Extending the boundaries of MS-based proteomics: towards comprehensive analysis of the proteome and the HLA ligandome

Fabio Marino

Dedicated to all my family and partner

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

Extending the boundaries of MS-based proteomics: towards comprehensive analysis of the proteome and the HLA ligandome

De grenzen van MS-gebaseerde proteomics verleggen: op weg naar een diepgaande analyse van het proteoom en het HLA ligandoom

(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 29 augustus 2016 des ochtends te 10.30 uur

door

Fabio Marino
geboren op 31 mei 1986 te Erice (TP), Italy

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

Copromotor: Dr. S. Mohammed

Table of Contents

CHAPTER 1

INTRODUCTION

THE COMPLEXITY OF THE PROTEOME

Understanding how cells function and communicate is one of the major goals in molecular biology; thus global analysis of proteins, which are key entities in the cell, is required to understand how cells operate. This research field is nowadays referred to as proteomics (1) and is aimed at determining the status of all proteins ('proteome') for an organism at a certain time under a specific condition. While the human genome remains unchanged to a large extent, the proteome is a far more complex and dynamic system which undergoes radical changes in response to environment. In fact, the human genome is estimated to contain approx. 25,000 genes (2, 3) which all potentially encode for proteins. The concentrations of these proteins span in a high dynamic range (i.e. difference between the most and least abundant proteins). Moreover, protein sequence variation such as alternative RNA splicing and post-translational modifications (PTMs) further increase the complexity of proteomes raising the number of possibly protein forms encoded in the human proteome to above one million (see Figure.1). (4) PTMs regulate protein functions by changing the chemical and physical characteristics of a protein, which affect its stability, cellular localization, and their interactions with other proteins and additional biomolecules, e.g. lipids and DNA. Identification, characterization, and mapping of these modifications to specific amino acid residues on proteins are critical towards understanding their functional significance in a biological context. PTMs such as phosphorylation, methylation and glycosylation have gained particular interest, due to their high occurrence in the human proteome, their association with many cellular functions; a number of diseases are thought to be caused by deregulation of specific PTMs (5–9).

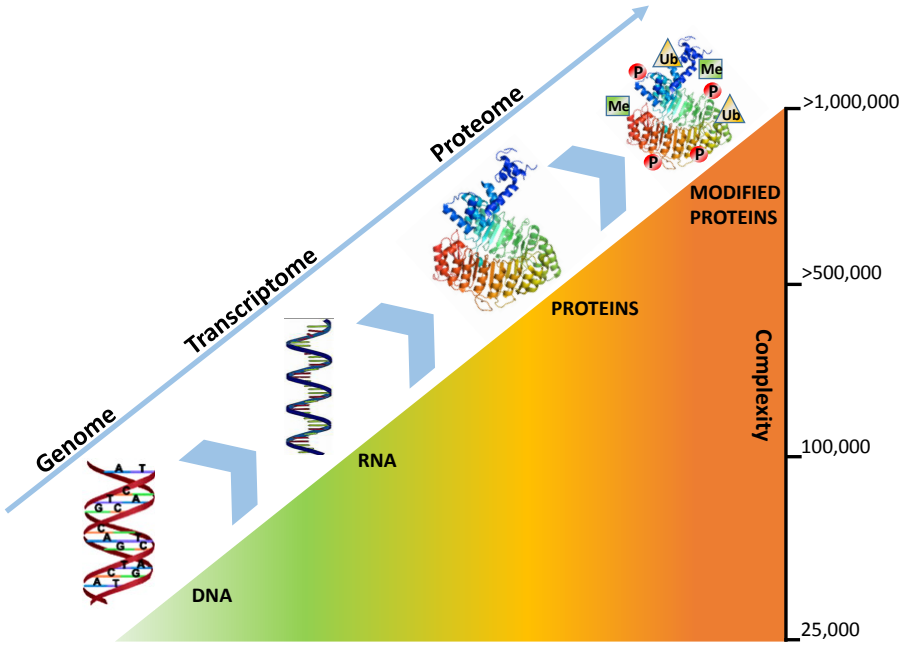


Figure.1: *The diverging complexity of the human genome, transcriptome and proteome.*

THE MOST COMMON PROTEIN POST-TRANSLATIONAL MODIFICATIONS

PHOSPHORYLATION

Protein phosphorylation is a reversible PTM in which the γ-phosphate from ATP is enzymatically transferred to a specific amino acid residue. In humans, phosphorylation is thought to occur in approx. 30% of the proteins (10) and it predominantly takes place at serine (Ser), threonine (Thr) and tyrosine (Tyr) residues, which constitute approximately 93-95%, 5-7% and <1%, respectively. (11) Although in-depth studies have also reported higher proportions of phospho Thr and Tyr. (12, 13) The introduction of the γ-phosphate might lead to, enable or hinder binding to a ligand, cofactor or substrate or might induce conformational changes that alter the function of the protein. This reversible modification is introduced by protein kinases, so far 518 kinases have been identified in the human genome (14) of which the majority are serine/threonine kinases while only 90 are tyrosine kinases.(16) Dephosphorylating enzymes are referred to as protein phosphatases which, so far, account for approx. 200 proteins in the proteome.(16)

GLYCOSYLATION

Protein glycosylation is also a reversible co- and post-translational modification which is thought to modify approximatively 40% of the proteome (17) and can occur in a number of

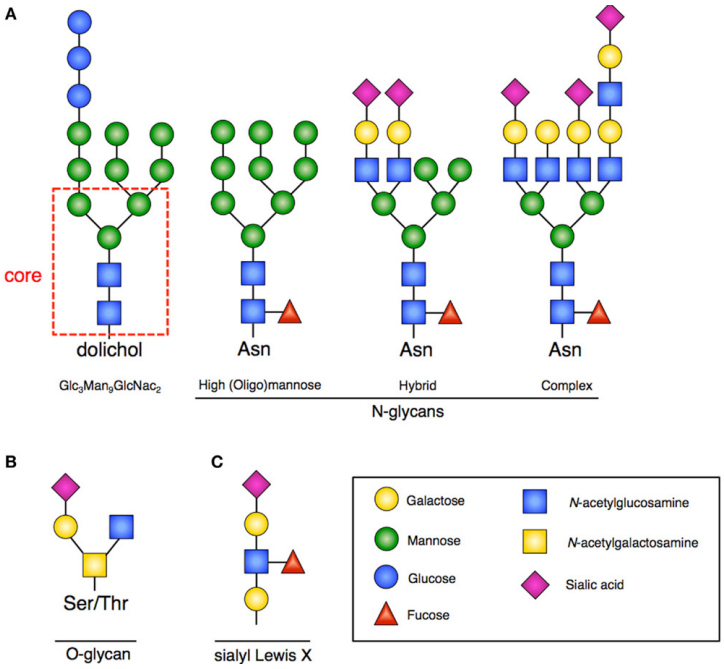


Figure.2: *A) representative example of N-glycans increasing in complexity from left to right. On the left is the sugar precursors with conserved core (dashed red box). This structure is transferred en bloc from dolichol to Asn to form, another N-glycan. Increasing branching and complexity is enabled by addition of GlcNAc to the core structure. B) Representative example of an O-GalNacylated initiated glycan with added other sugar blocks. C) Sialyl-Lewis-X structure that would be frequently found at the terminus of poly-N-acetyllactosamines. Figure adapted from Lyons et al. (18)*

distinct ways. In protein N-linked glycosylation, the oligosaccharide is covalently attached onto asparagine (Asn) residues. In mammalian systems N-linked glycans, a pre-assembled complex glycan is *en bloc* transferred from a membrane-bound lipid phosphate glycan to the Asn acceptor of a protein; the transfer occurs almost exclusively upon recognition of a consensus sequence, Asn-xxx-Ser/Thr (see Figure.2).(19, 20) The steps necessary for synthesis of N-glycans occur in different cellular compartments; therefore the final glycan composition depends on the accessibility to the different enzymes. (21) N-linked glycans can provide structural components to the extracellular matrix, modify protein properties, regulate direct trafficking of glycoproteins and mediate cell signaling. (19) In O-glycosylation the oligosaccharide is attached mainly onto Ser and Thr. Several different types of O-glycosylation exist in human including O-GalNAc, O-Mannose, O-Xylose, O-Fucose, O-Glucose and O-Galactose which are found on proteins passing through the secretory pathway (e.g. the ER and Golgi), while O-GlcNAc is mainly found on cytosolic and nuclear proteins.(22) O-GlcNAcylated and O-GalNAcylated proteins are the most frequent O-glycans, the latter ones are often termed mucins. The first step of mucin O-glycosylation is the transfer of N-acetylgalactosamine (GalNac) from UDP-GalNAc to Ser or Thr residues, which is catalyzed by a polypeptide-N-acetyl galactosaminyltransferase. Subsequently, with the addition of the next sugar, different mucin O-glycan core structures are synthesized (see Figure.2). (23) In the secreted mucins of the respiratory, gastrointestinal, and genitourinary tracts, as well as those of the eyes, the O-GalNAc glycans of mucous glycoproteins are essential for their ability to hydrate and protect the underlying epithelium. Mucins also trap bacteria via specific receptor sites within the O-glycans of the mucin. (24) O-GlcNAc is an uncharged acetylated hexosamine sugar attached through a glycosyl linkage; UDP-GlcNAc is used as a precursor to catalyze O-GlcNAc addition. O-GlcNAc is added to and removed from nucleocytoplasmic and mitochondrial target proteins by the intracellular enzymes O-GlcNAc transferase (OGT) and O-GlcNAcase (OGA). (25) O-GlcNAcylation also appears to be particularly abundant on proteins involved in processes such as signaling, stress responses, and energy metabolism. It also occurs on many cytoskeletal regulatory proteins. O-GlcNAc-ylation and O-phosphorylation exhibit a complex interplay on signaling, transcriptional, and cytoskeletal regulatory proteins within the cell.(26)

METHYLATION

In protein methylation a methyl moiety from S-adenosyl-L-methionine (SAM) is enzymatically transferred to acceptor groups on substrate proteins. This PTM is mainly found on arginine (Arg) and lysine (Lys) residues. The enzymes that regulate these enzymatic processes are called methyltransferases. Approximately 1-2% of genes from eukaryotic organisms encode methyltransferases.(28) The Ɛ-amino group of Lys may be mono-, di- and tri-methylated by the lysine methyl transferases (KMTs) (see Figure.3 lower part); while Arg may be only methylated or di-methylated by arginine methyl transferases (RMTs) (see Figure.3 upper part).(29) In humans RMTs are divided in three types.(30) Type I, II and III RMTs all generate mono-mehtylarginine (MMA) on one of the terminal nitrogen atoms; while type III can only produce MMA, type I and II RMTs respectively catalyze the formation of asymmetrical di-methylation (ADMA) and symmetrical di-methylation (SDMA). (31) Protein methylation was thought to be irreversible for many years, until recent identification of lysine demethylatases (KDMs), (32) while the apparent absence of protein arginine demethylases suggested that the only way to reverse the effects of the modification would be to degrade the protein

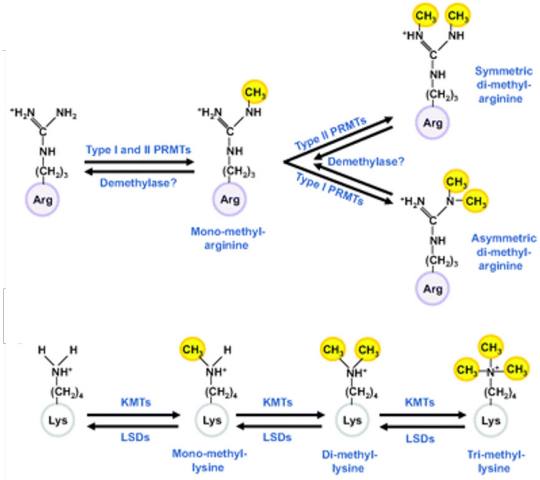


Figure.3: *(Top) Arginine methyl transferases (RMT) mediated mechanisms of mono-methylation and subsequent either symmetric (SDMA) or asymmetric (ASMA) di-methylation of arginine residues. (Bottom) Lysine methylation and di-methylation reactions are carried out by lysine methyl transferases (KMTs and KDMs), respectively. Adapted from Pek et al. (27)*

to its component amino acids and then make a new unmodified version by protein synthesis or dilute it out during cell replication.(30) Methylation of Arg and Lys of histone proteins have long been known to play a critical role in regulating gene expression, contributing to the so-called histone-code and the emerging field known as epigenetics,(33). More recently an increasing number of non-histone proteins involved in a variety of biological processes have been observed to harbor either/or lysine and arginine mono- or di-methylation.(30, 34)

MASS SPECTROMETRY (MS)-BASED PROTEOMICS

In its broadest definition proteomics encompasses a diverse number of techniques that allow different aspects of protein structure and function to be analyzed. These methods vary from immunoassays,(35) protein microarrays,(36) X-ray crystallography to NMR spectroscopy. (37) Among all proteomic techniques, liquid chromatography (LC) mass spectrometry (MS) based approaches have emerged as the prime methods for analyzing the abundance and function of proteins in given biological systems, sometimes also when perturbations are induced in the systems environment. Different strategies can be adopted in MS-based proteomics. The analysis of proteins can be approached at the peptide level, commonly referred to as bottom-up or shot-gun proteomics.(38) Alternatively, analysis can be performed at the intact protein level, known as top-down proteomics. Bridging the area between these two approaches a third approach is emerging, called middle-down proteomics whereby the analysis is focused on larger peptides and/or small proteins. (39) By far, the most used approach is still bottom-up proteomics, largely due to the ease of handling of smaller peptides, which can be relatively easily separated by LC and efficiently detected by ESI-MS, when compared to intact proteins or long peptides. (38) One of the major drawbacks of bottom-up proteomics analysis is the increase in complexity of the sample due to the generation of hundred thousands of peptides from the original mixture of thousands of proteins present in a cellular lysates. The continuous developments in sample preparation, PTM enrichment techniques, LC and MS over the last decade have enabled and facilitat-

ed large scale protein sequencing of samples. Such data subsequently allowed a deeper understanding of biological functions and cellular processes of proteins, their interacting partners and PTMs state. Despite much progress, the comprehensive characterization of proteomes by bottom-up workflows is still a major challenge and requires extensive resources in terms of time, sample amounts, instrumentation and costs. Three major factors hamper comprehensive protein identification by bottom-up proteomics: the complexity of cellular proteomes, the high dynamic range at which proteins are present in biological samples and the sensitivity of MS-based proteomics techniques for proteins of low abundance.

GENERAL WORKFLOW IN MASS SPECTROMETRY BASED PROTEOMICS

Mass-spectrometry (MS)-based bottom-up proteomic experiments generally involve several steps (see Figure.4). Starting with the extraction of proteins from their biological matrix, a peptide mixture can be obtained from the sample of interest by proteolytic digestion. The level of PTMs in specific signaling pathways can also be the object of a study; however, due to their sub-stoichiometric levels an extra step of enrichment for proteome-wide modification analysis is still essential. (40) The peptides can then be introduced into a one-dimensional (LC) or multi-dimensional (LC/LC) liquid-chromatography system. After separation, they are eluted into an electrospray ionization (ESI)-tandem mass spectrometer (MS/MS). The mass-to-charge ratios (m/z) of the peptide ions are measured first by mass spectrometry to determine the molecular mass of each precursor peptide. Then, peptide ions are isolated in the first mass analyzer (MS 1) and fragmented. The m/z values of the resultant fragments are measured in the second mass analyzer (MS/MS), producing a tandem mass spectrum. Peptide sequences can be automatically determined by matching the observed fragmentation spectra with theoretical fragmentation spectra generated *in silico* from genomic databases. As these databases can be rather large, robust statistical tools need to be applied in the form of scores for peptide matching, confidence

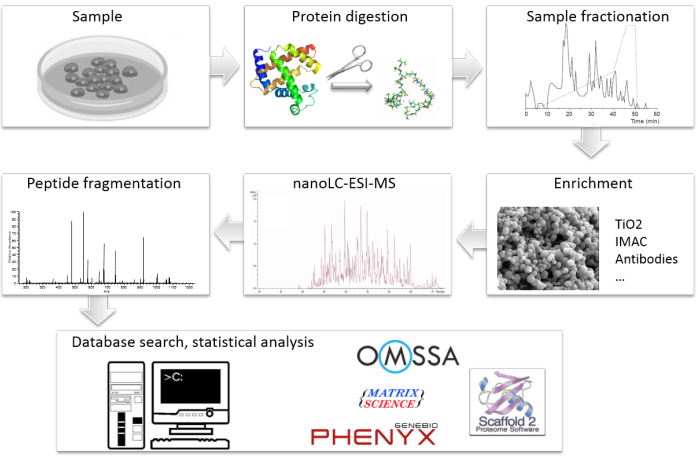


Figure.4: *Overview of several modules contributing to a generic bottom-up proteomics workflow.*

thresholds and false discovery rates (FDR). These steps are described in more detail below

SAMPLE PREPARATION

In bottom-up proteomics, there are several sample preparation steps that have to be followed to obtain peptides. These procedures can be subjected to variation based on the biological sample of interest. In general, the first step consists of extracting/isolating the desired proteins from the sample matrix, which can be composed of cultured cells, organelle extracts, tissue or even whole organisms. It's often necessary to disrupt the cell membrane/wall, and to solubilize and isolate the proteins. This can be achieved by using the right buffer conditions and performing sonication, followed by centrifugation. This process mainly allows recovery of the cytosolic proteins. To reach proteins from specific organelles and from the membrane more efficient and specific extraction steps are required. (41) To preserve protein modifications and also prevent premature protein degradation by endogenous proteases, mixtures of enzyme inhibitors are often added to the lysis buffer. Following protein extraction, a protein digestion has to be performed to obtain peptides. This is typically achieved after denaturation, disruption of the protein's disulfide bonds and alkylation of the resulting reduced ends, which prevents re-folding of the proteins. The solubilized proteins are then cleaved using a sequence-specific protease. Trypsin is most often used due to its highly specific cleavage at the C-terminus of Arg and Lys residue, (42) generating peptides suitable for most peptide separation technologies and subsequent identification by MS. In recent years several alternative enzymes to trypsin have been introduced and applied in large-scale proteomics experiments. (43) These enzymes have been shown to improve proteome coverage by generating data complementary to trypsin-based approaches by accessing different portions of the proteome neglected when only one enzyme is used. (43)

IONIZATION

Mass analysis is performed under vacuum. To introduce biological samples into a mass spectrometer the compounds must first be converted to gas-phase ions, which for biomolecules is nowadays typically performed by MALDI (matrix assisted laser dvesorption/ionization) or ESI (electro spray ionization). In MALDI, the biological sample is mixed with a matrix and co-crystallized on a MALDI target plate by vaporization of the solvents. (44) A laser is fired at the sample to desorb and ionize the matrix organic dye molecules. The charge is (partially) transferred from matrix originating ions to the analyte biomolecules resulting (in positive ion mode) in predominantly singly protonated peptide ion species. In ESI, the analyte is dissolved in a liquid solvent and sprayed through electrostatic charging. (45) At this stage, the volatile solvent in the droplet evaporates and as the solvent evaporates, the charge density of the droplets increases until the so-called Rayleigh limit is reached. At this juncture, Coulombic repulsion exceeds droplet surface tension and fission occurs. (46) Evaporation continues to occur leading to multiple rounds of fission. The gas-phase ion is formed when all the solvent is evaporated. ESI typically produces multiply charged ions, where the exact charge depends on the size of the peptide and its amino acid sequence. The advantage of MALDI over ESI is that MALDI is relatively more tolerant to salts, contaminations and detergents. Also, a sample can be reanalyzed as non-ionized compounds are conserved in the matrix spot. The largest advantage of ESI over MALDI and the reason why it is now the most prominent ionization technique in proteomics, is that it can be easily hyphenated with liquid chromatography (LC). The throughput is thereby increased. More importantly, the constant sample flow of ESI compared to the pulsed generation of ions in MALDI, allows maximal exploitation of the sequencing speed of current mass ana-

lyzers. In extremely complex biological samples, the effectiveness of an ESI-MS workflow depends also from its achievable sensitivity and dynamic range. (47) Thus, optimization of ESI has been a fundamental aspect in progression of MS-based approaches. The ESI process is strongly dependent on the physical and chemical properties of the sample liquid (surface tension, conductivity, polarity, vapor pressure) as well as on external parameters such as liquid flow rate and needle dimensions. To avoid problems induced by the high surface tension of pure water and to facilitate a stable spray, typically mixes of methanol (MeOH)-water or acetonitrile (ACN)-water are used as solvents. The ions signals observed in an ESI mass spectrum are typically protonated sample molecules in positive-ion mode, thus to promote uniform charging by protonation, acidified solutions are preferred. Operating ESI in a so called nano-ESI mode, with flow rates generally ranging from 50 to 500 nl/min has been shown to significantly improve ESI-MS sensitivity compared with ESI operated at >1 µl/min because the smaller charged droplets generated by nano-ESI lead to increased ionization efficiency. (48) The creation of smaller-diameter droplets through nano-ESI also contributed to have a more efficient charging of gas-phase analyte molecules; it also leads to lower competition and suppression effects due to the presence of less analytes in smaller droplets that ultimately lead to an extension of the dynamic range. (47) An essential advantage of nano-ESI is the very small amount of sample required for an analysis.(49) Nano-ESI has displayed an increased tolerance to salt, detergents, solvent composition and ion pairing agents.(49, 50)

ONE DIMENSIONAL PEPTIDE SEPARATION STRATEGY

Although MS technology has rapidly improved with the advent of higher sequencing speed, higher resolution and more sensitive instrumentation, sometimes even the best dynamic range capabilities tolerated by MS cannot compete with wide dynamic range of proteins present in samples. Moreover, even the fastest scanning instrument cannot sequence all the peptides contained in a complex mixture. Thus, the deeper and comprehensive characterization of proteomes routinely relies on efficient liquid chromatography technologies prior to nano-ESI-MS analysis. A LC separation occurs in a column packed with particles (stationary phase) while the mobile phase flows through the stationary phase. The samples are diluted in solution and added to the mobile phase. The principle of liquid chromatography is based on the interactions of the sample with the mobile and stationary phases. Since there are many stationary/mobile phase combinations that can be employed when separating a mixture, there are several different types of chromatography classified based on the physical properties of the used phases. For peptide separation the most common modes adopted are: reversed phase (RP), ion exchange chromatography (IEX), hydrophilic interaction liquid chromatography (HILIC) and electrostatic repulsion liquid chromatography (ERLIC) (Figure.6). In the previous paragraph the effect of flow rates on ESI process were mentioned. Thus current LC-MS configurations are performed at nanoliter per minute (nl/min) flow rates; (48, 51) columns internal diameters (IDs) have also been scaled down to 25 to 100 µm IDs to allow such flow rates.(52–57) The choice of mobile phases, (58) buffers and ion pairing agents (59) used in LC experiments are limited due to the needs and requirements for ESI and subsequent MS analysis. The most used type of chromatography for proteomics experiments directly coupled to nano-ESI-MS is RP because of the high compatibility of the buffers applied and its high peak capacity (usually defined as the theoretical maximum number

of peaks that can be obtained in a separation). The separation principle of RP is based on the hydrophobicity of the analyte and its capacity of partitioning between a hydrophobic stationary phase (which behaves almost as a hydrophobic solvent) and a polar hydrophilic mobile phase (Figure.6). RP chromatography involves loading of a peptide mixture onto a hydrophobic stationary phase in acidified water or a water-solvent mixture. The elution of bound components occurs by increasing the concentration of organic solvent. Commonly, for peptide separation, C18 materials are used as stationary phases, while mixtures of water-ACN are used as mobile phases. Solvents also contain ion-pairing agents that are used not only to modify the pH of the solution (thus the net charge of the peptides) but also to affect peptide retention on the column. There have been ongoing efforts to study the effects of chromatographic parameters such as particle size, column IDs, column and gradient lengths on LC-MS/MS performance in order to decrease the complexity of peptides eluting at any given time, with the aim to increase peptide identifications. Although better separation is achieved using a longer column and gradient, the gain in peak capacity obtained by increasing those two factors is negatively influenced by the peak broadening phenomena, which affects peak height and sensitivity. (55) The smaller the particle size, the higher the efficiency, through which a better separation may be achieved. (60) The drawback in using smaller particles is that the back pressure created by a column is inversely proportional to the square of the packing material diameter. (61) Small ID columns can provide higher separation efficiencies as well as optimized ESI sensitivity but also leading to increased backpressures of the system. Therefore smaller particle-diameter columns can easily exceed the maximum operating pressure (400 bars) of conventional HPLC (high pressure liquid chromatography) systems. The prospects of increasing the separation power by using sub-2 µm particles particle size and smaller column IDs was the main driving force that led Jorgenson et al. to pioneer ultra-high pressure liquid chromatography (UHPLC) system, which are capable of working in the range (or above) of 1000 bars.(62, 63) The developments in 1D RP-based UHPLC combined with the latest MS/MS technologies have routinely extended the achievable depth in proteome coverage to tens of thousands of peptides in few hours of analysis, by achieving peak capacities ranging from several hundreds to above one thousand, (53, 64–69) which makes RP one of the separations with the highest resolving power for peptides, today. (68, 70) Although the current 1D RP-MS/MS workflows can handle relatively complex mixtures, (71) it is still evident that the limited peak capacities in a single dimensional based approach do not have the right requirements to achieve the necessary depth in proteome coverage. An efficient way to address the limited peak capacity is to integrate RP as part of a multidimensional separation strategy.

MULTIDIMENSIONAL SEPARATION STRATEGY:

When two separation systems based on different retention mechanisms are coupled, the resulting 2D system has a higher resolving power than each single dimension. (72, 73) In theory if the two separation phases act perfectly orthogonal the final peak capacity for a 2D approach results from the product of each separation dimension. This drastic increase in separation power effectively creates a second opportunity for the peptides that co-elute in the first dimension to be resolved in the orthogonal second dimension. Multidimensional separation approaches can be roughly divided in off-line and on-line ones; on-line ones can be further divided into column switching and multidimensional protein identification technology (MudPIT) (74) set-ups (Figure.5 B-C).

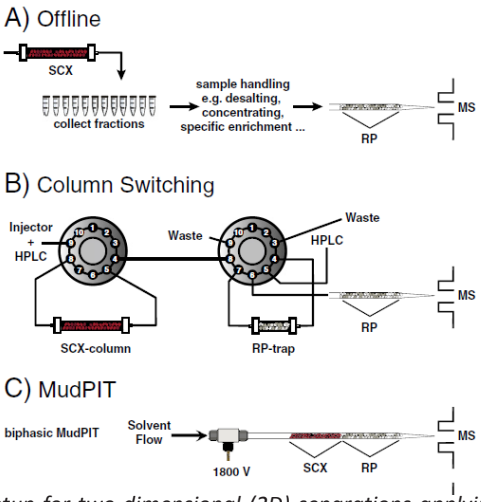


Figure.5: *Instrumental setup for two dimensional (2D) separations applying SCX and RP. A) In an off-line mode the sample is first separated by SCX and fractions are collected. The fractions can be processed if needed and are subsequently separated further by RP-LC and analyzed by MS. B) Schematic of an on-line column switching setup for SCX-RP peptide separation. The sample is first loaded onto the SCX column and eluted stepwise onto the trap column. The sample is then desalted and subsequently eluted onto the analytical RP column for further separation followed by MS analysis. C) In the MudPIT approach the SCX and RP materials are in one capillary that also functions as spray tip for direct MS analysis. In the triphasic setup an additional RP phase is packed before the SCX and functions as a trap for desalting the sample prior to SCX–RP-MS. Adapted from Di Palma et al.(75)*

In general in an off-line approach, a sample is pre-fractionated and collected after the first dimension and then further separated in the next dimension (Figure.5 A) while when a multidimensional separation is carried out in an on-line fashion there is a direct transfer of the eluent from the first dimension onto the second column (Figure.5 B-C). Off-line set-ups are often more simple in design and operation, and offer greater flexibility than on-line separations in the choice of LC modes, especially in the case of incompatibility of mobile phases between consecutive dimensions. (75, 76) In addition, fractions can be manipulated, i.e., diluted, concentrated or dissolved in a different solvent, chemically modified and, if necessary, reanalyzed. Major limitations of off-line set-ups are related to the higher sample amount required, higher risk of sample loss and contamination due to sample handling, and lack of automation of the system. When using on-line set-ups, sample consumption and handling is notably reduced, moreover on-line hyphenation can also decrease the overall analysis time. (77) However, online set-ups also have more stringent requirements, such as the solvent used for the elution in the first dimension must be a weak eluent in the second dimension, and the second dimension needs to be relatively fast in order not to lose the resolution achieved in the first dimension. In addition, whatever the on-line or off-line set-up, the column used in the second dimension must have a small ID in order not to compromise the sensitivity of the MS detection. In online valve switching approaches (Figure.5 B) the two columns are not directly connected, which provides higher flexibility in the choice of the solvent, column IDs and amount of sample that can be loaded on the first dimension. In MudPIT (Figure.5 C) the first and second dimension of separation are packed in the same column making this design less flexible. (78) Despite the excellent performance and several advantages of on-line set-ups (Figure.3 B-C), off-line systems (Fig

ure.5 A) are still more popular likely because of their enhanced flexibility. The most common technique combined with RP in a two dimensional separation is ion exchange chromatography (IEX) (Figure.6). In IEX charged analytes are mainly separated based on their charge, due to coulombic interactions between the molecules and the charged stationary phases. In order to elute the peptides, salts are generally mixed into the mobile phase. The salt's cation or anion population competitively displaces the peptide, and increasing concentrations of salt are needed to displace higher charged peptides. Since IEX separates predominantly based on charge, whereas RP separation is based on hydrophobicity, the two separation modes exhibit good orthogonality (Figure.6), (72) thus making these two techniques good candidates for multidimensional separation approaches. There are two main forms of IEX; cation exchange (CX) and anion exchange (AX) (Figure.6). In CX the stationary phase has an anionic functional group that enables binding of molecules with cationic groups (Figure.6). In strong cation exchange (SCX) chromatography the functional groups are strong acids, which enable work over a wide pH range. For proteomics purposes SCX separation has been largely employed at approximately pH 3, which results in most of the tryptic peptides having a positive net charge, which enables binding to the anionic stationary phase. (79–81) Primarily, Yates and co-workers pioneered the coupling of SCX to RP LC-MS/MS analysis, both in off-line and on-line workflows (Figure.5). (78, 82, 83) Another advantage of SCX separation (at low pH) is that it can be used for the selective enrichment of peptides harboring PTMs such as phosphorylated and N-acetylated peptides. (84) At pH 3 Lys, Arg, and the N-terminus of a peptide are positively charged while acidic residues are mainly protonated, on the other hand the phospho group at that pH still keeps one negative charge which reduces the overall positive net charge of the peptide harboring the modification. The phosphorylated tryptic peptides, that contain only a single basic amino acid in the sequence elute before the bulk of doubly and triply charged unmodified peptides. (85) Singly phosphorylated peptides containing multiple basic amino acids in their sequences elute in the bulk of the unmodified peptides, thus they require an extra step of enrichment. (40) N-terminal acetylated peptides have a reduced net charge due to the neutralization of the basic N-terminus by acetylation. Thus, N-terminal acetylated peptides with only one basic amino acid can be separated in the same way, as described for the phosphorylated peptides. It has been shown that N-terminal acetylated peptides can even be separated from singly phosphorylated peptides, even though they have the same net charge because of their orientation towards the SCX chromatographic material (84, 85) For peptides harboring methylated Arg, it has been shown that SCX and HILIC (86, 87) could provide useful separation modes for partially selective enrichment. In fact, trypsin does not cut efficiently after methylated Arg (88) and therefore such peptides are relatively highly charged because of the miss-cleavages. Combined with the frequent presence of hydrophilic residues surrounding the modified sites due to motifs recognized by RMTs these peptides can be well handled by SCX/HILIC combinations.(89) AX is another ion exchange chromatography method, which in contrast to CX, employs stationary phases positively charged, interacting with the anionic sites of the analyte molecules (Figure.6). Similiar to SCX, SAX enjoys popularity due to its applicability at a wide range of pHs and being nicely compatible with RP. Compared to SCX, SAX has not been employed as extensively as a first dimension of separation in multidimensional approaches for large-scale proteomics analysis. Only recently there has been work demonstrating SAX showing orthogonality to RP and achieving a remarkable separation power beneficial for proteome research.(90) AX has been exploited more frequently for phospho- and glyco- proteomics analysis. (91–94)

Besides IEX chromatography, other separation modes such as HILIC have been combined with RP in multidimensional setups for proteomics. In HILIC a sufficiently polar stationary phase and a typically low aqueous mobile phase (5 to 20% water in ACN) are used to create a water liquid layer, surrounding the polar stationary phase (Figure.6). The elution of the peptides is obtained by increasing the water content of the mobile phase. The exact nature of HILIC separation mechanisms has not been fully elucidated. It has been shown that the mechanism largely depends on the choice of the HILIC stationary phase and the pH of the solvent. (95, 96) HILIC has a high orthogonality with RP, which is highly beneficial for two-dimensional separations. Additionally, it allows to some extent enrichment of PTMs when their chemical properties modulate charge or polarity such as glycosylation, (97–99) phosphorylation, (100) N-acetylation (97) and methylation. (86, 101) In the case of phosphorylated peptides the additional charge on the phosphate groups renders peptides more hydrophilic thus enhancing their retention with HILIC columns. (102) Enrichment by HILIC of glyco-peptides is based on the fact that this class of peptides is more hydrophilic compared to their unmodified counterparts because of the attached carbohydrates. (99) ERLIC can be considered as a special variant of HILIC that departs from merely partitioning and exploits additionally superimposed electrostatic interactions. An IEX stationary phase is used with a highly organic mobile phase, to form a hydrophilic layer surrounding the polar stationary phase, generating hydrophilic interactions with polar analytes. In this way, all peptides in a mixture would be retained through hydrophilic partitioning, despite charged peptides being repelled to some extent by similar charges present on the stationary phase (Figure.6). ERLIC has shown to exhibit excellent orthogonality to RP and good resolving power. (103, 104) It has also been introduced as a potential phosphopeptide enrichment method, in fact the negative charge of phosphate groups

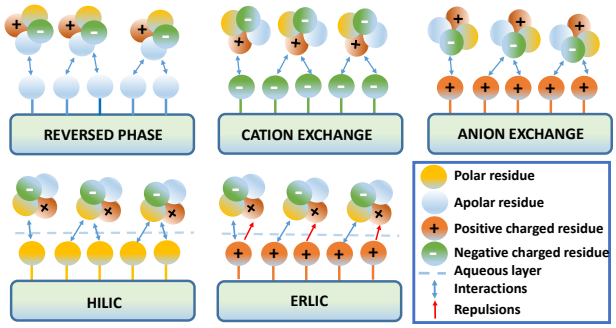


Figure.6: *Schematic illustration of possible interactions between polar, apolar, negatively charged and positively charged sites of a peptide and the most commonly used stationary phases in peptide chromatography.*

at low pH enhances the electrostatic interaction with the positively charged ERLIC stationary phase, allowing their selective enrichment from a peptide mixture. (91, 105) More recently, also RP has been combined with RP in a two dimensional separation design and this approach was shown to enable to reach very high peak capacities.(73, 96, 106, 107) In this multidimensional approach the difference between the first and second RP separation is the pH of the separation buffer. The pH influences the properties of charged residues in peptides due to pH change induced protonation or deprotonation. Neutralization of charged residues leads to a decreased hydrophilicity (increased hydrophobicity) and as a consequence changes the retention of peptides in RP. There-

fore RP-RP multi-dimensional approaches that exploits a high pH difference between the two dimensions have relatively good orthogonality. (72) A concatenated mixing scheme has also been introduced to effectively compensate for the imperfect orthogonality of the RP-RP dimensions, making more efficient use of a wider elution window in the second dimension separation compared to that of an individual fraction. (108)

METAL ION AFFINITY AND IMMUNOAFFINITY ENRICHMENT FOR PEPTIDES HARBORING PTMs

Despite the use of multidimensional separation techniques to tackle sample complexity and dynamic range issues prior to MS analysis, often other enrichment strategies, mainly based on metal affinity and immunoaffinity methods, are applied for dedicated analysis of peptides harboring PTMs. Modified peptides are quite often present at sub-stoichiometric levels when compared to the bulk of unmodified peptides. Therefore their identification can be hampered, even when compared to appraches based on the use of extensive multidimensional separation.. In the case of phosphopeptides, the most popular enrichment strategies are based on ionic interactions of the phosphate moiety with a metal ion. In immobilized metal ion affinity chromatography (IMAC) the trivalent metal ions, such as Fe3+, Al3+, or Ga3+ are chelated to a stationary phase through an immobilized poly-dentate ligand to form multiple coordinated bonds. (109–111) The immobilized positively charged metal ions form a stationary phase that can specifically accept coordination with the negatively charged phosphopeptides under acidic conditions. A major drawback of most IMAC materials is that peptides containing multiple acidic amino acid residues, i.e. aspartic acid (Asp) and glutamic acid (Glu), are also retained on the IMAC materials with the phosphorylated peptides, thereby reducing the phosphopeptide specificity. Metal oxide affinity chromatography (MOAC) provides an alternative to IMAC and uses metal oxides, such as TiO2 and ZrO2 to bind the phosphate group on phosphopeptides through the formation of bi-dentate bonds. (112, 113) Alike to most IMAC materials, the enrichment efficiency of MOAC materials is reduced by nonspecific binding of peptides containing multiple acidic residues. Ways to modulate and reduce the nonspecific binding are known such as esterification of acidic carboxylic groups (114) and the use of alternative protocols with optimized incubation buffers. (115) In the first case, due to extra steps in the procedure of esterification sample losses can be observed together with incomplete esterification (116) and presence of side reactions. (117) An alternative material using a phosphate group as the coordinating ligand for Ti4+ ions, has shown to possess superior specificity compared to other IMAC- or MOAC-based enrichment methods, (118) namely Ti4+-IMAC. Interestingly, not only phospho peptides can be enriched by metal ion affinity chromatography. In fact following the same principle N-sialylated glycopeptides (which also contain a negative net charge and thus the ability to coordinate) can be enriched by TiO2. (119) An alternative to the enrichment and separation methods described above is the immunoaffinity enrichment, which can be used for the analysis of phosphorylated proteins and is also very popular for isolation of phosphotyrosine modified peptide and proteins, (120) acetylated (121) and methylated proteins. (93) This technique relies on antibodies against specific amino acid residues and/or peptide sequence motifs containing PTMs. (122–124) For this method several milligrams of sample material are needed for an efficient enrichment due to the sub-stoichiometric levels of PTMs,

unfortunately often high-affinity antibodies are not available for PTMs of interest.

INSTRUMENTATION

In a proteomics workflow after cell lysis and protein extraction, the peptides originated from the digested proteins are separated before they are ionized by MALDI (matrix-assisted laser desorption/ionization) or ESI (electro spray ionization) to prepare them to fly into the mass spectrometer. Most mass spectrometers consist of the following parts: an ion source, transfer optics, mass analyzer(s), detector and finally data processing electronics.

MASS ANALYZERS

Mass analyzers separate ions based on their mass-to-charge-ratios (m/z) using a few different physical principles. Many types of mass analyzers exist, each having unique properties, such as mass range, duty cycle/speed, mass resolution, sensitivity, transmission efficiency, costs and dynamic range. Nowadays, quite often a combination of analyzers is used in hybrid instrumentation adding flexibility, and ideally combining the strengths of each mass analyzer. (Table.1)

Instrument	Resolution	Accuracy (p.p.m)	Fragmentation available
LTQ	2,000-20,000	<500	CID, ETD
TQ	<7,500	<100	CID
Q-TOF	<40,000	<2	CID
TripleTOF	<40,000	<3	CID
LTQ-FT-ICR	<1,000,000	<2	ECD, ETD, CID
LTQ-Orbitrap	≤120,000 (400 m/z)	<2	CID
LTQ-Orbitrap XL	≤120,000 (400 m/z)	<2	CID, HCD, ETD
LTQ-Elite	≤240,000 (400 m/z)	<2	CID, HCD, ETD, EThcD
Q-Orbitrap HF	≤240,000 (400 m/z)	<5	HCD
LTQ-Q-Orbitrap	≤500,000 (200 m/z)	<2	CID, HCD, ETD, EThcD

Table.1: *Overview of mass analyzers commonly used for the analysis of complex mixtures of peptides. Adapted from Thelen et al.(135)*

When selecting a mass spectrometer for in-depth proteomics analysis, PTMs detection and study of endogenous peptides various important parameters should be considered. In the analysis of complex mixtures, in which many peptides co-elute with very similar m/z, a sharp isolation is required to avoid the co-isolation of precursor ions, then a high resolution is essential for a precise m/z determination. For proteomics purposes, modern instruments are typically operated at mass resolutions higher than 20,000 FWHM (see Table.1). (125, 126) The mass accuracy directly determines the usefulness of MS by acting as a filter that directly reduces the number of potential false positive assignments. Nowadays instruments can reach accuracies of below parts per million (see Table.1). (127)

Fragmentation speed is very important in order to try to target for sequencing as many peptides as possible in the mixture (127, 128).The flexibility of an instrument can also be considered critical. Having the choice to switch between different analyzers adds flexibility, combining the strengths of each mass analyzer and might allow to increase the possible number of scans in a fix time. Moreover it can enable the use of complementary fragmentation techniques, which is fundamental for adapting the analysis to the analytical request. The most frequently used mass analyzers in proteomics for the above mentioned analysis are: ion traps (LTQ or LIT) (129), triple quadrupoles (TQ) (130), quadrupole-Time of flight (Q-TOF) (131), LTQ-Orbitrap (Orbitrap XL, Velos and Elite), Q-Orbitrap (Q-Exactive), Q-LTQ-Orbitrap (Orbitrap Fusion) (132), Quadrupole-Fourier transform ion cyclotron resonance LTQ-FT-ICR (133) and triple Q-TOF (TripleTOF) (134). An overview of some of their characteristics is shown in Table 1.

QUADRUPOLE MASS ANALYZERS

Quadrupole mass analyzers consist of 4 parallel, ideally hyperbolic metal rods (Figure.7). A combination of direct current (DC) and radio frequency (RF) voltages is applied to the rods of the quadrupole assembly. Gas-phase ions are introduced into the middle of the radially positioned rods, adjacent rods experience opposite potential while opposite rods have the same potentials. Based on the used DC and RF voltages, the analyzer allows only ions of a certain m/z to stabilize and pass through the device toward the detectors, while ions with a different m/z which are not stable at the applied voltages collide with the rods and become neutralized or expelled out of the quadrupole (Figure.7). By changing the voltages, the mass analyzer can be used as a mass filter, where it is either possible to isolate ions with a specific m/z or to obtain an entire spectrum by scanning all m/z values. (Figure.7). Quadrupole mass analyzers have relatively low resolution (below 10,000 FWHM) and they are often employed as mass filters in hybrid instruments.

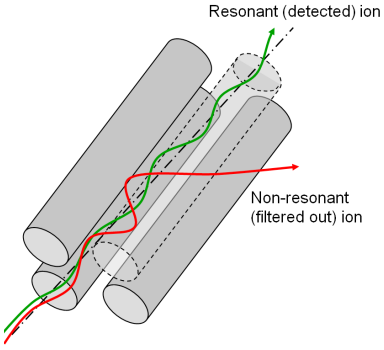


Figure.7: *Schematic of a quadrupole mass analyzer.*

QUADRUPOLE AND LINEAR ION TRAP

The quadrupole ion trap or 3D-trap consists of a circular electrode (ring electrode) and two hyperbolic caps/electrodes (endcaps), as depicted in Figure 8. An oscillating electric field is applied to the ring electrode and the end-cap electrodes are kept at ground potential. Since there is no DC potential applied, the trap will allow ions of a wide mass to charge range to have stable trajectories in three dimensions, which means the ions are effectively trapped.

