed by the Orbitrap analyzer. In addition, the implementation of an additional multipole collision cell in the LTQ-Orbitrap enabled also for beam-type Higher-energy Collisional Dissociation (HCD) [61]. In this case, the fragment ions can be analyzed in the Orbitrap with high mass accuracy.

PEPTIDE FRAGMENTATION AND IDENTIFICATION

In tandem mass spectrometry (MS/MS), peptide ions are isolated and subjected to an external stimulus to induce a controlled fragmentation, followed by the analysis of the *m/z*–values of the resulting fragment ions. The *m/z*–values of the intact peptide and of the fragment ions are required to obtain the amino acid sequence of the peptide. The fragmentation techniques commonly used in proteomics, dissociate peptides in a sequence-specific manner. These include Collision-Induced Dissociated (CID) and Electron-Transfer Dissociation (ETD). CID is the most frequently used fragmentation method because of its simplicity and ease of implementation [62]. In CID, a precursor ion is accelerated by an electric potential to a higher kinetic energy and allowed to collide with neutral gas atoms. Following the collisions, the kinetic energy of a peptide ion is converted into internal energy resulting in the cleavage of single amide bonds. The predominant fragment ions found in CID spectra of peptides are the so-called b-type and y-type ions, which are fragments derived from the N-terminus and C-terminus of the peptide, respectively (Figure 4B). Only charged fragment ions are detected by the mass spectrometer and the mass difference between the subsequent fragment ions can be used as readout of the (residual) mass of an amino acid. Dissociation of peptides is initiated by a 'mobile' proton, which is distributed over the molecule and predominantly weakens the amide bounds in the peptide backbone [62]. However, additional fragment ions are frequently observed in CID spectra, such as the loss of water, ammonia or labile Post-Translational Modifications (PTM). CID can be performed in ions traps and beam type collision cells, such as the triple quadruple (QqQ) or the earlier mentioned Higher-energy Collisional Dissociation cell (HCD) [63]. In ion trap-CID, a resonance excitation frequency is used to increase the kinetic energy of a preselected peptide ion to allow fragmentation by CID. This so-called tickle frequency is only specific for a particular *m/z*-value, preventing subsequent secondary fragmentation of primary product ions. In beam-type CID, precursor ions are typically isolated in a quadrupole mass filter and accelerated into a collision cell that is filled with a collision gas. Apart from hardware differences, fragmentation by resonant-excitation CID and beam-type CID occurs under different collision energies (eV vs. KeV) and activation times (30 ms vs. <1 ms), resulting in a different degree of fragmentation and ion intensities.

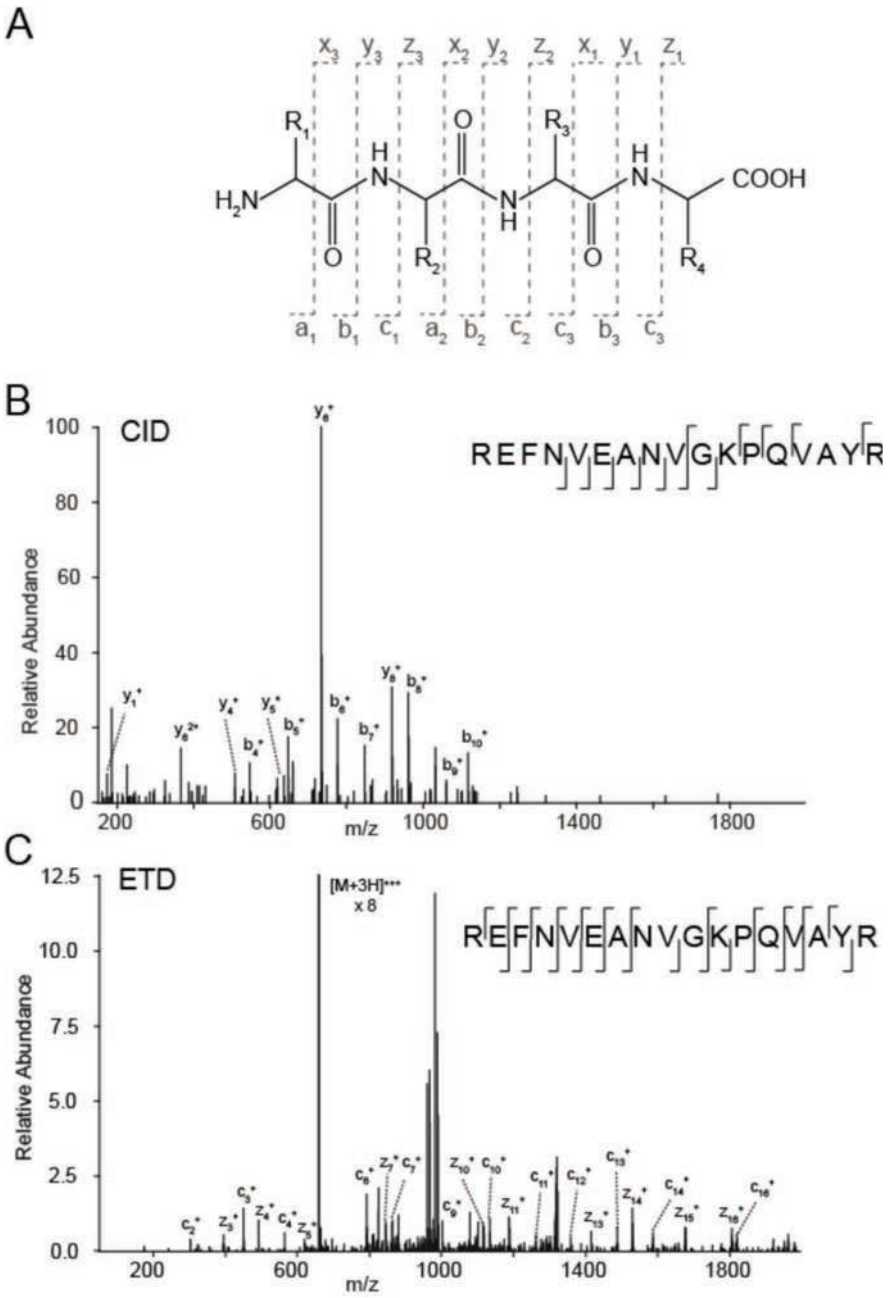


FIGURE 4. PEPTIDE FRAGMENTATION. (A) The dissociation of peptide ions can proceed via the cleavage of different bonds at the peptide backbone; a-, b- and c-type ions series are formed from the N-terminus, while x-, y- and z-type ions series are formed from the C-terminus. Nomenclature according to Roepstorff *et al.* [101]. (B) Example of a CID spectrum and (C) ETD spectrum, including the assignable fragment ions of the types b/y and c/z, respectively.

Electron Transfer Dissociation (ETD) involves an ion/ion interaction between a multiple protonated peptide and an anion radical molecule that, by transferring an electron to the peptide, induces fragmentation along the peptide backbone [64]. The charge-reduced peptide cation dissociates before energy randomization can occur, leaving labile modifications, such as phosphorylation and glycosylation, intact during ETD fragmentation. In CID, the labile bonds between these modifications and the peptide are usually preferentially fragmented, which limits the fragmentation across the peptide backbone and hampers peptide identification and localization of the Post-Translational Modification (PTM) sites. In contrast, ETD fragmentation still occurs along the peptide backbone for many PTM carrying peptides, enabling more effective sequence assignment and increased confidence in the site of modification. The predominant peptide fragmentation pathways proceeds via the cleavage of N-Cα bound of amino acids, resulting in c-type fragment ions from the N-terminus of the peptide and z-type fragment ions from C-terminus (Figure 4C). However, the major products of ETD are typically the charge-reduced precursor ions, which are the peptide precursor ions that have captured an electron but have not dissociated.

The MS/MS spectra provide direct information about peptide sequences, which could be extracted by manual interpretation or automated database search engines. Given the vast amount of spectra recorded in a standard LC-MS analysis, most researches rely on search engines, like Mascot or SEQUEST, for the assignment of an amino acid sequence to a spectrum, also referred to a Peptide-to-Spectrum Match (PSM) [30, 31]. These software programs compare the MS/MS spectra against theoretical spectra of peptides derived from *in silico* digestion of proteins from a specific protein database. Parameters necessary for database searching include the type of protease, the mass tolerance windows around the observed precursor and fragment ions, and possible (chemical) modifications (*e.g.* PTMs). Search algorithms report for each PSM a score that reflects the correlation between the experimental spectrum and the theoretical spectrum. Strict filter criteria are needed to control false positive identifications because search engines assign a peptide sequence to each spectrum, regardless the quality of the spectrum. The False Discovery Rate (FDR) is a generally established method to estimate false positive identifications and to determine the optimal filter criteria for individual datasets [32, 65, 66]. For FDR calculations, an additional database search is performed against a decoy database generated by reversing, shuffling or randomizing the target protein database. The FDR estimates for a certain threshold score the relative number of false positives in the data set by dividing the number of decoy PSMs by the total number of PSMs. Typically, a threshold score is chosen where the FDR is 1% [65].

QUANTITATIVE PROTEOMICS

Proteomes are highly dynamic systems; protein abundances constantly change due to regulation of synthesis and degradation (turnover), while their activities may be regulated via the addition or removal of Post-Translational Modifications (PTM). Hence, quantification of protein abundances is practically standard in every proteomics study aimed at understanding the biological system [67]. Many different methods have been described for absolute or relative quantification of proteins in biological systems. Absolute quantification relates to the absolute amounts of proteins in the sample, while relative quantification compares the differences in protein abundances between two or more physiological states, or in response to specific stimuli. Label free and stable isotope labeling methods can be applied to distinguish two or more biological samples and/or reference standards from each other.

In isotope labeling methods, proteins or peptides are labeled with heavy stable isotopes, such as ^{15}N, ^{13}C, ^{2}H or ^{18}O, to increase the mass compared to their native counterparts (^{14}N, ^{12}C, ^{1}H or ^{16}O) [68]. Since isotopically labeled peptides and/or proteins possess similar chemical properties, they (nearly) co-elute from the LC column and reach the MS (nearly) simultaneously. The MS is able to resolve the unique *m/z* difference between heavy and unlabeled form, due to their different physical properties (*i.e. m/z* values),and quantification is performed by comparing the mass spectral peak intensities of the corresponding isotopic analogues.

Incorporation of the stable isotopic labels can be realized at different stages in the proteomics workflow. In metabolic labeling, stable isotopes are introduced into proteins during cell growth and division, which generates a labeled standard for every protein in the sample of interest. This can be realized by substituting the nutrient source used for cell culturing with the heavy isotope analogous, typically by a ^{15}N-source (*e.g.* an isotopically labeled ammonium salt) or a ^{13}C-source of stable isotope labeled amino acids [69, 70]. The isotope label is introduced in the earliest point in the proteomic workflow, which offers high quantification accuracy and high precision because treated samples can be combined at the intact cell level, thus before lysis, digestion and fractionation (Figure 3B). In Stable Isotope Labeling of Amino acids in Culture medium (SILAC), typically isotopically labeled arginine and/or lysine are added to the culture medium to ensure that, following tryptic digestion, most peptides contain at least one labeled amino acid, allowing relative quantification of all the identified peptides. SILAC labeling is generally performed on cell culture systems although even plants and mice have been labeled using SILAC [71, 72]. In practice, a maximum of three samples can be differentially labeled by SILAC with at least a 4 Dalton mass difference between the labeled peptides. This 4 Dalton mass difference prevents accuracy problems in quantitative analysis, due to the possibility of overlapping isotope clusters in MS. The method is not applicable to autotrophic cells that are capable to synthesis all amino acids by themselves using inorganic compounds. For these groups of cells, labeling with ^{15}N, added *e.g.* as ammonium salt in the media, could be an attractive alternative. It is obvious that the number of incorporated ^{15}N labels and, hence, the resulting absolute mass differences between their native analogues depend on the peptide sequence as all nitrogen atoms are uniformly labeled, including backbone amide groups [73]. The relative mass shift effected upon the uniform incorporation of ^{15}N-atoms in proteins and/or peptides, however, is about 1.2% of the molecular weight of the analyte. This requires

alternative search strategies in peptide identification (due to the dynamic mass shifts), but it can be very useful in targeting subsets of proteins and/or peptides in complex samples [74].

Alternatively, a wide variety of chemical labeling methods is available for the modification of peptides and/or proteins with isotopically labeled reagents [75]. Basically, any reactive site on proteins and peptides can be targeted by chemical tags, but most methods rely on covalent labeling of primary amines on the protein/peptide N-terminus and lysine sides chains because this reaction is highly specific and largely complete. The two main groups of chemical reagents include labeling by isotopic tags and labeling by isobaric tags. Isobaric reagents, such as Isobaric Tags for Relative and Absolute Quantitation (iTRAQ) and Tandem Mass Tag (TMT), consist of an amine reactive group, a mass reporter and a balance group that counterbalances the differential weight of the reporter group. Consequently, identical peptides from differentially labeled samples have identical masses (isobaric) and will not further increase MS spectral complexity. The reporter moiety dissociates upon fragmentation, resulting in the generation of reporter ions in the low mass region of MS/MS spectrum. The signal intensities of the reporter ions are used for the peptide/protein quantification in multiple samples (typically 4-plex or 8-plex), while the fragment ions in the same MS/MS spectrum are used for peptide identification. Alternatively, isotope tags rely on the ability to distinguish labeled peptides and/or proteins in the MS spectra, rather than utilizing MS/MS. An example of such approach is stable-isotope dimethyl labeling, where the primary amines of the N-terminus and the lysine side chains are dimethylated by reactions with different combinations of isotopycially labeled formaldehyde and cyanoborohydride [76, 77]. Most commonly, duplex or triplex reactions are used with a mass difference of at least 4 Dalton between each sample to avoid issues with overlapping isotope envelopes. Raijmakers *et al.* [78] reported an online chemical labeling approach, which prevents elaborate sample handling and minimizing sample losses. This approach successively labels 3 samples by stable-isotope dimethylation on a trap column of the LC-MS system using an automated sample preparation protocol.

In label-free quantification, peptides and proteins are quantified without the use of an isotopic label by direct comparison of MS measurements of different samples, either by spectral counting or precursor ion signal intensities [67]. Spectral counting methods use the number of acquired Peptide-to-Spectrum Matches (PSM) as an indication for protein abundances. In intensity-based quantification, the extracted ion intensities (XIC) of any given peptide are obtained by integrating the chromatographic elution profile. The XICs are subsequently used to calculate relative protein abundances between separate runs of different sample types. The clear advantages of label-free methods include (*i*) the ability to compare an unlimited number of samples since each sample is analyzed separately, (*ii*) the applicability to any type of sample (*e.g.* human tissue, body fluid), and (*iii*) the straightforward sample preparation due to the absence of stable isotope labels. The disadvantage of label-free methods is the substantial increase in instrument time, not only because each sample is analyzed separately, but also because several replicates are often required to account for variations between samples due to errors in sample preprocessing and fluctuations in LC-MS/MS analysis. Therefore, label-free quantification also necessitates robust and reproducible sample preparation protocols and stable LC-MS platforms.

POSITIONAL PROTEOMICS

Positional proteomics is a relative new concept in proteomics, which refers to the selective enrichment and MS-based identification of peptides carrying the N-terminal or C-terminal part of the proteins [79]. Today, bottom-up proteomics typically relies on proteome digestion and analysis of the peptide fragments by various MS methods. However, it has been argued that many more peptides are analyzed than strictly necessary [79, 80]. Redundancies can be removed by depletion of all proteolytic peptides, with the exception of the positional-defined N-terminal or C-terminal peptide (Figure 5A). Since each protein is represented by one N-terminal and one C-terminal peptide, their selective enrichment significantly reduces sample complexity, while preserving the original proteome fingerprint. Positional proteomics approaches have therefore been proposed for system-wide analysis to maximize MS detection of relevant peptides, simplify data analysis and increase the dynamic range of analysis (Figure 5B). However, proteome coverage of such strategies is often limited due to the more elaborate sample enrichment procedures and because identification is based on a single peptide, which not always has the suitable characteristics for identification by MS (*e.g.* ionization or fragmentation efficiency). Currently, the main application of positional proteomics is the characterization of Post-Translational Modifications (PTMs), such as N-terminal acetylation and *in vivo* proteolysis (Figure 5C). Proteolysis, being the protease-mediated cleavage of proteins, is an irreversible modification that regulates many essential processes. These include not only the general degradation of proteins, but also highly specific processes such as proteolytic cascades (*e.g.* apoptosis) and highly specific cleavages that affect protein function, structure and cellular location [81]. Since proteolysis often proceeds at substoichiometric levels, the new N- and C-termini that are generated upon substrate cleavage get easily lost in the overwhelming background of peptides and enrichment is therefore desired. Protein N-acetylation is the transfer of an acetyl group from N-acetyltransferases to the N-terminus of newly synthesized proteins during protein translation, a very common process in eukaryotes of which the functional implications are not fully understood [82].

Several positional proteomics strategies have been designed, in particular for the enrichment of N-terminal peptides because the required chemical labeling steps proceed more efficiently on primary amines than on carboxyl groups [80, 82-87]. In general, there a two approaches that are able to discriminate between the N-terminal part and internal counterparts of proteins; direct enrichment of N-terminal peptides (positive selection) and depletion of internal peptides (negative selection) (Figure 5A). Positive selection methods are based on the incorporation of an affinity group (*e.g.* biotin) to the protein N-terminal amino groups, followed by digestion and enrichment of the modified N-terminal peptides by affinity-purification [87]. To overcome the inherent drawbacks of chemical labeling (*e.g.* incompleteness, side-reactivity), the group of Wells introduced an enzymatic biotinylation approach to selectively label the protein N-terminal amino group in a single step [84]. A severe drawback of positive selection approaches are incompletenesses in chemical or enzymatic labeling and the loss of N-terminal peptides for proteins having modified N-termini (*e.g.* acetylation) that blocks the affinity labeling reaction.

The combined fractional diagonal chromatography (COFRADIC) method is a well-established positional proteomics technology that enables the enrichment of N-terminal peptides by depletion of internal peptides (negative selection) [82, 88]. In this method, N-terminal sequences are

distinguished by several rounds of reversed phase LC separation from the internal peptides. This is accomplished by initial protection of protein N-terminal amino groups by acetylation, followed by proteome digestion and the differential labeling of internal peptides by 2,4,6-trinitrobenzenesulfonic acid (TNBS), such that the latter obtain a shift in retention on LC. Other negative selection approaches involve protective blocking of amino groups at the protein level followed by digestion and subsequent depletion of internal peptides by reaction with an amine reactive scavenger resin [79, 80] or the recently introduced high-molecular weight, water-soluble aldehyde polymer (TAILS methods) [85, 89].

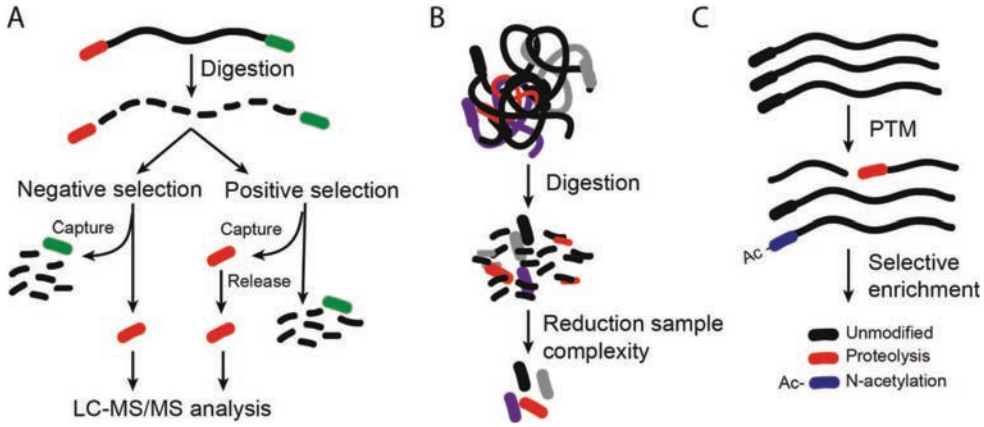


FIGURE 5. POSITIONAL PROTEOMICS. (A) Schematic representation of the isolation of N-terminal peptides (red bar), either by negative selection (left side) of positive selection (right side) methods. Figure adapted from [102]. Negative selection relies on the indirect enrichment of the N-terminal peptides by capturing all other peptides, while N-termini are directly enriched by positive selection methods. (B) Since each protein is represented by one N-terminal peptide, positional proteomics reduces sample complexity, while preserving the original proteome fingerprint. (C) Selective isolation of Post-Translational Modifications (PTM). Positional proteomes enable the isolating of unmodified N-terminal peptides (black bar), neo-N-terminal peptides resulting proteolysis (red bar) and acetylated N-terminal peptides (blue bar).

MASS SPECTROMETRY (MS)-BASED IDENTIFICATION OF MHC-ASSOCIATED PEPTIDES

Major Histocompatibility Complex (MHC) molecules are expressed on the cell surface and display short peptide fragments of self- and nonself proteins. Since these peptides have the ability to elicit an immune response upon recognition by cytotoxic T cells, they play an important role in the elimination of pathogens and certain cancer cells. The identity of MHC-associated peptides is therefore essential for our progressing insight in the antigen presentation pathway and pivotal for vaccine development and immune therapy design [7]. There are two different classes of MHC molecules, class I and class II, each communicating with a subset of T cells. MHC class I and II differ in the intercellular pathways of peptide presentation, as detailed in Figure 2.

Currently, mass spectrometry is the best technology for the high-throughput identification of peptides presented by MHC class I and class II molecules. The identification strategy for these peptides is rather similar to a bottom-up proteomics approach. The key difference is that proteomics typically relies on specific enzymes (*e.g.* trypsin) for the controlled and predictable decomposition of proteins, while MHC-associated peptides are endogenously processed and therefore analyzed without enzyme pre-treatment. Because MHC molecules are membrane proteins that can present peptides at low copy numbers per cell, enrichment procedures have been developed for the targeted analysis of MHC-associated peptides [8, 90, 91]. The most used approach is the affinity purification of MHC-peptide complexes from cell lysates utilizing anti-MHC antibodies. The source material for this method can be tissue material, blood cells or pellets from cultured cell lines [92, 93]. Typically, after detergent-based cell lysis, MHC-peptide complexes are isolated by subjecting the cell lysate to an affinity column to which a monoclonal antibody have been bound. Peptides are subsequently released from MHC-peptide complexes by strong acids elution (*e.g.* trifluoroacetic acid at pH 3) and further purified by high-molecular weight cut-off filtration (typically 10 kDa).

Alternatively, peptides can be isolated directly from MHC-peptide complexes that are presented at the cell surface by a short, mild acid treatment. This treatment results in the partial denaturation of the MHC molecule and the release of the bound peptides. Since the majority of the cells remain intact, cells can be regrown and the mild acid elution step repeated to improve the final peptide yield. However, the method suffers from specificity because cell damage during acid treatment can hardly be avoided [92, 93]. The group of Hildebrand has developed a soluble MHC construct for the milligram-scale enrichment of MHC-peptide complexes by affinity purification [94, 95]. As a result of stable transfection of cells with MHC molecules that lack the transmembrane domain, MHC-peptide complexes are not anchored in the plasma membrane and secreted in large amounts into the culture medium.

The enriched fraction of MHC-associated peptide is analyzed by a standard bottom-up proteomics workflow (Figure 3), typically using extensive fractionation, subsequent LC-MS/MS analysis and database search analysis. Although these workflows are continuously improved, particularly for tryptic peptides, challenges remain in MS-based sequencing and data analysis of MHC-presented peptides. The psychochemical properties of these endogenous peptides are enormously diverse because they are produced by a multi-enzyme process. The variable nature and occurrence of

internal basic residues represents difficulties in peptide sequencing and interpretation of the MS/MS spectra [96]. To address the limitations in peptide sequencing, sophisticated stable isotope labeling methods have been developed to allocate the peptides that change upon infection or malignant transformation [97, 98]. These methods enable the identification of disease-related peptides by high-accurate MS and subsequent comparison of the candidate peptide with a synthetic counterpart, even if fragmentation analysis (MS/MS) is not providing feasible information. Moreover, data analysis by automated search algorithms is not straightforward because MHC-presented peptides are processed by a variety of enzymes with multiple terminal cleavage specificities. As a consequence, MS/MS spectra are searched against the theoretical proteome database without an enzyme restriction. This leads to a large increase in database search space, accompanied with an increase in false positive identifications [12]. Implementation of strict filter criteria (*e.g.* < 1% FDR) is therefore often not possible without the loss of valuable data [12].

OUTLINE THESIS

In the past decades, immunoproteomics has greatly benefitted from the enormous progress in enabling technologies [8], such as the developments in soft ionization methods [25, 26], peptide separation [28, 34], MS instrumentation [49] and bioinformatics [30]. However, the tracing and identification of individual antigens, as well as global-scale analysis of complex proteomes remains highly challenging, but essential for a good understanding of the immune system and pivotal for vaccinology. Significant improvements in sample enrichment, peptide fractionation and MS instrumentation are still needed to obtain an in-depth view of proteomes [18]. This thesis describes several technological developments, aimed at improving specificity and specificity in mass spectrometry (MS)-based proteomics, and the application of these advancements in vaccine development and fundamental immunological studies.

In **CHAPTER 2,** we describe the development of a positional proteomics strategy for the selective isolation of N-terminal peptides of proteins. A novel affinity tag (PTAG) was used in combination with titanium dioxide affinity chromatography to reduce sample complexity. We demonstrate the relevance of such strategy for the N-proteome analysis of proteomes that are dominated by a single protein (dynamic range issues) and the identification of Post-Translational Modifications (PTM), *e.g.* acetylation and protease-generated cleavage products.

CHAPTER 3 describes the implementation of a quantification method to the PTAG approach. The method is based on duplexed stable-isotope dimethyl labeling of primary amines. We also introduced a common reference strategy that allows for quantitative analysis of many biological replicates. The quantitative PTAG approach was used to profile the relative protein content of outer membrane vesicles (OMV)-based vaccines, produced under different purification process conditions. The quantitative proteomics results were substantiated with serum blot proteomics to visualize and assess differences in immunogenic protein content.

In **CHAPTER 4**, we describe the development of a two dimensional (2D) chromatographic separation technique for the analysis complex peptide mixtures. The system utilizes mixed-bed ion exchange chromatography (*i.e.* weak anion and strong cation exchange) for first dimensional and reversed phase chromatography for second dimensional separations. To improve the overall sensitivity and compatibility with mass spectrometry, we introduced a new mobile phase system for ion exchange separations, optimized the 2D elution profile and miniaturized the column dimensions. The performance of the system was evaluated for the analysis of minute amounts (one microgram) of digested yeast proteins.

CHAPTER 5 describes the utility of a recently introduced dual fragmentation method for the identification of Human Leukocyte Antigens (HLA) class I-associated peptides. This dual fragmentation method employs Electron Transfer and Higher-energy Collision Dissociation (EThcD) to a single ion package. We demonstrate the need for advanced peptide sequencing by EThcD to provide more extensive peptide backbone dissociation and the more readily identification of endogenous peptides. The direct importance of this technique is provided by the generation of a large peptide inventory that contains new features in the sampling of peptides by HLA molecules.

This thesis covers major methodological aspects in the MS-based proteomics workflow, from advances in sample enrichment to peptide identification. Applied to different research questions in vaccinology, these methods address challenges in sample complexity, sensitivity in MS, as difficulties in peptide sequencing. In **CHAPTER 6**, the thesis ends with a summary and outlook.

CHAPTER 02

Unbiased selective isolation of protein N-terminal peptides from complex proteome samples using phospho tagging (PTAG) and TiO_2-based depletion

Geert P.M. Mommen 1,2,3,‡
Bas van de Waterbeemd 1,‡
Hugo D. Meiring 1
Gideon Kersten 1
Albert J. R. Heck 2,3
Ad P.J.M. de Jong 1

‡ Both authors contributed equally
1 Unit Vaccinology, Centre for Infectious Disease Control Netherlands, National Institute for Public Health and the Environment, The Netherlands
2 Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands
3 Netherlands Proteomics Centre, Utrecht, The Netherlands

Molecular and Cellular Proteomics, 2012. 11(9): p. 832-42

ABSTRACT

A positional proteomics strategy for global N-proteome analysis is presented based on phospho tagging (PTAG) of internal peptides followed by depletion by titanium dioxide (TiO_2) affinity chromatography. Therefore, N-terminal and lysine amino groups are initially completely dimethylated with formaldehyde at the protein level, after which the proteins are digested and the newly formed internal peptides modified with the PTAG reagent glyceraldhyde-3-phosphate in nearly perfect yields (> 99%). The resulting phosphopeptides are depleted through binding onto TiO_2, keeping exclusively a set of N-acetylated and/or N-dimethylated terminal peptides for analysis by LC-MS/MS. Analysis of peptides derivatized with differentially labeled isotopic analogous of the PTAG reagent revealed a high depletion efficiency (> 95%). The method enabled identification of 753 unique N-terminal peptides (428 proteins) in *N. meningitidis* and 928 unique N-terminal peptides (572 proteins) in *S. cerevisiae*. These included verified neo-N-termini from subcellular-relocalized membrane and mitochondrial proteins. The presented PTAG approach is therefore a novel versatile and robust method for mass spectrometry-based N-proteome analysis and identification of protease-generated cleavage products.

INTRODUCTION

In shotgun proteomics, proteins are digested into peptides, typically using trypsin as protease, separated by liquid chromatography and analyzed by online-coupled tandem mass spectrometry (LC-MS/MS). Identifying significant portions of all proteins present in complex samples by LC-MS remains a major challenge, even for advanced proteomics workflows [103]. To address the challenges, new concepts in sample preparation have been proposed, aiming at reduction of sample complexity while preserving the proteome fingerprint [79, 80, 88]. The most useful methods for this purpose yield a single, positional-defined peptide for each individual protein. McDonald and Beynon argued that the two most obvious positional locations within every protein are the extreme ends, thus the N-terminal and the C-terminal peptides (positional proteomics) [79]. As a result of drastic sample simplification, positional proteomics analysis provides insights in a variety of post translational modification (PTM) processes and proteolytic processing which proteins may undergo at their N-terminal and C-terminal ends [81, 104].

Positional proteomics strategies rely on the ability to differentiate between the N- or C-terminal parts of a protein and the internal counterparts [105, 106]. Protocols for target enrichment of C-terminal peptides have only recently been introduced [83, 86], mainly due to lower chemical reactivity of C-terminal carboxyl groups compared to N-terminal amino groups. Gevaert *et al.* [82, 88] introduced the well-established combined fractional diagonal chromatography (COFRADIC) technology. In their method, N-terminal sequences are distinguished and separated from the internal peptides by differential labeling of protein N-terminal amino groups on the one hand and the α-amino groups of internal proteolytic peptides on the other hand such that the latter obtain a shift in retention on reversed phase chromatography. To prevent the discriminative retention shift for N-terminal amino acid sequences, free amino groups of protein N-termini (α-amino) and lysine side chains (ε-amino group) are protected by acetylation prior to digestion. Another negative selection approach, proposed by McDonald *et al.* [79, 80] involves the protective blocking of amino groups at the protein level followed by digestion and subsequent depletion of internal peptides by reaction with an amine reactive scavenger resin. Kleifeld *et al.* [89, 107] have developed terminal amine isotope labeling of substrates (TAILS) for the negative selection of N-terminal peptides and identification and quantification of proteolytic events. They used a novel water-soluble aldehyde polymer for the selective binding of α-amine containing internal peptides [108]. Positive selection methods employ a reversed approach [109]. These protocols are based on the incorporation of an affinity group (*e.g.* biotin) to the protein N-terminal amino groups, followed by digestion and enrichment of the modified N-terminal peptides [87, 110]. Unwanted cross-reaction with the side chain amino group of lysines is prevented by guanidination (lysine to homoarginine conversion) at the protein level. Selective and complete lysine labeling on the protein level can be problematic, hence the group of Wells introduced an enzymatic approach to selectively label the protein N-terminal amino group in a single step [84]. A severe drawback of positive selection approaches is the loss of N-terminal peptides for proteins having naturally acetylated or otherwise modified N-termini, because these termini do not react with the affinity labeling agents.

Selective enrichment of N-terminal peptides constitutes a major challenge due to the consecutive sample preparation steps (*i.e.* protective labeling, purification and enrichment), each prone to side reactivity and sample losses. Moreover, tryptic protein digests contain many more internal

peptides than N-terminal peptides, therefore posing high demands on the efficiency of depletion to prevent significant contamination of the final sample fraction with internal peptides [109]. For example, Timmer *et al.* [87] used NHS-activated biotin for the positive selection of protein N-termini, however it was stated that a substantial portion of positive identifications were observed as a result of nonspecific biotinylation [81]. In case of amine-reactive scavenger beads, multiple incubation steps were needed for effective coupling of internal peptides [80, 111]. Zhang *et al.* reported specific loss of histidine-containing N-terminal peptides when using NHS-activated sepharose [111]. In addition, histidine- and arginine-containing N-terminal peptides are generally underrepresented in N-terminomics when SCX is used for pre-enrichment prior to depletion of internal peptide [112]. N-acetylated N-termini, widely present in higher eukaryotes, can be enriched more easily using SCX fractionation (without derivatization chemistry), but unfortunately such approaches are blind for unmodified protein N-termini. [37, 113, 114]. TAILS, however, used water-soluble aldehyde polymer for effective coupling and depletion of internal peptides in a single-step, thereby minimizing possible sample losses [107].

In view of the challenges associated with the enrichment of N-terminal peptides, a novel positional proteomics strategy was developed. The strategy utilizes a highly selective PTAG-labeling reagent for the modification of internal peptides, after initial protection of protein N-termini and lysine side chains. PTAG-derivatized peptides have similar properties as naturally phosphorylated peptides in terms of binding to titanium dioxide (TiO_2). Hence, the flow-through fraction of TiO_2 affinity chromatography is highly enriched in N-terminal peptides and could be directly analyzed by LC-MS/MS. It is demonstrated that PTAG is a straightforward and efficient N-proteome enrichment strategy due to the use of extreme selective derivatization chemistry, both at the protein and peptide level, in combination with robust and relative easy-to-implement TiO_2 technology.

MATERIALS AND METHODS

CELL CULTURE. *Neisseria meningitidis* strain used in this study is a recombinant non-encapsulated variant of the group B isolate H44/76, with a nonfunctional porB gene and truncated galE LPS [115]. Bacterial cultures were grown at 35 °C in a chemically defined medium (150 mL) in disposable, baffled 500-mL erlenmeyer shakeflasks with vented closure (Nalgene, Rochester, NY, U.S.A.) by shaking at 200 rpm [115]. Cells were harvested by centrifugation at 13,000x*g* for 2 min at 4 °C and resuspended in TE buffer (0.1 M EDTA, 1 M Tris-HCl pH 8.0, Sigma Aldrich, Zwijndrecht, The Netherlands) containing 0.5 mg/ml lysozyme (Sigma Aldrich, Zwijndrecht, The Netherlands) and incubated at 4 °C for 3 minutes in this medium. Next, proteins were extracted with Trizol (Invitrogen, Blijswijk, The Netherlands) according to the manufacturer's protocol and stored at -80 °C prior to use. Outer membrane vesicles (OMV) from *N. meningitidis* (grown as described above) were released by adding EDTA extraction buffer (0.01M EDTA, 0.1M Tris-HCl pH 8.6) and further purified by consecutive centrifugation and ultracentrifugation steps, as described by van de Waterbeemd *et al.* (NOMV protocol) [115]. Concentrated OMV (approximately 100 µl sample, containing 1 mg total protein) were mixed with 1 mL Trizol reagent and proteins were extracted and stored as described above.

Sacchoromyces cerevisiae strain BJ5460 (LGC Standards, Almere, The Netherlands) was cultured in 150 mL YPD medium in baffled 500-mL erlenmeyer shakeflasks with vented closure (Nalgene, Rochester, NY, U.S.A.) at 30 °C, by shaking at 200 rpm. Cells were harvested from 300 mL culture (OD_{590}=1.7) by centrifugation at 2,000x*g* for 5 min. Cells were washed three times with PBS and resuspended in 200 µL lysis buffer (2 M guanidine hydrochloride, 12 mM EDTA, 50 mM Tris-HCl, pH 7.5 (Sigma Aldrich, Zwijndrecht, The Netherlands) to which 5 µL protease inhibitor cocktail (Sigma Aldrich, Zwijndrecht, The Netherlands) was added. The cell suspension was subjected to three rounds of freeze-thaw cycles. Next, cleaned glass beads were added and cells were further disrupted by 6 vortex cycles with intermediate cooling steps (at 4 °C). Supernatants after each cycle were pooled and centrifuged at 2,000x*g* at 4 °C for 10 min. The resulting supernatant was incubated overnight with a 4-fold excess of acetone at -20 °C and proteins were subsequently pelleted at 13,000x*g* at 4 °C for 10 min. The protein pellet was washed twice with acetone/water 4/1 (v/v), pelleted after each wash step as described before and dried for 5 min by vacuum centrifugation at room temperature.

DIMETHYLATION OF PRIMARY AMINES. The protocol for dimethylation of primary amines was adapted from Boersema *et al.* [77]. An aliquot corresponding to 100 µg of protein was dissolved in 100 mM KH_2PO_4 (the pH adjusted to 7.5) containing 4 M guanidine hydrochloride. Disulfide bridges were reduced by adding dithiotreitol (Sigma Aldrich, Zwijndrecht, The Netherlands) to a final concentration of 20 mM and incubated 37 °C for 30 min. Free thiol groups were alkylated by adding iodoacetamide (Sigma Aldrich, Zwijndrecht, The Netherlands) to a final concentration of 100 mM and incubation at room temperature for 30 min in the dark. Excess reagent was exhausted by the addition of dithiotreitol at a final concentration of 100 mM (incubated at 37 °C for 30 min). The free N-terminal and lysine amino groups were dimethylated by formaldehyde at a final concentration of 180 mM (CH_2O, Sigma Aldrich, Zwijndrecht, The Netherlands) in the presence of 30 mM freshly prepared sodium cyanoborohydride ($NaCNBH_3$, Sigma Aldrich, Zwijndrecht, The Netherlands) at room temperature. Freshly prepared sodium cyanoborohydride at a final concentration of 30 mM was added after 1-h and 2-h time intervals and the sample was further incubated overnight.

Subsequently, the mixture was diluted 4 times with water to decrease the guanidine hydrochloride concentration to less than 1 M and proteins were extracted by acetone precipitation as described above. Precipitated proteins were reconstituted in 15 µL of 100 mM KH_2PO_4 (pH 7.5) containing 4 M guanidine hydrochloride. Excess formaldehyde was exhausted by the addition 50 µL of 1 M ammonia hydroxide (Zwijndrecht, The Netherlands) and incubation at room temperature for 1 h. Ammonium hydroxide was removed by vacuum centrifugation at room temperature till dryness.

PROTEIN DIGESTION. Dimethylated proteins were (parallel) digested in 50µl 100 mM KH_2PO_4 (pH 7.5) and guanidine hydrochloride with a concentration of less than 1 M with either chymotrypsin (Roche, Woerden, The Netherlands) in 4 mM calcium chloride (Sigma Aldrich, Zwijndrecht, The Netherlands) at 37 °C, trypsin (Promega, Leiden, The Netherlands) at 37 °C or endoprotease GluC (Roche, Woerden, The Netherlands) at room temperature, all with an enzyme/protein ratio of 1/20 (w/w). After 1 h, digest mixtures were diluted twice with 100 mM KH_2PO_4 (pH 7.5) to reduce to guanidine concentration to less than 0.5 M. Fresh enzyme was added at a enzyme/protein ratio of 1/20 w/w and the mixture was further incubated overnight at similar temperature conditions as described above.

REMOVAL PYROGLUTAMATE. N-terminal glutamine was enzymatically cyclized by glutamine cyclotransferase (Qcyclase, Qiagen, Venlo, The Netherlands) and the formed pyroglutamyl moiety was subsequently cleaved by the aminopeptidase pGAPase (Qiagen, Venlo, The Netherlands). This protocol was adapted from Staes *et al.* [112], with adjustment of the incubation time to 2 h at 37 °C.

PREPARATION OF PTAG DERIVATES. The free amino groups of the internal peptides were PTAG derivatized in 100 mM KH_2PO_4 (pH 7.5) with DL-glyceraldehyde-3-phosphate (Sigma Aldrich, Zwijndrecht, The Netherlands) at a final concentration of 90 mM and freshly prepared sodium cyanoborohydride in a final concentration of 30 mM at room temperature. Freshly prepared sodium cyanoborohydride was added at 1h and 2h intervals and the reaction mixture was further incubated overnight. Following PTAG derivatization, peptide mixtures were extensively purified by C18 solid phase extraction (SPE) chromatography (column dimensions 5 cm (L) x 200 µm inner diameter (ID), inhouse packed with 5 µm Reprosil Pur C18-AQ, Dr Maisch, Ammerbuch-Entringen, Germany).

DEPLETION OF PTAG-PEPTIDES. PTAG-peptides were depleted by TiO_2 affinity chromatography, essentially as previously described [116, 117]. Briefly, SPE-purified samples were evaporated to dryness, reconstituted in 0.1 M acetic acid in water (pH 2.7) and loaded onto a short TiO_2 column at a flow rate of 5 µl/min (100 µl injection loop). The short TiO_2 column comprises of a 1-mm ID PEEK tubing with an Upchurch (Oak Harbor, U.S.A.) 360-mm ID adapter at the front and end for connection to (fritted) microcapillary tubing and is slurry-packed with a 10-mm (L) bed of 5 µm Titansphere particles (10 mg) (GL Sciences, Tokyo, Japan). Unretained peptides were collected in the void volume. Next, the TiO_2 column was extensively washed with a 100-µl plug (3 column volumes) of acetonitrile/water/dimethyl sulfoxide in 0.1 M acetic acid (45/45/10/, v/v/v) (Sigma Aldrich, Zwijndrecht, The Netherlands). The TiO_2 flow-through fraction and the wash fraction were pooled, evaporated to dryness by vacuum centrifugation and reconstituted in formic acid/DMSO in water (5/5, v/v) and stored at -20 °C until analysis.

LC-MS/MS ANALYSIS. N-terminal peptides-enriched samples (TiO_2 flow-through fraction) were prefractionated offline (6-14 fractions) using a mixed bed anion-cation column as described by Motoyama *et al.* [118] or directly analyzed on an LTQ-Orbitrap XL instrument (Thermo Fisher Scientific, Bremen, Germany) and Agilent 1100 HPLC system (Agilent, Amstelveen, The Netherlands) modified for nanoflow LC separations as described previously by Meiring *et al.* [119]. All columns were packed in house. The trap column was a 100-µm ID fritted microcapillary packed with 20 mm, 5 µm particle size Reprosil Pur C18-AQ particles (Dr. Maisch, Ammerbuch-Entringen, Germany). The analytical column was a 50-µm ID microcapillary packed with 31 cm 3 µm particle size Reprosil Pur C18-AQ, with an integral-pulled tip and operated at a flowrate of 125 nL/min. ESI voltage, typically 1.8 kV, was applied by liquid junction at the top of the column. Solvent A consisted of 0.1 M acetic acid (Sigma Aldrich, Zwijndrecht, The Netherlands) in deionized water (Mili-Q, Millipore, Amsterdam, The Netherlands) and solvent B of 0.1 M acetic acid in acetonitrile (Biosolve, Valkenswaard, The Netherlands). Gradients were as follows: 100% solvent A during sample loading (0-10 min, flowrate 5 µL/min), 7% to 26% solvent B in 160 minutes followed by an increase to 60% solvent B in 20 minutes and reconditioning with solvent A for 10 min (total runtime 200 min). The mass spectrometer was set to acquire full MS spectra (*m/z* 350 to 1,500) for mass analysis in the orbitrap at 60,000 resolution (FWHM) followed by data-dependent MS/MS analysis (LTQ) for the top 5 or 7 abundant precursor ions above a threshold value of 10,000 counts. The normalized collision energy was set to 35%, isolation width to 2.0 Da, activation Q to 0.250 and activation time to 30 ms. The maximum ion time (dwell time) for MS scans was set to 200 ms and for MS/MS scans to 2500 s. Charge state screening and preview mode were enabled. Precursor ions with unknown and +1 charge states were excluded for subsequent MS/MS analysis. Dynamic exclusion was enabled (exclusion size list 500) with repeat set to 1 and an exclusion duration of 180 s. The background ion at *m/z* 391.284280 was used as lock mass for internal mass calibration.