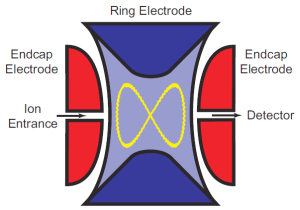


Figure.8: *Schematic picture of a quadrupole ion trap mass analyzer. The ring electrode is depicted in blue and the two endcap electrodes are colored red. Both endcap electrodes contain a hole to enable the entrance and exit of the ions. A schematic ion trajectory in the trap is depicted in yellow.*

Discrimination of m/z is performed using a changing amplitude of the RF field and resonance excitation. The linear ion trap (LIT) or 2D trap is another type of ion trap mass spectrometer (Figure.9). LIT uses a set of quadrupole rods to confine ions radially and a static electrical potential on the end electrodes to confine the ions axially (Figure.9).(136) LIT can be used as a mass filter or as a trap by creating a potential well for the ions along the axis of the trap. (137) In trapping mode, ions are confined by application of appropriate RF and DC voltages with their final position maintained within the center section of the ion trap. The RF voltage is adjusted and multi-frequency resonance ejection waveforms are applied to the trap to eliminate all but the desired ions (isolation) selecting on m/z in preparation for subsequent fragmentation and mass analysis. The design of the LIT enables to increase the ion storage volume compared to quadrupole ion traps. A benefit from an increased ion storage volume is a reduction of space charge effects. Space charge effects limit the maximum load of an ion trap, due to repulsion of ions, which in consequence disrupts their stable trajectories. (138)

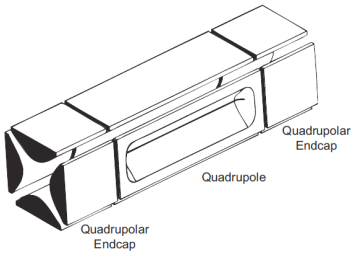


Figure.9: *Schematic of a linear ion trap mass analyzer. In this example the central section is end-capped by two small quadrupoles which works as end plates enabling axial trapping. Adapted from Schwarts et al. (138)*

ORBITRAP MASS ANALYZER

The Orbitrap is a mass analyzer composed of an inner spindle-shaped electrode, surrounded by a barrel-shaped outer electrode (Figure 10). Ions are trapped by an electrostatic field applied to the Orbitrap and move both in a rotational, radial and axial motion around the central electrode forming a harmonic orbit. Among the different ion motions, only the frequency of oscillation in the axial motion is entirely dependent on the m/z of the ions. The oscillating ions induce an image current that is detected with the help of a differential amplifier positioned on the two halves of the electrode encapsulating the Orbitrap. The detected image current is a superimposition of the currents induced by individual ion species and a Fourier transform has to be performed to deconvolute the image currents to obtain a mass spectrum. Longer transient recording times allow acquisition of more beats, which increase

the achievable resolution and resolving power. (139) The Orbitrap analyzer quickly became popular in proteomics, given its high analytical performances, such as the achievable high mass resolution (>200,000 FMHW) and mass accuracy (<5 p.p.m. with external calibration, <2 p.p.m. with internal calibration). In proteomics, Orbitrap mass spectrometers are typically combined with a linear ion trap such as the LTQ. MS measurements, where high mass resolution and accuracy are key, are performed in the Orbitrap, while MS/MS measurements are performed in the LTQ which has a faster scanning speed (see Table.1).(140)

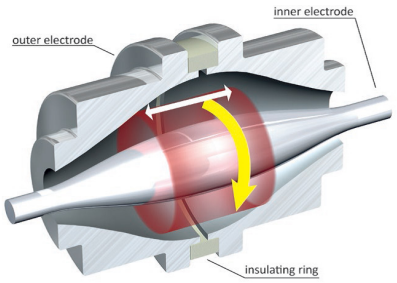


Figure.10: *Schematic of the Orbitrap mass analyzer. Injected ions move in a harmonic orbit around the inner spindle-shaped electrode (yellow arrow). They oscillate in axial direction due to the applied potential distribution (white arrow). Frequency of oscillation is only dependent on the m/z of the ions.*

PEPTIDE FRAGMENTATION TECHNIQUES

Complex mixtures of peptides easily contain thousands of ions with similar m/z, thus by only measuring the intact mass of a peptide, the amino acid sequence cannot be directly determined because isobaric/isomeric peptides cannot be distinguished (even at the higher resolving power and accuracy). To overcome this barrier peptide sequencing is required, allowing the sequence to be deduced from the detected fragment ions. Tandem MS (or MS/MS or MS2) is performed by isolating the peptide precursor ion followed by fragmenting the peptide ions and analyzing the resulting fragment ions in a second round of MS. These steps can be carried out by either tandem-in-time, or tandem-in-space approaches. In tandem-in-time instruments, the different MS/MS stages are carried out successively inside the same physical space but separately in time. On the contrary, tandem-in-space instruments involve two mass analyzers in series; thus the different stages of the process occur sequentially in separate physical regions.

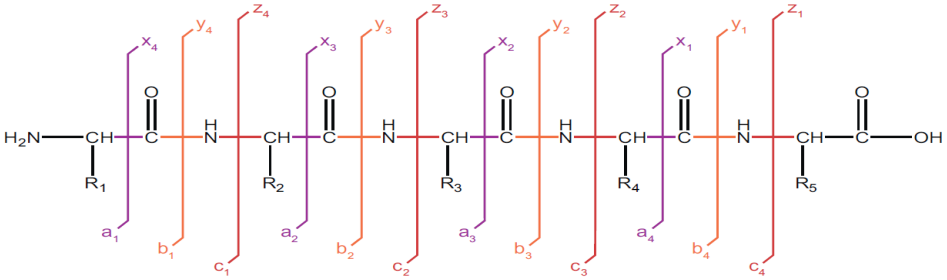


Figure.11: *Roepstorff nomenclature for the formation of peptide fragment ions. Illustration of fragment ions formed from the backbone cleavage of protonated peptides. Fragment ions retaining the positive charge on the amino terminus are termed a-, b-, or c- type ions. Fragment ions retaining the positive charge on the carboxy terminus are termed x-, y-, or z- type ions. (143, 144)*

COLLISION INDUCED DISSOCIATIONS

Collision-induced dissociation (CID), sometimes called collision-activated dissociation (CAD), is the most used fragmentation technique in proteomics for peptide sequencing. (141) In CID, the analyte ions are accelerated by an electric field and are subsequently subjected to collisions with an inert gas (often nitrogen, helium and argon). Upon collision of the precursor peptide ions with the gas, kinetic energy is partially converted into internal energy of the peptide ion that is distributed over the molecule bonds. Fragmentation of a precursor ion can occur if the collision energy is sufficiently high that the ion is excited beyond its threshold for dissociation. In the current model the fragmentation process in protonated peptide ions is initiated by a 'mobile' proton that weakens amide bonds. Migration to an amide bond nitrogen atom weakens the C-N bond and generates b- and y-ions as the internal energy increases during activation (Figure.11).(142) The energy that is necessary to mobilize a proton depends on the amino acid sequence of the peptide and its gas-phase basicity. It should be noted that the mobile protons do not reside stochastically at the different amide bonds throughout the peptide backbone, i.e. not all peptide bond cleavages are observed in CID. In fact the cleavage pattern depends (i) on the peptide charge state, (ii) on the position of the most basic residue and (iii) on the presence of proline residues.(145) The fragmentation pattern obtained by MS/MS depends also on the excitation energies and the time scales of fragmentation and detection. Depending on the used instruments we can distinguish tandem-in-space CID, which is performed in a separate collision cell, while tandem-in-time CID is performed typically in ion trap instruments. (142) CID processes occurring routinely can be further separated based on the kinetic energy of the precursor ion. In high energy CID, a kinetic energy of a few keV is applied. Because of the high kinetic energy of the ions, the time scale for dissociation in high-energy CID is usually on the order of sub-microsecond. The ions entering the collision cell, usually undergo a few collisions before mass analysis of the product ions resulting in a higher energy available for transfer to internal energy which ultimately leads to direct backbone cleavage products and side-chain fragmentation. (146) The term high-energy collisional dissociation (HCD) also refers to beam type CID. This kind of fragmentation is typically applied in TOF/TOF and Sector instruments. In low-energy CID (note that we often adopt the term HCD even for CID in the range of 20-200 eV) the kinetic energy of the precursor is in the range of 20-200 eV, thus the fragmentation process occurs in the order of a few hundred microseconds, where the ion precursor undergoes multiple collisions and sequentially absorbs more and more collision energy until the fragmentation threshold is reached. Threshold energies for several reactions are not readily achieved in the low-energy CID resulting in fewer fragment ions and in favored low-energy pathways that compete with co-occurring backbone fragmentation.(147) Low-energy CID is usually applied on triple quadrupole (QqQ) and in Q-TOF instruments. Moreover, in ion trap low-energy CID the kinetic energy of the precursor ion is only a few eV and the process takes in the order of milliseconds. For the latter kind of CID low-energy pathways are particularly favored.(147) Note that a known limitation of ion trap-CID is the loss of ions in the low m/z region. (148) Apart from the instrumentation used, the fragmentation behavior also depends on the peptide sequence and/or the presence of PTMs. Generally, CID has shown to produce limited sequence information of peptides containing multiple internal positively charged residues (i.e. non-tryptic, endogenous and tryptic miss-cleaved peptides) which can prevent random protonation along the peptide backbone inducing site specific dissociation and

few sequence ions. (149–151) Unfortunately, several peptide PTMs are labile especially under low-energy CID conditions (i.e. phosphorylation, nitrosylation, sulfonation, glycosylation) and could be difficult to identify both the modified sites and the peptide sequence when CID is applied (Figure.12 A-B). For example, in the case of phosphorylated Ser or Thr residues, the phosphate group competes with the peptide backbone as the preferred site of cleavage. The activation of these peptides (mainly upon ion trap CID) eliminates primarily phosphoric acid from the peptide, and frequently results in the peptide's backbone bonds remaining intact (Figure.12 A-B). Therefore alternative fragmentation techniques should be exploited to improve the identification of the above mentioned class of peptides.

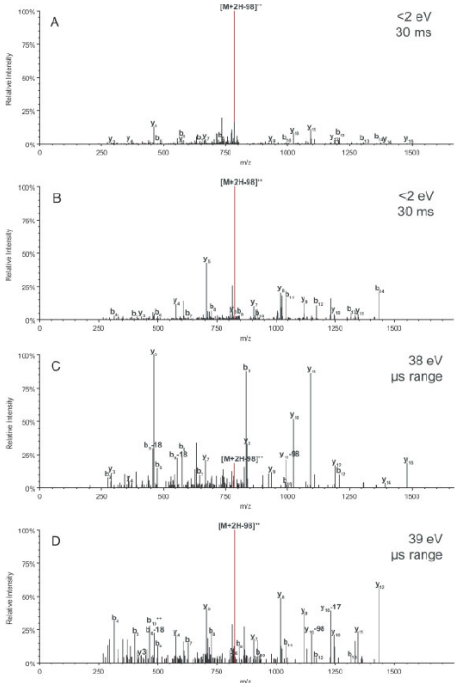


Figure.12: *Effect of collision energy regime and activation time on the extent of neutral loss. CID spectra of the same phosphopeptide precursor ions were obtained at low collision energy (<2 eV) and relatively long activation times (30 ms) in a linear ion trap (A and B) and at higher collision energy (38–39 eV) and shorter activation time (microsecond range) in a Q-TOF (C and D) with the neutral loss peak highlighted in red. A, C: SVSSNVASVpSPIPAGSK; B, D: GGSISVQVNSIKFDpSE. Adapted from Boersema et al. (152)*

ELECTRON-DRIVEN DISSOCIATION METHODS

Over the last decade, the development of electron capture induced dissociation (ECD) (153) and electron transfer induced dissociation (ETD) (154) have greatly advanced the capabilities of MS-based peptide and protein sequencing. Both are electron-driven dissociation techniques whereby the precursor protonated peptide ions capture an electron generating a highly excited odd-electron peptide radical cation. The available energy induces peptide backbone cleavage, especially of the N-Cα bonds. ECD and ETD generate thus primarily the so-called c- and z-ions (Figure.11 and 13). It is believed that both ETD and ECD follow similar mechanisms. However, the conditions under which ECD and ETD

are performed are quite different. In ECD, precursor ions are usually confined in the trap of an FTICR-MS and bombarded with near-thermal electrons (<0.2 eV). The capture of an electron by a protonated peptide ion is an exothermic process, releasing several eV of energy causing the peptide backbone to fragment promptly. (155) ECD in its most efficient form requires the immersion of the precursor sample ions in a dense population of near-thermal electrons. Emulating these conditions in the instruments used most commonly for peptide and protein analyses, those that trap ions with radio frequency (RF) electrostatic fields rather than with static magnetic and electric fields, has been technically challenging. In fact, thermal electrons introduced into the RF fields of ion trap or Q-TOF instruments maintain their energy only for a fraction of a microsecond and are not trapped. (154) As a result, ECD remained a technique almost exclusively used with FTICR. By using a different source of electrons, radical anions, ETD was first implemented on a RF based ion trap (154) later on it became available for Q-TOF and Orbitrap mass spectrometers.(156, 157) For ETD, typically a negative chemical ionization source is used to generate anions (e.g. from Fluoranthene or Azulene). (158) These anions with sufficiently low electron affinities function as a suitable one-electron donor to deliver electrons to multiply protonated peptides. Electron transfer to protonated peptides is an exothermic process which initiate fragmentation via similar pathways accessed in ECD. (154) Studies in ECD and ETD have revealed that the electron capture cross section is proportional to the square of the charge state of the precursor. This explains why ECD and ETD lead to more efficient fragmentation for higher charged peptides and to the poor dissociation of just doubly charged peptides (Figure.13 B). The latter represents a distressing problem in the analysis of tryptic peptides as the majority of them carries two charges following ESI (106). ETD can also lead to the formation of charge-reduced long-lived radical intermediates, which do not fully dissociate but are held together via non-covalent interactions. (159) The issue was partly overcome by increasing the internal energy of the precursor before or during the electron transfer reaction (160–162). Another approach adopted to solve the problem from Coon and co-workers was the implementation of a supplemental activation by applying resonant-excitation-CID following an initial electron transfer step, which resulted in a substantial increase in the overall product ion yield (Figure.13 C).(163) More recently Frese et al. (164)also implemented a HCD supplemental activation of ETD activated precursor and fragment ions, resulting in the so called EThcD hybrid fragmentation technique. Over the past years, several groups have reported comparisons between electron-driven and collision-activation methods. (165, 166) In summary, it is nowadays widely accepted that both methods have their strengths and weaknesses and their complementary use is highly recommended.(167) ETD and ECD fragmentations are more and more used because of their ability to cleave inter-residue bonds and for their virtual indifference to labile PTMs, which not only increases the chances to identify the modified peptide, but also to correctly localize the modified site. (168–173) Increasing reports describe the use of ETD or ECD techniques for the characterization of PTMs such as phosphorylation, glycosylation, sulfonation, ubiquitination and methylation. (12, 101, 168, 174, 175) ETD and ECD usefulness is not only limited to analysis of PTMs but a wide range of other application has been reported. Including the analysis of peptides possessing a higher charge density, in top-down proteomics and for endogenous (non-tryptic) peptides, such as neuropeptides and Human Leukocyte Antigens (HLA) peptides.(39, 150, 155, 176, 177) Compared to HCD and CID, ETD and ECD suffer from relatively slower duty cycles and higher costs and maintenance. In the new generation of hybrid instruments (132, 178)

some step forward have been made that significantly reduce maintenance and tuning requirements. For example in the new hybrid Orbitrap Fusion a novel front-end ETD reagent ion source has been placed,(179) making operation easier compared to the earlier filament-based ETD ionization source placed at the back of the HCD cell. (157) Moreover, Coon et al. recently implemented a calibration routine for ETD, which determines the correct number of reagent anions necessary to reach a defined ETD reaction rate. (180) These advances helped in reducing the duty cycle and increased product ion yield, thus lowering the performance gap between ETD fragmentations and collisional activation techniques in terms of scan speed, thereby also making their usage more user friendly.

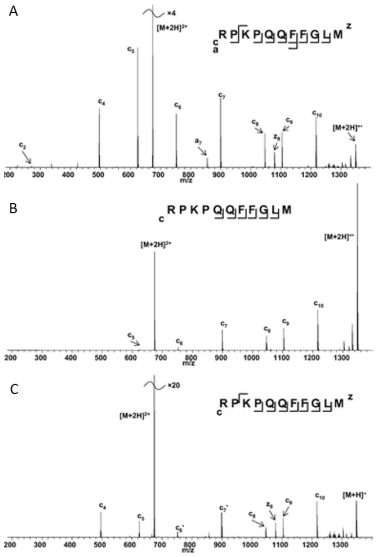


Figure.13: *(A) ECD MS/MS; (B) ETD MS/MS and (C) supplemental activation ETD MS/MS of doubly charged ions of the peptide substance P.*

DATABASE SEARCH AND STATISTICS

Although peptide sequences can be manually deduced from individual MS/MS spectra, as single LC-MS/MS runs typically generate tens of thousands of spectra, manual data interpretation of each spectrum is not really feasible. To assign spectra to peptides in an automated and high throughput manner, several database search algorithms have been developed (see Figure.14).(181–183) In principle, most database search algorithms operate in a similar manner: the first step involves the generation of a list of theoretical peptides and their corresponding masses (m/z) based on an in-silico proteome digest (Figure.14). The number of proteins in the database, the type of cleavage enzyme used, the enzyme specificity and the number of peptide modifications will largely influence the size of the peptide database. Subsequently, the m/z values and charge state of the acquired precursor masses are compared to all in-silico digested sequences. Only peptide candidates matching within a given mass tolerance are kept for further MS/MS comparison. Next, the theoretical fragment ion masses are calculated for each candidate peptide and compared to the fragmentation spectrum (Figure.14). Finally, a similarity or probability score, based on the number and type of fragment ion masses matched, is calculated for each candidate peptide and those with the highest score

are reported for each queried spectrum, the so called peptide spectrum matches (PSMs). An important factor in peptide and protein identification is the confidence with which they are identified, because when matching tens of thousands spectra to millions of possible candidates the chance of matching random incorrect hits is unavoidable. To limit the amount of false positive identifications, several statistical methods have been developed. The most used approach is to perform a second search against a so called "decoy" database, in which all protein sequences from the normal database are reversed or scrambled. (185) The number of identification found in such an artificial database estimates the number of false positives present in the authentic search. Several filtering criteria such as a minimum peptide score and/or a minimum sequence length are then applied until a desired false discovery rate, usually <1%, is reached. More sophisticated machine learning algorithms have been developed for improving the rate of confident peptide identifications, (186) which calculate the weight of multiple discriminating parameters, taking into account additional features of peptides that match the protein. Additional challenges are posed when the mass spectrometry based analysis targets the identification and localization of PTMs. To automate the interpretation of MS/MS spectra and to facilitate the comprehensive localization and scoring of phosphorylation (and other PTMs) sites, several strategies have been developed. (187–190) Two main strategies for scoring site localization reliability either try to assess the chance of a given peak that allows site determination to have been matched at random, or calculate a search engine score difference between peptide identifications with different site localizations.

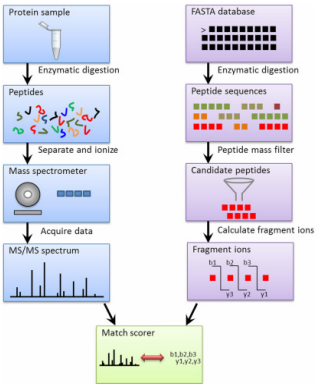


Figure.14: *Scheme of the data acquisition and database search workflow in a typical proteomics experiment. For a given experimental MS/MS spectrum, protein sequences from a database are in silico digested and peptides of the right mass are selected. Theoretical fragment ions from each candidate peptide are calculated and used to generate a similarity or probability score by comparing the theoretical fragment ion masses against the experimental spectra. Each candidate peptide is scored against the experimental spectrum and the best matching peptides and their scores are reported. Adapted from Eng et al. (184)*

MASS SPECTROMETRY (MS)-BASED IDENTIFICATION OF HUMAN LEUKOCYTE ANTIGEN-ASSOCIATED PEPTIDES

The work described in this thesis is largely focused on improving separation methods and MS based peptide analysis methods to enable a better understanding of our immune system, and in particular of the presentation of HLA peptides at the surface of cells. Following the introduction on the state-of-art peptide centric MS-based proteomics above,

in the next section a concise summary will be given to introduce the immune system.

THE ADAPTIVE IMMUNE SYSTEM

The vertebrate immune system consists of complex cellular networks that have evolved to defend the host against a wide array of pathogens and has been implicated in host protection from tumor-genesis. These processes are commonly separated into two categories; the innate immune response and the adaptive immune response. Cells of the innate immune system express pattern recognition receptors (PRRs),(191) which allow them to recognize pathogens. The innate response does not require prior contact with the pathogen and subsequent encounters with the same antigen do not lead to an increased activation response. In contrast, in order to be capable of engaging the key elements of adaptive immunity (specificity, memory, diversity, self/non-self discrimination), antigens have to be processed and presented to immune cells. Antigen presentation is mediated by MHC (Major histocompatibility complex) class I molecules, and class II molecules found on the surface of antigen-presenting cells (APCs) and certain other cells. MHC class I and II molecules are similar in function: they deliver peptides to the cell surface allowing these peptides to be recognized by CD8+ (cytotoxic) and CD4+ (helper) T cells, respectively. The difference is that the peptides originate from different sources – endogenous, or intracellular, for MHC class I; (192) and exogenous, or extracellular for MHC class II. (193) There is however also a so-called cross-presentation in which exogenous antigens can be presented by MHC class I molecules. (194) Endogenous antigens can also be presented by MHC class II complexes when they are degraded through autophagy.(195)

Human Leukocyte Antigen (HLA) CLASS I PROCESSING AND PRESENTATION

In most mammalian species, MHC class I heavy chains are encoded by three genes (HLA-A, HLA-B and HLA-C in humans), and all three are polymorphic, constituting the most unique characteristic of HLA molecules. Polymorphism of the HLA proteins results in different peptide-binding grooves that recognize unique peptides owing to variations in the anchor residues to which peptides dock (see Figure.15 A). The consequences of these polymorphisms are differential susceptibilities to infection and autoimmune diseases (196) that may result from the high diversity of peptides that can bind to HLA class I in different individuals.(197, 198) HLA class I molecules are expressed by all nucleated cells. HLA class I molecules are assembled in the endoplasmic reticulum (ER) and consist of two types of chain a polymorphic heavy chain and a chain called β2-microglobulin (Figure.16). The heavy chain is stabilized by the chaperone calnexin,(199) prior to association with the β2-microglobulin. A peptide is the third component required for stability, as it inserts itself into the HLA class I peptide-binding groove, which accommodates peptides of 8–12 amino acids (Figure.15 A-C). Peptides are derived from the degradation of proteins, which can be of viral or self-origin. Degradation of proteins is mediated by cytosolic- and nuclear proteasomes. (200) When not bound to peptides the HLA molecules can be stabilized by chaperone proteins and tapasin. The complex of TAP (transporter associated with antigen presentation), tapasin, HLA class I and the chaperone proteins is called the peptide-loading complex (PLC).(202) TAP translocates polypeptides, which may require additional trimming in the ER before binding to the HLA class I complexes. This is possibly due to the presence of the ER aminopeptidase (ERAAP) associated with antigen processing. (203) When peptides bind to HLA class I molecules, the

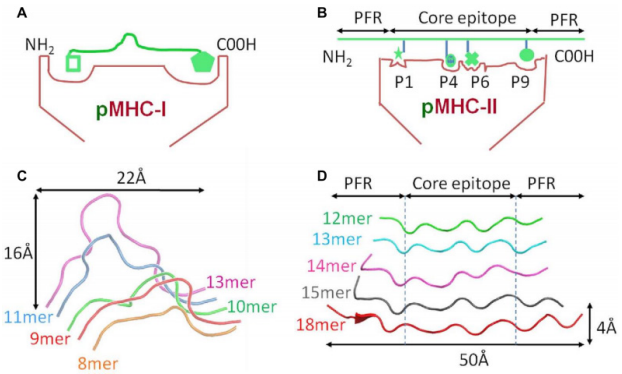


Figure.15: *Characteristics of peptides presented by MHC-I or MHC-II molecules. Schematic cross sections of the pMHC-I (A) and pMHC-II (B) binding grooves showing the key anchor sites in the groove. (C) pMHC-I complexes generally assume a central bulged conformation. As peptide length increases, the "closed" nature of the pMHC-I binding groove forces the central residues of the peptide out of the groove to accommodate the extra residues. (D) In contrast, the pMHC-II binding groove is "open" enabling longer peptides to extend out of the groove forming peptide flanking regions. Adapted from Holland et al. (201)*

chaperones are released and peptide–HLA class I complexes leave the ER for presentation at the cell surface (Figure.16). In some cases, peptides fail to associate with HLA class I and they are returned to the cytosol for degradation. On the surface of a single cell, HLA class I molecules provide a readout of the expression level of over 10,000 peptides. (192) The usual process of antigen presentation through the HLA I molecule is based on an interaction between the T-cell receptor and a peptide bound to the HLA class I molecule to monitor the events inside the cell and detect infection and/or tumor genesis (Figure.16).(204)

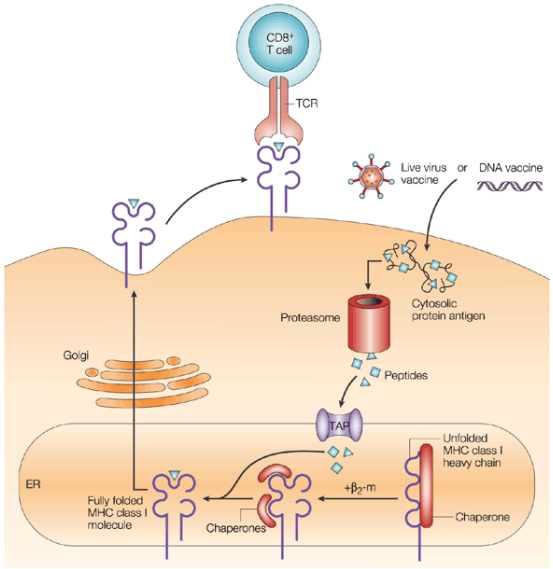


Figure.16: *Cartoon displaying HLA class I processing and presentation. Intracellular antigens, such as virus or tumor antigens, are processed into peptides by the immune-proteasome. Peptides are transported into the endoplasmic reticulum (ER), where they are loaded into the groove of the HLA class I complex. HLA class I complexes present antigens on the cell surface to CD8+ T cells. Adapted from Hanke et al. (205)*

HLA CLASS II PROCESSING AND PRESENTATION

MHC class II molecules in humans are encoded by three polymorphic genes: HLA-DR, HLA-DQ and HLA-DP. (206) HLA class II molecules are expressed by APCs, such as dendritic cells (DC), macrophages and B cells (Figure.17). HLA class II molecules bind to peptides that are derived from proteins degraded in the endocytic pathway.(207) HLA class II complexes consist of α- and β-chains that are assembled in the ER and are stabilized by the invariant chain (Ii). (208) The complex of HLA class II and Ii is transported through the Golgi into a compartment which is termed the HLA class II compartment (MIIC) (Figure.17).(209) Due to its acidic pH, proteases cathepsins (mainly S and L) are activated and digest Ii, leaving a residual class II-associated Ii peptide (CLIP) in the peptide-binding groove of the HLA class II.(210, 211) Later, the CLIP is exchanged for an antigenic peptide derived from a protein degraded in the endosomal pathway (Figure.17). This process requires the chaperone HLA-DM (Figure.17). (212) Unlike the more uniform peptide lengths (usually 8–12mers) bound in the HLA-I closed groove, HLA-II presented peptides are of a highly variable length (Figure.15 D). (213) The bound peptides consist of a core 9mer (reflecting the binding motif for the particular HLA-II type) but with variable peptide flanking residues (PFRs) that can extend from both the N- and C-terminus of the core (Figure.15 B-D). (213) HLA class II molecules loaded with peptide antigens are transported to the cell membrane to present their cargo to CD4+ T cells (Figure.17).

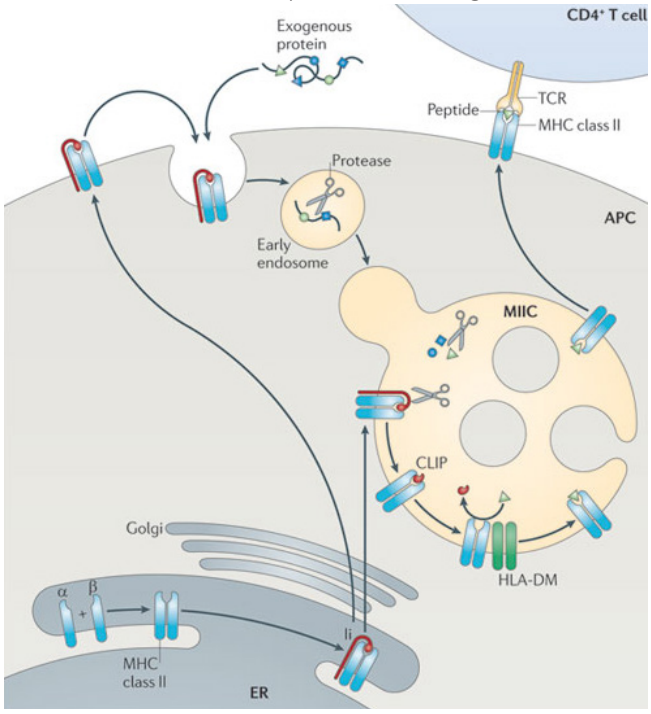


Figure.17: *Schematic of HLA class II antigen processing and presentation. Antigens from extracellular sources, such as bacterial antigens, are processed by endolysosomal enzymes into peptides of variable length. These peptides bind to the groove of the HLA class II complex by displacing the class II-associated invariant chain peptide (CLIP), which is derived from the HLA class II-associated invariant chain (Ii). HLA-DO and HLA-DM regulate the antigen-loading process. The HLA class II complex presents antigens to CD4+ T cells. Adapted from Neefjes et al. (214)*

IDENTIFICATION OF HLA CLASS I AND II ASSOCIATED PEPTIDES BY MASS SPECTROMETRY (MS)-BASED APPROACHES

To gain more insight in to antigen processing and presentation pathways we need to chart the identity of all HLA-associated peptides. Unambiguous identification of foreign peptides, being discriminated from the way more frequently occurring self-peptides, is pivotal for vaccine development and immune therapy design.(214) Ultra high-performance liquid chromatography (UHPLC)–ESI–tandem MS (MS/MS) is nowadays the state-of-the-art technology enabling the identification of several hundreds of HLA ligands in a single experiment. HLA molecules are membrane proteins that present peptides, thus enrichment procedures have been developed for the targeted analysis of HLA-associated peptides (Figure.18). (192, 215) Such approaches are initiated by affinity purification of the intact HLA-peptide complexes from cell lysates utilizing dedicated anti-HLA antibodies. Alternatively, direct mild acid elution from intact cells can be performed, but this approach often provides a high background of non-HLA peptides.(192) The starting source material the affinity purification approach can be of different nature, such as tissue material, blood circulating cells or pellets from cultured cell lines.(216, 217) Typically, after detergent-based cell lysis, HLA-peptide complexes are isolated by subjecting the cell lysate to an affinity column to which a monoclonal antibody has been coupled. Peptides are subsequently released from the purified HLA-peptide complexes by acid elution and further purified by high-molecular weight cut-off filtration (typically 10 kDa) or C18 stage tips.(218) The enriched fraction of HLA-associated peptides is analyzed by using relatively standard peptide-centric proteomics workflows (Figure.18), using extensive fractionation (when the peptide yield allows it), subsequent RP-MS/MS analysis and dedicated database searches. Although these workflows are continuously improved, challenges remain in the purification of the proteins and the peptides, MS-based sequencing and data analysis of the HLA-presented peptides. The physical chemical properties of these endogenous peptides are very diverse compared to for instance tryptic peptides from a digested lysate. In fact, the variable nature and occurrence of internal basic residues represents difficulties in peptide sequencing using collision dissociation-based techniques and interpretation of the MS/MS spectra.(219) Yet, CID has been the most used fragmentation technique by far for both HLA class I and II peptides. (193, 195, 220, 221) The use of HCD instead of CID has been shown to be beneficial in the identification of HLA class I bound peptides increasing identification rates. (218, 219) Recently, EThcD was compared to ETD, HCD and CID for the identification of HLA class I endogenous peptides (Figure.19 A-C),(150) revealing its particular advantages expanding the detectable ligandome. Although EThcD has proven to be particularly beneficial for the identification of certain peptides, the allele structural diversity of HLA class I peptides (i.e. different alleles have different anchor as amino acids, thus different peptide sequences) makes them even harder to analyse, thus combining fragmentation strategies, similar to what has been done for tryptic peptides, (176) could result in less biased analysis of the HLA peptidomes. Compared to class I HLA peptides improvements in the detection of HLA class II peptides have been somehow neglected. Here, the challenge is even more demanding due to the even higher diversity of HLA class II bound peptides (Figure.15 B-D). Thus, new analytical approaches are needed to expand the depth of analysis for this class of ligandomes. Data analysis by automated search algorithms is not straightforward because HLA-presented peptides are processed by a variety of enzymes with multiple terminal cleavage specificities. As a consequence, MS/MS spectra are searched against the the-

oretical proteome database without an enzyme restriction. This leads to a large increase in database search space, accompanied with an increase in false positive identifications. (222)

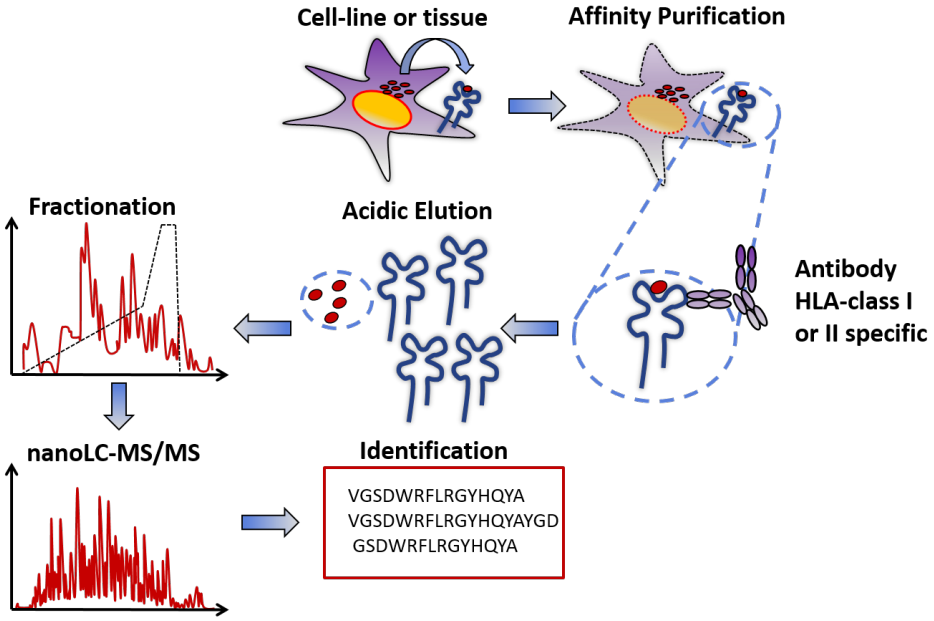


Figure.18: *General MS-based workflow for the identification of HLA-bound class I and II peptides.*

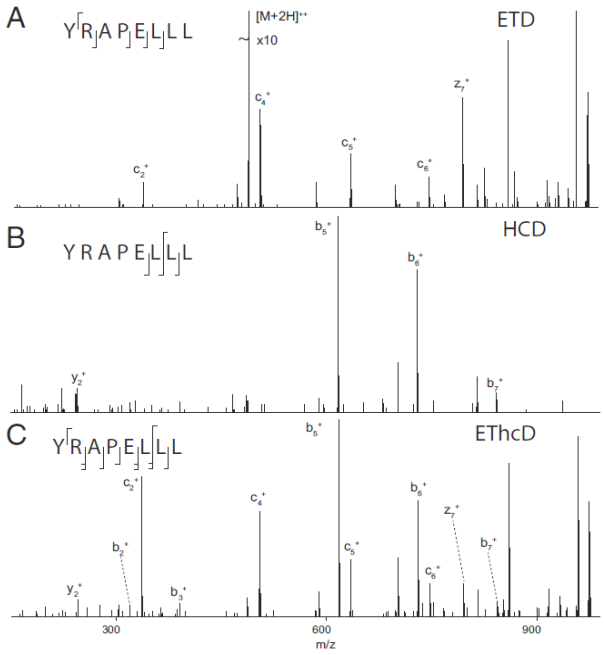


Figure.19: *HLA class I peptide fragmentation by ETD, HCD, and EThcD. Illustrative MS/MS spectra of the peptide YRAPELLL upon fragmentation by (A) ETD, (B) HCD, and (C) EThcD. The observed and/or assignable c/z and b/y fragment ions are indicated above and below the peptide sequence. Adapted from Mommen et al. (150)*

HLA-ASSOCIATED PEPTIDES HARBORING POST TRANSLATIONAL MODIFICATIONS

There have been a few reports revealing that some HLA ligands contain modified amino acid residues. For instance, an acetylated epitope from myelin basic protein was already reported over 20 years ago.(223) In the meantime, a long list of different PTMs have been identified for ligands of both HLA class I and class II.(224) For class I ligands, PTMs described up to now include O-linked glycosylation, (225) acetylation, (223) phosphorylation of serine or threonine, (226) deamidation induced by deglycosylation, (227) methylation, (228) and cysteinylation. (229) Although in some cases such modified peptides have been identified by analyzing the specificities of T cell lines, (223, 230) the standard method for sequencing has been MS. Since peptides with PTMs constitute only a small fraction of the ligandome, the application of enrichment strategies for glycopeptides (231) and phosphopeptides (232) has greatly facilitated their isolation and identification. The presence of PTMs in HLA ligands provides crucial insights into the antigen-presentation pathway. First of all, the modified amino acids obviously do not generally block the degradation, transport, and peptide-loading process, since otherwise such peptides could not be presented. In some N-glycosylated proteins, the enzymatic deamidation of the glycan-bearing asparagine to aspartate is a necessary precondition for the presentation on HLA class I molecules.(227, 233, 234) This reaction is catalyzed by the cytosolic enzyme peptide-N-glycanase, so the proteins have to be transported from the ER to the cytosol in their glycosylated form and there they become deglycosylated prior to proteasomal degradation. In this process, an epitope with a novel primary sequence is created, and one study has shown that the presentation of the non-deamidated epitope is possible if the glycosylation is defective.(235) Thus the glycosylation status can be rendered directly perceptible at the ligandome level. Furthermore, certain PTMs can enhance the presentation of the peptides carrying them. For example, it has been shown that phosphopeptides with suboptimal secondary anchor residues can be presented on HLA-A*02 because of stabilizing effects of the phosphate moiety. (236) Phosphopeptides as HLA ligands are especially of interest since phosphorylation in most cases is not a constitutive modification, but is rather intricately regulated. The presence or absence of the PTM thus creates the possibility for the immune system to recognize changes in different PTM status caused by inflammation, infection, or tumor genesis. Furthermore, large-scale MS analysis of the phospholigandome of cell lines from melanoma and ovarian carcinoma (226) has revealed several phosphopeptides presented on cancer cell lines but not on the EBV-transformed lymphoblast line JY. These results demonstrated that phosphorylated ligands could be presented in a tumor-specific manner. These findings highlight the potential important implications that the study of HLA ligands with PTMs has for immunotherapeutic approaches. In the case of a glycosylated HLA class II-restricted epitopes derived from type II collagen, it has already been established that differences in the PTM of the peptide is a critical factor in pathogenesis.(237) Thus PTMs can create new exciting possibilities for neo-antigens to be presented, and new opportunities for the development of therapies against pathogens, transformed cells and autoimmune diseases. However, responses to modified antigens will have to be carefully evaluated on an individual basis to determine if and through which mechanisms they contribute to disease pathogenesis.(224, 238, 239) Advances in MS not only gave the opportunity to increase the depth of the identified repertoires of HLA class I and II bound peptides, but it also rendered possible to discover several classes of PTMs harboring the associated HLA peptides. (224) In proteomics workflows the use of ETD fragmentation for the identi-

fication and site localization of PTMs has been shown to be fruitful. In the same way, the identity of the peptide, PTMs and exact site localization in immunopeptidomics are important to determine the possible mechanisms behind the presentation and recognition by T cells of a possible PTM peptide. Thus, ETD-based techniques such as EThcD or ETciD should be more explored to investigate the presence of PTMs on bound HLA bound peptides

REFERENCES

1. Wasinger VC1, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL, Humphery-Smith I.Wasinger VC1, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL, H.-S. I. Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. Electrophoresis 16, 1090–4

2. Baltimore, D. (2001) Our genome unveiled. Nature 409, 814–816

3. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Miklos, G. L. G., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., Mckusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Francesco, V. Di, Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V, Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z. Y., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S. C., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., Mccawley, S., Mcintosh, T., Mcmullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-stine, J., Caulk, P., Chiang, Y., Coyne, M., Dahlke, C., Mays, A. D., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., Mcdaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (2001) The Sequence of the Human Genome. 291,

4. Gstaiger, M., and Aebersold, R. (2009) Applying mass spectrometry-based proteomics to genetics, genomics and network biology. Nat Rev Genet 10, 617–627

5. Cohen, P. (2001) The role of protein phosphorylation in human health and disease: Delivered on June 30th 2001 at the FEBS meeting in Lisbon. Eur. J. Biochem. 268, 5001–5010

6. Gong, C.-X., Liu, F., Grundke-Iqbal, I., and Iqbal, K. (2006) Dysregulation of Protein Phosphorylation/Dephosphorylation in Alzheimer's Disease: A Therapeutic Target. J. Biomed. Biotechnol. 2006, 1–11

7. Stowell, S. R., Ju, T., and Cummings, R. D. (2015) Protein Glycosylation in Cancer. Annu. Rev. Pathol. Mech. Dis. 10, 473–510

8. Kölbl, A. C., Andergassen, U., and Jeschke, U. (2015) The Role of Glycosylation in Breast Cancer Metastasis and Cancer Control. Front. Oncol. 5, 1–5

9. Yang, Y., and Bedford, M. T. (2012) Protein arginine methyltransferases and cancer. Nat. Rev. Cancer 13, 37–50

10. Hubbard, M. J., and Cohen, P. (1993) On target with a new mechanism for the regulation of protein phosphorylation. Trends Biochem. Sci., 172–177

11. Matheron, L., van den Toorn, H., Heck, A. J. R., and Mohammed, S. (2014) Characterization of Biases in Phosphopeptide Enrichment by Ti(4+)-Immobilized Metal Affinity Chromatography and TiO2 Using a Massive Synthetic Library and Human Cell Digests. Anal. Chem. 86, 8312–20

12. Molina, H., Horn, D. M., Tang, N., Mathivanan, S., and Pandey, A. (2007) Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. Proc. Natl. Acad. Sci. U. S. A. 104, 2199–2204
13. Di Palma, S., Zoumaro-Djayoon, A., Peng, M., Post, H., Preisinger, C., Munoz, J., and Heck, A. J. R. (2013) Finding the same needles in the haystack? A comparison of phosphotyrosine peptides enriched by immuno-affinity precipitation and metal-based affinity chromatography. J. Proteomics 91, 331–337
14. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) The protein kinase complement of the human genome. Science (80-.). 298, 1912–1934
15. Alonso, A., Sasin, J., Bottini, N., Friedberg, I., Friedberg, I., Osterman, A., Godzik, A., Hunter, T., Dixon, J., and Mustelin, T. (2004) Protein tyrosine phosphatases in the human genome. Cell 117, 699–711
16. Sacco, F., Perfetto, L., Castagnoli, L., and Cesareni, G. (2012) The human phosphatase interactome: An intricate family portrait. FEBS Lett. 586, 2732–9
17. Rini J, Esko J, V. A. (2009) in Essentials of Glycobiology
18. Lyons, J. J., Milner, J. D., and Rosenzweig, S. D. (2015) Glycans Instructing Immunity: The Emerging Role of Altered Glycosylation in Clinical Immunology. Front. Pediatr. 3,
19. Lannoo, N., and Van Damme, E. J. M. (2015) Review/N-glycans: The making of a varied toolbox. Plant Sci. 239, 67–83
20. Kornfeld, R., and Kornfeld, S. (1985) Rosalind Kornfeld and Stuart Kornfeld. Annu. Rev. Biochem. 54, 631–664
21. Introduction to glycobiology / Maureen E. Taylor, Kurt Drickamer.- Version details- Trove
22. Schjoldager, K. T.-B. G., and Clausen, H. (2012) Site-specific protein O-glycosylation modulates proprotein processing — Deciphering specific functions of the large polypeptide GalNAc-transferase gene family. Biochim. Biophys. Acta- Gen. Subj. 1820, 2079–2094
23. Brockhausen, I., Schachter, H., and Stanley, P. (2009) O-GalNAc Glycans.
24. Fukuda, M. (2006) Roles of Mucin-Type O-Glycans Synthesized by Core2??1,6-N-Acetylglucosaminyltransferase. Methods Enzymol. 416, 332–346
25. Bond, M. R., and Hanover, J. A. (2015) A little sugar goes a long way: The cell biology of O-GlcNAc. J. Cell Biol. 208, 869–880
26. Hardivill??, S., and Hart, G. W. (2014) Nutrient regulation of signaling, transcription, and cell physiology by O- GlcNAcylation. Cell Metab. 20, 208–213
27. Pek, J. W., Anand, A., and Kai, T. (2012) Tudor domain proteins in development. Development 139, 2255–66
28. Petrossian, T. C., and Clarke, S. G. (2011) Uncovering the human methyltransferasome. Mol. Cell. Proteomics 10, M110.000976
29. Carr, S. M., Poppy Roworth, A., Chan, C., and La Thangue, N. B. (2015) Post-translational control of transcription factors: methylation ranks highly. FEBS J. 282, n/a–n/a
30. Bedford, M. T., and Clarke, S. G. (2009) Protein arginine methylation in mammals: who, what, and why. Mol Cell 33, 1–13
31. Bedford, M. T. (2007) Arginine methylation at a glance. J. Cell Sci. 120, 4243–4246
32. Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P., and Zhang, Y. (2006) Histone demethylation by a family of JmjC domain-containing proteins. Nature 439, 811–816
33. Weinhold, B. (2006) Epigenetics: the science of change. Environ. Health Perspect. 114, A160–7
34. Zhang, X., Wen, H., and Shi, X. (2012) Lysine methylation: beyond histones. Acta Biochim Biophys Sin 44, 14–27
35. Betzen, C., Alhamdani, M. S. S., Lueong, S., Schr??der, C., Stang, A., and Hoheisel, J. D. (2015) Clinical proteomics: Promises, challenges and limitations of affinity arrays. Proteomics- Clin. Appl. 9, 342–347
36. Hu, B., Niu, X., Cheng, L., Yang, L.-N., Li, Q., Wang, Y., Tao, S.-C., and Zhou, S.-M. (2015) Discovering cancer biomarkers from clinical samples by protein microarrays. Proteomics. Clin. Appl. 9, 98–110
37. Politis, A., and Borysik, A. J. (2015) Assembling the pieces of macromolecular complexes: Hybrid structural biology approaches. Proteomics 15, 2792–2803
38. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C., and Yates, J. R. (2013) Protein analysis by shotgun/bottom-up proteomics. Chem Rev 113, 2343–2394
39. Moradian, A., Kalli, A., Sweredoski, M. J., and Hess, S. (2014) The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. Proteomics 14, 489–497
40. Huang, J., Wang, F., Ye, M., and Zou, H. (2014) Enrichment and separation techniques for large-scale proteomics analysis of the protein post-translational modifications. J. Chromatogr. A 1372C, 1–17
41. Mayne, J., Ning, Z., Zhang, X., Starr, A. E., Chen, R., Deeke, S., Chiang, C. K., Xu, B., Wen, M., Cheng, K., Seebun, D., Star, A., Moore, J. I., and Figeys, D. (2016) Bottom-Up Proteomics (2013-2015): Keeping up in the Era of Systems Biology. Anal. Chem. 88, 95–121
42. Olsen, J. V, Ong, S.-E., and Mann, M. (2004) Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues. Mol. Cell. Proteomics 3, 608–614

43. Swaney, D. L., Wenger, C. D., and Coon, J. J. (2010) NIH Public Access. Digestion 9, 1323–1329
44. Hillenkamp, F., and Karas, M. (1990) Mass Spectrometry of Peptides and Proteins by Matrix-Assisted Ultraviolet Laser Desorption/Ionization. Methods Enzym. 193, 280–295
45. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. Science (80-.)., 64–71
46. Li, K. Y., Tu, H., and Ray, A. K. (2005) Charge limits on droplets during evaporation. Langmuir 21, 3786–3794
47. Cech, N. B., and Enke, C. G. (2001) Practical implications of some recent studies in electrospray ionization fundamentals. Mass Spectrom. Rev. 20, 362–387
48. Schmidt, A., Karas, M., and Dülcks, T. (2003) Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI? J. Am. Soc. Mass Spectrom. 14, 492–500
49. Karas, M., Bahr, U., and Dülcks, T. (2000) Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. Fresenius. J. Anal. Chem. 366, 669–676
50. Annesley, T. M. (2003) Ion suppression in mass spectrometry. Clin. Chem. 49, 1041–1044
51. Chervet, J. P., Ursem, M., and Salzmann, J. P. (1996) Instrumental requirements for nanoscale liquid chromatography. Anal. Chem. 68, 1507–12
52. Köcher, T., Pichler, P., Swart, R., and Mechtler, K. (2012) Analysis of protein mixtures from whole-cell extracts by single-run nanoLC-MS/MS using ultralong gradients. Nat Protoc 7, 882–890
53. Cristobal, A., Hennrich, M. L., Giansanti, P., Goerdayal, S. S., Heck, A. J., and Mohammed, S. (2012) In-house construction of a UHPLC system enabling the identification of over 4000 protein groups in a single analysis. Analyst 137, 3541–3548
54. Pirmoradian, M., Budamgunta, H., Chingin, K., Zhang, B., Astorga-Wells, J., and Zubarev, R. A. (2013) Rapid and deep human proteome analysis by single-dimension shotgun proteomics. Mol Cell Proteomics,
55. Köcher, T., Swart, R., and Mechtler, K. (2011) Ultra-high-pressure RPLC hyphenated to an LTQ-Orbitrap Velos reveals a linear relation between peak capacity and number of identified peptides. Anal Chem 83, 2699–2704
56. Xiang, R., Horváth, C., and Wilkins, J. a (2003) Elution-modified displacement chromatography coupled with electrospray ionization-MS: on-line detection of trace peptides at low-femtomole level in peptide digests. Anal. Chem. 75, 1819–27
57. Ficarro, S. B., Zhang, Y., Lu, Y., Moghimi, A. R., Askenazi, M., Hyatt, E., Smith, E. D., Boyer, L., Schlaeger, T. M., Luckey, C. J., and Marto, J. a. (2009) Improved electrospray ionization efficiency compensates for diminished chromatographic resolution and enables proteomics analysis of tyrosine signaling in embryonic stem cells. Anal. Chem. 81, 3440–3447
58. Kostiainen, R., and Kauppila, T. J. (2009) Effect of eluent on the ionization process in liquid chromatography-mass spectrometry. J. Chromatogr. A 1216, 685–699
59. GARCIA, M. (2005) The effect of the mobile phase additives on sensitivity in the analysis of peptides and proteins by high-performance liquid chromatography–electrospray mass spectrometry. J. Chromatogr. B 825, 111–123
60. Jorgenson, J. W. (2010) Capillary liquid chromatography at ultrahigh pressures. Annu Rev Anal Chem (Palo Alto Calif) 3, 129–150
61. Wixom, R. L., and Gehrke, C. W. eds. (2010) Chromatography (John Wiley & Sons, Inc., Hoboken, NJ, USA)
62. MacNair, J. E., Patel, K. D., and Jorgenson, J. W. (1999) Ultrahigh-pressure reversed-phase capillary liquid chromatography: isocratic and gradient elution using columns packed with 1.0-micron particles. Anal Chem 71, 700–708
63. Shen, Y., Zhao, R., Berger, S. J., Anderson, G. A., Rodriguez, N., and Smith, R. D. (2002) High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics. Anal Chem 74, 4235–4249
64. Thakur, S. S., Geiger, T., Chatterjee, B., Bandilla, P., Fröhlich, F., Cox, J., and Mann, M. (2011) Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. Mol Cell Proteomics 10, M110.003699
65. Zhou, F., Lu, Y., Ficarro, S. B., Webber, J. T., and Marto, J. A. (2012) Nanoflow low pressure high peak capacity single dimension LC-MS/MS platform for high-throughput, in-depth analysis of mammalian proteomes. Anal Chem 84, 5133–5139
66. Yamana, R., Iwasaki, M., Wakabayashi, M., Nakagawa, M., Yamanaka, S., and Ishihama, Y. (2013) Rapid and deep profiling of human induced pluripotent stem cell proteome by one-shot NanoLC-MS/MS analysis with meter-scale monolithic silica columns. J Proteome Res 12, 214–221
67. Iwasaki, M., Sugiyama, N., Tanaka, N., and Ishihama, Y. (2012) Human proteome analysis by using reversed phase monolithic silica capillary columns with enhanced sensitivity. J Chromatogr A 1228, 292–297
68. Shen, Y., Zhang, R., Moore, R. J., Kim, J., Metz, T. O., Hixson, K. K., Zhao, R., Livesay, E. A., Udseth, H. R., and Smith, R. D. (2005) Automated 20 kpsi RPLC-MS and MS/MS with chromatographic peak capacities of 1000-1500 and capabilities in proteomics and metabolomics. Anal Chem 77, 3090–3100
69. Iwasaki, M., Sugiyama, N., Tanaka, N., and Ishihama, Y. (2012) Human proteome analysis by using reversed

phase monolithic silica capillary columns with enhanced sensitivity. J Chromatogr A 1228, 292–297
70. Miyamoto, K., Hara, T., Kobayashi, H., Morisaka, H., Tokuda, D., Horie, K., Koduki, K., Makino, S., Núñez, O., Yang, C., Kawabe, T., Ikegami, T., Takubo, H., Ishihama, Y., and Tanaka, N. (2008) High-efficiency liquid chromatographic separation utilizing long monolithic silica capillary columns. Anal. Chem. 80, 8741–8750
71. Hebert, A. S., Richards, A. L., Bailey, D. J., Ulbrich, A., Coughlin, E. E., Westphall, M. S., and Coon, J. J. (2014) The one hour yeast proteome. Mol Cell Proteomics 13, 339–347
72. Gilar, M., Fridrich, J., Schure, M. R., and Jaworski, A. (2012) Comparison of orthogonality estimation methods for the two-dimensional separations of peptides. Anal Chem 84, 8722–8732
73. Gilar, M., Olivova, P., Daly, A. E., and Gebler, J. C. (2005) Orthogonality of separation in two-dimensional liquid chromatography. Anal Chem 77, 6426–6434
74. Washburn, M. P., Wolters, D., and Yates, J. R. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol 19, 242–247
75. Di Palma, S., Hennrich, M. L., Heck, A. J., and Mohammed, S. (2012) Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. J Proteomics 75, 3791–3813
76. Fournier, M. L., Gilmore, J. M., Martin-Brown, S. A., and Washburn, M. P. (2007) Multidimensional separations-based shotgun proteomics. Chem Rev 107, 3654–3686
77. Vollmer, M., Hörth, P., and Nägele, E. (2004) Optimization of two-dimensional off-line LC/MS separations to improve resolution of complex proteomic samples. Anal. Chem. 76, 5180–5
78. Washburn, M. P., Wolters, D., and Yates, J. R. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol 19, 242–247
79. Isobe, T., Takayasu, T., Takai, N., and Okuyama, T. (1982) High-performance liquid chromatography of peptides on a macroreticular cation-exchange resin: Application to peptide mapping of Bence-Jones proteins. Anal. Biochem. 122, 417–425
80. Mant, C. T., and Hodges, R. S. (1985) Separation of peptides by strong cation-exchange high-performance liquid chromatography. J Chromatogr 327, 147–155
81. Alpert, A. J., and Andrews, P. C. (1988) Cation-exchange chromatography of peptides on poly(2-sulfoethyl aspartamide)-silica. J Chromatogr 443, 85–96
82. Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villén, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci U S A 101, 12130–12135
83. McDonald, W. H., Ohi, R., Miyamoto, D. T., Mitchison, T. J., and Yates, J. R. (2002) Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. Int. J. Mass Spectrom. 219, 245–251
84. Mohammed, S., and Heck, A. (2011) Strong cation exchange (SCX) based analytical methods for the targeted analysis of protein post-translational modifications. Curr Opin Biotechnol 22, 9–16
85. Alpert, A. J., Petritis, K., Kangas, L., Smith, R. D., Mechtler, K., Mitulović, G., Mohammed, S., and Heck, A. J. (2010) Peptide orientation affects selectivity in ion-exchange chromatography. Anal Chem 82, 5253–5259
86. Uhlmann, T., Geoghegan, V. L., Thomas, B., Ridlova, G., Trudgian, D. C., and Acuto, O. (2012) A method for large-scale identification of protein arginine methylation. Mol Cell Proteomics 11, 1489–1499
87. Bedford, M. T., Frankel, A., Yaffe, M. B., Clarke, S., Leder, P., and Richard, S. (2000) Arginine methylation inhibits the binding of proline-rich ligands to Src homology 3, but not WW, domains. J Biol Chem 275, 16030–16036
88. Smith, J. J., Rücknagel, K. P., Schierhorn, A., Tang, J., Nemeth, A., Linder, M., Herschman, H. R., and Wahle, E. (1999) Unusual sites of arginine methylation in Poly(A)-binding protein II and in vitro methylation by protein arginine methyltransferases PRMT1 and PRMT3. J Biol Chem 274, 13229–13234
89. Guo, A., Gu, H., Zhou, J., Mulhern, D., Wang, Y., Lee, K. A., Yang, V., Aguiar, M., Kornhauser, J., Jia, X., Ren, J., Beausoleil, S. A., Silva, J. C., Vemulapalli, V., Bedford, M. T., and Comb, M. J. (2014) Immunoaffinity enrichment and mass spectrometry analysis of protein methylation. Mol Cell Proteomics 13, 372–387
90. Ritorto, M. S., Cook, K., Tyagi, K., Pedrioli, P. G. A., and Trost, M. (2013) Hydrophilic strong anion exchange (hSAX) chromatography for highly orthogonal peptide separation of complex proteomes. J. Proteome Res. 12, 2449–2457
91. Alpert, A. J., Hudecz, O., and Mechtler, K. (2015) Anion-Exchange Chromatography of Phosphopeptides: Weak Anion Exchange versus Strong Anion Exchange and Anion-Exchange Chromatography versus Electrostatic Repulsion–Hydrophilic Interaction Chromatography. Anal. Chem.,
92. Han, G., Ye, M., Zhou, H., Jiang, X., Feng, S., Jiang, X., Tian, R., Wan, D., Zou, H., and Gu, J. (2008) Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography. Proteomics 8, 1346–1361
93. Guo, a., Gu, H., Zhou, J., Mulhern, D., Wang, Y., Lee, K. a., Yang, V., Aguiar, M., Kornhauser, J., Jia, X., Ren, J., Beausoleil, S. a., Silva, J. C., Vemulapalli, V., Bedford, M. T., and Comb, M. J. (2014) Immunoaffinity Enrichment and Mass Spectrometry Analysis of Protein Methylation. Mol. Cell. Proteomics 13, 372–387

94. Deguchi, K., Keira, T., Yamada, K., Ito, H., Takegawa, Y., Nakagawa, H., and Nishimura, S. I. (2008) Two-dimensional hydrophilic interaction chromatography coupling anion-exchange and hydrophilic interaction columns for separation of 2-pyridylamino derivatives of neutral and sialylated N-glycans. J. Chromatogr. A 1189, 169–174
95. Di Palma, S., Mohammed, S., and Heck, A. J. (2012) ZIC-cHILIC as a fractionation method for sensitive and powerful shotgun proteomics. Nat Protoc 7, 2041–2055
96. Boersema, P. J., Mohammed, S., and Heck, A. J. R. (2008) Hydrophilic interaction liquid chromatography (HILIC) in proteomics. Anal. Bioanal. Chem. 391, 151–159
97. Boersema, P. J., Divecha, N., Heck, A. J. R., and Mohammed, S. (2007) Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. J. Proteome Res. 6, 937–946
98. Zhu, J., Wang, F., Chen, R., Cheng, K., Xu, B., Guo, Z., Liang, X., Ye, M., and Zou, H. (2012) Centrifugation assisted microreactor enables facile integration of trypsin digestion, hydrophilic interaction chromatography enrichment, and on-column deglycosylation for rapid and sensitive N-glycoproteome analysis. Anal. Chem. 84, 5146–5153
99. Sugahara, D., Kaji, H., Sugihara, K., Asano, M., and Narimatsu, H. (2012) Large-scale identification of target proteins of a glycosyltransferase isozyme by Lectin-IGOT-LC/MS, an LC/MS-based glycoproteomic approach. Sci. Rep. 2, 680
100. McNulty, D. E., and Annan, R. S. (2008) Hydrophilic Interaction Chromatography Reduces the Complexity of the Phosphoproteome and Improves Global Phosphopeptide Isolation and Detection. Mol. Cell. Proteomics 7, 971–980
101. Snijders, A. P., Hung, M. L., Wilson, S. A., and Dickman, M. J. (2010) Analysis of arginine and lysine methylation utilizing peptide separations at neutral pH and electron transfer dissociation mass spectrometry. J Am Soc Mass Spectrom 21, 88–96
102. Zhou, H., Di Palma, S., Preisinger, C., Peng, M., Polat, A. N., Heck, A. J., and Mohammed, S. (2012) Towards a comprehensive characterization of a human cancer cell phosphoproteome. J Proteome Res,
103. Hao, P., Qian, J., Ren, Y., and Sze, S. K. (2011) Electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) versus strong cation exchange (SCX) for fractionation of iTRAQ-labeled peptides. J. Proteome Res. 10, 5568–5574
104. Hao, P., Guo, T., Li, X., Adav, S. S., Yang, J., Wei, M., and Sze, S. K. (2010) Novel application of electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) in shotgun proteomics: Comprehensive profiling of rat kidney proteome. J. Proteome Res. 9, 3520–3526
105. Alpert, A. J. (2008) Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. Anal. Chem. 80, 62–76
106. Sun, R.-X., Dong, M.-Q., Song, C.-Q., Chi, H., Yang, B., Xiu, L.-Y., Tao, L., Jing, Z.-Y., Liu, C., Wang, L.-H., Fu, Y., and He, S.-M. (2010) Improved peptide identification for proteomic analysis based on comprehensive characterization of electron transfer dissociation spectra. J. Proteome Res. 9, 6354–67
107. Song, C., Ye, M., Han, G., Jiang, X., Wang, F., Yu, Z., Chen, R., and Zou, H. (2010) Reversed-phase-reversed-phase liquid chromatography approach with high orthogonality for multidimensional separation of phosphopeptides. Anal. Chem. 82, 53–56
108. Dwivedi, R. C., Spicer, V., Harder, M., Antonovici, M., Ens, W., Standing, K. G., Wilkins, J. A., and Krokhin, O. V. (2008) Practical implementation of 2D HPLC scheme with accurate peptide retention prediction in both dimensions for high-throughput bottom-up proteomics. Anal. Chem. 80, 7036–7042
109. Li, X., Gerber, S. a, Rudner, A. D., Beausoleil, S. a, Haas, W., Villén, J., Elias, J. E., and Gygi, S. P. (2007) Large-scale phosphorylation analysis of alpha-factor-arrested Saccharomyces cerevisiae. J. Proteome Res. 6, 1190–1197
110. Posewitz, M. C., and Tempst, P. (1999) Immobilized Gallium (III) Affinity Chromatography of Phosphopeptides peptides , as a front end to mass spectrometric analysis , the use of an immobilized metal affinity chromatography. Biosystems 71, 29520–29529
111. Villén, J., and Gygi, S. P. (2008) The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat. Protoc. 3, 1630–8
112. Pinkse, M. W. H., Uitto, P. M., Hilhorst, M. J., Ooms, B., and Heck, A. J. R. (2004) Selective Isolation at the Femtomole Level of Phosphopeptides from Proteolytic Digests Using 2D-NanoLC-ESI-MS / MS and Titanium Oxide Precolumns Selective Isolation at the Femtomole Level of Phosphopeptides from Proteolytic Digests Using 2D-NanoLC-ESI-MS . Anal. Chem. 76, 3935–3943
113. Kweon, H. K., and Håkansson, K. (2006) Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. Anal. Chem. 78, 1743–9
114. Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F., and White, F. M. (2002) Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. Nat. Biotechnol. 20, 301–5
115. Kokubu, M., Ishihama, Y., Sato, T., Nagasu, T., and Oda, Y. (2005) Specificity of immobilized metal affinity-based IMAC/C18 tip enrichment of phosphopeptides for protein phosphorylation analysis. Anal. Chem. 77, 5144–54
116. Trinidad, J. C., Specht, C. G., Thalhammer, A., Schoepfer, R., and Burlingame, A. L. (2006) Comprehensive

identification of phosphorylation sites in postsynaptic density preparations. Mol. Cell. Proteomics 5, 914–22
117. Larsen, M. R. (2005) Highly Selective Enrichment of Phosphorylated Peptides from Peptide Mixtures Using Titanium Dioxide Microcolumns. Mol. Cell. Proteomics 4, 873–886
118. Zhou, H., Ye, M., Dong, J., Corradini, E., Cristobal, A., Heck, A. J., Zou, H., and Mohammed, S. (2013) Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. Nat Protoc 8, 461–480
119. Larsen, M. R., Jensen, S. S., Jakobsen, L. a, and Heegaard, N. H. H. (2007) Exploring the sialiome using titanium dioxide chromatography and mass spectrometry. Mol. Cell. Proteomics 6, 1778–87
120. Artemenko, K. A., Lind, S. B., Elfineh, L., Mayrhofer, C., Zubarev, R. A., Bergquist, J., and Pettersson, U. (2011) Optimization of immunoaffinity enrichment and detection: toward a comprehensive characterization of the phosphotyrosine proteome of K562 cells by liquid chromatography-mass spectrometry. Analyst 136, 1971
121. Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V, and Mann, M. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science (80-.). 325, 834–840
122. Rush, J., Moritz, A., Lee, K. A., Guo, A., Goss, V. L., Spek, E. J., Zhang, H., Zha, X. M., Polakiewicz, R. D., and Comb, M. J. (2005) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. Nat Biotechnol 23, 94–101
123. Zhang, H., Zha, X., Tan, Y., Hornbeck, P. V., Mastrangelo, A. J., Alessi, D. R., Polakiewicz, R. D., and Comb, M. J. (2002) Phosphoprotein analysis using antibodies broadly reactive against phosphorylated motifs. J. Biol. Chem. 277, 39379–39387
124. Of, I., and Protein, A. N. (2002) A Mass Spectrometry-based Proteomic Approach for Identification of Serine/ Threonine-phosphorylated Proteins by Enrichment with Phospho-specific Antibodies. Mol. Cell. Proteomics, 517–527
125. Andrews, G. L., Simons, B. L., Young, J. B., Hawkridge, A. M., and Muddiman, D. C. (2011) Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). Anal Chem 83, 5442–5446
126. Michalski, A., Damoc, E., Hauschild, J. P., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., and Horning, S. (2011) Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. Mol Cell Proteomics 10, M111.011015
127. Zubarev, R., and Mann, M. (2007) On the proper use of mass accuracy in proteomics. Mol. Cell. Proteomics 6, 377–81
128. Mann, M., and Kelleher, N. L. (2008) Precision proteomics: the case for high resolution and high mass accuracy. Proc Natl Acad Sci U S A 105, 18132–18138
129. Henion, J., Wachs, T., and Mordehai, a (1993) Recent developments in electrospray mass spectrometry including implementation on an ion trap. J. Pharm. Biomed. Anal. 11, 1049–61
130. Wang, H., Lim, K. B., Lawrence, R. F., Howald, W. N., Taylor, J. A., Ericsson, L. H., Walsh, K. A., and Hackett, M. (1997) Stability enhancement for peptide analysis by electrospray using the triple quadrupole mass spectrometer. Anal. Biochem. 250, 162–168
131. Morris, H. R., Paxton, T., Panico, M., McDowell, R., and Dell, A. (1997) A novel geometry mass spectrometer, the Q-TOF, for low- femtomole/attomole-range biopolymer sequencing. J. Protein Chem. 16, 469–479
132. Eliuk, S., and Makarov, A. (2015) Evolution of Orbitrap Mass Spectrometry Instrumentation. Annu. Rev. Anal. Chem. 8, 61–80
133. Chen, J., Canales, L., and Neal, R. E. (2011) Multi-Segment Direct Inject nano-ESI-LTQ-FT-ICR-MS/MS For Protein Identification. Proteome Sci 9, 38
134. Andrews, G. L., Simons, B. L., Young, J. B., Hawkridge, A. M., and Muddiman, D. C. (2011) Performance Characteristics of a New Hybrid Triple Quadrupole Time-of-Flight Tandem Mass Spectrometry. Anal. Chem. 83, 5442–5446
135. Thelen, J. J., and Miernyk, J. A. (2012) The proteomic future: where mass spectrometry should be taking us. Biochem. J. 444, 169–181
136. Douglas, D. J., Frank, A. J., and Mao, D. (2005) Linear ion traps in mass spectrometry. Mass Spectrom. Rev. 24, 1–29
137. March, R. E. (2000) Quadrupole ion trap mass spectrometry: a view at the turn of the century. Int. J. Mass Spectrom. 200, 285–312
138. Schwartz, J. C., Senko, M. W., and Syka, J. E. P. (2002) A two-dimensional quadrupole ion trap mass spectrometer. J. Am. Soc. Mass Spectrom. 13, 659–669
139. Makarov, A. (2000) Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. Anal. Chem. 72, 1156–1162
140. Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K., and Horning, S. (2006) Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. Anal. Chem. 78, 2113–2120
141. Yates, J. R., Ruse, C. I., and Nakorchevsky, A. (2009) Proteomics by mass spectrometry: approaches, advances,

and applications. Annu Rev Biomed Eng 11, 49–79
142. Paizs, B., and Suhai, S. (2005) Fragmentation pathways of protonated peptides. Mass Spectrom. Rev. 24, 508–548
143. Roepstorff P, F. J. (1984) No TitleProposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed. Mass Spectrom.,
144. Biemann, K. (1990) Nomenclature for Peptide Fragment Ions (Positive Ions)*. Methods Enzymol. 193, 886–887
145. Tabb, D. L., Huang, Y., Wysocki, V. H., and Yates, J. R. (2004) Influence of basic residue content on fragment ion peak intensities in low-energy collision-induced dissociation spectra of peptides. Anal Chem 76, 1243–1248
146. Fabris, D., Kelly, M., Murphy, C., Wu, Z., and Fenselau, C. (1993) High-energy collision-induced dissociation of multiply charged polypeptides produced by electrospray. J. Am. Soc. Mass Spectrom. 4, 652–661
147. McLuckey, S. A., and Goeringer, D. E. (1997) Slow heating methods in tandem mass spectrometry. J. Mass Spectrom. 32, 461–474
148. Cunningham, C., Glish, G. L., and Burinsky, D. J. (2006) High amplitude short time excitation: a method to form and detect low mass product ions in a quadrupole ion trap mass spectrometer. J. Am. Soc. Mass Spectrom. 17, 81–4
149. Frese, C. K., Boender, A. J., Mohammed, S., Heck, A. J., Adan, R. A., and Altelaar, A. F. (2013) Profiling of diet-induced neuropeptide changes in rat brain by quantitative mass spectrometry. Anal Chem 85, 4594–4604
150. Mommen, G. P. M., Frese, C. K., Meiring, H. D., van Gaans-van den Brink, J., de Jong, A. P. J. M., van Els, C. a C. M., and Heck, A. J. R. (2014) Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proc. Natl. Acad. Sci. U. S. A. 111, 4507–12
151. Kim, S., Mischerikow, N., Bandeira, N., Navarro, J. D., Wich, L., Mohammed, S., Heck, A. J. R., and Pevzner, P. A. (2010) The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search. Mol. Cell. Proteomics 9, 2840–52
152. Boersema, P. J., Mohammed, S., and Heck, A. J. R. (2009) Phosphopeptide fragmentation and analysis by mass spectrometry. J. Mass Spectrom. 44, 861–878
153. Zubarev, R. A. (2004) Electron-capture dissociation tandem mass spectrometry. Curr. Opin. Biotechnol. 15, 12–16
154. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A 101, 9528–9533
155. Zubarev, R. a, Horn, D. M., Fridriksson, E. K., Kelleher, N. L., Kruger, N. a, Lewis, M. a, Carpenter, B. K., and McLafferty, F. W. (2000) Electron capture dissociation for structural characterization of multiply charged protein cations. Anal. Chem. 72, 563–573
156. Xia, Y., Chrisman, P. A., Erickson, D. E., Liu, J., Liang, X., Londry, F. A., Yang, M. J., and McLuckey, S. A. (2006) Implementation of ion/ion reactions in a quadrupole/time-of-flight tandem mass spectrometer. Anal. Chem. 78, 4146–54
157. McAlister, G. C., Phanstiel, D., Good, D. M., Berggren, W. T., and Coon, J. J. (2007) Implementation of electron-transfer dissociation on a hybrid linear ion trap-orbitrap mass spectrometer. Anal. Chem. 79, 3525–34
158. Coon, J. (2009) NIH Public Access. 79, 3525–3534
159. O'connor, P. B., Lin, C., Cournoyer, J. J., Pittman, J. L., Belyayev, M., and Budnik, B. A. (2006) Long-lived electron capture dissociation product ions experience radical migration via hydrogen abstraction. J. Am. Soc. Mass Spectrom. 17, 576–585
160. Ledvina, A. R., McAlister, G. C., Gardner, M. W., Smith, S. I., Madsen, J. A., Schwartz, J. C., Jr., G. C. S., Syka, J. E. P., Brodbelt, J. S., and Coon, J. J. (2009) Infrared Photoactivation Reduces Peptide Folding and Hydrogen-Atom Migration following ETD Tandem Mass Spectrometry. Angew. Chemie Int. Ed., 8526–8528
161. Pitteri, S. J., Chrisman, P. A., and McLuckey, S. A. (2005) Electron-transfer ion/ion reactions of doubly protonated peptides: effect of elevated bath gas temperature. Anal. Chem. 77, 5662–5669
162. Horn, D. M., Breuker, K., Frank, a. J., and McLafferty, F. W. (2001) Kinetic intermediates in the folding of gaseous protein ions characterized by electron capture dissociation mass spectrometry. J. Am. Chem. Soc. 123, 9792–9799
163. Swaney, D. L., McAlister, G. C., Wirtala, M., Schwartz, J. C., Syka, J. E. P., and Coon, J. J. (2007) Supplemental Activation Method for High-Efficiency Electron-Transfer Dissociation of Doubly Protonated Peptide Precursors. Anal. Chem. 79, 477–485
164. Frese, C. K., Altelaar, A. F., van den Toorn, H., Nolting, D., Griep-Raming, J., Heck, A. J., and Mohammed, S. (2012) Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. Anal Chem 84, 9668–9673
165. Good, D. M., Wirtala, M., McAlister, G. C., and Coon, J. J. (2007) Performance characteristics of electron transfer dissociation mass spectrometry. Mol Cell Proteomics 6, 1942–1951
166. Frese, C. K., Altelaar, A. F., Hennrich, M. L., Nolting, D., Zeller, M., Griep-Raming, J., Heck, A. J., and Moham-

med, S. (2011) Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. J Proteome Res 10, 2377–2388
167. Coon, J. J. (2009) Collision or Electron? Protein Sequence Analysis on the 21st Century. Anal. Chem. 81, 3208–3215
168. Wang, H., Straubinger, R. M., Aletta, J. M., Cao, J., Duan, X., Yu, H., and Qu, J. (2009) Accurate localization and relative quantification of arginine methylation using nanoflow liquid chromatography coupled to electron transfer dissociation and orbitrap mass spectrometry. J Am Soc Mass Spectrom 20, 507–519
169. Mischerikow, N., Altelaar, a F. M., Navarro, J. D., Mohammed, S., and Heck, A. J. R. (2010) Comparative assessment of site assignments in CID and electron transfer dissociation spectra of phosphopeptides discloses limited relocation of phosphate groups. Mol. Cell. Proteomics 9, 2140–2148
170. Witze, E. (2007) Mapping Protein Post-Translational Modifcications With Mass Spectrometry. Nat. Methods 4, 798–806
171. Olsen, J. V, and Mann, M. (2013) Status of large-scale analysis of post-translational modifications by mass spectrometry. Mol Cell Proteomics,
172. Elviri, L. (2012) ETD and ECD Mass Spectrometry Fragmentation for the Characterization of Protein Post Translational Modifications. Tandem Mass Spectrom.- Appl. Princ., 163–178
173. Guthals, A., and Bandeira, N. (2012) Peptide identification by tandem mass spectrometry with alternate fragmentation modes. Mol Cell Proteomics 11, 550–557
174. Manuscript, A., Free, T., Landscapes, E., and Conformational, G. (2008) NIH Public Access. 15, 1203–1214
175. Sarbu, M., Ghiulai, R. M., and Zamfir, A. D. (2014) Recent developments and applications of electron transfer dissociation mass spectrometry in proteomics. Amino Acids 46, 1625–1634
176. Frese, C. K., Altelaar, A. F., Hennrich, M. L., Nolting, D., Zeller, M., Griep-Raming, J., Heck, A. J., and Mohammed, S. (2011) Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. J Proteome Res 10, 2377–2388
177. Zhurov, K. O., Fornelli, L., Wodrich, M. D., Laskay, Ü. a, and Tsybin, Y. O. (2013) Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. Chem. Soc. Rev. 42, 5014–30
178. C. Stoermer, D. A. K. R. H. M. L. O. R. M. A. P. J. F. (2010) Electron Transfer Dissociation on Small Intact Proteins in an Ultra High Resolution Quadrupole Time of Flight Mass Spectrometer. J. Biomol. Tech. 21, S38
179. Earley, L., Anderson, L. C., Bai, D. L., Mullen, C., Syka, J. E. P., English, A. M., Dunyach, J.-J., Stafford, G. C., Shabanowitz, J., Hunt, D. F., and Compton, P. D. (2013) Front-end electron transfer dissociation: a new ionization source. Anal. Chem. 85, 8385–90
180. Rose, C. M., Rush, M. J. P., Riley, N. M., Merrill, A. E., Kwiecien, N. W., Holden, D. D., Mullen, C., Westphall, M. S., and Coon, J. J. (2015) A Calibration Routine for Efficient ETD in Large-Scale Proteomics. J. Am. Soc. Mass Spectrom., doi: 10.1007/s13361–015–1183–1
181. Eng, J. K., McCormack, A. L., and Yates, J. R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 5, 976–89
182. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551–67
183. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V, and Mann, M. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. 10, 1794–805
184. Eng, J. K., Searle, B. C., Clauser, K. R., and Tabb, D. L. (2011) A face in the crowd: recognizing peptides through database search. Mol. Cell. Proteomics 10, R111.009522
185. Elias, J. E., and Gygi, S. P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat. Methods 4, 207–14
186. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., and MacCoss, M. J. (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods 4, 923–925
187. Savitski, M. M., Lemeer, S., Boesche, M., Lang, M., Mathieson, T., Bantscheff, M., and Kuster, B. (2011) Confident phosphorylation site localization using the Mascot Delta Score. Mol. Cell. Proteomics 10, M110.003830
188. Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C., and Mechtler, K. (2011) Universal and confident phosphorylation site localization using phosphoRS. J Proteome Res 10, 5354–5362
189. Beausoleil, S. A., Villén, J., Gerber, S. A., Rush, J., and Gygi, S. P. (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nat. Biotechnol. 24, 1285–92
190. Chalkley, R. J., and Clauser, K. R. (2012) Modification site localization scoring: strategies and performance. Mol. Cell. Proteomics 11, 3–14
191. Takeuchi, O., and Akira, S. (2010) Pattern recognition receptors and inflammation. Cell 140, 805–20
192. Mester, G., Hoffmann, V., and Stevanović, S. (2011) Insights into MHC class I antigen processing gained from large-scale analysis of class I ligands. Cell. Mol. Life Sci. 68, 1521–1532
193. Falk, K., Rötzschke, O., Stevanović, S., Jung, G., and Rammensee, H. G. (1994) Pool sequencing of natural

HLA-DR, DQ, and DP ligands reveals detailed peptide motifs, constraints of processing, and general rules. Immunogenetics 39, 230–242
194. Li, Y., Wang, L. X., Yang, G., Hao, F., Urba, W. J., and Hu, H. M. (2008) Efficient cross-presentation depends on autophagy in tumor cells. Cancer Res 68, 6889–6895
195. Dengjel, J., Schoor, O., Fischer, R., Reich, M., Kraus, M., Müller, M., Kreymborg, K., Altenberend, F., Brandenburg, J., Kalbacher, H., Brock, R., Driessen, C., Rammensee, H.-G., and Stevanovic, S. (2005) Autophagy promotes MHC class II presentation of peptides from intracellular source proteins. Proc. Natl. Acad. Sci. U. S. A. 102, 7922–7927
196. Spurkland, A., Sollid, L. M., Polanco, I., Vartdal, F., and Thorsby, E. (1992) HLA-DR and-DQ genotypes of celiac disease patients serologically typed to be non-DR3 or non-DR5/7. Hum. Immunol. 35, 188–92
197. Rao, X., Hoof, I., Costa, A. I. C. A. F., van Baarle, D., and Keşmir, C. (2011) HLA class I allele promiscuity revisited. Immunogenetics 63, 691–701
198. Sidney, J., del Guercio, M. F., Southwood, S., Engelhard, V. H., Appella, E., Rammensee, H. G., Falk, K., Rötzschke, O., Takiguchi, M., and Kubo, R. T. (1995) Several HLA alleles share overlapping peptide specificities. J. Immunol. 154, 247–59
199. Diedrich, G., Bangia, N., Pan, M., and Cresswell, P. (2001) A role for calnexin in the assembly of the MHC class I loading complex in the endoplasmic reticulum. J. Immunol. 166, 1703–9
200. Belich, M. P., and Trowsdale, J. (1995) Proteasome and class I antigen processing and presentation. Mol. Biol. Rep. 21, 53–6
201. Holland, C. J., Cole, D. K., and Godkin, A. (2013) Re-Directing CD4(+) T Cell Responses with the Flanking Residues of MHC Class II-Bound Peptides: The Core is Not Enough. Front Immunol 4, 172
202. Blees, A., Reichel, K., Trowitzsch, S., Fisette, O., Bock, C., Abele, R., Hummer, G., Schäfer, L. V, and Tampé, R. (2015) Assembly of the MHC I peptide-loading complex determined by a conserved ionic lock-switch. Sci. Rep. 5, 17341
203. Kanaseki, T., and Shastri, N. (2008) Endoplasmic reticulum aminopeptidase associated with antigen processing regulates quality of processed peptides presented by MHC class I molecules. J. Immunol. 181, 6275–82
204. Stevanović, S. (2002) Structural basis of immunogenicity. Transpl. Immunol. 10, 133–6
205. McMichael, A., and Hanke, T. (2002) The quest for an AIDS vaccine: is the CD8+ T-cell approach feasible? Nat. Publ. Gr. 2, 283–291
206. Salamon, H., Klitz, W., Easteal, S., Gao, X., Erlich, H. A., Fernandez-Viña, M., Trachtenberg, E. A., McWeeney, S. K., Nelson, M. P., and Thomson, G. (1999) Evolution of HLA class II molecules: Allelic and amino acid site variability across populations. Genetics 152, 393–400
207. Pathak, S. S., and Blum, J. S. (2000) Endocytic recycling is required for the presentation of an exogenous peptide via MHC class II molecules. Traffic 1, 561–9
208. Neumann, J., and Koch, N. (2005) Assembly of major histocompatibility complex class II subunits with invariant chain. FEBS Lett. 579, 6055–9
209. Calafat, J., Nijenhuis, M., Janssen, H., Tulp, A., Dusseljee, S., Wubbolts, R., and Neefjes, J. (1994) Major histocompatibility complex class II molecules induce the formation of endocytic MIIC-like structures. J. Cell Biol. 126, 967–77
210. Riese, R. J., Mitchell, R. N., Villadangos, J. A., Shi, G. P., Palmer, J. T., Karp, E. R., De Sanctis, G. T., Ploegh, H. L., and Chapman, H. A. (1998) Cathepsin S activity regulates antigen presentation and immunity. J. Clin. Invest. 101, 2351–63
211. Honey, K., Nakagawa, T., Peters, C., and Rudensky, A. (2002) Cathepsin L regulates CD4+ T cell selection independently of its effect on invariant chain: a role in the generation of positively selecting peptide ligands. J. Exp. Med. 195, 1349–58
212. Schulze, M.-S. E. D., and Wucherpfennig, K. W. (2012) The mechanism of HLA-DM induced peptide exchange in the MHC class II antigen presentation pathway. Curr. Opin. Immunol. 24, 105–11
213. Cole, D. K., Gallagher, K., Lemercier, B., Holland, C. J., Junaid, S., Hindley, J. P., Wynn, K. K., Gostick, E., Sewell, A. K., Gallimore, A. M., Ladell, K., Price, D. A., Gougeon, M. L., and Godkin, A. (2012) Modification of the carboxy-terminal flanking region of a universal influenza epitope alters CD4+ T-cell repertoire selection. Nat Commun 3, 665
214. Neefjes, J., Jongsma, M. L., Paul, P., and Bakke, O. (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol 11, 823–836
215. Hillen, N., and Stevanovic, S. (2006) Contribution of mass spectrometry-based proteomics to immunology. Expert Rev Proteomics 3, 653–664
216. Bassani-Sternberg, M., Barnea, E., Beer, I., Avivi, I., Katz, T., and Admon, A. (2010) Soluble plasma HLA peptidome as a potential source for cancer biomarkers. Proc. Natl. Acad. Sci. U. S. A. 107, 18769–18776
217. Bozzacco, L., Yu, H., Zebroski, H. A., Dengjel, J., Deng, H., Mojsov, S., and Steinman, R. M. (2011) Mass spectrometry analysis and quantitation of peptides presented on the MHC II molecules of mouse spleen dendritic

cells. J Proteome Res 10, 5016–5030
218. Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L. J., and Mann, M. (2015) Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. Mol Cell Proteomics 14, 658–673
219. Escobar, H., Reyes-Vargas, E., Jensen, P. E., Delgado, J. C., and Crockett, D. K. (2011) Utility of characteristic QTOF MS/MS fragmentation for MHC class i peptides. J. Proteome Res. 10, 2494–2507
220. Adamopoulou, E., Tenzer, S., Hillen, N., Klug, P., Rota, I. a, Tietz, S., Gebhardt, M., Stevanovic, S., Schild, H., Tolosa, E., Melms, A., and Stoeckle, C. (2013) Exploring the MHC-peptide matrix of central tolerance in the human thymus. Nat. Commun. 4, 2039
221. Fissolo, N., Haag, S., de Graaf, K. L., Drews, O., Stevanovic, S., Rammensee, H. G., and Weissert, R. (2009) Naturally presented peptides on major histocompatibility complex I and II molecules eluted from central nervous system of multiple sclerosis patients. Mol Cell Proteomics 8, 2090–2101
222. Hassan, C., Kester, M. G. D., de Ru, A. H., Hombrink, P., Drijfhout, J. W., Nijveen, H., Leunissen, J. a M., Heemskerk, M. H. M., Falkenburg, J. H. F., and van Veelen, P. a (2013) The human leukocyte antigen-presented ligandome of B lymphocytes. Mol. Cell. Proteomics 12, 1829–43
223. Zamvil, S. S., Mitchell, D. J., Moore, A. C., Kitamura, K., Steinman, L., and Rothbard, J. B. (1986) T-cell epitope of the autoantigen myelin basic protein that induces encephalomyelitis. Nature 324, 258–260
224. Engelhard, V. H., Altrich-Vanlith, M., Ostankovitch, M., and Zarling, A. L. (2006) Post-translational modifications of naturally processed MHC-binding epitopes. Curr. Opin. Immunol. 18, 92–97
225. Haurum, J. S., Høier, I. B., Arsequell, G., Neisig, a, Valencia, G., Zeuthen, J., Neefjes, J., and Elliott, T. (1999) Presentation of cytosolic glycosylated peptides by human class I major histocompatibility complex molecules in vivo. J. Exp. Med. 190, 145–150
226. Zarling, A. L., Polefrone, J. M., Evans, A. M., Mikesh, L. M., Shabanowitz, J., Lewis, S. T., Engelhard, V. H., and Hunt, D. F. (2006) Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. Proc. Natl. Acad. Sci. U. S. A. 103, 14889–14894
227. Skipper, J. C., Hendrickson, R. C., Gulden, P. H., Brichard, V., Van Pel, A., Chen, Y., Shabanowitz, J., Wolfel, T., Slingluff, C. L., Boon, T., Hunt, D. F., and Engelhard, V. H. (1996) 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. 183, 527–34
228. Jarmalavicius, S., Trefzer, U., and Walden, P. (2010) Differential arginine methylation of the G-protein pathway suppressor GPS-2 recognized by tumor-specific T cells in melanoma. FASEB J. 24, 937–946
229. Meadows, L., Wang, W., den Haan, J. M. ., Blokland, E., Reinhardus, C., Drijfhout, J. W., Shabanowitz, J., Pierce, R., Agulnik, A. I., Bishop, C. E., Hunt, D. F., Goulmy, E., and Engelhard, V. H. (1997) The HLA-A*0201-Restricted H-Y Antigen Contains a Posttranslationally Modified Cysteine That Significantly Affects T Cell Recognition. Immunity 6, 273–281
230. Xu, Y., Gendler, S. J., and Franco, A. (2004) Designer glycopeptides for cytotoxic T cell-based elimination of carcinomas. J. Exp. Med. 199, 707–716
231. Kastrup, I. B., Stevanovic, S., Arsequell, G., Valencia, G., Zeuthen, J., Rammensee, H. G., Elliott, T., and Haurum, J. S. (2000) Lectin purified human class I MHC-derived peptides: evidence for presentation of glycopeptides in vivo. Tissue Antigens 56, 129–35
232. Abelin, J. G., Trantham, P. D., Penny, S. a, Patterson, A. M., Ward, S. T., Hildebrand, W. H., Cobbold, M., Bai, D. L., Shabanowitz, J., and Hunt, D. F. (2015) Complementary IMAC enrichment methods for HLA-associated phosphopeptide identification by mass spectrometry. Nat. Protoc. 10, 1308–1318
233. Mosse, C. A., Hsu, W., and Engelhard, V. H. (2001) Tyrosinase degradation via two pathways during reverse translocation to the cytosol. Biochem. Biophys. Res. Commun. 285, 313–9
234. Hudrisier, D., Riond, J., Mazarguil, H., Oldstone, M. B. A., and Gairin, J. E. (1999) Genetically Encoded and Post-translationally Modified Forms of a Major Histocompatibility Complex Class I-restricted Antigen Bearing a Glycosylation Motif Are Independently Processed and Co-presented to Cytotoxic T Lymphocytes. J. Biol. Chem. 274, 36274–36280
235. Altrich-VanLith, M. L., Ostankovitch, M., Polefrone, J. M., Mosse, C. A., Shabanowitz, J., Hunt, D. F., and Engelhard, V. H. (2006) Processing of a class I-restricted epitope from tyrosinase requires peptide N-glycanase and the cooperative action of endoplasmic reticulum aminopeptidase 1 and cytosolic proteases. J. Immunol. 177, 5440–50
236. Mohammed, F., Cobbold, M., Zarling, A. L., Salim, M., Barrett-Wilt, G. A., Shabanowitz, J., Hunt, D. F., Engelhard, V. H., and Willcox, B. E. (2008) Phosphorylation-dependent interaction between antigenic peptides and MHC class I: a molecular basis for the presentation of transformed self. Nat. Immunol. 9, 1236–43
237. Dzhambazov, B., Holmdahl, M., Yamada, H., Lu, S., Vestberg, M., Holm, B., Johnell, O., Kihlberg, J., and Holmdahl, R. (2005) The major T cell epitope on type II collagen is glycosylated in normal cartilage but modified by arthritis in both rats and humans. Eur. J. Immunol. 35, 357–66