DATA ANALYSIS. The analysis of mass spectrometric RAW data was carried out using Proteome Discoverer 1.2 software (Thermo Fisher Scientific, Bremen, Germany) applying standard settings unless otherwise noted. MS/MS scans were searched against the *N. meningitidis* strain MC58 database (containing 2,003 entries, 2010, Uniprot) or the *S. cerevisiae* SGD database (http://www.yeastgenome.org, 2010, containing 5,821 entries) using SEQUEST (Proteome Discoverer 1.2, Thermo Fisher Scientific, Bremen, Germany). Precursor ion and MS/MS tolerances were set to 10 ppm and 0.8 Da, respectively. Decoy database searches were performed with False Discovery Rate (FDR) tolerances set to 5% and 1% for modest and high confidence filtering settings, respectively. The data were searched separately with either no enzyme, C-terminal trypsin cleavage specificity, C-terminal chymotrypsin cleavage specificity or C-terminal GluC cleavage specificity allowing 5 miss-cleavages because lysine cleavage is prevented by dimethyl modification. Cysteine carbamidomethyl, N-terminal dimethylation and lysine dimethylation were set as fixed modifications while aspargine deamidation and methionine oxidation were set as variable modifications. Similar searches were performed for alternative modifications by substituting N-terminal dimethylation modification with acetylation, ammonia loss, no modification, glyceraldehyde-3-phosphate, and monomethylation (Supplemental tables 3-5). N-terminal dimethylated, N-terminal acetylated and N-terminal monomethyled proline-starting peptide sequences with high confidence (Xcorr values > 2.2, false discovery rates < 1%), rank No. 1 and linear sequences within the first 70 amino acids

were considered for manual data analysis. Peptides possessing charge states of 6+ and higher were excluded. True N-terminal peptides (initiator methionine and methionine cleavage) were kept in the final dataset as well as proteins with signal peptide or transit peptide cleavage sites. Peptidase cleavages sites were verified by prediction software [120] or previous data [121] (Supplemental table 6-7). Annotated spectra are provided for proteins with only a single confident peptide identification (Supplemental spectra 1-2). The raw data files and protocols associated with this manuscript are available to the reader if requested.

ASSESSMENT OF BINDING EFFICIENCY OF PTAG-PEPTIDES. Synthesis of isotopically deuterium-labeled PTAG reagents was started from either D_0-tetrahydrofuran or D_8-tetrahydrofuran (Sigma Aldrich, Zwijndrecht, The Netherlands). 4-bromobutyl acetate was obtained by ring-opening of tetrahydrofuran and nucleophilic substitution (into the acetate ester) by incubation with 33% hydrobromic acid in acetic acid (Sigma Aldrich, Zwijndrecht, The Netherlands) at room temperature for 1 h. Tetramethylammonium salt of dibenzylphosphate was prepared from the dropwise addition of tetramethylammonium hydroxide into a solution of dibenzylphosphate in methanol/acetone 1/1 (v/v) (Sigma Aldrich, Zwijndrecht, The Netherlands) at -10 °C. Tetramethylammonium salt of dibenzylphosphate was refluxed with 4-bromobutyl acetate for 5 h in dioxane (Sigma Aldrich, Zwijndrecht, The Netherlands). The resulting dibenzyl-4-acetatebutyl phosphate was purified on silica and hydrogenated into dibenzyl-4-hydroxybutyl phosphate with 0.4 M sodium carbonate (Sigma Aldrich, Zwijndrecht, The Netherlands) in 1/1 (v/v) ethanol/water at room temperature for 24 h. The product was extracted from the ethanol/water mixture with dichloromethane and subsequently dried on anhydrous magnesium sulphate (Baker, Deventer, The Netherlands). Dibenzyl-4-hydroxybutyl phosphate was oxidized into dibenzyl-4-oxobutyl phosphate by the incubation with a catalytic amount of pyridinium chlorochromate (PCC, Sigma Aldrich, Zwijndrecht, The Netherlands) in dichloromethane at room temperature for 1 h. Dibenzyl-4-oxobutyl phosphate was purified on silicagel and the dibenzylphosphate functionality was reduced by a catalytic amount of 10% palladium on carbon (Sigma Aldrich, Zwijndrecht, The Netherlands) under a hydrogen atmosphere at room temperature for 1 h. The structure of the final products, D_0-4-oxobutyl dihydrogen phosphate and D_5-4-oxobutyl dihydrogen phosphate, were verified by NMR (Joel (400 Mhz), Tokyo, Japan). Partial hydrogen/deuterium exchange was observed as a result of enolization of the deuterated carbonyl functionality. After equilibration of the enolization reaction (storage in water), a $\Delta M = 5.0$ Da between the isotopic variants was obtained. The TiO_2 binding efficiencies of the PTAG-peptides were evaluated using the inhouse-synthesized, isotopically-labeled PTAG-reagents. An aliquot corresponding to 100 µg of protein from the OMV fraction of *N. meningitidis* was dimethylated as described above. Proteins were digested with trypsin and free α-amines of internal peptides were PTAG derivatized with an equimolar mixture of D_0-4-oxobutyl dihydrogen phosphate and D_5-4-oxobutyl dihydrogen phosphate and the addition of freshly prepared sodium cyanoborohydride (as described above). Peptide mixtures were subjected to TiO_2 affinity chromatography (as described above) and the flow-through fraction was analyzed by LC-MS. Raw data files were converted to the NetCDF file format and imported into MsXelerator software package (v.2.9, MsMetrix, Maarssen, The Netherlands). Co-eluting mass spectral doublets with a ΔM of 5.0306 (PTAG-peptides) were extracted by the search algorithm as previously described by Meiring *et al.* [46].

RESULTS

GENERAL DESCRIPTION OF N-PROTEOME ENRICHMENT. A schematic overview of phospho tagging (PTAG) for global N-proteome analysis is depicted in Fig. 1. The strategy starts with the denaturation of proteins by reduction and alkylation of cysteines to enhance the accessibility of amino groups for chemical labeling and erratic depletion of cysteine linkages in N-terminal peptides. Reductive dimethylation of primary amines using formaldehyde and sodium cyanoborohydride simultaneously blocks the free α-amines of protein N-termini, except when they are already *in vivo* blocked by N-acetylation, as well as ε-amines of the lysine side chains [77]. Subsequent, trypsination of the proteome will result in an ArgC like digestion pattern as lysine cleavage is prevented due to the dimethyl modification, as previously shown [107]. The free N-terminal α-amines of the, upon digestion, newly generated internal peptides are susceptible for tagging with the PTAG reagent glyceraldehyde-3-phoshate (GAP3) (Fig 2.). PTAG-labeled peptides are subsequently depleted through binding to TiO_2, with the flow-through fraction being highly enriched in N-acetylated and N-dimethylated peptides is analyzed by LC-MS/MS (Supplemental Fig 1.).

Confident protein assignment may be problematic in N-terminal peptide enrichment strategies because identification is inherently based on a single peptide, referred to as 'one-hit wonders' [122]. For this reason, parallel replicates of each proteome sample were digested with three different proteases (trypsin, endoprotease GluC and chymotrypsin) to generate overlapping N-terminal peptides with different lengths. In addition, N-proteome coverage is increased by the parallel use of proteases, because trypsin alone may generate N-terminal peptides of inappropriate length or

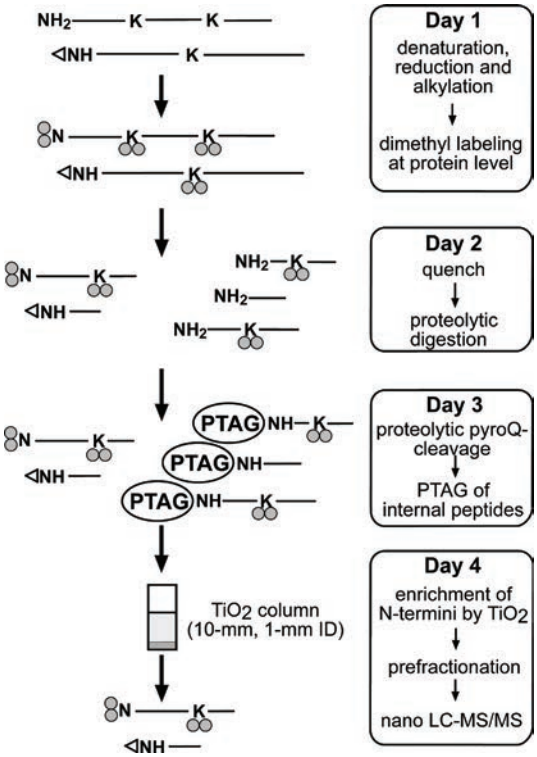


FIGURE 1. WORKFLOW FOR RECOVERY OF N-TERMINAL PEPTIDES BY THE PTAG STRATEGY. Protein samples are denaturated, cysteines reduced and alkylated, following by the reductive dimethylation of free α- and ε-amines (double grey dot) (day 1). Proteins are purified from excess reagent by acetone precipitation and remaining residual amounts are quenched by ammonium hydroxide. After proteolytic digestion (day 2), peptides with an N-terminal glutamine are enzymatically converted into pyroglutamate by Qcyclase and subsequently cleaved by pGAPase. Protein internal peptides are Phospho-tagged (PTAG) by the reaction with DL-glyceraldehyde-3-phosphate (GAP3) in the presence of cyanoborohydride (day 3). Peptide mixtures are subsequently desalted and excess reagents were removed by C18 solid phase extraction (SPE) chromatography. PTAG-peptides are depleted by TiO_2 affinity chromatography (day 4) for enrichment of dimethylated (grey double dot) and naturally acetylated (open triangle) in the flow-through fraction.

poor ionization and fragmentation properties to be identified by LC-MS/MS [79, 114]. Despite the multiple digestion strategy, many proteins are expected to be identified based on a single peptide. To enhance confidence in protein assignment for these single peptide identifications, offline ion exchange prefractionation (typically 6 fractions) was performed in combination with high accurate MS analysis using an LTQ-Orbitrap to increase the number of MS/MS identification of a single peptide from technical replicates, different charge states and deamidation or oxidation states [107]. Also, stringent database search criteria were used (false discovery rates < 1%) to obtain high confidence in sequence assignment and annotated spectra for proteins with a single peptide identification are provided (Supplemental spectra 1-2).

DIMETHYL LABELING ON THE PROTEIN LEVEL. High efficiency in reductive dimethylation of N-termini and lysine side chains is critical to minimize erratic depletion of N-terminal peptides [89]. Protective labeling of these amino groups with formaldehyde was investigated for the standard protein Cytochrome C. Progress of the reaction was determined on the 6 available ε-amine groups of the chymotryptic acetylated N-terminal peptide Ac-GDVEKGKKIF. At 3-hour, the reaction yielded an incorporation of 5.9 methyl groups (98%) (Supplemental Fig. 2). The reaction was driven to essentially quantitative by the addition of freshly prepared cyanoborohydride after 1-h and 2-h points followed by overnight incubation (data not shown).

PHOSPHO TAGGING (PTAG) OF INTERNAL PEPTIDES. PTAG-labeling of proteolytically generated internal peptides should reach completion to minimize co-purification of unmodified internal peptides in the final sample mixture. Evaluation of the derivatization reaction for the internal peptides generated from chymotryptic digestion of Cytochome C yielded > 99% efficiency without any side reactions. Similar results were found for a more complex (100 µg) trypsinated Outer Membrane Vesicle (OMV) fraction from *N. meningitidis*. OMV are nano-sized spherical protein-lipid vesicles which can be used as a vaccine against serogroup B meningitis [123]. The main antigen present in OMV is the porin A protein (PorA), which constitutes approximately 75% of the total protein content. For proteolytically digested OMV proteome, with abundances spanning several orders of magnitude, PTAG derivatization chemistry was nearly quantitative (> 99%). Despite the high labeling yields, unmodified residues could still be detected however at greatly reduced abundances (< 1%) compared to their PTAG-labeled analogues (Supplemental Fig. 3). Slow and/or incomplete reactivity of proline-starting internal peptides, as reported by others [88, 112], was not observed for the PTAG reagent.

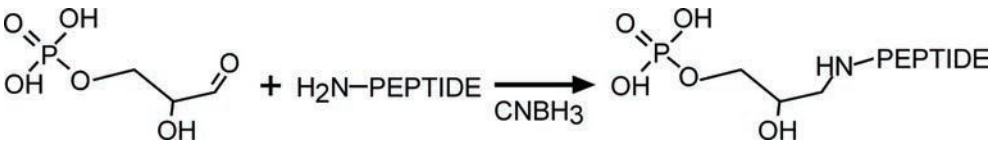


FIGURE 2. PHOSPHO-TAGGING REACTION SCHEME. N-terminal α-amines of proteolytically generated peptides react with glyceraldehyde-3-phosphate (GAP3) into Schiffs bases that are rapidly reduced in the presence of cyanoborohydride. Note, ε-amines of lysine side chains are protected by dimethylation with formaldehyde and cyanoborohydride on the protein level.

DEPLETION EFFICIENCY OF PTAG-PEPTIDES. The most critical step in positional proteomics is the effective depletion of internal peptides while maintaining maximum recovery of N-terminal peptides. The depletion efficiency of internal peptides for the PTAG strategy was evaluated for the chymotryptic digest of Cytochrome C (Supplemental Fig. 4). The naturally acetylated N-terminal peptide of Cytochrome C (Ac-GDVEKGKKIF) was detected as the most predominant ion trace in the flow-through fraction.

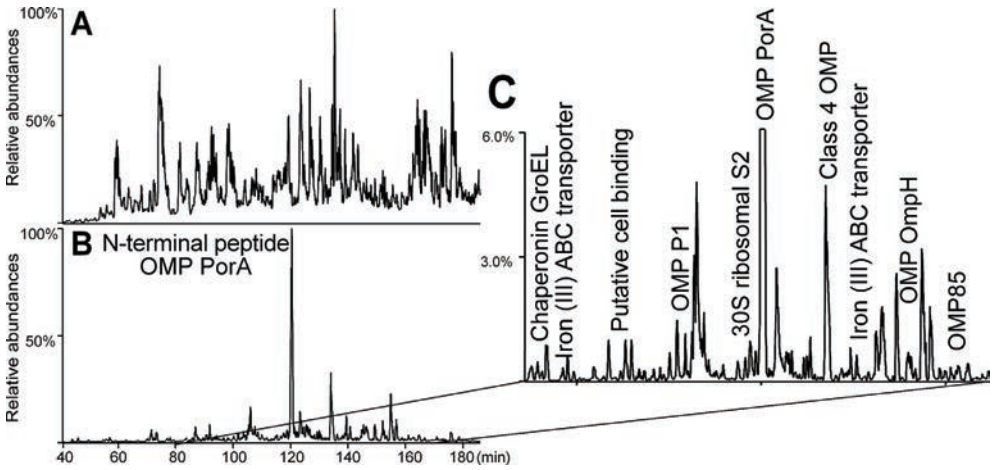


FIGURE 3. LC-MS MAPS OF PROTEOLYTIC PEPTIDES (GLUC) OF *N. MENINGITIDIS* OUTER MEMBRANE VESICLE (OMV) FRACTION BEFORE (A) AND AFTER (B) ENRICHMENT OF N-TERMINAL PEPTIDES. The protein composition of OMV is dominated by the outer membrane porA protein, accounting for approximately 75% of the total protein content. N-terminal peptides of several proteins, including the high abundant N-terminus of PorA were recovered as predominant ion trace in the enriched sample fraction (B and C), indicating effective depletion of internal peptides.

For the N-proteome analysis of complex proteome samples, a high capacity, offline TiO_2 affinity chromatography column was prepared in-house (slurry packed). Stringent washing conditions were included to minimize nonspecific binding [124] (Supplemental table 2). A 100-µg proteolytic digest of OMV was subjected to TiO_2 affinity chromatography (10 mg beads) and effective depletion of PTAG peptides was established in less than 20 minutes at a 1:100 peptide-to-beads ratio (w/w). This is illustrated by the substantial sample simplification of a proteolytic digest of OMV fraction of *N. meningitidis* (Fig 3). After depletion of internal peptides, N-termini of several low abundant proteins were recovered as most predominant base peak traces in the chromatogram along with the high abundant N-terminus of PorA.

Evaluation of MS/MS-data revealed that PTAG-peptides with poor binding affinity were occasionally identified in the final sample (data not shown). However, identification of these peptides by data-dependent MS/MS analysis may not reflect the total fraction because identification is hampered by substantial neutral loss of phosphate during CID fragmentation [125]. For this purpose, D_0- and D_5-labeled analogues of the PTAG reagent were synthesized inhouse and employed to discriminate PTAG peptides by their unique 5-Da mass difference (doublets) from N-terminal peptides (singlets). In a single LC-MS/MS analysis of a N-terminally enriched OMV sample, approximately 5% of all

peptides ions were assigned as poorly retained PTAG peptides (Supplemental Fig. 5), thereby again demonstrating the high efficiency in depletion even for complex and high dynamic range proteome samples.

PEPTIDE MODIFICATIONS. The result of the PTAG strategy applied to a tryptic whole cell lysate of *N. meningitidis* is summarized in table 1. In total, 645 unique N-terminal dimethylated peptides were identified by offline ion exchange prefractionation (6 fractions) of the TiO_2 flow-through fraction and subsequent LC-MS/MS analysis. Of these N-terminal-dimethylated peptides, 423 peptides (312 proteins) were annotated as true protein N-terminal peptides with initiator methionine residues, methionine cleavage sites and proteins with signal peptide cleavage sites. Redundancy was especially observed for high abundant proteins. Multiple unique N-terminal peptides per protein were identified as a result of incomplete initiator methionine or signal peptide processing or unspecific C-terminal cleavage specificity of trypsin. The remaining 222 N-terminal-dimethylated peptides are so-called neo-N-termini derived from internal cleavage sites throughout the protein. Neo-cleavage sites may originate from unknown protolytic activity before (*in vivo*) or during harvesting (cell lysis), or from protein degradation during sample preparation [126].

TABLE 1. N-proteome analysis of tryptic *N. meningitidis* and tryptic *S. cerevisiae* using the PTAG strategy. The number of unique[a] N-dimethyl- or N-acetyl-modified peptides are listed. Dimethylated N-terminal peptides include (true) protein N-terminal peptides and neo-N-terminal peptides with unknown cleavage specificity from *e.g.* (*in vivo*) proteolytic activity. Only N-dimethylated peptides which are assigned to as (true) protein N-termini indicated by retained initiator methionine, methionine cleavage and proteins with signal/transit peptide cleavage are used for further data evaluation. The TiO_2 flow-through fraction was co-purified with N-unmodified (internal) peptides (lacking a PTAG due to incomplete derivatization), N-terminally cyclized peptides (N-pyroglutamate, N-pyrocarbamidomethyl cysteine and N-aspargine (P2)) and N-monomethylated peptides.

N-dimethylated or N-acetylated peptide modifications	*N. meningitidis* whole cells	*N. meningitidis* OMV fraction	*S. cerevisiae* whole cells
N-dimethylated peptides, incl. internal cleavages sites	645	897	776
N-dimethylated peptides assigned to protein N-termini	423	215	388
N-acetylated peptides assigned to protein N-termini			333
Number of proteins identified by N-terminal peptides	312	170	470
Co-purified peptide modifications			
N-unmodified (internal) peptides	45	154	187
N-pyroglutamate[b]	14 (12)	6 (5)	18 (17)
N-pyrocarbamidomethyl cysteine[c]	35	23	130
N-aspargine (P2) cyclized[d]	30	17	45
N-monomethyl[e]	56	164	95

[a] Peptide deamination and oxidation states are not considered unique.
[b] N-terminal glutamate was enzymatically cyclized by Qcyclase to N-pyroglutamate and subsequently cleaved by pGAPase generating an α-amine for PTAG and depletion. This enzymatic approach is especially inefficient for peptides with a peptidase inhibiting proline (P2) residue next to N-pyroglutamate (depicted between brackets).
[c] Cyclization of N-terminal carbamidomethylated (iodoacetamide) cysteine residues.
[d] N-terminal cyclization of the aspargine side chain is restricted for position 2 (P2) residues.
[e] N-terminal monomethylation of internal peptides as a consequence of residual activity of the reagent formaldehyde during trypsin digestion.

The high labeling efficiency of the method is underlined by the identification of only 45 unmodified internal peptides (free α-amine). These peptides lack a PTAG affinity label as a result of incomplete derivatization chemistry, but are detected in dramatically reduced abundance compared to their PTAG-labeled analogues (Supplemental Fig. 3). Cyclized N-termini are also enriched in the TiO_2 flow-through fraction because these peptides lack a free α-amine for PTAG derivatization. Especially N-pyroglutamyl residues were frequently identified during method development (data not shown). Staes *et al.* [112] introduced a method to generate a free α-amine (for subsequent labeling) by enzymatic cyclization of N-terminal glutamine (Qcyclase) and subsequent cleavage (pGAPase) of the formed N-pyroglutamyl moiety. This method was successfully implemented in the PTAG strategy because only 14 N-pyroglutamyl peptides were identified of which 12 peptides had a pGAPase inhibiting proline residue following the pyroglutamyl cleavage site. Furthermore, 35 cyclized peptides with a N-terminal, iodoacetamine-alkylated cysteine residue (pyrocarbamidomethyl cysteine) [112] and 30 peptides where the position 2 aspargine side chain is N-terminally cyclized (X-Asn-X motif) [127] were recovered in the flow-through fraction.

For proline starting proteins [128], the N-terminus (initiator methionine cleavage) is incorporated by a single methyl group by protective labeling with formaldehyde (N-monomethylation). Data evaluation revealed that 56 peptides were N-monomethylated. Several peptides could successfully be assigned as true proline starting protein N-termini, however the vast majority of peptides was identified as internal peptides. It appeared that, due to residual activity of the reagent formaldehyde, proteolytically generated internal peptides were (inefficiently) labeled at the N-terminus by a single methyl group, thereby blocking subsequent PTAG derivatization.

The peptide identification results for tryptic OMV fractions from *N. meningitidis* and tryptic *S. cerevisiae* whole cells are in general similar as discussed above (Table 1). As expected, also quite a few N-acetylated termini (333) were detected in *S. cerevisiae* and none in *N. meningitidis*. Of note, a high number of neo-N-termini from internal cleavage sites were found for OMV in comparison to whole cell lysates. This high number may have been resulted from proteolytic activity upon cell lysis to stimulate OMV release or from protein degradation during the OMV-purification procedure.

N-PROTEOME CHARACTERIZATION. The results of N-proteome enrichment of *N. meningitidis* and *S. cerevisiae* by the parallel use of trypsin, GluC and chymotrypsin are summarized in Fig 4. The results of the OMV fraction and the whole cell fraction of *N. meningitidis* largely overlap in terms of protein and peptide identifications (Supplemental table 1). These data sets were combined to reduce the number of single peptide identifications ('one hit wonders') and hence reduce the chance of false positive peptide and protein identifications. In total, 753 unique N-terminal peptides were identified for *N. meningitidis* (428 proteins) predominantly by tryptic samples (Fig 4). The majority of the proteins (75%) were assigned by multiple MS/MS spectra of the same peptide at different charge states, deamidation or oxidation states or, more importantly, overlapping peptides with different lengths (Supplemental table 1).

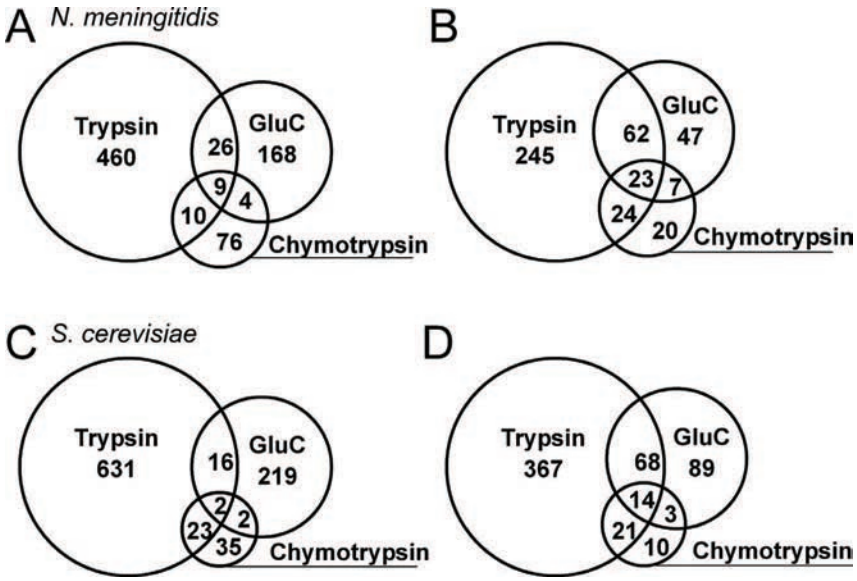


FIGURE 4. VENN DIAGRAM COMPARING THE NUMBER OF IDENTIFIED UNIQUE N-TERMINAL PEPTIDES (A, C) AND ASSIGNED PROTEINS (B, D) IN *N. MENINGITIDIS* (COMBINED OMV AND WHOLE CELLS FRACTION) AND *S. CEREVISIAE*, RESPECTIVELY, BY THE (PARALLEL) USE OF DIFFERENT PROTEASES IN THE PTAG PROTOCOL. Both in *N. meningitidis* and *S. cerevisiae* the largest number of confident sequence assignments was obtained for trypsin with little overlap between GluC and chymotrypsin. Parallel digestion with GluC added approximately 25% unique N-terminal peptides and a 15% increase in proteome coverage, while the additional use of chymotrypsin may be superfluous for global N-proteome analysis.

N-termini were assigned with retained initiator methionine, methionine cleavage sites and proteins with signal peptide cleavage (Fig. 5A). Analysis of the native protein N-termini provide a profile for *in vivo* specificity of methionine removal (Fig 5B). This process depends on the amino acid at position 2 and is preferentially associated with the small and uncharged residues serine, alanine, proline, glycine and cysteine [128]. Furthermore, 123 neo-N-terminal peptides (43 proteins) were identified for proteins of which the signal peptide was removed (Fig 5A). The secretory signal peptide that targets a protein for translocation across membranes is typically 14-30 amino acids long and is removed by a signal peptidase upon translocation [120]. Signal peptide cleavage sites of 43 identified proteins could be accurately verified by prediction software [120] because cleavage is governed by distinct zones (basic N-terminus, a hydrophobic region and a C-terminal region) with high sequence consistency around the cleavage site.

N-proteome analysis of *S. cerevisiae* resulted in the identification of 928 unique N-terminal peptides that could be assigned to 572 proteins (Fig 4). Approximately 45% of the proteins were fully or partially N-acetylated, preferentially at alanine, serine, threonine and methionine (Fig. 5C-D). Methionine cleavage specificity is conserved in prokaryotes and eukaryotes and preferential for small, uncharged residues at position 2 (Fig. 5E) [120]. In addition, 125 neo-N-terminal peptides (71 proteins) were identified for mitochondrial proteins of which the transit peptide was cleaved. Hallmarks for transit peptides cleavages are less well-described and predictable than those for signal peptides in prokaryotes [120]. The sequence consistency around the cleavage site is low, with arginine in position -3 or -2 relative to the cleavage site as the most common motif. For this purpose, accurate verification of peptidase cleavage site specificity was not performed by prediction [120], but instead a comparison was made with a recently published COFRADIC study. Vogtle *et al.* [121] characterized transit peptide cleavage specificity of an enriched mitochondrial protein fraction of *S. cerevisiae*. Identical cleavage sites were found for both positional proteomics strategies for the majority of the PTAG-identified mitochondrial proteins (46 out of 71 proteins).

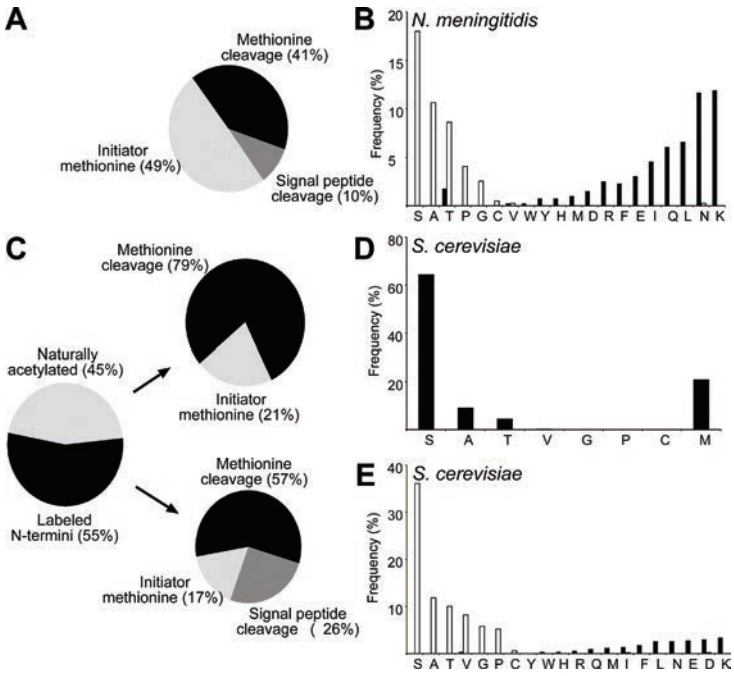


FIGURE 5. DISTRIBUTION OF N-TERMINAL MODIFICATIONS IN MATURE PROTEIN N-TERMINI. Protein N-termini in *N. meningitidis* are assigned as retained initiator methionine, methionine cleavage and proteins with signal peptide cleavage (A). Frequency distribution of the position 2 (P2) amino acid for removed initiator methionine (open bars) and retained methionine (black bars) (B). Methionine removal is preferred for small and uncharged residues at the P2 position. For *S. cerevisiae*, protein N-termini were identified as naturally acetylated (45%) and free N-termini, which are dimethylated in the PTAG protocol (labeled N-termini, 55%). From the pool of acetylated proteins, 79% showed the methionine cleavage while 21% retained the methionine residue. Frequency distribution of naturally acetylated amino acids at the protein N-termini (D). Labeled N-termini are annotated as retained initiator methionine, methionine cleavage and proteins with signal/transit peptide cleavage (C). Methionine cleavage (open bars) is similar to prokaryotes preferred for small and uncharged residues at the P2 position (E).

Recent N-proteome enrichment procedures suffered from underrepresentation of histidine and arginine containing N-terminal peptides as a result of charge retention on SCX pre-enrichment or side reactivity of histidine residues with NHS-activated sepharose beads [111, 112]. To investigate possible undersampling, the frequency distribution of the amino acids (P2) following the N-terminal methionine was calculated for the experimental data as the theoretical proteome (protein-coding genes). The experimental data correlates reasonably well with the frequency distribution in the complete proteome set for both *N. meningitidis* as *S. cerevisiae*, demonstrating that the obtained data are representative for the global N-proteome (Fig 6).

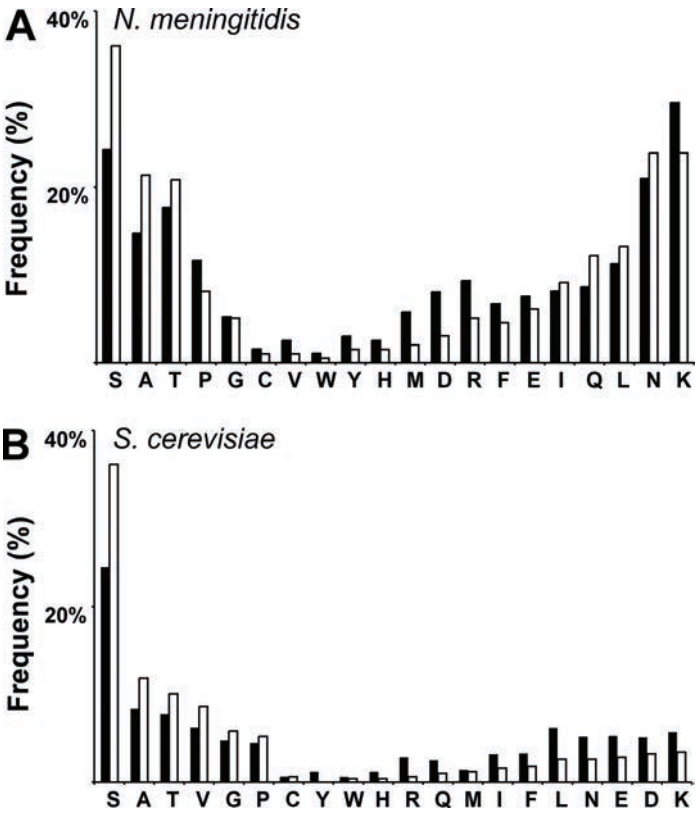


FIGURE 6. N-PROTEOME WIDE FREQUENCY DISTRIBUTION OF THE POSITION 2 (P2) AMINO ACID FOLLOWING THE INITIATOR METHIONINE. The frequency distribution calculated for the complete set of proteins (protein-coding genes) is depected in black bars and the frequency distribution for the experimentally identified protein N-termini is depicted in open bars. For both *N. meningitidis* (A) and *S. cerevisiae* (B), the experimental data correlates with theoretically data, indicating that the outcome the PTAG strategy is representative for the global N-proteome.small and uncharged residues at the P2 position (E).

DISCUSSION

In MS-based proteomics, complexity in protein samples greatly limits proteome coverage and identification of low abundant proteins [129]. Several advanced enrichment protocols have been introduced over the recent years to address this problem, including N-terminal positional proteomics strategies [106, 130]. Enrichment of N-terminal peptides by complete removal of internal peptides constitutes a major challenge [109]. PTAG was developed for specific labeling of free α-amines of internal peptides after proteolytic digestion. Subsequent, PTAG-labeled peptides were depleted by TiO_2 affinity chromatography (Fig. 1). High labeling efficiency was established, both at the protein level and at the peptide level, essential for minimizing the loss of N-terminal peptides. Dimethylation by formaldehyde was preferred because it is inexpensive, resistant to hydrolysis and full labeling of protein N-termini and lysine side chains was accomplished (Supplemental Fig. 2). In addition, the availability of stable isotopically labeled analogues of formaldehyde and cyanoborohydride enables multiplexed relative quantification in combination with PTAG [77, 107]. Internal peptides were selectively modified by the commercially available PTAG reagent glyceraldehyde-3-phosphate (GAP3). This reagent is similar to formaldehyde, resistant to hydrolysis and showed a high conversion yields (> 99%), thereby reducing possible background contamination of internal peptides (Supplemental Fig 3.). Excellent compatibility to reversed phase C18 ensured easy removal of salts and excess reagents prior to either TiO_2 affinity chromatography or direct LC-MS/MS analysis.

For the effective depletion of proteolytically generated internal peptides from a 100-µg proteome sample, a straightforward TiO_2-based depletion strategy was developed using in-house slurry-packed columns and including stringent washing conditions to minimize nonspecific adsorption [124] (Supplemental table 2). The ability to efficiently reduce sample complexity and identify a large number of proteins (207 proteins) in the Outer Membrane Vesicle (OMV) fraction of *N. meningitidis*, with dynamic range spanning several orders of magnitude, underlines the excellent recovery and selectivity of this strategy (Fig. 3). It should be noted that the final enriched sample fraction of this complex OMV digest is slightly biased by breakthrough of poorly retained PTAG peptides (5%). These peptides were assigned by employing stable isotopically labeled variants of the PTAG reagent (Supplemental Fig. 5 and 6).

N-proteome analysis of *N. meningitidis* and *S. cerevisiae* by the PTAG strategy resulted in the identification of 753 (428 proteins) and 928 (572 proteins) unique N-terminal peptides, respectively. Characterization of native protein N-termini provided a profile for *in vivo* specificity of methionine removal. This process is conserved in prokaryotes and eukaryotes and preferentially associated to small and uncharged residues (Fig 5) [131]. Furthermore, in *S. cerevisiae*, about 45% of the proteins were fully or partially N-terminally acetylated on alanine, serine, threonine or methionine (Fig. 5), which is a similar percentage as recently found by others [132]. N-proteome data also include neo-N-termini from subcellular-relocalized membrane (43 proteins, *N. meningitidis*) or mitochondrial proteins (71 proteins, *S. cerevisiae*). Accurate assignment of protease substrate cleavage sites could be problematic because these cleavage sites are typically not annotated in protein sequence databases [133]. Moreover, co-purification of neo-N-termini generated from background proteolytic activity with unknown specificity could result in the false positive identification of cleavage sites [107]. For this purpose, signal peptide and transit peptide cleavage specificity of the reported proteins were confirmed by either prediction [120] or by experimental data from a recent COFRADIC study [84].

Confidence in protein assignment may be problematic in N-proteome analysis as the majority of proteins are identified by a single peptide. Similar to other positional proteomics strategies [80, 134], prefractionation of the N-terminally enriched peptide fraction was performed to increase the number of MS/MS identifications per peptide and combined with strict database search criteria (FDR < 1%) to minimize the number false positive peptide identifications. Also, a number of proteases were used to generate N-terminal peptides with different lengths and hence not only enhance the confidence in protein assignment but also increase proteome coverage (Fig 3). It was estimated by *in silico* analysis that 50% of trypsin-generated N-terminal peptides could be analyzed by LC-MS/MS and parallel use of trypsin and GluC would increase coverage up to 80% [79, 109]. In practice, the increase in unique protein identifications for the use GluC in addition to trypsin is ~15% (Fig 3). This emphasizes one of the major challenges in N-proteomics: the identification of proteins by single N-terminal peptides with suitable LC-MS/MS characteristics, *e.g.* peptide length, hydrophobicity, ionization efficiencies and fragmentation properties. Alternative fragmentation techniques complementary to Collision Induced Fragmentation (CID) may be in particular beneficial in N-proteomics. A substantial fraction of the sample consists of long and highly charged N-terminal peptides with poor CID-fragmentation behavior. These long and highly charged peptides are generated by the ArgC digestion pattern of trypsin in combination with the remained ionic state of dimethylated lysine residues. Electron Transfer Dissociation (ETD) and Higher-energy Collisional Dissociation (HCD) have shown to be more effective for these highly charged peptides [59].

Recently, a number of large scale N-proteome studies for prokaryotes and *S. cerevisiae* were published. Timmer *et al.* [87, 110] characterized the N-proteome of *E. coli* by positive selection and identified 393 proteins, while Mc Donald *et al.* [79] identified more than 300 proteins in *E. coli* utilizing NHS-activated sepharose beads for depletion of internal peptides. N-proteome analysis of *S. cerevisiae* by COFRADIC in combination with SCX resulted in the identification of 650 unique N-terminal peptides [135]. Helbig *et al.* [136] characterized protein N-acetylation in *S. cerevisiae* with SCX and detected 756 N-acetylated protein termini, including acetylated neo-N-termini from internal cleavage sites. PTAG enabled the identification and characterization of 753 unique N-terminal peptides (428 proteins) in *N. meningitidis* and 928 unique N-terminal peptides (572 proteins), thereby representing one of the largest N-proteome datasets for these organisms so far. More detailed evaluations of *S. cerevisiae* N-proteome data provided by Helsens *et al.* [135], Helbig *et al.* [136] and N-proteome analysis by PTAG, revealed relative small overlaps between the studies (Supplemental Fig 6). The small overlaps are most likely due to undersampling of the full proteome or differences associated with the used method. For example, the study of Helbig *et al.* [136] was restricted to N-acetylated peptides while Helsens *et al.* [135] used SCX pre-enrichment with known undersampling of histidine containing N-terminal peptides [111, 112]. For N-proteome analysis by PTAG there is no indication of undersampling of certain amino acid sequences (Fig 6.). Of note, a clear disadvantage of the PTAG strategy is the inability to recover N-termini of proteins of which the N-terminal domain is phosphorylated by nature. N-terminal phosphorylation has for example been demonstrated in the p53 tumor suppression protein [137] and several crucial proteins of the photosystem II (PSII) in A. thaliana [138].

In conclusion, the PTAG positional proteomics strategy greatly reduces sample complexity, while maintaining the N-proteome fingerprint of whole cell lysates. The use of commercially available and highly reactive PTAG reagents in combination with high performance TiO_2 affinity chromatography provide a robust platform for global N-proteome analysis and MS-based profiling of proteolytic events. PTAG for unbiased selective isolation of protein N-termini is therefore a welcome alternative to well-established positional proteomics strategies.

ACKNOWLEDGEMENTS

We thank Martin Hamzink and Bert Zomer from the National Institute for Public Health and the Environment for the development and synthesis of the isotopically labeled PTAG reagents. We acknowledge Dr. Shabaz Mohammed from Utrecht University for helpful discussion. The Netherlands Proteomics Centre, embedded in the Netherlands Genomics Initiative, is kindly acknowledged for financial support.

ASSOCIATED CONTENT

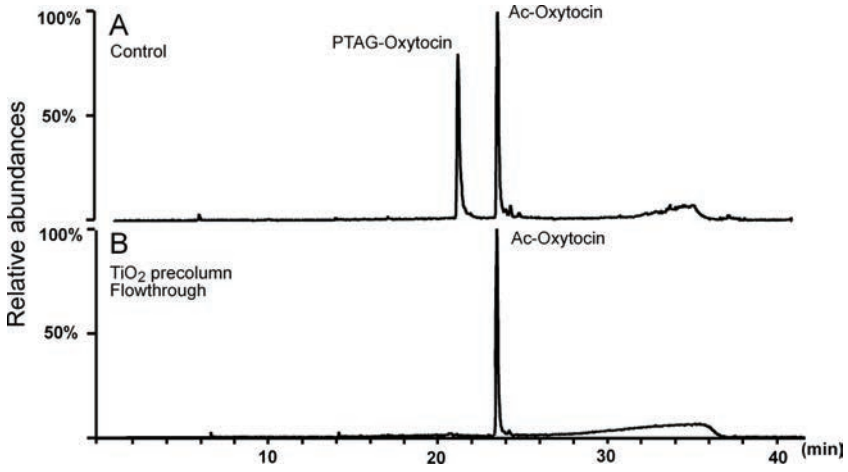
Additional information as noted in the text.

SUPPLEMENTAL TABLE 1: N-TERMINAL PEPTIDE ENRICHMENT BY THE PTAG STRATEGY OF THE *N. MENINGITIDIS* AND *S. CEREVISIAE* PROTEOME SAMPLES. Peptide identification data are obtained from the parallel use of trypsin, GluC and chymotrypsin. Mass different[a] and unique[b] N-terminal peptide identifications and assigned proteins are listed for proteins occurring in Outer Membrane Vesicle (OMV) fraction of *N. meningitidis* and whole cells lysates of *N. meningitidis* and *S. cerevisiae*. To enhance reliability in peptide and protein assignment in *N. meningitidis*, the largely overlapping data sets of the OMV and whole cells fractions were combined, thereby reducing the number of protein assignments based on a single peptide ('one hit wonders').

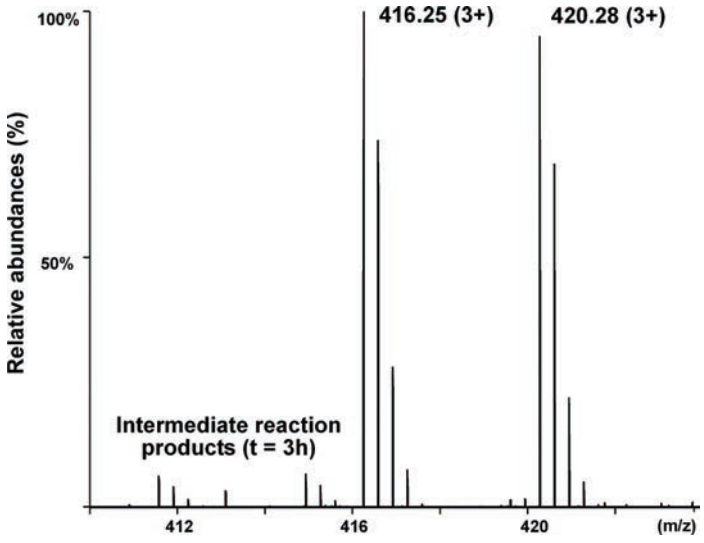
	N. meningitidis whole cells	*N. meningitidis* OMV fraction	*N. meningitidis* combined	*S. cerevisiae* whole cells
Mass different N-terminal peptides	695	337	835	1129
Unique N-terminal peptides	636	297	753	928
Proteins (true N-termini)	390	207	428	572
'One hit wonders'	31%	47%	25%	49%

[a] Mass different N-terminal peptides include deamination, oxidation and N-acetylation states

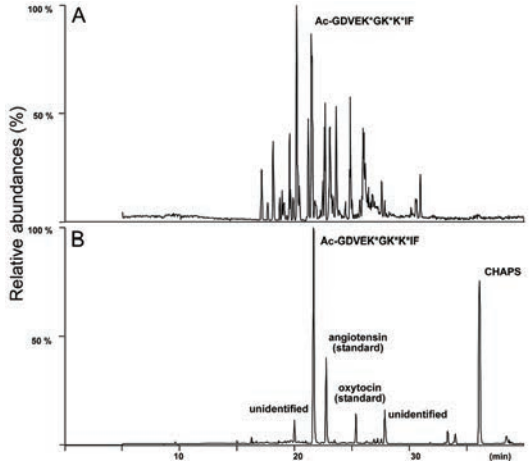
[b] Unique peptides include peptides with unique amino acid sequences (N-terminal peptides with different lengths).