238. Petersen, J., Purcell, A. W., and Rossjohn, J. (2009) Post-translationally modified T cell epitopes: immune recognition and immunotherapy. J Mol Med 87, 1045–1051
239. DOYLE, H. A., and MAMULA, M. J. (2005) Posttranslational Modifications of Self-Antigens. Ann. N. Y. Acad. Sci. 1050, 1–9
240. Yates, J. R. (1998) Mass spectrometry and the age of the proteome. J Mass Spectrom 33, 1–19
241. Zubarev, R. A., and Makarov, A. (2013) Orbitrap mass spectrometry. Anal Chem 85, 5288–5296
242. Makarov, A., Denisov, E., Lange, O., and Horning, S. (2006) Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. J Am Soc Mass Spectrom 17, 977–982
243. Olsen, J. V, Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E. R., Senko, M., Makarov, A., Mann, M., and Horning, S. (2009) A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. Mol Cell Proteomics 8, 2759–2769
244. Altelaar, A. F., Munoz, J., and Heck, A. J. (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. Nat Rev Genet 14, 35–48
245. Olsen, J. V, and Mann, M. (2013) Status of large-scale analysis of post-translational modifications by mass spectrometry. Mol Cell Proteomics,
246. Andrews, G. L., Simons, B. L., Young, J. B., Hawkridge, A. M., and Muddiman, D. C. (2011) Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). Anal Chem 83, 5442–5446
247. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A 101, 9528–9533
248. Stoeckli, M., Chaurand, P., Hallahan, D. E., and Caprioli, R. M. (2001) Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. Nat Med 7, 493–496
249. Second, T. P., Blethrow, J. D., Schwartz, J. C., Merrihew, G. E., MacCoss, M. J., Swaney, D. L., Russell, J. D., Coon, J. J., and Zabrouskov, V. (2009) Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures. Anal Chem 81, 7757–7765
250. Jorgenson, J. W. (2010) Capillary liquid chromatography at ultrahigh pressures. Annu Rev Anal Chem (Palo Alto Calif) 3, 129–150
251. MacNair, J. E., Patel, K. D., and Jorgenson, J. W. (1999) Ultrahigh-pressure reversed-phase capillary liquid chromatography: isocratic and gradient elution using columns packed with 1.0-micron particles. Anal Chem 71, 700–708
252. MacNair, J. E., Lewis, K. C., and Jorgenson, J. W. (1997) Ultrahigh-pressure reversed-phase liquid chromatography in packed capillary columns. Anal Chem 69, 983–989
253. Shen, Y., Zhao, R., Berger, S. J., Anderson, G. A., Rodriguez, N., and Smith, R. D. (2002) High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics. Anal Chem 74, 4235–4249
254. Thakur, S. S., Geiger, T., Chatterjee, B., Bandilla, P., Fröhlich, F., Cox, J., and Mann, M. (2011) Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. Mol Cell Proteomics 10, M110.003699
255. Gilar, M., Fridrich, J., Schure, M. R., and Jaworski, A. (2012) Comparison of orthogonality estimation methods for the two-dimensional separations of peptides. Anal Chem 84, 8722–8732
256. Di Palma, S., Hennrich, M. L., Heck, A. J., and Mohammed, S. (2012) Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. J Proteomics 75, 3791–3813
257. Mommen, G. P., Meiring, H. D., Heck, A. J., and de Jong, A. P. (2013) Mixed-bed ion exchange chromatography employing a salt-free pH gradient for improved sensitivity and compatibility in MudPIT. Anal Chem 85, 6608–6616
258. Motoyama, A., Xu, T., Ruse, C. I., Wohlschlegel, J. A., and Yates, J. R. (2007) Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. Anal Chem 79, 3623–3634
259. McDonald, W. H., Ohi, R., Miyamoto, D. T., Mitchison, T. J., and Yates, J. R. (2002) Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. Int. J. Mass Spectrom. 219, 245–251
260. Motoyama, A., Venable, J. D., Ruse, C. I., and Yates, J. R. (2006) Automated ultra-high-pressure multidimensional protein identification technology (UHP-MudPIT) for improved peptide identification of proteomic samples. Anal Chem 78, 5109–5118
261. Mitulović, G., Stingl, C., Smoluch, M., Swart, R., Chervet, J. P., Steinmacher, I., Gerner, C., and Mechtler, K. (2004) Automated, on-line two-dimensional nano liquid chromatography tandem mass spectrometry for rapid analysis of complex protein digests. Proteomics 4, 2545–2557
262. Nägele, E., Vollmer, M., and Hörth, P. (2004) Improved 2D nano-LC/MS for proteomics applications: a comparative analysis using yeast proteome. J Biomol Tech 15, 134–143
263. Liu, H., Finch, J. W., Luongo, J. A., Li, G.-Z., and Gebler, J. C. (2006) Development of an online two-dimensional nano-scale liquid chromatography/mass spectrometry method for improved chromatographic performance and hydrophobic peptide recovery. J. Chromatogr. A 1135, 43–51

264. Taylor, P., Nielsen, P. A., Trelle, M. B., Hørning, O. B., Andersen, M. B., Vorm, O., Moran, M. F., and Kislinger, T. (2009) Automated 2D peptide separation on a 1D nano-LC-MS system. J Proteome Res 8, 1610–1616
265. BURKE, T. W. L., MANT, C. T., BLACK, J. A., and HODGES, R. S. (1989) STRONG CATION-EXCHANGE HIGH-PERFORMANCE LIQUID-CHROMATOGRAPHY OF PEPTIDES- EFFECT OF NON-SPECIFIC HYDROPHOBIC INTERACTIONS AND LINEARIZATION OF PEPTIDE RETENTION BEHAVIOR. J. Chromatogr. 476, 377–389
266. Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C., and Mechtler, K. (2011) Universal and confident phosphorylation site localization using phosphoRS. J Proteome Res 10, 5354–5362
267. Grossmann, J., Roschitzki, B., Panse, C., Fortes, C., Barkow-Oesterreicher, S., Rutishauser, D., and Schlapbach, R. (2010) Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods. J Proteomics 73, 1740–1746
268. Silva, J. C., Gorenstein, M. V, Li, G. Z., Vissers, J. P., and Geromanos, S. J. (2006) Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics 5, 144–156
269. Vizcaíno, J. A., Côté, R. G., Csordas, A., Dianes, J. A., Fabregat, A., Foster, J. M., Griss, J., Alpi, E., Birim, M., Contell, J., O'Kelly, G., Schoenegger, A., Ovelleiro, D., Pérez-Riverol, Y., Reisinger, F., Ríos, D., Wang, R., and Hermjakob, H. (2013) The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res 41, D1063–9
270. Giambruno, R., Grebien, F., Stukalov, A., Knoll, C., Planyavsky, M., Rudashevskaya, E. L., Colinge, J., Superti-Furga, G., and Bennett, K. L. (2013) Affinity purification strategies for proteomic analysis of transcription factor complexes. J. Proteome Res. 12, 4018–27
271. Shen, Y., Moore, R. J., Zhao, R., Blonder, J., Auberry, D. L., Masselon, C., Pasa-Tolić, L., Hixson, K. K., Auberry, K. J., and Smith, R. D. (2003) High-efficiency on-line solid-phase extraction coupling to 15-150-microm-i.d. column liquid chromatography for proteomic analysis. Anal Chem 75, 3596–3605
272. Chamrád, I., Rix, U., Stukalov, A., Gridling, M., Parapatics, K., Müller, A. C., Altiok, S., Colinge, J., Superti-Furga, G., Haura, E. B., and Bennett, K. L. (2013) A miniaturized chemical proteomic approach for target profiling of clinical kinase inhibitors in tumor biopsies. J. Proteome Res. 12, 4005–17
273. King, R., Bonfiglio, R., Fernandez-Metzler, C., Miller-Stein, C., and Olah, T. (2000) Mechanistic investigation of ionization suppression in electrospray ionization. J. Am. Soc. Mass Spectrom. 11, 942–50
274. Blume-Jensen, P., and Hunter, T. (2001) Oncogenic kinase signalling. Nature 411, 355–365
275. Stenberg, K. A., Riikonen, P. T., and Vihinen, M. (1999) KinMutBase, a database of human disease-causing protein kinase mutations. Nucleic Acids Res 27, 362–364
276. Ubersax, J. A., and Ferrell, J. E. (2007) Mechanisms of specificity in protein phosphorylation. Nat Rev Mol Cell Biol 8, 530–541
277. Songyang, Z., Blechner, S., Hoagland, N., Hoekstra, M. F., Piwnica-Worms, H., and Cantley, L. C. (1994) Use of an oriented peptide library to determine the optimal substrates of protein kinases. Curr Biol 4, 973–982
278. Rodriguez, M., Li, S. S., Harper, J. W., and Songyang, Z. (2004) An oriented peptide array library (OPAL) strategy to study protein-protein interactions. J Biol Chem 279, 8802–8807
279. Yaffe, M. B. (2004) Novel at the library. Nat Methods 1, 13–14
280. Manning, B. D., and Cantley, L. C. (2002) Hitting the target: emerging technologies in the search for kinase substrates. Sci STKE 2002, pe49
281. Hutti, J. E., Jarrell, E. T., Chang, J. D., Abbott, D. W., Storz, P., Toker, A., Cantley, L. C., and Turk, B. E. (2004) A rapid method for determining protein kinase phosphorylation specificity. Nat Methods 1, 27–29
282. Wiesner, J., Premsler, T., and Sickmann, A. (2008) Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. Proteomics 8, 4466–4483
283. Frese, C. K., Altelaar, A. F. M., Hennrich, M. L., Nolting, D., Zeller, M., Griep-Raming, J., Heck, A. J. R., and Mohammed, S. (2011) Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. J. Proteome Res. 10, 2377–88
284. Xue, L., Wang, W. H., Iliuk, A., Hu, L., Galan, J. A., Yu, S., Hans, M., Geahlen, R. L., and Tao, W. A. (2012) Sensitive kinase assay linked with phosphoproteomics for identifying direct kinase substrates. Proc Natl Acad Sci U S A 109, 5615–5620
285. Singh, S. A., Winter, D., Bilimoria, P. M., Bonni, A., Steen, H., and Steen, J. A. (2012) FLEXIQinase, a mass spectrometry-based assay, to unveil multikinase mechanisms. Nat Methods 9, 504–508
286. Schilling, O., auf dem Keller, U., and Overall, C. M. (2011) Protease specificity profiling by tandem mass spectrometry using proteome-derived peptide libraries. Methods Mol Biol 753, 257–272
287. Schilling, O., and Overall, C. M. (2008) Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. Nat Biotechnol 26, 685–694
288. Kettenbach, A. N., Wang, T., Faherty, B. K., Madden, D. R., Knapp, S., Bailey-Kellogg, C., and Gerber, S. A. (2012) Rapid determination of multiple linear kinase substrate motifs by mass spectrometry. Chem Biol 19, 608–618
289. Gevaert, K., Staes, A., Van Damme, J., De Groot, S., Hugelier, K., Demol, H., Martens, L., Goethals, M., and Vandekerckhove, J. (2005) Global phosphoproteome analysis on human HepG2 hepatocytes using reversed-phase

diagonal LC. Proteomics 5, 3589–3599
290. Staes, A., Impens, F., Van Damme, P., Ruttens, B., Goethals, M., Demol, H., Timmerman, E., Vandekerckhove, J., and Gevaert, K. (2011) Selecting protein N-terminal peptides by combined fractional diagonal chromatography. Nat Protoc 6, 1130–1141
291. Staes, A., Van Damme, P., Helsens, K., Demol, H., Vandekerckhove, J., and Gevaert, K. (2008) Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). Proteomics 8, 1362–1370
292. Taouatas, N., Heck, A. J., and Mohammed, S. (2010) Evaluation of metalloendopeptidase Lys-N protease performance under different sample handling conditions. J Proteome Res 9, 4282–4288
293. Jelluma, N., Brenkman, A. B., McLeod, I., Yates, J. R., Cleveland, D. W., Medema, R. H., and Kops, G. J. (2008) Chromosomal instability by inefficient Mps1 auto-activation due to a weakened mitotic checkpoint and lagging chromosomes. PLoS One 3, e2415
294. Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A. J. (2009) Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protoc 4, 484–494
295. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009) Improved visualization of protein consensus sequences by iceLogo. Nat Methods 6, 786–787
296. Peng, M., Taouatas, N., Cappadona, S., van Breukelen, B., Mohammed, S., Scholten, A., and Heck, A. J. (2012) Protease bias in absolute protein quantitation. Nat Methods 9, 524–525
297. Mohammed, S., and Heck, A. J. (2011) Strong cation exchange (SCX) based analytical methods for the targeted analysis of protein post-translational modifications. Curr. Opin. Biotechnol. 22, 9–16
298. Hennrich, M. L., van den Toorn, H. W., Groenewold, V., Heck, A. J., and Mohammed, S. (2012) Ultra acidic strong cation exchange enabling the efficient enrichment of basic phosphopeptides. Anal Chem 84, 1804–1808
299. Shabb, J. B. (2001) Physiological substrates of cAMP-dependent protein kinase. Chem Rev 101, 2381–2411
300. Ruppelt, A., and Tasken, K. Physiological substrates of PKA and PKG. Three-Volu,
301. Kops, G. J., and Shah, J. V (2012) Connecting up and clearing out: how kinetochore attachment silences the spindle assembly checkpoint. Chromosoma 121, 509–525
302. Vleugel, M., Hoogendoorn, E., Snel, B., and Kops, G. J. (2012) Evolution and function of the mitotic checkpoint. Dev Cell 23, 239–250
303. Kang, J., Chen, Y., Zhao, Y., and Yu, H. (2007) Autophosphorylation-dependent activation of human Mps1 is required for the spindle checkpoint. Proc Natl Acad Sci U S A 104, 20232–20237
304. Mattison, C. P., Old, W. M., Steiner, E., Huneycutt, B. J., Resing, K. A., Ahn, N. G., and Winey, M. (2007) Mps1 activation loop autophosphorylation enhances kinase activity. J Biol Chem 282, 30553–30561
305. Dou, Z., von Schubert, C., Körner, R., Santamaria, A., Elowe, S., and Nigg, E. A. (2011) Quantitative mass spectrometry analysis reveals similar substrate consensus motif for human Mps1 kinase and Plk1. PLoS One 6, e18793
306. Mok, J., Kim, P. M., Lam, H. Y., Piccirillo, S., Zhou, X., Jeschke, G. R., Sheridan, D. L., Parker, S. A., Desai, V., Jwa, M., Cameroni, E., Niu, H., Good, M., Remenyi, A., Ma, J. L., Sheu, Y. J., Sassi, H. E., Sopko, R., Chan, C. S., De Virgilio, C., Hollingsworth, N. M., Lim, W. A., Stern, D. F., Stillman, B., Andrews, B. J., Gerstein, M. B., Snyder, M., and Turk, B. E. (2010) Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. Sci Signal 3, ra12
307. Mann, M., Ong, S. E., Grønborg, M., Steen, H., Jensen, O. N., and Pandey, A. (2002) Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. Trends Biotechnol 20, 261–268
308. Holinger, E. P., Old, W. M., Giddings, T. H., Wong, C., Yates, J. R., and Winey, M. (2009) Budding yeast centrosome duplication requires stabilization of Spc29 via Mps1-mediated phosphorylation. J Biol Chem 284, 12949–12955
309. Friedman, D. B., Kern, J. W., Huneycutt, B. J., Vinh, D. B., Crawford, D. K., Steiner, E., Scheiltz, D., Yates, J., Resing, K. A., Ahn, N. G., Winey, M., and Davis, T. N. (2001) Yeast Mps1p phosphorylates the spindle pole component Spc110p in the N-terminal domain. J Biol Chem 276, 17958–17967
310. Jelluma, N., Brenkman, A. B., van den Broek, N. J., Cruijsen, C. W., van Osch, M. H., Lens, S. M., Medema, R. H., and Kops, G. J. (2008) Mps1 phosphorylates Borealin to control Aurora B activity and chromosome alignment. Cell 132, 233–246
311. London, N., Ceto, S., Ranish, J. A., and Biggins, S. (2012) Phosphoregulation of Spc105 by Mps1 and PP1 regulates Bub1 localization to kinetochores. Curr Biol 22, 900–906
312. Shepperd, L. A., Meadows, J. C., Sochaj, A. M., Lancaster, T. C., Zou, J., Buttrick, G. J., Rappsilber, J., Hardwick, K. G., and Millar, J. B. (2012) Phosphodependent recruitment of Bub1 and Bub3 to Spc7/KNL1 by Mph1 kinase maintains the spindle checkpoint. Curr Biol 22, 891–899
313. Yamagishi, Y., Yang, C. H., Tanno, Y., and Watanabe, Y. (2012) MPS1/Mph1 phosphorylates the kinetochore protein KNL1/Spc7 to recruit SAC components. Nat Cell Biol 14, 746–752
314. Kemmler, S., Stach, M., Knapp, M., Ortiz, J., Pfannstiel, J., Ruppert, T., and Lechner, J. (2009) Mimicking Ndc80 phosphorylation triggers spindle assembly checkpoint signalling. EMBO J 28, 1099–1110

315. Araki, Y., Gombos, L., Migueleti, S. P., Sivashanmugam, L., Antony, C., and Schiebel, E. (2010) N-terminal regions of Mps1 kinase determine functional bifurcation. J Cell Biol 189, 41–56
316. Shimogawa, M. M., Graczyk, B., Gardner, M. K., Francis, S. E., White, E. A., Ess, M., Molk, J. N., Ruse, C., Niessen, S., Yates, J. R., Muller, E. G., Bloom, K., Odde, D. J., and Davis, T. N. (2006) Mps1 phosphorylation of Dam1 couples kinetochores to microtubule plus ends at metaphase. Curr Biol 16, 1489–1501
317. Roche, P. A., and Furuta, K. (2015) The ins and outs of MHC class II-mediated antigen processing and presentation. Nat Rev Immunol 15, 203–216
318. Suri, A., Lovitch, S. B., and Unanue, E. R. (2006) The wide diversity and complexity of peptides bound to class II MHC molecules. Curr Opin Immunol 18, 70–77
319. Ovsyannikova, I. G., Johnson, K. L., Bergen, H. R., and Poland, G. A. (2007) Mass spectrometry and peptide-based vaccine development. Clin Pharmacol Ther 82, 644–652
320. Thibodeau, J., Bourgeois-Daigneault, M. C., and Lapointe, R. (2012) Targeting the MHC Class II antigen presentation pathway in cancer immunotherapy. Oncoimmunology 1, 908–916
321. Purcell, A. W. (2004) Isolation and characterization of naturally processed MHC-bound peptides from the surface of antigen-presenting cells. Methods Mol Biol 251, 291–306
322. Lippolis, J. D., White, F. M., Marto, J. A., Luckey, C. J., Bullock, T. N., Shabanowitz, J., Hunt, D. F., and Engelhard, V. H. (2002) Analysis of MHC class II antigen processing by quantitation of peptides that constitute nested sets. J Immunol 169, 5089–5097
323. Münz, C. (2012) Antigen processing for MHC class II presentation via autophagy. Front. Immunol. 3, 1–6
324. Deretic, V., Saitoh, T., and Akira, S. (2013) Autophagy in infection, inflammation and immunity. Nat Rev Immunol 13, 722–737
325. Stenger, R. M., Meiring, H. D., Kuipers, B., Poelen, M., van Gaans-van den Brink, J. A., Boog, C. J., de Jong, A. P., and van Els, C. A. (2014) Bordetella pertussis proteins dominating the major histocompatibility complex class II-presented epitope repertoire in human monocyte-derived dendritic cells. Clin Vaccine Immunol 21, 641–650
326. Bergseng, E., Dørum, S., Arntzen, M. Ø., Nielsen, M., Nygård, S., Buus, S., de Souza, G. a., and Sollid, L. M. (2014) Different binding motifs of the celiac disease-associated HLA molecules DQ2.5, DQ2.2, and DQ7.5 revealed by relative quantitative proteomics of endogenous peptide repertoires. Immunogenetics 67, 73–84
327. Depontieu, F. R., Qian, J., Zarling, A. L., McMiller, T. L., Salay, T. M., Norris, A., English, A. M., Shabanowitz, J., Engelhard, V. H., Hunt, D. F., and Topalian, S. L. (2009) Identification of tumor-associated, MHC class II-restricted phosphopeptides as targets for immunotherapy. Proc Natl Acad Sci U S A 106, 12073–12078
328. Benham, H., Nel, H. J., Law, S. C., Mehdi, A. M., Street, S., Ramnoruth, N., Pahau, H., Lee, B. T., Ng, J., G Brunck, M. E., Hyde, C., Trouw, L. A., Dudek, N. L., Purcell, A. W., O'Sullivan, B. J., Connolly, J. E., Paul, S. K., Lê Cao, K. A., and Thomas, R. (2015) Citrullinated peptide dendritic cell immunotherapy in HLA risk genotype-positive rheumatoid arthritis patients. Sci Transl Med 7, 290ra87
329. Hill, J. a, Southwood, S., Sette, A., Jevnikar, A. M., Bell, D. a, and Cairns, E. (2003) Cutting edge: the conversion of arginine to citrulline allows for a high-affinity peptide interaction with the rheumatoid arthritis-associated HLA-DRB1*0401 MHC class II molecule. J. Immunol. 171, 538–541
330. Chornoguz, O., Gapeev, a., O'Neill, M. C., and Ostrand-Rosenberg, S. (2012) Major Histocompatibility Complex Class II+ Invariant Chain Negative Breast Cancer Cells Present Unique Peptides That Activate Tumor-specific T Cells From Breast Cancer Patients. Mol. Cell. Proteomics, 1457–1467
331. Dengjel, J., Decker, P., Schoor, O., Altenberend, F., Weinschenk, T., Rammensee, H. G., and Stevanovic, S. (2004) Identification of a naturally processed cyclin D1 T-helper epitope by a novel combination of HLA class II targeting and differential mass spectrometry. Eur J Immunol 34, 3644–3651
332. Collado, J. A., Alvarez, I., Ciudad, M. T., Espinosa, G., Canals, F., Pujol-Borrell, R., Carrascal, M., Abian, J., and Jaraquemada, D. (2013) Composition of the HLA-DR-associated human thymus peptidome. Eur J Immunol 43, 2273–2282
333. Granados, D. P., Laumont, C. M., Thibault, P., and Perreault, C. (2015) The nature of self for T cells—a systems-level perspective. Curr. Opin. Immunol. 34, 1–8
334. van Haren, S. D., Herczenik, E., ten Brinke, A., Mertens, K., Voorberg, J., and Meijer, A. B. (2011) HLA-DR-presented peptide repertoires derived from human monocyte-derived dendritic cells pulsed with blood coagulation factor VIII. Mol Cell Proteomics 10, M110.002246
335. Masterson, A. J., Sombroek, C. C., de Gruijl, T. D., Graus, Y. M. F., van der Vliet, H. J. J., Lougheed, S. M., van den Eertwegh, A. J. M., Pinedo, H. M., and Scheper, R. J. (2002) MUTZ-3, a human cell line model for the cytokine-induced differentiation of dendritic cells from CD34+precursors. Blood 100 , 701–703
336. Hoefnagel, M. H., Vermeulen, J. P., Scheper, R. J., and Vandebriel, R. J. (2011) Response of MUTZ-3 dendritic cells to the different components of the Haemophilus influenzae type B conjugate vaccine: towards an in vitro assay for vaccine immunogenicity. Vaccine 29, 5114–5121
337. Hoonakker, M. E., Verhagen, L. M., Hendriksen, C. F., van Els, C. A., Vandebriel, R. J., Sloots, A., and Han, W. G. (2015) In vitro innate immune cell based models to assess whole cell Bordetella pertussis vaccine quality: a proof

of principle. Biologicals 43, 100–109
338. Meiring, H. D., Soethout, E. C., Poelen, M. C. M., Mooibroek, D., Hoogerbrugge, R., Timmermans, H., Boog, C. J., Heck, A. J. R., de Jong, A. P. J. M., and van Els, C. a C. M. (2006) 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 5, 902–913
339. Meiring, H. D., Soethout, E. C., de Jong, A. P., and van Els, C. A. (2007) Targeted identification of infection-related HLA class I-presented epitopes by stable isotope tagging of epitopes (SITE). Curr Protoc Immunol Chapter 16, Unit 16.3
340. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., and MacCoss, M. J. (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods 4, 923–925
341. Mommen, G. P., Frese, C. K., Meiring, H. D., van Gaans-van den Brink, J., de Jong, A. P., van Els, C. A., and Heck, A. J. (2014) Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proc Natl Acad Sci U S A 111, 4507–4512
342. Yu, C. S., Cheng, C. W., Su, W. C., Chang, K. C., Huang, S. W., Hwang, J. K., and Lu, C. H. (2014) CELLO2GO: a web server for protein subCELlular LOcalization prediction with functional gene ontology annotation. PLoS One 9, e99368
343. Karosiene, E., Rasmussen, M., Blicher, T., Lund, O., Buus, S., and Nielsen, M. (2013) NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. Immunogenetics 65, 711–724
344. Andreatta, M., Lund, O., and Nielsen, M. (2013) Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach. Bioinformatics 29, 8–14
345. Impens, F., Colaert, N., Helsens, K., Ghesquière, B., Timmerman, E., De Bock, P. J., Chain, B. M., Vandekerckhove, J., and Gevaert, K. (2010) A quantitative proteomics design for systematic identification of protease cleavage events. Mol Cell Proteomics 9, 2327–2333
346. Sun, H., Lou, X., Shan, Q., Zhang, J., Zhu, X., Wang, Y., Xie, Y., Xu, N., and Liu, S. (2013) Proteolytic characteristics of cathepsin D related to the recognition and cleavage of its target proteins. PLoS One 8, e65733
347. Nelson, C. A., Vidavsky, I., Viner, N. J., Gross, M. L., and Unanue, E. R. (1997) Amino-terminal trimming of peptides for presentation on major histocompatibility complex class II molecules. Proc Natl Acad Sci U S A 94, 628–633
348. Larsen, S. L., Pedersen, L. O., Buus, S., and Stryhn, A. (1996) T cell responses affected by aminopeptidase N (CD13)-mediated trimming of major histocompatibility complex class II-bound peptides. J Exp Med 184, 183–189
349. West, M. A., Wallin, R. P., Matthews, S. P., Svensson, H. G., Zaru, R., Ljunggren, H. G., Prescott, A. R., and Watts, C. (2004) Enhanced dendritic cell antigen capture via toll-like receptor-induced actin remodeling. Science (80-.). 305, 1153–1157
350. Mi, H., Muruganujan, A., Casagrande, J. T., and Thomas, P. D. (2013) Large-scale gene function analysis with the PANTHER classification system. Nat Protoc 8, 1551–1566
351. Caron, E., Espona, L., Kowalewski, D. J., Schuster, H., Ternette, N., Alpízar, A., Schittenhelm, R. B., Ramarathinam, S. H., Lindestam Arlehamn, C. S., Chiek Koh, C., Gillet, L. C., Rabsteyn, A., Navarro, P., Kim, S., Lam, H., Sturm, T., Marcilla, M., Sette, A., Campbell, D. S., Deutsch, E. W., Moritz, R. L., Purcell, A. W., Rammensee, H.-G., Stevanovic, S., and Aebersold, R. (2015) An open-source computational and data resource to analyze digital maps of immunopeptidomes. Elife 4, 1–17
352. Seward, R. J., Drouin, E. E., Steere, A. C., and Costello, C. E. (2011) Peptides presented by HLA-DR molecules in synovia of patients with rheumatoid arthritis or antibiotic-refractory Lyme arthritis. Mol. Cell. Proteomics 10, M110.002477
353. Alvarez, I., Collado, J., Daura, X., Colomé, N., Rodríguez-García, M., Gallart, T., Canals, F., and Jaraquemada, D. (2008) The rheumatoid arthritis-associated allele HLA-DR10 (DRB1*1001) shares part of its repertoire with HLA-DR1 (DRB1*0101) and HLA-DR4 (DRB*0401). Arthritis Rheum. 58, 1630–1639
354. Sasaki, K., Osaki, T., and Minamino, N. (2013) Large-scale identification of endogenous secretory peptides using electron transfer dissociation mass spectrometry. Mol. Cell. Proteomics 12, 700–9
355. Marino, F., Bern, M., Mommen, G. P. M., Leney, A. C., van Gaans-van den Brink, J. a. M., Bonvin, A. M. J. J., Becker, C., van Els, C. a. C. M., and Heck, A. J. R. (2015) Extended O-GlcNAc on HLA Class-I-Bound Peptides. J. Am. Chem. Soc., 150819112655001
356. Milner, E., Barnea, E., Beer, I., and Admon, A. (2006) The Turnover Kinetics of Major Histocompatibility Complex Peptides of Human Cancer Cells. Mol. Cell. Proteomics, 357–365
357. Weinzierl, A. O., Lemmel, C., Schoor, O., Muller, M., Kruger, T., Wernet, D., Hennenlotter, J., Stenzl, A., Klingel, K., Rammensee, H.-G., and Stevanovic, S. (2006) Distorted Relation between mRNA Copy Number and Corresponding Major Histocompatibility Complex Ligand Density on the Cell Surface. Mol. Cell. Proteomics 6, 102–113
358. Fortier, M.-H., Caron, É., Hardy, M.-P., Voisin, G., Lemieux, S., Perreault, C., and Thibault, P. (2008) The MHC class I peptide repertoire is molded by the transcriptome. J. Exp. Med. 205, 595–610

359. Blanco, P., Palucka, A. K., Pascual, V., and Banchereau, J. (2008) Dendritic cells and cytokines in human inflammatory and autoimmune diseases. Cytokine Growth Factor Rev 19, 41–52
360. Watts, C. (2012) The endosome-lysosome pathway and information generation in the immune system. Biochim. Biophys. Acta- Proteins Proteomics 1824, 14–21
361. Chicz, R. M., Urban, R. G., Gorga, J. C., Vignali, D. A., Lane, W. S., and Strominger, J. L. (1993) Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. J Exp Med 178, 27–47
362. Rotzschke, O., and Falk, K. (1994) Origin, structure and motifs of naturally processed MHC class II ligands. Curr Opin Immunol 6, 45–51.
363. Juncker, A. S., Larsen, M. V, Weinhold, N., Nielsen, M., Brunak, S., and Lund, O. (2009) Systematic characterisation of cellular localisation and expression profiles of proteins containing MHC ligands. PLoS One 4, e7448
364. Schmid, D., Dengjel, J., Schoor, O., Stevanovic, S., and Münz, C. (2006) Autophagy in innate and adaptive immunity against intracellular pathogens. J Mol Med 84, 194–202
365. Jaraquemada, D., Marti, M., and Long, E. O. (1990) An endogenous processing pathway in vaccinia virus-infected cells for presentation of cytoplasmic antigens to class II-restricted T cells. J Exp Med 172, 947–954
366. Jacobson, S., Sekaly, R. P., Jacobson, C. L., McFarland, H. F., and Long, E. O. (1989) HLA class II-restricted presentation of cytoplasmic measles virus antigens to cytotoxic T cells. J Virol 63, 1756–1762
367. Li, Y., Wang, L. X., Pang, P., Cui, Z., Aung, S., Haley, D., Fox, B. A., Urba, W. J., and Hu, H. M. (2011) Tumor-derived autophagosome vaccine: mechanism of cross-presentation and therapeutic efficacy. Clin Cancer Res 17, 7047–7057
368. Wang, R. F., Wang, X., Atwood, A. C., Topalian, S. L., and Rosenberg, S. A. (1999) Cloning genes encoding MHC class II-restricted antigens: mutated CDC27 as a tumor antigen. Science (80-.). 284, 1351–1354
369. Dengjel, J., Schoor, O., Fischer, R., Reich, M., Kraus, M., Müller, M., Kreymborg, K., Altenberend, F., Brandenburg, J., Kalbacher, H., Brock, R., Driessen, C., Rammensee, H. G., and Stevanovic, S. (2005) Autophagy promotes MHC class II presentation of peptides from intracellular source proteins. Proc Natl Acad Sci U S A 102, 7922–7927
370. Münz, C. (2009) Enhancing immunity through autophagy. Annu. Rev. Immunol. 27, 423–449
371. Crotzer, V. L., and Blum, J. S. (2009) Autophagy and its role in MHC-mediated antigen presentation. J Immunol 182, 3335–3341
372. Herberts, C. a., Meiring, H. M., Van Gaans-van den Brink, J. a M., Van der Heeft, E., Poelen, M. C. M., Boog, C. J. P., De Jong, A. P. J. M., and Van Els, C. a C. M. (2003) Dynamics of measles virus protein expression are reflected in the MHC class I epitope display. Mol. Immunol. 39, 567–575
373. Yewdell, J. W. (2006) Confronting complexity: real-world immunodominance in antiviral CD8+ T cell responses. Immunity 25, 533–543
374. Blum, J. S., Wearsch, P. a, and Cresswell, P. (2013) Pathways of Antigen Processing
375. Villadangos, J. A., Schnorrer, P., and Wilson, N. S. (2005) Control of MHC class II antigen presentation in dendritic cells: a balance between creative and destructive forces. Immunol Rev 207, 191–205
376. Novellino, L., Renkvist, N., Rini, F., Mazzocchi, A., Rivoltini, L., Greco, A., Deho, P., Squarcina, P., Robbins, P. F., Parmiani, G., and Castelli, C. (2003) Identification of a mutated receptor-like protein tyrosine phosphatase kappa as a novel, class II HLA-restricted melanoma antigen. J. Immunol. 170, 6363–6370

CHAPTER 2

Characterization and usage of the EASY-spray technology as part of an online 2D SCX-RP ultra-high pressure system

Fabio Marino[‡†], Alba Cristobal[‡†], Nadine A. Binai[‡†], Nicolai Bache[§], Albert J.R. Heck[‡†], Shabaz Mohammed[‡†#]

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

[‡]Netherlands Proteomics Centre, Padualaan 8, 3584 CH, Utrecht, The Netherlands.

[§]Thermo Fisher Scientific, Edisonsvej 4, DK-5000 Odense C, Denmark

[#]Department of Biochemistry, University of Oxford, South Parks Road, OX1 3QU, Oxford, United Kingdom.

Analyst. 2014 Dec 21;139(24):6520-8.

ABSTRACT

Ultra-high pressure liquid chromatography (UHPLC) systems combined with state-of-the-art mass spectrometers have pushed the limit of deep proteome sequencing to new heights making it possible to identify thousands of proteins in a single LC-MS experiment within a few hours. The recently released EASY-spray technology allows one to implement nano UHPLC with straightforwardness. In this work we initially characterized the EASY-spray containing a 50 cm column containing <2 µm particles and found the system allowed 3000 proteins to be identified in 90 minutes. We then asked the question whether a fast and sensitive online 2D SCX-RP UHPLC-MS/MS workflow, could compete with 1D long gradient analyses, using total analysis time- versus proteome coverage and sample used as benchmark parameters. The 2D LC-MS strategy consisted of EASY-spray system that had been augmented by the addition of an SCX column. The conversion was made facile since no additional valves were required and by the use of components containing viper fittings. We benchmarked the system using a human cell lysate digest (<10 µg). The 2D SCX-RP UHPLC-MS/MS workflow allowed the identification of almost 37000 unique peptides and 6000 proteins in a total analysis time of ~7 hours. On the same system a 1D RP UHPLC-MS/MS workflow plateaued at only 20000 peptides and 4400 unique proteins and required approx. 8 hours of analysis time. Furthermore, the 2D workflow could continue to increase the proteome coverage with longer analysis times, in fact with a 21 hour analysis we identified 56600 unique peptides and >7500 proteins. We report, here, that with this fast online SCX-RP UHPLC-MS/MS workflow, the proteome coverage can be substantially extended without significantly compromising analysis time and sample usage.

1. INTRODUCTION

A major goal in mass spectrometry based proteomics is the complete characterization of the proteome that is to detect/observe all proteins in all their proteoforms. The main analytical obstacles to reach this goal can be classified into three areas; lack of sensitivity, complexity and insufficient dynamic range. At present most in-depth analyses of proteomes employ 'shotgun strategies'; (1) where all proteins are extracted from the cells, digested with trypsin, where after the resulting complex peptide mixture is subjected to a form of liquid chromatography (LC)-mass spectrometry (MS) analysis. (2) These two symbiotic analytical technologies (liquid chromatography and mass spectrometry) have improved immensely in the last decade. Mass spectrometry has advanced in many aspects, i.e. new peptide sequencing techniques, sequencing speed, mass resolution, sensitivity and dynamic range in ion detection. (2–12) The advances in MS, particularly speed, have meant that the overall performance of a proteomic LC-MS analysis would progress with increased resolution power of the upstream separation approach. Furthermore, higher resolving power would increase the chance of detecting low-abundant peptides, including those that are very similar in mass to charge (m/z) ratio. In single dimensional (1D) LC-MS experiments, for which normally C18 materials are used, longer columns with smaller internal diameter (ID) and particle size have been introduced to increase the resolving power, even though such modifications lead to a dramatic increase of the back pressure on the system. (13) Jorgenson et al. pioneered ultra-high pressure liquid chromatography (UHPLC) using long columns with smaller ID and sub-2 µm particles. (14, 15) Subsequently, Smith and co-workers provided

compelling evidence of the utility of UHPLC for the proteomics community. (16) Nowadays, UHPLC systems which can hold up pressures of about 1000 bar have become commercially available and have been implemented in proteomics workflows by an increasing number of laboratories. Illustratively, Kocher et al. showed that a peak capacity of almost 700 (with a 10 hour gradient) can be achieved through the use of 50 cm columns and 2 µm particles. (17) One-dimensional RP-based UHPLC combined with the latest MS/MS technologies have extended the achievable depth in proteome coverage. Several groups have demonstrated that a few thousand proteins can now be detected in a single LC-MS/MS run using UHPLC with long columns, <2 micron particles and extended gradients. (18–22) There is a very direct link between proteome depth and peak capacity of the chromatographic system. (17) This relationship highlights the need for even higher peak capacities since full proteome analyses (in terms of the number of proteins and peptides) are still not being achieved. Although the current 1D LC-MS/MS workflows are powerful, it is still evident that no single dimensional approach has the right requirements to achieve this ultimate goal. Offline multidimensional approaches, ideally using orthogonal separation principles, (23, 24) provide the ideal way of dramatically increasing the overall peak capacity, but typically that comes at the expense of sample loss and total MS analysis time. (25) The pioneering work of Yates and co-workers (26) whereby they coupled SCX to RP LC-MS/MS analysis ('MudPIT'), both in offline and online workflows, became the template for many 2D LC-MS/MS approaches. In MudPIT, the peptides bind first onto the SCX material and are eluted to the RP through several steps of solvent injections containing increasing concentrations of volatile salt. In between the salt steps a gradient of ACN is applied for the separation of the peptides that became bound to the RP material. (26) The original MudPIT approach has been further developed, for instance, by introducing a thriphasic version of the capillary containing RP-SCX-RP particles or using mixed phase ion exchange columns. (27, 28) Here, the first RP operates as a trap, thus avoiding the sample loss during the needed offline desalting step prior to LC-MS/MS analysis. (29) Yates and co-workers also demonstrated a MudPIT system operating in the UHPLC regime (30) several years before UHPLC systems were commercialized. Despite the excellent performance of MudPIT, offline systems are still more popular because they are perceived as less difficult to build and control. In valve switching approaches, the two columns are not directly connected, allowing higher flexibility in the choice of the solvents. Additionally, first dimension SCX columns with an ID bigger than the RP second dimension column can be employed, so that a higher (and more appropriate) amount of sample can be loaded. (25, 31) Often, both in MudPIT and column switching approaches, a RP-based trap column is placed in between the SCX and RP columns used for the peptide separation. This allows to concentrate and desalt the peptides eluted from the SCX column, adding extra flexibility and giving the possibility to use solvents containing either volatile (32) or non-volatile salts. (33) Gebler et al. (34)and more recently, Kislinger et al. (35) demonstrated that commercial systems can be modified without too much effort into a 2D LC system where salt plugs were injected directly from the auto-sampler. In this way, the setup could be simplified because the solvent containing the salt does not need to be delivered by additional pumps. Here, we demonstrate that the EASY-spray design can easily be augmented to be a 2D LC configuration. We extend the platform by introducing an SCX trap (with very few new connections) and by using nanoViper fittings. Our design requires no extra pumps for the release of the SCX solvents, and can be switched rapidly from 1D to 2D and vice versa. We also show that this fast online 2D SCX-RP UHPLC-MS/MS

workflow remarkably extends the proteome coverage remarkable when compared to 1D approaches, notably without compromising analysis time and sample usage.

2. EXPERIMENTAL SECTION

Materials

Complete Mini EDTA-free protease inhibitor cocktail and phosStop phosphatase inhibitor cocktail were obtained from Roche Diagnostic (Mannheim, DE), lysyl endopeptidase (Lys-C) from Wako (Richmond, VA, USA) and trypsin endopeptidase from Promega (Madison, WI, USA). iodacetamide, dl-dithiothreitol (DTT), ammonium bicarbonate (AMBIC) and ammonium acetate were purchased from Sigma Aldrich (Steinheim, DE). Urea, formic acid (FA) and dimethylsulfoxyde (DMSO) were obtained from Merck (Darmstadt, DE). HPLC grade acetonitrile (ACN) was obtained from Biosolve BV (Valkenswaard, NL). Bradford protein assay was purchased from BioRad Laboratories (Hercules, CA) and high purity water was obtained from a Millipore Milli-Q system (Billerica, MA).

Preparation of the HEK293 digest

HEK293 cells were re-suspended in lysis buffer composed of 8M urea, 50 mM AMBIC pH 8, 1 tablet of PhosStop phosphatase inhibitors, and 1 tablet of complete Mini EDTA-free protease inhibitor cocktail. The cell lysate was sonicated 3 times on ice. After, centrifugation at 20000 g at 4°C for 20 minutes the soluble protein fraction was separated from the insoluble protein fraction. The soluble fraction was collected and the protein concentration of the lysate was determined by a Bradford protein assay. Proteins were reduced with 2 mM DTT at 56°C for 25 minutes, followed by alkylation with iodacetamide (4 mM) at room temperature for 30 minutes in the dark. After the alkylation step the digestion was carried out with a first step of Lys-C for 4 hours at 37°C with a protein to enzyme ratio of 75:1 (w/w). Subsequently, the sample was diluted 4 times with 50 mM AMBIC to a urea concentration of 2 M. The second step of digestion was performed with trypsin overnight at 37°C with a substrate to enzyme ratio of 100:1 (w/w). After digestion the sample was acidified with 10% FA.

1D LC setup and experiments

LC separation was performed with an EASY-spray system (Thermo Scientific, Odense, DK) consisting of a 50 cm, 75 µm ID PepMap RSLC, C18, 100 Å, 2 µm particles which was connected to an Easy-nLC Ultra UHPLC system (Thermo Scientific, Odense, DK). The spray emitter was set at 1.9 kV and the column was heated to 30 °C. The sample was picked up from the auto-sampler vial plate and loaded into the loop (Figure.1A) using the Pump S at 20 µl/min with solvent A (99.9% water, 0.1% FA). During the sample pick up, 5 µl and 1 µl of solvent A were respectively used to equilibrate the pre-column and the analytical column at the controlled back pressure of 700 bar. The sample was loaded (Figure.1B) from the loop on the back-flushed trap column (Thermo, PepMap RSLC, C18, 100 Å, 5 µm particles packed in 5 mm trap column with 300 µm ID) and the trap was connected to waste for the time needed to wash with 20 µl of solvent A. The pre-column, analytical column equilibration and the loading were back pressure controlled, therefore the time employed to perform these steps can be prone to little variance. After the loading/washing step the back flushed trap column was then switched online with the analytical column (Figure.1C) and a gradient of solvent A and B (99.9% ACN, 0.1% FA) was started. The gradient for the separation ranged from 7 to 30% of solvent B respectively in 22,

37, 67, 157, 214, 334, 454 and 574 minutes at a flow rate of 150 nl/min. After the gradient, the column was washed for 2 minutes by increasing the buffer B concentration to 100% followed by conditioning the system with 93% buffer A for at least 15 minutes. All the steps led to an analysis time of ~45, ~60, ~90, ~180, ~240, ~360, ~480 and ~600 min. Total analysis time includes washing steps, pre-column equilibration, analytical column equilibration and loading, which in each run approximately took 20-25 min. With analysis times of 45, 60, 90 and 180 minutes lengths, 1 µg of protein digest was injected while for the longer runs of 240, 360,480 and 600 minutes, 4 µg were injected in order to have a similar total ion current in every run.

Figure 1.

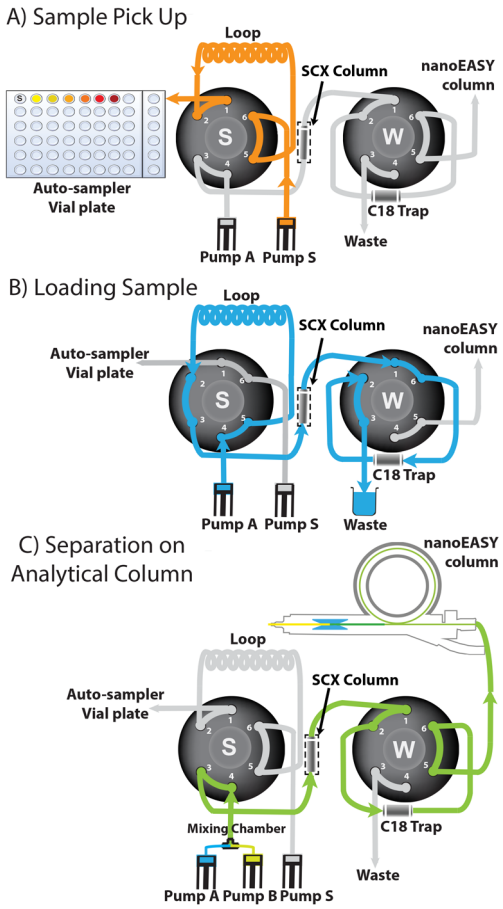


Figure.1: *Description of the 1D and 2D EASY-spray configurations. An SCX trap column (dashed box) between valves S and W is the difference between the 1D and 2D setups. (A) Sample pick up: the sample or the salt plugs are picked up from the auto-sampler vial plate and loaded on the loop using pump S. (B) The sample is carried from the loop to the C18 trap in the 1D fashion and from the loop to the SCX column for the 2D strategy. In the 2D setup the salt plugs are subsequently injected, populations of peptides with increasing net charges are displaced from the SCX column and bind the C18 trap. (C) The peptides are then eluted from the back-flushed C18 trap column and separated on the nanoEASY-spray column with a gradient of A and B solvents.*

2D LC Setup and automated 6 salt plugs experiments

The 2D experiments were performed using the same UHPLC system, trap and analytical column as described for the 1D experiments (Figure.1). In order to implement a 2D setup we added a SCX trap column (Luna, 5 µm, 100 Å, 100 µm ID, 5 cm, Phenomenex, Utrecht, NL) (Dashed box in Figure.1) in between the sample line and the W valve (Figure.1). (An SCX material with similar performance to the one used in this work, is also commercially available and can be purchased as a readymade column from Thermo Scientific). In all experiments 10 µg of HEK293 digest was used. The sample was transferred from the auto-sampler to the sample loop (Figure.1A) with solvent S (0.05% FA and 5% DMSO). The sample was loaded on the SCX trap column (Figure.1B) with 25 µl and at 1 µl/min. Typical flow rates are an order of magnitude higher for loading but we found that decreasing the flow rate for the SCX trapping helped to quantitatively trap the peptides on the SCX material. The peptides that did not bind the SCX trap column (flow through) were trapped on the back-flushed reversed phase trap column. Switching the W valve switch connects the trap column with the analytical column and a gradient of solvent A and B, as described above for the 1D experiments, was applied for the separation (Figure.1C). We used 37 min gradients for short 2D experiments and 157 min gradients for the long 2D experiments. The above described steps were repeated for each of the injected salt steps. The six salt plugs contained ammonium acetate at concentrations of 5mM, 10 mM, 20 mM, 50 mM, 100 mM, 500 mM, 5% of ACN (36) and 0.1% of FA. Each injection contained 18 µl of salt solution (Figure.1A) which was transferred on to the SCX trap column. Total volume of sample loading (Figure.1B) was 40 µl (i.e. 18 µL of salt and 22 µL of solvent A) to ensure an extensive washing. The total analysis time for the 2D short experiment amounted approx. to ~420 min while for the long 2D experiment to ~1260 min.

Mass Spectrometric Conditions

Mass spectra were acquired with an Orbitrap Q-Exactive mass spectrometer (Thermo Scientific, San Jose, CA) in a data-dependent mode, with automatic switching between MS and MS/MS scans using a top 10 method, where the 10 most abundant precursors were chosen for fragmentation in every MS scan. MS spectra were acquired in positive mode at a resolution of 35000 with a scan range going from 350 to 1500 m/z. For the full scans the AGC target was set to 3x106 ions and maximum injection time to 250 ms. The precursor ions selected for MS2 scans were then fragmented by high-energy collision dissociation (HCD) with the energy set at 25 NCE. Ion selection was performed at 1.5 m/z and the intensity threshold was set to 4.2 x 103 with charge exclusion of z=1 ions. The MS/MS spectra were acquired with fixed first mass of 180 m/z, resolution of 17500, AGC value of 5x104 ions and maximum injection time of 120 ms. The dynamic exclusion varied in each method due to the chromatographic performance. Increasing gradients generated wider peak widths, thus in each method the dynamic exclusion was changed accordingly to the average peak width. For the methods with 45, 60, 90, 180, 240, 360 480 and 600 minute analysis times, the dynamic exclusion was set to 10, 15, 20, 30, 40, 40, 40, 50 and 50 s respectively.

Data Analysis

Each raw data file was processed and quantified by Proteome Discoverer (version 1.3.0.339, Thermo Scientific). Top N Peaks filter was selected, where the 10 most abundant peaks in a mass window of 100 Da alongside a signal-to-noise threshold of 1.5 were parsed. All generated peak lists were searched using Mascot software (version 2.4.1 Matrix

Science). Data were searched against the UniProt-SwissProt version 2010-12 for Human. The database search was performed with the following parameters: a mass tolerance of ±50 ppm for precursor masses; ±0.05 Da fragment ions, allowing two missed cleavages and cysteine carbamidomethylation as fixed modification. Methionine oxidation, N-terminal Acetylation, phosphorylation on serine, threonine and tyrosine were set as variable modifications. The scoring of phosphorylation sites of the identified phosphopeptides was performed by the phosphoRS (37) algorithm (version 2.0) implemented in Proteome Discoverer. The enzyme was specified as trypsin while the fragment ion type as ESI-QUAD-TOF. Percolator (38) calculated the target FDR with a strict cut-off of 0.01. The identified and quantified peptides were first filtered for high confidence (FDR below 1%) then all the results were combined and further filtered with the following criteria: Mascot ion score of at least 20 on peptides and proteins, maximum peptide rank 1, maximum search engine rank 1 and pRS isoform confidence probability of at least 75%. Precursor ion area detection node was added for the 1D 180 and 600 min triplicates and the 2D experiments (1260 min and 420 min total analysis times), in order to obtain the area under the curve of the proteins as sum of the three most intense peptides of a given protein. (39, 40) The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://www.proteomexchange.org) via the PRIDE partner repository (41) with the dataset identifier PXD000705.

3. RESULTS AND DISCUSSION

Evaluation of the performance of the EASY-spray platform in a 1D workflow with back-flush configuration

The EASY-spray technology (42) consists of a plug a play source and a single piece that contains an integrated (and disposable) analytical column, column heater and emitter (Figure.1). The configuration used in this work contained a 50 cm, 75 µm ID analytical column (Acclaim PepMap RSLC C18, 2 µm) that was kept at 30oC. In order to decouple the SCX from the RP analytical column on the commercial EasyLC system, and therefore increase the flexibility of our 2D system, we opted to use a valve switching configuration requiring the RP trap column to operate in back-flush mode (Figure.1), which is already a well-established configuration in proteomic workflows. (43–45) We took a human cell lysate digest to benchmark the performance and assessed variable gradient times ranging from 22 to 574 min (see materials and methods). All these 1D analyses were performed in triplicate. The results as given by the number of unique peptides and proteins detected are represented in Figure.2 and the Supplementary Table.1. For instance, using a 157 minute gradient (~180 min analysis time) we identified, on average, 14575 peptides per run and 3520 proteins, while extending the gradient time to 574 min (~600 min analysis time) led to an increase in identifications to 24367 peptides and 4626 proteins. The data, including column peak capacity, follows closely the results by Kocher et al., (46) unsurprising since the column in the EASY-spray platform is constructed of the same packing material, by the same manufacturer and has the same dimensions. Moreover, these results are in line with recently reported data of alternative UHPLC designs (18, 19, 42, 46) from which we conclude that 1D back-flush configuration shows a highly competitive performance. We further tested the stochastic nature of data dependent acquisition (DDA) in MS analysis, wherein peptide fragmentation analysis is triggered on the most abundant ions in the full scans. This behavior is evident in our data, as when combine the triplicate analysis of the ~180 min runs (i.e.

~540 min total analysis time), we increased our coverage to a total of 24367 peptides and 4369 proteins. Similarly 37805 peptides, and 5646 proteins could be detected combining the data of three 600 min runs (i.e. 1800 min total analysis time) (Supplementary Table.2). As also shown by us (18) and others (12, 20, 46) and visualized by our data in Figure 2, in a 1D LC-MS/MS approach using ultra-long gradients the amount of detected "new" unique peptides and proteins levels off rapidly, certainly beyond 214 min of gradient time in our system. These data indicate that other dimensions of separation are needed to increase the proteome coverage.

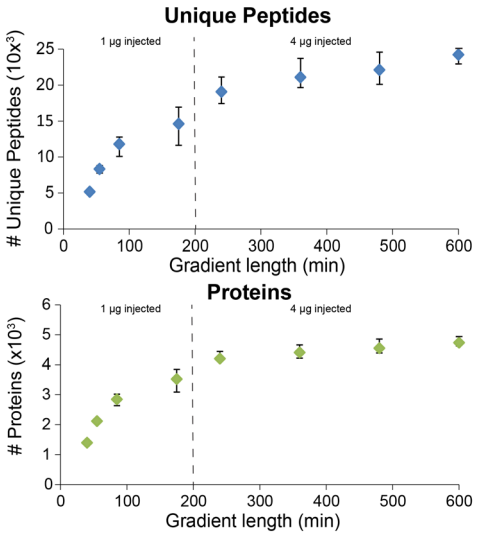


Figure.2: *Average number of peptides and proteins identified in the 1D triplicates. The top graph reports the number of unique peptides identified (Y axis), an average of three replicates, for each of the gradient lengths applied (X axis). The graph on the bottom represents the average number of unique proteins identified, from three replicates, for each applied gradient length.*

Evaluation of the EASY-spray platform in a 2D SCX-RP workflow

Our aim was to design an automated online multidimensional UHPLC workflow, requiring minimal facile changes, having competitive performance in parameters such as proteome depth, total analyses time and sensitivity. In our design no extra pumps are needed to supply the salt solution as the plugs are injected directly from the auto-sampler (Figure.1A). In this system, switching from the 1D to 2D setup is straightforward, as just an SCX trap column is connected via a nanoViper fitting using a zero dead volume connector (dashed box Figure.1). After an initial evaluation we found that the optimal salt plug concentrations and compositions were consistent with previous MudPIT experiments, (20, 30, 35) unsurprising since we were utilizing an SCX column of similar composition and dimensions. To further evaluate the system we used aliquots (10 µg) of the same HEK293 digest. The 2D configuration was tested with two types of analyses; one set were aimed at evaluating the performance when employing short (37 min) second dimension RP gradients, while the second set of experiments used longer RP gradients (157 min). Total analysis times for the two 2D SCX-RP experiments were ~420 and ~1260 min, respectively. The MS analyses of the flow through revealed relatively few regular unmodified tryptic peptides indicating good retention by the SCX material (Figure.3). Moreover, a more careful observation of the peptides eluted in the flow through

and 5 mM salt plug, showed that the majority contained either an acetyl group on the N-terminus or a phosphate group (Supplementary Figure.1), in agreement with earlier observations by Alpert et al. (47) Salt plugs of 10, 20 and 50 mM ammonium acetate produced highly complex peptide mixtures, causing identification rates to be essentially controlled by the length of the reversed phase gradient. The number of unique peptides and proteins identified per fraction, and cumulatively, are summarized in Figure.3. Each subsequent salt plug produced diminishing returns. Pleasingly,

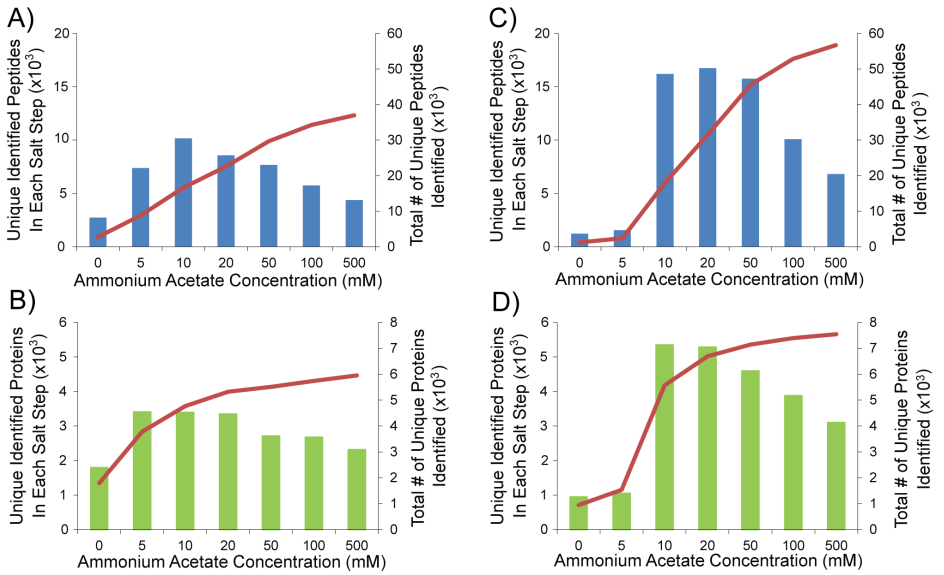


Figure.3: *Total and per fraction number of unique peptides and proteins identified in the 2D experiments. (A) The plot shows for the 2D short experiment (420 min of analysis time) the number of unique peptides identified for each salt plug injected (0 mM to 500 mM) and with the red line the cumulative number of unique peptides identified. (B) The number of unique proteins for each salt plug and the cumulative number of unique proteins identified is presented for the 2D short experiment. (C) The number of unique peptides for each salt plug and the cumulative number of unique peptides identified is shown for the 2D long experiment (1260 min analysis time). (D) The number of unique proteins for each salt plug and the cumulative number of unique proteins identified is shown for the 2D long experiment.*

we observed that the peak capacity of the RP column (above 400 for 157 min gradients using peak widths at 4σ or at 13.4% peak height) remained intact after the injection of salt plugs and typical for this material and column dimensions. (17) Peptides identified in multiple fractions, numbered no more than 20% of any fraction, suggesting a mild issue with carryover or poor fractionation. Based on this data, we attempted a third experiment where we chose a gradient length that is linked to the complexity of the salt plug. We chose a 37 min gradient for the 0 and 5 mM fractions, 214 min for the 10, 20, 50 and 100 mM fractions and 157 min for the 500 mM salt plug, for a total analysis time of ~1260 min. Unfortunately, optimal gradient for each fraction had a negligible effect on the end result (Supplementary Figure.2). In the short 2D experiment we identified 36943 unique peptides and 5958 proteins while with the long 2D experiment we identified 56600 unique peptides and 7565 proteins (Figure.3, Supplementary Table.3). These results are not only favorable in total analysis time compared to the triplicate 1D LC-MS/MS analyses (~540 and ~1800

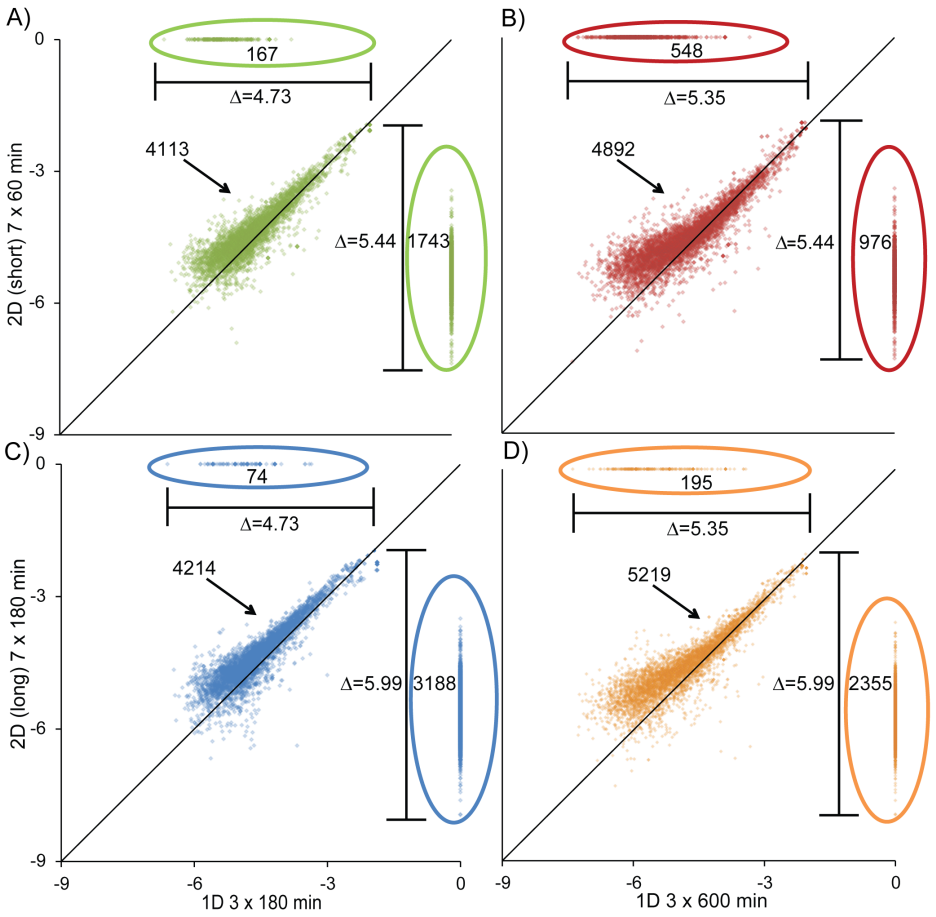


Figure.4: *Normalized protein intensities (log10) for the 2D and 1D combined triplicate experiments. The plots show on the Y axis the normalized intensity (log10) of proteins quantified with the 2D short (A and B) and long (C and D) experiment and on the X axis the normalized protein intensities on the logarithmic scale for the 1D 180 (A and C) and 600 min (B and D) combined triplicates. The clouds in the middle of the graphs are the overlapping quantified proteins among the 2D and 1D experiments. On the X axis (top part) the graphs show the number of proteins uniquely quantified by the 1D 180 min (A and C) and 600 min (B and D) combined triplicates while on the Y axis (right side) the number of proteins only quantified with the 2D short (A and B) and long (C and D) experiment are shown. The horizontal bars illustrate the dynamic range of the 1D 180 min (A and C) and 600 min (B and D) triplicate experiments while the vertical the dynamic ranges for the 2D experiments.*

min) described above, but also in the achieved proteome depth. Using ~1260 min of total analysis time our 2D LC-MS/MS approach identifies 34% more proteins than the ~1800 min triplicate 1D LC-MS/MS approach. See Supplementary Table.4 for an overview of all results in the context of sample amount, time and configuration. We next investigated the overlap in identified peptides comparing one-to-one the cumulative data of the three short and long analysis time triplicate 1D experiments, with the short and more extended 2D experiments, respectively (Supplementary Figure.3). The choice of 1D runs was based on analysis time. We not only identified, in the extended 2D LC ex-

periment, the majority of proteins and peptides found in the two 1D experiments, but we also significantly increased the number of peptides and proteins identified (Supplementary Figure.3C,D,G,H). This performance of the longer gradient time 2D experiment is likely due to the combined effect of fractionation and longer RP gradient used i.e. a larger effective peak capacity. It is worthy to note that even the short version 2D experiment could rival with the long-gradient 1D combined triplicates (Supplementary Figure.3 A, B, E, F). To further demonstrate the advantageous effect of the online multidimensional UHPLC separation we next investigated the proteins abundances as reflected by the ion currents of their top-3 peptides. (39, 40) In Figure.4 the intensities of the proteins are plotted as extracted from the data out of the 2D and 1D experiments. The dots at the top represent proteins only observed in the 1D experiment while the dots at the right side of the plots represent proteins uniquely identified in the 2D experiment. Pleasingly, in general the data points follow a straight line, indicating that the derived protein abundance is in agreement among the different experiments. However, it is clear from this data that high abundant proteins have typical ratios close to 1:1, while most of the medium to low abundant proteins have higher estimated abundance in the 2D experiments, making them appear above the straight line. This observation is likely the results of the higher resolution obtained in the 2D approach, which decreases ion suppression elevating (48, 49) ion currents for peptides co-eluting with high abundant ones in the 1D workflow. Figure.4 also provides the dynamic range of the different experiments calculated as the difference in the intensity between the highest abundant protein and the lowest. As expected the highest dynamic range belonged to the long 2D experiment followed by the short 2D one. The 1D 600 min triplicate showed a similar dynamic range to the short 2D experiment despite its much longer analysis time, while 1D 180 min triplicate performed poorest. It seems that a single 2D experiment of similar analytical time outperforms a set of 1D experiments. Moreover, the mass spectrometer itself is a key component which heavily influences overall proteome performance. Recent improvements in mass spectrometers such as those found in the Orbitrap Fusion (50) will also have a major influence. We envisage that these mass spectrometers will equally benefit (in terms of proteome coverage boost) when the peak capacity of the separation is increased by switching from 1D to 2D LC setups.

4. CONCLUSION

Here, we demonstrate that the EASY-spray system is a powerful and versatile platform. We report on a plug and play online multidimensional SCX-RP UHPLC system that does not require additional pumps for the elution of the salt plugs and is capable of identifying over 7500 proteins when using a Q-Exactive. A major benefit of our setup is that minimal changes are required to modify the 1D LC-MS system. Therefore, we believe this system can be easily constructed by any proteomics laboratory.

5. ACKNOWLEDGEMENTS

The authors acknowledge all members of the Heck-group. This research was performed within the framework of PRIME-XS, grant number 262067, funded by the European Union 7th Framework Program. Additionally, the Netherlands Organization for Scientific Research (NWO) supported large scale proteomics facility Proteins@Work (project 184.032.201) embedded in the Netherlands

Proteomics Centre is kindly acknowledged for financial support as well as the Netherlands Organization for Scientific Research (NWO) with the VIDI grant for SM (700.10.429).

6. REFERENCES

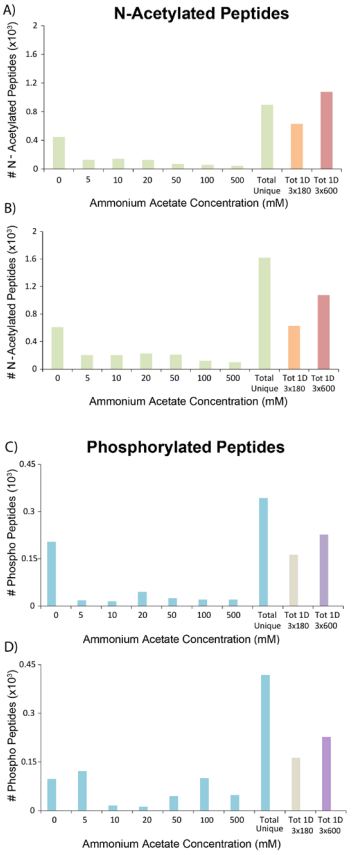
1. Yates, J. R. (1998) Mass spectrometry and the age of the proteome. J Mass Spectrom 33, 1–19
2. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C., and Yates, J. R. (2013) Protein analysis by shotgun/bottom-up proteomics. Chem Rev 113, 2343–2394
3. Zubarev, R. A., and Makarov, A. (2013) Orbitrap mass spectrometry. Anal Chem 85, 5288–5296
4. Michalski, A., Damoc, E., Hauschild, J. P., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., and Horning, S. (2011) Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. Mol Cell Proteomics 10, M111.011015
5. Makarov, A., Denisov, E., Lange, O., and Horning, S. (2006) Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. J Am Soc Mass Spectrom 17, 977–982
6. Olsen, J. V, Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E. R., Senko, M., Makarov, A., Mann, M., and Horning, S. (2009) A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. Mol Cell Proteomics 8, 2759–2769
7. Altelaar, A. F., Munoz, J., and Heck, A. J. (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. Nat Rev Genet 14, 35–48
8. Olsen, J. V, and Mann, M. (2013) Status of large-scale analysis of post-translational modifications by mass spectrometry. Mol Cell Proteomics,
9. Andrews, G. L., Simons, B. L., Young, J. B., Hawkridge, A. M., and Muddiman, D. C. (2011) Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). Anal Chem 83, 5442–5446
10. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A 101, 9528–9533
11. Stoeckli, M., Chaurand, P., Hallahan, D. E., and Caprioli, R. M. (2001) Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. Nat Med 7, 493–496
12. Second, T. P., Blethrow, J. D., Schwartz, J. C., Merrihew, G. E., MacCoss, M. J., Swaney, D. L., Russell, J. D., Coon, J. J., and Zabrouskov, V. (2009) Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures. Anal Chem 81, 7757–7765
13. Jorgenson, J. W. (2010) Capillary liquid chromatography at ultrahigh pressures. Annu Rev Anal Chem (Palo Alto Calif) 3, 129–150
14. MacNair, J. E., Patel, K. D., and Jorgenson, J. W. (1999) Ultrahigh-pressure reversed-phase capillary liquid chromatography: isocratic and gradient elution using columns packed with 1.0-micron particles. Anal Chem 71, 700–708
15. MacNair, J. E., Lewis, K. C., and Jorgenson, J. W. (1997) Ultrahigh-pressure reversed-phase liquid chromatography in packed capillary columns. Anal Chem 69, 983–989
16. Shen, Y., Zhao, R., Berger, S. J., Anderson, G. A., Rodriguez, N., and Smith, R. D. (2002) High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics. Anal Chem 74, 4235–4249
17. Köcher, T., Swart, R., and Mechtler, K. (2011) Ultra-high-pressure RPLC hyphenated to an LTQ-Orbitrap Velos reveals a linear relation between peak capacity and number of identified peptides. Anal Chem 83, 2699–2704
18. Cristobal, A., Hennrich, M. L., Giansanti, P., Goerdayal, S. S., Heck, A. J., and Mohammed, S. (2012) In-house construction of a UHPLC system enabling the identification of over 4000 protein groups in a single analysis. Analyst 137, 3541–3548
19. Thakur, S. S., Geiger, T., Chatterjee, B., Bandilla, P., Fröhlich, F., Cox, J., and Mann, M. (2011) Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. Mol Cell Proteomics 10, M110.003699
20. Zhou, F., Lu, Y., Ficarro, S. B., Webber, J. T., and Marto, J. A. (2012) Nanoflow low pressure high peak capacity single dimension LC-MS/MS platform for high-throughput, in-depth analysis of mammalian proteomes. Anal Chem 84, 5133–5139
21. Yamana, R., Iwasaki, M., Wakabayashi, M., Nakagawa, M., Yamanaka, S., and Ishihama, Y. (2013) Rapid and deep profiling of human induced pluripotent stem cell proteome by one-shot NanoLC-MS/MS analysis with meter-scale monolithic silica columns. J Proteome Res 12, 214–221
22. Iwasaki, M., Sugiyama, N., Tanaka, N., and Ishihama, Y. (2012) Human proteome analysis by using reversed phase monolithic silica capillary columns with enhanced sensitivity. J Chromatogr A 1228, 292–297
23. Gilar, M., Olivova, P., Daly, A. E., and Gebler, J. C. (2005) Orthogonality of separation in two-dimensional liquid chromatography. Anal Chem 77, 6426–6434

24. Gilar, M., Fridrich, J., Schure, M. R., and Jaworski, A. (2012) Comparison of orthogonality estimation methods for the two-dimensional separations of peptides. Anal Chem 84, 8722–8732
25. Di Palma, S., Hennrich, M. L., Heck, A. J., and Mohammed, S. (2012) Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. J Proteomics 75, 3791–3813
26. Washburn, M. P., Wolters, D., and Yates, J. R. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol 19, 242–247
27. Mommen, G. P., Meiring, H. D., Heck, A. J., and de Jong, A. P. (2013) Mixed-bed ion exchange chromatography employing a salt-free pH gradient for improved sensitivity and compatibility in MudPIT. Anal Chem 85, 6608–6616
28. Motoyama, A., Xu, T., Ruse, C. I., Wohlschlegel, J. A., and Yates, J. R. (2007) Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. Anal Chem 79, 3623–3634
29. McDonald, W. H., Ohi, R., Miyamoto, D. T., Mitchison, T. J., and Yates, J. R. (2002) Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. Int. J. Mass Spectrom. 219, 245–251
30. Motoyama, A., Venable, J. D., Ruse, C. I., and Yates, J. R. (2006) Automated ultra-high-pressure multidimensional protein identification technology (UHP-MudPIT) for improved peptide identification of proteomic samples. Anal Chem 78, 5109–5118
31. Fournier, M. L., Gilmore, J. M., Martin-Brown, S. A., and Washburn, M. P. (2007) Multidimensional separations-based shotgun proteomics. Chem Rev 107, 3654–3686
32. Mitulović, G., Stingl, C., Smoluch, M., Swart, R., Chervet, J. P., Steinmacher, I., Gerner, C., and Mechtler, K. (2004) Automated, on-line two-dimensional nano liquid chromatography tandem mass spectrometry for rapid analysis of complex protein digests. Proteomics 4, 2545–2557
33. Nägele, E., Vollmer, M., and Hörth, P. (2004) Improved 2D nano-LC/MS for proteomics applications: a comparative analysis using yeast proteome. J Biomol Tech 15, 134–143
34. Liu, H., Finch, J. W., Luongo, J. A., Li, G.-Z., and Gebler, J. C. (2006) Development of an online two-dimensional nano-scale liquid chromatography/mass spectrometry method for improved chromatographic performance and hydrophobic peptide recovery. J. Chromatogr. A 1135, 43–51
35. Taylor, P., Nielsen, P. A., Trelle, M. B., Hørning, O. B., Andersen, M. B., Vorm, O., Moran, M. F., and Kislinger, T. (2009) Automated 2D peptide separation on a 1D nano-LC-MS system. J Proteome Res 8, 1610–1616
36. BURKE, T. W. L., MANT, C. T., BLACK, J. A., and HODGES, R. S. (1989) STRONG CATION-EXCHANGE HIGH-PERFORMANCE LIQUID-CHROMATOGRAPHY OF PEPTIDES- EFFECT OF NON-SPECIFIC HYDROPHOBIC INTERACTIONS AND LINEARIZATION OF PEPTIDE RETENTION BEHAVIOR. J. Chromatogr. 476, 377–389
37. Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C., and Mechtler, K. (2011) Universal and confident phosphorylation site localization using phosphoRS. J Proteome Res 10, 5354–5362
38. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., and MacCoss, M. J. (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods 4, 923–925
39. Grossmann, J., Roschitzki, B., Panse, C., Fortes, C., Barkow-Oesterreicher, S., Rutishauser, D., and Schlapbach, R. (2010) Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods. J Proteomics 73, 1740–1746
40. Silva, J. C., Gorenstein, M. V, Li, G. Z., Vissers, J. P., and Geromanos, S. J. (2006) Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics 5, 144–156
41. Vizcaíno, J. A., Côté, R. G., Csordas, A., Dianes, J. A., Fabregat, A., Foster, J. M., Griss, J., Alpi, E., Birim, M., Contell, J., O'Kelly, G., Schoenegger, A., Ovelleiro, D., Pérez-Riverol, Y., Reisinger, F., Ríos, D., Wang, R., and Hermjakob, H. (2013) The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res 41, D1063–9
42. Pirmoradian, M., Budamgunta, H., Chingin, K., Zhang, B., Astorga-Wells, J., and Zubarev, R. A. (2013) Rapid and deep human proteome analysis by single-dimension shotgun proteomics. Mol Cell Proteomics,
43. Giambruno, R., Grebien, F., Stukalov, A., Knoll, C., Planyavsky, M., Rudashevskaya, E. L., Colinge, J., Superti-Furga, G., and Bennett, K. L. (2013) Affinity purification strategies for proteomic analysis of transcription factor complexes. J. Proteome Res. 12, 4018–27
44. Shen, Y., Moore, R. J., Zhao, R., Blonder, J., Auberry, D. L., Masselon, C., Pasa-Tolić, L., Hixson, K. K., Auberry, K. J., and Smith, R. D. (2003) High-efficiency on-line solid-phase extraction coupling to 15-150-microm-i.d. column liquid chromatography for proteomic analysis. Anal Chem 75, 3596–3605
45. Chamrád, I., Rix, U., Stukalov, A., Gridling, M., Parapatics, K., Müller, A. C., Altiok, S., Colinge, J., Superti-Furga, G., Haura, E. B., and Bennett, K. L. (2013) A miniaturized chemical proteomic approach for target profiling of clinical kinase inhibitors in tumor biopsies. J. Proteome Res. 12, 4005–17
46. Köcher, T., Pichler, P., Swart, R., and Mechtler, K. (2012) Analysis of protein mixtures from whole-cell extracts by single-run nanoLC-MS/MS using ultralong gradients. Nat Protoc 7, 882–890
47. Alpert, A. J., Petritis, K., Kangas, L., Smith, R. D., Mechtler, K., Mitulović, G., Mohammed, S., and Heck, A. J. (2010) Peptide orientation affects selectivity in ion-exchange chromatography. Anal Chem 82, 5253–5259

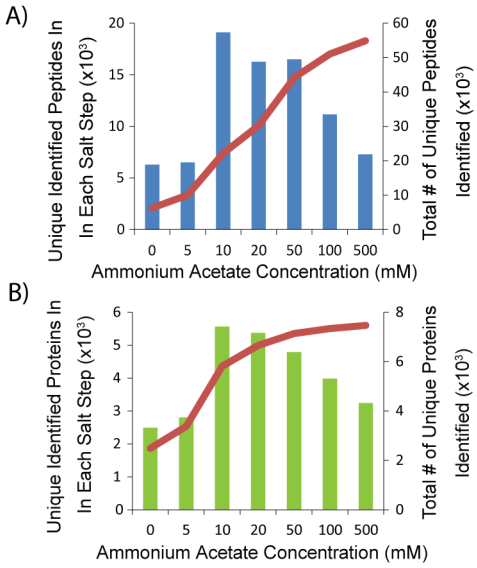
48. King, R., Bonfiglio, R., Fernandez-Metzler, C., Miller-Stein, C., and Olah, T. (2000) Mechanistic investigation of ionization suppression in electrospray ionization. J. Am. Soc. Mass Spectrom. 11, 942–50
49. Schmidt, A., Karas, M., and Dülcks, T. (2003) Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI? J. Am. Soc. Mass Spectrom. 14, 492–500
50. Hebert, A. S., Richards, A. L., Bailey, D. J., Ulbrich, A., Coughlin, E. E., Westphall, M. S., and Coon, J. J. (2014) The one hour yeast proteome. Mol Cell Proteomics 13, 339–347

7. SUPPLEMENTARY

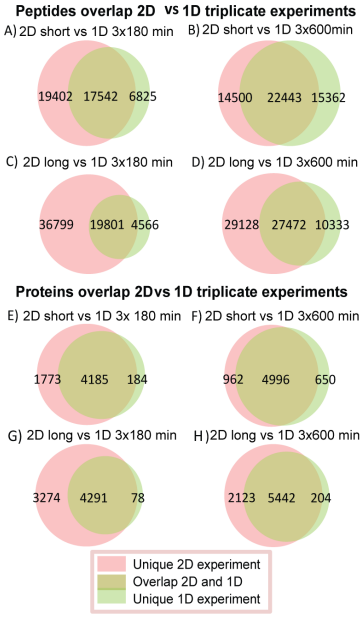
Supplementary Tables.1 to 4 are available as excel files via the internet at http://pubs.rsc.org/



Supplementary Figure.1: *The chart shows in light green N-terminus acetylated peptides identified in each salt step for the 2D short (A) and long experiment (B) and the total unique ones identified. In orange and light red the total unique N-terminus acetylated peptides for the combined triplicates of the 1D 180 and 600 min experiments. In light blue are represented the phosphopeptides identified in each salt step for the 2D short (C) and long experiment (D) and the total unique ones identified. While respectively in gray and violet the total unique phosphopeptides are shown covered by the combined triplicates of the 1D 180 and 600 min experiments.*



Supplementary Figure.2: *Total and per fraction number of unique peptides and proteins identified in the 2D optimized experiment. A) The plot shows for the 2D short experiment the number of unique peptides identified for each salt plug injected (0 mM to 500 mM) and with the red line the cumulative number of unique peptides identified. B) Number of unique proteins for each salt plug and the cumulative number of unique proteins identified.*



Supplementary Figure.3: *Protein and peptide overlap between 2D and 1D combined triplicates experiments. A-B) The 2D short experiment is respectively compared for the number of peptides covered to the 1D 180 and 600 min combined triplicates. While the same comparison in E-F is made for the proteins. C-D) The 2D long experiment is respectively compared for the number of peptides covered to the 1D 180 and 600 min combined triplicates. The same comparison in G-H is made at protein level.*

CHAPTER 3

Universal quantitative kinase assay based on diagonal SCX chromatography and stable isotope dimethyl labeling provides high-definition kinase consensus motifs for PKA and human Mps1

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

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

[‡]Netherlands Proteomics Centre, Padualaan 8, 3584CH, Utrecht, The Netherlands.

[§]Department of Medical Oncology and Department of Molecular Cancer Research, University Medical Center, Strateneum 2.231, Universiteitsweg 100, 3584CG Utrecht, The Netherlands.

[#]These authors contributed equally to this wotk

J Proteome Res. 2013 May 3;12(5):2214-24

ABSTRACT

In order to understand cellular signaling a clear understanding of kinase-substrate relationships is essential. Some of these relationships are defined by consensus recognition motifs present in substrates making them amendable for phosphorylation by designated kinases. Here, we explore a method that is based on two sequential steps of strong cation exchange chromatography combined with differential stable isotope labeling, to define kinase consensus motifs with high accuracy. We demonstrate the value of our method by evaluating the motifs of two very distinct kinases; cAMP regulated protein kinase A (PKA) and human monopolar spindle 1 (Mps1) kinase, also known as TTK. PKA is a well-studied basophilic kinase with a relatively well-defined motif and numerous known substrates in vitro and in vivo. Mps1, a kinase involved in chromosome segregation, has been less well characterized. Its substrate specificity is unclear and here we show that Mps1 is an acidophilic kinase with a striking tendency for phosphorylation of threonines. The final outcomes of our work are high-definition kinase consensus motifs for PKA and Mps1. Our generic method, which makes use of proteolytic cell lysates as source for peptide-substrate libraries, can be implemented for any kinase present in the kinome.

1. INTRODUCTION

Protein phosphorylation is one of the most explored and important post translational modifications (PTM). The level of protein phosphorylation in processes such as metabolism, transcription, cell cycle progression, cytoskeletal rearrangement, and apoptosis is tightly controlled by protein kinases and phosphatases. (1) Not surprisingly, mutations and dysfunctions of protein kinases play causal roles in major human diseases including cancer, immunodeficiencies and endocrine disorders. (2, 3) In order to fulfill their critical roles, kinases need to recognize their substrates among a vast pool of candidates with high specificity. Distal docking sites, scaffolding proteins, allosteric regulation, localization in given cellular compartments, and recognition of consensus phosphorylation motifs are some of the underlying principles for substrate recognition. Of these, the substrate consensus motif is one of the most important. It is defined as the N- and C-terminal amino acid sequence stretch adjacent to the potential site of phosphorylation. (4) In the 1990ies, the identification of in vitro kinase substrate consensus motifs was performed by the use of large oriented synthetic peptide libraries in which the phosphorable serine or threonine amino acids were fixed and the surrounding amino acids were varied. The substrates phosphorylated by the kinase were then separated from the bulk of non-phosphorylated peptides using a ferric iminodiacetic acid column (IDA), and peptide sequences were obtained by Edman sequencing. (5) This method suffered from severe drawbacks including lack in selectivity at the sites of secondary importance in the consensus sequence, negligence of other possible interactions further located from the phospho-site, and incomplete Edman sequencing. When solid-phase in vitro assays could be implemented some of those issues could be resolved. In these methodologies, peptide libraries were immobilized on membranes and incubated with the kinase of interest in presence of radiolabeled ATP. (6) Unfortunately, these strategies exhibited artifacts due to the nonspecific binding of radiolabeled ATP on the solid support and non-covalent association of the kinase with less optimal substrates due to their high local concentrations. (7, 8) A more recent

in vitro approach makes use of synthetic peptide libraries whereby the kinase assay with radiolabeled ATP is performed in solution followed by immobilization. (9) This contributed to a significant improvement in the prediction of in vitro kinase substrate consensus motifs. Emerging methods using high-resolution mass spectrometry (MS) (10–12) for the in-depth analysis of phosphopeptides, further improved the identification of putative kinase targets and thus, yielded to a more extensive comprehension of signaling networks. (13, 14) Still, discovering putative substrates for kinases is challenging. Partially, this is caused by the fact that MS-based identification of phosphopeptides suffers from the sub-stoichiometric level of phosphorylation on peptides compared to their non-modified counterparts. (15) Wide-scale MS-based proteomics approaches have the great advantage to render accessible database-searchable large peptide libraries consisting of natural biological sequences derived from appropriate cellular proteomes. This for instance allowed the determination of protease cleavage site consensus motifs. (16, 17) A similar MS-based approach for determining kinase consensus motifs (18) also made use of peptide libraries allowing protein kinases to choose among thousands of different peptide sequences, targeting only the likely endogenous chemical and positional preferences. This latter method implemented in its workflow several steps to overcome the notorious problems linked to the use of MS-based techniques on complex peptide libraries as represented by a full proteome digest. Not only strong cation exchange (SCX) (19) was used to fractionate and reduce the complexity of the sample, also a phosphatase induced dephosphorylation was implemented to minimize the presence of endogenous phosphate groups. Following the in vitro kinase reaction, a final TiO2 enrichment was used to pull down the phosphopeptides prior to LC-MS/MS analysis. Despite the developments brought in by this approach, once more, the trivial issue of unambiguously assessing endogenously phoshopeptides from the kinase substrates was not completely solved. Here, we evaluate an alternative high throughput LC-MS/MS-based proteomics method for the definition of consensus motifs of protein kinases. Our approach involves the use of two stages of SCX chromatography (equivalent to combined fractional diagonal chromatography; COFRADIC) (20–22) where the separation step is used to distinguish between regular and phosphorylated peptides. Our approach is independent of the choice of peptide mixture and thus one can make a peptide library using any cell lysate and any protease allowing access to diverse libraries of natural peptides. The added flexibility in our workflow will help to increase the chance of generating appropriate substrates and improve the chance of identifying the consensus motifs of structurally very diverse protein kinases, but also to enquire those whose specificity is still unknown. Our strategy does not require phosphatase treatment, since it can distinguish between originally present and newly formed phosphorylation sites. We demonstrate the performance of our approach by first determining the kinase motif of the quite well-characterized cAMP regulated protein kinase A (PKA). Subsequently, we define the preferred substrate motif for phosphorylation by human Mps1 kinase, an essential regulator of chromosome segregation.

2. EXPERIMENTAL SECTION

Materials

Complete Mini EDTA-free protease inhibitor cocktail and phosStop phosphatese inhibitor cocktail were obtained from Roche Diagnostic (Germany), metallo-endopeptidase Lys-N from Seikagaku Corporation (Japan), lysyl endopeptidase (Lys-C) from Wako (Richmond, VA)

and trypsin endopeptidase from Promega (The Netherlands). Iodacetamide, DL-dithiothreitol (DTT), formaldehyde (37% in water), formaldehyde 13CD2 (20 % in D2O) and sodium cyanoborodeuterate were purchased from Sigma Aldrich (Germany). Urea, potassium dihydrogen phosphate (KH2PO4), potassium chloride (KCl) and formic acid (FA) were obtained from Merck KGaA (Germany). Ammonium bicarbonate (AMBIC) and sodium cyanoborohydrate were obtained from Fluka Analytical (Germany), while HPLC-S grade acetonitrile (ACN) was obtained from Biosolve BV (The Netherlands). Bradford protein assay was purchased from BioRad (Germany), high purity water was obtained from a Millipore Milli-Q system (Billerica, MA) and deuterated formaldehyde CD2O (20% solution in D2O) from Isotec (Miamisburg, OH).

Lys-N library preparation for PKA

HEK293 cells were re-suspended in lysis buffer composed of 8M urea, 50 mM AMBIC pH 8, 1 tablet of PhosStop phosphatase inhibitors, and 1 tablet of cOmplete Mini EDTA-free protease inhibitor cocktail. The cell lysate was sonicated 3 times on ice. Subsequently, centrifugation at 20000 g at 4°C for 20 minutes separated the soluble from the insoluble protein fraction. The soluble fraction was collected and the protein concentration of the lysate was determined by a Bradford protein assay. Proteins were reduced with 2 mM DTT at 56°C for 25 minutes, followed by alkylation with iodacetamide (4 mM) at room temperature for 30 minutes in the dark. An incomplete digestion was carried out for 4 hours at 37°C using a substrate to Lys-N ratio of 100:1 (w/w). (23) After digestion the sample was acidified with 10% FA, then desalted using Sep-Pak 1CC 50 mg C18 cartridges (Waters Corporation, Milford, MA), and dried down. Dried peptides were re-suspended in 10% FA before the SCX separation.

Trypsin library preparation for Mps1

Hela S3 cells were digested as described for the PKA experiment with a few modifications. After the alkylation step the digestion was carried out with a first step of lysyl endopeptidase (Lys-C) for 4 hours at 37°C with a protein to enzyme ratio of 75:1 (w/w). Subsequently, the sample was diluted 4 times with 50 mM AMBIC to a urea concentration of 2 M. The second step of digestion was performed with trypsin overnight at 37°C with a substrate to enzyme ratio of 100:1 (w/w). The resulting solution was acidified with 10% FA to quench the reaction, desalted on a Sep-Pak 1CC 50 mg C18 cartridge and the eluate was dried down. Dried peptides were re-suspended in 10% FA before the SCX separation.

SCX separation before the kinase assay

All SCX experiments were performed on an Agilent 1100/1200 HPLC system (Agilent Technologies, Germany). Peptides from the digests corresponding to 1 mg of protein material for the PKA experiment and approx. 3 mg for the Mps1 experiment were loaded onto a C18 trap column (strata-x 33 µm Polymeric Reversed Phase, 50x4.6 mm, Phenomenex, The Netherlands) for 5 minutes at 300 µL/min using aqueous 0.05% FA as solvent. Subsequently, peptides were eluted for 5 minutes from the trapping column with 80% ACN containing 0.05% FA onto a PolySULFOETHYL A column 200 × 2.1 mm, 5 µm particles, and 200 Å pore size (PolyLC Inc., Columbia, MD) at the same flow rate. Separation was performed using a nonlinear 65 min. gradient: isocratic for 2 min. at 100% solvent A (5 mM KH2PO4, 30% ACN and 0.05% FA, pH 2.7); from 2 to 10 min. at 3% solvent B (5 mM KH2PO4, 30% ACN, 350 mM KCl and 0.05% FA at pH 2.7); from 10 to 40 min. a gradient to 35% solvent B, and from 40 to 45 min. to 100% solvent B. The column was subsequently washed for 10 min. with solvent B and finally equilibrated with 100% solvent A. Fractions were collected in 1 min. intervals for the

first 40 min. and for 3 min. intervals in the last 15 minutes. After evaporation of the solvents, fractionated peptides were resuspended in both cases in kinase buffer as described below.

Protein expression and purification of PKA

The cAMP-dependent PKA catalytic subunit (New England Biolabs) had a concentration of 2500000 U/mL and an activity of 5000000 U/mg. It was supplied in 50 mM NaCl, 20 mM Tris-HCl at pH 7.5, 1 mM Na2EDTA, 2 mM DTT and 50% of glycerol.

Protein expression and purification of Mps1

P2 baculovirus (Invitrogen) was used to facilitate direct transfer of the gene of interest in vitro into SF9 cells. pFASTBAC plasmid was used to generate GST-Mps1-wild-type (WT) and –kinase-dead (KD: D664A). The enzymes were then purified using glutathion beads. Detailed procedures are described in. (24)

In vitro PKA phosphorylation

Selected fractions from the initial SCX separation were pooled together in order to form 6 samples containing peptides of net charge from +2 to +7. These pools were de-salted using Sep-Pak cartridges followed by drying down and re-suspension in 45 µL of 100 mM Tris-HCl and 20 mM MgCl2. The 6 samples were split in two, one half would be used for the in vitro assay and the other half as a control for the assay. 25 µL of kinase reaction buffer composed of 10 mM Tris-HCL, pH 7.5, 20 mM MgCl2 and 1 µL of the PKA solution containing 2500 U of kinase were added to both aliquots, while 20 mM ATP was added only to the in vitro aliquots, whereas no ATP was added to the control aliquots. All the samples were incubated at 37°C for 10 minutes and subsequently, the reaction was quenched with 10% FA.