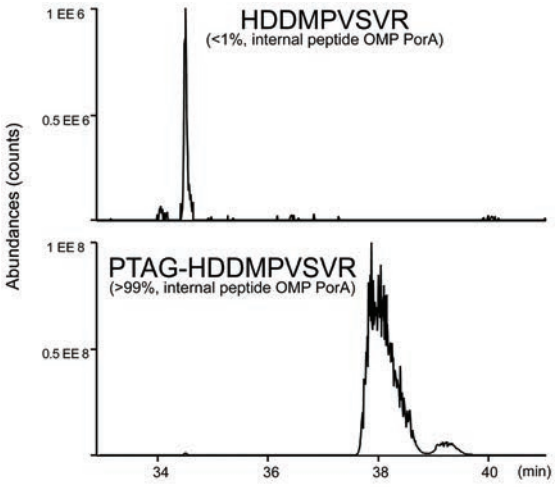
SUPPLEMENTAL FIG. 1. BASIC PRINCIPLE OF THE PTAG STRATEGY FOR ENRICHMENT OF N-TERMINALLY BLOCKED PEPTIDES. Mass trace chromatograms of a LC-MS analysis of an equimolar mixture of N-acetylated Oxytocin (Ac-Oxytocin) and PTAG-modified Oxytocin (PTAG-Oxytocin) (A). Enrichment of Ac-Oxytocin was accomplished by depletion of PTAG-modified Oxytocin through binding onto a TiO_2 precolumn (Pinkse, M.W. *et al*, J. Proteome Res., 2008) (B).



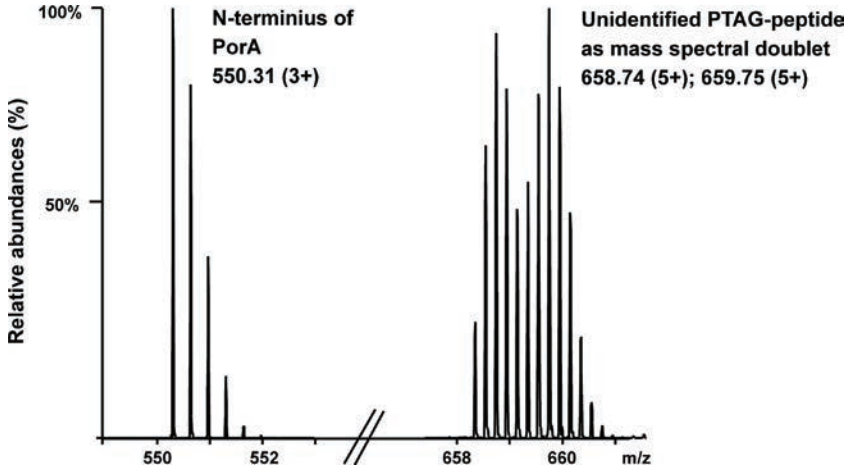
SUPPLEMENTAL FIG. 2. DIMETHYL LABELING EFFICIENCY AT THE PROTEIN LEVEL. After 3-hours of incubation of Cytochrome C with formaldehyde (180 mM) and cyanoborohydride, the naturally acetylated N-terminus of Cytochrome C (Ac-GDVEK*GK*K*IF) was observed predominantly as fully dimethylated by the triply charged peptide ions at m/z 416.25 and m/z 420.27 when using D_0-formaldehyde and D_2-formaldehyde, respectively. In case of a 3-hour incubation, intermediate reaction products of 14.01 Da lower in mass are detected (lacking one methyl group), implying incomplete dimethylation of one Lys side chains. By evaluation of the signal intensities, it is estimated that on average 5.9 methyl groups are incorporated on the N-terminal peptides of Cytochrome C (98%).



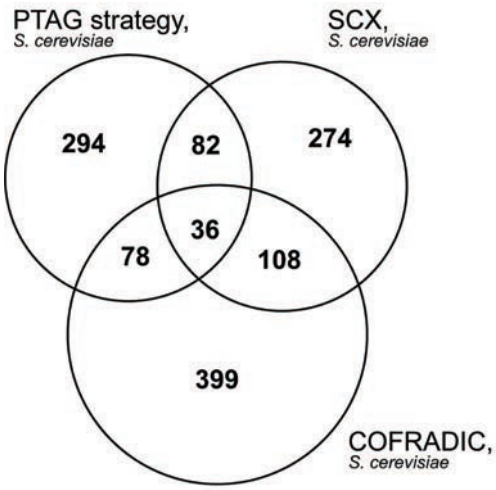
SUPPLEMENTAL FIG. 3. ENRICHMENT OF THE NATURALLY ACETYLATED N-TERMINUS OF CYTOCHROME C BY THE PTAG STRATEGY. Base peak chromatogram of the chymotrypsin-digested dimethyl-labeled Cytochrome C prior to enrichment (A). Enrichment of N-terminally blocked peptides by depletion of PTAG-modified internal peptides onto a TiO_2-based precolumn (Pinkse, M.W. *et al.*, J. Proteome Res., 2008) (B). The naturally acetylated N-terminal peptide of Cytochrome C (Ac-GDVEK*GK*K*IF) was recovered as the most predominant ion trace, while other ion traces correspond to unidentified molecules, angiotensin (spiked standard), oxytocin (spiked standard) and CHAPS (present as contaminant in DL-glyceraldehyde 3-phosphate).



SUPPLEMENTAL FIG. 4. EVALUATION OF PTAG LABELING EFFICIENCY BY COMPARING THE ABSOLUTE ABUNDANCES OF THE UNMODIFIED (A) AND PTAG-MODIFIED (B) INTERNAL PEPTIDE HDDMPVSVR FROM THE TRYPTIC-DIGESTED POR A PROTEIN IN THE OMV FRACTION OF *N. MENINGITIDIS*. PTAG labeling of a 100-µg tryptic proteome by overnight incubation with the PTAG reagent DL-glyceraldehyde 3-phosphate in addition of cyanoborohydride (after 0 h, 1 h and 2 h time points) yielded on average > 99% conversion. Despite this high labeling efficiency, unmodified residues (< 1%) can be detected at extremely reduced abundances.



SUPPLEMENTAL FIG. 5. ISOTOPICALLY LABELED PTAG-REAGENTS FOR ASSESSMENT OF POORLY RETAINED PTAG-LABELED PEPTIDES IN THE TIO_2 FLOWTHROUGH FRACTION. Alter initial blocking of protein N-termini and lysine side chains with D_0-formaldehyde, internal peptides were PTAG labeled with an equimolar mixture of D_0-oxobuthyl dihydrogen phosphate and D_5-oxobuthyl dihydrogen phosphate. After depletion by TiO_2 affinity chromatography, the N-terminus of porA appeared as singlet in the mass spectrum, while poorly retained PTAG labeled peptides were detected by their unique $\Delta M = 5.0$ Da mass spectral doublets.



SUPPLEMENTAL FIG 6. VENN DIAGRAM COMPARING N-PROTEOME DATA OF *S. CEREVISIAE* BY STUDIES PERFORMED USING THE PTAG STRATEGY (490 PROTEINS), SCX (500 PROTEINS) (HELBIG, A.O. *ET AL.*, BMC GENOMICS, 2010) AND COFRADIC (621 PROTEINS) (HELSENS, K. *ET AL.*, J PROTEOME RES., 2011), REVEALING A SMALL OVERLAP BETWEEN THE STUDIES. The comparison is restricted to annotated protein N-termini with initiator methionine or methionine cleavage, while (un-annotated) neo-N-termini from internal cleavage sites are discarded. Of note, the SCX study of Helbig *et al.* provided only N-acetylated termini, while PTAG and COFRADIC data include both N-acetylated and N-labeled termini.

CHAPTER 03

Quantitative proteomics reveals distinct differences in the protein content of outer membrane vesicle vaccines

Geert P.M. Mommen 1,2,3‡,
Bas van de Waterbeemd 1‡
Jeroen L.A. Pennings 1
Michel H. Eppink 4
René H. Wijffels 4
Leo A. van der Pol 1
Ad P.J.M. de Jong 1

‡ Both authors contributed equally
1 National Institute for Public Health and the Environment (RIVM), Vaccinology, The Netherlands
2 Utrecht Institute for Pharmaceutical Sciences and Bijvoet Center for Biomolecular Research, The Netherlands
3 Netherlands Proteomics Centre, The Netherlands
4 Wageningen University, Bioprocess Engineering, The Netherlands

Journal of Proteome Research, 2013. 12(4): p. 1898-08

INTRODUCTION

The use of outer membrane vesicles (OMV) is a promising approach for vaccine development against *Neisseria meningitidis* serogroup B, which causes acute and severe meningitis [139-143]. OMV consist of a phospholipid bilayer with outer membrane proteins, endotoxin and a lumen with periplasmic proteins [144, 145]. Outer membrane porin A protein (PorA) was identified as the immunodominant antigen in OMV, but is antigenically variable between circulating strains [146, 147]. To obtain a broadly protective vaccine, recombinant strains with multiple PorA subtypes were developed [148, 149]. Recent studies demonstrated that conserved minor antigens, like factor H binding protein (fHbp) or iron-regulated proteins, can complement PorA to further improve cross-protection [150-152]. In addition to well-described antigens the OMV proteome contains a considerable number of other proteins that may be relevant for immunogenicity [153].

The first OMV vaccines were prepared with detergent-extraction (DOMV) and have successfully stopped *N. meningitidis* serogroup B epidemics in several countries [141-143, 154-157]. The detergent-extraction was required to remove endotoxin, but also caused partially intact and aggregated vesicles [154, 158]. The lpxL1 mutation successfully attenuated meningococcal endotoxin and allowed a detergent-free process for vaccine development [159, 160]. A detergent-free process either uses extraction with a chelating agent to improve yield (native OMV; NOMV) [161-165], or no extraction at all (spontaneous OMV; SOMV) [166-168]. It was found that detergent-free OMV retain lipoproteins like fHbp, which improved cross-protection and functional immunogenicity in mice [150, 158, 169, 170]. These immunological properties of OMV vaccines were measured with SBA (serum bactericidal activity), which is an established correlate of protection in humans but does not provide in-depth information at the protein level [171, 172].

Proteomics has been used to assess the protein content of OMV in more detail. Initial studies on DOMV from *N. meningitidis* serogroup B used gel electrophoresis combined with LC-MS/MS peptide identification [166, 173-175]. One study revealed that DOMV and SOMV vaccines have a different protein content [169]. More recently, quantitative 2D gel analysis with fluorescent labeling identified a limited number of differential proteins in DOMV after growth on two media [176]. Gel-based proteomics however is labour intensive and less compatible with hydrophobic membrane proteins from OMV than a gel-free approach [177]. It also has a strong bias toward high abundant proteins [35]. To overcome these limitations, several gel-free quantitative proteomics methods have been developed [77, 103, 178]. Compared to other quantitative methods, multiplex dimethyl labeling of amino groups on N-termini and lysine residues is fast, robust and inexpensive [76, 77]. When dimethyl labeling is performed at the protein level, the N-terminal part of the protein (blocked α-amino group) can be selectively purified from the internal peptides after proteolytic cleavage (free α-amino groups) [79, 80, 85, 88]. Such a positional proteomics strategy strongly reduces sample complexity and uncovers low-abundant peptides, while preserving the original proteome fingerprint.

Positional proteomics successfully addressed dynamic range issues observed with complex proteomes that are dominated by a few proteins, like human plasma or OMV [179, 180]. Recently the PTAG strategy (phospho-tag) was developed for this purpose, using glyceraldehyde-3-phosphate reagent to derivatize free amino groups of internal peptides after proteolytic cleavage[181].

PTAG-modified internal peptides were depleted with efficient titanium dioxide (TiO_2) affinity chromatography, which purified protein N-terminal peptides for consecutive LC-MS/MS analysis. This PTAG strategy identified 572 unique proteins of *S. cerevisiae* and 428 unique proteins of *N. meningitidis* by their N-terminal peptides, representing one of largest N-proteome data sets available for these organisms. The study included 170 unique proteins from *N. meningitidis* NOMV vaccine, but dimethyl quantification was not implemented.

This work describes the first quantitative proteome comparison of detergent-extracted OMV vaccines with detergent-free alternatives. A novel quantification method was developed, based on dimethylation with stable isotopes and selective purification of N-terminal peptides. Distinct differences in protein content were observed, including several immunogenic proteins. These findings demonstrate that purification processes can change the protein content of OMV vaccines and support previously observed differences in functional immunogenicity.

MATERIALS AND METHODS

STRAIN, GROWTH CONDITIONS AND OMV PURIFICATION. The *N. meningitidis* vaccine strain that was used is a recombinant variant of isolate H44/76 (serogroup B) [182], combining one wild type and two recombinant PorA antigens (trivalent PorA; subtypes P1.7,16; P1.5-1,2-2 and P1.19,15-1) with a non-functional porB gene [148]. The cps locus was deleted, resulting in a non-encapsulated phenotype with galE-truncated LPS. Additional deletions in lpxL1 and rmpM genes were made to attenuate LPS toxicity and improve OMV yield, respectively [158]. Cultures were grown in chemically defined medium [183]. Erlenmeyer shake flasks with 150 mL medium were inoculated with 10 mL working seedlot (cells at OD_{590} = 1.5 ± 0.1; stored at -135°C with glycerol). Pre-culture shake flasks were incubated at 35°C, 200 rpm and 10 mL portions (OD_{590} of 1.5 ± 0.3) were used to inoculate secondary shake flasks. OD_{590} of the secondary flasks was monitored and bacteria were harvested for OMV purification after 5 hours of stationary growth. OMV vaccines were purified as described previously [158]. For OMV quality control, total protein concentration (>1.0 mg/mL; Lowry with Peterson's modification), PorA content (>50% of total protein) and vesicle size distribution (average size 70-110 nm; polydispersity index <0.20) were performed [161, 184]. As reported before, detergent-extraction caused aggregation of DOMV samples which resulted in high polydispersity. NOMV and SOMV vaccines did pass all quality criteria [158].

DIMETHYL LABELING AND N-PROTEOME ENRICHMENT. Protein was extracted from OMV vaccines with Trizol reagent (Invitrogen, The Netherlands). An OMV amount corresponding to 500 µg of total protein was used for each extraction. OMV sample volume was first reduced to 20-50 µL with a vacuum dryer before adding 500 µL Trizol. Protein was extracted according to manufacturer's protocol and the resulting pellets were stored at -80°C. Thawed pellets were dissolved in 50 µl buffer (100 mM KH_2PO_4, pH 7.5) containing 4M Guanidine-HCl. Total protein concentration was measured using the Lowry assay with Peterson's modification (sample was diluted at least 100-fold to prevent interference of guanidine with the assay). SDS-PAGE analysis confirmed comparable protein compositions before and after Trizol extraction (data not shown). Protein labeling, digestion and purification protocols (including the PTAG strategy) were described previously [181]. The dimethyl labeling strategy was adapted for relative quantification purposes. The common reference (equimolar mixture of all experimental protein samples) and the individual experimental samples were incubated with either light formaldehyde label (CH_2O) or heavy label (CD_2O; Sigma Aldrich), respectively, with addition of sodium cyanoborohydride (Sigma Aldrich) for 24 h at room temperature. Each heavy labeled sample was pooled individually with light labeled common reference in a 1:1 (w/w) ratio and digested by trysin under previously described conditions [181]. A sample aquilot was diluted 100 times in formic acid/DMSO in water (5/5/90, v/v) and stored at -20°C until further analysis. Peptide mixtures after trypsin digestion were PTAG derivatized and the N-terminal peptides were recovered/enriched in flow-through fraction of TiO_2 affinity chromatography. Samples were dried and reconstituted in formic acid/DMSO in water (5/5/90, v/v) and stored at -20°C until analysis.

PEPTIDE SEARCH LIST COMPILATION USING DATA-DEPENDENT LC-MS/MS. A search list containing entries with unique combinations of peptide sequence, accurate mass and retention time window was obtained by data-dependent LC-MS/MS analysis [185, 186]. Common reference sample aliquots were collected both before and after PTAG depletion and subsequently fractionated by

strong cation exchange (SCX). The common reference samples were loaded onto a biphasic 200µm ID trapping column packed with 20 mm 5 µm C18 resin (Reprosil Pur C18-AQ (Dr. Maisch, Ammerbuch-Entringen, Germany) and 20 mm 5µm SCX resin (Polysulphoethyl A, PolyLC Inc, Colombia, USA) in 0.1M HOAc at 5µL/min. Following RP-to-SCX transfer using 50% acetonitrille in 0.1M HOAc, peptide fractions were recovered by flushing the SCX bed step-wise with 2 µL plugs of potassium chloride in 0.1 mM acetic acid containing acetonitrile. Seven different acetonitrile concentrations were used in the plugs, ranging from 10 to 500 mM. Fractions were dried and reconstituted in formic acid/DMSO in water (5/5/90, v/v) and analyzed by LC-MS/MS. An LTQ-Orbitrap XL instrument (Thermo Fisher Scientific, The Netherlands) and Agilent 1100 HPLC system (Agilent, The Netherlands) was modified for nanoflow LC separations as previously described [187]. The trap column was a 100-µm ID fritted microcapillary packed with 20 mm, 5 µm particle size Reprosil Pur C18-AQ particles (Dr. Maisch, Germany). The analytical column was a 50-µm ID fritted microcapillary packed with 31 cm 3 µm particle size Reprosil Pur C18-AQ. The column effluent was directly electrosprayed into the MS using an in-house prepared, gold and conductive carbon coated fused silica tapered tip of ~2 µm (typically at 2.0 kV) [187]. Solvent A consisted of 0.1 M acetic acid in deionized water and solvent B of 0.1 M acetic acid in acetonitrile. Gradients were as follows: 100% solvent A during sample loading (0-10 min, flowrate 5 µL/min), 7% to 26% solvent B in 160 minutes followed by an increase to 60% solvent B in 20 minutes and reconditioning with solvent A for 10 min (total runtime 200 min). The mass spectrometer was set to acquire full MS spectra (*m/z* 350 to 1,500) for mass analysis in the orbitrap at 60,000 resolution (FWHM) followed by data-dependent MS/MS analysis (LTQ) for the top 7 abundant precursor ions above a threshold value of 10^4 counts. The normalized collision energy was set to 35%, isolation width to 2.0 Da, activation Q to 0.250 and activation time to 30 ms. The maximum ion time (dwell time) for MS scans was set to 200 ms and for MS/MS scans to 2500 s. Charge state screening and preview mode were enabled. Precursor ions with unknown and +1 charge states were excluded for subsequent MS/MS analysis. Dynamic exclusion was enabled (exclusion size list with 500 entries) with repeat set to 1 and an exclusion duration of 180 s. The background ion at *m/z* 391.284280 was used as lock mass for internal mass calibration. Analysis of LC-MS/MS raw data was carried out with Proteome Discoverer 1.2 software (Thermo Fisher Scientific, Bremen, Germany) applying standard settings unless otherwise noted. MS/MS scans were searched against the *N. meningitidis* serogroup B database (Uniprot Knowledgebase, July 18th 2012) using SEQUEST. Precursor ion and MS/MS tolerances were set to 10 ppm and 0.8 Da, respectively. Decoy database searches were performed with False Discovery Rate (FDR) tolerances set to 5% and 1% for modest and high confidence filtering settings, respectively. The data were searched with full trypsin cleavage specificity, allowing 5 miss-cleavages (lysine cleavage is prevented by dimethyl modification). Cysteine carbamidomethyl and lysine dimethylation were set as fixed modifications, while aspargine deamidation and methionine oxidation were set as variable modifications. An additional search was performed using C-terminal trypsin cleavage specificity and implementation a fixed N-terminal dimethyl modification. High confidence peptide sequence identifications (Xcorr values >2.2, false discovery rate <1%, rank No.1) were exported to an Excel data file (the peptide search list). Raw data files and protocols associated with this manuscript are available for the reader upon request.

HIGH RESOLUTION PEAK QUANTIFICATION USING LC-MS. Quantification experiments of light (common reference) and heavy (experimental sample) labeled peptide mixtures were performed on the full MS level, relying on the high mass accuracy of the Orbitrap (data-dependent MS/MS disabled). All analyses were performed with the same nano LC column and identical gradient conditions as described above for maximal chromatographic reproducibility. Light/heavy peptide mixtures prior to PTAG-labeling and TiO_2 purification (15 samples) and N-terminally enriched peptide mixtures after TiO_2 purification (15 samples) were directly analyzed by LC-MS.

DATA PROCESSING FOR ANALYSIS OF RELATIVE PROTEIN ABUNDANCE. Relative peptide abundances were determined by using the high resolution LC-MS data in combination with the precompiled peptide search list (accurate mass and retention time approach) [185, 186]. Raw LC-MS data files were deconvoluted to mono-isotopic masses to minimize isobaric interference of contaminants using Xcalibur software (Thermo Scientific; XtractAll plugin; MH+ mode; S/N threshold = 2)[188]. The output was saved in NetCDF format (Thermo file converter) and imported in MS-Xelerator (MSMetrix, The Netherlands). The MSX-Quant plugin of this software traced and extracted the peak areas of all peptide entries of the pre-compiled Excel search list using accurate mass (±0.01 Da) and retention time (±10 min) information. Light/heavy peak area ratios were calculated from extracted ion chromatograms. Further data processing was performed in R (http://www.R-project.org). Individual MSX-Quant result files were compiled and stringent quality criteria were applied. A peptide search list entry was positively identified when both light and heavy labeled chromatographic peaks were extracted for at least 3 out of 5 independent biological replicates of at least 1 OMV vaccine type at an intensity threshold of $>10^4$ counts (above background noise). In addition, all chromatographic peaks were required to have an accurate mass deviation of <10 ppm and retention time window ±10 min compared to the search list entry. Starting with entries that were positively identified in any OMV vaccine type, it was determined whether or not these entries were identified in each of the three OMV vaccine types, by using the same quality criteria. The resulting lists of positively identified entries were mapped to corresponding peptides and proteins. The overlap between the respective lists per OMV vaccine type were visualized as Venn diagrams.

STATISTICAL ANALYSIS OF RELATIVE PROTEIN ABUNDANCE. For quantification of relative protein abundance, the light/heavy ratio values of positively identified peptide search list entries were 2log transformed. For entries with a positive identification (found in ≥3 out of 5 replicates) for an OMV type, missing values (due to quality issues) were imputed as the mean of detected values. OMV types with a negative identification for a search list entry (found in ≤2 out 5 replicates) were imputed with background values. Background values were calculated from the mean of the minimal value for that entry across all experimental samples and the minimal value for that experimental sample across all search list entries. This ensured that both sample and entry-dependent background levels were taken into account. Next, log-transformed values of search list entries that belonged to the same peptide were averaged, followed by averaging of values from peptides that belonged to the same protein. The resulting protein expression data was analyzed across all OMV types with one-way ANOVA. This identified proteins that were differentially expressed between any of the OMV types (maximal Foldratio >2 and $p<0.001$). This corresponded to a False Discovery

Rate (FDR) of <1%. The expression values of differentially expressed proteins were visualized as a heat map using Genemaths XT (Applied Maths, Belgium). Functional annotations and keywords of identified proteins were adopted from Uniprot Knowledgebase, *N. meningitidis* strain MC58 (http://www.uniprot.org). A few entries were substituted with homologs from strain H44/76 or strain Alpha710, due to an outdated MC58 annotation. Predicted cellular locations were obtained with the PSORTb algorithm (http://www.psort.org/psortb) [189].

SERUM BLOT PROTEOMICS. Female BALB/c mice were immunized subcutaneously on day 0 and 28 with either DOMV, NOMV or SOMV vaccine and sera were collected on day 42, as previously described [158]. SDS gel electrophoresis was performed with 4 µg total protein of each OMV vaccine per lane [158]. Gels were either stained with Novex Collodial Blue (Invitrogen, U.S.A.) or blotted to nitrocellulose membranes and blocked with buffer containing 0.5% Protifar (Nutricia, The Netherlands). Blot membranes were incubated with pooled mice sera after immunization with the corresponding OMV vaccine (200x diluted), with monoclonal antibody against fHbp (NIBSC, United Kingdom) or with PorA P1.19 monoclonal antibody (RIVM, The Netherlands). Goat-anti-mouse antibody conjugated to alkaline phosphatase was used as a secondary antibody for staining. The protocol for direct surface digestion of blotted proteins was adapted from Luque-Garcia *et al.* [190]. Differential serum bands were excized and incubated in 100 mM KH_2PO_4 (pH 7.5), 10% acetonitrille with trypsin at an enzyme/protein ratio of 1/20 (w/w) at 37°C for 16 h. Peptide mixtures were purified with RP/SCX solid phase extraction as described above (peptide search list compilation paragraph). LC-MS/MS conditions and data analysis were also similar as above, however with 90 min. LC gradients (7%–40% solvent B in 80 min). Cleavage specificity was set to full trypsin allowing 2 miss-cleavages. Aspargine deamidation and methionine oxidation were set as variable modifications. At least 2 unique, high-confidence peptides were required for protein identification and large deviations between theoretical and observed protein molecuar weight on the immunoblot were not tolerated. Protein identifications that matched the immunoblot pattern were verified manually, to confirm presence of the corresponding tryptic peptides in the raw chromatographic data.

RESULTS

DEVELOPMENT OF QUANTITATIVE PROTEOMICS METHOD. A novel quantitative proteomics method was developed based on duplex dimethyl labeling, combined with unbiased selection of N-terminal peptides (PTAG; phospho tagging) [181]. A detailed workflow is shown in Figure 1. Relative quantification was based on comparison of 15 experimental samples (3 different OMV vaccines; 5 biological replicates per vaccine) against a common reference (CR). The CR contained equal amounts of all experimental samples. Samples for high-throughput quantification were collected after proteolytic digestion (all dimethylated peptides) and after PTAG depletion (N-terminal peptides). Quantification was performed with a two-step approach, using stringent quality criteria (Figure 2) [185]. The CR allowed an unrestricted number of biological replicates despite inherent multiplex restrictions of the dimethyl labeling (either duplex or triplex [76, 77]). As a result, each experimental group (OMV vaccine) had 3–5 independent measurements per peptide, allowing robust statistical analysis. Prior to OMV proteome analysis, the quantification method was successfully validated with a standard protein mixture containing known ratios (S-Table 1, Supporting Information).

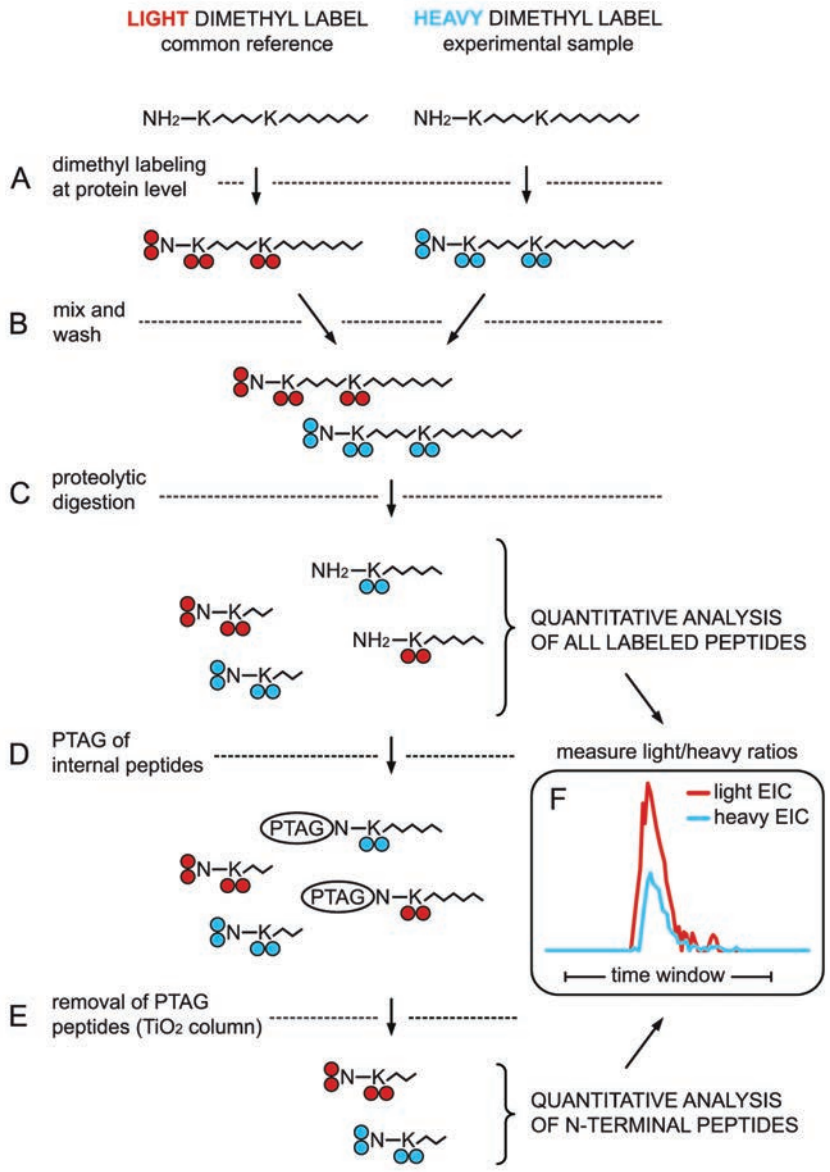


FIGURE 1. WORKFLOW FOR PURIFICATION OF DIMETHYLATED N-TERMINAL PEPTIDES WITH PTAG. (A) First, the common reference is mixed from equal amounts of all experimental samples. Free N-terminal and lysine amino groups of the common reference are dimethylated at the protein level with formaldehyde (CH_2O, light label). In parallel, experimental samples are dimethylated with heavy formaldehyde (CD2O isotope, heavy label). (B) Light and heavy labeled samples are mixed and washed with acetone precipitation. (C) Proteolytic digestion with trypsin generates internal peptides with free N-terminal amino groups. (D) Internal peptides are derivatized on the N-termini with PTAG reagent. (E) PTAG peptides are captured with TiO_2 columns for selective purification of dimethylated N-terminal peptides. (F) Samples are collected after step C (all dimethylated peptides) and after step E (N-terminal peptides) for quantification of light/heavy peak ratios with nano LC-MS (Figure 2).

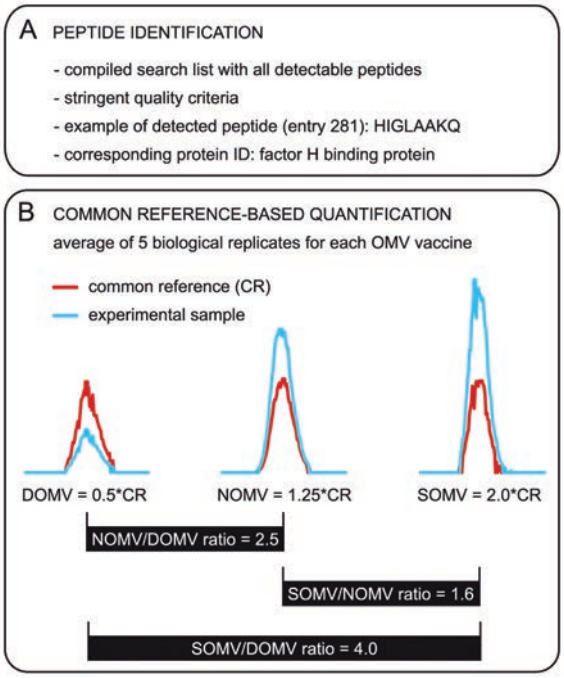


FIGURE 2. COMMON REFERENCE-BASED PEPTIDE QUANTIFICATION USING STRINGENT QUALITY CRITERIA. (A) High-throughput peptide identification was done with a two-step approach. First, the common reference (CR) was analysed to compile a search list with accurate mass, retention time window and sequence ID of all detectable peptides[185, 186]. Second, full LC-MS chromatograms were acquired for all light/heavy mixtures. The light (CR) peptides were traced with the peptide search list. Corresponding heavy (experimental) peptides were matched based on calculated mass and retention time. Stringent criteria were required for a positive identification (see Material and Methods section). Search list entries that complied to these criteria nearly always provided a unique hit. (B) Peak areas were extracted for quantification of light/heavy ratios. The CR was used as an internal standard to calculate ratios between experimental samples, as illustrated for search list entry 281, a peptide from factor H binding protein. This peptide had a low expression in DOMV compared to NOMV (2.5-fold higher) and SOMV (4.0-fold higher). Each detectable peptide in the search list was quantified with 3–5 independent biological replicates, allowing robust statistical analysis.

OMV PROTEIN IDENTIFICATION. The proteome analysis was performed on OMV vaccines from three different purification processes [158]. A total of 618 unique peptides were positively identified, 400 peptides in DOMV (detergent extraction), 509 peptides in NOMV (detergent-free extraction) and 497 peptides in SOMV (no extraction), with high reproducibility but moderate overlap between the three vaccines (282 peptides; Figure 3A). Merging of peptides with identical accession numbers yielded 185 unique proteins, of which 76 proteins were shared between the three OMV vaccines (Figure 3B). These results were obtained after merging of datasets from internal peptides and N-terminal peptides. Added value of the PTAG strategy was evaluated by comparing the contribution of each dataset to the total number of detected peptides (Figure 3C) and proteins (Figure 3D). At the peptide level, the PTAG strategy resulted in detection of 86 unique N-terminal peptides, of which 62% was not detected before this procedure (53 N-terminal peptides). These 53 additional N-terminal peptides however accounted for only 7 additional proteins. This indicates that most of the N-termini found with the PTAG strategy belonged to proteins that were already detected by their internal peptides (before PTAG).

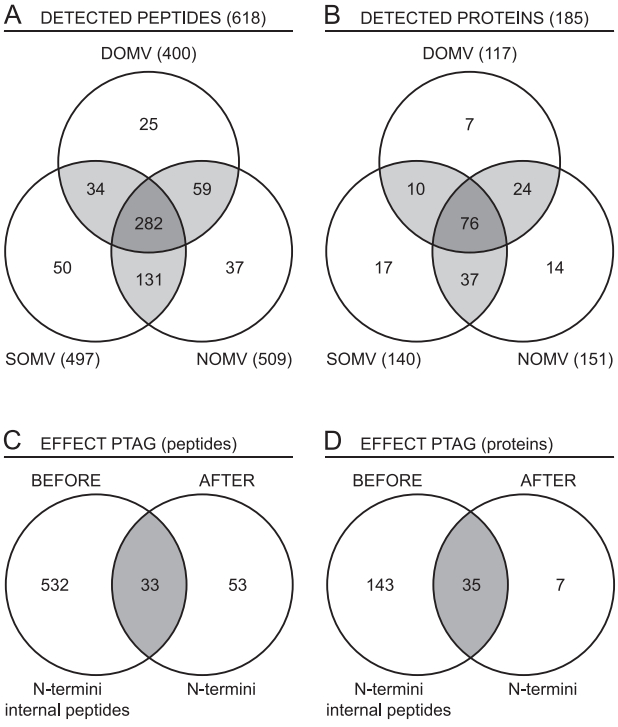


FIGURE 3. VENN DIAGRAMS OF IDENTIFIED PEPTIDES AND PROTEINS. (A) Peptide overlap between DOMV vaccine (detergent-extraction), NOMV vaccine (detergent-free extraction) and SOMV vaccine (no extraction). A moderate overlap of 282 out of 618 unique peptides was observed. (B) The peptides corresponded to 185 unique proteins, of which 76 were shared. (C) Assessment of the added value of PTAG depletion at the peptide level. A total of 53 N-terminal peptides out of 86 (62%) were were uncovered by the PTAG procedure. (D) The additional N-terminal peptides represented only 7 additional proteins, since most proteins were already detected by their internal peptides. This indicates that OMV samples may not have sufficient complexity to challenge the PTAG procedure.

DOMV VACCINES CONTAIN CYTOPLASMIC PROTEINS AS A RESULT OF LYSIS. Relative protein abundance was quantified to identify differences between the OMV vaccines. Proteins were clustered in 7 groups based on their expression profile (Figure 4). Group VII contained 39 proteins that were detected without significant differences in relative protein abundance (therefore not shown in Figure 4). This group included several well-known membrane proteins like PorA, OpcA and FetA [191, 192]. Groups I to VI each had a specific expression pattern (*i.e.* downregulated in DOMV). Functional annotation revealed interesting differences. The 146 differentially expressed proteins in group I to VI together had 33 ribonucleoprotein annotations (from Uniprot keywords) of which 17 were found in group II (upregulated in DOMV; enrichment p-value <0.05). Group II was also significantly enriched in RNA-binding (14/23 annotations), nucleotide binding (9/17), cytoplasm (9/16) and transferase functions (7/12). These proteins originate from the cytoplasm and may be present exclusively in DOMV as a result of lysis, an undesired side-effect of detergent extraction. Group III (upregulated in DOMV and NOMV) confirmed this hypothesis because it was also significantly enriched in ribonucleoprotein (12/33) and RNA-binding functions (7/23). This group however contained less proteins since NOMV purification (mild, detergent-free extraction) indeed caused less lysis than DOMV, but clearly more than SOMV (no extraction).

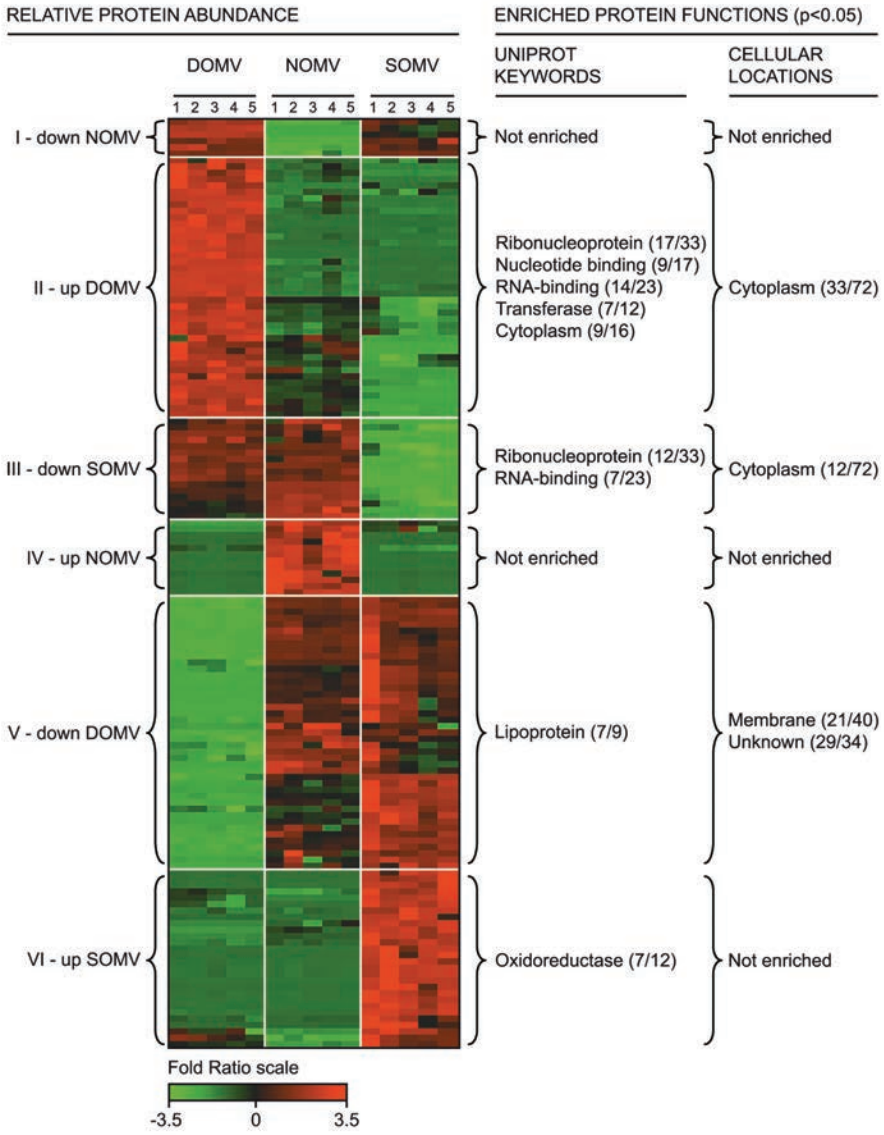


FIGURE 4. DIFFERENTIALLY EXPRESSED OMV PROTEINS. Quantitative proteome analysis was performed on OMV vaccines from 3 different purification processes, using 5 biological replicates per vaccine. Proteins were clustered in expression pattern groups, *i.e.* group V (downregulated in DOMV). The color scale covers 3.5-fold downregulation (green), via no regulation (black) to 3.5-fold upregulation (red). Functional anntation was based on Uniprot keywords and predicted cellular locations (PSORTb algorithm). Enriched protein functions are depicted for each expression group ($p<0.05$). The results demonstrate that OMV vaccines from biomass extraction processes (especially DOMV and to a lesser extent NOMV) are enriched with lysis-derived cytoplasmic proteins, including proteins with ribonucleoprotein, RNA-binding or nucleotide binding function. NOMV and SOMV vaccines (detergent-free processes) are enriched with membrane proteins and proteins with unknown location, including lipoproteins like fHbp. Therefore detergent-free OMV vaccines, SOMV in particular, have a preferred protein composition.

DETERGENT-FREE OMV VACCINES ARE SIGNIFICANTLY ENRICHED WITH MEMBRANE (LIPO) PROTEINS. The lysis-related results were further substantiated with cellular location predictions (PSORTb algoritm [189]), which confirmed that groups II and III indeed were enriched with cytoplasmic proteins. Group V (upregulated in detergent-free NOMV and SOMV vaccines) however was enriched with proteins that had a membrane location (21/40 annotations) or unknown location (29/34 annotations). Also, this group contained 6 out of 8 proteins with a lipoprotein annotation (all enrichment p-values <0.05). Proteins with such locations or functions are more likely to be relevant for the immunogenicity of OMV vaccines than lysis-derived cytoplasmic proteins. Notably, several oxidoreductases were found specifically in SOMV vaccines (group VI; 7/12 annotations). In addition to these redox-related proteins, SOMV vaccines contained proteins with non-enriched functions that are involved in iron uptake (*i.e.* heme utilization protein and bacterioferritin BfrA, which is regulated in response to iron availability [193, 194]). Since SOMV are spontaneously released and most similar to natural OMV, these virulence-related proteins may be relevant for the pathogenicity of *N. meningitidis* [195]. Detailed protein information with expression data and functional annotation is provided in S-Table 2 (Supporting Information). An overview of the PSORTb location distribution per OMV vaccine is shown in S-Figure 1 (Supporting Information).

SERUM BLOT PROTEOMICS SUPPORTS THE QUANTITATIVE PROTEOMICS RESULTS FOR SEVERAL IMMUNOGENIC PROTEINS. Serum blot proteomics was performed to identify immunogenic proteins in the different OMV vaccines. Figure 5A shows the overall protein composition of the OMV vaccines after 1D gel electrophoresis, visualizing the major contributions of PorA (41 kD), Omp85 (85 kD) and several proteins at 27 kD (*i.e.* OpcA protein). As discussed above, these proteins were found in all OMV types without significant expression differences. Next, mice sera after two vaccinations were used for immunoblotting against corresponding OMV vaccines. Despite a heavy immuno-dominance of PorA at 41 kD the serum blots revealed clear differences in immunogenicity, especially for DOMV compared to NOMV and SOMV (Figure 5B). This indicated that the different OMV processes retained or removed specific immunogenic proteins. These immunogenic proteins were not visible with 1D gel electrophoresis, therefore they represent a small but potentially important portion of the total protein content.