In vitro Mps1 phosphorylation

Selected fractions from the initial SCX separation were selected and pooled together in order to form 3 samples containing peptides of net charge from +2 to +4. The 3 pools were de-salted on a Sep-Pak cartridge, dried down and then re-suspended in 30 µL of 500 mM Tris-HCl, 30 µL of 100 mM MgCl2, 30 µL of 10 mM DTT and 180 µL of Milli-Q water. The reconstituted samples were split into three. A third of the sample was used for the in vitro assay with wild-type Mps1, a second third with the Mps1 kinase-dead mutant and the last aliquot with wild-type Mps1 without addition of ATP. Essentially, the two latter aliquots are expected to contain no phosphorylation activity and are controls. The kinase reaction was performed by adding 10 µL of 2 mM ATP (or 10 µL of Milli-Q water for the control without ATP), 1µL of wild-type kinase solution to one sample, 1µL of kinase-dead solution to another sample and 1 µL of wild-type kinase solution to the sample without ATP. The aliquots were incubated for 30 minutes at 37°C and then quenched with 10% formic acid.

On column stable isotope dimethyl labeling

The on column dimethyl labeling of the samples was performed as described previously. (25) The PKA sample was labeled with the light dimethyl label and the control with the intermediate label. For the Mps1 experiment the samples containing wild-type kinase, wild-type kinase without ATP and the kinase-dead mutant were labeled with the heavy, intermediate and light label, respectively. The eluates of the PKA and Mps1 samples were combined according to the SCX pool of origin at a 1:1 and 1:1:1 ratio, respectively.

SCX separation following the *in vitro* kinase assays

In both experiments each combined labeled sample was subjected to SCX separation as already explained for the first SCX experiment. Newly formed phosphopeptides would elute earlier, due to their decrease in net charge, than peptides that were not phosphorylated by the kinases. Selected fractions were desalted, dried down and then resuspended in 10% FA.

LC-MS/MS analysis

Selected fractions from the second SCX separation were subjected to nanoLC-MS/MS analysis using an LTQ Orbitrap Velos Mass Spectrometer equipped with an electron transfer dissociation (ETD) source (Thermo Fisher, Germany). Mobile phase buffers for LC separation consisted of 0.1 M acetic acid in water (buffer A) and 0.1 M acetic acid in 80% (buffer B). For the Mps1 experiment trapping and washing of the sample was performed at 5 µL/min for 10 min with 100% buffer A. The gradient for the separation was from 10 to 25% of buffer B within 107 min. and from 25 to 50% in the following 35 minutes. After the gradient, the column was washed by increasing the buffer B concentration to 100% followed by conditioning the system with 100% buffer A for at least 15 minutes. For the PKA samples a total analysis time of 90 min. was chosen. The gradient ranged from 0 to 30% buffer B in 61 min., then for 3 min. from 30% to 100% buffer B, followed by washing and conditioning the column. The flow rate over the column was kept at 100 nL/min and the column effluent was directly introduced into the ESI source of the MS using an in-house pulled fused silica emitter gold-coated, biased to 1.7 kV. The mass spectrometer was operated in positive ion mode, from 350 to 1500 m/z in MS mode. We carried out the analysis with a data-dependent mode with MS operating at a resolution of 30,000. MS/MS dependent data were acquired with a Top5 method, where each precursor was fragmented with both HCD and ETD (a total of 10 MS/MS events). Charge state screening was enabled and precursors with unknown charge state or a charge state of 1 were excluded. The normalized collision energy was set to 40% for HCD. The resulting fragments were detected with a resolution of 7500 in the Orbitrap. Parent ions were sequentially isolated to a target value of 5,000 and fragmented in the linear ion trap collision cell by ETD and supplemental activation was enabled. The ETD reagent target value was set to 200,000 and the reaction time to 50 ms. The resulting fragments were detected in the LTQ and dynamic exclusion was enabled (exclusion size list 500, exclusion duration 60 s).

Data Analysis

Each raw data file was processed and quantified by Proteome Discoverer (version 1.3, Thermo Scientific). Top N Peaks filter was selected, where the 10 most abundant peaks in a mass window of 100 Da alongside a signal-to-noise threshold of 1.5 were parsed. ETD non-fragment filter was also implemented with the precursor peak removal asset to 4 Da, charge-reduced precursor removal to 2 Da, and removal of known neutral losses from the charge-reduced precursor to 2 Da and a maximum neutral loss mass set to 120 Da. Only high resolution (HCD) MS/MS spectra were charge de-convoluted with default options using a node of Proteome Discover called “MS2 Spectrum Processor”. All generated peak lists were searched using Mascot software (version 2.3.02 Matrix Science). Both Mps1 and PKA data were searched against the Uniprot-Human database. The database search was performed with the following parameters: a mass tolerance of ±50 ppm for precursor masses; ±0.8 Da for ETD-ion trap fragment ions; ±0.08 Da for HCD-Orbitrap fragment ions, allowing two missed cleavages, and cysteine carbamidomethylation as fixed modification. Light, in-

termediate and heavy dimethylation were set as a variable modification on peptide N-termini and lysine residues in the case of the Mps1 experiment and only light and intermediate dimethylation for PKA. Methionine oxidation, phosphorylation on serine, threonine and tyrosine were set as variable modifications. The enzyme was specified as Lys-N (metalloendopeptidase) in the case of PKA and trypsin for Mps1. The fragment ion type was specified as electrospray ionization ETD-TRAP, and ESI-QUAD-TOF. The scoring of phosphorylation sites of the identified phosphopeptides was performed by the phosphoRS algorithm (version 2.0) (26) implemented in Proteome Discoverer. The dimethyl-based quantitation method was chosen in Proteome Discoverer, with a mass precision requirement of 2 ppm for consecutive precursor measurements. We also allowed spectra with 1 missing channel to be quantified in the case of PKA and 2 missing channels for Mps1. We enabled the options for replacing the missing quantification values with minimum intensity which resulted to be fundamental in our experiment. Percolator (27) calculated the target FDR with a strict cut-off of 0.01. After identification and quantification, we combined all results and filtered them with the following criteria: only the PSMs with a site localization probability for the phospho-group of at least 0.75 were selected, Mascot ion score of at least 20, maximum peptide rank 1, high peptide confidence, lowest peptide length 7 and highest unlimited, labeling ratio of at least <0.2 intermediate/light (I/L) in the case of PKA and ratio <0.2 light/heavy and intermediate/heavy (L/H and I/H) for Mps1. If multiple PSMs showed the same sequence then only the one with higher Mascot ion score and pRS site localization probability where selected.

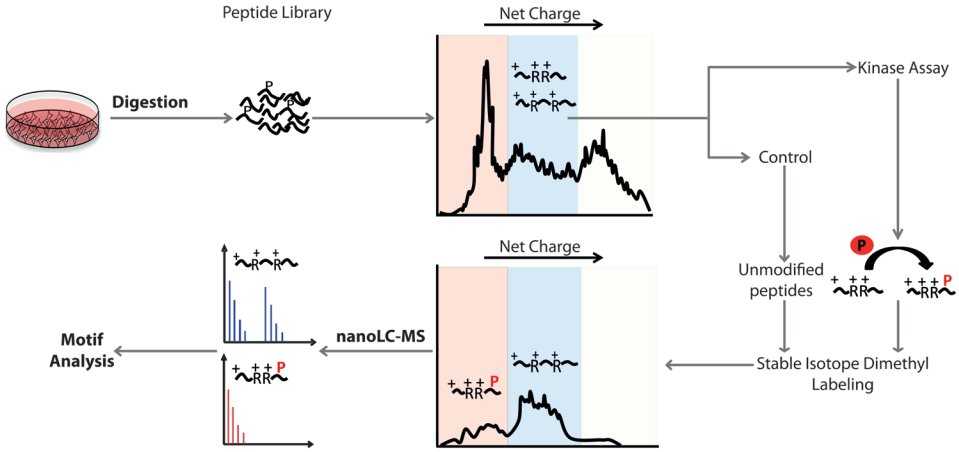


Figure.1: *Experimental workflow. Natural peptide libraries optimized for the individual kinases are created using different proteases. The resulting proteolytic peptides are separated by SCX chromatography. Fractions containing peptides with the same net charge are pooled together and then split to create aliquots for the controls and the in vitro kinase assay. After the assay, the peptides are labeled by stable isotope dimethyl labeling. Corresponding labeled peptides are pooled back together and are separated again by SCX chromatography. Unmodified peptides elute with identical retention times, while the newly phosphorylated peptides will elute earlier due to their decreased net charge after addition of the phosphate group. Selected SCX fractions are then analyzed by nanoLC–MS/MS. Targets of the kinase will show a high ratio or only the channel of the kinase assay. The putative targets are analyzed by IceLogo to retrive the sequence motif of the kinase.*

Motif Analysis

IceLogo (28) was used to obtain the most probable consensus motif of phosphopeptides identified in the in vitro kinase assays. Briefly, IceLogo calculates the chance of

occurrence (p) of every amino acid in every position in the experimental set. Our experimental set of phosphopeptides was created aligning those at their common phosphorylation site. The reference set for calculating the occurrence was the Uniprot complete proteome set. IceLogo uses statistics to find over- and under-represented amino acids of the experimental dataset compared to the background database in each position flanking the phosphorylation-site. A confidence interval can be set using specific p-values, which in our case was set to 0.001 in order to have a high certainty for the determination of the most likely phosphorylation motifs of the enzymes. In addition, IceLogo was used for visualizing manually selected subsets of peptides.

3. RESULTS AND DISCUSSION

Current proteomics strategies which exploit peptide libraries originating from natural sources for finding consensus motifs of protein kinases often still face interferences from endogenously phosphorylated peptides. In the here presented strategy, we diminish this issue, which enables us to define in vitro kinase motifs with high precision. Unique to our approach is the controlled generation of peptide libraries from natural sources by using an appropriate choice of protease for the kinase being studied. Therefore, our approach is suitable to study kinases with unknown specificity by combining peptide libraries, generated by using a range of proteases, enabling the coverage of more proteome space. (29) The created peptide libraries were first separated based on their net charge by SCX chromatography at pH 3. Importantly, at pH 3 carboxyl groups are predominantly protonated and have only a minor influence on the net charge, whereas phospho-sites on the peptides are still negatively charged, thus reducing the net charge of peptides by one in comparison to their unphosphorylated counterparts. Thus, the difference in net charge state enables the separation of phosphopeptides from their non-phosphorylated counterparts. (30–32) Libraries of peptides for the in vitro kinase assays were selected from the SCX fractions, discarding the SCX fractions that in our set-up contain mainly phosphopeptides. Fractions carrying peptides with the same net charge were pooled together and subsequently split in equal parts for in vitro treatment with the kinase and for the control. Prior to mixing the peptides from the kinase assays and controls, they were differentially isotopically dimethyl labeled. The labeling represents one of the key-steps in our workflow enabling to distinguish unambiguously kinase targets from endogenously phosphorylated peptides. After the kinase assay is performed, a second round of SCX chromatography is performed. Newly formed phosphopeptides will have a lower net charge, and will thus elute earlier. This enables the separation from the bulk of peptides. Phosphopeptides that were originally present will not have their net charge changed and so will also elute with the bulk of peptides. Thus, the second SCX separation and the dimethyl labeling will allow us to clearly identify newly formed phosphopeptides. Selected SCX fractions were then analyzed by LC-MS/MS. The resulting data were then filtered for several criteria including confidence of site assignment and observed ratio. Our exact criteria can be found in the material and methods section. In order to identify enrichment of selected amino acids at positions surrounding the phosphorylation site, IceLogo was used as a predictive tool, where the data-set of phosphopeptides was compared to a background dataset based on the whole proteome (Figure.1).

In vitro kinase substrate assay for the cAMP regulated protein kinase A

In an effort to validate our approach we studied cAMP regulated protein kinase A (PKA).

It is a well-known lysine/arginine directed serine/threonine kinase. Over 160 physiological PKA substrates have already been described mostly confirmed by thorough biochemical assays and by using strict criteria of eligibility, e.g. the phosphorylation of the targeted protein was confirmed both in vitro and in vivo. Despite the already high number of substrates described, it has been hypothesized that these represent only 1% of the total pool of substrates of PKA. (33, 34) Nevertheless, PKA represents a kinase that is difficult to target by in vitro kinase assays and MS-based proteomics for a number of reasons. PKA targets basophilic substrates rich in basic residues, a challenge for SCX separations. Furthermore, trypsin is not an ideal enzyme to create a peptide library which contains substrates of PKA, since trypsin cleaves specifically at lysine or arginine, which are part of the predicted phospho-site motif of PKA. We argued that using trypsin would eliminate a vast number of potential substrate peptides. The need of a proteolytic enzyme able to generate PKA suitable libraries containing relatively long and more basic peptides amendable for LC-MS/MS analysis, directed our attention towards Lys-C or Lys-N.

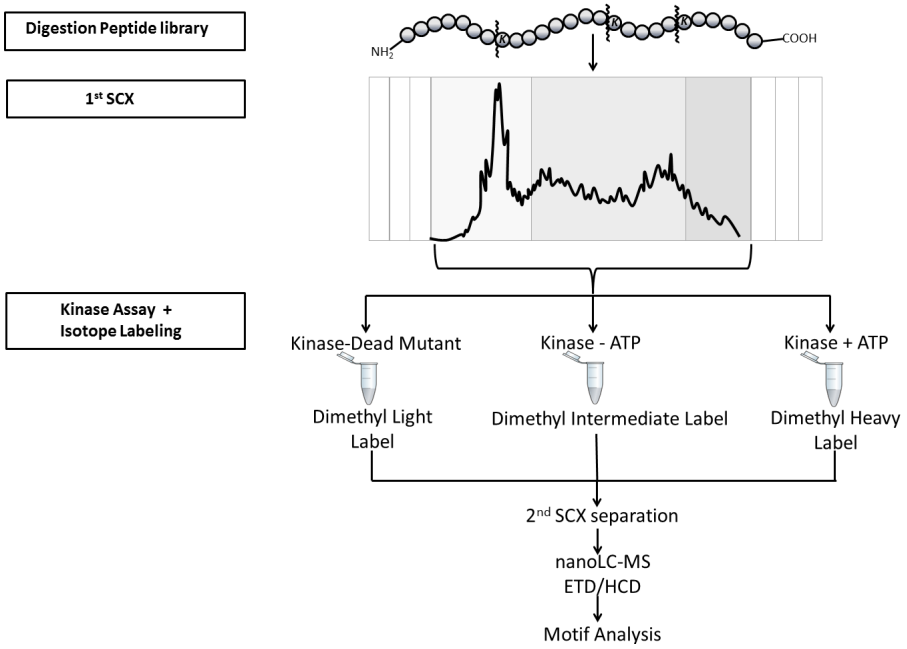


Figure.2: *Workflow for the in vitro kinase assays of PKA and Mps1. Lys-N and trypsin are used to create a suitable peptide library for PKA and Mps1, respectively. Peptides are separated by SCX chromatography and the fractions containing mainly peptides with a net charge of +2, +3, +4, +5, +6 and +7 for PKA and +2, +3 and +4 for Mps1 are pooled together. The pools are then split in two aliquots for the control and the kinase assay with PKA and in three aliqouts for the two controls and the Mps1 kinase assay. Subsequently, the samples are differentially dimethyl labeled. Light and intermediate labels are used for the PKA kinase assay and the assay without addition of ATP (control), while light, intermediate and heavy isotope lables are used for the kinase-dead Mps1 control experiment, the wild-type Mps1 assay without addition of ATP (control 2) and the wildtype Mps1 kinase assays, respectively. Next, samples were mixed again and separated by SCX, enabling the separation of phosphorylated peptides from their unmodified counterparts. Enriched phosphofractions are analyzed by LC–MS/MS using both ETD and HCD. After database searching and filtering the list of phosphopeptides is processed with IceLogo for decifering the amino acid distribution on the surrounding positions of the phospho-site, generating a high-precision consensus motif*

We chose Lys-N due to its cleavage specificity at the N-terminus of lysine, which will lead to a slightly higher number of potential substrates when compared to Lys-C. Furthermore, we performed a sub-optimal digestion allowing miss-cleavage formation and increasing the propensity of RKXS(T), KKXS(T) etc. sequences. The generated proteolytic peptides were subsequently separated by SCX, where only the late fractions were further used, since they contain mainly highly basic peptides. Fractions, that would contain the same net charge, were pooled together. Prior to the in vitro kinase assay, the activity of PKA was tested with its proto-typical substrate kemptide (data not shown). The pools were split in two, half for the control reaction without the addition of ATP, and half for the in vitro kinase assay. After the assay, the peptide pools were intermediate and light dimethyl labeled for the control and for the assayed aliquots, respectively. Post incubation and labeling, the pools were appropriately mixed together again and concentrated in a vacuum centrifuge. The reconstituted peptide pools were once again fractionated by SCX and subjected to analysis by LC-MS/MS. Considering we were expecting our newly formed phosphopeptides to be highly basic and highly charged, we opted for the use of both ETD and HCD for sequencing (12) (see Figure.2 for outline of workflow). As expected, we identified the majority of newly formed phosphopeptides from SCX pools corresponding to a higher net charge i.e. multiple basic residues. Moreover, almost all identified putative PKA substrate peptides were identified by using ETD. An example of MS and MS/MS spectra of a putative PKA target is shown in Figure.3. From the analysis we could quantify over 2100 peptides (Supplementary Table.1) and more

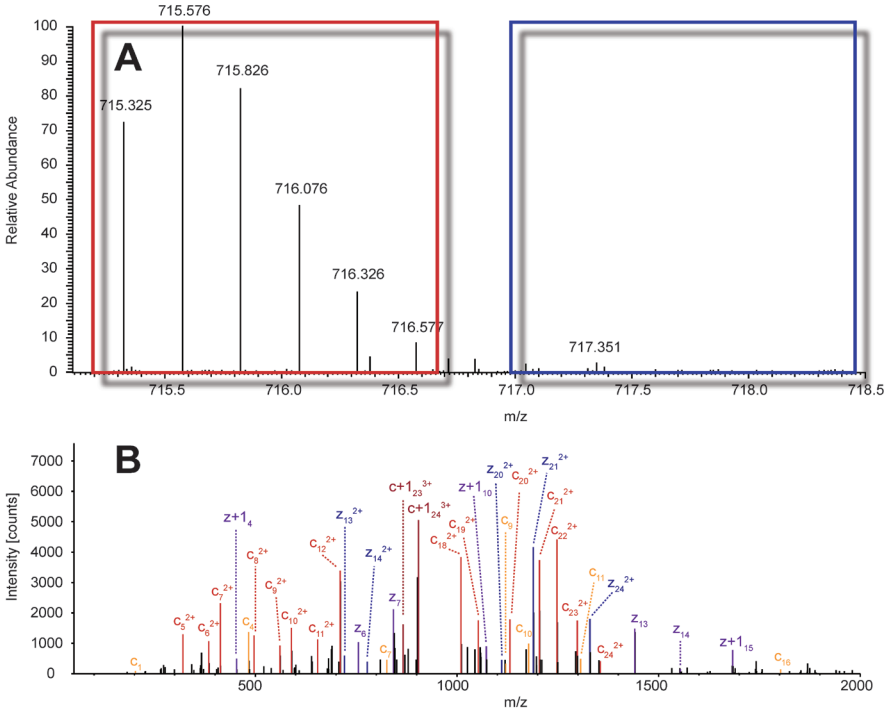


Figure.3: *MS and MS/MS spectra of an in vitro phosphorylated peptide. (A) MS spectrum of the putative PKA target KSPPREGpSQGELTPANSQSR(O)MSTN(O)M. In the red box the light labeled phosphorylated peptide is observed, whereas the corresponding intermediate labeled peptide (in the blue box) is absent. (B) Corresponding ETD fragmentation spectrum of the light peptide. The spectrum contains extensive series of z and c ions from which the peptide and phospho-site can be easily identified.*

than 150 phosphopeptides having a phosphoRS site probability greater than 75% (Supplementary Table.2). After filtering for a cutoff ratio (Light/Medium < 0.2, Supplementary Figure.1), we identified more than 120 putative targets of PKA in our in vitro kinase assay (Supplementary Table.3). The resulting dataset of phosphorylated kinase substrates was evaluated with IceLogo to determine the enrichment of specific amino acids at positions surrounding the phospho-site. The overall motif of PKA detected by IceLogo, considering as position 0 the phospho-site, showed a very high enrichment in arginine at positions -2 and/or -3, while at +1 position hydrophobic amino acid such as I, L and V were over-represented (Figure.4). These data are in very good agreement with known literature, whereby the predominant motif found in in vivo studies is RRXS/TX also with a preference of hydrophobic amino acids at position +1. (34) Furthermore, due to the sensitivity and selectivity of our method, we could investigate in-depth and confirm the positional preference of PKA. We noticed that only 5% of all our in vitro targets were phosphorylated at threonine with 95% at serine. Indeed, PKA has also been predicted to have, both in vivo and in vitro, a marked preference for serine. (33) The selectivity of PKA was shown to be XRXS(T)X > RXXS(T)X > RRXS(T)X. Other motifs were also found in our data-set, albeit with much lower frequency, such as KRXS(T)X, RKXS(T)X, RQXS(T)X and others. Also several of these motifs have already been described as bona fide in vivo PKA targets. (33, 34) The observation of motifs such as RKXS(T)X and KKXS(T)X in our dataset confirmed that the choice of Lys-N and the digestion conditions chosen were optimal for generating the PKA peptide library.

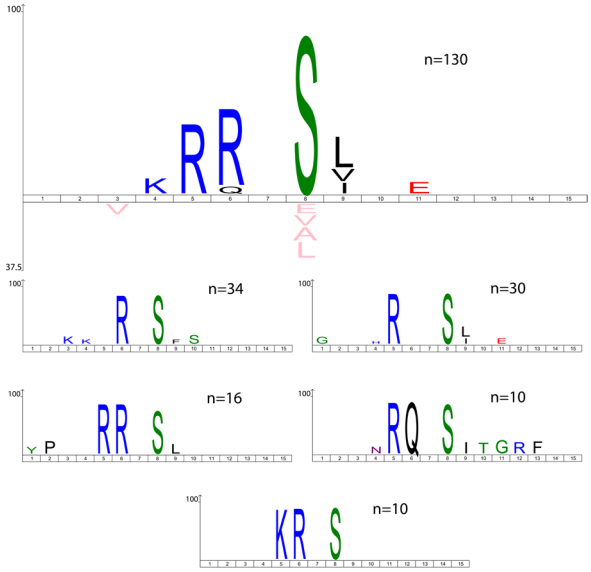


Figure.4: *High precision PKA consensus motifs and statistically relevant submotifs. The unambiguously identified 130 PKA target peptides all exhibited a pRS site probability >0.75, a Mascot ion score ≥20 and a light/intermediate ratio >5. These phosphopeptides were converted into aligned sequences with the phosphorylation site at position 8. The enrichment of each amino acid surrounding the phospho-sites was calculated using IceLogo with a p value of 0.001. Over-represented amino acids are displayed on the top of the figure with their relative frequencies (y axe) in each position, while under-represented ones appear at the bottom. In each figure, "n" represents how many sequences were used to calculate the statistics with IceLogo. The overall consensus motif is displayed at the top. In the five lower panels submotifs are depicted which were manually selected out of the 130 PKA target. These were illustrated with IceLogo.*

In vitro kinase substrate assay for Mps1, a kinase involved in mitotic spindle orientation

In order to demonstrate the wide applicability of our approach, we next sought to define sequence preference for phosphorylation by a poorly characterized but essential kinase, human Mps1, also known as TKK. Mps1 ensures correct chromosome segregation by establishing proper chromosome attachments to the mitotic spindle and by activating the mitotic checkpoint that delays cell division when such attachments have not been established. (35, 36) Compared to PKA, human Mps1 is less well characterized. Some genuine Mps1 substrates have been identified, including the conserved centromeric protein KNL1. (36) In addition, Mps1 auto-phosphorylates in cis and trans. (24, 37, 38) Still, a high-definition prediction of the authentic Mps1 phosphorylation motif is missing. Mps1 is thought to phosphorylate substrate peptides enriched in acidic amino acids (E and D), (39, 40) which makes it distinct from basophilic kinases like PKA. In order to create suitable peptides examination as Mps1 substrates, we picked trypsin as protease for the library generation. The resulting tryptic peptides were separated by SCX and fractions that would contain peptides of the same net charge were pooled. Subsequently, each pool was split into three aliquots. One third was used for the in vitro assay with wild-type Mps1, another third with a kinase-deficient mutant of Mps1 and the remaining part with wild-type Mps1 without addition of ATP. Essentially, the two latter aliquots, the controls, are expected to contain no phosphorylation activity. Light and intermediate dimethyl labels were chosen for the controls and the heavy form for the assayed aliquot. The reconstituted pools were then separated by SCX chromatography again. Subsequently, selected fractions were analyzed by LC-MS/MS. After the data analysis, the putative targets were defined as phosphorylated peptides with the heavy channel being at least 5 times more abundant than the other two channels (controls) (Figure.2). From the analysis we could quantify more than 3500 peptides (Supplementary Table.4) and more than 560 phosphopeptides having a phosphoRS site probability greater than 75% (Supplementary Table.5). After filtering for a ratio greater than five (Heavy/Light > 5 and Medium/Light > 5, Supplementary Figure.2), we gathered more than 500 potential substrates (Supplementary Table.6). Once again, we used the IceLogo program to identify amino acids that appeared to be enriched in the putative targets of Mps1. From this data, we could derive a putative consensus sequence motif for human Mps1, considering as position 0 the phospho-site. This motif contains a high enrichment in acidic amino acids in the positions -2, and/or -3 with a preference at -2, while in the +3 position a predominant preference for the hydrophobic branched-chain amino acids leucine, valine and isoleucine is observed (Figure.5). We noticed that over 90% of the peptide targets were phosphorylated on a threonine, and only a few on a serine. This clear preference for threonine is considerable and highly selective, since threonine phosphorylations are less common in human. (41) The few studies reporting on the substrate repertoire of Mps1 are consistent with these preferences: the target sequences in the budding yeast Mps1p substrates Spc29 and Spc110 are phosphorylated mostly on threonines and have a predominance of acidic residues, most commonly at position -2. (42, 43) Similarly, phosphorylation of human Borealin as well as yeast and human KNL1/Spc105 by Mps1 occurs exclusively on threonines and particularly on those preceded by acidic residues. (44–47) Moreover, the related auto-phosphorylation sites of Mps1 (24, 37–39) also have an E/D/N/Q preference at -2. (40, 42, 48–50) However, although a bias for I/V/L at position +3 can be seen in the auto-phosphorylated sequences, no such bias can be detected in KNL1/Spc105. Some of the numerous MELT-

like motifs in this protein are bona fide Mps1 substrate sequences but the +3 position is highly variable. Mps1 may therefore select substrates based mostly on threonine preceded by acidic residues. Our description of preferred Mps1 substrate sequences will greatly aid identification of bona fide functional Mps1 targets, and as such will help define the molecular regulatory network that guards against chromosome segregation errors.

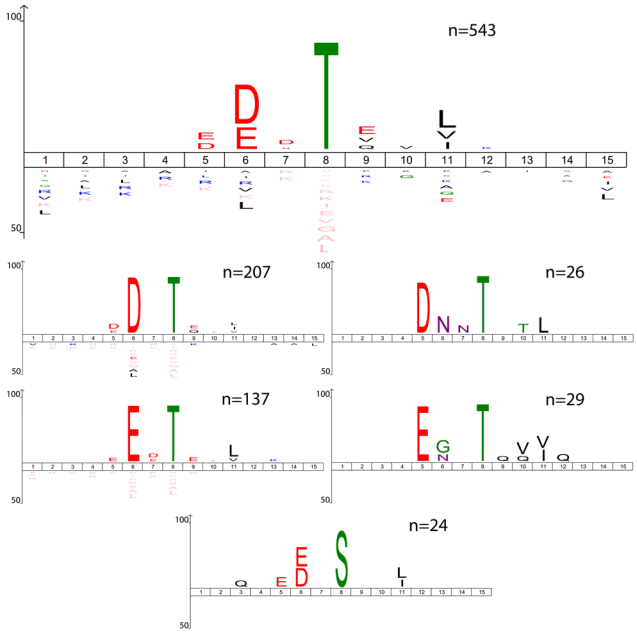


Figure.5: *High precision Mps1 consensus motif. The unambiguously identified over 500 Mps1 target peptides all exhibited a pRS site probability >0.75, a Mascot ion score ≥20 and a light/heavy and intermediate/heavy ratio >5. These phosphopeptides were converted into aligned sequences with the phosphorylation site at position 8. Statistics to find over- and under-represented amino acids of the experimental data set compared to the background database in each position flanking the phosphorylation-site were calculated using IceLogo with a p value of 0.001. Over-represented amino acids are displayed on the top figure with their relative frequencies (y axis) in each position, while under-represented ones appear on the bottom. In each figure "n" represents how many sequences were used to calculate the statistics with IceLogo. The overall consensus motif is displayed at the top. In the five lower panels submotifs are depicted which were manually selected out of the Mps1 target peptides. These were illustrated with IceLogo.*

4. CONCLUSIONS

We present a sensitive method designed for evaluating kinase consensus motifs with high accuracy. In this method, we create peptide libraries originating from the same organism as the studied kinase and combine it with a COFRADIC approach based on SCX chromatography. In order to make this strategy suitable to large peptide libraries, we combined it with stable isotope labeling. We evidenced its merits by finely discerning and thus, identifying several in vitro targets of two totally different protein kinases, namely the basophilic PKA and the acidophilic Mps1. Furthermore, from the statistical analysis of the two datasets, we could confirm the already well defined kinase consensus motif of human PKA and retrieved a high accuracy kinase consensus motif for human Mps1. Therefore, the use of our flexible

and generic method can be extended potentially to any kind of kinase existing in the kinome.

5. ACKOWLEDGEMENTS

The authors acknowledge all members of the Heck-group. This research was performed within the framework of the PRIME-XS project, grant number 262067, funded by the European Union 7th Framework Program. Additionally, the Netherlands Proteomics Centre, a program embedded in the Netherlands Genomics Initiative, is kindly acknowledged for financial support as well as the Netherlands Organization for Scientific Research (NWO) with the VIDI grant for SM (700.10.429).

6. REFERENCES

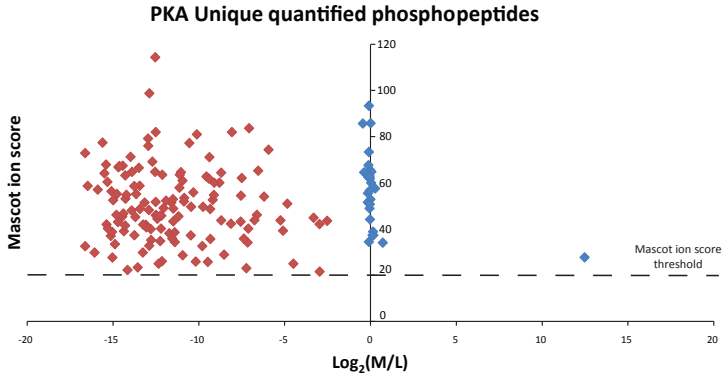
1. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) The protein kinase complement of the human genome. Science (80-.). 298, 1912–1934
2. Blume-Jensen, P., and Hunter, T. (2001) Oncogenic kinase signalling. Nature 411, 355–365
3. Stenberg, K. A., Riikonen, P. T., and Vihinen, M. (1999) KinMutBase, a database of human disease-causing protein kinase mutations. Nucleic Acids Res 27, 362–364
4. Ubersax, J. A., and Ferrell, J. E. (2007) Mechanisms of specificity in protein phosphorylation. Nat Rev Mol Cell Biol 8, 530–541
5. Songyang, Z., Blechner, S., Hoagland, N., Hoekstra, M. F., Piwnica-Worms, H., and Cantley, L. C. (1994) Use of an oriented peptide library to determine the optimal substrates of protein kinases. Curr Biol 4, 973–982
6. Rodriguez, M., Li, S. S., Harper, J. W., and Songyang, Z. (2004) An oriented peptide array library (OPAL) strategy to study protein-protein interactions. J Biol Chem 279, 8802–8807
7. Yaffe, M. B. (2004) Novel at the library. Nat Methods 1, 13–14
8. Manning, B. D., and Cantley, L. C. (2002) Hitting the target: emerging technologies in the search for kinase substrates. Sci STKE 2002, pe49
9. Hutti, J. E., Jarrell, E. T., Chang, J. D., Abbott, D. W., Storz, P., Toker, A., Cantley, L. C., and Turk, B. E. (2004) A rapid method for determining protein kinase phosphorylation specificity. Nat Methods 1, 27–29
10. Good, D. M., Wirtala, M., McAlister, G. C., and Coon, J. J. (2007) Performance characteristics of electron transfer dissociation mass spectrometry. Mol Cell Proteomics 6, 1942–1951
11. Wiesner, J., Premsler, T., and Sickmann, A. (2008) Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. Proteomics 8, 4466–4483
12. Frese, C. K., Altelaar, A. F. M., Hennrich, M. L., Nolting, D., Zeller, M., Griep-Raming, J., Heck, A. J. R., and Mohammed, S. (2011) Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. J. Proteome Res. 10, 2377–88
13. Xue, L., Wang, W. H., Iliuk, A., Hu, L., Galan, J. A., Yu, S., Hans, M., Geahlen, R. L., and Tao, W. A. (2012) Sensitive kinase assay linked with phosphoproteomics for identifying direct kinase substrates. Proc Natl Acad Sci U S A 109, 5615–5620
14. Singh, S. A., Winter, D., Bilimoria, P. M., Bonni, A., Steen, H., and Steen, J. A. (2012) FLEXIQinase, a mass spectrometry-based assay, to unveil multikinase mechanisms. Nat Methods 9, 504–508
15. Zhou, H., Di Palma, S., Preisinger, C., Peng, M., Polat, A. N., Heck, A. J., and Mohammed, S. (2012) Towards a comprehensive characterization of a human cancer cell phosphoproteome. J Proteome Res,
16. Schilling, O., auf dem Keller, U., and Overall, C. M. (2011) Protease specificity profiling by tandem mass spectrometry using proteome-derived peptide libraries. Methods Mol Biol 753, 257–272
17. Schilling, O., and Overall, C. M. (2008) Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. Nat Biotechnol 26, 685–694
18. Kettenbach, A. N., Wang, T., Faherty, B. K., Madden, D. R., Knapp, S., Bailey-Kellogg, C., and Gerber, S. A. (2012) Rapid determination of multiple linear kinase substrate motifs by mass spectrometry. Chem Biol 19, 608–618
19. Li, X., Gerber, S. a, Rudner, A. D., Beausoleil, S. a, Haas, W., Villén, J., Elias, J. E., and Gygi, S. P. (2007) Large-scale phosphorylation analysis of alpha-factor-arrested Saccharomyces cerevisiae. J. Proteome Res. 6, 1190–1197
20. Gevaert, K., Staes, A., Van Damme, J., De Groot, S., Hugelier, K., Demol, H., Martens, L., Goethals, M., and Vandekerckhove, J. (2005) Global phosphoproteome analysis on human HepG2 hepatocytes using reversed-phase diagonal LC. Proteomics 5, 3589–3599
21. Staes, A., Impens, F., Van Damme, P., Ruttens, B., Goethals, M., Demol, H., Timmerman, E., Vandekerckhove, J., and Gevaert, K. (2011) Selecting protein N-terminal peptides by combined fractional diagonal chromatography.

Nat Protoc 6, 1130–1141
22. Staes, A., Van Damme, P., Helsens, K., Demol, H., Vandekerckhove, J., and Gevaert, K. (2008) Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). Proteomics 8, 1362–1370
23. Taouatas, N., Heck, A. J., and Mohammed, S. (2010) Evaluation of metalloendopeptidase Lys-N protease performance under different sample handling conditions. J Proteome Res 9, 4282–4288
24. Jelluma, N., Brenkman, A. B., McLeod, I., Yates, J. R., Cleveland, D. W., Medema, R. H., and Kops, G. J. (2008) Chromosomal instability by inefficient Mps1 auto-activation due to a weakened mitotic checkpoint and lagging chromosomes. PLoS One 3, e2415
25. Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A. J. (2009) Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protoc 4, 484–494
26. Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C., and Mechtler, K. (2011) Universal and confident phosphorylation site localization using phosphoRS. J Proteome Res 10, 5354–5362
27. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., and MacCoss, M. J. (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods 4, 923–925
28. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009) Improved visualization of protein consensus sequences by iceLogo. Nat Methods 6, 786–787
29. Peng, M., Taouatas, N., Cappadona, S., van Breukelen, B., Mohammed, S., Scholten, A., and Heck, A. J. (2012) Protease bias in absolute protein quantitation. Nat Methods 9, 524–525
30. Di Palma, S., Hennrich, M. L., Heck, A. J., and Mohammed, S. (2012) Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. J Proteomics 75, 3791–3813
31. Mohammed, S., and Heck, A. J. (2011) Strong cation exchange (SCX) based analytical methods for the targeted analysis of protein post-translational modifications. Curr. Opin. Biotechnol. 22, 9–16
32. Hennrich, M. L., van den Toorn, H. W., Groenewold, V., Heck, A. J., and Mohammed, S. (2012) Ultra acidic strong cation exchange enabling the efficient enrichment of basic phosphopeptides. Anal Chem 84, 1804–1808
33. Shabb, J. B. (2001) Physiological substrates of cAMP-dependent protein kinase. Chem Rev 101, 2381–2411
34. Ruppelt, A., and Tasken, K. Physiological substrates of PKA and PKG. Three-Volu,
35. Kops, G. J., and Shah, J. V (2012) Connecting up and clearing out: how kinetochore attachment silences the spindle assembly checkpoint. Chromosoma 121, 509–525
36. Vleugel, M., Hoogendoorn, E., Snel, B., and Kops, G. J. (2012) Evolution and function of the mitotic checkpoint. Dev Cell 23, 239–250
37. Kang, J., Chen, Y., Zhao, Y., and Yu, H. (2007) Autophosphorylation-dependent activation of human Mps1 is required for the spindle checkpoint. Proc Natl Acad Sci U S A 104, 20232–20237
38. Mattison, C. P., Old, W. M., Steiner, E., Huneycutt, B. J., Resing, K. A., Ahn, N. G., and Winey, M. (2007) Mps1 activation loop autophosphorylation enhances kinase activity. J Biol Chem 282, 30553–30561
39. Dou, Z., von Schubert, C., Körner, R., Santamaria, A., Elowe, S., and Nigg, E. A. (2011) Quantitative mass spectrometry analysis reveals similar substrate consensus motif for human Mps1 kinase and Plk1. PLoS One 6, e18793
40. Mok, J., Kim, P. M., Lam, H. Y., Piccirillo, S., Zhou, X., Jeschke, G. R., Sheridan, D. L., Parker, S. A., Desai, V., Jwa, M., Cameroni, E., Niu, H., Good, M., Remenyi, A., Ma, J. L., Sheu, Y. J., Sassi, H. E., Sopko, R., Chan, C. S., De Virgilio, C., Hollingsworth, N. M., Lim, W. A., Stern, D. F., Stillman, B., Andrews, B. J., Gerstein, M. B., Snyder, M., and Turk, B. E. (2010) Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. Sci Signal 3, ra12
41. Mann, M., Ong, S. E., Grønborg, M., Steen, H., Jensen, O. N., and Pandey, A. (2002) Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. Trends Biotechnol 20, 261–268
42. Holinger, E. P., Old, W. M., Giddings, T. H., Wong, C., Yates, J. R., and Winey, M. (2009) Budding yeast centrosome duplication requires stabilization of Spc29 via Mps1-mediated phosphorylation. J Biol Chem 284, 12949–12955
43. Friedman, D. B., Kern, J. W., Huneycutt, B. J., Vinh, D. B., Crawford, D. K., Steiner, E., Scheiltz, D., Yates, J., Resing, K. A., Ahn, N. G., Winey, M., and Davis, T. N. (2001) Yeast Mps1p phosphorylates the spindle pole component Spc110p in the N-terminal domain. J Biol Chem 276, 17958–17967
44. Jelluma, N., Brenkman, A. B., van den Broek, N. J., Cruijsen, C. W., van Osch, M. H., Lens, S. M., Medema, R. H., and Kops, G. J. (2008) Mps1 phosphorylates Borealin to control Aurora B activity and chromosome alignment. Cell 132, 233–246
45. London, N., Ceto, S., Ranish, J. A., and Biggins, S. (2012) Phosphoregulation of Spc105 by Mps1 and PP1 regulates Bub1 localization to kinetochores. Curr Biol 22, 900–906
46. Shepperd, L. A., Meadows, J. C., Sochaj, A. M., Lancaster, T. C., Zou, J., Buttrick, G. J., Rappsilber, J., Hardwick, K. G., and Millar, J. B. (2012) Phosphodependent recruitment of Bub1 and Bub3 to Spc7/KNL1 by Mph1 kinase maintains the spindle checkpoint. Curr Biol 22, 891–899
47. Yamagishi, Y., Yang, C. H., Tanno, Y., and Watanabe, Y. (2012) MPS1/Mph1 phosphorylates the kinetochore protein KNL1/Spc7 to recruit SAC components. Nat Cell Biol 14, 746–752

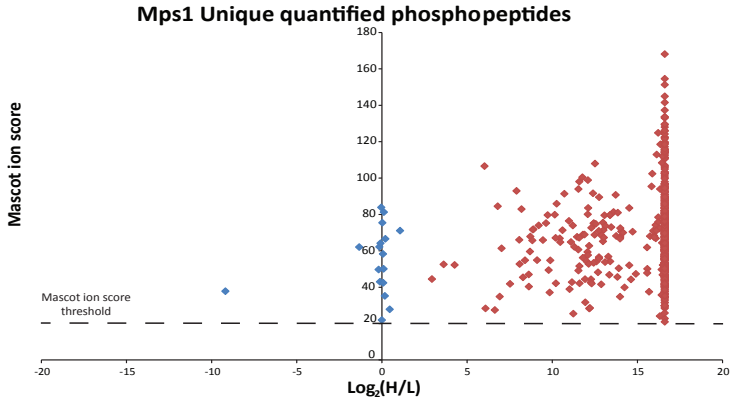
48. Kemmler, S., Stach, M., Knapp, M., Ortiz, J., Pfannstiel, J., Ruppert, T., and Lechner, J. (2009) Mimicking Ndc80 phosphorylation triggers spindle assembly checkpoint signalling. EMBO J 28, 1099–1110
49. Araki, Y., Gombos, L., Migueleti, S. P., Sivashanmugam, L., Antony, C., and Schiebel, E. (2010) N-terminal regions of Mps1 kinase determine functional bifurcation. J Cell Biol 189, 41–56
50. Shimogawa, M. M., Graczyk, B., Gardner, M. K., Francis, S. E., White, E. A., Ess, M., Molk, J. N., Ruse, C., Niessen, S., Yates, J. R., Muller, E. G., Bloom, K., Odde, D. J., and Davis, T. N. (2006) Mps1 phosphorylation of Dam1 couples kinetochores to microtubule plus ends at metaphase. Curr Biol 16, 1489–1501

7. SUPPLEMENTARY

Supplementary Table.1 to 6 are available via the Internet at http://pubs.acs.org.



Supplementary Figure.1: *Scatter plot of the distribution of phosphopeptides identified in the PKA kinase assay. The plotted over 150 unambiguously identified phosphopeptides (Supplementary Table.2) all exhibited a pRS site probability greater than 75% and a Mascot ion score equal to or greater than 20. The Log2 ratios (Intermediate/Light) of the phosphopeptides are plotted against the Mascot ion score. Two distinct distributions corresponding to the PKA targets (red dots) and non-targets (blue dots) can be distinguished. The dashed line indicates the applied Mascot ion score threshold of 20. The maximum allowed ratio in the quantification software was 100000 corresponding to a Log2 of 16.6.*



Supplementary Figure.2: *Scatter plot of the distribution of phosphopeptides identified in the Mps1 kinase assay. The plotted over 560 unambiguously identified phosphopeptides all exhibited a pRS site probability greater than 75% and a Mascot ion score equal to or greater than 20. The Log2 ratios (Heavy/Light) of the phosphopeptides are plotted against the Mascot ion score. Two distinct distributions corresponding to the Mps1 targets (red dots) and non-targets (blue dots) can be distinguished. The dashed line indicates the applied Mascot ion score threshold of 20. The maximum allowed ratio in the quantification software was 100000 corresponding to a Log2 of 16.6.*

CHAPTER 4

Arginine (di)methylated Human Leukocyte Antigen class I peptides are favorably presented by HLA-B*07

Fabio Marino[1,2], Geert P.M. Mommen[1,2,3], Anita Jeko[1,2], Hugo D. Meiring[3], Jacqueline A.M. van Gaans-van den Brink[4], Richard A. Scheltema[1,2], Cécile A.C.M. van Els[4], Albert J. R. Heck[1,2]

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

[2]Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands.

[3]Institute for Translational Vaccinology, Bilthoven, Netherlands.

[4]Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands.

Manuscript under revision

ABSTRACT

Specific alterations in protein post-translational modification (PTMs) are recognized hallmarks of diseases. These modifications potentially provide a unique disease-related source of Human Leukocyte Antigen (HLA) class I-presented peptide antigens that can elicit specific immune responses. However, the complexity and frequency of presentation of arginine methylated HLA class I peptides have not been explored in detail. In a model human B-cell line we detected by mass spectrometry (MS) 149 HLA class I peptides harboring mono- and/or di-methylated arginine residues. The source proteins of these antigens play important roles in signal transduction, gene transcription and DNA repair. A striking preference was observed in presentation of arginine (di)methylated peptides predicted to bind HLA-B*07 molecules, most likely because the binding motifs of this allele resemble the substrates for arginine methyl-transferases. The HLA-B*07 peptides were preferentially di-methylated at the P3 position in the sequence, thus successively to the proline anchor residue at position P2. This proline-arginine substrate has been associated with the arginine methyl-transferases CARM1 and PRMT5. Making use of the specific neutral losses in the MS/MS spectra we could further assign most of the peptides to be asymmetrically di-methylated by CARM1. The here presented data expands our knowledge of processing and presentation of arginine (di)methylated HLA class I peptides, indicating that this type of modification is frequently presented for recognition by T-cells and might be a potential target for immunotherapy.

1. INTRODUCTION

Human Leukocyte Antigen (HLA) class I molecules display short peptides derived from the degradation of cellular proteins on the cell surface.(1) Recognition of pathogen- or disease-related HLA class I-presented peptides by CD8 T lymphocytes cells leads to the activation of a cytotoxic response and clearance of the affected cells. The repertoire of peptides presented by HLA class I molecules is dominated by self-peptides that could encompass over 14,000 different species,(2, 3) all with a defined binding motif that allows binding to the expressed class I molecules. The overwhelming majority of the naturally processed and HLA class I-presented peptides are unmodified, albeit that a small fraction of peptides harbor posttranslational modifications (PTMs) (4). The importance of HLA class I presentation of PTMs peptides and their specific recognition by T cells has been demonstrated in infectious,(5)autoimmune diseases,(6, 7) and cancer.(8) Protein arginine methylation is a common PTM catalyzed by a family of enzymes known as protein arginine methyl transferases (PRMTs), yielding mono-methylated, symmetric (SDMA) or asymmetric (ADMA) di-methylated arginine residues (Figure.1A). Methylation of arginine is known to play a critical role in regulating gene expression(9) and has been implicated in signal transduction and DNA repair.(10) By using mass spectrometry (MS)-based approaches, few HLA class I peptides carrying arginine methylated residues have been reported. (11–13) Yagüe et al. (11) identified for the first time a naturally processed and HLA-B*39 presented di-methylated peptide derived from a RNA-binding nucleoprotein. Immunological recognition of the side-chain modification of arginine residues was then shown by Jarmalavicus et al.,(12) demonstrating that T cells of specific melanoma patients only reacted against the mono-methylated HLA-A*11 peptide of the GPS-2 protein. Apart from these interesting findings, little is known about the nature of arginine methylated HLA class I

peptides, their source protein origin and frequency of presentation. Here, we report 149 arginine (di)methylated HLA class I peptides, indicating that this PTM is frequently processed and presented for scrutiny by T cells. We found that arginine methylated peptides were presented with a marked allele-specific preference and a favorable P3 sequence position, thereby resembling the underlying substrate motif of arginine methyltransferases.

2. EXPERIMENTAL SECTION

Cell culturing, isolation of HLA class I-associated peptides and LC-MS/MS analysis.
Cell culture conditions and isolation of HLA-peptides complexes was described previously. (14) Briefly, two biological replicates of the HLA-A*01, A*03, B*07,-B*27,-C*02, and-C*07–positive B-lymphoblastoid cell line GR were grown in RPMI-1640 medium to a total number of 9 × 109 cells. HLA class I-peptide complexes were immunoprecipitated from lysed GR cells, using the HLA-A–,-B–, and-C–specific mouse monoclonal IgG2a antibody W6/32. HLA class I-peptides complexes were eluted using 10% (vol/vol) acetic acid and peptides were further purified by passage over a 10-kDa molecular weight cutoff membrane. HLA class I-eluted peptides were fractionated by SCX chromatography. The system comprises a Hypercarb trapping column (5 × 0.2 mm i.d., 7 µm particle size; Thermo Fisher) and SCX column (12 × 0.02 cm i.d. polysulfoethyl aspartamide, 5 µm; Poly LC). For each biological replicate a total number of nine SCX fractions were collected. The SCX fractions were analyzed directly by nanoscale LC-MS/MS using a Thermo Scientific EASY-nLC 1000 (Thermo Fisher Scientific) in combination with an ETD-enabled LTQ Orbitrap Elite (first biological replicate) or an Orbitrap Fusion (second biological replicate) mass spectrometer (Thermo Fisher Scientific). The LC system comprises a 20 × 0.1 mm i.d. trapping column (Reprosil C18, 3 µm; Dr. Maisch) and a 50 × 0.005 cm i.d. analytical column (Poroshell 120 EC-C18; 2.7 µm). For the Orbitrap Elite, full MS spectra were acquired in the Orbitrap at a resolution of 60,000 (FWHM) while fragment ions were detected in the Orbitrap at a resolution of 15,000 (FWHM). The 10 most abundant precursor ions were selected either for data-dependent EThcD, CID, ETD,or HCD as previously described (14). The maximum ion accumulation time for MS scans was set to 200 ms and for MS/MS scans to 2,500 ms. For the Orbitrap Fusion, both full MS and MS/MS spectra were acquired in an Orbitrap with a resolution of 60,000 (FWHM) and 15,000 (FWHM), respectively. The Top Speed method was enabled for fragmentation where all most abundant precursor ions in 3 seconds were selected for data-dependent EThcD. The maximum ion accumulation time for MS and MS/MS scans was set to 50 ms and 250 ms, respectively.

Data analysis
All raw data files were analyzed using Proteome Discoverer 1.4 software package (Thermo Fisher Scientific, Bremen, Germany). MS/MS scans were searched against the Swissprot human reviewed database (September 2015, 20203 entries) with no enzyme specificity using the SEQUEST HT mode. Precursor ion and MS/MS tolerances were set to 10 ppm and 0.05 Da. Methionine oxidation, arginine mono-methylation and arginine di-methylation were set as variable modifications. The peptides-to-spectrum matches were further filtered for precursor tolerance 5 ppm, < 1% FDR using Percolator,(15) XCorr > 1.7 and peptide rank 1. Only peptides between 8 and 14 amino acid long were selected for further analysis. Precursor ion area detection node was added in order to obtain the area

under the curve of the LC elution profiles of methylated, di-methylated and unmodified peptides. The areas under the curve of differently methylated peptides (none, mono-, di-) were only compared when detected in the same LC run. The NetMHC 3.4 algorithm(16) was used to predict the HLA-peptide binding affinities for each of the identified peptide sequence. Predictions were enabled for HLA-A*01, -A*03, -B*07 and –B*27, and peptides were assigned to bind a particular allele when IC50 < 1000 nM. Sequence logo's were generated by the IceLogo (17) algorithm with the Uniprot-Swissprot protein database as reference set and a p-value of 0.05. Gene Ontology (GO) analysis of the source proteins of the (di)methylated HLA class I peptides was performed by PANTHER.(18)

Distinction between asymmetrical and symmetrical di-methyl arginine

Asymmetric Di-Methyl Arginine (ADMA) and Symmetrical Di-Methyl Arginine (SDMA) modified peptides are positional isomers, which can be distinguished by their specific neutral loss ions generated upon peptide fragmentation using collision induced dissociation techniques.(19) Fragmentation spectra of ADMA peptides contain specific losses of 45 Da corresponding to di-methylamine (DMA), whereas SMDA peptides display a neutral loss of 31 Da corresponding to mono-methylamine loss (MMA).(20) Evaluation of symmetry was automated by annotating the neutral losses with a mass tolerance of 0.05 and calculating a likelihood of finding that many neutral losses in the background of annotated 'normal' fragment ions, which acts as a p-value.(21) The specific neutral loss (DMA or MMA) with the best likelihood below 0.05 was selected as the indicator for symmetry.

Synthesis of (di)methylated peptides and their unmodified counter parts

Peptide building blocks were purchased from Novabiochem and appropriately functionalized resins were from Applied Biosystems. Peptides were synthesized using standard Fmoc-based solid-phase peptide synthesis protocols and appropriately functionalized polyethylene glycol–polysterene Wang resins. Functionalized resins were subjected to coupling cycles, in which deprotection of the Fmoc group with piperidine/NMP (1:4 [v/v]) was followed by coupling with four equivalents each of Fmoc-protected amino acid, di-isopropylethylamine, and (benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate. Reactions were carried out in NMP at a volume of 1 ml/0.1 g resin. After the final coupling step, the Fmoc group was removed and peptides were directly fully deprotected and released from the resin directly by treating the resin with trifluoroacetic acid (TFA)/H2O/triisopropylsilane (93:5:2 [v/v]) for 2.5 h. Alternatively, peptides were treated with four equivalents each of acetic anhydride and di-isopropylethylamine for 40 min to acetylate the N terminus prior to deprotection and release from the resin.

3. RESULTS

Identification of arginine methylated peptides presented by HLA class I molecules

The repertoire of HLA class I peptides were isolated from two replicate cell cultures of the HLA-A*01, -A*03, -B*07, -B*27, C*02, C*07-positive human B-cell line GR by immunopurifcation. Isolated HLA class I peptides were fractionated by strong cation exchange and analyzed by LC-MS/MS using an LTQ Orbitrap Elite (replicate 1) and Orbitrap Fusion (replicate 2). By searching the MS/MS data files for side-chain modification of arginine residues we detected, in the first biological replicate, 38 arginine methylated and 46 ar-

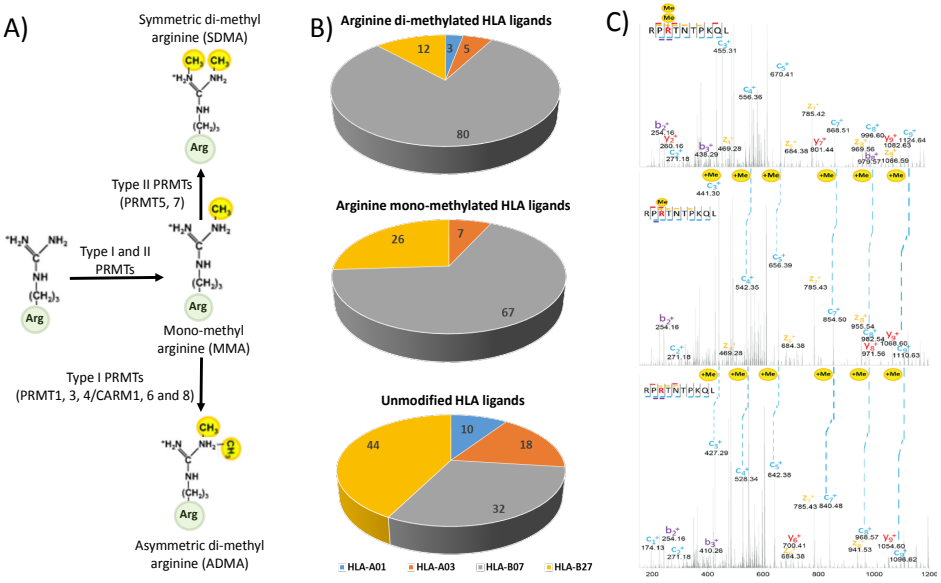


Figure. 1: *Overview of enzymatic arginine (di)methylation processes, the observed allele preference and exemplary EThcD fragmentation spectra. A) The family of protein arginine methylation transferases (PRMTs) mediate methylation and successive symmetric or asymmetric di-methylation. The latter, respectively by type II or I PRMTs. B) Distribution of detected HLA class I peptides over the four most prominent expressed alleles. From top to bottom are displayed the distribution of di-methylated peptides, mono-methylated peptides and unmodified peptides. C) Exemplary EThcD peptide fragmentation spectra of the di-methylated, mono-methylated and unmodified form of the peptide RPRTNTPQL, originating from the protein NCoA-3. EThcD enabled complete peptide sequence coverage and unambiguous PTM site localization at P3 of the peptide sequence (blue dashed line).*

ginine di-methylated and 10 peptides with multiple (di)methylation sites. In the second replicate, 25 arginine methylated and 44 arginine di-methylated and 13 peptide with multiple modifications sites were detected. Cumulatively, a total of 55 unique arginine methylated, 74 unique arginine di-methylated and 20 unique peptides with multiply arginine methylation or di-methylation sites were found. Arginine (di)methylated peptides were found to be between 8 and 14 amino acids long, with a preference for the canonical length of 9 and 10 amino acids. Compared to the cumulative number of about 17,000 unmodified HLA class I peptides that were detected in both data sets, the 149 arginine (di) methylated peptides represent approximately 1% of the total HLA class I ligandome. For the accurate site assignment of the arginine (di)methylated HLA class I peptides we mainly relied on the dual fragmentation technique EThcD (Figure.1C). EThcD generates complementary ion series of b/y-ions and c/z-ions has been shown to be beneficial for the accurate assignment of the PTM site.(22, 23) The NetMHC algorithm(16) was used to predicted the binding affinity and assign the identified peptides to the expressed HLA-A*01,-A*03,-B*07 or –B*27 alleles. The HLA-C*02 and –C*07 alleles were excluded from this analysis because of the low surface expression and lack in accuracy in binding affinity predictions.(14) This analysis revealed that approximately 80% of the identified arginine (di)methylated peptides were predicted to bind to one of the four most prominent expressed alleles. To further validate the peptide assignments and their site of modification, we compared our data with

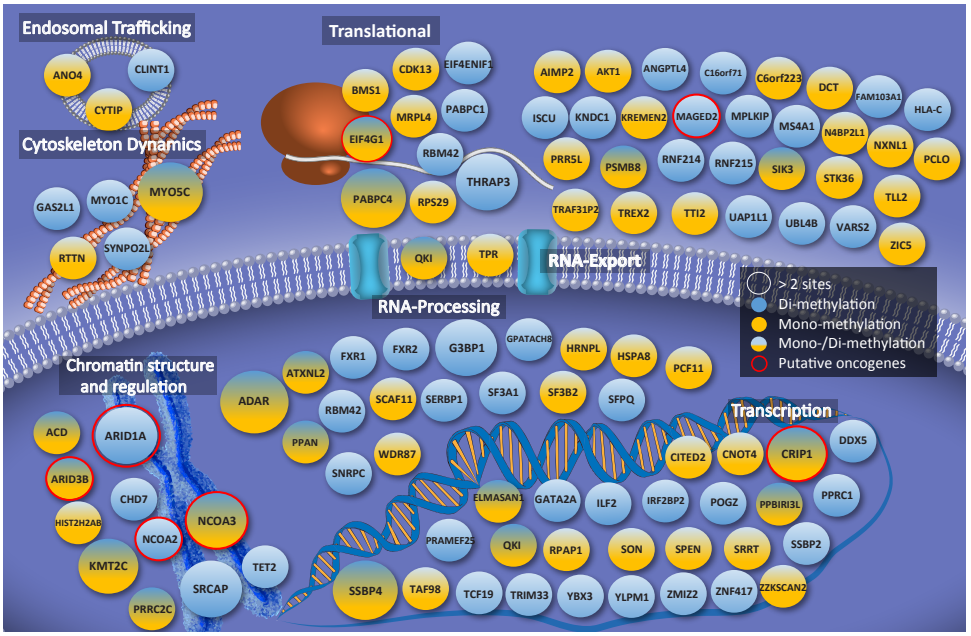


Figure.2: *Overview of the source proteins of arginine (di)methylated HLA class I peptides. Proteins are grouped according to their annotated function (using the UniprotKB database and Panther).(18) Proteins for which only methylated HLA class I peptides were found are drawn in yellow, only di-methylated HLA class I peptides are colored in blue, while proteins that were found to be both mono-methylated and di-methylated are blue/yellow. Proteins for which more than two modified peptides were detected are annotated with larger circles, while the ones annotated with red circles are from putative oncogenes, previously found to be involved in various cancers.*

previously identified arginine methylation sites deposited in the PhosphoSite.org database, one of the largest publicly available inventories of protein modifications. Notably, 13 protein arginine methylation and 18 protein arginine di-methylation sites have been reported previously. Furthermore, we verified the identification and PTM site localization of 12 endogenous peptides (methylated, di-methylated and unmodified) using synthetic peptide analogous. Supplementary Figures 3 to 14 display the nearly perfect match between the annotated MS/MS spectra of the endogenous peptides and their synthetic counterparts. The arginine (di)methylated HLA class I-associated peptides originated from 105 non-redundant source proteins. Gene ontology (GO) annotation of these source proteins revealed enrichment of RNA/DNA processing and binding, and regulation of transcription (Figure. 2). A complete overview of the here identified source proteins is given in Figure. 2. Next to the above-mentioned categories, we also observed proteins implicated in other cellular functions such as RNA-export (i.e. TPR is a component of the nucleopore complex), endosomal trafficking and cytoskeletal re-arrangements.

Allele-specific presentation of arginine (di)methylated peptides

The arginine (di)methylated peptides were found to be associated with the expressed HLA-A*01, -A*03, -B*07 and -B*27 alleles. However, we found that the HLA-B alleles present larger number of (di)methylated peptides than the HLA-A alleles. Figure.1B shows the frequency distribution of the identified arginine di-methylated, the arginine methylated and the unmodified peptides over the four investigated HLA alleles. In con-

trast to the set of unmodified peptides, both arginine methylated and arginine di-methylated HLA class I peptides exhibited a remarkable preference for HLA-B*07, with respectively 67% and 80% of the predicted binders being associated to HLA-B*07. The favorable presentation of arginine (di)methylated peptides by HLA-B*07 could be partly explained by the higher frequency of arginine residues in the consensus peptide sequence for this allele. The sequence logo's in Supplementary Figure.1 show that arginine residues are favored at P1, P3, P5 and P6 in the set of unmodified HLA-B*07-associated peptides. HLA-A*03 and HLA-B*27 prefer a basic amino acid residue at their N-termini, while HLA-A*01 does not favor an arginine residue in any part of the peptide sequence. To further investigate the seemingly enhanced presentation of (di)methylated peptides by HLA-B*07, we evaluated the positional frequency of the modification sites and compared these with the positional frequency of non-modified arginine residues in of the total set of identified HLA class I peptides (Figure.3). This analysis revealed for the set of unmodified peptides (black bars) a preference for arginine on position P1 and P2, which corresponds to the previously discussed consensus sequence motifs mainly for HLA-A*03 and HLA-B*27 (Supplementary Figure.1). For both arginine methylated and arginine di-methylated peptides, however, we found that these peptides are presented by HLA class I with a high preference for the modification site on position P3 of the peptide sequence: frequency distributions of 36% and 48%, respectively (Figure.3). Note, the arginine (di)methylated peptide with a position P3 modification site were found to be exclusively associated with the HLA-B*07 allele. This preferred presentation of position P3 modified peptides by HLA-B*07 could reflect structural features of this HLA allele or represent specific motifs required for the enzymatic modification by arginine methyltransferases. In addition, a moderate preference was detected for (mono-) methylation sites at position P2 of the presented HLA class I peptides (Figure.3, 21%). The observation that the position P2 arginine methylated peptides were associated with HLA-B*27, and that nearly no di-methylated sites were detected at this position for this allele, indicates that only one additional methyl group on the arginine side-chain seem to be tolerated at the anchor position of HLA-B*27 (Supplementary Figure.1).

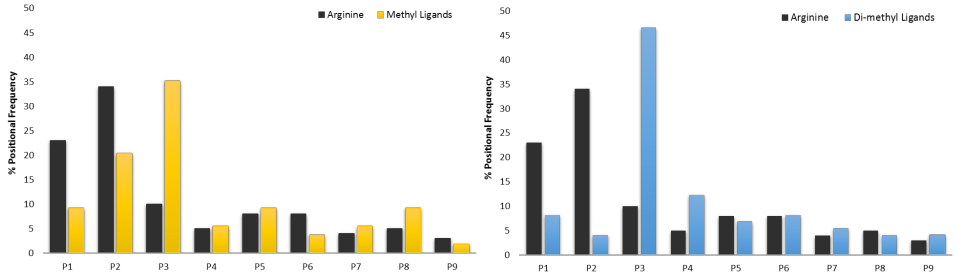


Figure.3: *Positional preference in the site of modification. The positional frequency distributions (%) the site of modification for methylated (orange bars) and di-methylated (blue bars) HLA class I peptide were compared to the distribution of unmodified arginine residues in total data set of HLA class I peptides (black bars).*

Arginine di-methylated peptides are more frequently and more abundantly presented by HLA-B*07

To further characterize the preferential presentation of arginine (di)methylated peptides by HLA-B*07 allele, we first performed a qualitative comparison analysis with a recent proteomics study.(24) Using antibody enrichment and LC-MS/MS analysis, Guo et al. identified

1500 arginine (di)methylation sites in a cellular proteome sample from a human cell line. Evaluating this proteomics date set revealed that 66% of the identified protein modification sites were found to be arginine methylated and 33% to be arginine di-methylated. In contrast, in our study a higher number arginine di-methylated HLA class I peptides were identified compared to arginine methylated peptides (57% versus 43%). This could indicate that the sampling of arginine-modified proteins by HLA class I pathway is biased for arginine di-methylation. To confirm this observation, we quantified the relative abundance of the arginine methylated and di-methylated HLA class I peptides in comparison to their corresponding unmodified peptide counterparts. Relative abundances were obtained from 22 non-modified peptides with their corresponding arginine methylated peptides, 22 non-modified with their corresponding di-methylated peptides and a further 14 peptides with mono- and di-methylation sites. The relative abundance varied considerably from peptide to peptide, but when we further clustered the data with respect to the preferred

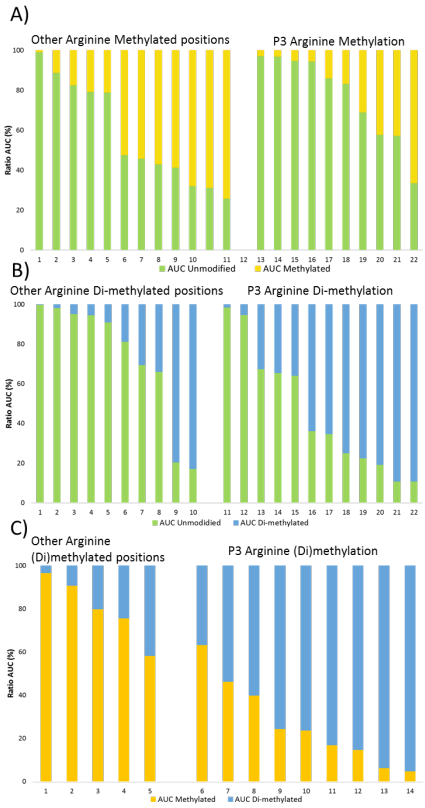


Figure.4: *Ratios of ion abundances of arginine (di)methylated HLA class I peptides and their unmodified counterparts in P3 and other modified positions. A) The relative (%) extracted area under the curves (AUCs) of P3 methylated ligands (right panel, orange bars) were compared to their unmodified counterparts (right panel, green bars). The same trend was studied for the modification harboring other positions than P3 (left panels). B) The relative AUCs of P3 di-methylated ligands (right panel, blue bars) were compared to their unmodified counterparts (right panel, green bars). The same trend was studied for the modification harboring other positions than P3 (left panels). C) The relative AUCs of P3 di-methylated ligands (right panel, blue bars) were compared to their methylated counterparts (right panel, orange bars). The same trend was studied for the modification harboring other positions than P3 (left panels).*

arginine modification at the position P3 for HLA-B*07, a strikingly trend was observed. On average, the non-modified peptides seems to be presented at higher levels compared to their arginine methylated variants regardless of the site of modification. This is also the case for the comparison of non-modified and di-methylated peptides, although a large fraction of the peptides carrying a modification at position P3 displayed higher abundances compared to their non-modified counterparts. An even more striking trend was found when we compared the same peptides harboring differential methylated arginine forms. In nearly every comparison arginine di-methylated peptides were found to be presented at higher levels compared to their methylated variants only when site of modification was located at position P3. Altogether, the qualitative and quantitative results clearly show a preferential bias in the presentation of arginine di-methylated peptides with the modification site at position P3. To investigate whether the observed biases in arginine P3 (di)methylation is imposed by an underlying substrate motif of PRMTs, we extracted all human protein (di)methylated sites from the PhosphoSite.org database and used the NetMHC algorithm (16) to predicted whether this large set of (di)methylation could be presented by HLA-B*07. For this analysis, potential 9-mer HLA class I peptides were predicted from the source protein sequence with a window of 9 residues around the protein arginine (di)methylation site. In total, respectively 650 and 506 arginine methylated and arginine di-methylated 9-mers from previously reported modifications sites were predicted to bind to the HLA-B*07 allele (Figure.5). The positional frequency of the site of modification from the predicted set of HLA-B*07 binders were compared against our endogenous dataset (Figure.5). In this analysis, we excluded the position P2 and C-term anchor residues because no endogenous HLA-B*07 peptides was found to be modified at these positions. The extracted set of predicted HLA-B*07 binders (black bars) showed a preference for arginine (di)methylation at P1, P3, P5 and P6, which corresponds nicely to enrichment of arginine residues at these positions in the consensus binding motif of this allele (Supplementary Figure.1). More importantly, preference for the arginine (di)methylation site at position P3 (< 30%) was found for the extracted date set, but only at a moderate enrichment in comparison the identified set of endogenous peptides (> 60%). These results indicate that the enhanced presentation of arginine (di)methylated peptides by HLA-B*07 could be partially

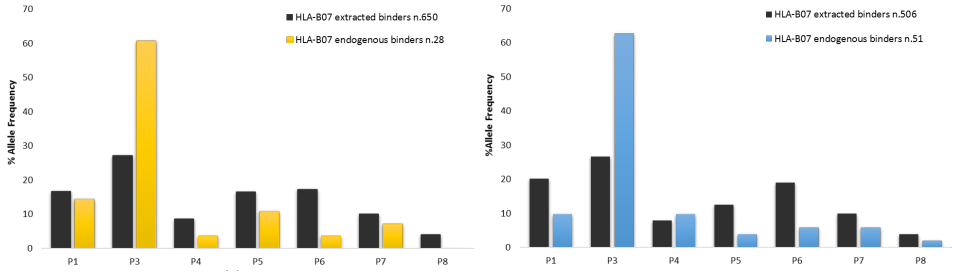


Figure.5: *Positional frequency of modified residues in HLA-B*07-associated arginine (di)methylated peptides. Positional frequency of the site of modification of the identified peptides in comparison to an extracted set of predicted HLA-B*07 peptides with a previously identified protein arginine (di)methylation site (PhosphoSite.org). Comparative analysis is shown for methylated (left panel, orange) and di-methylated (right panel, blue) arginine residues. For the extracted set of peptides, all known human protein (di)methylation sites were extracted from PhosphoSite.org database and potential HLA-B*07 binding peptides (9mers) were predicted using the NetMHC algorithm. The positional frequency of the (di)methylated arginine residues in this datasets is show by black bars.*

attributed to underlying PRMTs substrate motifs that resemble the peptide binding motif. Still, the potential substrate motifs do not fully explain the more frequent and more abundant presentation of arginine di-methylated peptides, suggesting that additional structural features in HLA-B*07-peptide binding favor a di-methyl group at the position P3 arginine residue.

Arginine (di)methylated HLA class I ligands exhibit substrate motifs linked to specific PRMTs

We then examined the amino residues flanking the identified modified sites to investigate whether the modification had been performed by specific PRMTs. The aligned sequences of the HLA class I arginine methylated and di-methylated ligands (Figure.6, upper panels) were submitted to IceLogo(17) to calculate the over-representation of the amino acids flaking the modified arginine. For comparison, we extracted from PhophoSite.org all from previous studies deposited protein arginine methylated and di-methylated sites and subjected them to an alike analysis (Figure.6, bottom panels). The sites extracted from this latter analysis reveals that arginine methylation is primarily occurring at glycine-arginine-rich (GAR) stretches.(25) In our HLA class I ligandome both methylated and di-methylated data sets displayed a distinctively different preference, namely in PRP and PRG-rich motifs. This was expected, as this PRP and PRG substrate motif resembles nicely the proline anchor residue at position 2 and the favored (di)methylated arginine residue at position P3 of HLA-B*07-associated peptides. It has been established that most PRMTs methylate on glycine- and arginine-rich (GAR) motifs in their substrates, except PRMT4/CARM-1 and PRMT5 that both seem to be able to methylate also proline-rich motifs.(26) Distinctively, PRMT4/CARM-1 asymmetrically di-methylates arginine residues (ADMA), whereas PRMT5 symmetrically di-methylates arginine (SDMA) residues (Figure.1A). To investigate which of these two PRMTs are specifically involved in the arginine modification of the identified HLA class I ligands, we further inspected the MS/MS fragmentation spectra of our di-methylated HLA peptides as described in the methods section. Making use of distinct patterns in neutral losses (27) we were able to annotate 30 ADMA peptides and 4 SDMA peptides (some examples of the spectra are in Supplementary Figures 15 to 18). The verified 30 ADMA peptides were mostly predicted to be HLA-B*07 binders (83%) (Supplementary Figure.2). One 13 of the verified ADMA peptides showed '-RP' motifs, thus most likely being modified by CARM1. (28) The number of '-RG' motifs detected in this subset of peptides was 9. The remaining motifs contained L/Q/I/Y in position +1 of the methylation site, which is line with other reports which showed 'unconventional' motifs recognized by PRMTs type I.(10, 29) The 4 SDMA verified

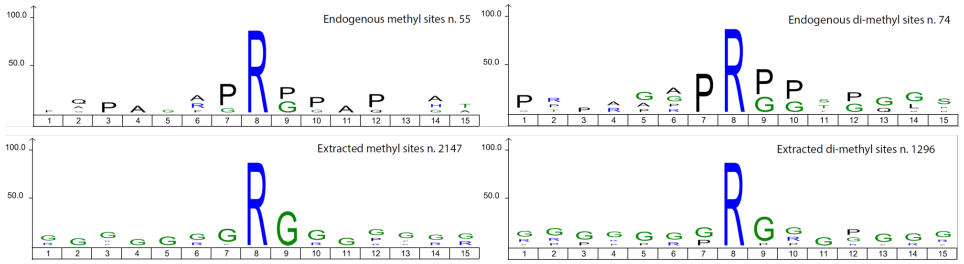


Figure.6: *Sequence motif analysis of the methylated and di-methylated sites. Motif analysis of preferred amino acids flanking the arginine methylation and di-methylation sites identified on HLA class I (top panels). For comparison, on the bottom panels are shown the motif analysis of the arginine methylated and di-methylated sites annotated in the PhoSphosite.org database.*

peptides, which are likely substrates of PRMT5, contained both '-RG' and '-RP' rich motifs.

A subset of (di)methylated HLA ligands is derived from source proteins associated with cancer

As depicted in Figure 2 quite a few of the source proteins of especially the (di)methylated HLA peptides in GR cells included proteins that have been linked to cancer (Figure.2, red circles). Illustrative examples include the proteins ARID1A, ARID3B and CRIP1 which were found in our HLA class I ligandome both with arginine mono-methylated and di-methylated peptides. These proteins have roles as a tumor suppressor, (30) and to increase tumor growth in ovarian cancer, respectively.(31) Additionally, CRIP1 is overexpressed in several tumors.(32),(33),(34),(35). Moreover, the melanoma-associated antigen, MAGED2, as other MAGE proteins, act as anti-tumoral immune targets. (36, 37) Furthermore, NCoA-3(38) is known to be deregulated in cancer and found in our data set multiple times, both di-methylated and methylated HLA antigens.

4. DISCUSSION

Post-translationally modified HLA class I peptide antigens can be specifically recognized by the immune system.(39–41) As certain PTMs (e.g. phosphorylation) represent hallmarks of human diseases (37, 42–46) these modified HLA class I peptides have attained significant attention as candidates for immunotherapy or vaccination. Here, we report the identification of 55 unique arginine methylated, 74 arginine di-methylated and 20 HLA class I peptides with multiple (di)methylation sites from two replicate ligandome studies on a heterozygous B cell line. The total number of 149 arginine (di)methylated peptide antigens represent approximately 1% of the total repertoire of unique HLA class I peptides, a frequency of presentation which is in agreement with other detected PTMs, such as phosphorylation,(47) glycosylation, (13) asparagine deamidation. (14) Although the number of identified arginine (di)methylated peptides is relatively small, the results reported here still indicate that this type of modification is frequently presented by HLA class I for recognition by T cells. The (di)methylated peptide ligands originated from 105 source proteins. GO analysis revealed that the source proteins were highly enriched in translation, RNA-processing, transcription and chromatin regulation, which is in close agreement with the main regulatory functions of arginine methylation. (48, 49) A comparison of the number of arginine (di)methylated peptides predicted to be bound to the expressed HLA-A and –B alleles revealed enhanced presentation of peptides by HLA-B*07. Interestingly, in our data set we also found a strikingly high positional prevalence for arginine (di)methylation sites at P3 of the HLA-B*07-presented peptide sequences. Re-analysis of the HLA-B*07 positive B-cell ligandome data set published (data not shown) by Hassan et al (3) revealed a comparable allele-specific enrichment and positional preference in arginine (di)methylated peptides. We argue that this enhanced presentation is related to the similarity between the binding motif of HLA-B*07 and the substrate amino acids for PRMTs. The consensus binding motif of HLA-B*07 is highly enriched in proline residues at position P2 and prefers arginine at position P3. (50) Thus, not surprisingly many of the here detected position P3 arginine (di)methylated peptides exhibit so-called Proline-Arginine rich domains of PRMTs.(48, 49, 51) Such a Proline-Arginine rich motif is not the most common motif observed for PRMTs (that is the Glycine-Arginine rich motif), but has been linked to a few less common, albeit selective PRMTs.

The underlying Proline-Arginine rich motif cannot fully explain the high selectivity of HLA-B*07 in presenting arginine di-methylated ligands. First, a higher number of arginine di-methylated than arginine methylated peptide ligands were found, which is in contradiction with the overall frequency of protein arginine (di)methylation sites found in-depth studies of cellular proteome. (24) Second, the positional preference for P3 arginine (di)methylation was only partially observed in a large set of in silico predicted 9-mer HLA-B*07-associated peptides harboring previously identified modification sites (PhosphoSite.org database). Third, quantitative comparison of peptides pairs with differentially modified methylation sites revealed that arginine di-methylated peptides were presented at higher abundances compared to their arginine methylated peptide counterparts and in most of the cases of their unmodified counterparts too. Hence, we reasoned that additional structural features in HLA-B*07 could play a role in the selective presentation arginine di-methylated peptides, but additional binding assays or crystallization experiments are required to understand the underlying mechanisms of binding. Large scale proteomics studies have also revealed that the majority of protein arginine methylations are mediated on glycine- and arginine-rich motifs (GAR motifs), (25) although a few motifs beyond the 'RGG' paradigm have been recognized. (29) In the study performed by Yagüe et al. the identified 10-mer was found methylated in the context of HLA-A*39 and also contained the canonical 'RGG motif'. While our analysis of the enriched amino acids flanking the modified methylated and di-methylated sites in the HLA class I peptide dataset reveal distinct motifs beyond the 'RGG' paradigm. In fact we observed the enrichment of proline- and arginine-rich motifs, which, as discussed above, relates to the consensus binding motif of HLA-B*07. Specific PRMTs are known to target proline-enriched motifs, notably CARM1 and PRMT5. CARM1 preferentially modifies protein substrates containing proline-, glycine- and methionine-rich motifs (PGM motifs), (28) while PRMT5 can also modify substrates with PGM motifs (51). At first glance, we therefore hypothesized that PRMT5 and CARM1 are the likely enzymes that have modified the majority of the here reported HLA class I peptides. Type I PRMT enzymes catalyze protein mono-methylation (MMA) and asymmetrical di-methylation (ADMA), while type II PRMT enzymes catalyze mono-methylation and symmetrical di-methylation (SDMA)(52) (Figure.1A). It has been shown that neutral losses occurring in the MS/MS spectra of arginine di-methylated peptides can be used to assign ADMA and SMDA modified peptide. (19) We employed this approach to our data and classified 30 of the di-methylated HLA class I peptides to harbor ADMA and 4 to harbor SMDA. Since CARM1 leads to ADMA, whereas PRMT5 produces SMDA, most of the here detected di-methylated HLA class I peptides are thus putative CARM1 (28) substrates. While the 3 confirmed SDMA peptides, displaying mixed '-RG' and '-RP' motifs, are putative PRMT5 substrates. (52) Protein arginine (di)methylation has been linked to carcinogenesis and metastasis. (53) For instance, the transcriptional co-activator NCoA-3 (38) is known to be deregulated in cancer and found in our data set with representative arginine methylated and arginine di-methylated HLA class I peptide antigens (Figure.2). Notably, aggressive breast tumors overexpressing CARM1 also have high levels of the oncogenic co-activator NCoA-3. (54) It has been hypothesized that the oncogenic properties of NCoA-3 may be attenuated by inhibiting CARM1 activity.(53) Another notable example is ARID1A (a tumor suppressor), for which decreased of levels has been linked to poor prognosis. (30),(55, 56),(57),(58) We observed both the arginine methylated and arginine di-methylated HLA class I peptides originating from ARID1A. Evidently the here reported arginine (di)methylated peptides are self-antigens, presented on GR B lymphoblasts. Although the GR cell line does not represent cancerous cell line, it

has an immortalized phenotype after transformation with Epstein Barr virus. Therefore, we argue that GR-derived methylated peptides could represent interesting therapeutic antigens since their proteins of origin are often overexpressed in cancer tissues, combined with the fact that methylation or di-methylation may affect the protein turnover and transcription. Translation of these findings would however first require the demonstration that there is a T cell repertoire recognizing the HLA class I presented methylated epitopes involved, as was first demonstrated by Jarmalavicus. (12) More generally, the rules and selectivity's of (di)methylated peptide antigens presentation that we extract from our data will be of help to predict and identify disease-related HLA class I antigens.

5. ACKNOWLEDGEMENTS

This work was partly supported by the project Proteins At Work (project 184.032.201), a program of the Netherlands Proteomics Centre financed by the Netherlands Organization for Scientific Research (NWO) as part of the National Roadmap Large-scale Research Facilities of the Netherlands, and by the projects Immunoproteomics and Correlates of Protection, financed by the Dutch Ministry of Health. This work was further supported by the Institute for Chemical Immunology, an NWO Gravitation project funded by the Ministry of Education, Culture and Science of the Netherlands.

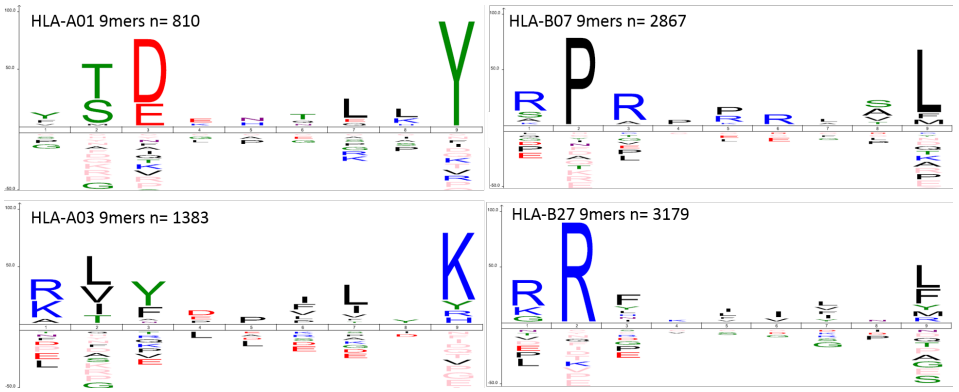
6. REFERENCES

1. Neefjes, J., Jongsma, M. L., Paul, P., and Bakke, O. (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol 11, 823–836
2. Mester, G., Hoffmann, V., and Stevanović, S. (2011) Insights into MHC class I antigen processing gained from large-scale analysis of class I ligands. Cell. Mol. Life Sci. 68, 1521–1532
3. Hassan, C., Kester, M. G. D., de Ru, A. H., Hombrink, P., Drijfhout, J. W., Nijveen, H., Leunissen, J. a M., Heemskerk, M. H. M., Falkenburg, J. H. F., and van Veelen, P. a (2013) The human leukocyte antigen-presented ligandome of B lymphocytes. Mol. Cell. Proteomics 12, 1829–43
4. Engelhard, V. H., Altrich-Vanlith, M., Ostankovitch, M., and Zarling, A. L. (2006) Post-translational modifications of naturally processed MHC-binding epitopes. Curr Opin Immunol 18, 92–97
5. van Els, C. A. C. M., Corbière, V., Smits, K., van Gaans-van den Brink, J. A. M., Poelen, M. C. M., Mascart, F., Meiring, H. D., and Locht, C. (2014) Toward Understanding the Essence of Post-Translational Modifications for the Mycobacterium tuberculosis Immunoproteome. Front. Immunol. 5, 361
6. Sollid, L. M., Pos, W., and Wucherpfennig, K. W. (2014) Molecular mechanisms for contribution of MHC molecules to autoimmune diseases. Curr. Opin. Immunol. 31, 24–30
7. McGinty, J. W., Marré, M. L., Bajzik, V., Piganelli, J. D., and James, E. A. (2015) T cell epitopes and post-translationally modified epitopes in type 1 diabetes. Curr. Diab. Rep. 15, 90
8. Cobbold, M., Peña, H. D. La, Norris, A., Polefrone, J., Qian, J., Michelle, a, Cummings, K., Penny, S., Turner, J. E., Cottine, J., and Jennifer, G. (2014) Memory-Like Immunity in Leukemia. 5,
9. Weinhold, B. (2006) Epigenetics: the science of change. Environ. Health Perspect. 114, A160–7
10. Uhlmann, T., Geoghegan, V. L., Thomas, B., Ridlova, G., Trudgian, D. C., and Acuto, O. (2012) A method for large-scale identification of protein arginine methylation. Mol Cell Proteomics 11, 1489–1499
11. Yagüe, J., Vázquez, J., and López de Castro, J. A. (2000) A post-translational modification of nuclear proteins, N(G),N(G)-dimethyl-Arg, found in a natural HLA class I peptide ligand. Protein Sci. 9, 2210–7
12. Jarmalavicius, S., Trefzer, U., and Walden, P. (2010) Differential arginine methylation of the G-protein pathway suppressor GPS-2 recognized by tumor-specific T cells in melanoma. FASEB J. 24, 937–946
13. Marino, F., Bern, M., Mommen, G. P. M., Leney, A. C., van Gaans-van den Brink, J. a. M., Bonvin, A. M. J. J., Becker, C., van Els, C. a. C. M., and Heck, A. J. R. (2015) Extended O-GlcNAc on HLA Class-I-Bound Peptides. J. Am. Chem. Soc., 150819112655001
14. Mommen, G. P., Frese, C. K., Meiring, H. D., van Gaans-van den Brink, J., de Jong, A. P., van Els, C. A., and Heck, A. J. (2014) Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proc Natl Acad Sci U S A 111, 4507–4512
15. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., and MacCoss, M. J. (2007) Semi-supervised learning for