Immunoblot bands with a differential pattern between OMV sera were excized and separately digested with trypsin, for qualitative LC-MS/MS analysis (serum blot proteomics). The analysis identified several proteins that matched the blot pattern (Figure 5C). At 41 kD, PorA was found abundantly on all blots. Lipoprotein E6MY46 (14 kD) was found on NOMV and SOMV blots, while LptD (solvent tolerance protein; 85 kD) was unique for DOMV. In other sections no matching proteins were detected (100 kD, 50-70 kD and 15-20 kD) despite a differential serum blot pattern. Around 27 kD all serum blots had a visible band, but the bands for NOMV and SOMV were clearly more pronounced. Serum blot proteomics of this 27 kD band provided several interesting matches like PilQ, a protein involved in type IV pilus biogenesis and only found in NOMV and SOMV. Proteins Q7DDR0 (putative cell-binding factor) and Q9JYQ6 (uncharacterized protein) were found in NOMV and SOMV, and only in SOMV, respectively, but have not been described in literature. Two important identifications on the 27 kD blot section were fHbp and a C-terminal PorA fragment cleaved-off at the Pro268 position. These two immunogenic proteins were verified with monoclonal antibodies

(Figure 5D). Detergent-extraction during DOMV purification is known to remove lipoproteins like fHbp [150, 167]. The anti-fHbp immunoblot confirmed that fHbp indeed was absent in DOMV. The blot with anti-PorA monoclonal confirmed an overall strong band at 41 kD and confirmed the C-terminal PorA fragment. This PorA fragment explains the weaker 27 kD band on the DOMV serum blot, while the presence of fHbp explains the stronger 27 kD bands on NOMV and SOMV serum blots. Notably, the serum blot identifications were in full agreement with the quantitative proteomics results. Only LptD was somewhat ambiguous. LptD was 5-fold higher in DOMV which matched the serum blot results, but upregulation was not significant (p-value just above threshold).

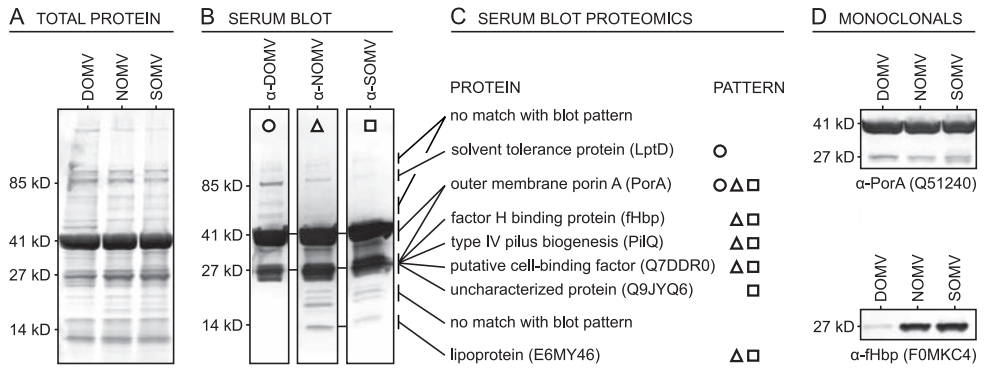


FIGURE 5. SERUM BLOT PROTEOMICS REVEALS DIFFERENCES IN IMMUNOGENIC PROTEIN CONTENT. (A) Protein composition of OMV from three purification processes after SDS gel electrophoresis. (B) Sera of immunized mice were blotted against corresponding OMV vaccines. Despite a comparable SDS page pattern, the serum blots reveal differential immunogenicity (several variable bands at 50–85 kD and 14–27 kD) and confirm immuno-dominance of PorA (41 kD). (C) Serum blot proteomics identified proteins that matched the differential pattern (circle: found in DOMV; triangle: NOMV; square: SOMV). This demonstrates that NOMV and SOMV vaccines (detergent-free process) are enriched with immunogenic proteins. (D) The results were verified with monoclonal antibodies against two protective antigens, PorA and fHbp. This confirms that PorA immunogenicity is not dependent on the purification process, while fHbp lipoprotein is largely removed after detergent-extraction (DOMV process).

DISCUSSION

To investigate previously observed changes in functional immunogenicity, the proteomes of OMV vaccines from different purification processes were compared and quantified [158]. Dimethyl labeling of free amines at the protein level was followed by selective purification of N-terminal peptides wih the PTAG strategy[181]. This novel protocol combined the analysis of internal peptides before PTAG and N-terminal peptides after the PTAG procedure, using a straightforward workflow (Figure 1). Dimethylated peptides were quantified relative to a common reference, which represented an internal standard between all experimental samples (Figure 2). Strong cation exchange (SCX) fractionation followed by LC-MS/MS analysis was only required to compile a peptide search list, linking accurate mass and retention time to peptide and protein identifications [185, 186, 196]. Once the search list was available, a single LC-MS run for each experimental sample was sufficient to identify and quantify all detectable peptides in the OMV proteome. This accurate mass and retention time approach reduced total acquisition time and allowed a high number of independent biological replicates, resulting in robust statistical analysis of relative protein abundance. These advantages are not available with other quantative proteomics methods, which mainly rely on weighed peptide fold-changes in a single biological sample for statistics [178].

The PTAG strategy was included to reduce sample complexity and address the dynamic range issues expected for PorA-dominated OMV proteomes [181]. This allowed detection of 618 peptides from 185 unique OMV proteins (Figure 3). The majority of N-terminal peptides was not detected before PTAG depletion, indicating that this method indeed uncovered additional peptides. OMV sample complexity however was too low to take full advantage of PTAG. This is illustrated by the fact that the uncovered N-terminal peptides accounted for only 7 additional proteins, since most proteins were already detected by their internal peptides. PTAG apparently had little added value for quantitative analysis of OMV samples but the procedure is likely to be beneficial for highly complex protein samples like cell lysates. Other studies identified between 25 and 166 OMV proteins with qualitative methods, like 1D or 2D gel electrophoresis and one gel-free approach [169, 173, 174, 177]. A quantitative 2D electrophoresis study identified 74 OMV proteins, of which 10 were differentially expressed [176]. Overlap of these studies with the current results was moderate to high (S-Table 3, Supporting Information), therefore this study is representative and comprises the largest quantitative OMV dataset available to date.

Quantitative analysis of relative protein abundance revealed distinct differences between OMV vaccines (Figure 4). Several proteins were absent in one or two vaccine types, resulting in large fold ratios (>10). The dataset also contained a variety of subtle changes that would remain unnoticed with a qualitative approach. DOMV vaccines contained substantially more cytoplasmic proteins than NOMV or SOMV vaccines. These proteins may be released during detergent-extraction, which removes endotoxin from DOMV but apparently caused lysis and contamination with cytoplasmic proteins. A smaller but distinct set of cytoplasmic proteins was shared between DOMV and NOMV vaccines, while SOMV remained largely free of cytoplasmic contamination. This supported the extraction hypothesis, since NOMV are purified with a milder, detergent-free extraction and SOMV purification does not require any extraction at all. Other studies also found cytoplasmic proteins in DOMV, including one qualitative comparison of DOMV and SOMV from the New-Zealand vaccine

strain[169, 174-177]. The present study added an intermediate vaccine (NOMV) to the comparison, which demonstrated that cytoplasmic protein contamination of OMV is directly related to the purification process that was used.

In addition to less cytoplasmic contamination, detergent-free OMV vaccines were significantly enriched with membrane proteins and proteins with an unknown cellular location. For SOMV vaccines, the difference was most pronounced (S-Figure 1, Supporting Information). Such proteins are more likely to be important for the immunogenicity of OMV vaccines than cytoplasmic proteins, which predominantly represent well-characterized aspects of microbial metabolism. The importance of unknown proteins is illustrated by the observation that 6 out of 9 lipoproteins in the dataset had an unknown predicted location rather than a membrane location. The data also showed that most lipoproteins were selectively removed after detergent extraction. This included factor H binding protein (fHbp), which contributes to functional immunogenicity and is currently investigated in clinical trials as a purified protein vaccine [170, 197, 198]. SOMV vaccines contained an additional set of proteins with virulence-related functions. Therefore detergent-free OMV, in particular SOMV, have a preferred protein composition compared to DOMV.

The detergent-free OMV vaccines in this study previously had improved cross-protection and functional immunogenicity in mice [158], which was confirmed by others [150, 169, 170]. Functional immunogenicity is measured with SBA, which correlates to protection in humans but does not provide in-depth information on the protein content of vaccines. Therefore this study supplemented SBA results with quantitative proteomics data. The results were refined with serum blot proteomics to visualize and assess differences in immunogenic protein content. Differential serum blot bands were excized, digested with trypsin and analyzed with qualitative LC-MS/MS. Several antigens matched the immunogenicity pattern on the serum blot, including PorA (constant pattern) and proteins that were only found exclusively in detergent-free OMV like fHbp, type IV pilus assembly protein PilQ or putative cell-binding factor (Figure 5). Notably, the pattern of serum blot identifications was in full agreement with the quantitative proteomics results, indicating that both novel methods produced reproducible data. Serum blot proteomics demonstrated that detergent-free OMV are enriched with immunogenic proteins, but this does not necessarily translate to functional immunogenicity [199]. Differential endotoxin content of OMV vaccines should also be taken into account, since attenuated lpxL1 endotoxin in detergent-free OMV vaccines adjuvates the immune response against proteins in the vaccine [159, 200]. Also the endotoxin itself can contribute to functional immunogenicity [162]. Even though these aspects are not covered with proteomics, the current results provided valuable insight in the immunogenic protein content of different OMV vaccines. Such an approach can support functional assays like SBA.

CONCLUSIONS

A novel proteomics method was developed, providing a quantitative fingerprint of complex protein products. Quantification was based on a common reference sample, which enabled a high number of independent biological replicates for robust statistics. The method has a broad potential applications in the field of biotechnology, like time course analysis of relative protein abundance in whole-cells, or batch-to-batch comparison of product quality. This study compared the proteomes of OMV vaccines from different purification processes and revealed distinct protein profiles. Serum blot proteomics substantiated these results by identifying differential immunogenic proteins. The results indicate that the (immunogenic) protein content of OMV vaccines is at least partially determined by the purification process. This supports previously observed functional differences between OMV vaccines and illustrates that detergent-free OMV have a preferred protein composition.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Marco Ruijken (MSMetrix, The Netherlands) for customizing MS-Xelerator software, Hugo Meiring for help with troobleshooting, Lonneke van Keulen for OMV purification and serum blot analysis, Gideon Kersten for helpful discussion. This work was funded by the Dutch Ministry of Health, Welfare and Sport and The Netherlands Proteomics Centre (embedded in the Netherlands Genomics Initiative).

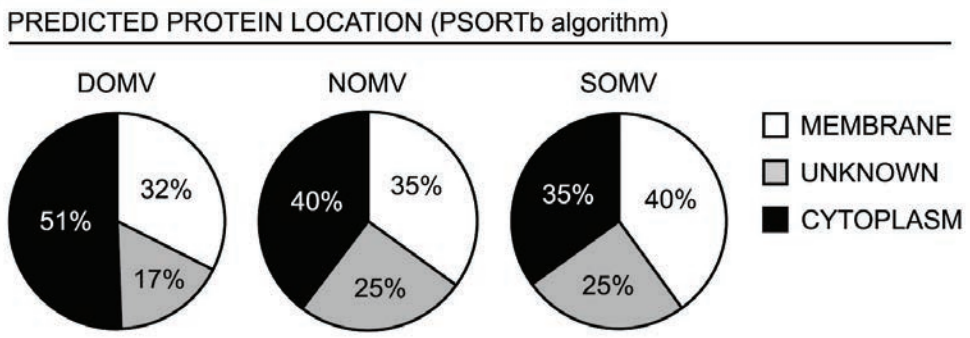
ASSOCIATED CONTENT

Additional information as noted in the text.

SUPPORTING TABLE 1. VALIDATION OF QUANTITATIVE PROTEOMICS METHOD. The quantitative proteomics method was validated with a protein mixture containing known ratios. The mixture included carbonic anhydrase (1:1 ratio), lysozyme (1:1 ratio), insulin (1:2 ratio), alpha casein (1:5 ratio) and serum albumin (1:10 ratio) and was analyzed with 5 independent replicates. Internal peptides before PTAG-labeling and TiO_2 depletion allowed quantification of all proteins except carbonic anhydrase, because this protein has no lysine residues in its internal peptides (required for labeling). N-termini after depletion could be quantified for 3 out of 5 proteins. The N-terminus of lysozyme is too short for detection because trypsin cleaves off the N-terminus at the Asn2 residue. The serum albumin N-terminus was only detected for the light peptide, but not for the 10-fold less abundant heavy labeled peptide. The abundance was most likely below noise threshold, since several internal peptides of serum albumin had the same problem (data not shown). The noise threshold is determined by the detection sensitivity, but also by the accuracy of mono-isotopic mass deconvolution. Deconvolution errors are most likely to occur for isotope pairs with a greatly downregulated heavy labeled peptide (like serum albumin), due to overlapping isotope envelopes from the light labeled peptide. Other serum albumin peptides that were correctly detected had a variable ratio (25 ± 18). The large standard deviation indicates that the 1:10 ratio of serum albumin represents the outer limit of the dynamic range. Therefore, larger ratios are scored as >10 in the remainder of this study. Proteins with ratios between 1:1 and 1:5 were quantified with good to high accuracy, indicating that the method can be used to measure unknown ratios.

Protein	Uniprot ID	Internal peptides	N-terminus	Fold ratio	
				Expected	Observed
Carbonic anhydrase	P00921	not quantified*	quantified	1.0	1.0 ± 0.0
Lysozyme	P00698	quantified	not detectable**	1.0	1.3 ± 0.4
Insulin	P013017	quantified	quantified	2.0	2.1 ± 0.2
Alpha casein	P02662	quantified	quantified	5.0	5.1 ± 1.1
Serum albumin	P02769	quantified	not quantified***	10.0	25 ± 18

*Labelling not possible (no lysine residues).
**N-terminal peptide too short for detection.
***Heavy labelled peptide below noise threshold.



SUPPORTING FIGURE 1. DISTRIBUTION OF PREDICTED PROTEIN LOCATIONS. The different OMV vaccines in this study had a process-dependent protein location distribution. Detergent-extraction of DOMV vaccines resulted in enrichment with cytoplasmic proteins (51% of all proteins) due to bacterial lysis, while SOMV (no extraction) contained 35% cytoplasmic proteins. The mild, detergent-free extraction of NOMV resulted in an intermediate amount (40% cytoplasmic proteins). Due to the low amount of cytoplasmic proteins, detergent-free OMV vaccines (NOMV and SOMV) were enriched with membrane and unknown proteins, which are more likely to be relevant for immunogenicity. Cellular locations were predicted using the PSORTb algorithm (http://www.psort.org/psortb).

SUPPORTING TABLE 3. OVERLAP OF THIS STUDY WITH AVAILABLE LITERATURE. Available studies on OMV vaccines identified 25–166 proteins with several different proteomics methods (gel-free processing versus 1D or 2D gel electrophoresis, combined with qualitative or quantitative LC-MS/MS). Overlap of these studies with the current results was moderate to high, 43–61% of all detected proteins. Therefore the results of this study are likely to be representative and comprise the largest OMV dataset available to date.

OMV study	Proteomics method	# OMV proteins	Overlap
(A) Vipond *et al.* (2006)	Qualitative 2D electrophoresis; LC-MS/MS	25	60%
(B) Vipond *et al.* (2005)	Qualitative 1D electrophoresis; LC-MS/MS	41	61%
(C) Tsolakos *et al.* (2010)	Quantitative 2D electrophoresis; LC-MS/MS	74	58%
(D) Gil *et al.* (2009)	Qualitative gel-free LC-MS/MS	101	50%
(E) Ferrari *et al.* (2006)	Qualitative 1D/2D electrophoresis; LC-MS/MS	166	43%
This study	Quantitative gel-free LC-MS/MS	185	100%

(A) Proteomics. 2006, 6(11): p3400-13.
(B) Human Vaccines. 2005, 1(2): p80-4.
(C) Vaccine. 2010, 28(18): p3211-8.
(D) Human Vaccines. 2009, 5(5): p347-56.
(E) Proteomics. 2006, 6(6): p1856-66.

CHAPTER 04

Mixed-bed ion exchange chromatography employing a salt-free pH gradient for improved sensitivity and compatibility in MudPIT

Geert P. M. Mommen 1,2,3
Hugo D. Meiring 1
Albert J. R. Heck 2,3
Ad P. J. M. de Jong 1

1 Formulation and Analytical Research, Institute for Translational Vaccinology, Bilthoven, The Netherlands
2 Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands
3 Netherlands Proteomics Centre, Utrecht, The Netherlands

Analytical Chemistry, 2013. 85(14): p. 6608-16

ABSTRACT

In proteomics, comprehensive analysis of peptides mixtures necessitates multiple dimensions of separation prior to mass spectrometry analysis to reduce sample complexity and increase the dynamic range of analysis. The main goal of this work was to improve the performance of (on-line) multidimensional protein identification technology (MudPIT) in terms of sensitivity, compatibility and recovery. The method employs weak anion and strong cation mixed-bed ion exchange chromatography (ACE) in the first separation dimension and reversed phase chromatography (RP) in the second separation dimension (Motoyama *et.al.*, Anal Chem 2007, 79, 3623-34.). We demonstrated that the chromatographic behavior of peptides in ACE chromatography depends on both the WAX/SCX mixing ratio as the ionic strength of the mobile phase system. This property allowed us to replace the conventional salt gradient by a (discontinuous) salt-free, pH gradient. First dimensional separation of peptides was accomplished with mixtures of aqueous formic acid and dimethylsulfoxide with increasing concentrations. The overall performance of this mobile phase system was found comparable to ammonium acetate buffers in application to ACE chromatography, but clearly outperformed strong cation exchange for use in first dimensional peptide separation. The dramatically improved compatibility between (salt-free) ion exchange chromatography and reversed phase chromatography-mass spectrometry allowed us to downscale the dimensions of the RP analytical column down to 25 µm I.D for an additional 2 to 3 fold improvement in performance compared to current technology. The achieved level of sensitivity, orthogonality and compatibility demonstrates the potential of salt-free ACE MudPIT for the ultra-sensitive, multidimensional analysis of very modest amounts of sample material.

INTRODUCTION

Current shot-gun proteomics methods are generally based on peptide separation using reversed phase (RP) liquid chromatography (LC) where identification is typically accomplished by tandem mass spectrometry (MS/MS). It is generally accepted that no single chromatographic procedure is capable of resolving extremely complex peptide mixtures such as whole cell digests[103]. Combining two (2D) or more orthogonal separation procedures significantly improves the overall peak capacity and dynamic range of the analysis and results in a larger number of identified peptides [201]. Ideally, in multidimensional separations, the displacement mechanisms of each method are orthogonal, meaning that little correlation exists between the retention of compounds in each dimension [41]. In most 2D workflows, reversed phase (RP) is utilized as the second dimension in combination with various separation methods in the first dimension. Popular methods for the first dimension are Strong Cation eXchange (SCX), Hydrophilic Interaction Liquid Chromatography (HILIC), perfluorinated reversed phase chromatography, mixed mode (MM) chromatography and RP-RP with high pH and low pH elution [37, 201, 202]. More recently, the need for multidimensional separation in complex proteome analysis has been challenged [203-205], reporting significant proteome coverage in a single run analysis using ultra-high-pressure-liquid chromatography (UHPLC) and high speed sequencing mass spectrometry.

Notwithstanding, deep proteome profiling requires multiple dimensions of separation. These multidimensional peptide separation methods can be used in both off-line and on-line mode, each with their inherent advantages and disadvantages [40]. Off-line separations offer greater flexibility in optimizing the chromatographic conditions for each dimension and the ability to re-analyze individual sample fractions. Advantages of on-line formats are the ease of automation and a reduction of sample loss and contamination. The earliest reported and still most frequently used 2D proteomics approach today is the so-called multidimensional protein identification technology (MudPIT), which couples SCX and RP chromatography on-line with mass spectrometry [28]. Volatile salts like ammonium acetate are used in the elution solvents to fractionate SCX-bound peptides for subsequent analysis by RP LC-MS. SCX-RP chromatography combines fairly good orthogonality, albeit with only moderate separation efficiency in the first separation dimension. The chromatographic performance of SCX was improved by the use of continuous salt and/or pH gradient resulting in less overlap between fractions (carry-over) [206, 207], but such analysis required a more sophisticated on-line setup. Of all 2D methods, RP/RP at high pH and low pH offers the highest peak capacity [38, 208], however it lacks mobile phase compatibility for on-line applications since organic modifiers are required in both dimensions.

Recently, Motoyama *et al.* [118] reported on an interesting variant of ion exchange chromatography. They mixed weak anion exchange and strong cation exchange resins (ACE) in a single column for the first dimensional separation of peptides. The ACE mixed-bed required lower salt concentrations for elution and yielded a two-fold higher recovery compared to SCX alone. Improved recoveries were attributed to higher fluxes of salt cations proximal to the SCX surface by repulsion by the positive charges on the anion exchange resin surface (WAX), a mechanism they termed by the Donnan effect [118].

Here, we further improved the compatibility of ion exchange chromatography with downstream reversed phase LC-MS analysis. The key achievement of our study is the finding that elution of peptides retained by an anion exchange and cation exchange (ACE) mixed-bed can be accomplished by elution with mixtures of formic acid (FA) and dimethylsulfoxide (DMSO). We describe a novel, completely salt-free, two-dimensional LC-MS/MS system based on the ACE mixed-bed technology for the analysis of complex peptide mixtures.

MATERIALS AND METHODS

PEPTIDES. The peptides Angiotensine III (RVHYIHPF) and Oxytocin (CYIQNCPLG) were purchased from Sigma Aldrich (Zwijndrecht, The Netherlands). KLWESPQEI and RRYPDAVYL were prepared in-house by FMOC chemistry. Evaluation of 2D LC-MS/MS was performed with a tryptic peptide mixture comprising the standard proteins Bovine Serum Albumin, Yeast Enolase, Phosphorylase b, Bovine Hemoglobin, and Yeast Alcohol Dehydrogenase (Waters, Milford, USA). The digests were mixed in an equimolar ratio, diluted to a final concentration of 10 fmol/µl in water/formic acid (FA)/ dimethylsulfoxide (DMSO) 90/5/5 (v/v/v). Typically, 10 µl of this protein digestion mixture was used for analysis.

SAMPLE PREPARATION. *Saccharomyces cerevisiae* (strain BJ5460, LGC Standards, Almere, The Netherlands) protein extracts were prepared as described previously [209]. An aliquot corresponding to 100 µg of *S. cerevisiae* proteins was dissolved in 100 mM phosphate buffer pH 7.5 containing < 1 M guanidine hydrochloride. Disulfide bridges were reduced (dithiotreitol, Sigma Aldrich) and free thiol groups were alkylated (Iodoacetamide, Sigma Aldrich). Proteins were digested by trypsin (Promega, Leiden, The Netherlands) at 37 °C, an enzyme/protein ratio of 1/20 (w/w) and overnight incubation. The peptide mixture was diluted to 0.1 µg/µl with water/FA/DMSO 90/5/5 (v/v/v) and typically 10 µL of this mixture was used for analysis (1 µg).

SALT-FREE ACE MUDPIT ANALYSIS. 2D LC-MS/MS experiments were performed using an Agilent 1200 HPLC system (Agilent, Amstelveen, The Netherlands) equipped with a vented column switching system (Figure 1) [119]. Solvent A was 1 mM acetic acid (HOAc) (pH ~3.8) (Sigma Aldrich) in deionized water (MiliQ, Millipore, Amsterdam, The Netherlands) and solvent B was 1 mM acetic acid in acetonitrile (AcN) (Biosolve, Valkenswaard, The Netherlands). Micro-capillary trapping and analytical columns were prepared by slurry packing using fritted capillaries [119]. The biphasic trapping column was composed of a 20 mm ACE mixed-bed as downstream and a 20 mm 5 µm Reprosil-Pur C18-AQ (Dr. Maisch, Ammerbuch-Etringen, Germany) as upstream packed in a 100 µm I.D. capillary (preparation ACE blends described below). This trapping column was connected to a 50-µm I.D. trapping column packed with 20 mm 5 µm Reprosil Pur C18-AQ using an in-house modified zero dead-volume column connector (Upchurch, Oak Harbor, USA). The analytical column was a 25-µm I.D. micro-capillary packed with 40 cm of 3 µm Reprosil Pur C18-AQ (Dr. Maisch). Unless stated, fresh ACE blends were prepared for each experiment by mixing dry 5µm PolyWAX LP (PolyLC Inc, Colombia, USA) and 5 µm polysulfoethyl A (PolyLC Inc, Colombia, USA) resins in a 1/4 (w/w) ratio. Mixtures were suspended in 2-propanol for slurry packing. In cases were a 50-µm I.D. analytical column was used, a 20-mm ACE bed was sandwiched between two C18 compartments (15 and 20 mm, respectively) in a single 100-µm I.D. fritted fused silica capillary.

SAMPLE LOADING. Samples were loaded onto the upstream RP bed of the biphasic trapping column at a flow rate of 3 µL/min with the switching valve set to the vent position (port 1 connected to port 6, see Figure 1) and subsequently desalted by flushing with 100% solvent A for the remaining of the 10 min injection period. Following injection and desalting, peptides retained by the upstream RP bed of the biphasic trapping column were transferred into the ACE bed for (electrostatic) binding by running a 90-min gradient run with the analytical column in series (port 2 connected to port 3, see Figure 1). The pump flow rate was set to generate a column head pressure of approximately 500 bar and a column flow rate of approximately 30 nL/min.

ANALYTICAL RUNS. Salt-free elution buffers (Table 1) were injected at a flow rate of 3 µL/min. Eluted peptides were trapped and refocused on the downstream C18 bed for subsequent gradient RP LC-MS/MS analysis. Gradient profile: step increase at the gradient start to 8% B, a 160-min gradient from 8 to 26% B, a 20-min gradient from 26 to 60% B followed by a step-decrease to 0% B for reconditioning of the column for 20 min. In the first fraction, the ACE bed is extensively flushed (90 µL or approximately 500 bed volumes) with 1 mM HOAc to elute peptides with very low affinity to the ACE resin. Subsequent fractions were obtained by the injection of small volumes (3-10 µL) of aqueous buffers of formic acid (FA) and dimethylsulfoxide (DMSO) with increasing concentration (Table 1).

TABLE 1. COMPOSITION ELUTION BUFFERS IN SALT-FREE ACE MUDPIT. Concentration of formic acid (FA) and dimethylsulfoxide (DMSO) used to elute peptides retained by weak anion exchange and strong cation exchange (ACE) mixed-bed chromatography.

Fraction	FA (mol/L)	DMSO (mol/L)	pH	Volume (µL)
1[a]	-	-	3.9	90
2	0.0014	0.0007	3.6	10
3	0.0135	0.007	2.9	10
4	0.135	0.07	2.3	10
5	1.35	0.7	2.0	10
6	3.38	0.7	1.5	10
7	6.75	0.7	1.0	3
8	13.5	0.7	0.6	3

-[a] Fraction 1: elution with 1 mM acetic acid

SALT-BASED MUDPIT ANALYSIS. Salt-based SCX-RP and ACE-RP experiments were performed using a vented column switching system [119] with otherwise identical conditions. Solvent A was 0.1% FA in deionized water and solvent B 0.1% FA in acetonitrile. The biphasic trapping column was a 20-mm SCX or 20-mm ACE bed, sandwiched between two C18 compartments (15 and 20 mm, respectively) in a single 100-µm I.D. fritted fused silica capillary. The analytical column was a 50-µm I.D. micro-capillary packed with 40 cm of 3 µm Reprosil Pur C18-AQ. Following sample loading, peptide fractions were recovered by flushing the SCX or ACE column with pulses (5 µL) of ammonium acetate (10 mM – 500 mM) in 0.1% FA containing 2% AcN at a flow rate of 5 µL/min. After each pulse, the trapping column was extensively washed by flushing with 100% solvent A for the remaining of a 20 min injection period. The discontinuous ammonium acetate gradient was increased in eight steps from 10 mM to 500 mM (10 mM, 25 mM, 50 mM, 75 mM, 100 mM, 175 mM, 250 mM and 500 mM).

MASS SPECTROMETRY CONDITIONS. A LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) was used in this study. The LC column effluent was directly electrosprayed into the MS using a gold and conductive carbon-coated fused silica tapered tip of ~5 µm I.D. [119], at a potential of typically 2 kV. The MS was set to acquire full scan MS spectra (*m/z* 250 to 1,500) for mass analysis in the orbitrap at 60,000 resolution (FWHM), each followed by data-dependent collisional induced dissociation MS/MS analysis (LTQ) for the top 10 abundant precursor ions above an abundance threshold of 100,000 or 50,000 counts. The normalized collision energy was set to 35 %, isolation width to 2.0 Th, activation Q to 0.250 and activation time to 30 ms. The maximum ion collection time (dwell time) for MS scans was set to 200 ms and for MS/MS scans to 500 ms. Precursor ions with unknown and +1 charge state were excluded for MS/MS analysis. Dynamic exclusion was enabled (exclusion size list 500) with repeat set to 1 and an exclusion duration of 90 s. The background ion at *m/z* 391.284280 was used as lock mass for internal mass calibration.

DATA ANALYSIS. Analysis of mass spectrometric RAW data was carried out using Proteome Discoverer 1.2 software package (Thermo Fisher Scientific, Bremen, Germany) applying standard settings unless otherwise stated. MS/MS scans were searched against the *S. cerevisiae* SGD database (http://www.yeastgenome.org, 2010, containing 5,821 entries) using SEQUEST (Proteome Discoverer 1.2, Thermo Fisher Scientific, Bremen, Germany). Precursor ion and MS/MS tolerances were set to 5 ppm and 0.8 Da, respectively. Decoy database searches were performed with False Discovery Rate (FDR) tolerances set to 5% and 1% for modest and high confidence filtering settings, respectively. The data were searched with full trypsin cleavage specificity allowing 2 miss-cleavages. Cysteine carbamidomethyl was set as fixed modification while asparagine deamidation and methionine oxidation were set as variable modifications. An additional search was performed for N-acetylated peptides with N-terminal acetyl as fixed modification and C-terminal trypsin cleavage specificity. Peptide identifications were filtered to an FDR less than 1 % and Rank no. 1. On request, the raw data files and protocols associated with this manuscript are available to the reader. Possible peptide features (LC peaks) [210] were determined by MsXelerator software (MsMetrix, Maarssen, The Netherlands) [46] from full MS scans for *m/z* traces with a signal-to-noise ratio (S/N) > 10 and at least 6 scans to determine peak shape. Singly charged peptide features were discarded from the final list.

LC-UV ANALYSIS. To evaluate the peptide elution window of SCX and ACE mixed bed columns, tryptic BSA samples (Waters, Milford, USA) were analyzed by ion exchange chromatography with UV detection. The system comprises of a quaternary pump (HP1050, Agilent, Amstelveen, Netherlands) equipped with an (nanoflow cell) UV detector [46]. Solvent A was 0.1 M HOAc and Solvent B 500 mM KCl in 0.1 M HOAc containing 35% AcN. Samples were loaded onto a Hypercarb™ (7 µm particle size, Thermo Fisher, 5 mmL × 200 µm I.D.) trapping column at a flow rate of 5 µL/min with the effluent directed to waste. After 4 min, the trapping column was switched in series with the column. Peptides were transferred onto the top of the IE column by a plug injection of 10 µL 85% AcN in 0.1 M HOAc. Peptides were eluted by a KCl/AcN gradient ramping from 0-50% B in 33 min followed by a further increase to 100% B in 10 min at a flow rate of 2 µL/min.

RESULTS AND DISCUSSION

EVALUATION OF PEPTIDE RETENTION BY ACE MIXED-BEDS. In this study we build upon the concept introduced by Motoyama *et al.* [118] using weak anion exchange and strong cation exchange (ACE) mixed-bed for the first dimensional separation of complex peptide mixtures. To evaluate the elution characteristics of peptides from both SCX and ACE mixed-bed ion exchange (IE) resins, we analyzed a tryptic BSA digest by LC-UV using ion exchange columns with different WAX/SCX mixing ratios (0/1, 2/1, 4/1, 6/1 (w/w)). Supplementary Figure S1 demonstrates a clear shift to lower salt concentrations required to elute peptides from columns with increasing WAX/SCX mixing ratios. According to Motoyama *et al.* [118], such a shift to lower salt concentrations relates to the corresponding length of the SCX bed and the higher fluxes of salt cations and lower fluxes of salt anions proximal to the SCX resin (Donnan effect). We reasoned that this shift in peptide retention to lower buffer strengths could be used in the search for alternative elution buffers for use in on-line multidimensional LC-MS applications. The use of inorganic salts is generally considered undesirable because of possible ion suppression in electrospray ionization and limiting lifetime of the column [202, 211] (Supplementary Figure S2). Volatile salts are, to some extent, tolerated when trapping columns are applied for extensive washing and desalting of the LC system prior to MS analysis [212]. Although H_3O^+ is a weak eluotrope in SCX chromatography, initial experiments showed that peptides can be recovered from ACE resins by elution with aqueous formic acid (FA). Furthermore, we found that dimethylsulfoxide (DMSO) exhibits favorable eluotropic features for SCX bound peptides (Supplementary Figure S3). DMSO is a polar aprotic solvent capable of stabilizing ions in solution and known as an excellent solvent for poorly water-soluble peptides. Hence, DMSO may be counteracting hydrophobic interactions, a phenomenon recognized as a major source for peak broadening and peptide losses in SCX chromatography [211]. The eluotropic properties of DMSO is not restricted to SCX alone and concentrations were kept below 10% to prevent losses of polar compounds in reversed phase (RP) chromatography (Supplementary Figure S4). We combined the eluotropic properties of aqueous mixtures of FA/DMSO with the decreased retention of peptides by ACE chromatography and established a genuinely salt-free MudPIT separation method.

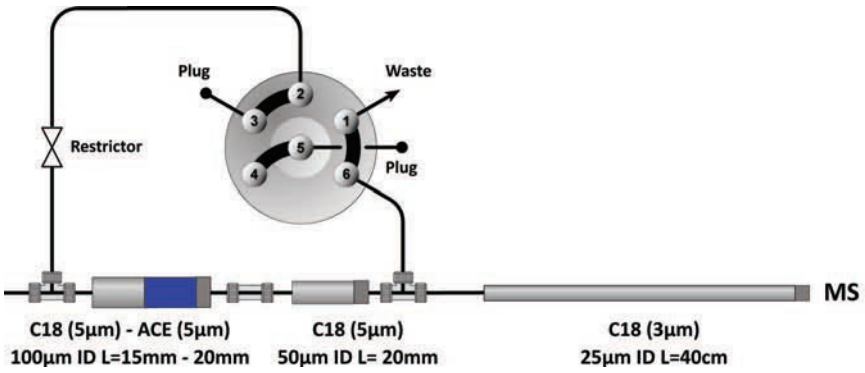


FIGURE 1. SCHEMATIC REPRESENTATION OF THE ACE-RP WORKFLOW. The system comprises of a 100-µm I.D. biphasic RP-ACE trapping column coupled to a 50-µm I.D. RP trapping column and a true nanoflow 25-µm I.D. analytical RP column. Samples and elution buffers are injected with the valve set to waste (port 1 connected to 6). Subsequent gradient analysis is accomplished using passive splitting (port 2 and 1 connected) of the primary pump flow rate (typically 0.3 mL/min), resulting in a column head pressure of approximately 500 bar and a flow rate through the column of approximately 30 nL/min.

WORKFLOW FOR SALT-FREE ACE-RP SEPARATION OF PEPTIDES. On-line, 2D peptide separation experiments were carried out using an ACE mixed-bed trapping column in the first separation dimension combined with a RP analytical column in the second dimension (Figure 1). A RP solid phase extraction (SPE) bed was placed in front of the ACE resin to extract the analytes from the sample solvent, salts and other crude sample constituents, that otherwise may adversely affect the binding by the ACE resin. In addition, pre-trapping allowed the loading of analytes onto the ACE resin in a well-controlled manner. We found that the pH of the LC solvent (1 mM HOAc, pH 3.8) and residence time are critical in the binding efficiency by the ACE resin, particularly for peptides with low charge states. The residence time in a bed equals the void volume divided by the flow rate. Low transfer flow rates are readily obtained by a gradient elution with the analytical column in series. A typical analytical column flow rate of 30 nL/min results in a residence time of 2 minutes in the ACE mixed-bed, sufficiently long for effective binding of weak cationic species (Supplementary Figure S5). To preserve the overall separation efficiency, the I.D. ratio of the trapping column and analytical column was kept at a value of 2. Hence, a downstream 50 µm I.D. trapping column is appropriate for use in combination with a 25 µm I.D. analytical column. Owing to the solvent-compatibility between the first (ACE, water) and the second separation dimension (RP, organic), the I.D. of the upstream RP/ACE biphasic trapping columns can be chosen independently. Possible band broadening in this part of the system due to lower than optimal linear velocities of the mobile phase will be effectively counteracted by refocusing on the top of the downstream RP bed. For practical reasons, *e.g.* speed of sample loading, capacity and residence time, we opted for a 100 µm I.D. upstream RP/ACE trapping column.

The flow-through fraction during the RP-to-ACE transfer virtually comprised of non-peptidic substances only. Such a distinct selectivity remains a desirable feature of ion exchange chromatography compared to other 2D methods [38, 213] in analysis of crude genuine biological samples that often suffer from poor analyte-to-debris ratios (Supplementary Figure S6). Following the transfer, the ACE resin was flushed with relatively large volumes (90 µL) of 1 mM acetic acid to collect weakly bound peptides in a single fraction. The content of this fraction comprises of singly charged (no basic residues) and short polar, doubly charged (single basic residue) peptides (Supplementary Figure S5). For example, the singly charged tryptic C-terminal peptide of BSA (LVVSTQTALA charge state 0.1 at pH 3.9), was recovered in this fraction. Elution of consecutive peptide fractions was accomplished by pulses of aqueous mixtures of FA and DMSO in strength increasing over four orders of magnitude (Table 1), each followed by standard RP gradient separation and MS analysis.

OPTIMIZATION OF WAX/SCX BLEND RATIO. To evaluate the ability to recover a given analyte from the ACE resin by salt-free elution buffers, we first determined the eluotropic strength of various acids. An ACE trapping column (WAX/SCX, 4/1 (w/w)) was challenged with a tryptic peptide mixture (5 standard proteins) and successively eluted with equimolar (0.1 M) concentration of acetic acid (HOAc), formic acid (FA), trifluoroacetic acid (TFA) and hydrochloric acid (HCl). As expected, the eluotropic strength parallels the acidity (pKa) of the acid (HOAc < FA < TFA < HCl with pKa values of 4.76 < 3.77 < 0.23 < -6.2, respectively). TFA and HCl were rejected as elution buffers because of retention time shift by ion-pairing (TFA), chlorine ion clusters in electrospray MS and deterioration of the resins (HCl). The optimal mobile phase system consists of aqueous HOAc or FA because of their buffering properties and excellent compatibility with RP-LC-MS.

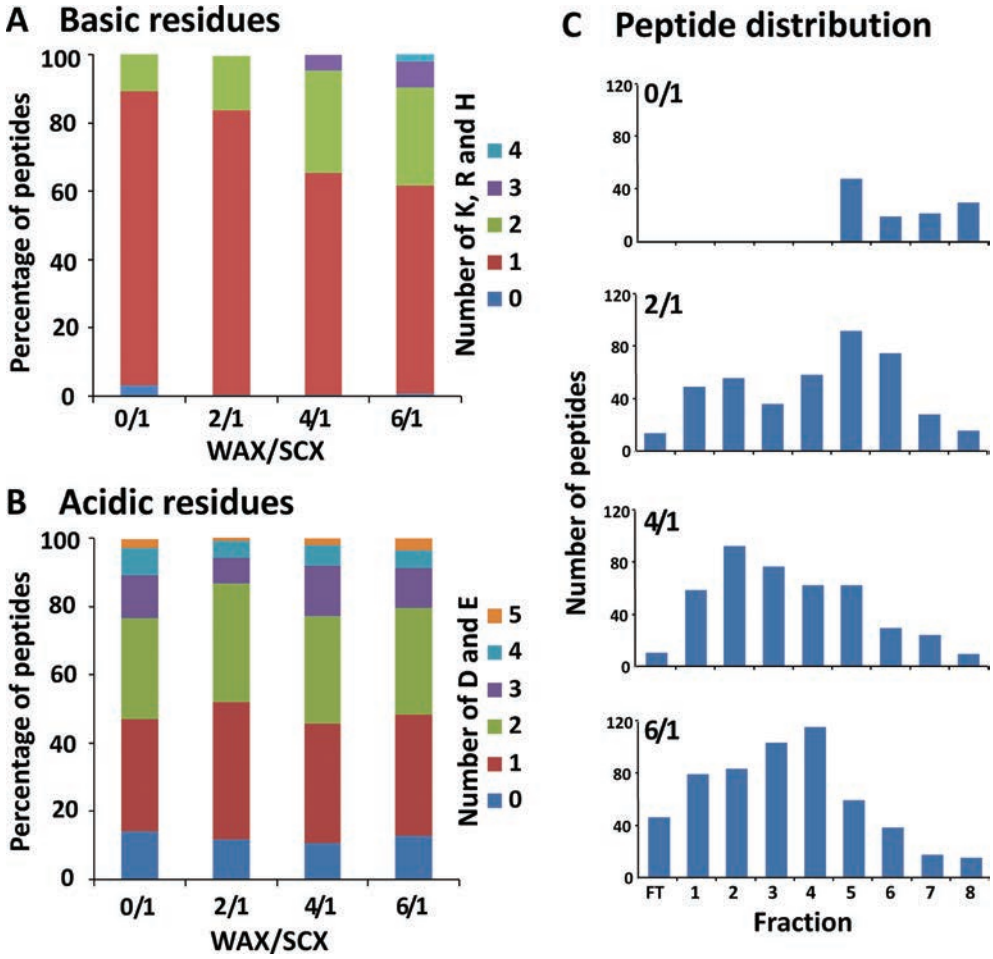


FIGURE 2. OPTIMIZATION OF WAX/SCX MIXING RATIO IN SALT-FREE ACE MUDPIT. Distribution of the number of basic (A) and acidic (B) amino acid residues in peptides recovered by elution with salt-free buffers (see Table 1) of strong cation exchange (SCX), and weak anion exchange and strong cation exchange (ACE) mixed-bed columns. The WAX/SCX mixing ratios investigated were 0/1 (100% SCX), 2/1 (33% SCX), 4/1 (20% SCX) and 6/1 (14% SCX), as denoted by a/b. (C) Number of peptides recovered in individual fractions during a discontinuous pH gradient elution of the different ion exchange columns. Fractions were analyzed by reversed phase LC-MS/MS (60 min gradient). Sample: tryptic 5 protein digest.

Next, we evaluated the recovery of tryptic peptides from both SCX and ACE with different WAX/SCX mixing ratios (2/1 to 6/1 (w/w)) by elution with aqueous FA/DMSO. Tryptic peptides were eluted from the biphasic column by an 8-step discontinuous pH gradient (Table 1), each step followed by RP LC-MS/MS analysis. The results were evaluated on the ability to recover cationic peptides and coverage of the 2D elution plane. Figure 2A shows the basicity of the recovered peptides as represented by the number of basic residues (His, Lys, Arg) during a discontinuous eight-step gradient elution of the blends investigated. The results showed that densely charged (> 3 basic residues) species were not recovered by elution with salt-free buffers of mixed-bed columns high in SCX content (2/1 (w/w)). Vice versa, evidence was obtained for the ability to recover multiple (+) charged peptides with up to 4 to 5 basic residues from mixed-bed columns relatively rich in WAX (4/1 and 6/1 (w/w)).

The coverage of the 2D separation plane varied for the mixed-bed columns investigated and revealed a clear shift in peptide elution towards early fractions for columns with increasing WAX/SCX mixing ratios (Figure 2C). For example, only a few peptides were recovered during the early fractions of the SCX column (strong retention), while these early fractions were rich in peptides for the WAX/SCX 6/1 (w/w) column (weak retention). The most evenly distribution of peptides across the fractions, together with a minimum of flow-through peptides during transfer, was obtained for the WAX/SCX 4/1 (w/w) mixing ratio. The elution profile showed a good correlation between the average charge state of recovered peptides and the pH of the elution buffer (Supplementary Figure S8). Moreover, the recovery of peptides from an ACE (WAX/SCX, 4/1 (w/w)) trapping column was estimated at 80%. This estimation was based on the comparison of the RP-LC base peak chromatograms of peptides eluted from the ACE bed by a single, high-concentration FA/DMSO pulse with a control (single-phase) RP analysis (data not shown).

As expected, acidic residues (D, E; phosphorylated peptides not considered) show no significant effect on the retention of peptides by ACE chromatography (Figure 2B). During FA/DMSO pulses, the acidic medium largely blocks the dissociation of carboxylic acid groups, preventing (-) electrostatic interactions with the WAX resin. The experimental data suggest that the retention mechanism of ACE, under acidic conditions, is determined by (+) electrostatic interaction between the cationic peptides and the negatively charged groups of the resin. The comparison of the ACE trapping columns with different WAX/SCX mixing ratios revealed a clear shift in elution profile towards earlier fractions (Figure 2C). To investigate whether this shift is related to length of the SCX bed [118], we compared the elution profiles of SCX columns with lengths ranging from 3 mm to 20 mm. The result showed a minor shift (by one fraction) in the elution of peptides, but no improvement in the recovery of peptides with multiple basic residues (Supplementary Figure S9). This means that the WAX resin governs the displacement of peptides during the elution of ACE columns by salt-free buffers. The positive charges of the WAX resin promotes peptide elution as a result of the higher fluxes of buffer cations and lower fluxes of buffer anions proximal to the surface of the SCX resin (Donnan effect) [118]. The positive charges may also interact with the negative charged groups of the SCX resin and therefore may interfere with the electrostatic interaction between peptides and the mixed-bed column. However, the individual effect of the WAX resin on the electrostatic interaction (peptide binding) or the Donnan effect (peptide elution) is unclear since it could not be examined separately.

Moreover, our results revealed a clear relation between WAX/SCX mixing ratio and ionic strength of the mobile phase system required for peptide elution. For example, the separation of peptides by employing a FA/DMSO gradient necessitates an ACE column higher in WAX content than needed for a salt buffer system (4/1 and 2/1 (w/w), respectively) [118].

TABLE 2. PERFORMANCE SALT-FREE ACE MUDPIT. Number of sequence attempts (MS/MS spectra), peptide-to-spectrum matches (PSM), number of peptides and unique peptides identified in each fraction in the 2D analysis of a tryptic digest of *S. cerevisiae* (1 µg consumed). First dimensional separation was achieved by a discontinuous pH gradient elution (FA/DMSO, Table 1) of a weak anion exchange and strong cation exchange mixed-bed (ACE, WAX/SCX 4/1, (w/w)) coupled on-line to a true nanoflow (25 µm I.D.) reversed phase (RP) column in the second separation dimension. The average charge state of the identified peptides is compared to the average charge state of peptide features (LC peaks, see Experimental section).

Fraction	# sequence attempts	# PSM	# peptide identifications	# unique peptides	avg. charge state identified peptides	avg. charge state peptide features
1	13303	3035	1980	1980	2.0 + 0.1	2.1 + 0.5
2	15010	4210	1967	1006	2.0 + 0.1	2.2 + 0.5
3	16815	4333	2481	1812	2.1 + 0.4	2.4 + 0.6
4	13667	5626	3010	2111	2.3 + 0.5	2.7 + 0.8
5	14282	5732	3455	2208	2.5 + 0.6	2.9 + 0.9
6	8579	3751	2355	1116	2.8 + 0.6	3.3 + 1.1
7	6105	2567	1672	429	2.9 + 0.7	3.6 + 1.3
8	4541	2086	1347	313	2.9 + 0.8	3.6 + 1.3

EVALUATION OF SALT-FREE ACE MUDPIT FOR GLOBAL PROTEOME ANALYSIS. To evaluate the performance of salt-free ACE MudPIT in applications to complex proteome samples, we performed a triplicate 2D analysis of a tryptic digest of *S. cerevisiae* (1 µg/replicate and WAX/SCX 4/1 (w/w) format). Each replicate consisted of an eight-step discontinuous pH gradient, each followed by a 3-h RP LC-MS/MS gradient run (24 h total elution window). The evaluation included the chromatographic performance, the orthogonality and the peptide and protein identification results. The chromatograms in Supplementary Figure S10 demonstrate that the tryptic peptides were nicely distributed over the ACE fractions. To determine the orthogonality of the separation in more detail, all detectable peptide features (LC peaks) were extracted from the MS data and plotted in a heatmap view (Figure 3A). The plot demonstrates significant 2D separation orthogonality, particularly for the first 6 fractions (~ 90% of peptide content in the sample). The more densely covered upper/right area of the 2D elution plane is largely due to the nature of tryptic peptides. For tryptic peptides, only peptides containing histidine(s) and/or missed cleavage peptides carry a charge greater than two. Obviously, missed cleavage products are more cationogenic and are generally longer and are more hydrophobic and therefore stronger retained in both separation dimensions.

The chromatographic performance was critically evaluated since resins in both traps are transiently exposed to buffers with pH values below manufacturer's recommendation (pH 2). The stability of the resins was demonstrated by the high reproducibility in the triplicate 2D analysis of tryptic *S. cerevisiae* (Supplementary Figure S11). The widths of chromatographic peaks in the second separation dimension are also a good parameter to estimate changes in separation efficiency and chromatographic integrity. The median peak-width increased by not more than 5% in the third replicate compared to the first, demonstrating that little deterioration of the resins is taking place (Supplementary Figure S12).

Peak capacities were calculated from the widths (baseline, 4σ) of the peptide features (Figure 3B), yielding an estimated value for the 3-h gradient of ~300, a value comparable to values reported by others [204, 211]. According to Giddings [214], the peak capacity for 2D systems equals the sum of the peak capacity in individual cycles, adding up to ~2400 for the 8-step gradient. This value is an overestimation since the calculation does not take into account imperfect orthogonality, carry-over between fractions and other issues that lower the duty cycle.

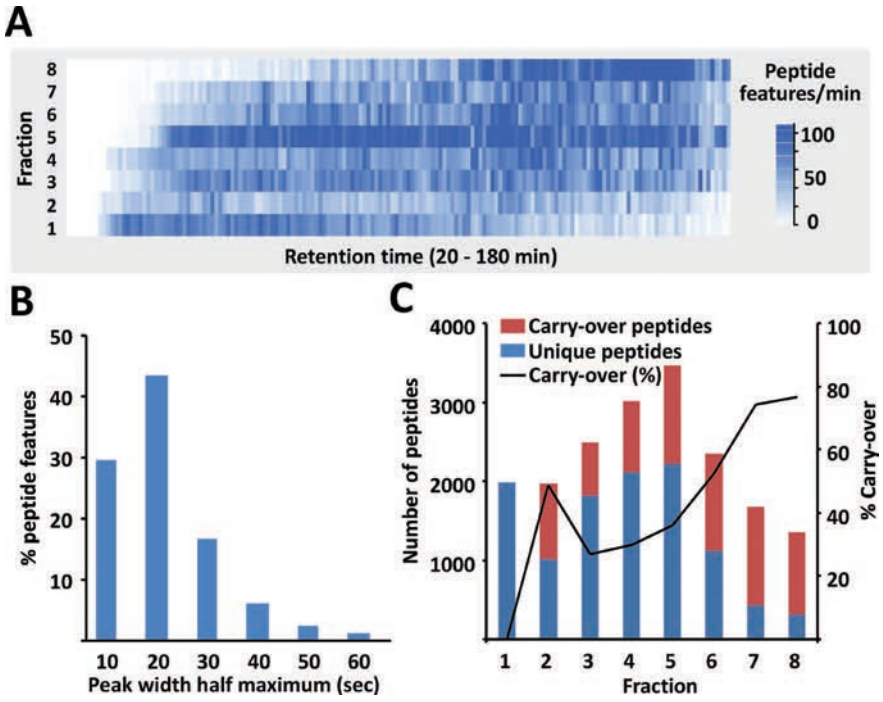


FIGURE 3. PERFORMANCE OF SALT-FREE ACE MUDPIT APPLIED TO THE ANALYSIS OF TRYPTIC *S. CEREVISIAE*. (A) Heat map view of the number of peptide features eluting per minute during an eight-step discontinuous pH gradient elution (see Table 1) of a weak anion exchange and strong cation exchange (ACE) mixed-bed. The WAX/SCX ratio was 4/1 (w/w). Fractions were analyzed by reversed phase LC-MS/MS employing a 25 µm I.D. analytical column eluted by a shallow, 3-h gradient. (B) Distribution of peak width at half maximum (FWHM) obtained in second separation dimension. (C) Unique (blue bars) and carry-over (red bars) peptides identified across first dimensional fractions. The line graph depicts the percentage carry-over peptides in each fraction.

The 2D analysis of a tryptic digest of a *S. cerevisiae* (1 µg) resulted in the cumulative identification of 10,976 unique peptides corresponding to 2355 proteins (FDR < 1%). The significant orthogonality of 2D ACE-RP separation was also reflected by the fairly equal distribution of the number of peptide-to-spectrum matches (PSM) and the number of peptide identifications across the fractions (Table 2). The majority of unique peptides were identified in the first 6 fractions, mainly due to the high degree of peptide carry-over in the latter two fractions (Figure 3C). The carry-over in the present study is still about just half of that observed in a more conventional MudPIT setup [212].

The triplicate 2D analysis of tryptic S. cerevisia, using a first generation LTQ-Orbitrap instrument, yielded together more than 15,000 unique peptides and with 2700 protein identifications already covered a significant part of the yeast proteome (41%, 6717 ORFs) (Supplementary Figure 11). However, we believe these impressive numbers could be even further boosted using an even more advanced (*i.e.* faster and more sensitive) instrument (*e.g.* Q-Exactive, Orbitrap Elite, or triple-TOF).

DIRECT COMPARISON TO SALT-BASED MUDPIT. The performance of salt-free ACE regarding peptide and protein identifications was evaluated in direct comparison to salt-based MudPIT. However, in our hands the refined, 25 µm I.D. analytical column and the stability in electrospray-MS appeared less tolerant to the exposure to high salt burden (Supplementary Figure S2/S13). These problems were to some extent circumvented by the use of larger I.D. trapping and analytical columns with more sufficient washing capabilities [212]. A tryptic digest of *S. cerevisiae* was analyzed by 2D LC-MS using a 100 µm I.D. biphasic trapping column and a 50 µm I.D. analytical column. The experiment consisted of an eight-step discontinuous pH (FA/DMSO) elution of an ACE column with a 4/1 (w/w) WAX/SCX mixing ratio. Salt-based MudPIT experiments were performed by employing an eight-step ammonium acetate gradient elution (10 – 500 mM) of a SCX column or ACE columns with different WAX/SCX mixing ratios. The fractions were analyzed by 3-h RP-LC-MS gradient runs.

Consistent with previous experiments, the salt-based elution of ACE columns showed a clear shift in peptide elution towards earlier fractions for mixed-bed columns increasing in WAX/SCX mixing ratios (Figure 4C). The most optimal salt-based ACE MudPIT (1/1 (w/w)) analysis yielded a total of 4503 unique peptides and 1083 proteins. Similar identification numbers were found with salt-free ACE MudPIT (4393 unique peptides, 1347 proteins). Moreover, salt-free and salt-based ACE performed equally well in terms of the ion-exchange separation profile and the recovery of cationogenic peptides, but both clearly outperformed SCX in the first dimensional separation of peptides (Figure 4). The improved compatibility of the salt-free mobile phase system with RP-LC-MS allowed us to hyphenate ACE to an ultra-narrow 25 µm I.D. analytical column (previous section). This yielded an additional 2 to 3 fold more peptide and protein identifications, which we attributed to the overall improvements in sensitivity (*e.g.* 25µm I.D. column), recovery (*e.g.* highly cationic peptides) and compatibility (*e.g.* no performance loss).

To investigate the enhancement in peptide recovery by ACE in comparison to SCX in more detail [118], we comparatively evaluated the number of basic residues in the peptide sequences for the individual fractions (Supplementary Figure S7). The results suggest significant losses of highly cationic peptides by SCX, most probably due to the insufficient buffer strength of ammonium

acetate for eluting these peptides. Gilar *et al.* [41] provided evidence for this assumption. They demonstrated that missing, predominantly highly charged peptides during salt gradient SCX analysis could be recovered by elution with mixtures of aqueous ammonium hydroxide and formic acid, a buffer less high in salt cations but fortified with H_3O^+ ions.

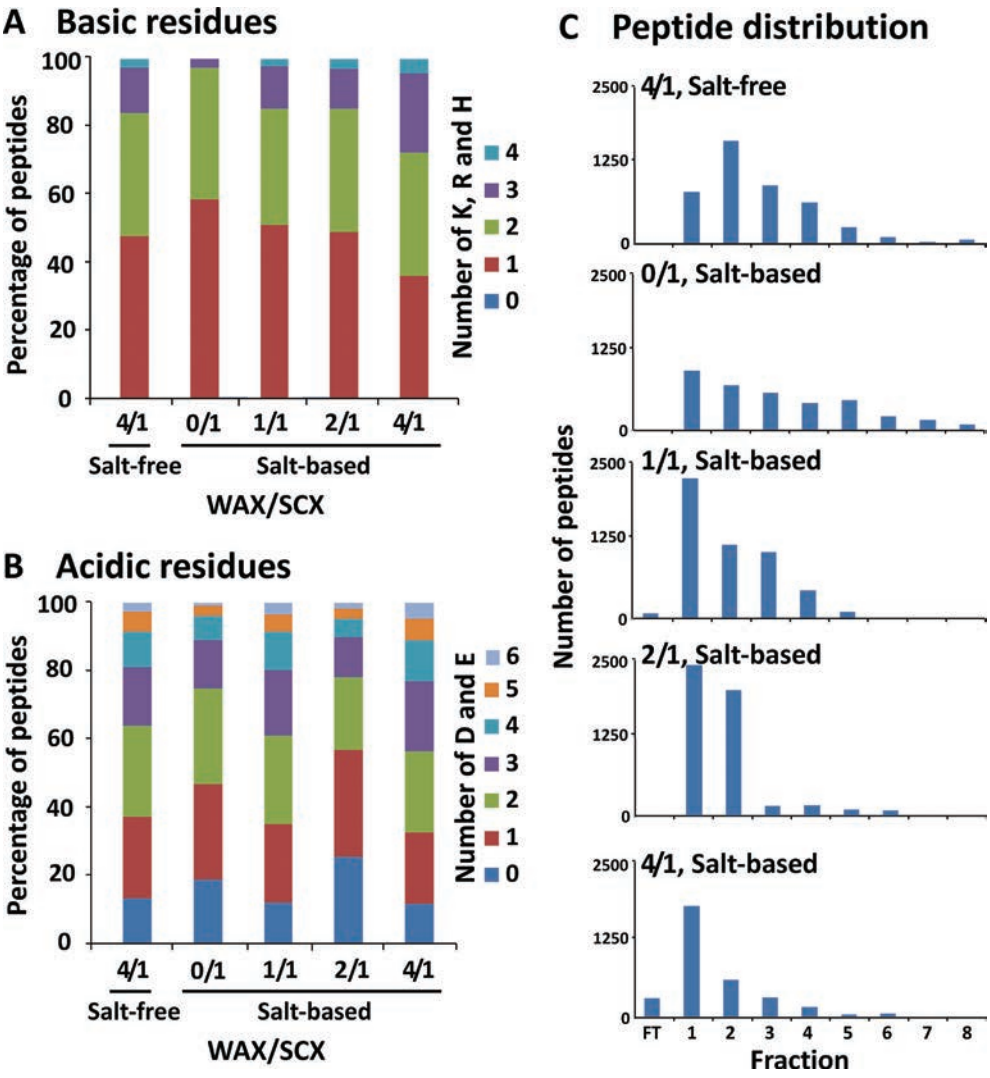


FIGURE 4. COMPARISON OF THE ELUTION PROFILE OF SALT-FREE AND SALT-BASED ACE MUDPIT. Distribution of the number of basic (A) and acidic (B) amino acid residues in peptides recovered by elution with salt-free and salt-based buffers of ion exchange columns with different WAX/SCX mixing ratios. The WAX/SCX ratios investigated were 0/1 (SCX only), 1/1, 2/1 and 4/1 (w/w). (C) The number of peptides recovered in individual fractions by step-wise elution of the different ion exchange columns with salt-free (formic acid/dimethylsulfoxide, see table 1) and salt-based (10-500 mM ammonium acetate) elution buffers. Second dimension separations were carried out on a 50 µm I.D. reversed phase analytical column (3-h gradient).

CONCLUSIONS

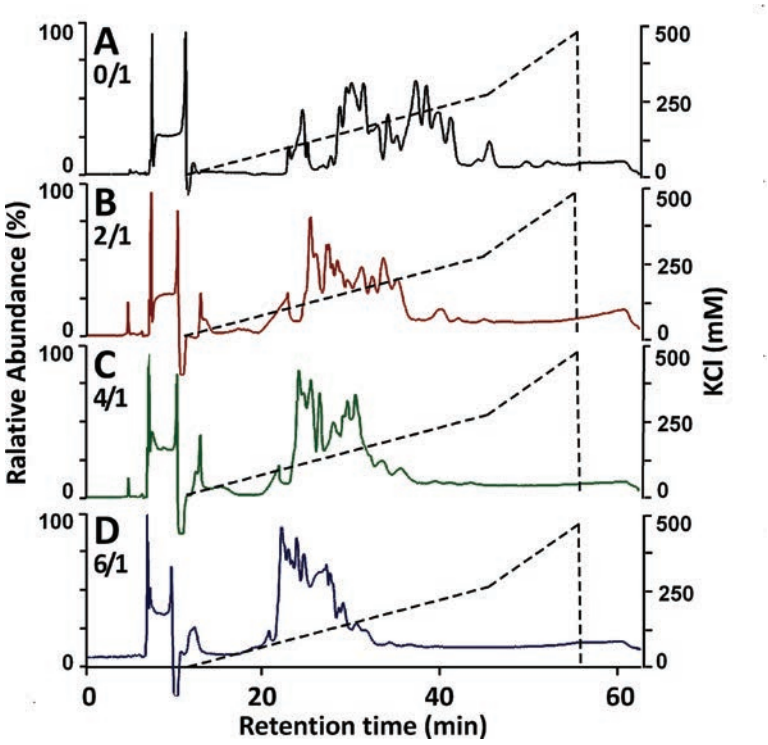
An improved alternative for the on-line, 2-dimensional LC-MS/MS analysis of complex whole cell digests is reported. The method is based on the previously described weak anion and strong cation exchange mixed-bed (ACE) for first dimensional separation and revered phase (RP) for the second dimensional separation. We demonstrate that the balance between peptide retention and elution is dependent on both the ionic strength of the mobile phase system and WAX/SCX mixing ratio. This property appeared to have several advantages in the 2D analysis of complex peptide mixtures. First, one can tailor the chromatographic behavior of peptides to the nature of the sample (*e.g.* high pI proteins) or, as demonstrated in this report, to the eluotropic strength of salt-free elution buffers, simply by changing the WAX/SCX ratio in the mixed-bed column. Second, we found significantly improved peptide recoveries in comparison to SCX in particular for peptides possessing multiple basic amino acids. The loss of these highly cationogenic peptides in SCX chromatography is most likely due to the insufficient elution strength of ammonium acetate. Third, by omitting the use of undesirable salts in LC-MS we addressed one of the main criticisms of MudPIT. Moreover, the excellent compatibility with downstream RP-LC-MS analysis allowed us to hyphenate the ACE trapping column with a true nanoflow 25 µm I.D. analytical column for significant improvements in sensitivity and peptide and protein identification results. The proposed 2D separation method combines ion exchange chromatography and RP chromatography at salt-free, MS compatible conditions and allowed the ultra-sensitive proteomic analysis of thousands of proteins form a very modest amount of sample material.

ACKNOWLEDGEMENTS

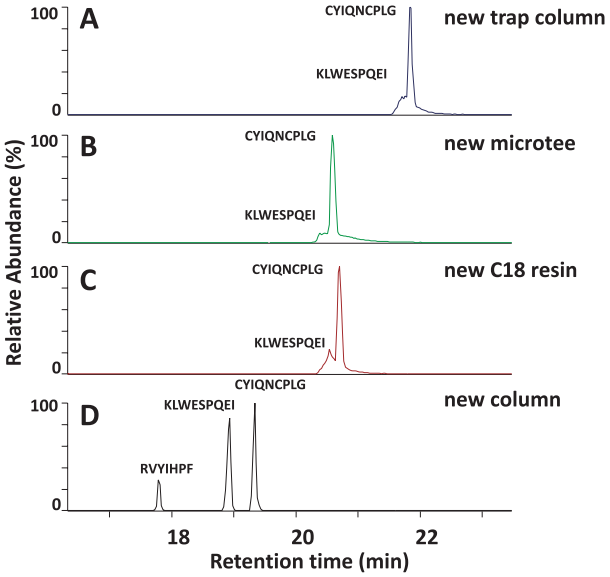
The Netherlands Proteomics Centre, embedded in the Netherlands Genomics Initiative, is kindly acknowledged for financial support.

ASSOCIATED CONTENT

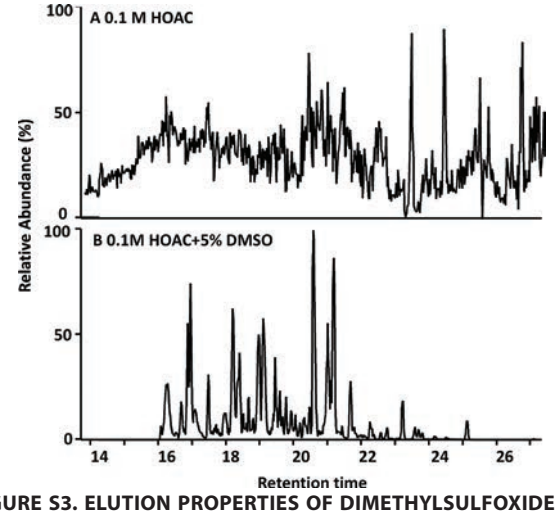
Additional information as noted in the text.



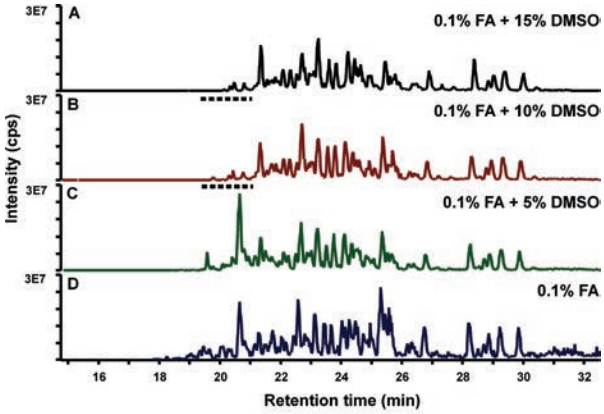
SUPPLEMENTARY FIGURE S1. ION EXCHANGE CHROMATOGRAPHY OF TRYPTIC PEPTIDES. Comparison of the elution window for tryptic peptides (bovine serum albumin) eluted by a salt gradient (KCl, 0-500 mM, dotted line) from columns packed (10 cm long, 200 µm I.D) with strong cation exchange (SCX, A) and weak anion exchange - strong cation exchange (ACE) mixed-beds (traces B-D). The ratio WAX/SCX in traces B-D was 2/1 (w/w), 4/1 (w/w) and 6/1 (w/w), respectively. Detection, UV at 215 nm.



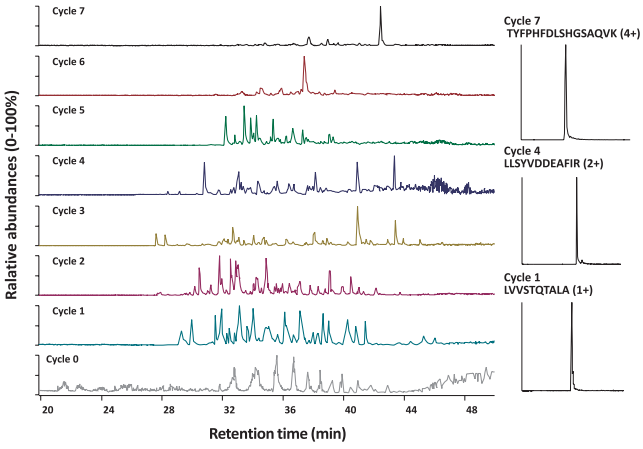
SUPPLEMENTARY FIGURE S2. ADVERSE EFFECTS OF EXPOSURE TO INORGANIC SALT OF REVERSED PHASE CHROMATOGRAPHY. Shown is the loss of performance of the RP column for 2 out of 3 standard peptides ((RVYIHPF, KLWESPQEI, CYIQNCOLG(NH_2)) after exposure to 0.5 M KCl (10µL). Parts of the LC system were replaced by unexposed materials. No improvement was obtained after replacement of the trapping column (A), the microtee's (B) and repacking the analytical column (C). The performance of the system could only be restored with a brand new analytical column (D).



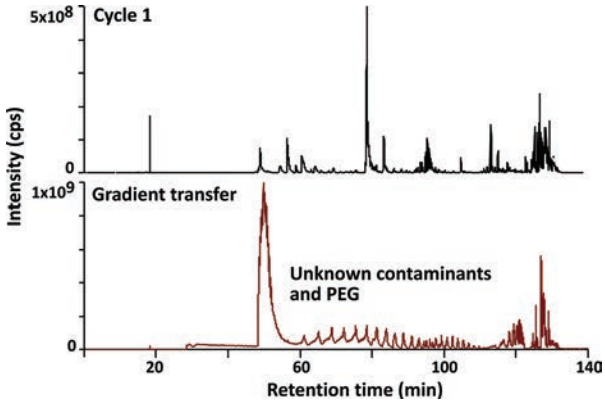
SUPPLEMENTARY FIGURE S3. ELUTION PROPERTIES OF DIMETHYLSULFOXIDE (DMSO) FOR STRONG CATION EXCHANGE (SCX)-BOUND PEPTIDES. Shown are base peak traces of the analysis of peptides recovered from a SCX trap by elution with acetic acid (0.1 M) without (A) and with (B) DMSO (5 %, v/v). The SCX bed (20 mm) was sandwiched between two C18 beds of 20 mm each, packed in a 100 µm I.D. microcapillary and coupled to a 50 µm I.D. RP analytical column in the second separation dimension. RP gradient, 0 to 60% acetonitrile (AcN) in 30 min.



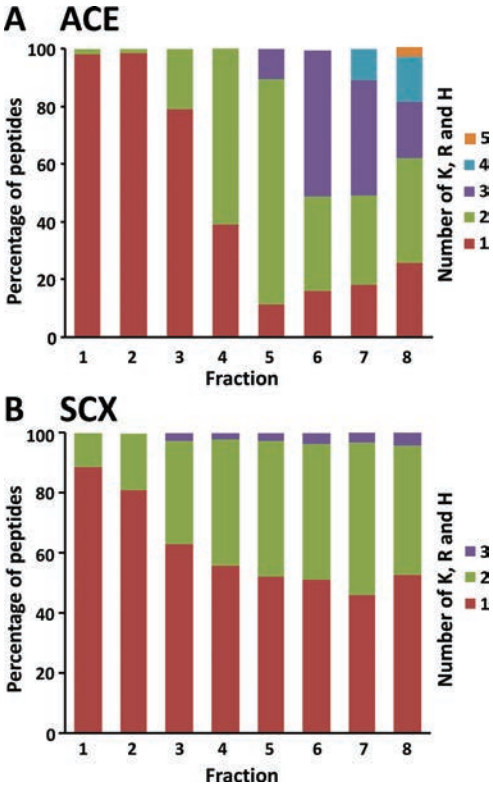
SUPPLEMENTARY FIGURE S4. EFFECT OF DIMETHYLSULFOXIDE (DMSO) ON THE TRAPPING OF PEPTIDES BY REVERSED PHASE (RP) SOLID PHASE EXTRACTION. A standard tryptic peptide mixture (5 protein digest) was dissolved in 0.1% formic acid (FA) fortified with different DMSO concentrations and analysed by RP chromatography using a vented column system. Samples were loaded onto a RP trapping column (100 µm I.D., 20 mm long) at a flow rate of 5 µL/min with the sample effluent directed to waste. Subsequent gradient analysis was performed using a 50 µm I.D RP analytical column. Peptide losses during trapping are indicated by the dotted lines.



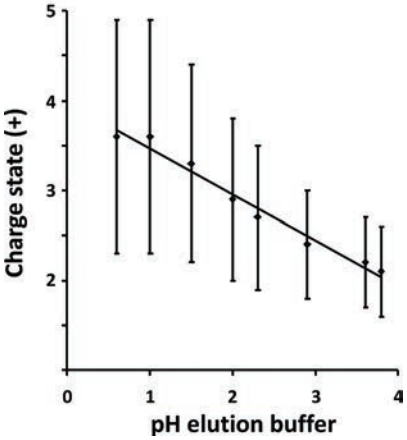
SUPPLEMENTARY FIGURE S5. SALT-FREE ACE MUDPIT SEPARATION OF TRYPTIC PEPTIDES. Shown are the LC-MS base peak traces of analysis of subsequent fractions obtained by elution of a weak anion exchange-strong cation exchange mixed bed (ACE, WAX/SCX ratio 4/1, (w/w)) by pulses of salt-free elution buffers containing formic acid and dimethylsulfoxide of increasing strength (see Table 1). The flow through during RP to ACE sample transfer (cycle 0) comprised of neutral (uncharged) sample constituents, like poly ethylene glycols (PEG), detergents and other neutral chemicals. Peptide (+) charge states gradually increase in subsequent first dimensional fractions as illustrated by the extracted ion traces of arbitrary chosen peptides (right hand panels).



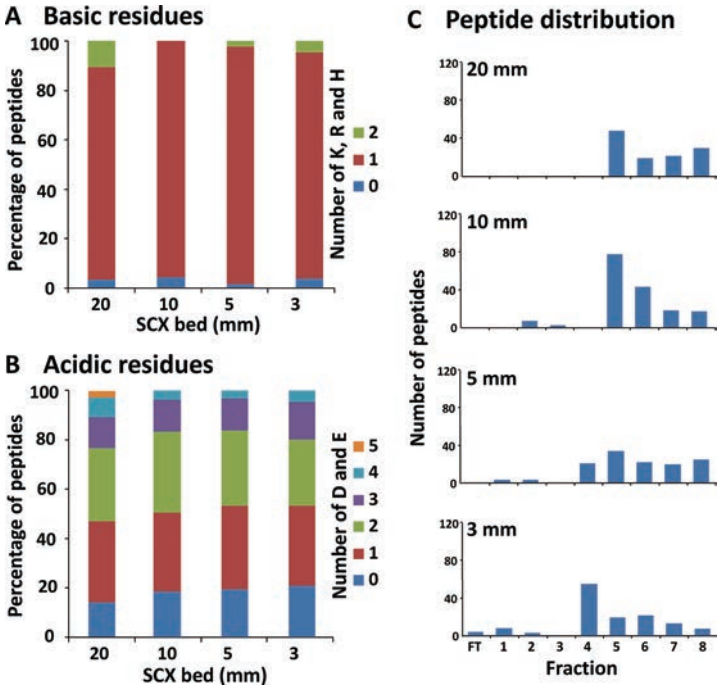
SUPPLEMENTARY FIGURE S6. EFFECTIVE SAMPLE CLEANUP OF UNCHARGED SAMPLE CONTAMINANTS BY ACE-RP. Shown are the base peak chromatograms after sample loading and gradient transfer (lower trace) and the first peptide fraction (upper trace). Samples obtained from immunoaffinity purification of detergent-solubilized protein-protein complexes which are dominated by uncharged sample contaminants (e.g polyethylene glycols (PEG)) can be effectively purified by ACE-RP. These contaminates are not retained by the ACE resin and hence end up in the flow-through fraction (Cycle 0), while ACE-bound peptides are subsequently recovered by a discontinuous pH gradient (e.g. Cycle 1).



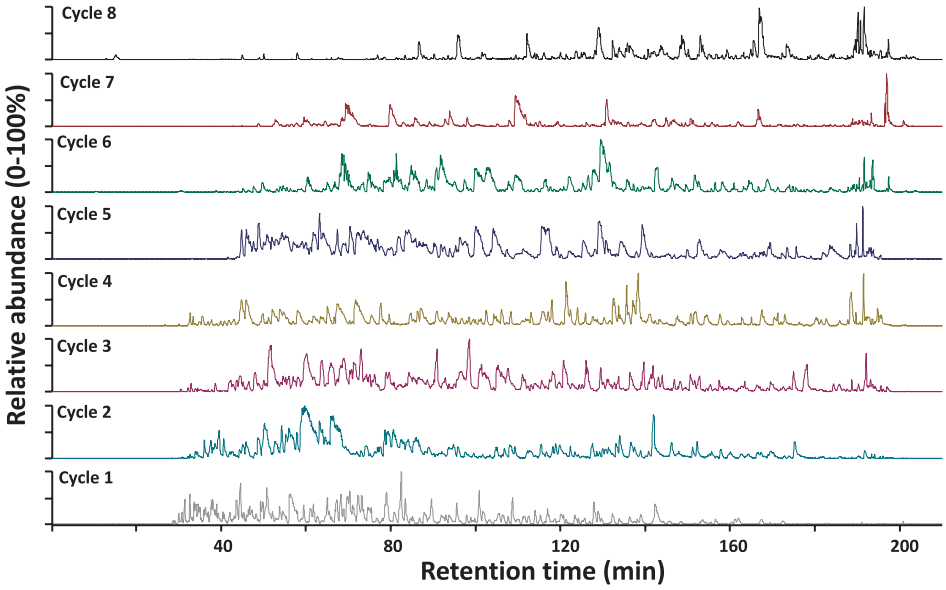
SUPPLEMENTARY FIGURE S7. RECOVERY OF MULTIPLE CHARGED PEPTIDES FROM ION EXCHANGE COLUMNS. Distribution (normalised) of the number of basic amino acid residues (Arg, Lys and His) in peptides recovered by elution of a weak anion exchange-strong cation exchange (ACE) mixed bed (A) and a strong cation exchange (SCX) column (B) in the first separation dimension. The WAX/SCX ratio in the ACE column was 4/1, (w/w). The SCX column was fractionated by a discontinuous salt gradient of ammonium acetate (0-500 mM) in 1 mM acetic acid / 2% (v/v) acetonitril and the ACE column by a discontinuous gradient of formic acid and dimethylsulfoxide in water (Table 1). Fractions were analysed by reversed phase LC-MS/MS using a shallow, 3-h gradient. Singly charged (0 basic residues) peptide features were excluded from MS/MS targetting. Sample: 1 ug tryptic digest of *S. cerevisiae.*



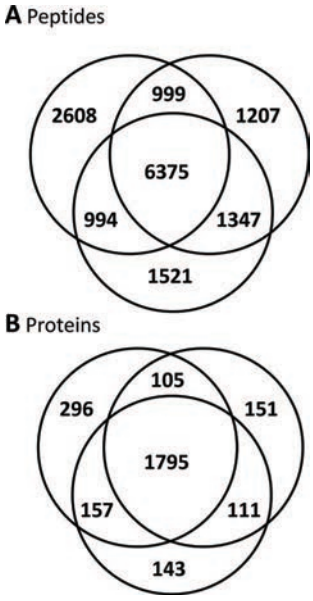
SUPPLEMENTARY FIGURE S8. CHARGE STATE (AVERAGE +/- SD) OF PEPTIDES RECOVERED FROM A WEAK ANION EXCHANGE AND STRONG CATION EXCHANGE (ACE) MIXED BED (WAX/SCX 4/1,(W/W)) BY ELUTION WITH AQUEOUS FORMIC ACID /DIMETHYLSULFOXIDE (TABLE 1).



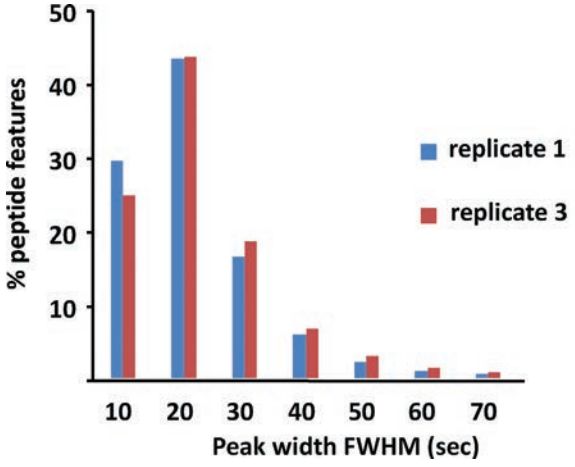
SUPPLEMENTARY FIGURE S9. EVALUATION OF THE SCX COLUMN LENGTH IN SALT-FREE MUDPIT. (A/B) Distribution of the number of basic (A) and acidic (B) amino acid residues in peptides recovered by elution with acidic, salt-free buffers of strong cation exchange (SCX) columns. The SCX columns investigated were 20 mm, 10 mm, 5 mm and 3 mm in length. (C) Number of peptides recovered in individual fractions during a discontinuous gradient pH elution (Table 1). Fractions were analysed by reversed phase LC-MS/MS (60 min gradient). Sample: standard 5 protein tryptic digest.



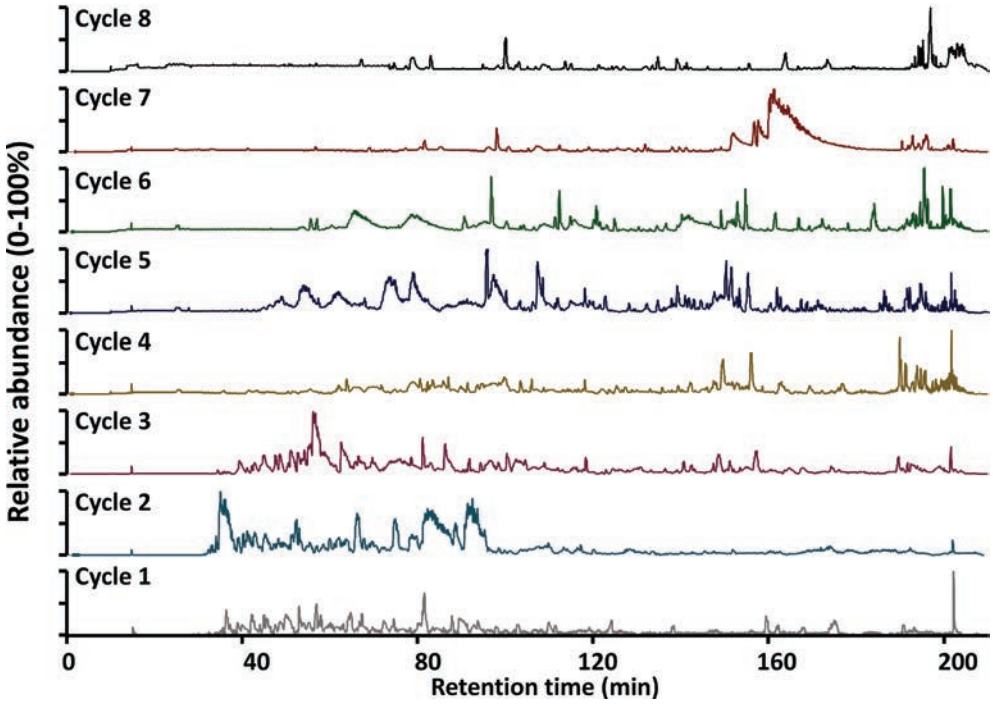
SUPPLEMENTARY FIGURE S10. BASE PEAK CHROMATOGRAMS OF SALT-FREE ACE MUDPIT ANALYSIS OF A TRYPTIC DIGEST OF S. CEREVISIAE (1 µg CONSUMED). Peptides were fractionated by employing weak anion exchange and strong cation exchange (ACE) mixed-bed (WAX/SCX 4/1 (w/w)) column in the first separation dimension and a reversed phase bed column in the second (180 min gradients). Cycle 1 through 8 represent the consecutive RP LC-MS runs following each first dimensional fractionation step with formic acid/dimethylsulfoxide elution buffers (Table 1).



SUPPLEMENTARY FIGURE S11. VENN DIAGRAMS OF THE OVERLAP OF (A) IDENTIFIED UNIQUE PEPTIDES AND (B) PROTEINS BETWEEN TRIPLICATE ANALYSIS OF A TRYPTIC *S. CEREVISIAE* SAMPLE BY SALT-FREE ACE MUDPIT.



SUPPLEMENTARY FIGURE S12. REPEATABILITY OF SALT-FREE ACE MUDPIT. Peak width (FWHM) distribution of peptide features separated by second-dimension reversed phase chromatography (180 min gradients). Comparison of peak widths obtained in first (bleu bars) and third (red bars) replicate salt-free ACE MudPIT analysis. Peptides features were extracted from raw mass chromatograms by the MS-Xelerator software package for multiple charged peak clusters with intensities > 10^5 counts. Tryptic peptides of *S. cerevisiae* were fractionated by employing weak anion exchange and strong cation exchange (ACE) mixed-bed (WAX/SCX 4/1 (w/w)) and a discontinuous pH gradient (Table 1).



SUPPLEMENTARY FIGURE S13. BASE PEAK CHROMATOGRAMS OF SALT-BASED SCX MUDPIT ANALYSIS OF A TRYPTIC DIGEST OF *S. CEREVISIAE* (1 µg). Peptide fractionation was accomplished by SCX using a 8-step discontinuous salt gradient (5mM – 500mM ammonium acetate in 100mM HOAc and 2% acetonitrile). Each fraction was analyzed by reversed phase LC-MS analysis (180 min gradients) using an ultra-narrow, 25µm ID analytical column (40 cm long, 3µm Reprosil Pur C18).

CHAPTER 05

Expanding the detectable HLA peptide repertoire using Electron-Transfer / higher-energy Collision Dissociation (EThcD)

Geert P. M. Mommen 1,2,3
Christian K. Frese 2,3
Hugo D. Meiring 1
Jacqueline van Gaans-van den Brink 4
Ad P. J. M. de Jong 1
Cecile A.C.M. van Els 4
Albert J. R. Heck 2,3

1 Formulation and Analytical Research, Institute for Translational Vaccinology, Bilthoven , The Netherlands
2 Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, The Netherlands
3 Netherlands Proteomics Centre, Utrecht, The Netherlands
4 Center for immunology of infectious diseases and vaccines, National Institute for Public health and the Environment, Bilthoven, The Neterlands

Submitted

ABSTRACT

The identification of peptides presented by human leukocyte antigens (HLA) class I is tremendously important for the understanding of antigen presentation mechanisms under healthy or diseased conditions. Currently, mass spectrometry (MS)-based methods represent the best methodology for the identification of HLA class I associated peptides. However, the HLA class I peptide repertoire remains largely unexplored because the variable nature of endogenous peptides represents difficulties in conventional peptide fragmentation technology. Here, we substantially enhanced (about 3-fold) the identification success rate of peptides presented by HLA class I using combined electron-transfer/higher-energy collision dissociation (EThcD), reporting over 12,000 high-confident (FDR <1%) peptides from a single human B cell line. The direct importance of such unprecedented large data set is highlighted by the discovery of new features in antigen presentation. The observation that a substantial part of proteins is sampled across different HLA alleles, and the common occurrence of HLA class I 'nested sets', suggest that the constraints of HLA class I to comprehensively present the health states of cells are not as tight as previously thought. Our data set contains a substantial set of peptides bearing a variety of post-translational modifications presented with marked allele-specific differences. We propose that EThcD should become the method of choice in analyzing HLA class I presented peptides.

INTRODUCTION

Class I molecules of the Human Leukocyte Antigen (HLA) complex present short peptides, typically 8 to 11 amino acids in length at the cell surface for scrutiny by the immune system [5]. These peptide fragments are generated in the cytoplasm by proteasomal degradation of source proteins, translocated into the endoplasmic reticulum (ER) and subjected to N-terminal trimming to a size that is suitable for loading onto HLA [7]. Loading is governed by physico-chemical binding motifs typical for each HLA class I allele [215]. Depending on the motif required for the HLA class I allele(s) expressed, an ER residing peptide may become presented or not. Recognition of specific HLA class I peptide complexes by CD8 T lymphocytes on pathogen infected or cancerous cell leads to the activation of a cytotoxic response and the clearance of the diseased cell. The identification of these HLA class I associated peptides has important consequences for understanding the biology of cells, vaccine design and tumor immunotherapy [9, 216].

Mass spectrometry (MS) is nowadays the core technology for the analysis of HLA class I presented peptides. These peptides are typically enriched from cell lysates by affinity purification of HLA class I peptide complexes, released from the HLA by acid elution and separated by liquid chromatography prior to introduction into the mass spectrometer. Identification is commonly accomplished by MS sequencing using collision-induced dissociation (CID) or beam-type higher-energy collision induced dissociation (HCD) [91]. Both methods generate the peptide fragment ions that can be used for sequence identification in automated database search strategy or de novo sequence analysis. These methods have been thoroughly optimized and work particularly well for tryptic peptides that are produced by *in vitro* digestion of proteomes. However, for endogenously processed peptides, such as HLA class I associated peptides, only a small fraction of the acquired tandem MS spectra contains sufficient sequence-diagnostic information for correct assignment of the peptide sequence [91, 96]. To improve peptide identifications, the alternative fragmentation method electron-transfer dissociation (ETD) can be used, which complements CID particularly for longer and more basic peptides [59, 217]. However, both CID and ETD may suffer from the limited sequence information concealed in the short HLA class I associated peptide sequences and incomplete peptide fragmentation due to the occurrence of certain amino acid residues that are known to hamper efficient backbone dissociation [217]. As a consequence, we hypothesized that a 'treasure' of peptides presented by HLA class I molecules might still be in oblivion for identification by current MS technology.

Recently, we introduced a novel fragmentation scheme termed electron-transfer and higher-energy collision dissociation, EThcD [218, 219]. This method employs dual fragmentation to a single ion package to generate both the fragment ions induced by ETD (c/z) and HCD (b/y) in a single spectrum. The generation of dual fragment ion series results in more informative MS/MS spectra that enable highly confident peptide assignment and localization of post-translational modifications [218, 219]. In this study, we explore the power of dual fragmentation EThcD for the analysis of peptides presented by HLA class I molecules on human B lymphoblastoid cells. Our results revealed a supreme performance of EThcD in comparison to all established methods (CID, HCD, ETD), resulting in an unprecedented inventory of HLA class I associated peptides.

EXPERIMENTAL

CULTURED CELL LINE AND ISOLATION OF HLA CLASS I ASSOCIATED PEPTIDES. The HLA-A*01,-03, B*07,-27, -C*02,-07 positive B lymphoblastoid cell-line GR was grown in RPMI-1640 medium to a total number of 9 x 10^9 cells. HLA class I peptide complexes were immunoprecipitated from lysed GR cells essentially as described previously [220], using the HLA-A,-B,-C specific mouse monoclonal IgG2a antibody W6/32. HLA class I associated peptides were eluted with 10% acetic acid and collected by passage over a 10-kDa mw cutoff membrane.

Mass spectrometry. For 1D analysis, the HLA elution sample was analyzed directly by nanoscale LC-MS using a Thermo Scientific EASY-nLC 1000 (Thermo Fisher Scientific, Odense, Denmark) and ETD enabled LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Bremen, Germany), as described previously [218]. The system comprises an 20 mm x 100 µm ID trapping column (Reprosil C18, 3 µm, Dr Maisch, Ammerbuch, Germany) and a 40 cm x 75 µm ID analytical column (Zorbax SB-C18, 1.8 µm). Full MS spectra (*m/z* 300 to 1,500) were acquired in the Orbitrap at 60,000 resolution (FWHM). The 10 most abundant precursor ions were selected for either data-dependent EThcD, CID, HCD or ETD fragmentation (unknown and 1+ charge state excluded). For sequential CID/HCD analysis, each precursor ion was sequentially selected for CID and HCD (top 5 method). The maximum ion accumulation time for MS scans was set to 200 ms and for MS/MS scans to 2500 ms. Fragment ions were detected in the Orbitrap analyzer at 15,000 resolution (FWHM). Dynamic exclusion was enabled with a repeat count of 1 and 60 s exclusion duration. The background ions at *m/z* 391.2843 and 445.1200 were used as lock mass.

For the 2D strategy, HLA class I eluted peptides were fractionated by strong cation exchange (SCX) chromatography, as described previously [98]. The system comprises an Hypercarb™ trapping column (200 µm I.D., 5 mm, 7 µm particle size, Thermo Fisher) and SCX column (200 µm I.D., 12 cm, polysulfoethyl aspartamide, 5 µm, Poly LC). From a total number of 26 SCX fractions, the 9 most informative were subjected to LC-MS/MS, as described for 1D analysis.

DATA ANALYSIS. The analysis of mass spectrometric RAW data was carried out using Proteome Discoverer 1.4 software package (Thermo Fisher Scientific, Bremen, Germany). The nonfragment filter was used to simplify ETD spectra and the spectrum grouper function to merge consecutive CID and HCD spectra. MS/MS scans were searched against the human Uniprot database with no enzyme specificity using the SEQUEST HT mode. Precursor ion and MS/MS tolerances were set to 3 ppm and 0.02 Da, respectively. Asparagine deamidation, methionine oxidation and cysteinylation were set as variable modifications. Additional searches included phosphorylation (S, T, Y), or N-terminal glutamate cyclization (ammonia loss). Results were filtered to < 1% FDR using percolator [221], Xcorr > 1.75 and rank = 1. NetMHC 3.4 or the NetMHCpan 2.8 algorithm was used to predict the HLA-peptide binding affinities for each identified peptide sequence [222, 223].

RESULTS

EXPERIMENTAL APPROACH FOR THE IDENTIFICATION OF HLA CLASS I ASSOCIATED PEPTIDES. To evaluate the full and unbiased potential of EThcD for the analysis of the HLA class I associated peptides, we analyzed a complex repertoire of peptides presented by various class I molecules on the surface of an HLA-A, -B and -C heterozygous B lymphoblastoid cell line, GR. HLA class I peptide complexes were isolated from a GR lysate by affinity purification, bound peptides were released by acid elution and analyzed by reversed phase liquid chromatography – tandem mass spectrometry (LC-MS/MS) [19]. This 1D strategy was employed to assess the performance of all possible peptide fragmentation techniques currently available on the here employed Orbitrap Elite instrument. To evaluate whether sequencing short, endogenously processed peptides benefits from the combined information generated by two different fragmentation modes, we compared EThcD fragmentation with sequential CID/HCD (*in silico* spectral merging) or single fragmentation methods CID, HCD and ETD (Fig. S1). For each fragmentation technique, fragment ion spectra (MS/MS) were acquired with high mass-accuracy and resolution using the Orbitrap analyzer, not only to ensure consistency between the datasets, but also to improve specificity in the database search analysis [59, 60]. MS/MS spectra were searched against the human proteome using SEQUEST and filtered to a <1% false discovery rate (FDR) using the percolator algorithm [221]. We found that percolator performed approximately a 2-fold better than a standard target-decoy approach [32], irrespective of the fragmentation method used.

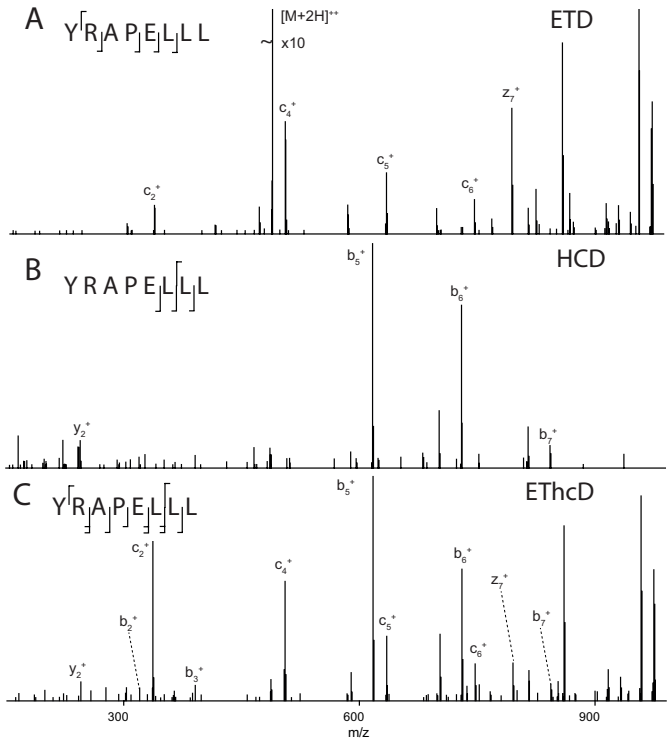


FIGURE. 1. PEPTIDE FRAGMENTATION BY ETD, HCD AND ETHCD. Illustrative MS/MS spectra of the HLA-B27-associated peptide YRAPELLL upon fragmentation by (A) ETD, (B) HCD and (C) EThcD. The observed and/or assignable fragment ions of the types c/z and b/y are indicated above and below the peptide sequence.

ELECTRON TRANSFER / HIGHER-ENERGY COLLISION DISSOCIATION (ETHCD) BOOSTS THE IDENTIFICATION OF HLA CLASS I ASSOCIATED PEPTIDES. The peptide identification results from the 1D strategy are summarized in Table 1 (*i.e.* single LC runs). Although the number of MS/MS spectra were roughly identical (~ 12,000), the identification success rate and the number of uniquely identified peptides clearly indicate that EThcD fragmentation by far outperforms all other fragmentation techniques. We found that 39% of the MS/MS attempts by EThcD lead to high-confident peptide assignments, which is approximately a factor 3 higher in comparison to HCD and ETD alone. The excellent performance of EThcD can be directly attributed to the more extensive backbone fragmentation and the generation of information-rich MS/MS spectra. As exemplified in Fig. 1, dissociation of the HLA-B27 restricted peptide YRAPELLL by HCD and ETD, respectively, generated limited sequence information, primarily due to the near complete absence of fragment ions derived from the peptide C-terminus. Dual fragmentation EThcD generated the complementary c/z and b/y ions, enabling high-confidence sequence assignment. The improvement in spectral quality is also reflected in the SEQUEST Xcorr distributions of the complete data set, as significant higher scores were found for EThcD due to the assignment of both c/z and b/y ions (Fig. S2) [218, 219]. These findings were further sustained by sequential CID/HCD, which performed significantly better than single stage fragmentation (CID and HCD alone) (Table 1).

To benchmark and compare these results, we analyzed a tryptic *E. coli* digest (5 ng) and found that the identification success rate of EThcD for sequencing endogenous peptides agrees well with the success rate for low abundant tryptic peptides (39% versus 30%). However, in contrary to endogenous peptides, for tryptic peptides the success rate of EThcD and HCD is comparable (30 and 34% respectively), which is in agreement with previous data [218, 219]. Together, the data unambiguously demonstrate that sequencing short, endogenous peptides is rather cumbersome when using standard HCD or ETD, and considerably benefits from combining the complementary sequence information induced by EThcD.

TABLE 1. Summary of data obtained in the LC-MS/MS analysis of HLA-associated peptides presented by GR B cells. [a] The number of merged CID/HCD spectra used for database search analysis. b Identification rate was calculated from twice the number of merged CID/HCD spectra.

	1D strategy					2D strategy		
Data set	EThcD	CID/HCD	CID	HCD	ETD	EThcD	CID/HCD	Combined data
MS/MS spectra	11,777	10,214[a]	13,394	16,054	11,223	39,156	42,227[a]	196,485
PSM (<1% FDR)	4,532	2,418	1,319	1,948	2,545	13,555	8,794	35,111
Identification rate	39%	12%b	10%	12%	23%	35%	10%b	19%
Unique peptides	3,454	2,142	1,273	1,737	2,288	9,255	6,554	12,699
Unique peptides 8 -14 a.a	3,381	2,027	1,174	1,622	2,215	9,015	6,381	12,199
Proteins	2,205	1,517	982	1,268	1,650	4,536	3,706	5,603
Analysis time	3 h	3 h	3 h	3 h	3 h	9 x 2 h	9 x 2 h	51 h

GLOBAL MAPPING OF THE HLA CLASS I PEPTIDE REPERTOIRE. To more comprehensively profile the repertoire of peptides presented by HLA class I, we next employed a two-dimensional (2D) peptide separation strategy using SCX fractionation followed by LC-MS/MS analysis. Here, we focused on exclusively using EThcD and sequential CID/HCD because these two methods performed best based on the initial 1D strategy. In total, 9,015 and 6,381 unique 8-14mer peptides were identified by EThcD and sequential CID/HCD, respectively (Table 1). The most optimal balance between peptide identification coverage and the required analysis time was found by SCX fractionation and LC-MS/MS analysis using EThcD (Fig. 2D), covering 74% of the complete set of unique HLA class I peptides in less than one day of instrument time.

To validate whether the identified peptides were indeed initially bound to HLA molecules, we predicted the HLA class I binding affinities using the NetMHC algorithm [222]. Importantly, 90% of the identified peptides were predicted as strong binders (< 1000nM IC50) to a single HLA-A and B allele on GR cells, which further strengthens our confidence in the correctness of the peptide assignments. Using sequence-based rules on the cumulative dataset, the NetMHC algorithm assigned 1,170 peptides to HLA-A1 (9%), 2,132 peptides to HLA-A3 (17%), 4,314 peptides to HLA-B7 (35%), and 3,530 peptides to HLA-B27 (29%), while 1,234 peptides could not be classified on these terms (N.A.). The latter peptides include besides weak predicted binders also a low contribution of peptides bound to the HLA-C alleles expressed. The less accurate NetMHCpan algorithm [223] allowed assignment of 277 peptide sequences to HLA-C (Fig. S3), but we found their numbers relative to the total assigned repertoire too low to further specify in this study.

Remarkably, comparison of the complete data set with recent large-scale studies revealed that 81% of the peptides have not been reported before (Table S2) [224-226]. To our knowledge, our data set comprises the largest inventory of unique and unambiguously identified HLA class I associated peptides based on a single cell line to date.

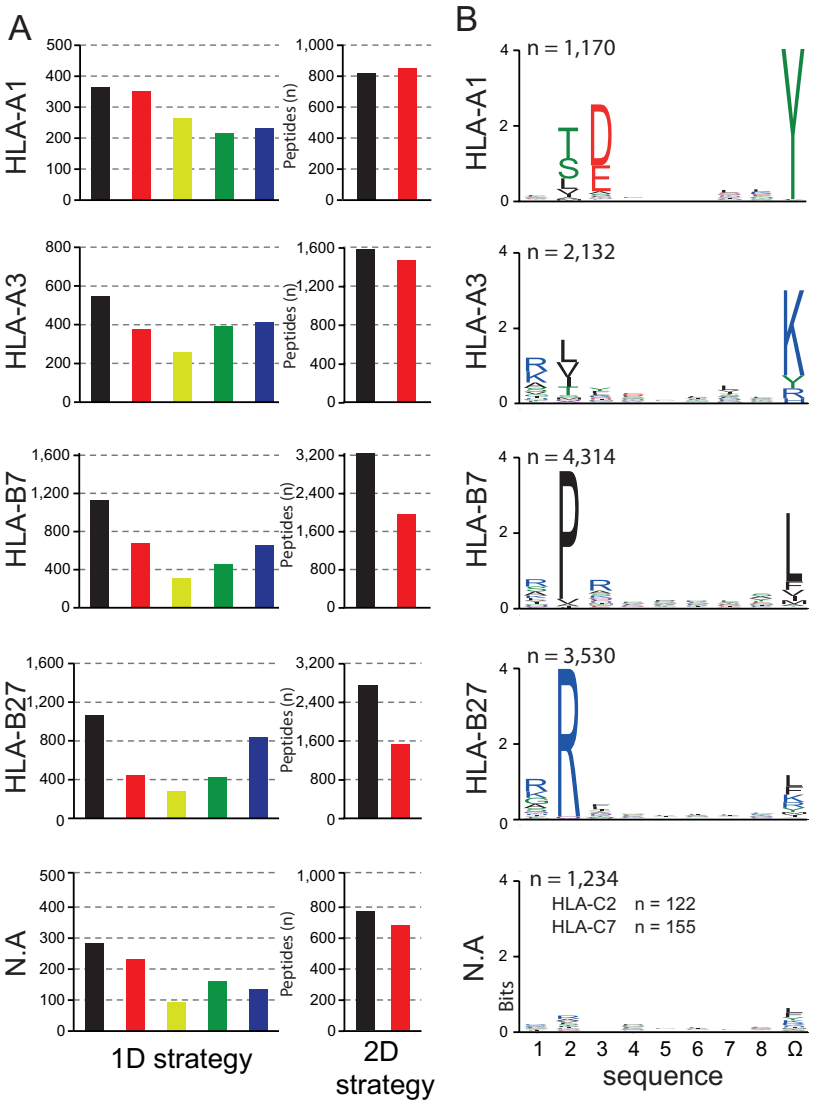


FIGURE 2. PERFORMANCE CHARACTERISTICS OF PEPTIDE FRAGMENTATION TECHNIQUES FOR HLA CLASS I ASSOCIATED PEPTIDES. Shown are identification results and binding motifs of peptides presented by different HLA class I molecules on the surface of GR cells. (A) The number of unique 8-14 mer peptides identified by the 1D and 2D peptide separation strategy, assigned to HLA-A1, -A3, -B7, -B27 molecules, or not assigned to any of the above (N.A). The 1D strategy comprised consecutive single LC-MS/MS analysis runs, employing either EThcD (black), sequential CID/HCD (red), CID (yellow), HCD (green) or ETD (blue). The 2D strategy included the analysis of strong cation exchange (SCX) fractions by LC-MS/MS using EThcD (black) or sequential CID/HCD (red). (B) Sequence logo's showing that peptides are bound to a particular HLA molecule through a distinct binding motif, typically by anchor residues on position P2 (P3) and the C-terminal PΩ residue.

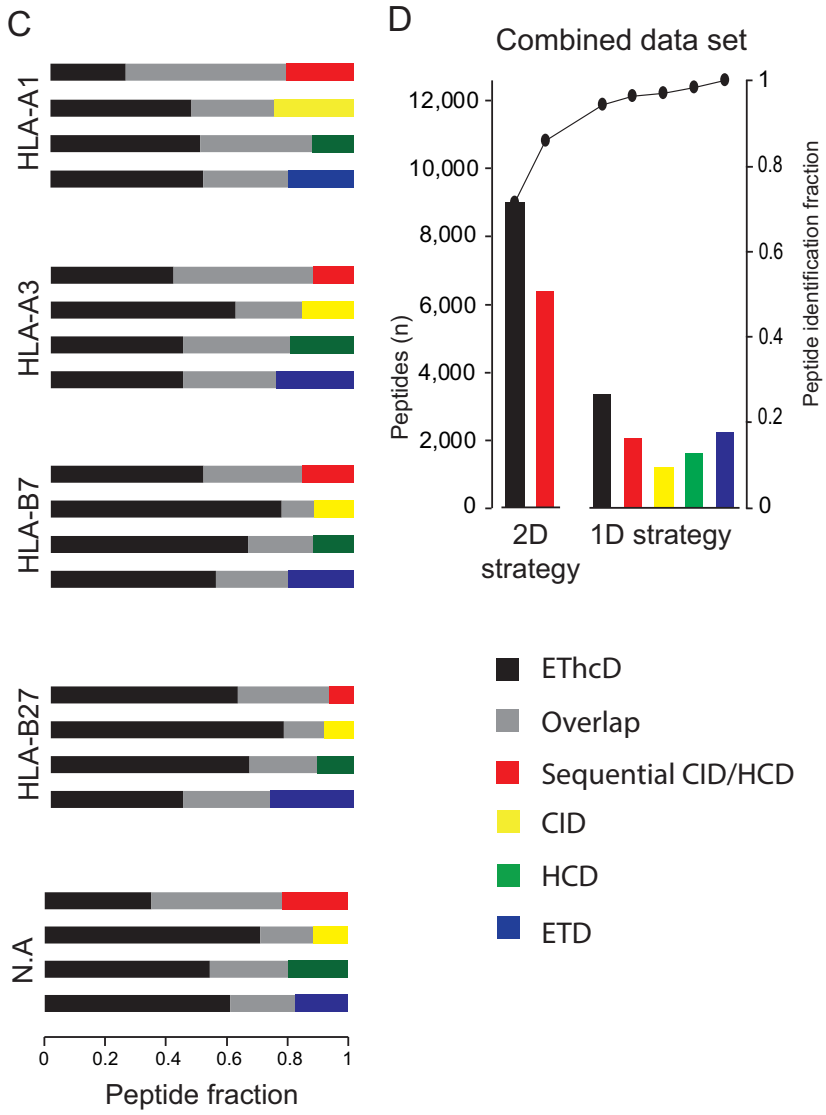


FIGURE 2 (CONTINUED). (C) Bar diagrams depicting the peptide fraction identified by EThcD (black), the overlap between the methods (grey) and uniquely by either CID/HCD (red), CID (yellow), HCD (green), or ETD (blue). (D) Number of unique peptides identified by the different peptide fragmentation methods (color-coded) in the 1D and 2D strategy. The line graph depicts the cumulative number of unique HLA class I associated peptides (left y-axis) and the peptide identification fraction (right y-axis).

FRAGMENTATION-RELATED BIAS IN PROFILING THE HLA CLASS I PEPTIDE REPERTOIRE. To test for fragmentation-related biases, we further specified the identification results for each of the applied fragmentation method with respect to the different HLA class I alleles (Fig. 2). This evaluation included, for each HLA allele, the number of uniquely assigned peptides (Fig. 2A), the peptide binding consensus (Sequence logo's, Fig. 2B), and the overlap in peptide identification between the fragmentation techniques (Fig. 2C). The observed sequence logo's clearly matched the known binding motifs of the HLA alleles, through amino acids anchor residues at position P2 (P3) and the C-terminal amino acid (PΩ). EThcD provided the largest number of peptide identifications across all HLA alleles (Fig. 2A). The unmatched performance of EThcD was most pronounced for the HLA-B7 and HLA-B27. The peptides bound to these alleles are more difficult to sequence by conventional methods due to the preferred presence of a proline or an arginine residue at the P2 position [217]. A known limitation of ETD is the inability to cleave the N-Cα bond of proline, while CID/HCD suffers from inefficient peptide dissociation due to internal arginine residues that are known to hamper random backbone protonation [64]. The generation of dual ion series in EThcD overcomes these limitations, resulting in an almost 2-fold improvement in the identification of peptides bound to HLA-B7 and HLA-B27.

To compare the performance of EThcD with the conventional fragmentation techniques in more detail, we used the 1D data set to determine the normalized overlap between the peptide identification results for each HLA allele. Fig. 2C shows that the added value of the conventional fragmentation techniques is limited because EThcD accounts for ~80% of the peptide identifications across all HLA alleles. It is worth mentioning, but not surprising, that ETD alone performed particularly well for these HLA-B27-bound peptides. This was reflected by the absolute number of identified peptides (Fig. 2A), as well as the complementary advantage of ETD for sequencing of triply charged peptides (Fig. S4). These findings are in line with previous reports describing that peptides with a strong basic N-terminus generate more straightforward and easy-to-interpret MS/MS spectra in ETD [227].

CHARACTERISTICS OF THE HLA CLASS I PEPTIDE REPERTOIRE. The general characteristics of the HLA class I peptide repertoire identified here add to existing knowledge on the self peptidome. The 12,199 unique peptides originate from 5,603 source proteins, which are primarily located in the ER, nucleus, and cytoplasm (data not shown). Distribution analysis of the number of peptides identified per protein revealed that 48% (2,636 proteins) were represented by a single peptide, while the remaining proteins were represented by multiple unique peptides (Fig. S5). As reported previously [224, 228], the HLA class I sampling rate seems to correlate with protein length (Fig. S6). For proteins that were represented by multiple peptides, we found that 2,295 proteins were represented by at least two different HLA alleles (Fig. S5). Moreover, for the peptides related to these proteins, no relation was found between peptide intensity, allele specificity or predicted binding strength (data not shown), confirming that the final level of presentation depends on many factors (*e.g.* proteolytic activity, TAP transport efficiency) [224, 229].

HLA class I molecules predominantly present peptides of 9 to 11 a.a. in length (Fig. S5). However, there are numerous reports of longer peptides that can also elicit cytotoxic T-cell responses, and some of these unusually long peptides fully overlap in sequence with their counterparts of more conventional length [230]. In our dataset we observed 1,376 peptides pairs with an identical core sequence but having 1 - 5 a.a. length differences. These additional amino acids were detected either at the N-terminus (n=341), the C-terminus (n=829), or with no consensus at their termini (*i.e.* ragged ends at both termini, n=206). Mapping these nested peptide species to different HLA alleles revealed unexpected differences between the frequency of N-terminal- and C-terminal-extended peptide pairs (Fig. 3A-B). For HLA-B alleles, significantly more C-terminal-extended peptide pairs were detected compared to N-terminal extended pairs, similar as reported for HLA-B27 bound peptides [225]. In contrast to these findings in HLA-B, a more equal frequency distribution between N- and C-terminal-extended length variants was found for HLA-A. Frequency analysis of the C-terminal anchor residues revealed more strict constraints on the C-terminal anchor residue of the HLA-A versus the HLA-B alleles (Fig. 3C). For example, the C-terminus of HLA-A1 bound peptide pairs are predominated by a tyrosine residue, while there is less restriction for variants bound to HLA-B27 (Fig. 2B, Fig. 3C). For HLA class I 'nested sets', no relation was found between peptide intensity, predicted binding strength or peptide length (data not shown).

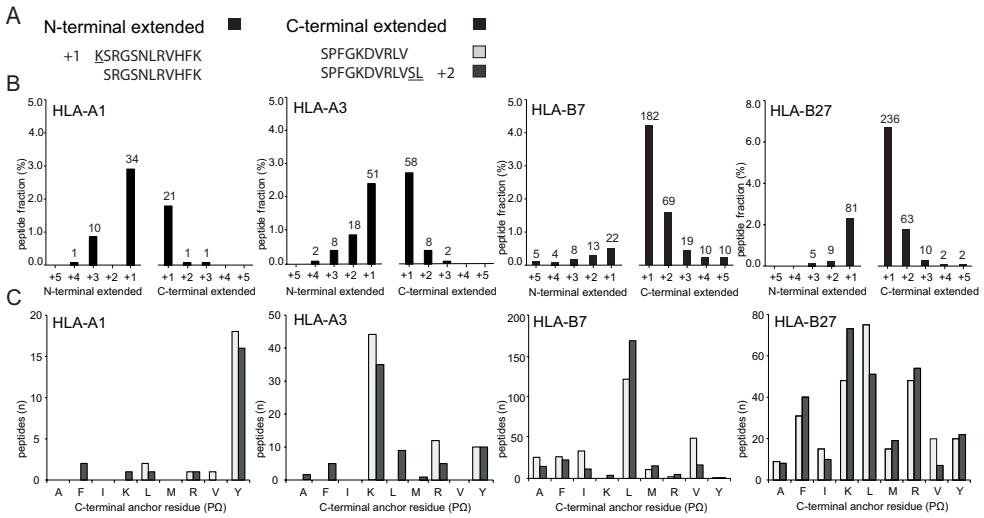


FIGURE 3. FREQUENCY DISTRIBUTION OF HLA CLASS I ASSOCIATED PEPTIDE PAIRS EXHIBITING LENGTH VARIATIONS. (A) N-terminal-extended peptide pairs are defined as having an extension of between 1 to 5 amino acids at the N-terminus, but a preserved C-terminus. Visa versa holds for the C-terminal-extended peptides. (B) Allele-specific frequency distribution of the N-terminal- and C-terminal-extended peptide pairs normalized to the total numbers of peptides assigned to that HLA allele. The absolute number of peptide pairs for each length extension is depicted. (C) Frequency of the C-terminal anchor residues (PΩ) of C-terminal-extended peptide pairs. The PΩ-frequency of the shortest peptide is highlighted in gray and C-terminal-extended peptide in light black. The amino acids depicted in figures represent 90-95% of the identified C-terminal anchor residues.

POST-TRANSLATIONALLY MODIFIED HLA CLASS I PEPTIDES. Peptides presented by HLA molecules can harbor post-translational modifications (PTMs) [231]. Table 2 summarizes the PTMs identified on HLA class I associated peptides in this study. The here frequently observed asparagine deamidation reaction can proceed either spontaneously (aging) or enzymatically. It has been demonstrated that N-glycosylated peptides may be presented by HLA class I molecules, but only after back transport of ER proteins to the cytosol, with concomitant enzymatic deglycosylation, resulting in the deamidation of the glycosylated asparagine [232]. Approximately 45% of the deamidated peptides were derived from membrane-associated glycoproteins with the correct site-specific glycosylation motifs (Fig. S7). Interestingly, for HLA-A1 only, we found a relatively high proportion of deamidated peptides in which the asparagine residue on position 3 was converted to an aspartic acid, the anchor residue for this class I allele (Table S3). For example, the LSNISHLNY sequence of glycosyltransferase BGnT-2 was only detected in the deamidated form LSDISHLNY.

We also detected a total of 59 serine or threonine phosphorylated peptides. Phosphorylated peptides were assigned to both HLA-A and B alleles at similar (normalized) frequencies (Table S3). A preferred bias was found for a basic residue on P1 and the phosphate group on P4 of peptide sequence (Fig. S7), characteristics similar to those previously reported [233, 234]. Mapping the data to phosphorylation-specific motifs that are commonly associated with specific kinases resulted in the assignment of 72% of the phosphopeptides to a proline-directed kinase motif (PxTP, PxSP), similar as reported by Cobbold *et al.* [233], or basophilic kinase motif (RxxS) (Fig. S7). Comparison with previous studies revealed that 16 HLA-B7-associated phosphopeptides had been reported previously [224, 233, 235], while a total of 24 phosphorylation sites are annotated in the Uniprot protein database.

Cysteinylation is a spontaneous reaction where a cysteine forms a disulfide bond to a free cysteine molecule, either *in vivo* or *in vitro* [236, 237]. In assigning the 196 cysteinylated peptides to the HLA-A and B alleles we unexpectedly found a bias towards the HLA-A1. Compared to unmodified peptides, 6% of the HLA-A1-assocatiated peptides were cysteinylated, while the frequency of this modification was only 1-2% for the remaining alleles (Table S3).

TABLE 2: Overview of HLA-associated peptides harboring post-translational modifications. [a]Annotated N-glycosylation sites.

Post-translational modifications	Modified Peptides	Unmodified counterpart	Uniprot annotated	Preferred Position
Asparagine deamination	83	47	26[a]	P3-P6 (76%)
Serine phosphorylation	44	24	22	P4-P7 (78%)
Threonine phosphorylation	15	6	2	P4 (67%)
Methionine oxidation	988	725		
Cysteinylation	196	12		
N-Glutamine cyclization (pyroQ)	246	166		

DISCUSSION

Identification of HLA class I associated peptides by mass spectrometry has contributed substantially to the knowledge of the antigen processing and presentation mechanisms involved in the cellular immune response [91]. However, the identity of a considerable part of the repertoire of peptides presented by HLA class I molecules remains largely unknown due to current limitations in MS sequencing. The immediate interest of this study was therefore to uncover peptides that escaped identification by currently employed sequencing methods (CID, HCD, ETD). Therefore, we applied EThcD fragmentation for the identification of HLA class I associated peptides, a dual fragmentation method that combines ETD-derived and HCD-derived fragment ions in a single spectrum for enhanced peptide fragmentation and sequence identification [218].

The most important finding of this comprehensive evaluation study was the excellent performance of EThcD for the global analysis of the HLA peptide repertoire, revealing a larger HLA class I peptidome for a single cell line than previously reported. Strikingly, almost 40% of the MS/MS spectra generated by EThcD led to highly-confident assignments of peptide sequences, which is, to our knowledge, the most advanced peptide identification success rate for endogenous peptides to date. Evaluation of the mass spectrometric data revealed that EThcD generates more extensive backbone fragmentation and increased coverage of the peptide sequence, resulting in more readily identified HLA class I associated peptides (Fig. 1). In comparison, Hassan *et al.* [224] reported recently a large inventory of HLA class I peptides using CID-based MS, but implemented less strict confidence criteria (FDR <10% instead of <1% here). The advantage of EThcD in generating more informative-rich MS/MS spectra for high-confident peptide assignments was supported by *in silico* spectral merging of separately acquired CID and HCD spectra. Contrary to endogenous peptides, EThcD performs equally well as HCD for tryptic peptides generated by *in vitro* digestion of proteomes [218]. This discrepancy is primarily attributed to the fact that endogenous peptides are more difficult to sequence by HCD [96], while current MS technology and database search algorithms work particularly well for tryptic peptides and do not *per se* require improvements in spectral quality just for their identification.

The selection of a B lymphoblastoid cell-line that was heterozygous for HLA-A, -B and -C allowed us to evaluate allele-specific performance of the different types of fragmentation techniques. A remarkable finding was that EThcD provided an unbiased insight into the repertoire of presented peptides. In contrast, biases in allele-specific peptide identification results could be directly translated to the known limitations of conventional peptide fragmentation techniques. This suggests that the outcome of HLA class I peptide identification studies that are based on CID (HCD) [92, 224, 238, 239] or ETD [240, 241] alone may only partially reflect the actual landscape of peptides presented at the cell surface. EThcD as universal fragmentation method is therefore an attractive alternative to current MS sequencing technology.

The high discovery rate in this study of peptides that constitute the HLA class I repertoire allowed us to confirm typical known features, but also to add new findings. On a global scale, we assigned HLA class I peptides to approximately 5,600 source proteins, which is in agreement with the findings of Hassan *et al.* [224]. Considering that cultured human cells express approximately 11,000 protein-coding genes [242], mapping the HLA class I peptidome is now becoming possible for 50%

of the expressed proteins. Moreover, approximately 40% of the source proteins are represented by at least two peptides bound to different HLA molecules. This broad representation of the cell's proteome by HLA class I is in agreement with the general belief that equitable sampling is necessary to comprehensively reflect the health status of the cell [243]. Yet, although the major part of the proteome seemed suited for HLA sampling, the actual small number of peptides identified per source protein confirms that considerable selection of peptides takes place, related to proteolytic cleavage, efficiency of transporter associated with antigen processing (TAP) and the binding affinity of the available HLA molecule [224, 229]. Earlier studies reporting fewer numbers of peptides (hundreds) demonstrated that most proteins were represented by only a single peptide [226, 228]. Hence, fragmentation techniques that can capture the full landscape of HLA class I peptides in cells of interest, such as EThcD, can make an important step into a next stage of discovery.

Our study furthermore revealed various allele-specific or peptide-specific features that can lead to a better understanding of the peptide processing and peptide specificity to various HLA alleles. We identified a large set of nested peptides that differ in length either at the N- or C-terminus. The presence of N-terminal-extended length variants have been reported previously [225, 239]. The activity of aminopeptidases in the ER effectively edits the majority of endogenously processed peptides to a length suitable for loading onto the HLA class I molecules, but a minority of peptides escape this efficient trimming process [244, 245]. Only a few studies report C-terminally extended peptide pairs. Ben Dror *et al.* [225] attributed these C-terminal-extended peptides to the broad cleavage specificity of the proteasome, while Lorente *et al.* [246] suggested an additional role of C-terminal trimming by carboxypeptidase. The global picture emerging from this study revealed an important role of allele-specific physico-chemical binding motifs. The here observed differential C-terminal length variation for the various HLA alleles were directly attributed to the different constraints for C-terminal binding between the HLA alleles, in particular because we could assume equal enzymatic activities since the peptides were extracted from a single cell line.

In addition, detailed findings included marked differences in the number of post-translationally modified peptides presented by different HLA class I molecules. The presentation of certain HLA-A1 associated peptides is enhanced by the deamidation of asparagine into aspartic acid. This modification creates a negatively charged side chain with the ability to form hydrogen bonds and hence servers as a high-affinity anchor residue for HLA-A1. Such mechanism has been previously reported for glutamine containing peptides in HLA class II [247]. The cysteinylation of cysteine residues is a non-enzymatic reaction which has been shown to play a role in T-cell recognition [236, 237]. However, the biological relevance of this modification is rather unclear because this reaction can proceed *in vivo* or *in vitro*, and can both positively and negatively affect HLA binding [248]. We found an altered (4-fold higher) frequency of cysteinylated peptides bound to HLA-A1 compared to the other alleles investigated in this study. This suggest that cysteinylation of peptides might occur prior to binding to the HLA molecule in the ER, and thus modulate the ability of peptides to bind to certain HLA alleles.

Finally, the presentation of HLA class I phosphopeptides and their recognition by T-cells is increasingly been reported to play an important role in human diseases as the deregulation of signaling pathways is a hallmark of malignant transformation [233, 234]. Although high-confident assignment of HLA class I phosphopeptides is challenging [224], we combined the advantage of EThcD for unambiguous phosphosite localization [219] with bioinformatics validation. The fraction of phosphorylated peptides identified in this study is approximately 0.5% of all peptides, irrespectively of the alleles investigated. This suggests that there is no significant bias towards a certain HLA class I molecule for presenting phosphorylated peptides. The preference of a basic residue on P1 and the phosphate group on P4 of the sequence confirms previously reported roles of these biases in the preferential binding of phosphopeptides to HLA class I molecules [233, 234, 249]. The ability to map underlying kinase motifs further reflects the accuracy in phophopeptide assignments [233].

In summary, we present a large-scale analysis of HLA class I associated peptides by EThcD fragmentation. The advantage of EThcD as universal fragmentation technique was demonstrated by the improved number of identified peptides and the increased coverage of fragment ions over conventional fragmentation techniques. The data generated here further expand our knowledge of peptide presentation by HLA class I molecules. We therefore foresee a predominant role of EThcD for comprehensive analysis of endogenously processed peptides, also beyond HLA class I peptides repertoire.

ACKNOWLEDGEMENTS

G.P.M.M, C.F, A.P.J.M.deJ. and A.J.R.H. acknowledge support from the Netherlands Proteomics Centre. C.F and A.J.R.H. are partly financed by the European Community's Seventh Framework Programme (FP7/2007-2013) by the PRIME-XS project grant agreement number 262067.

ASSOCIATED CONTENT

Additional information as noted in the text.

TABLE S1. Summary of data obtained in the LC-MS/MS analysis of trypsin digested *E.coli* proteome (5ng each analysis). IT = ion trap read out of the fragment ions.

E. coli data set	EThcD	CID-IT	HCD	ETD-IT
MS/MS spectra	5,326	20,182	9.127	12,988
PSM (<1% FDR)	1,064	4,723	3,125	3,443
Identification rate	30	23	34	27
Unique peptides	1,447	3,737	2,696	2,795
Proteins	387	864	606	692
Analysis time	3 h	3 h	3 h	3 h

TABLE S2. Comparison of our data to previously reported HLA class I-associated peptides.
[a] Granados *et al.*, Blood, 2012; 119(26): 181-91
[b] Hassan *et al.*, Mol Cell Proteomics, 2013; 12(7):1829-43
[c] Ben Dror *et al.*, Arthritis Rheum, 2010; 62(2):420-9

	This study, Combined data	Reference, data			
	Unique peptides	Unique peptides	Strategy	ETD-IT	Overlap
HLA-A1	1170	304[a]	2D SCX-RP	12,988	235
HLA-A3	2132	136[a]	2D SCX-RP	3,443	52
HLA-B7	4314	3272[b]	2D SCX-RP, IEF-RP and RP-RP	27	1544
HLA-B27	3530	1268[c]	1D RP	2,795	528

TABLE S3. Overview of the number and fraction of post-translationally modified peptides. Altered frequencies are highlighted. [a]For HLA-A1, 9 out of 23 of deamidated peptides have a deamidated aspargine residue on position 3, the anchor residue for this class I allele.

	HLA-A1		HLA-A3		HLA-B7		HLA-B27		N.A.	
HLA data set	Peptides	Fraction	Peptides	Fraction	Peptides	Fraction	Peptides	Fraction	Peptides	Fraction
Native peptides	1170	100 %	2132	100 %	4314	100 %	3530	100 %	1234	100 %
Asp deamination	23 (9)[a]	2 %	9	0.4 %	34	0.8 %	10	0.3 %	9	0.7 %
Phosphorylation (S/T)	4	0.3 %	8	0.4 %	27	0.6 %	16	0.5 %	7	0.6 %
Met oxidation	118	10 %	169	8 %	306	7 %	345	10 %	75	6 %
Cysteinylation	64	6 %	17	0.8 %	63	2 %	38	1 %	14	1 %
N-cyclization	15	1 %	38	2 %	73	2 %	43	1 %	29	2 %

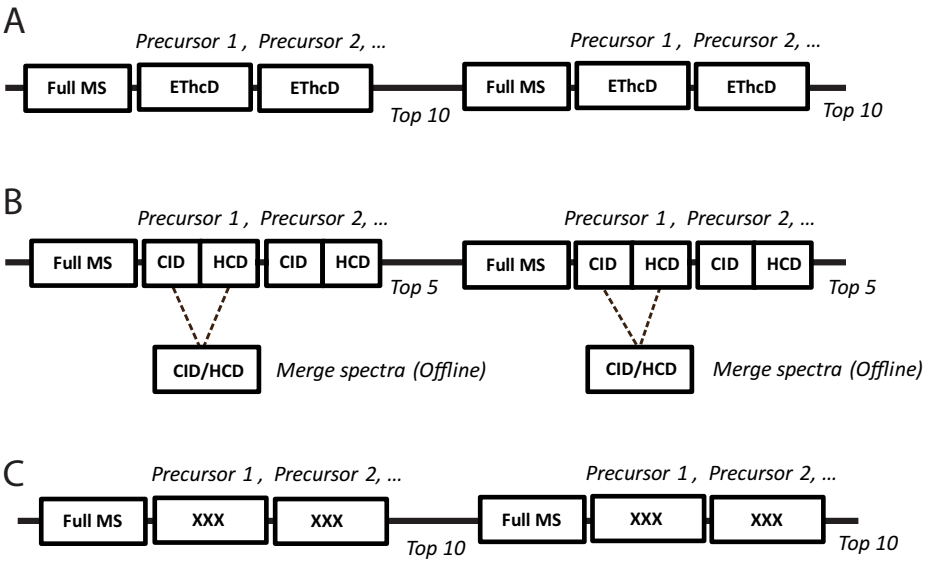


FIGURE. S1: FRAGMENTATION METHODS FOR THE IDENTIFICATION OF HLA CLASS I ASSOCIATED PEPTIDES. The workflow for the data-dependent peptide sequencing using an ETD-enabled LTQ-Orbitrap Elite mass spectrometer. As indicated, the top 10 or top 5 most intense precursor ions from each full scan were selected for (A) EThcD, (B) sequential CID/HCD, and (C) CID, HCD or ETD fragmentation. In sequential CID/HCD, each precursor is selected for CID and HCD fragmentation and obtained spectra are merged post-acquisition into a single spectrum before database search analysis. All spectra and fragmentation spectra were acquired in the Orbitrap analyzer.

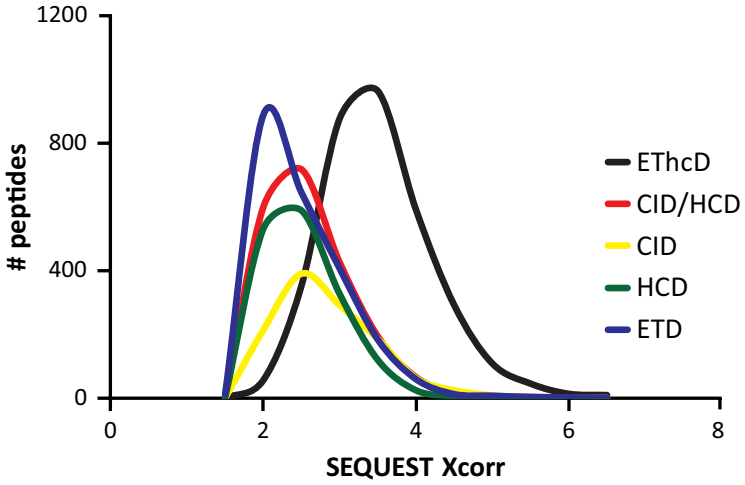


FIGURE. S2: COMPARISON OF SPECTRUM QUALITY SCORES. Frequency distribution of the SEQUEST Xcorr scores from the unique peptides identified by EThcD (black), sequential CID/HCD (red), CID (yellow), HCD (green) and ETD (blue). Fragmentation spectra were acquired in five consecutive LC-MS/MS runs (1D strategy) and searched against the human proteome database by SEQUEST. Peptide identification results were filtered to a <1% false discovery rate (FDR) and only peptides with a Xcorr > 1.75 and a canonical length of 8-14 amino acids were included.

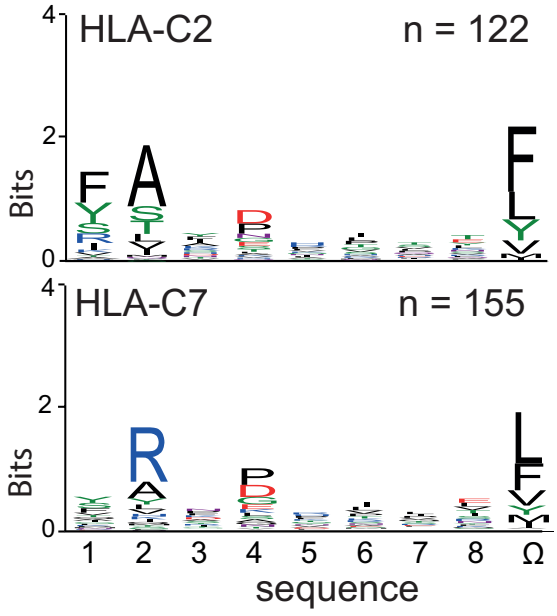


FIGURE. S3. SEQUENCE LOGO'S OF THE PEPTIDES BOUND TO HLA-C02 AND C07. The logo's display the peptide binding motifs through anchor residues at positions P2, P4 and the C-terminal PΩ. The number of peptides assigned to the HLA-C alleles is depicted in the upper right corner,

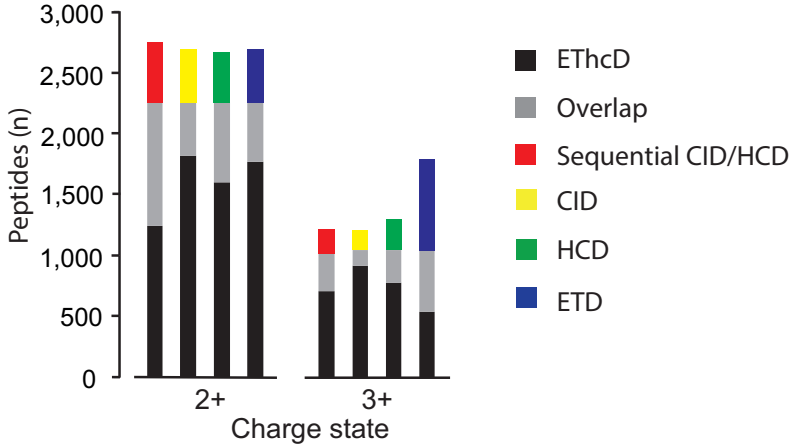


FIGURE. S4. PERFORMANCE CHARACTERISTICS OF PEPTIDE FRAGMENTATION TECHNIQUES FOR DOUBLY AND TRIPLY CHARGED HLA-ASSOCIATED PEPTIDES. Number of peptides identified from each precursor charge state (2+, 3+) by EThcD (black), overlap between both methods (grey) and either sequential CID/HCD (red), CID (yellow), HCD (green) or ETD (blue).

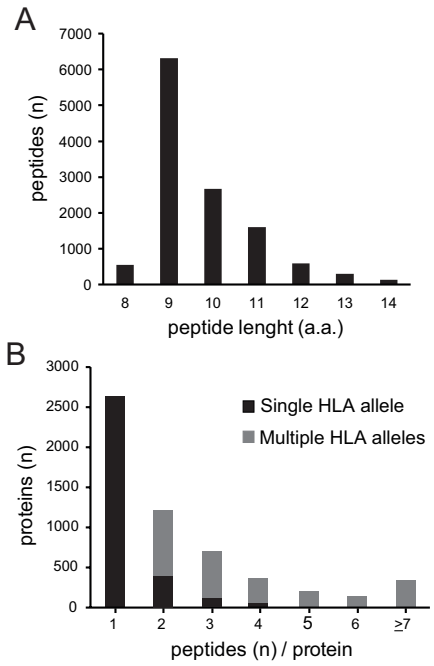


FIGURE. S5. OVERALL CHARACTERISTICS OF THE IDENTIFIED HLA PEPTIDE REPERTOIRE. (A) Distribution of peptide lengths for the complete data set (n=12,199). (B) Frequency distribution of the number of peptides identified per protein. The peptides that are derived from the source proteins can be presented by a single HLA allele (black) or multiple different HLA alleles (grey).

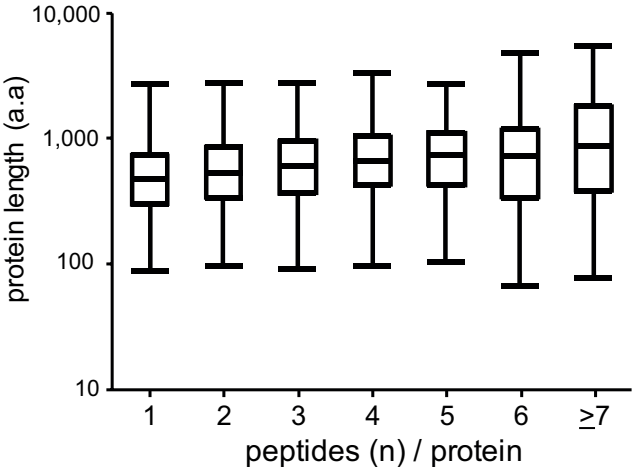


FIG.URE S6. PROTEIN LENGTH DISTRIBUTION (LOG-SCALE) AS FUNCTION OF THE NUMBER OF PEPTIDES IDENTIFIED PER PROTEIN. The boxplots summarizes the median (middle line), the interquartile range (box), and the 1st and 99th percentile of the protein sequences lengths (in amino acids).

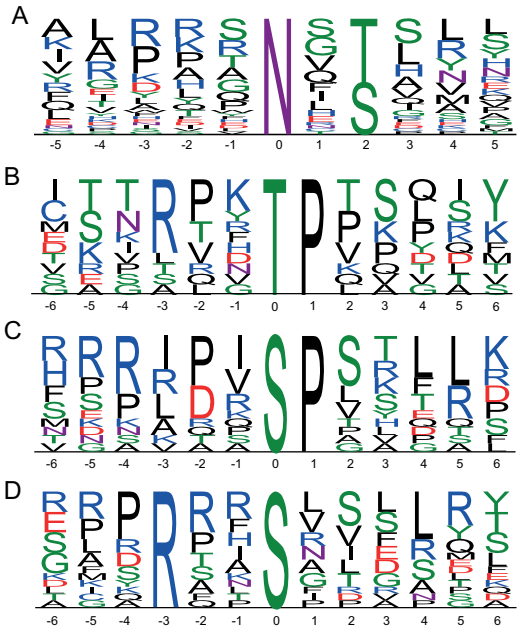


FIGURE. S7. SEQUENCE LOGO OF THE GLYCOSYLATION-SPECIFIC (A) AND PHOSPHORYLATION-SPECIFC MOTIFS (B-D). (A) Mapping of deamidated aspargine (centered) containing HLA-associated peptides to the N-glycosylation motif (38 out of 83). Deamidation indirectly suggest initial glycosylation since deglycosylated by N-glucanase results in the conversion to aspartic acid. (B) Proline-directed kinase motif centered on threonine and with a strong preference for proline C-terminal to the phosphorylation site (n=11). C) Proline-direct kinase motif centered on serine (n=15). D) Basophilic kinase motif centered on serine with a strong preference for arginine N-terminal to the phosphorylation site (n=17).

CHAPTER 06

SUMMARY

The immune system is a complex system of molecules, cells, tissues, and organs that protects higher organisms from pathogens or cancerous cells. A proper understanding of the immune system is vital for the discovery of new vaccines against infectious diseases or the development of immunotherapy against cancer. Among other molecular biology approaches, proteomics is nowadays an important analytical strategy in vaccinology and immunology. Proteomics mostly rely on Mass Spectrometry (MS) to identify and quantify proteins in cells or tissue, under normal or treated (diseased) conditions. Moreover, MS-based proteomics enables the identification of antigenic proteins or peptides (epitopes) that are able to elicit an immune response, or the profiling essential host–pathogen interactions, all aimed at finding new leads for vaccine development or to enhance the understanding of the immune system. Although MS-based proteomics has advanced rapidly in the last decades, it still faces technological challenges. The aim of this thesis was the development of dedicated MS-based proteomics methods with improved sensitivity and selectivity in detection for application in vaccinology.

CHAPTER 1 provides a general introduction to mass spectrometry (MS)-based proteomics. The most common workflow involves the analysis of peptide mixtures (obtained from proteome digestion) by liquid chromatography-tandem mass spectrometry. In tandem MS, peptide ions are isolated and fragmented into sequence-informative ions and further analyzed by database search algorithms to identify the corresponding peptide sequence. Nowadays, the proteomics toolbox encompasses a large variety of sample preparation protocols, quantitative strategies, multidimensional LC formats and MS instrumentation. The most prevalent technologies and their inherent advantages are highlighted in this chapter. In addition, to isolate subsets of peptides or proteins of interest, dedicated enrichment techniques can be applied. These include for example positional proteomics strategies that are aimed at the selective isolation of N-terminal peptides or antibody-based immunoprecipitation for targeted enrichment of Major Histocompatibility Complex (MHC) molecules.

In **CHAPTER 2**, we introduce a new positional proteomics strategy for the selective isolation of N-terminal peptides. The method employs protective labeling of primary amines at the protein level, subsequent proteolytic digestion and phosphotagging (PTAG) of all peptides except those derived from the protein N-terminus. The resulting phosphopeptides are depleted by titanium dioxide affinity chromatography, while N-terminal peptides are recovered in the flow-through for analysis by liquid chromatography-mass spectrometry. We demonstrate that the PTAG method is broadly applicable (*e.g.* prokaryotes, eukaryotes) and effectively reduces the complexity of proteomes dominated by a single protein (*i.e.* dynamic range issues). Moreover, large-scale N-proteome analysis allowed global characterization of Post Translation Modifications (PTMs) which proteins may undergo at their N-terminal ends, including initiator methionine removal, N-terminal acetylation and the substrate identification of signal and transit peptide cleavage.

CHAPTER 3 describes the development of a quantitative proteomics strategy to profile the relative protein content of *Neisseria meningitidis* Outer Membrane Vesicles (OMV)-based vaccines. Based on the findings of chapter 2, we used the PTAG method to address dynamic range issues and added a

stable-isotope dimethyl labeling strategy to quantify OMV products. We found distinct differences in protein content between OMV vaccines produced under different purification conditions. The OMV vaccines extracted from biomass under milder (detergent-free) conditions were enriched in potentially immunogenic membrane (lipo)proteins. These results were sustained by serum blot proteomics, confirming a differential pattern in antigenic protein content.

In **CHAPTER 4**, we describe the development of a two dimensional (2D) chromatographic separation technique for highly sensitive analysis of complex peptide mixtures. The method employs, in an online configuration, weak anion and strong cation exchange (ACE) mixed-bed ion exchange chromatography for first dimensional separation and reversed phase chromatography for second dimensional separations. Elution of peptides retained by ACE mixed-bed was accomplished by a novel, salt-free buffer system instead of a conventional salt gradient. This allowed us to omit the undesirable salts, resulting in an excellent compatibility between ion exchange chromatography and reversed phase chromatography-mass spectrometry. The achieved level of sensitivity and compatibility demonstrates the potential of this system for the 2-dimensional analysis of very modest amounts (µg) of sample material.

In **CHAPTER 5**, we investigated the utility of a recently introduced peptide fragmentation method, the so-called Electron-Transfer and Higher-energy collision Dissociation (EThcD) technique, for the identification of Human Leukocyte Antigens (HLA) class I-associated peptides. To date, the repertoire of HLA class I peptides remains largely unexplored, although their identity could lead to new targets for vaccine development or cancer immunotherapy. EThcD employs dual fragmentation to a single ion package to generate both the fragment ions induced by ETD (c/z) and HCD (b/y) in a single spectrum. We demonstrate that sequencing of HLA class I peptides considerably benefits from the complementary and informative-rich fragmentation spectra produced by EThcD. The importance of this next-generation peptide sequencing technology is provided, expanding the detectable peptide repertoire to an unprecedented depth (> 12,000 peptides) and revealing novel features in the antigen processing and presentation machinery. Our data includes furthermore a variety of post-translational modifications for which evidence is accumulating that they play important roles in human diseases.

NEDERLANDSE SAMENVATTING

Het immuunsysteem is een complex systeem van moleculen, cellen, weefsels en organen dat als doel heeft pathogenen of kankercellen te bestrijden. Het begrijpen van het immuunsysteem is erg belangrijk voor de ontwikkeling van nieuwe vaccins tegen infectieziekten of immunotherapieën tegen kanker. Naast een aantal andere moleculaire biologie methoden is proteomics een belangrijke analytische strategie in immunologie en vaccinologie. Proteomics maakt voornamelijk gebruik van massaspectrometrie (MS) om op zeer gevoelig niveau proteïnen in cellen of tissue te identificeren en kwantificeren. Proteomics kan tevens specifieke proteïnen of peptiden (epitopen) identificeren die een immuunrespons veroorzaken of belangrijke interacties tussen de gastheer cel en het pathogeen bepalen. Dit kan lijden tot nieuwe leads in vaccinontwikkeling of een beter begrip van het immuunsysteem. Ondanks de snelle technologische ontwikkelingen van de laatste jaren wordt proteomics nog altijd geconfronteerd met analytische uitdagingen. Het doel van dit proefschrift was het ontwikkelen van speciale proteomics methoden met verbeterde gevoeligheid en specificiteit in detectie die dienen voor toepassing in de vaccinologie.

HOOFDSTUK 1 geeft een algemene inleiding over proteomics onderzoek en geeft een overzicht van analysestrategieën waarbij men gebruik maakt van massaspectrometrie (MS) om proteïnen te karakteriseren. In de meest gebruikte workflow worden peptidenmengsels (verkregen uit digestie van proteïnen) geanalyseerd met behulp van vloeistofchromatografie gekoppeld aan tandem massaspectrometrie. In tandem MS, worden (geïoniseerde) peptide geïsoleerd en gefragmenteerd in sequentie-informatieve ionen. Deze tandem MS spectra worden vervolgens geanalyseerd met zoekalgoritmen om de overeenkomstige peptidesequentie te achterhalen. Tegenwoordig omvat de proteomics toolbox een grote verscheidenheid aan monstervoorbereidingsprotocollen, kwantificeringsstrategieën, multidimensionale LC systemen en MS instrumentatie. De meest gebruikte technologieën en hun inherente voor- en nadelen worden benadrukt in dit hoofdstuk. Om bepaalde subpopulaties van peptiden, of specifieke proteïnen van interesse te isoleren kunnen verrijkingstechnieken worden toegepast. Voorbeelden hiervan zijn positional proteomics methoden (gericht op de selectieve isolatie van N-terminale peptiden) of op antilichamen gebaseerde immunoprecipitaties (verrijking van bijvoorbeeld 'Major Histocompatibility Complex' (MHC) moleculen).

In **HOOFDSTUK 2** introduceren we een nieuwe positional proteomics strategie voor het selectief isoleren van N-terminale peptiden. Deze strategie past een chemische derivatiseringsreactie toe om de reactiviteit van primaire amines te blokkeren. Vervolgens worden de proteïnen enzymatisch gedigesteerd en alle gevormde peptiden gederivatiseerd met een phosphotagging (PTAG) reagens (met uitzondering van de geblokkeerde N-terminale peptiden). De resulterende phosphopeptiden worden weggevangen door titaniumdioxide affiniteitschromatografie, terwijl N-terminale peptiden worden verrijkt in de flow-through fractie. De flow-through fractie kan direct worden geanalyseerd met vloeistofchromatografie gekoppeld aan massaspectrometrie. We toonden aan dat de PTAG methode breed toepasbaar is (prokaryoten, eukaryoten) en zeer effectief de complexiteit vermindert van eiwitmengsels die gedomineerd worden door één enkele proteïne. Bovendien laten we zien dat met globale N-terminale proteïne analyse het mogelijk is om zogenaamde 'post-translationele modificaties' (PTMs) te karakteriseren. Voorbeelden hiervan zijn de verwijdering van

de initiator methionine, N-terminale acetylering en de substraatidentificatie van het enzymatisch knippen van signal en transit peptides.

HOOFDSTUK 3 beschrijft de ontwikkeling van een kwantitatieve proteomics strategie om de relatieve expressieniveaus van proteïnen te bepalen van vaccins tegen *Neisseria meningitidis* (meningokokken). Deze vaccins zijn gebaseerd zijn op buitenmembraan deeltjes, ook wel 'outer membrane vesicles' (OMV) genoemd. Op basis van de bevindingen van hoofdstuk 2 gebruikten we de PTAG methode om dynamisch bereik-problemen aan te pakken en voegden daar een op stabiele isotopen gebaseerde dimethyleringsreactie aan toe om de OMV producten te kwantificeren. We vonden duidelijke verschillen in de expressieniveaus van individuele proteïnen tussen OMV-vaccins die geproduceerd zijn onder verschillende condities. De OMV-vaccins verkregen uit biomassa extractie onder mildere condities (detergentia-vrij) zijn verrijkt met potentieel immunogene membraanproteïnen. Deze verschillende proteïneprofielen die we vonden met de kwantitatieve methode werden ondersteund door serum blot proteomics omdat hiermee het differentiële patroon in immunogeen proteïnegehalte werd bevestigd.

In **HOOFDSTUK 4** beschrijven we de ontwikkeling van een tweedimensionale (2D) chromatografische scheidingstechniek waarmee op zeer gevoelig niveau peptiden in complexe monsters geanalyseerd kunnen worden. De methode maakt gebruik van een zogenaamd 'weak anion/strong cation exchange' (ACE) gemengde ionenwisselingschromatografie voor scheidingen in de eerste dimensie en reversed phase chromatografie voor scheidingen in de tweede dimensie. Elutie van peptiden die gebonden zijn op het ACE mixed-bed was mogelijk door gebruik te maken van een nieuw, zoutvrij buffersysteem in plaats van een conventioneel zout-bevattend systeem. Dit resulteerde in een optimale afstemming tussen ionenwisselingschromatografie en reversed phase chromatografie gekoppeld aan massaspectrometrie. Tevens toonden we met de bereikte mate van gevoeligheid en compatibiliteit de kracht van dit systeem aan voor de 2-dimensionale analyse van zeer geringe hoeveelheden (µg) aan monstermateriaal.

In **HOOFDSTUK 5** onderzochten we het belang van een recent geïntroduceerde fragmentatietechniek, de zogenaamde 'Electron-Transfer / Higher-energy collisional Dissociation' (EThcD) techniek, voor de identificatie van 'Human Leukocyt Antigenen (HLA) klasse I geassocieerde peptiden. Tot op heden blijft het repertoire van HLA klasse I peptiden door technische beperkingen grotendeels onontgonnen, terwijl de identificatie van deze peptiden essentieel is voor de ontwikkeling van vaccins of immunotherapieën tegen kanker. EThcD maakt gebruik van een gecombineerde fragmentatietechniek waarbij de peptidefragmenten van zowel ETD (c/z) als HCD (b/y) worden gegenereerd. We laten zien dat de identificatie van HLA klasse I peptiden aanzienlijk profiteert van deze complementaire en informatieve spectra. Het belang van deze nieuwe technologie wordt aangetoond doordat we het repertoire aan peptiden verder hebben weten uit te breiden, resulterend in nieuwe leads in de antigeenpresentatie-mechanismen. Omdat post-translationele modificaties met de nieuwe fragmentatietechniek ook bijzonder goed kunnen worden gekarakteriseerd, laat onze dataset ook zien dat deze gemodificeerde peptiden een substantiële bijdrage leveren in de antigeenpresentatie.

OUTLOOK

The ultimate goal of proteomics is to understand complex biological systems by analyzing protein amount, function, localization, modifications and possible (mutual) interactions/complexes [52, 250-252]. One of the objectives of proteomics is the search for changes in the pattern of protein expression associated with the onset of diseases [250, 253]. To achieve this goal, proteomics can take advantage of the immune system of higher organisms. The immune system has the ability to identify minute amounts of abnormal and foreign proteins, which serve as alarm signals indicating the onset of diseases. These alarm signals reflect a molecular signature that is related to infection or cancer. Thus, under the large umbrella of proteomics, new and optimized methodologies emerge constantly that are aimed at the identification of disease- or host cell-related proteins, as reviewed frequently [2, 4, 9, 91, 254, 255]. This chapter highlights some exciting advances in proteomics technologies and discusses the perspective of these applications in vaccinology.

MS-based proteomics typically relies on a 'bottom-up' approach, where proteins are isolated form biological samples of interest, enzymatically cleaved into peptides and analyzed by liquid chromatography coupled to mass spectrometry (LC-MS). Currently, advanced MS methods enable the identification of the entire proteome in biological systems, even to a depth of more than 10,000 proteins for cultured human cells [242, 256]. This extended proteome coverage will further expand our knowledge of the immune system and will have its impact on vaccinology.

In recent years, there is an increasing emphasis on systems biology in immunology and vaccine research [257-259]. Systems biology is an interdisciplinary approach that uses 'omics' technologies (*e.g.* genomics, transcriptomics and proteomics) for the large-scale characterization of biological systems to elucidate biological rules and develop predictive models. Recent studies of the yellow fever virus and the influenza virus have provided early molecular signatures that can be used to predict immunogenicity and vaccine efficacy [260-262]. These studies mostly rely on measurement of mRNA employing microarray (transcriptomics), functional essays and bioinformatics analysis. From a systems biology perspective, however, MS-based proteomics delivers distinct types of information [103]. This includes for example the understanding of viral pathogenesis at the posttranscriptional level [263] and the characterization of post-translational modification mechanisms (*e.g.* phosphorylation) that regulate the expression of host-cell immune factors [264]. In a recent systems biology study, the physiological state of a single person was monitored for a 14-month period by 'omics' approaches [265]. They found distinct molecular events and pathways that where activated by viral infection, mostly by transcriptomics profiling, although some were detected with the proteomics data only. Thus, there is a great promise that proteomics can provide essential information on pathogen-host interactions and signaling pathways in response to infection or vaccination that can be used in an integrative 'omic' approach to develop the next-generation vaccines [266].

Proteomics is not only of use in the early phase of vaccine development (*i.e.* lead finding and immunology) but also during process development. Development of a reproducible production process for instance is particularly important in order to guarantee constant and high product quality. Proteomics tools have been extensively used to characterize and compare final products,

both for antiviral and antibacterial vaccines [254, 255]. The importance hereof is provided in chapter 3 of this thesis, where we demonstrated that the antigenic protein composition of *Neisseria meningitidis* outer membrane vesicle vaccines depends at least partially on the purification process. However, in order to understand how bacteria and eukaryotic cells inside a bioreactor respond to medium composition and physical or biological process parameters, monitoring of the cellular composition throughout the cultivation process can provide the desired insights for process development. Costs, expertise and throughput time have precluded the routine use of MS-based proteomics in process development [267], but recent developments offer solutions to these restraints. Target proteomics approaches enable the high-throughput characterization of predefined proteins across many conditions [268]. These methods use an internal peptide or protein standard in combination with Selected Reaction Monitoring (SRM) mass spectrometry to accurately quantify proteins in complex biological samples [269]. Because SRM assays are now easily developed and, in most cases, outperform traditional gel-based or antibody methods, it is expected that SRM technology becomes the 'golden standard' for the routine quantification of proteins in experimental biology, as recently posited by Aebersold *et al.* [270]. SRM technology is, however, less suited for the global and unbiased profiling of biological systems because only a few hundreds of predefined peptides can be quantified in a single run. Recently, the group of M. Mann developed a powerful platform that enables the rapid and in-depth profiling of proteomes by single shot analysis [203, 205]. This methodology, which already has been adopted by others [271-273], couples ultra-high performance liquid chromatography with fast sequencing mass spectrometry (UHPLC-MS) and enabled the identification of almost 4000 yeast proteins in a single run analysis. Considering the advantages, *e.g.* speed and proteome coverage, this technology may gain a dominant position in many biological applications, including in process and product development of biopharmaceuticals. Ideally, single-shot proteomics technology is combined with an appropriate labeling strategy that enables the accurate quantification and robust statistical validation of multiple biological replicates across multiple time points or different process variables. Among several other approaches [67], the Common Reference approach as described in chapter 3 enables multiplex quantification.

The comprehensive and detailed profiling of proteomes comes with a critical note because large amounts of starting material are typically required to achieve such results. However, biological processes are often proceed via complex system of cells, tissue or organs. The involved proteins are often found at substoichiometric abundances and therefore require dedicated enrichment procedures prior to MS analysis [18]. Fluorescence-Activated Cell Sorting (FACS) has become the tool of choice for the separation of subsets of immune cells from heterogeneous cell populations, such as blood and tissue samples [274]. FACS uses fluorescent-labeled antibodies to identify and sort cells by their specific surface-expressed markers. Application of MS-based proteomic techniques to FACS-sorted cells has long been hampered by the small sample quantities [275], but with recent advances in sensitivity it is now possible to identify around 5,000 proteins from a limited number of purified cells [14, 17, 276]. Luber *et al.* [14] used FACS sorting to purify dendritic cell (DC) subsets from mouse spleen, which were subsequently analyzed by MS-based proteomics, combined with label-free quantification. A key finding was that differences exist in viral recognition between DC subsets, whereby the CD8α+ DCs largely lacked the receptor required to sense certain viruses in the cytoplasm. Despite these exciting developments in combining FASC sorting with MS-based

proteomics, the limited numbers (*i.e.* several thousands) of enriched cells from *in vivo* samples still pose a significant challenge to proteomic sample preparation. Di Palma *et al.* [17] analyzed the minute amounts of sample material obtained from FACS sorting by highly sensitive 2-dimensional LC-MS, while Martin *et al.* [277] developed an integrated system, coupling FACS sorting online with proteome digestion and LC-MS analysis. Chapter 4 of this thesis describes the development of an online 2-dimensional LC-MS system for highly sensitive proteome analysis, a methodology that can find application in the analysis of the microgram amounts of proteins obtained after FASC sorting.

Proteolysis, *i.e.* the enzymatic degradation of target proteins, is an irreversible post-translation modification (PTM) that affects every protein at some point. Initially, proteases were considered primarily to be protein-degrading enzymes [278]. These protein-degrading enzymes are located in special compartments (*e.g.* lysosome, proteasome) and play an important role in protein recycling and antigen processing and presentation to Major Histocompatibility Complex (MHC) class I and II molecules as part of the immune system [7, 279]. However, proteases are also extremely important signaling molecules that are involved in numerous vital processes. For example, granzymes are a family of serine proteases that are stored in cytotoxic granules of immune cells (T cells and natural killer cells) [280]. These enzymes induce apoptosis to eliminate virus-infected cells and tumor cells. Another example are the inflammatory caspases, which are essential proteases for the initiation of apoptosis, and the processing and maturation of inflammatory cytokines [281]. Given the biological importance, there is an increasing interest in the specific functions proteases and their substrate repertoires [282]. Currently, there are only a limited number of MS-based proteomics methods that enable the identification of protease substrates. These include gel-based methods [283, 284] and positional proteomics methods, the latter enriches for the newly generated N-termini that result from protease cleavage [102, 285]. In chapter 2 of this thesis, we describe the development and application of phosphotagging (PTAG), a selective isolation strategy for protein N-terminal peptides that could be implementable for the identification of protease substrates. Target proteomics methodologies have provided a wealth of protease-specific substrate cleavage data, including the substrates of granzymes and inflammatory caspases [282, 286, 287]. However, the full functional repertoire of proteases is just being elucidated. Methodological developments could be expected in future, in particular because only a fraction of the total N-proteome is currently covered by positional proteomics [82]. Since these techniques rely on the identification of single peptide per proteins, inappropriate lengths, poor ionizability or poor fragmentation characteristics make a significant part of N-terminal peptides unsuitable for identification by MS [109]. Alternative fragmentation techniques that are less influenced by the peptide sequence, such as Electron Transfer Dissociation (ETD) [64], or a multiplex digestion approach that generate N-termini with various lengths may partially address these issues [121].

The challenges associated with the accurate identification of peptides are also found for endogenously processed and Major Histocompatibility Complex (MHC) presented peptides. Although the repertoire of MHC-presented peptides serves as a valuable source of information for the development of vaccines and cancer immunotherapy [9, 253], their identity remains largely unexplored due to several technical limitations. For example, approximately 500,000 MHC

class I molecules are expressed on almost all nucleated mammalian cells [91]. They represent a heterogeneous mixture of processed self and nonself peptides. Only a fraction (< 2%) of the associated peptides can be identified by MS-based proteomics. MHC-peptide complexes are typically purified by immunoprecipitation using MHC-specific antibodies. Since the overall work-up efficiency is about 5% for cultured cell systems (10^9-10^{10} cells) [224], efficient purification methods without major sample losses are still needed. In particular for MHC-peptide complexes presented on cells of small organelles or tumor biopsies because for these types of samples it is extremely difficult to enrich sufficient peptide material for LC-MS analysis [10, 288]. Moreover, the identification of MHC-presented peptides by tandem MS is complicated because the spectra are often of insufficient quality for high confident assignment of the peptide sequence [90]. To address this problem, we successfully introduced Electron Transfer / Higher Energy Dissociation (EThcD) for the analysis of MHC class I peptides in chapter 5 of this thesis. This method employs dual fragmentation to a single ion package to generate more informative MS/MS spectra that enable high confident peptide assignments. Database searching of MS/MS data is also not straightforward because the MHC peptides are endogenously processed by multi-step digestion. Therefore, MS/MS data needs to be analyzed without enzyme specificity, which dramatically enlarges the search space and increases the number of potential false positive identifications within the defined peptide mass tolerance. Sophisticated bioinformatics tools emerge constantly that are able to improve the reliability in MHC-peptide identifications. These methods, for example, combine *in silico* the information of multiple fragmentation techniques to generate more informative MS/MS spectra [289] or rescoring the database results using a machine learning method to improve confidence in peptide assignments [221].

In conclusion, mass spectrometry-based proteomics is an indispensable tool in biological research such as vaccinology, which enables the fairly routine profiling of complex proteomes to a large depth of information. In the near future, more applications of MS can be expected in systems biology approaches to obtain a global picture of the immune system in response to infection or vaccination. Moreover, complete coverage of all expressed proteins is now possible by 'single-shot proteomics' for moderate complex samples (*i.e.* bacteria, yeast). This makes proteomics a powerful tool for the high-throughput characterization of cultivation processes to ensure vaccine quality and product yield. However, future technological developments, such as described in this thesis, are still required in order to unravel the relevant proteome data (*i.e.* antigens and epitopes) from complex biological systems (*i.e.* immune system).

PUBLICATIONS

Mommen GPM, Frese CK, Meiring HD, van Gaans-van den Brink J, de Jong APJM, van Els CACM, Heck AJR, Expanding the detectable HLA peptide repertoire using Electron-Transfer / higher-energy Collision Dissociation (EThcD), Submitted (2013)

Mommen GPM, Meiring HD, Heck AJR, de Jong APJM, Mixed-bed ion exchange chromatography employing a salt-free pH gradient for improved sensitivity and compatibility in MudPIT. Anal Chem, 2013. 85(14): p. 6608-16.

van de Waterbeemd B‡, Mommen GPM‡, Pennings, Eppink MH, Wijffels RH, van der Pol LA, de Jong APJM , Quantitative Proteomics Reveals Distinct Differences in the Protein Content of Outer Membrane Vesicle Vaccines. J Proteome Res, 2013. 12(4): p. 1898-08

Mommen GPM‡, van de Waterbeemd B‡, Meiring HD, Kersten G, Heck AJR, de Jong APJM, Unbiased selective isolation of protein N-terminal peptides from complex proteome samples using phospho tagging (PTAG) and TiO_2-based depletion. Mol Cell Proteomics, 2012. 11(9): p. 832-42.

Mommen GPM, Hamzink MRJ, Zomer G, de Jong APJM, Meiring HD, Protein affinity tag and uses thereof, Patent Number 2012081978, 2012

‡Both authors contributed equally

CURRICULUM VITAE

Geert Mommen was born on 23 August 1981 in Venray, the Netherlands. He was educated at the Radboud University Nijmegen, The Netherlands, where he received his M.Sc. degree in Chemistry in 2007. During his master project, he specialized in analytical chemistry and chemometrics under the supervision of Prof. Dr. Lutgarde Buydens. After graduation, Geert started as senior engineer at NXP Semiconductors (Nijmegen, The Netherlands), where he was introduced to mass spectrometry for the monitoring of trace elements in a variety of industrial applications. In 2009, he decided to pursue his career in biomolecular mass spectrometry and proteomics, and started as PhD candidate at the Institute of Translational Vaccinology (Intravacc) (formerly the Netherlands Vaccine Institute) under the supervision of Dr. ir. Ad de Jong and Prof. Dr. Albert Heck (Utrecht University, The Netherlands). His research is mainly focused on methodology development in mass spectrometry-based proteomics for applications in immunology and vaccinology.

DANKWOORD

Het is een druilerige zondagmiddag in de kerstvakantie. Het boekje is drukklaar. Het enige wat er nog mist is mijn dankwoord. Het is rustig in huis. Onze dochter Sanne van bijna 2 jaar ligt te slapen, ze was doodop na een ochtend zwemmen. 'Papa swemme, grote bad?', zoals ze zelf altijd zegt. Achter het beeldscherm van mijn laptop bedenk ik me dat er in het afgelopen jaar veel te weinig van dit soort momenten zijn geweest. Vaker was het: 'papa werke?' Het resultaat hiervan is dit proefschrift. Natuurlijk ben ik zeer tevreden en trots op dit boekje, maar het wordt nu de hoogste tijd dat ze op zondag voortaan 'papa swemme' zegt in plaats van 'papa werke'.

Uiteraard ben ik niet de enige die veel tijd en energie in dit proefschrift heeft gestoken. Allereerst wil ik mijn co-promotor Ad bedanken voor zijn eindeloze inzet om mij het vak van biomoleculaire massaspectrometrie bij te brengen. Ik realiseer me dat je een belangrijke rol hebt gespeeld in mijn persoonlijke ontwikkeling tot jonge onderzoeker. Veel waardering heb ik voor je oprechtheid. Of een manuscript nu in 'slecht engels' of 'goed leesbaar' is, duidelijk ben je altijd. Daarnaast heb ik zeer veel bewondering voor je betrokkenheid, zelfs op de momenten dat ik dat niet van je kon verwachten. Dit laatste heb ik als groot voorrecht ervaren.

Terugdenkend aan de afgelopen jaren bedenk ik me dat mijn promotor Albert misschien wel de belangrijkste bijdrage heeft geleverd aan het slagen van mijn promotietraject. In eerste instantie heb je me vrijgelaten in mijn keuzes met betrekking tot het onderzoek, en op cruciale momenten wist je me te behoeden voor valkuilen en me feilloos door de eindfase van het promoveren te loodsen. Hier ben ik je zeer dankbaar voor. Jouw enthousiasme en onuitputtelijke energie voor het onderzoek is zeer bewonderingswaardig en werkt zeer motiverend.

Bas, ik wil je bedanken voor onze prettige samenwerking. Met veel plezier kijk ik terug op de vele inhoudelijke discussies, praktisch werk en publicaties, maar inderdaad ook op het vertrouwen in elkaars expertise. Tijdens je eigen promotie had je een duidelijke lijn in je onderzoek, waar je niet vanaf week en waar je met veel enthousiasme en inzet aan werkte. Hiervan heb ik veel geleerd. Hugo, je hebt een niet te onderschatten bijdrage geleverd aan dit werk. Je hebt me ontelbare keren uit de brand geholpen. Van troubleshooting tot het schrijven van manuscripten, werkelijk in elk facet van het werk kan ik op je hulp rekenen.

Cécile, voor immunologische vraagstukken kan ik altijd bij je aankloppen. Je kunt als geen ander complexe materie in voor mij begrijpelijke taal uitleggen, altijd met evenveel gedrevenheid en enthousiasme. Bedankt voor al je hulp. Christian, in der letzten Phase meiner Untersuchung haben wir zusammen sehr gute Resultate erzielt. Deine Expertise in der Massenspektrometrie hat dazu beigetragen ein Teil unseres Immunsystems besser zu verstehen. Ohne dein Wissen und deine Kreativität, hätten wir nie ein solch gutes Ergebnis erreicht, dafür meinen Dank. Gideon, ik heb altijd veel steun en vertrouwen ervaren, in het bijzonder op de momenten dat het onderzoek stroef verliep. Uiteraard ben ik je ook dankbaar voor de mogelijkheid die je me hebt geboden om mijn carrière bij Intravacc voort te zetten.

Bernard, het is bijzonder prettig om samen te werken met iemand met zo'n brede kennis en interesse. Of het nu vragen zijn over chemie, biologie of immunologie, altijd kun je me weer een stapje verder helpen. Joost, als kamergenoot heb je niet alleen de hoogtepunten, maar ook de tegenslagen van mijn promotie van dichtbij meegemaakt. Bedankt voor je luisterend oor. Je creativiteit en passie voor fotografie kwam tevens goed van pas bij de coverfoto, ook hiervoor bedankt. Cas, ik ben je zeer dankbaar voor het vormgeven van dit proefschrift. Het resultaat is prachtig. Daarnaast was het prettig om de layout uit handen geven met de wetenschap dat het goed gebeurt.

Mijn collega's wil ik bedanken, in het bijzonder de collega's van de biomoleculaire massaspectrometrie clubs van zowel Intravacc als de Universiteit Utrecht (UU). Peter H., Grazyna, Piero (UU), Sietske en Henk (UU) bedankt voor de support. Martin, bedankt voor alle synthese producten waar je aan gewerkt hebt. Dit was een erg leuke en enerverende tijd. Bert, het abrupte einde van onze samenwerking en collegialiteit kan ik nog steeds niet bevatten. Ik herinner me je als een markante onderzoeker met eindeloos veel ideeën. Jamila en Kevin, bedankt voor jullie hulp. Met veel zelfstandigheid wisten jullie de afstudeerprojecten succesvol af te ronden. Als laatste wil ik René, Peter S., Thomas (TU/e), Sietske, Gaurav, Heleen en Piero (UU) veel succes wensen bij het afronden van jullie eigen promotieonderzoek.

De herfstachtige middag loopt ten einde. Ook het schrijven van het dankwoord is nu bijna achter de rug. Cindy komt net binnen lopen met ons dochtertje Sanne, ze heeft haar slaap net uit. Goh, wat ben ik trots op die twee meiden. Het zal zeker niet altijd gemakkelijk zijn geweest, zo'n promovendus in huis die alleen maar met zijn gedachte bij de wetenschap is. Maar ook zij hielden vol. Papa en mama, bedankt dat jullie zo'n trotse (groot)ouders zijn. Trots zijn klinkt op papier zo gemakkelijk, maar in werkelijkheid is het de mooiste vorm van onvoorwaardelijke steun. Broertjes, bedankt dat jullie mij tot het laatste moment zullen helpen. Ondanks dat ik niet twijfel aan jullie deskundigheid, ben ik toch blij dat paranimfen alleen ceremoniële taken hebben. Cindy, lieve schat, je geduld is de laatste jaren meermaals op de proef gesteld, maar deze hindernis is nu eindelijk genomen. Er zal nog meer op ons afkomen, niet alleen mooie en bijzondere dingen maar ook tegenslagen. Maar één ding is zeker, samen kunnen we de toekomst aan.

ABBREVIATION

ACE	Anion exchange – Cation Exchange mixed-bed
Antigen	A substance that is recognized by the immune system
CID	Collision-Induced Dissociation
COFRADIC	Combined Fractional Diagonal Chromatography
Epitope	A part of an antigen that is recognized by the immune system
ETD	Electron-Transfer Dissociation
EThcD	Electron-Transfer / Higher-energy collisional Dissociation
FACS	Fluorescence-Activated Cell Sorting
FDR	False Discovery Rate
HCD	Higher-energy Collisional Dissociation
HLA	Human Leukocyte Antigen
LC	Liquid Chromatography
MHC	Major Histocompatibility Complex
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
PTAG	Phospho TAGging
OMV	Outer Membrane Vesicles
QqQ	Triple quadrupole
QTOF	Quadrupole Time Of Flight
RP	Reversed Phase
SILAC	Stable Isotope Labeling by Amino Acids in Culture
SCX	Strong Cation eXchange
TiO_2	Titanium Dioxide

REFERENCES

1. Jungblut, P.R., Proteome analysis of bacterial pathogens. Microbes Infect, 2001. 3(10): p. 831-40.
2. Tjalsma, H., R.M. Schaeps, and D.W. Swinkels, Immunoproteomics: From biomarker discovery to diagnostic applications. Proteomics Clin Appl, 2008. 2(2): p. 167-80.
3. James, P., Protein identification in the post-genome era: the rapid rise of proteomics. Q Rev Biophys, 1997. 30(4): p. 279-331.
4. Purcell, A.W. and J.J. Gorman, Immunoproteomics: Mass spectrometry-based methods to study the targets of the immune response. Mol Cell Proteomics, 2004. 3(3): p. 193-208.
5. Parkin, J. and B. Cohen, An overview of the immune system. Lancet, 2001. 357(9270): p. 1777-89.
6. Gordon, S., Pattern recognition receptors: doubling up for the innate immune response. Cell, 2002. 111(7): p. 927-30.
7. Neefjes, J., *et al.*, Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol, 2011. 11(12): p. 823-36.
8. Engelhard, V.H., The contributions of mass spectrometry to understanding of immune recognition by T lymphocytes. Int J Mass Spectrom, 2007. 259(1-3): p. 32-39.
9. Ovsyannikova, I.G., *et al.*, Mass spectrometry and peptide-based vaccine development. Clin Pharmacol Ther, 2007. 82(6): p. 644-52.
10. Dutoit, V., *et al.*, Exploiting the glioblastoma peptidome to discover novel tumour-associated antigens for immunotherapy. Brain, 2012. 135(Pt 4): p. 1042-54.
11. Wahl, A., *et al.*, HLA class I molecules consistently present internal influenza epitopes. Proc Natl Acad Sci U S A, 2009. 106(2): p. 540-5.
12. Hassan, C., *et al.*, The human leukocyte antigen-presented ligandome of B lymphocytes. Mol Cell Proteomics, 2013.
13. Reineke, U. and M. Schutkowski, Epitope mapping protocols. Preface. Methods Mol Biol, 2009. 524: p. v-vi.
14. Luber, C.A., *et al.*, Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. Immunity, 2010. 32(2): p. 279-89.
15. Ferreira, G.B., C. Mathieu, and L. Overbergh, Understanding dendritic cell biology and its role in immunological disorders through proteomic profiling. Proteomics Clin Appl, 2010. 4(2): p. 190-203.
16. Dennehy, R. and S. McClean, Immunoproteomics: the key to discovery of new vaccine antigens against bacterial respiratory infections. Curr Protein Pept Sci, 2012. 13(8): p. 807-15.
17. Di Palma, S., *et al.*, Highly sensitive proteome analysis of FACS-sorted adult colon stem cells. J Proteome Res, 2011. 10(8): p. 3814-9.
18. Altelaar, A.F. and A.J. Heck, Trends in ultrasensitive proteomics. Curr Opin Chem Biol, 2012. 16(1-2): p. 206-13.
19. Hunt, D.F., *et al.*, Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. Science, 1992. 255(5049): p. 1261-3.
20. Cox, B. and A. Emili, Tissue subcellular fractionation and protein extraction for use in mass-spectrometry-based proteomics. Nat Protoc, 2006. 1(4): p. 1872-8.
21. Shevchenko, A., *et al.*, In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc, 2006. 1(6): p. 2856-60.

22. Wisniewski, J.R., *et al.*, Universal sample preparation method for proteome analysis. Nat Methods, 2009. 6(5): p. 359-62.
23. Luche, S., V. Santoni, and T. Rabilloud, Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. Proteomics, 2003. 3(3): p. 249-53.
24. Olsen, J.V., S.E. Ong, and M. Mann, Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Mol Cell Proteomics, 2004. 3(6): p. 608-14.
25. Karas, M. and F. Hillenkamp, Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem, 1988. 60(20): p. 2299-301.
26. Fenn, J.B., *et al.*, Electrospray ionization for mass spectrometry of large biomolecules. Science, 1989. 246(4926): p. 64-71.
27. Di Palma, S., *et al.*, Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. J Proteomics, 2012. 75(13): p. 3791-813.
28. Washburn, M.P., D. Wolters, and J.R. Yates, 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol, 2001. 19(3): p. 242-7.
29. Xie, F., R.D. Smith, and Y. Shen, Advanced proteomic liquid chromatography. J Chromatogr A, 2012. 1261: p. 78-90.
30. MacCoss, M.J., C.C. Wu, and J.R. Yates, 3rd, Probability-based validation of protein identifications using a modified SEQUEST algorithm. Anal Chem, 2002. 74(21): p. 5593-9.
31. Perkins, D.N., *et al.*, Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis, 1999. 20(18): p. 3551-67.
32. Elias, J.E. and S.P. Gygi, Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods, 2007. 4(3): p. 207-14.
33. Yates, J.R., C.I. Ruse, and A. Nakorchevsky, Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng, 2009. 11: p. 49-79.
34. Gygi, S.P., *et al.*, Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. Proc Natl Acad Sci U S A, 2000. 97(17): p. 9390-5.
35. Gevaert, K., *et al.*, A la carte proteomics with an emphasis on gel-free techniques. Proteomics, 2007. 7(16): p. 2698-718.
36. Boersema, P.J., *et al.*, Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. J Proteome Res, 2007. 6(3): p. 937-46.
37. Mohammed, S. and A. Heck, Jr., Strong cation exchange (SCX) based analytical methods for the targeted analysis of protein post-translational modifications. Curr Opin Biotechnol, 2011. 22(1): p. 9-16.
38. Gilar, M., *et al.*, Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. J Sep Sci, 2005. 28(14): p. 1694-703.
39. Desmet, G. and S. Eeltink, Fundamentals for LC miniaturization. Anal Chem, 2013. 85(2): p. 543-56.
40. Motoyama, A. and J.R. Yates, 3rd, Multidimensional LC separations in shotgun proteomics. Anal Chem, 2008. 80(19): p. 7187-93.
41. Gilar, M., *et al.*, Orthogonality of separation in two-dimensional liquid chromatography. Anal Chem, 2005. 77(19): p. 6426-34.

42. Sandra, K., *et al.*, Highly efficient peptide separations in proteomics Part 1. Unidimensional high performance liquid chromatography. J Chromatogr B Analyt Technol Biomed Life Sci, 2008. 866(1-2): p. 48-63.
43. Sandra, K., *et al.*, Highly efficient peptide separations in proteomics. Part 2: bi- and multidimensional liquid-based separation techniques. J Chromatogr B Analyt Technol Biomed Life Sci, 2009. 877(11-12): p. 1019-39.
44. Shen, Y., *et al.*, High-efficiency on-line solid-phase extraction coupling to 15-150-microm-i.d. column liquid chromatography for proteomic analysis. Anal Chem, 2003. 75(14): p. 3596-3605.
45. Shen, Y., *et al.*, Ultra-high-efficiency strong cation exchange LC/RPLC/MS/MS for high dynamic range characterization of the human plasma proteome. Anal Chem, 2004. 76(4): p. 1134-44.
46. Meiring, H.D., *et al.*, Targeted identification of infection-related HLA class I-presented epitopes by stable isotope tagging of epitopes (SITE). Curr Protoc Immunol, 2007. Chapter 16: p. Unit 16 3.
47. Zhou, F., *et al.*, Nanoflow low pressure high peak capacity single dimension LC-MS/MS platform for high-throughput, in-depth analysis of mammalian proteomes. Anal Chem, 2012. 84(11): p. 5133-9.
48. Wilm, M., Principles of electrospray ionization. Mol Cell Proteomics, 2011.
49. Thelen, J.J. and J.A. Miernyk, The proteomic future: where mass spectrometry should be taking us. Biochem J, 2012. 444(2): p. 169-81.
50. Han, X., A. Aslanian, and J.R. Yates, 3rd, Mass spectrometry for proteomics. Curr Opin Chem Biol, 2008. 12(5): p. 483-90.
51. Aebersold, R. and M. Mann, Mass spectrometry-based proteomics. Nature, 2003. 422(6928): p. 198-207.
52. Walther, T.C. and M. Mann, Mass spectrometry-based proteomics in cell biology. J Cell Biol, 2010. 190(4): p. 491-500.
53. Olsen, J.V., *et al.*, Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol Cell Proteomics, 2005. 4(12): p. 2010-21.
54. Cox, J. and M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol, 2008. 26(12): p. 1367-72.
55. Mann, M. and N.L. Kelleher, Precision proteomics: the case for high resolution and high mass accuracy. Proc Natl Acad Sci U S A, 2008. 105(47): p. 18132-8.
56. Zubarev, R.A. and A. Makarov, Orbitrap mass spectrometry. Anal Chem, 2013. 85(11): p. 5288-96.
57. Andrews, G.L., *et al.*, Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). Anal Chem, 2011. 83(13): p. 5442-6.
58. Hu, Q., *et al.*, The Orbitrap: a new mass spectrometer. J Mass Spectrom, 2005. 40(4): p. 430-43.
59. Frese, C.K., *et al.*, Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. J Proteome Res, 2011. 10(5): p. 2377-88.
60. Shen, Y., *et al.*, Improving collision induced dissociation (CID), high energy collision dissociation (HCD), and electron transfer dissociation (ETD) fourier transform MS/MS degradome-peptidome identifications using high accuracy mass information. J Proteome Res, 2012. 11(2): p. 668-77.

61. Olsen, J.V., *et al.*, A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. Mol Cell Proteomics, 2009. 8(12): p. 2759-69.
62. Jones, A.W. and H.J. Cooper, Dissociation techniques in mass spectrometry-based proteomics. Analyst, 2011. 136(17): p. 3419-29.
63. Olsen, J.V., *et al.*, Higher-energy C-trap dissociation for peptide modification analysis. Nat Methods, 2007. 4(9): p. 709-12.
64. Syka, J.E., *et al.*, Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A, 2004. 101(26): p. 9528-33.
65. Elias, J.E., *et al.*, Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. Nat Methods, 2005. 2(9): p. 667-75.
66. Nesvizhskii, A.I., O. Vitek, and R. Aebersold, Analysis and validation of proteomic data generated by tandem mass spectrometry. Nat Methods, 2007. 4(10): p. 787-97.
67. Bantscheff, M., *et al.*, Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Anal Bioanal Chem, 2012. 404(4): p. 939-65.
68. Gevaert, K., *et al.*, Stable isotopic labeling in proteomics. Proteomics, 2008. 8(23-24): p. 4873-85.
69. Ong, S.E., *et al.*, Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics, 2002. 1(5): p. 376-86.
70. Oda, Y., *et al.*, Accurate quantitation of protein expression and site-specific phosphorylation. Proc Natl Acad Sci U S A, 1999. 96(12): p. 6591-6.
71. Gruhler, A., *et al.*, Stable isotope labeling of Arabidopsis thaliana cells and quantitative proteomics by mass spectrometry. Mol Cell Proteomics, 2005. 4(11): p. 1697-709.
72. Wu, C.C., *et al.*, Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. Anal Chem, 2004. 76(17): p. 4951-9.
73. Van Oudenhove, L. and B. Devreese, A review on recent developments in mass spectrometry instrumentation and quantitative tools advancing bacterial proteomics. Appl Microbiol Biotechnol, 2013. 97(11): p. 4749-62.
74. Meiring, H.D., *et al.*, Mass tag-assisted identification of naturally processed HLA class II-presented meningococcal peptides recognized by CD4+ T lymphocytes. J Immunol, 2005. 174(9): p. 5636-43.
75. Leitner, A. and W. Lindner, Chemical tagging strategies for mass spectrometry-based phospho-proteomics. Methods Mol Biol, 2009. 527: p. 229-43, x.
76. Boersema, P.J., *et al.*, Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. Proteomics, 2008. 8(22): p. 4624-32.
77. Boersema, P.J., *et al.*, Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protoc, 2009. 4(4): p. 484-94.
78. Raijmakers, R., *et al.*, Automated online sequential isotope labeling for protein quantitation applied to proteasome tissue-specific diversity. Mol Cell Proteomics, 2008. 7(9): p. 1755-62.
79. McDonald, L., *et al.*, Positional proteomics: selective recovery and analysis of N-terminal proteolytic peptides. Nat Methods, 2005. 2(12): p. 955-7.
80. McDonald, L. and R.J. Beynon, Positional proteomics: preparation of amino-terminal peptides as a strategy for proteome simplification and characterization. Nat Protoc, 2006. 1(4): p. 1790-8.

81. Doucet, A., *et al.*, Metadegradomics: toward *in vivo* quantitative degradomics of proteolytic post-translational modifications of the cancer proteome. Mol Cell Proteomics, 2008. 7(10): p. 1925-51.
82. Staes, A., *et al.*, Selecting protein N-terminal peptides by combined fractional diagonal chromatography. Nat Protoc, 2011. 6(8): p. 1130-41.
83. Van Damme, P., *et al.*, Complementary positional proteomics for screening substrates of endo- and exoproteases. Nat Methods, 2010. 7(7): p. 512-5.
84. Mahrus, S., *et al.*, Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. Cell, 2008. 134(5): p. 866-76.
85. Kleifeld, O., *et al.*, Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. Nat Biotechnol, 2010. 28(3): p. 281-8.
86. Schilling, O., *et al.*, Proteome-wide analysis of protein carboxy termini: C terminomics. Nat Methods, 2010. 7(7): p. 508-11.
87. Timmer, J.C., *et al.*, Profiling constitutive proteolytic events *in vivo*. Biochem J, 2007. 407(1): p. 41-8.
88. Gevaert, K., *et al.*, Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. Nat Biotechnol, 2003. 21(5): p. 566-9.
89. Kleifeld, O., *et al.*, Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. Nat Protoc, 2011. 6(10): p. 1578-611.
90. Hillen, N. and S. Stevanovic, Contribution of mass spectrometry-based proteomics to immunology. Expert Rev Proteomics, 2006. 3(6): p. 653-64.
91. Mester, G., V. Hoffmann, and S. Stevanovic, Insights into MHC class I antigen processing gained from large-scale analysis of class I ligands. Cell Mol Life Sci, 2011. 68(9): p. 1521-32.
92. Bassani-Sternberg, M., *et al.*, Soluble plasma HLA peptidome as a potential source for cancer biomarkers. Proc Natl Acad Sci U S A, 2010. 107(44): p. 18769-76.
93. Bozzacco, L., *et al.*, Mass spectrometry analysis and quantitation of peptides presented on the MHC II molecules of mouse spleen dendritic cells. J Proteome Res, 2011. 10(11): p. 5016-30.
94. Wahl, A., J. Weidanz, and W. Hildebrand, Direct class I HLA antigen discovery to distinguish virus-infected and cancerous cells. Expert Rev Proteomics, 2006. 3(6): p. 641-52.
95. Prilliman, K., *et al.*, Large-scale production of class I bound peptides: assigning a signature to HLA-B*1501. Immunogenetics, 1997. 45(6): p. 379-85.
96. Escobar, H., *et al.*, Utility of characteristic QTOF MS/MS fragmentation for MHC class I peptides. J Proteome Res, 2011. 10(5): p. 2494-507.
97. Lemmel, C., *et al.*, Differential quantitative analysis of MHC ligands by mass spectrometry using stable isotope labeling. Nat Biotechnol, 2004. 22(4): p. 450-4.
98. Meiring, H.D., *et al.*, Stable isotope tagging of epitopes: a highly selective strategy for the identification of major histocompatibility complex class I-associated peptides induced upon viral infection. Mol Cell Proteomics, 2006. 5(5): p. 902-13.
99. Meiring, H.D., Targeted Identification of Disease-related MHC-presented Epitopes. 2005.
100. Ong, S.E. and M. Mann, Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol, 2005. 1(5): p. 252-62.
101. Roepstorff, P. and J. Fohlman, Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed Mass Spectrom, 1984. 11(11): p. 601.

102. Plasman, K., P. Van Damme, and K. Gevaert, Contemporary positional proteomics strategies to study protein processing. Curr Opin Chem Biol, 2013. 17(1): p. 66-72.
103. Cox, J. and M. Mann, Quantitative, high-resolution proteomics for data-driven systems biology. Annu Rev Biochem, 2011. 80: p. 273-99.
104. Van Damme, P., T. Arnesen, and K. Gevaert, Protein alpha-N-acetylation studied by N-terminomics. FEBS J, 2011. 278(20): p. 3822-34.
105. Nakazawa, T., *et al.*, Terminal proteomics: N- and C-terminal analyses for high-fidelity identification of proteins using MS. Proteomics, 2008. 8(4): p. 673-85.
106. Huesgen, P.F. and C.M. Overall, N- and C-terminal degradomics: new approaches to reveal biological roles for plant proteases from substrate identification. Physiol Plant, 2011.
107. Kleifeld, O., *et al.*, Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. Nat Biotechnol. 28(3): p. 281-8.
108. Beaudette, P., *et al.*, Development of soluble ester-linked aldehyde polymers for proteomics. Anal Chem, 2011. 83(17): p. 6500-10.
109. Agard, N.J. and J.A. Wells, Methods for the proteomic identification of protease substrates. Curr Opin Chem Biol, 2009. 13(5-6): p. 503-9.
110. Timmer, J.C., *et al.*, Structural and kinetic determinants of protease substrates. Nat Struct Mol Biol, 2009. 16(10): p. 1101-8.
111. Zhang, X., *et al.*, A proteome-scale study on *in vivo* protein Nalpha-acetylation using an optimized method. Proteomics. 11(1): p. 81-93.
112. Staes, A., *et al.*, Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). Proteomics, 2008. 8(7): p. 1362-70.
113. Dormeyer, W., *et al.*, Targeted analysis of protein termini. J Proteome Res, 2007. 6(12): p. 4634-45.
114. Helbig, A.O., *et al.*, Profiling of N-acetylated protein termini provides in-depth insights into the N-terminal nature of the proteome. Mol Cell Proteomics, 2010. 9(5): p. 928-39.
115. van de Waterbeemd, B., *et al.*, Improved OMV vaccine against *Neisseria meningitidis* using genetically engineered strains and a detergent-free purification process. Vaccine. 28(30): p. 4810-6.
116. Lemeer, S., *et al.*, Online automated *in vivo* zebrafish phosphoproteomics: from large-scale analysis down to a single embryo. J Proteome Res, 2008. 7(4): p. 1555-64.
117. Pinkse, M.W., *et al.*, Highly robust, automated, and sensitive online TiO_2-based phosphoproteomics applied to study endogenous phosphorylation in Drosophila melanogaster. J Proteome Res, 2008. 7(2): p. 687-97.
118. Motoyama, A., *et al.*, Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. Anal Chem, 2007. 79(10): p. 3623-34.
119. Meiring, H.D., *et al.*, Nanoscale LC–MS(n): technical design and applications to peptide and protein analysis. Sep Sci, 2002. 25(9): p. 557-568.
120. Emanuelsson, O., *et al.*, Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc, 2007. 2(4): p. 953-71.
121. Vogtle, F.N., *et al.*, Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability. Cell, 2009. 139(2): p. 428-39.

122. Veenstra, T.D., T.P. Conrads, and H.J. Issaq, What to do with "one-hit wonders"? Electrophoresis, 2004. 25(9): p. 1278-9.
123. Holst, J., *et al.*, Properties and clinical performance of vaccines containing outer membrane vesicles from *Neisseria meningitidis*. Vaccine, 2009. 27 Suppl 2: p. B3-12.
124. Raijmakers, R., *et al.*, Exploring the human leukocyte phosphoproteome using a microfluidic reversed-phase-TiO_2-reversed-phase high-performance liquid chromatography phosphochip coupled to a quadrupole time-of-flight mass spectrometer. Anal Chem. 82(3): p. 824-32.
125. Boersema, P.J., S. Mohammed, and A.J. Heck, Phosphopeptide fragmentation and analysis by mass spectrometry. J Mass Spectrom, 2009. 44(6): p. 861-78.
126. Plasman, K., *et al.*, Probing the efficiency of proteolytic events by positional proteomics. Mol Cell Proteomics. 10(2): p. M110 003301.
127. Zhang, X. and P. Hojrup, Cyclization of the N-Terminal X-Asn-Gly Motif during Sample Preparation for Bottom-Up Proteomics. Anal Chem. 82(20): p. 8680-5.
128. Frottin, F., *et al.*, The proteomics of N-terminal methionine cleavage. Mol Cell Proteomics, 2006. 5(12): p. 2336-49.
129. Michalski, A., J. Cox, and M. Mann, More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. J Proteome Res. 10(4): p. 1785-93.
130. Mischerikow, N. and A.J. Heck, Targeted large-scale analysis of protein acetylation. Proteomics, 2011. 11(4): p. 571-89.
131. Mogk, A., R. Schmidt, and B. Bukau, The N-end rule pathway for regulated proteolysis: prokaryotic and eukaryotic strategies. Trends Cell Biol, 2007. 17(4): p. 165-72.
132. Arnesen, T., *et al.*, Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. Proc Natl Acad Sci U S A, 2009. 106(20): p. 8157-62.
133. Lange, P.F. and C.M. Overall, TopFIND, a knowledgebase linking protein termini with function. Nat Methods, 2011. 8(9): p. 703-4.
134. auf dem Keller, U., *et al.*, A statistics-based platform for quantitative N-terminome analysis and identification of protease cleavage products. Mol Cell Proteomics. 9(5): p. 912-27.
135. Helsens, K., *et al.*, Bioinformatics analysis of a *Saccharomyces cerevisiae* N-terminal proteome provides evidence of alternative translation initiation and post-translational N-terminal acetylation. J Proteome Res, 2011. 10(8): p. 3578-89.
136. Helbig, A.O., *et al.*, Perturbation of the yeast N-acetyltransferase NatB induces elevation of protein phosphorylation levels. BMC Genomics, 2010. 11: p. 685.
137. Miller Jenkins, L.M., *et al.*, p53 N-Terminal Phosphorylation: A Defining Layer of Complex Regulation. Carcinogenesis, 2012.
138. Vener, A.V., *et al.*, Mass spectrometric resolution of reversible protein phosphorylation in photosynthetic membranes of Arabidopsis thaliana. J Biol Chem, 2001. 276(10): p. 6959-66.
139. Girard, M.P., *et al.*, A review of vaccine research and development: meningococcal disease. Vaccine, 2006. 24(22): p. 4692-700.
140. Jodar, L., *et al.*, Development of vaccines against meningococcal disease. Lancet, 2002. 359(9316): p. 1499-508.

141. Bjune, G., *et al.*, Effect of outer membrane vesicle vaccine against group B meningococcal disease in Norway. Lancet, 1991. 338(8775): p. 1093-6.
142. Sierra, G.V., *et al.*, Vaccine against group B *Neisseria meningitidis*: protection trial and mass vaccination results in Cuba. NIPH Ann, 1991. 14(2): p. 195-207; discussion 208-10.
143. Thornton, V., *et al.*, Safety and immunogenicity of New Zealand strain meningococcal serogroup B OMV vaccine in healthy adults: beginning of epidemic control. Vaccine, 2006. 24(9): p. 1395-400.
144. Ellis, T.N. and M.J. Kuehn, Virulence and immunomodulatory roles of bacterial outer membrane vesicles. Microbiol Mol Biol Rev, 2010. 74(1): p. 81-94.
145. Deatherage, B.L., *et al.*, Biogenesis of Bacterial Membrane Vesicles. Mol Microbiol, 2009.
146. van der Ley, P., *et al.*, Topology of outer membrane porins in pathogenic Neisseria spp. Infect Immun, 1991. 59(9): p. 2963-71.
147. Martin, D.R., *et al.*, The VR2 epitope on the PorA P1.7-2,4 protein is the major target for the immune response elicited by the strain-specific group B meningococcal vaccine MeNZB. Clin Vaccine Immunol, 2006. 13(4): p. 486-91.
148. van der Ley, P., J. van der Biezen, and J.T. Poolman, Construction of *Neisseria meningitidis* strains carrying multiple chromosomal copies of the porA gene for use in the production of a multivalent outer membrane vesicle vaccine. Vaccine, 1995. 13(4): p. 401-7.
149. van den Dobbelsteen, G.P., *et al.*, Immunogenicity of a combination vaccine containing pneumococcal conjugates and meningococcal PorA OMVs. Vaccine, 2007. 25(13): p. 2491-6.
150. Zollinger, W.D., *et al.*, Design and evaluation in mice of a broadly protective meningococcal group B native outer membrane vesicle vaccine. Vaccine, 2010.
151. Urwin, R., *et al.*, Distribution of surface protein variants among hyperinvasive meningococci: implications for vaccine design. Infect Immun, 2004. 72(10): p. 5955-62.
152. Zollinger, W.D., J.T. Poolman, and M.C. Maiden, Meningococcal serogroup B vaccines: will they live up to expectations? Expert Rev Vaccines, 2011. 10(5): p. 559-61.
153. Wheeler, J.X., C. Vipond, and I.M. Feavers, Exploring the proteome of meningococcal outer membrane vesicle vaccines. Proteomics Clin Appl, 2007. 1(9): p. 1198-210.
154. Claassen, I., *et al.*, Production, characterization and control of a *Neisseria meningitidis* hexavalent class 1 outer membrane protein containing vesicle vaccine. Vaccine, 1996. 14(10): p. 1001-8.
155. Fredriksen, J.H., *et al.*, Production, characterization and control of MenB-vaccine "Folkehelsa": an outer membrane vesicle vaccine against group B meningococcal disease. NIPH Ann, 1991. 14(2): p. 67-79; discussion 79-80.
156. de Moraes, J.C., *et al.*, Protective efficacy of a serogroup B meningococcal vaccine in Sao Paulo, Brazil. Lancet, 1992. 340(8827): p. 1074-8.
157. Boslego, J., *et al.*, Efficacy, safety, and immunogenicity of a meningococcal group B (15:P1.3) outer membrane protein vaccine in Iquique, Chile. Chilean National Committee for Meningococcal Disease. Vaccine, 1995. 13(9): p. 821-9.
158. van de Waterbeemd, B., *et al.*, Improved OMV vaccine against *Neisseria meningitidis* using genetically engineered strains and a detergent-free purification process. Vaccine, 2010. 28(30): p. 4810-6.

159. van der Ley, P., *et al.*, Modification of lipid A biosynthesis in *Neisseria meningitidis* lpxL mutants: influence on lipopolysaccharide structure, toxicity, and adjuvant activity. Infect Immun, 2001. 69(10): p. 5981-90.
160. van der Ley, P. and G. van den Dobbelsteen, Next-generation outer membrane vesicle vaccines against *Neisseria meningitidis* based on nontoxic LPS mutants. Hum Vaccin, 2011. 7(8): p. 886-90.
161. van de Waterbeemd, B., *et al.*, Identification and optimization of critical process parameters for the production of NOMV vaccine against *Neisseria meningitidis*. Vaccine, 2012. 30(24): p. 3683-90.
162. Keiser, P.B., *et al.*, A phase 1 study of a meningococcal native outer membrane vesicle vaccine made from a group B strain with deleted lpxL1 and synX, over-expressed factor H binding protein, two PorAs and stabilized OpcA expression. Vaccine, 2010.
163. Pinto, V.B., *et al.*, An experimental outer membrane vesicle vaccine from *N. meningitidis* serogroup B strains that induces serum bactericidal activity to multiple serogroups. Vaccine, 2011.
164. Prachayasittikul, V., *et al.*, EDTA-induced membrane fluidization and destabilization: biophysical studies on artificial lipid membranes. Acta Biochim Biophys Sin (Shanghai), 2007. 39(11): p. 901-13.
165. Haque, H. and A.D. Russell, Effect of chelating agents on the susceptibility of some strains of gram-negative bacteria to some antibacterial agents. Antimicrob Agents Chemother, 1974. 6(2): p. 200-6.
166. Post, D.M., *et al.*, Biochemical and functional characterization of membrane blebs purified from *Neisseria meningitidis* serogroup B. J Biol Chem, 2005. 280(46): p. 38383-94.
167. Koeberling, O., A. Seubert, and D.M. Granoff, Bactericidal antibody responses elicited by a meningococcal outer membrane vesicle vaccine with overexpressed factor H-binding protein and genetically attenuated endotoxin. J Infect Dis, 2008. 198(2): p. 262-70.
168. Koeberling, O., *et al.*, Immunogenicity of a meningococcal native outer membrane vesicle vaccine with attenuated endotoxin and over-expressed factor H binding protein in infant rhesus monkeys. Vaccine, 2011. 29(29-30): p. 4728-34.
169. Ferrari, G., *et al.*, Outer membrane vesicles from group B *Neisseria meningitidis* delta gna33 mutant: proteomic and immunological comparison with detergent-derived outer membrane vesicles. Proteomics, 2006. 6(6): p. 1856-66.
170. Koeberling, O., I. Delany, and D.M. Granoff, A critical threshold of meningococcal factor h binding protein expression is required for increased breadth of protective antibodies elicited by native outer membrane vesicle vaccines. Clin Vaccine Immunol, 2011. 18(5): p. 736-42.
171. Borrow, R., P. Balmer, and E. Miller, Meningococcal surrogates of protection--serum bactericidal antibody activity. Vaccine, 2005. 23(17-18): p. 2222-7.
172. Holst, J., *et al.*, Serum bactericidal activity correlates with the vaccine efficacy of outer membrane vesicle vaccines against *Neisseria meningitidis* serogroup B disease. Vaccine, 2003. 21(7-8): p. 734-7.
173. Vipond, C., *et al.*, Proteomic analysis of a meningococcal outer membrane vesicle vaccine prepared from the group B strain NZ98/254. Proteomics, 2006. 6(11): p. 3400-13.

174. Vipond, C., *et al.*, Characterization of the protein content of a meningococcal outer membrane vesicle vaccine by polyacrylamide gel electrophoresis and mass spectrometry. Hum Vaccin, 2005. 1(2): p. 80-4.
175. Williams, J.N., *et al.*, Proteomic analysis of outer membranes and vesicles from wild-type serogroup B *Neisseria meningitidis* and a lipopolysaccharide-deficient mutant. Infect Immun, 2007. 75(3): p. 1364-72.
176. Tsolakos, N., *et al.*, Characterization of meningococcal serogroup B outer membrane vesicle vaccines from strain 44/76 after growth in different media. Vaccine, 2010. 28(18): p. 3211-8.
177. Gil, J., *et al.*, Proteomic study via a non-gel based approach of meningococcal outer membrane vesicle vaccine obtained from strain CU385: A road map for discovery new antigens. Hum Vaccin, 2009. 5(5): p. 347-56.
178. Bantscheff, M., *et al.*, Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem, 2007. 389(4): p. 1017-31.
179. Guryca, V., *et al.*, Qualitative improvement and quantitative assessment of N-terminomics. Proteomics, 2012. 12(8): p. 1207-16.
180. Wildes, D. and J.A. Wells, Sampling the N-terminal proteome of human blood. Proc Natl Acad Sci U S A, 2010. 107(10): p. 4561-6.
181. Mommen, G.P., *et al.*, Unbiased Selective Isolation of Protein N-terminal Peptides from Complex Proteome Samples Using Phospho Tagging (PTAG) and TiO_2-based Depletion. Mol Cell Proteomics, 2012. 11(9): p. 832-42.
182. Holten, E., Serotypes of *Neisseria meningitidis* isolated from patients in Norway during the first six months of 1978. J Clin Microbiol, 1979. 9(2): p. 186-8.
183. Baart, G.J., *et al.*, Scale-up for bulk production of vaccine against meningococcal disease. Vaccine, 2007. 25(34): p. 6399-408.
184. Baart, G.J., *et al.*, Modeling *Neisseria meningitidis* B metabolism at different specific growth rates. Biotechnol Bioeng, 2008. 101(5): p. 1022-35.
185. Smith, R.D., *et al.*, An accurate mass tag strategy for quantitative and high-throughput proteome measurements. Proteomics, 2002. 2(5): p. 513-23.
186. Conrads, T.P., *et al.*, Utility of accurate mass tags for proteome-wide protein identification. Anal Chem, 2000. 72(14): p. 3349-54.
187. Meiring, H.D., *et al.*, Nanoscale LC–MS(n): technical design and applications to peptide and protein analysis. Journal of Separation Science, 2002. 25(9): p. 557-68.
188. Cappadona, S., *et al.*, Current challenges in software solutions for mass spectrometry-based quantitative proteomics. Amino Acids, 2012. 43(3): p. 1087-108.
89. Yu, N.Y., *et al.*, PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics, 2010. 26(13): p. 1608-15.
190. Luque-Garcia, J.L., *et al.*, Analysis of electroblotted proteins by mass spectrometry: protein identification after Western blotting. Mol Cell Proteomics, 2008. 7(2): p. 308-14.
191. Tsai, C.M., C.E. Frasch, and L.F. Mocca, Five structural classes of major outer membrane proteins in *Neisseria meningitidis*. J Bacteriol, 1981. 146(1): p. 69-78.
192. Frasch, C.E., Vaccines for prevention of meningococcal disease. Clin Microbiol Rev, 1989. 2 Suppl: p. S134-8.

193. van Ulsen, P., *et al.*, Identification of proteins of *Neisseria meningitidis* induced under iron-limiting conditions using the isobaric tandem mass tag (TMT) labeling approach. Proteomics, 2009. 9(7): p. 1771-81.
194. Grifantini, R., *et al.*, Identification of iron-activated and -repressed Fur-dependent genes by transcriptome analysis of *Neisseria meningitidis* group B. Proc Natl Acad Sci U S A, 2003. 100(16): p. 9542-7.
195. Kulp, A. and M.J. Kuehn, Biological functions and biogenesis of secreted bacterial Outer Membrane Vesicles. Annual Review of Microbiology, 2010. 64: p. 163-84.
196. Andreev, V.P., *et al.*, Label-Free Quantitative LC-MS Proteomics of Alzheimer's Disease and Normally Aged Human Brains. J Proteome Res, 2012.
197. Jiang, H.Q., *et al.*, Broad vaccine coverage predicted for a bivalent recombinant factor H binding protein based vaccine to prevent serogroup B meningococcal disease. Vaccine, 2010. 28(37): p. 6086-93.
198. Snape, M.D., *et al.*, Immunogenicity of 2 Investigational Serogroup B Meningococcal Vaccines in the First Year of Life: A Randomized Comparative Trial. Pediatr Infect Dis J, 2010.
199. Arigita, C., *et al.*, Restored functional immunogenicity of purified meningococcal PorA by incorporation into liposomes. Vaccine, 2003. 21(9-10): p. 950-60.
200. Steeghs, L., *et al.*, Teasing apart structural determinants of 'toxicity' and 'adjuvanticity': implications for meningococcal vaccine development. J Endotoxin Res, 2004. 10(2): p. 113-9.
201. Manadas, B., *et al.*, Peptide fractionation in proteomics approaches. Expert Rev Proteomics, 2010. 7(5): p. 655-63.
202. Di Palma, S., *et al.*, Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. J Proteomics, 2012.
203. Thakur, S.S., *et al.*, Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. Mol Cell Proteomics, 2011. 10(8): p. M110 003699.
204. Cristobal, A., *et al.*, In-house construction of a UHPLC system enabling the identification of over 4000 protein groups in a single analysis. Analyst, 2012.
205. Nagaraj, N., *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. Mol Cell Proteomics, 2012. 11(3): p. M111 013722.
206. 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(23): p. 4311-9.
207. Winnik, W.M., Continuous pH/salt gradient and peptide score for strong cation exchange chromatography in 2D-nano-LC/MS/MS peptide identification for proteomics. Anal Chem, 2005. 77(15): p. 4991-8.
208. Song, C., *et al.*, Reversed-phase-reversed-phase liquid chromatography approach with high orthogonality for multidimensional separation of phosphopeptides. Anal Chem, 2010. 82(1): p. 53-6.
209. Mommen, G.P., *et al.*, Unbiased selective isolation of Protein N-terminal peptides from Complex Proteome Samples using Phospho Tagging (PTAG) and TiO_2-based depletion. Mol Cell Proteomics, 2012.

210. Michalski, A., J. Cox, and M. Mann, More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. J Proteome Res, 2011. 10(4): p. 1785-93.
211. Motoyama, A., *et al.*, Automated ultra-high-pressure multidimensional protein identification technology (UHP-MudPIT) for improved peptide identification of proteomic samples. Anal Chem, 2006. 78(14): p. 5109-18.
212. Wolters, D.A., M.P. Washburn, and J.R. Yates, 3rd, An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem, 2001. 73(23): p. 5683-90.
213. Boersema, P.J., S. Mohammed, and A.J. Heck, Hydrophilic interaction liquid chromatography (HILIC) in proteomics. Anal Bioanal Chem, 2008. 391(1): p. 151-9.
214. Giddings, J.C., Concepts and comparisons in multidimensional separation. J. High Resol. Chromatogr., 1987 10: p. 10: 319–323.
215. Rammensee, H.G., T. Friede, and S. Stevanoviic, MHC ligands and peptide motifs: first listing. Immunogenetics, 1995. 41(4): p. 178-228.
216. Kessler, J.H. and C.J. Melief, Identification of T-cell epitopes for cancer immunotherapy. Leukemia, 2007. 21(9): p. 1859-74.
217. 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(7): p. 2193-8.
218. Frese, C.K., *et al.*, Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. Anal Chem, 2012. 84(22): p. 9668-73.
219. Frese, C.K., *et al.*, Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). J Proteome Res, 2013. 12(3): p. 1520-5.
220. van Els, C.A., *et al.*, A single naturally processed measles virus peptide fully dominates the HLA-A*0201-associated peptide display and is mutated at its anchor position in persistent viral strains. Eur J Immunol, 2000. 30(4): p. 1172-81.
221. Kall, L., *et al.*, Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods, 2007. 4(11): p. 923-5.
222. Lundegaard, C., *et al.*, NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. Nucleic Acids Res, 2008. 36(Web Server issue): p. W509-12.
223. Zhang, H., C. Lundegaard, and M. Nielsen, Pan-specific MHC class I predictors: a benchmark of HLA class I pan-specific prediction methods. Bioinformatics, 2009. 25(1): p. 83-9.
224. Hassan, C., *et al.*, The Human Leukocyte Antigen-presented Ligandome of B Lymphocytes. Mol Cell Proteomics, 2013. 12(7): p. 1829-43.
225. Ben Dror, L., *et al.*, The HLA-B*2705 peptidome. Arthritis Rheum, 2010. 62(2): p. 420-9.
226. Granados, D.P., *et al.*, MHC I-associated peptides preferentially derive from transcripts bearing miRNA response elements. Blood, 2012. 119(26): p. e181-91.
227. Altelaar, A.F., *et al.*, Database independent proteomics analysis of the ostrich and human proteome. Proc Natl Acad Sci U S A, 2012. 109(2): p. 407-12.
228. Hoof, I., *et al.*, Proteome sampling by the HLA class I antigen processing pathway. PLoS Comput Biol, 2012. 8(5): p. e1002517.

229. Yewdell, J.W., Confronting complexity: real-world immunodominance in antiviral CD8+ T cell responses. Immunity, 2006. 25(4): p. 533-43.
230. Burrows, S.R., J. Rossjohn, and J. McCluskey, Have we cut ourselves too short in mapping CTL epitopes? Trends Immunol, 2006. 27(1): p. 11-6.
231. Petersen, J., A.W. Purcell, and J. Rossjohn, Post-translationally modified T cell epitopes: immune recognition and immunotherapy. J Mol Med (Berl), 2009. 87(11): p. 1045-51.
232. Skipper, J.C., *et al.*, An HLA-A2-restricted tyrosinase antigen on melanoma cells results from posttranslational modification and suggests a novel pathway for processing of membrane proteins. J Exp Med, 1996. 183(2): p. 527-34.
233. Cobbold, M., *et al.*, MHC Class I-Associated Phosphopeptides Are the Targets of Memory-like Immunity in Leukemia. Sci Transl Med, 2013. 5(203): p. 203ra125.
234. Zarling, A.L., *et al.*, Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. Proc Natl Acad Sci U S A, 2006. 103(40): p. 14889-94.
235. Meyer, V.S., *et al.*, Identification of natural MHC class II presented phosphopeptides and tumor-derived MHC class I phospholigands. J Proteome Res, 2009. 8(7): p. 3666-74.
236. Meadows, L., *et al.*, The HLA-A*0201-restricted H-Y antigen contains a posttranslationally modified cysteine that significantly affects T cell recognition. Immunity, 1997. 6(3): p. 273-81.
237. Pierce, R.A., *et al.*, Cutting edge: the HLA-A*0101-restricted HY minor histocompatibility antigen originates from DFFRY and contains a cysteinylated cysteine residue as identified by a novel mass spectrometric technique. J Immunol, 1999. 163(12): p. 6360-4.
238. Fortier, M.H., *et al.*, The MHC class I peptide repertoire is molded by the transcriptome. J Exp Med, 2008. 205(3): p. 595-610.
239. Escobar, H., *et al.*, Large scale mass spectrometric profiling of peptides eluted from HLA molecules reveals N-terminal-extended peptide motifs. J Immunol, 2008. 181(7): p. 4874-82.
240. Ostrov, D.A., *et al.*, Drug hypersensitivity caused by alteration of the MHC-presented self-peptide repertoire. Proc Natl Acad Sci U S A, 2012. 109(25): p. 9959-64.
241. Depontieu, F.R., *et al.*, Identification of tumor-associated, MHC class II-restricted phosphopeptides as targets for immunotherapy. Proc Natl Acad Sci U S A, 2009. 106(29): p. 12073-8.
242. Nagaraj, N., *et al.*, Deep proteome and transcriptome mapping of a human cancer cell line. Mol Syst Biol, 2011. 7: p. 548.
243. Hickman, H.D., *et al.*, Toward a definition of self: proteomic evaluation of the class I peptide repertoire. J Immunol, 2004. 172(5): p. 2944-52.
244. Hearn, A., I.A. York, and K.L. Rock, The specificity of trimming of MHC class I-presented peptides in the endoplasmic reticulum. J Immunol, 2009. 183(9): p. 5526-36.
245. Schatz, M.M., *et al.*, Characterizing the N-terminal processing motif of MHC class I ligands. J Immunol, 2008. 180(5): p. 3210-7.
246. Lorente, E., *et al.*, Diversity of natural self-derived ligands presented by different HLA class I molecules in transporter antigen processing-deficient cells. PLoS One, 2013. 8(3): p. e59118.
247. Kim, C.Y., *et al.*, Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. Proc Natl Acad Sci U S A, 2004. 101(12): p. 4175-9.

248. Kittlesen, D.J., *et al.*, Human melanoma patients recognize an HLA-A1-restricted CTL epitope from tyrosinase containing two cysteine residues: implications for tumor vaccine development. J Immunol, 1998. 160(5): p. 2099-106.
249. Mohammed, F., *et al.*, Phosphorylation-dependent interaction between antigenic peptides and MHC class I: a molecular basis for the presentation of transformed self. Nat Immunol, 2008. 9(11): p. 1236-43.
250. Altelaar, A.F., J. Munoz, and A.J. Heck, Next-generation proteomics: towards an integrative view of proteome dynamics. Nat Rev Genet, 2013. 14(1): p. 35-48.
251. Angel, T.E., *et al.*, Mass spectrometry-based proteomics: existing capabilities and future directions. Chem Soc Rev, 2012. 41(10): p. 3912-28.
252. Bensimon, A., A.J. Heck, and R. Aebersold, Mass spectrometry-based proteomics and network biology. Annu Rev Biochem, 2012. 81: p. 379-405.
253. Admon, A., E. Barnea, and T. Ziv, Tumor antigens and proteomics from the point of view of the major histocompatibility complex peptides. Mol Cell Proteomics, 2003. 2(6): p. 388-98.
254. Adamczyk-Poplawska, M., S. Markowicz, and E.K. Jagusztyn-Krynicka, Proteomics for development of vaccine. J Proteomics, 2011. 74(12): p. 2596-616.
255. Jagusztyn-Krynicka, E.K., *et al.*, Proteomic technology in the design of new effective antibacterial vaccines. Expert Rev Proteomics, 2009. 6(3): p. 315-30.
256. Beck, M., *et al.*, The quantitative proteome of a human cell line. Mol Syst Biol, 2011. 7: p. 549.
257. Oberg, A.L., *et al.*, Systems biology approaches to new vaccine development. Curr Opin Immunol, 2011. 23(3): p. 436-43.
258. Buonaguro, L., *et al.*, Systems biology applied to vaccine and immunotherapy development. BMC Syst Biol, 2011. 5: p. 146.
259. Pulendran, B., S. Li, and H.I. Nakaya, Systems vaccinology. Immunity, 2010. 33(4): p. 516-29.
260. Nakaya, H.I., *et al.*, Systems biology of vaccination for seasonal influenza in humans. Nat Immunol, 2011. 12(8): p. 786-95.
261. Querec, T.D., *et al.*, Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. Nat Immunol, 2009. 10(1): p. 116-25.
262. Brandes, M., *et al.*, A systems analysis identifies a feedforward inflammatory circuit leading to lethal influenza infection. Cell, 2013. 154(1): p. 197-212.
263. Pastorino, B., *et al.*, Identification of cellular proteome modifications in response to West Nile virus infection. Mol Cell Proteomics, 2009. 8(7): p. 1623-37.
264. Nakayasu, E.S., *et al.*, Comparative phosphoproteomics reveals components of host cell invasion and post-transcriptional regulation during Francisella infection. Mol Cell Proteomics, 2013.
265. Chen, R., *et al.*, Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell, 2012. 148(6): p. 1293-307.
266. Koff, W.C., *et al.*, Accelerating next-generation vaccine development for global disease prevention. Science, 2013. 340(6136): p. 1232910.
267. Gupta, P. and K.H. Lee, Genomics and proteomics in process development: opportunities and challenges. Trends Biotechnol, 2007. 25(7): p. 324-30.
268. Williams, T.L., *et al.*, Quantification of influenza virus hemagglutinins in complex mixtures using isotope dilution tandem mass spectrometry. Vaccine, 2008. 26(20): p. 2510-20.

269. Lange, V., *et al.*, Selected reaction monitoring for quantitative proteomics: a tutorial. Mol Syst Biol, 2008. 4: p. 222.
270. Aebersold, R., A.L. Burlingame, and R.A. Bradshaw, Western blots versus selected reaction monitoring assays: time to turn the tables? Mol Cell Proteomics, 2013. 12(9): p. 2381-2.
271. Cristobal, A., *et al.*, In-house construction of a UHPLC system enabling the identification of over 4000 protein groups in a single analysis. Analyst, 2012. 137(15): p. 3541-8.
272. Pirmoradian, M., *et al.*, Rapid and deep human proteome analysis by single-dimension shotgun proteomics. Mol Cell Proteomics, 2013.
273. Kocher, T., *et al.*, Analysis of protein mixtures from whole-cell extracts by single-run nanoLC-MS/MS using ultralong gradients. Nat Protoc, 2012. 7(5): p. 882-90.
274. Maecker, H.T., J.P. McCoy, and R. Nussenblatt, Standardizing immunophenotyping for the Human Immunology Project. Nat Rev Immunol, 2012. 12(3): p. 191-200.
275. Klimmeck, D., *et al.*, Proteomic cornerstones of hematopoietic stem cell differentiation: distinct signatures of multipotent progenitors and myeloid committed cells. Mol Cell Proteomics, 2012. 11(8): p. 286-302.
276. Da Silva, N., *et al.*, Proteomic analysis of V-ATPase-rich cells harvested from the kidney and epididymis by fluorescence-activated cell sorting. Am J Physiol Cell Physiol, 2010. 298(6): p. C1326-42.
277. Martin, J.G., T. Rejtar, and M. Stephen, Integrated microscale analysis system for targeted LC/MS proteomics on limited amounts of enriched cell populations. Anal Chem, 2013.
278. Turk, B., Targeting proteases: successes, failures and future prospects. Nat Rev Drug Discov, 2006. 5(9): p. 785-99.
279. Blum, J.S., P.A. Wearsch, and P. Cresswell, Pathways of antigen processing. Annu Rev Immunol, 2013. 31: p. 443-73.
280. Chowdhury, D. and J. Lieberman, Death by a thousand cuts: granzyme pathways of programmed cell death. Annu Rev Immunol, 2008. 26: p. 389-420.
281. Scott, A.M. and M. Saleh, The inflammatory caspases: guardians against infections and sepsis. Cell Death Differ, 2007. 14(1): p. 23-31.
282. Plasman, K., *et al.*, Probing the efficiency of proteolytic events by positional proteomics. Mol Cell Proteomics, 2011. 10(2): p. M110 003301.
283. Stoehr, G., *et al.*, A SILAC-based approach identifies substrates of caspase-dependent cleavage upon TRAIL-induced apoptosis. Mol Cell Proteomics, 2013. 12(5): p. 1436-50.
284. Dix, M.M., G.M. Simon, and B.F. Cravatt, Global mapping of the topography and magnitude of proteolytic events in apoptosis. Cell, 2008. 134(4): p. 679-91.
285. Rogers, L. and C.M. Overall, Proteolytic post translational modification ofproteins: proteomic tools and methodology. Mol Cell Proteomics, 2013.
286. Agard, N.J., D. Maltby, and J.A. Wells, Inflammatory stimuli regulate caspase substrate profiles. Mol Cell Proteomics, 2010. 9(5): p. 880-93.
287. Van Damme, P., *et al.*, Analysis of protein processing by N-terminal proteomics reveals novel species-specific substrate determinants of granzyme B orthologs. Mol Cell Proteomics, 2009. 8(2): p. 258-72.
288. Adamopoulou, E., *et al.*, Exploring the MHC-peptide matrix of central tolerance in the human thymus. Nat Commun, 2013. 4: p. 2039.

289. Hayakawa, E., *et al.*, Improving the Identification Rate of Endogenous Peptides Using Electron Transfer Dissociation and Collision-Induced Dissociation. J Proteome Res, 2013.