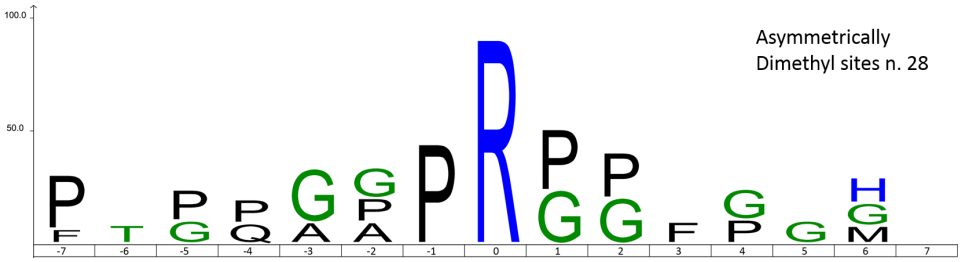
peptide identification from shotgun proteomics datasets. Nat Methods 4, 923–925
16. Nielsen, M., Lundegaard, C., Worning, P., Lauemøller, S. L., Lamberth, K., Buus, S., Brunak, S., and Lund, O. (2003) Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci. 12, 1007–1017
17. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009) Improved visualization of protein consensus sequences by iceLogo. Nat Methods 6, 786–787
18. Mi, H., Muruganujan, A., Casagrande, J. T., and Thomas, P. D. (2013) Large-scale gene function analysis with the PANTHER classification system. Nat Protoc 8, 1551–1566
19. Gehrig, P. M., Hunziker, P. E., Zahariev, S., and Pongor, S. (2004) Fragmentation pathways of N(G)-methylated and unmodified arginine residues in peptides studied by ESI-MS/MS and MALDI-MS. J Am Soc Mass Spectrom 15, 142–149
20. Brame, C. J., Moran, M. F., and McBroom-Cerajewski, L. D. (2004) A mass spectrometry based method for distinguishing between symmetrically and asymmetrically dimethylated arginine residues. Rapid Commun Mass Spectrom 18, 877–881
21. Olsen, J. V, and Mann, M. (2004) Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. Proc. Natl. Acad. Sci. U. S. A. 101, 13417–22
22. Frese, C. K., Altelaar, A. F. M., van den Toorn, H., Nolting, D., Griep-Raming, J., Heck, A. J. R., and Mohammed, S. (2012) Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. Anal. Chem. 84, 9668–9673
23. Frese, C. K., Zhou, H., Taus, T., Altelaar, A. F., Mechtler, K., Heck, A. J., and Mohammed, S. (2013) Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). J Proteome Res 12, 1520–1525
24. Guo, A., Gu, H., Zhou, J., Mulhern, D., Wang, Y., Lee, K. A., Yang, V., Aguiar, M., Kornhauser, J., Jia, X., Ren, J., Beausoleil, S. A., Silva, J. C., Vemulapalli, V., Bedford, M. T., and Comb, M. J. (2014) Immunoaffinity enrichment and mass spectrometry analysis of protein methylation. Mol Cell Proteomics 13, 372–387
25. Najbauer, J., Johnson, B. a., Young, a. L., and Aswad, D. W. (1993) Peptides with sequences similar to glycine, arginine-rich motifs in proteins interacting with RNA are efficiently recognized by methyltransferase(s) modifying arginine in numerous proteins. J. Biol. Chem. 268, 10501–10509
26. Bedford, M. T. (2007) Arginine methylation at a glance. J. Cell Sci. 120, 4243–4246
27. Wang, H., Straubinger, R. M., Aletta, J. M., Cao, J., Duan, X., Yu, H., and Qu, J. (2009) Accurate localization and relative quantification of arginine methylation using nanoflow liquid chromatography coupled to electron transfer dissociation and orbitrap mass spectrometry. J Am Soc Mass Spectrom 20, 507–519
28. Cheng, D., Côté, J., Shaaban, S., and Bedford, M. T. (2007) The Arginine Methyltransferase CARM1 Regulates the Coupling of Transcription and mRNA Processing. Mol. Cell 25, 71–83
29. Wooderchak, W. L., Zang, T., Zhou, Z. S., Acuña, M., Tahara, S. M., and Hevel, J. M. (2008) Substrate profiling of PRMT1 reveals amino acid sequences that extend beyond the “RGG” paradigm. Biochemistry 47, 9456–9466
30. Kim, M. J., Gu, M. J., Chang, H.-K., and Yu, E. (2015) Loss of ARID1A expression is associated with poor prognosis in small intestinal carcinoma. Histopathology 66, 508–516
31. Bobbs, A., Gellerman, K., Hallas, W. M., Joseph, S., Yang, C., Kurkewich, J., and Cowden Dahl, K. D. (2015) ARID3B Directly Regulates Ovarian Cancer Promoting Genes. PLoS One 10, e0131961
32. Ma, X.-J., Salunga, R., Tuggle, J. T., Gaudet, J., Enright, E., McQuary, P., Payette, T., Pistone, M., Stecker, K., Zhang, B. M., Zhou, Y.-X., Varnholt, H., Smith, B., Gadd, M., Chatfield, E., Kessler, J., Baer, T. M., Erlander, M. G., and Sgroi, D. C. (2003) Gene expression profiles of human breast cancer progression. Proc. Natl. Acad. Sci. U. S. A. 100, 5974–5979
33. Groene, J., Mansmann, U., Meister, R., Staub, E., Roepcke, S., Heinze, M., Klaman, I., Brümmendorf, T., Hermann, K., Loddenkemper, C., Pilarsky, C., Mann, B., Adams, H.-P., Buhr, H. J., and Rosenthal, A. (2006) Transcriptional census of 36 microdissected colorectal cancers yields a gene signature to distinguish UICC II and III. Int. J. Cancer 119, 1829–1836
34. Wang, Q., Williamson, M., Bott, S., Brookman-Amissah, N., Freeman, a, Nariculam, J., Hubank, M. J. F., Ahmed, a, and Masters, J. R. (2007) Hypomethylation of WNT5A, CRIP1 and S100P in prostate cancer. Oncogene 26, 6560–6565
35. Baumhoer, D., Elsner, M., Smida, J., Zillmer, S., Rauser, S., Schoene, C., Balluff, B., Bielack, S., Jundt, G., Walch, A., and Nathrath, M. (2011) CRIP1 expression is correlated with a favorable outcome and less metastases in osteosarcoma patients. Oncotarget 2, 970–975
36. Xiao, J., and Chen, H. (2004) Biological functions of melanoma-associated antigens. 10, 1849–1853
37. Sadanaga, N., Nagashima, H., Mashino, K., Tahara, K., Yamaguchi, H., Ohta, M., Fujie, T., Tanaka, F., Inoue, H., Takesako, K., Akiyoshi, T., and Mori, M. (2001) Dendritic cell vaccination with MAGE peptide is a novel therapeutic approach for gastrointestinal carcinomas. Clin. Cancer Res. 7, 2277–2284
38. Burandt, E., Jens, G., Holst, F., Jänicke, F., Müller, V., Quaas, a., Choschzick, M., Wilczak, W., Terracciano, L., Si-

mon, R., Sauter, G., and Lebeau, a. (2013) Prognostic relevance of AIB1 (NCoA3) amplification and overexpression in breast cancer. Breast Cancer Res. Treat. 137, 745–753
39. Engelhard, V. H., Altrich-Vanlith, M., Ostankovitch, M., and Zarling, A. L. (2006) Post-translational modifications of naturally processed MHC-binding epitopes. Curr. Opin. Immunol. 18, 92–97
40. Skipper, J. C., Hendrickson, R. C., Gulden, P. H., Brichard, V., Van Pel, A., Chen, Y., Shabanowitz, J., Wolfel, T., Slingluff, C. L., Boon, T., Hunt, D. F., and Engelhard, V. H. (1996) 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. 183, 527–34
41. Meadows, L., Wang, W., den Haan, J. M. ., Blokland, E., Reinhardus, C., Drijfhout, J. W., Shabanowitz, J., Pierce, R., Agulnik, A. I., Bishop, C. E., Hunt, D. F., Goulmy, E., and Engelhard, V. H. (1997) The HLA-A*0201-Restricted H-Y Antigen Contains a Posttranslationally Modified Cysteine That Significantly Affects T Cell Recognition. Immunity 6, 273–281
42. Girbal-neuhauser, E., Durieux, J., Dalbon, P., Sebbag, M., Vincent, C., Simon, M., Senshu, T., Masson-bessière, C., Jolivet-reynaud, C., Jolivet, M., Serre, G., Girbal-neuhauser, E., Durieux, J., Arnaud, M., Dalbon, P., Sebbag, M., Vincent, C., Simon, M., Senshu, T., Masson-bessie, C., Jolivet-reynaud, C., Jolivet, M., and Serre, G. (2015) This information is current as of November 5, 2015.
43. Arentz-Hansen, H., Körner, R., Molberg, O., Quarsten, H., Vader, W., Kooy, Y. M., Lundin, K. E., Koning, F., Roepstorff, P., Sollid, L. M., and McAdam, S. N. (2000) The intestinal T cell response to alpha-gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase. J Exp Med 191, 603–612
44. Utz, P. J., Hottelet, M., Schur, P. H., and Anderson, P. (1997) Proteins Phosphorylated during Stress-induced Apoptosis Are Common Targets for Autoantibody Production in Patients with Systemic Lupus Erythematosus. J. Exp. Med. 185, 834–854
45. Doyle, H. a., Zhou, J., Wolff, M. J., Harvey, B. P., Roman, R. M., Gee, R. J., Koski, R. a., and Mamula, M. J. (2006) Isoaspartyl Post-translational Modification Triggers Anti-tumor T and B Lymphocyte Immunity. J. Biol. Chem. 281, 32676–32683
46. Taguchi, A., Taylor, A. D., Rodriguez, J., Celiktaş, M., Liu, H., Ma, X., Zhang, Q., Wong, C.-H., Chin, A., Girard, L., Behrens, C., Lam, W. L., Lam, S., Minna, J. D., Wistuba, I. I., Gazdar, A. F., and Hanash, S. M. (2014) A search for novel cancer/testis antigens in lung cancer identifies VCX/Y genes, expanding the repertoire of potential immunotherapeutic targets. Cancer Res. 74, 4694–705
47. Mohammed, F., Cobbold, M., Zarling, A. L., Salim, M., Barrett-Wilt, G. A., Shabanowitz, J., Hunt, D. F., Engelhard, V. H., and Willcox, B. E. (2008) Phosphorylation-dependent interaction between antigenic peptides and MHC class I: a molecular basis for the presentation of transformed self. Nat. Immunol. 9, 1236–43
48. Geoghegan, V., Guo, A., Trudgian, D., Thomas, B., and Acuto, O. (2015) Comprehensive identification of arginine methylation in primary T cells reveals regulatory roles in cell signalling. Nat. Commun. 6, 6758
49. Guo, a., Gu, H., Zhou, J., Mulhern, D., Wang, Y., Lee, K. a., Yang, V., Aguiar, M., Kornhauser, J., Jia, X., Ren, J., Beausoleil, S. a., Silva, J. C., Vemulapalli, V., Bedford, M. T., and Comb, M. J. (2014) Immunoaffinity Enrichment and Mass Spectrometry Analysis of Protein Methylation. Mol. Cell. Proteomics 13, 372–387
50. Sidney, J., del Guercio, M. F., Southwood, S., Engelhard, V. H., Appella, E., Rammensee, H. G., Falk, K., Rötzschke, O., Takiguchi, M., and Kubo, R. T. (1995) Several HLA alleles share overlapping peptide specificities. J. Immunol. 154, 247–59
51. Bedford, M. T., and Clarke, S. G. (2009) Protein Arginine Methylation in Mammals: Who, What, and Why. Mol. Cell 33, 1–13
52. Bedford, M. T., and Clarke, S. G. (2009) Protein arginine methylation in mammals: who, what, and why. Mol Cell 33, 1–13
53. Yang, Y., and Bedford, M. T. (2012) Protein arginine methyltransferases and cancer. Nat. Rev. Cancer 13, 37–50
54. El Messaoudi, S., Fabbrizio, E., Rodriguez, C., Chuchana, P., Fauquier, L., Cheng, D., Theillet, C., Vandel, L., Bedford, M. T., and Sardet, C. (2006) Coactivator-associated arginine methyltransferase 1 (CARM1) is a positive regulator of the Cyclin E1 gene. Proc. Natl. Acad. Sci. U. S. A. 103, 13351–6
55. Wang, D., Chen, Y., Pan, K., Wang, W., Chen, S., Chen, J., Zhao, J., Lv, L., Pan, Q., Li, Y., Wang, Q., Huang, L., Ke, M., He, J., and Xia, J. (2012) Decreased Expression of the ARID1A Gene Is Associated with Poor Prognosis in Primary Gastric Cancer. PLoS One 7, e40364
56. Yan, H.-B., Wang, X.-F., Zhang, Q., Tang, Z.-Q., Jiang, Y.-H., Fan, H.-Z., Sun, Y.-h., Yang, P.-Y., and Liu, F. (2014) Reduced expression of the chromatin remodeling gene ARID1A enhances gastric cancer cell migration and invasion via downregulation of E-cadherin transcription. Carcinogenesis 35, 867–876
57. Jones, S., Wang, T., Shih, I., Mao, T., Nakayama, K., Glas, R., Slamon, D., Jr, L. a D., Vogelstein, B., Kenneth, W., Velculescu, V. E., and Papadopoulos, N. (2011) NIH Public Access. 330, 228–231
58. He, F., Li, J., Xu, J., Zhang, S., Xu, Y., Zhao, W., Yin, Z., and Wang, X. (2015) Decreased expression of ARID1A associates with poor prognosis and promotes metastases of hepatocellular carcinoma. J. Exp. Clin. Cancer Res. 34, 47

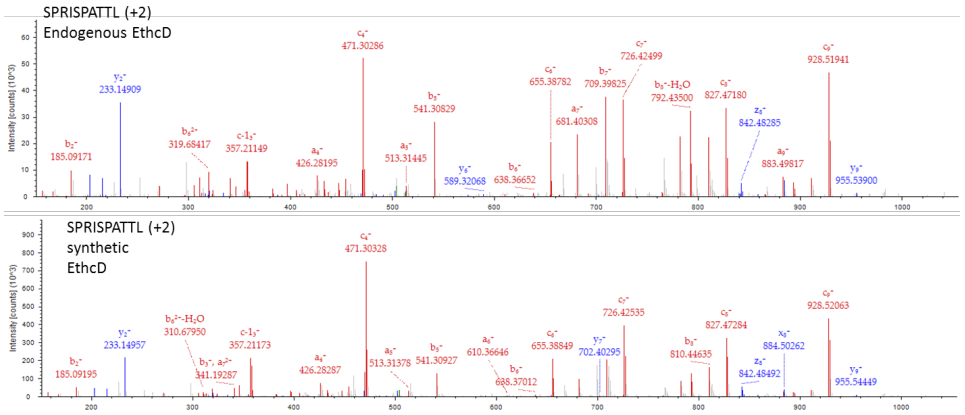
7. SUPPLEMENTARY



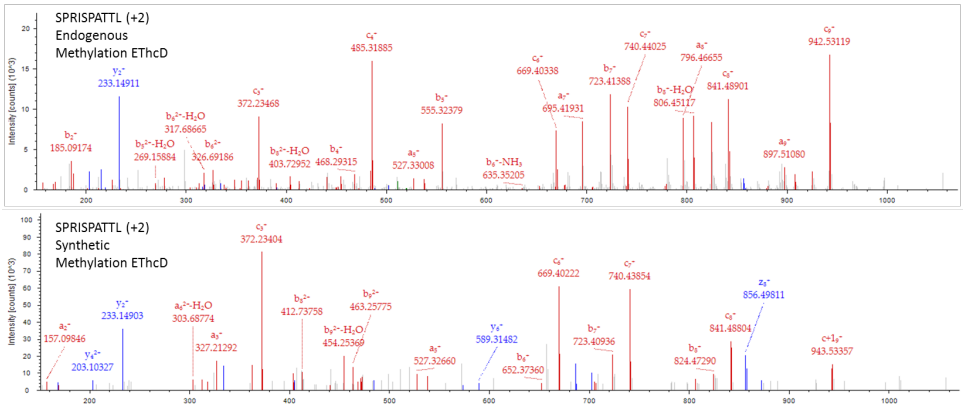
Supplementary Figure.1: *Sequence motifs of the identified unmodified HLA class I peptides. The NetMHC algorithm was used for binding affinity prediction and assignment of the set of identified (unmodified) peptides to HLA-A*01, -A*03, -B*07 and –B*27. Analysis was restricted to peptides 9 amino acids in length.*



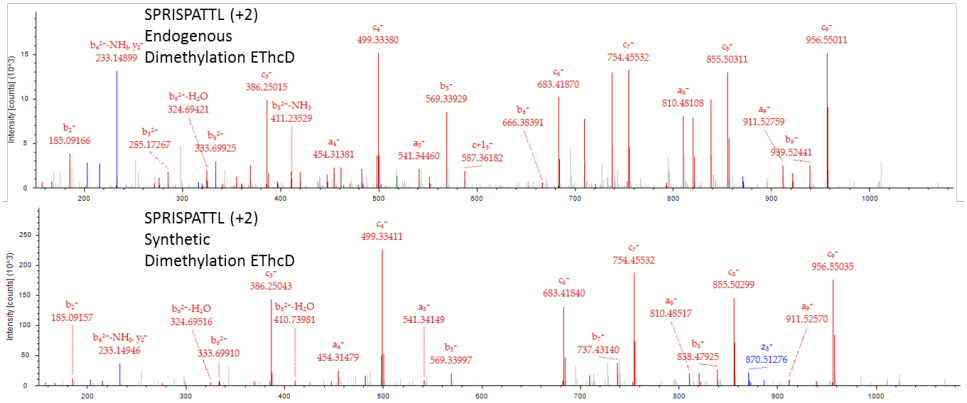
Supplementary Figure.2: *Sequence motif of asymmetrically di-methylated sites. The sequence motif shows the preference of amino acid flanking the asymmetrically di-methylation (ADMA) sites identified on HLA class I peptides.*



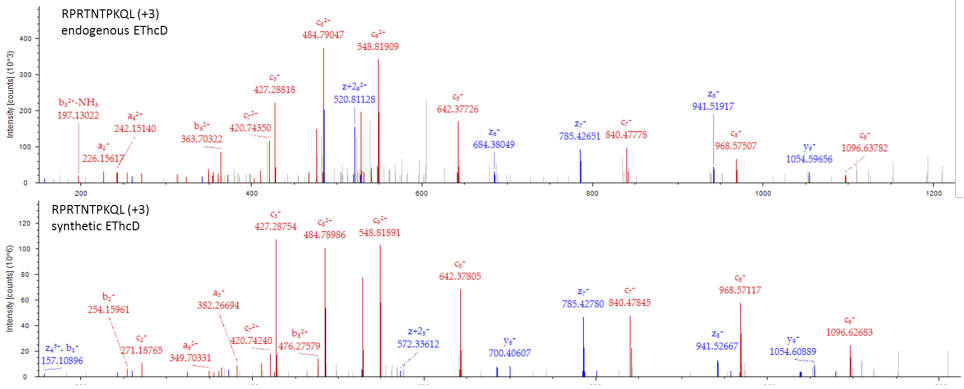
Supplementary Figure.3: *EThcD spectra of the endogenous unmodified HLA class I peptide SPRISPATTL (top). For comparison the EThcD spectrum of the synthetic peptide analogue is given (bottom). This peptide originates from the AT-Rich Interactive Domain-Containing Protein 3B.*



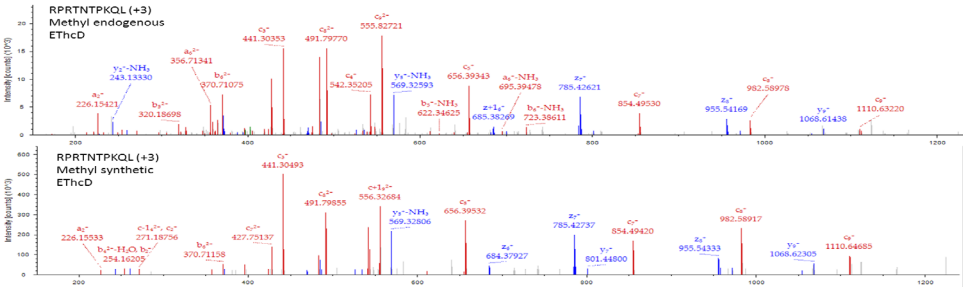
Supplementary Figure.4: *EThcD spectra of the endogenous P3 methylated HLA class I peptide SPRISPATTL (top). For comparison the EThcD spectra of the synthetic peptide analogue is given (bottom). The peptide originates from the protein AT-Rich Interactive Domain-Containing Protein 3B.*



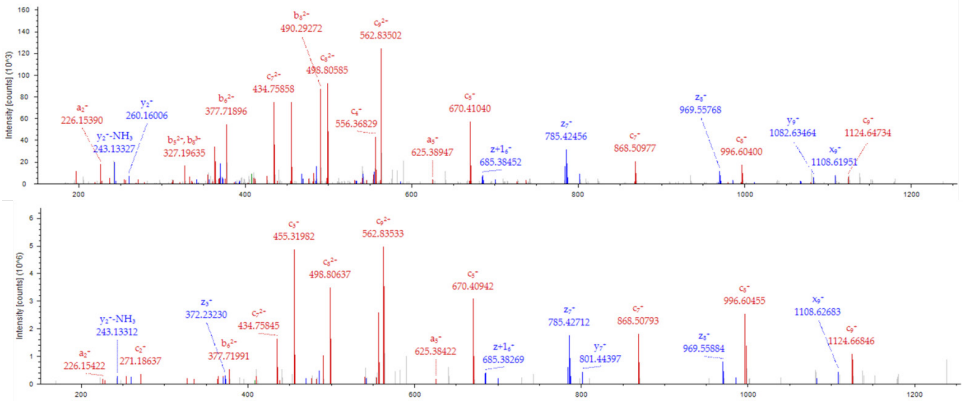
Supplementary Figure.5: *EThcD spectra of the endogenous P3 di-methylated HLA class I peptide SPRISPATTL (top). For comparison the EThcD spectra of the synthetic peptide analogue is given (bottom). The peptide originates from the protein AT-Rich Interactive Domain-Containing Protein 3B.*



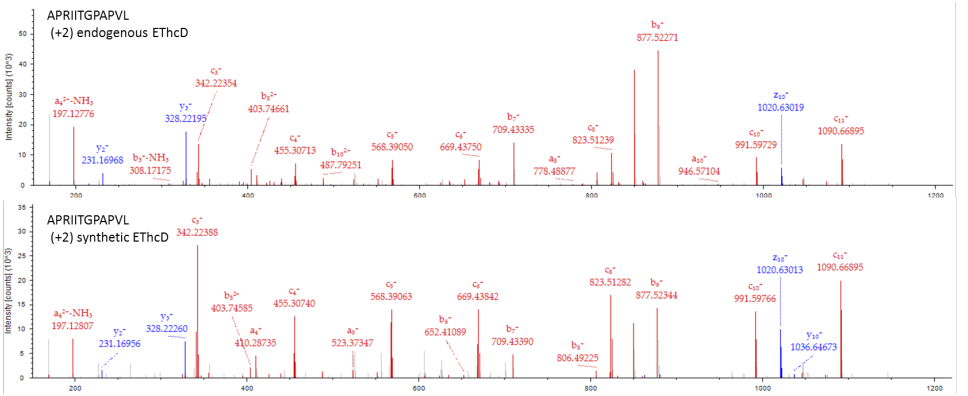
Supplementary Figure.6: *EThcD spectra of the endogenous unmodified HLA class I peptide RPRTNTPKQL (top). For comparison the EThcD spectrum of the synthetic peptide analogue is given (bottom). This peptide originates from the protein Nuclear Receptor Coactivator 3.*



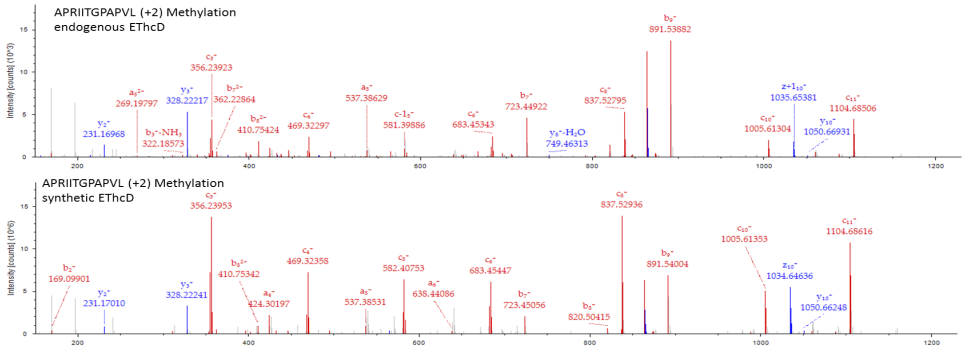
Supplementary Figure.7: *EThcD spectra of the P3 methyl modified peptide RPRTNTPKQL HLA class I peptide RPRTNTPKQL (top). For comparison the EThcD spectrum of the synthetic peptide analogue is given (bottom). This peptide originates from the protein Nuclear Receptor Coactivator 3.*



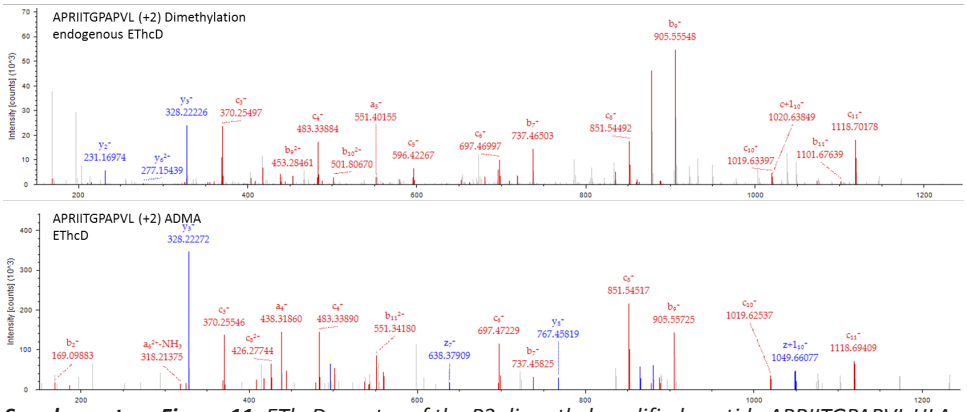
Supplementary Figure.8: *EThcD spectra of the P3 di-methyl modified peptide RPRTNTPKQL HLA class I peptide RPRTNTPKQL (top). For comparison the EThcD spectrum of the synthetic peptide analogue is given (bottom). This peptide originates from the protein Nuclear Receptor Coactivator 3.*



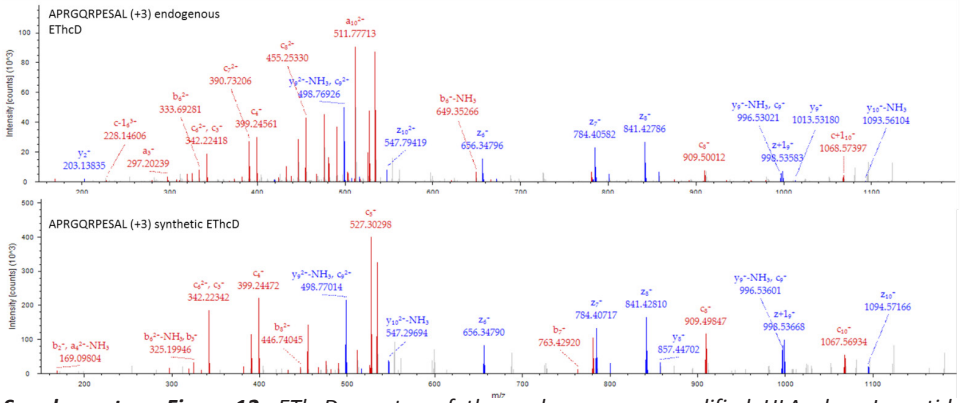
Supplementary Figure.9: *EThcD spectra of the endogenous unmodified HLA class I peptide APRIITGPAPVL (top). For comparison the EThcD spectrum of the synthetic peptide analogue is given (bottom). This peptide originates from the protein of origin the protein Protein Quaking.*



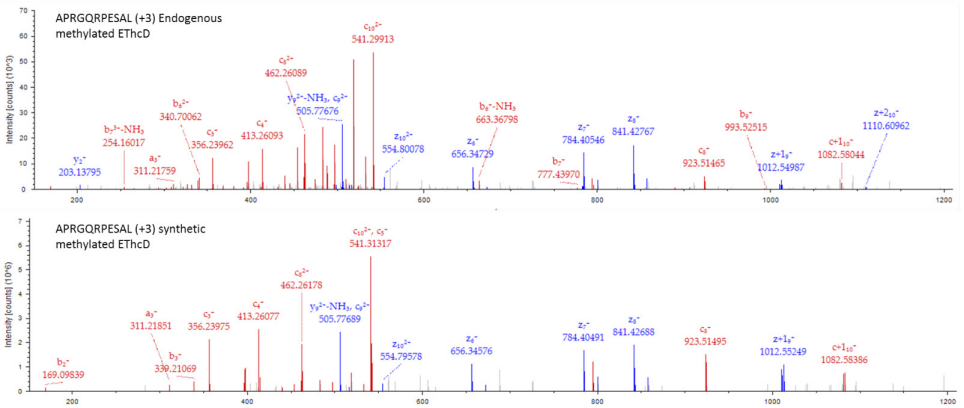
Supplementary Figure.10: *EThcD spectra of the P3 methyl modified peptide APRIITGPAPVL HLA class I peptide (top). For comparison the EThcD spectrum of the synthetic peptide analogue is given (bottom). This peptide originates from the protein Protein Quaking.*



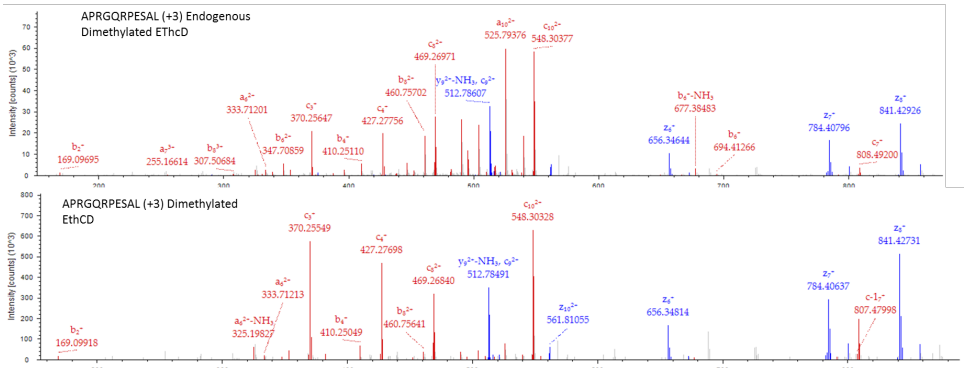
Supplementary Figure.11: *EThcD spectra of the P3 di-methyl modified peptide APRIITGPAPVL HLA class I peptide (top). For comparison the EThcD spectrum of the synthetic peptide analogue is given (bottom). This peptide originates from the protein Protein Quaking.*



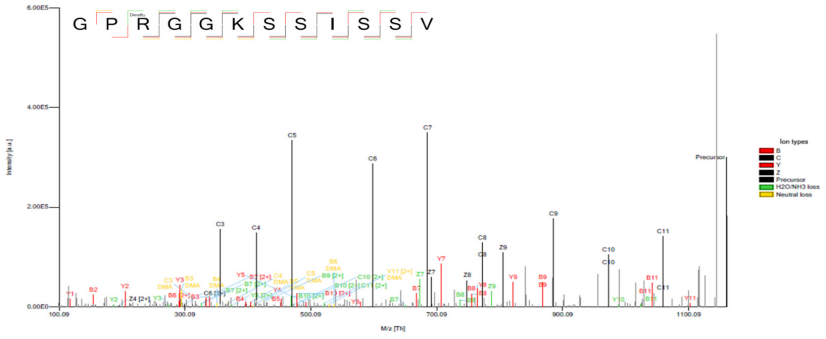
Supplementary Figure.12: *EThcD spectra of the endogenous unmodified HLA class I peptide APRGQRPESAL (top). For comparison the EThcD spectrum of the synthetic peptide analogue is given (bottom). This peptide originates from the protein of origin the protein Proteasome Subunit Beta Type-8.*



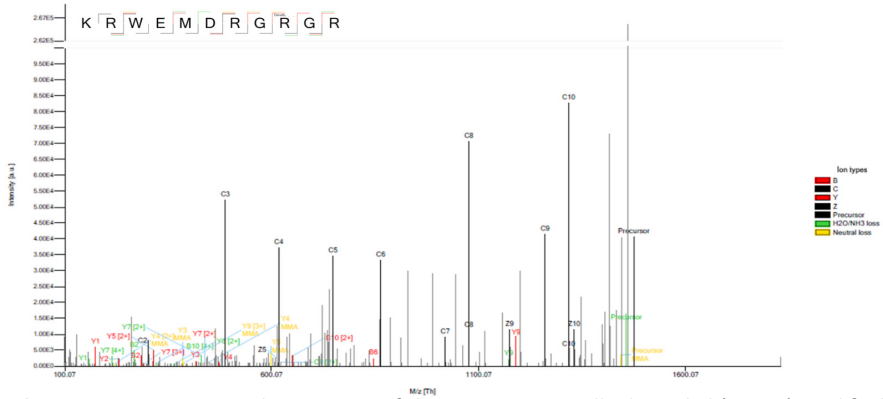
Supplementary Figure.13: *EThcD spectra of the P3 methyl modified peptide APRGQRPESAL HLA class I peptide (top). For comparison the EThcD spectrum of the synthetic peptide analogue is given (bottom). This peptide originates from the protein Proteasome Subunit Beta Type-8.*



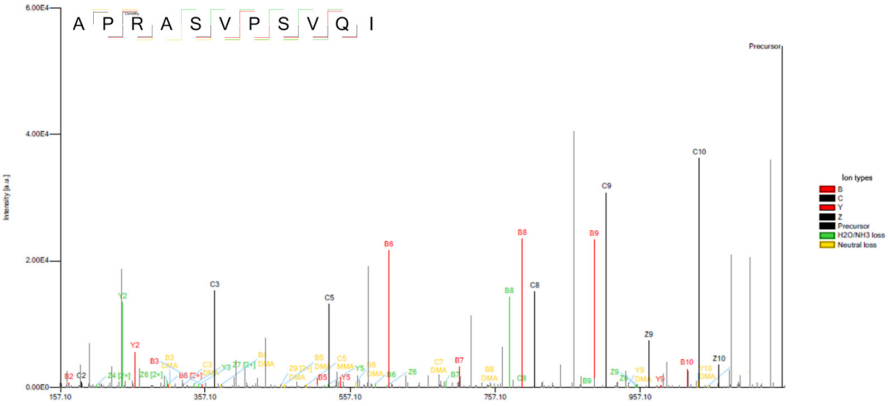
Supplementary Figure.14: *EThcD spectra of the P3 di-methyl modified peptide APRGQRPESAL HLA class I peptide (top). For comparison the EThcD spectrum of the synthetic peptide analogue is given (bottom). This peptide originates from the protein Proteasome Subunit Beta Type-8.*



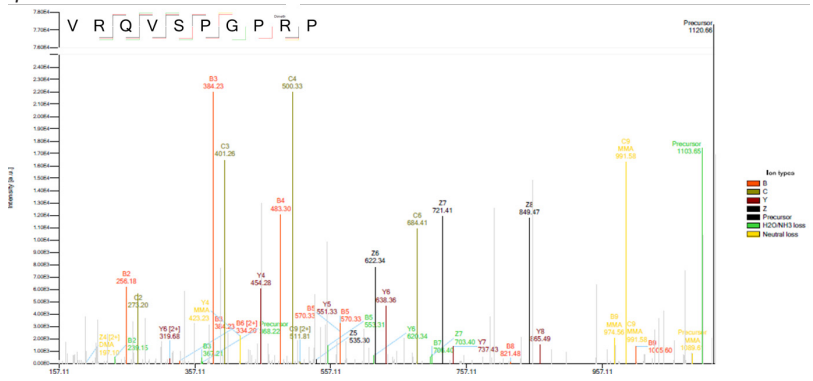
Supplementary Figure.15: *EThcD spectra of the P3 asymmetrycally di-methyl (ADMA) modified peptide GPRGGKSSISSV from the protein of origin Fragile X mental retardation syndrome-related protein 1 (hFXR1p). The neutral losses specific for asymmetrical di-methylation (DMA) are pointed in yellow in the spectra. The peptide was predicted as a HLA-B*07 binder.*



Supplementary Figure.16: *EThcD spectra of the P9 symmetrycally di-methyl (SDMA) modified peptide KRWEMDRGRGR from the protein Suppressor of SWI4 1 homolog (Ssf-1). The neutral losses specific for symmetrical di-methylation (MMA) are pointed in yellow in the spectra. The peptide was predicted as a HLA-B*27 binder.*



Supplementary Figure.17: *EThcD spectra of the P3 asymmetrycally di-methyl (ADMA) modified peptide APRASVPSVQI from the protein Transcriptional repressor p66-alpha (Hp66alpha). The neutral losses specific for asymmetrical di-methylation (DMA) are pointed in yellow in the spectra. The peptide was predicted as a HLA-B*07 binder.*



Supplementary Figure.18: *EThcD spectra of the P9 symmetrycally di-methyl (SDMA) modified peptide VRQVSPGRP from the Peroxisome proliferator-activated receptor gamma coactivator-related protein 1. The neutral losses specific for symmetrical di-methylation (MMA) are pointed in yellow in the spectra.*

Chapter 5

CHAPTER 5

Extended O-GlcNAc on HLA class-I-bound peptides

Fabio Marino[1,2], Marshall Bern[3], Geert P.M. Mommen[1,2,4], Aneika C. Leney[1,2], Jacqueline A.M. van Gaans-van den Brink[5], Alexandre M.J.J. Bonvin[6], Christopher Becker[3], Cécile A.C.M. van Els[5], Albert J. R. Heck[1,2]

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.
[2] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands.
[3] Protein Metrics Inc., San Carlos, CA 94070.
[4] Institute for Translational Vaccinology, Bilthoven, Netherlands.
[5] Centre for Infectious Disease Control, National Institute for Public Health and the Environment, The Netherlands.
[6] Computational Structural Biology, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

Journal of the American Chemical Society (JACS) Communication: August 17, 2015

ABSTRACT

We report unexpected mass spectrometric observations of glycosylated human leukocyte antigen (HLA) class I-bound peptides. Complemented by molecular modeling, in vitro enzymatic assays, and oxonium ion patterns, we propose that the observed O-linked glycans carrying up to five monosaccharides are extended O-GlcNAc's rather than GalNAc-initiated O-glycans. A cytosolic O-GlcNAc modification is normally terminal and does not extend to produce a polysaccharide, but O-GlcNAc on an HLA peptide presents a special case because the loaded HLA class I complex traffics through the endoplasmic reticulum and Golgi apparatus on its way to the cell membrane, and is hence exposed to glycosyltransferases. In addition we report for the first time natural HLA class I presentation of O- and N-linked glycopeptides derived from membrane proteins. HLA class I peptides with centrally located oligosaccharides have been shown to be immunogenic and may therefore be important targets for immune-surveillance.

1. INTRODUCTION

Identification of peptide antigens presented by human leukocyte antigen (HLA) class I molecules is of great importance for the development of vaccines and immunotherapies, (1) including breakthrough T-cell therapies.(2) Along with peptide sequence and protein of origin, it is also important to identify posttranslational modifications (PTMs), because alteration of PTMs is a recognized hallmark of many diseases, including cancer and autoimmune diseases, and altered PTMs can strongly regulate immune system recognition of HLA class I peptides. A small fraction of the HLA class I peptides have been reported to carry phosphorylation (3) or O-GlcNAcylation, (4) indicating that these regulatory modifications pass intact through processing and presentation events in antigen presenting cells. O-GlcNAc is a dynamic PTM that occurs in the cytosol,(5) so O-GlcNAc-ylated proteins may degrade through a route involving the proteasome and classical HLA class I loading. (6) Membrane glycoproteins may also contribute HLA class I peptides via cross-presentation, (7) but HLA class I peptides carrying extracellular O- or N-linked glycosylation have not been previously observed. Here we report a variety of glycosylated HLA class I peptides presented by the HLA-A, -B, and -C heterozygous B-lymphoblastoid cell line GR. The most surprising observations are of O-linked glycans with up to five monosaccharides on serines and threonines in nuclear and cytosolic proteins which we assign as primarily O-GlcNAc, along with extensions by Gal, Gal-NeuAc, and further monosaccharides.

2. EXPERIMENTAL SECTION

HLA peptides purification and LC-MS/MS methods

The HLA-A*01,-03, B*07,-27, and -C*02,-07–positive B-lymphoblastoid cell line GR was grown in RPMI-1640 medium to a total number of 9 × 109cells. HLA class I peptide complexes were immunoprecipitated from lysed GR cells, using the HLA-A–,-B–, and-C–specific mouse monoclonal IgG2a antibody W6/32. HLA class I-associated peptides were eluted with 10% (vol/vol) acetic acid and collected by passage over a 10-kDa molecular weight cutoff membrane. HLA class I-eluted peptides were fractionated by SCX chromatography. The system comprises a Hypercarb trapping column (5 × 0.2 mm i.d., 7 µm particle size; Thermo

Fisher) and SCX column (12 × 0.02 cm i.d. polysulfoethyl aspartamide, 5 µm; Poly LC). A total number of nine SCX fractions were subjected to LC-MS/MS.The HLA elution fractions were analyzed directly by nanoscale LC-MS/MS using a Thermo Scientific EASY-nLC 1000 (Thermo Fisher Scientific) in combination with an ETD-enabled LTQ Orbitrap Elite and an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific). The chromatographical system comprises of a 20 × 0.1 mm i.d. trapping column (Reprosil C18, 3 µm; Dr. Maisch) and a 50 × 0.005 cm i.d. analytical column (Poroshell 120 EC-C18; 2.7 µm). For the Elite analysis the full MS spectra were acquired in the Orbitrap at a resolution of 60,000 (FWHM) while fragment ions were detected in the Orbitrap at a resolution of 15,000 (FWHM). In each case, the 10 most abundant precursor ions were selected for data-dependent EThcD. The maximum ion accumulation time for MS scans was set to 200 ms and for MS/MS scans to 2,500 ms. For the Orbitrap Fusion analysis, the full MS spectra were acquired in an Orbitrap at a resolution of 60,000 (FWHM) with a resolution of 15,000 (FWHM) used to detect the fragment ions. The Top Speed method was enabled for fragmentation where all most abundant precursor ions in 3 seconds were selected for data-dependent EThcD. The maximum ion accumulation time for MS and MS/MS scans was set to 50 ms and 250 ms, respectively.

Data analysis

All EThcD spectra were searched in Byonic 2.6.46 (Protein Metrics, San Carlos, CA) using a 10 ppm mass accuracy for precursors and 0.05 Da (for Orbitrap MS2) for fragment ions. We chose to perform an unspecific search and set HexNAc, HexNAc(1)Hex(1), HexNAc(1)Hex(1)Fuc(1), HexNAc(1)Hex(1)NeuAc(1), HexNAc(2)Hex(2), HexNAc(2)Hex(2)NeuAc(1) of serine and threonine; HexNAc of asparagine; phosphorylation of serine, threonine and tyrosine; methylation and dimehtylation of arginine were set as variable. The search used the full SwissProt database FDR was calculated using target-decoy reverse-database searching and restricted to 1%. Allele predictions and affinities (IC50s) were predicted using NetMHC (8) version 3.4 (www.cbs.dtu.dk/services/NetMHC/), taking as input the unmodified peptides. In this program a strong binder threshold is set at 50 nM. The weak binder threshold score is 500 nM.

Probing O-GlcNAcylation in vitro using the O-GlcNAc transferase enzyme and synthetic substrate peptides

The 11 synthetic peptides (10 µM each) (provided by Nederlands Kanker Instituut), OGT (250 µM) (provided by Nederlands Kanker Instituut) and UDP-GlcNAc (500 µM) (Sigma-aldrich) were incubated overnight at 37 °C, shaking at 300 rpm. O-GlcNAcylation was quenched by rapid dilution into 10% FA and the samples stored at -20 °C prior to LC-MS/MS analysis on the Orbitrap Fusion.

Computational modeling of the structures of the HLA class I / peptide complexes

The modeling of the protein-peptide complexes was performed using the HADDOCK web server,(9) starting from an ensemble of two conformations for each peptide: extended and poly-proline II conformations, following the previously published protein-peptide docking protocol. (10) HADDOCK (High Ambiguity Driven DOCKing) (11, 12) is a semi-flexible docking protocol that can make use of bioinformatics predictions and biochemical/biophysical interaction data to drive the docking process. It uses CNS (Crystallographic and NMR system (13)) as its structure calculation engine. The protocol consists of three steps: i) randomization of orientation and rigid body docking by energy minimization driven by interaction restraints (it0), ii) semi-flexible refinement in torsion angle space in which

side-chains and backbone atoms of the interface residues are allowed to move (it1) and iii) Cartesian dynamics refinement in explicit solvent, typically water. The final structures are clustered using the pairwise backbone ligand interface RMSD and the resulting clusters ranked according to the HADDOCK score (weighted sum of the restraint energy, van der Waals and electrostatic energies based on OPLS parameters) (14) and a desolvation energy term. In order to drive the docking, the known anchor residues were targeted to their respective binding pockets by defining ambiguous distance restraints based on an analysis of existing MRC-peptide complexes. The HLA structures were taken from PDB entries 3VCL for HLA-B*07 and 3RL2 for HLA-A*0301. The peptides were built with PyMOL (15) and treated as fully flexible throughout the docking protocol. The number of steps for flexible refinement was increased from the default values to 2000/2000/2000/1000 for the four consecutive stages of the simulated annealing. Considering the rather deep binding pocket on the MHCs, the intermolecular interactions were scaled down to 0.01 in the initial rigid-body docking stage to allow for better peptide insertion. The final models were clustered based on interface ligand RMSDs using a 2.5 Å cutoff. The final clusters were subject to a final refinement round using the refinement interface of the web server to remove any strain introduced by the restraints. The average amino acid solvent accessibilities of the models from all top ranking clusters whose score were not significantly different from each other were calculated using NACCESS . (16) The model of the glycosylated peptide complex was generated following the same protocol, but starting from a modified threonine to which GlcNAc was attached. The parameters and topologies for the oligosaccharide were obtained using the GLYCANS web server (Krzeminki and Bonvin unpublished, http://haddock.science.uu.nl/enmr/services/GLYCANS).

3. RESULTS AND DISCUSSION

Identification of HLA-class I peptides harboring Glycans

In this study we focused on HLA-class I peptides carrying posttranslational modified (PTM) antigens presented at the cell surface as HLA peptides and be particularly enriched on specific alleles and/or harboring a defined position of the peptide sequence due to structural constraints, alike to other PTM antigens. (17) The purified peptides from HLA-A, -B, and -C heterozygous B-lymphoblastoid cell line GR were analyzed on an Orbitrap Fusion instrument (Thermo Fisher Scientific) with high-mass accuracy and high resolution. We used a rather novel fragmentation mode denoted EThcD, (18) which employs both electron-transfer dissociation (ETD) and higher-energy collisional dissociation (HCD). (19, 20) In total we observed 32 HLA class I-peptides carrying glycan moieties (Supplementary Table.1). The most surprising observations are 18 O-linked glycans with up to five monosaccharides on serines and threonines in nuclear and cytosolic proteins (Table 1). HexNAc(1)Hex(1)NeuAc(1), 730 = HexNAc(2)Hex(2), 1021 = HexNAc(2) Hex(2)NeuAc(1). Site of modification is indicated by g = glycosylation, p = phosphorylation, m = methylation and d = dimethylation. Apart from the 18 HLA class I-peptides carrying extended HexNAc moieties, we also identified 14 other glycopeptides having a non-extended HexNac (Supplementary Table.1).

EThcD enables unambiguous localization and validation of the glycan structures

The use of EThcD fragmentation provided several advantages, in fact the spectra not only in-

clude c- and z-ions with labile modifications intact for confident modification site localization, (20) but also oxonium ions from glycan fragments which are necessary for the validationof the glycan structure. Figure.1 shows an EThcD spectra of two forms of an HLA class I glycopeptide, as indicated, the c3 to c9 ions enabled unambiguous localization of the glycan, while the oxonium ions at 204 and 274 Da, the O-GlcNAc diagnostic oxonium ion fragmentation pattern, (21) and neutral losses of the HexNAc acetyl group and the labile sialic acid NeuAc corroborate the exact monosaccharide composition. The protein of origin for the peptide in Figure.1 is the Zinc finger protein 281. Uniprot lists this protein as localized in the nucleus. Figure.2 shows an EThcD spectrum of a HLA class I glycopeptide from the RNA-binding protein 27 carrying HexNAc(1)Hex(1). This protein has previously been reported to be O-GlcNAcylated.(22, 23) We observe the 11-residue, IPRPPITQSSL peptide, with and without arginine (di)methylation, with 5 different glycans located at the same threonine: HexNAc, HexNAc(1)Hex(1), HexNAc(1)Hex(1)NeuAc(1), HexNAc(2)Hex(2), and HexNAc(2)Hex(2)NeuAc(1)

Peptide	Glycan Masses	Protein Name (Gene)	HLA Allele	Affinity (IC50)	YinOYang Score
LPKPANgTSAL	365	DNA-binding protein RFX7 (RFX7)	B*07:02	17 nM	0.73
RPPVgTKASSF	203, 656	R3H domain-containing protein 2 (R3HD2)	B*07:02	27 nM	0.82
RVKpTPTgSQSY	0, 203, 365, 730, 1021	Zinc finger protein 281 (ZN281)	A*03:01	284 nM	0.7
RVKpTPTgSQSYR	0, 203, 365, 656, 730, 1021	Zinc finger protein 281 (ZN281)	A*03:01	445 nM	0.7
APVgSPSSQKL	656	Telomeric repeat-binding factor 2-interacting protein 1 (TE2IP)	B*07:02	70 nM	0.58
IPdRPPIgTQSSL	203, 365, 656, 730, 1021	RNA-binding protein 27 (RBM27)	B*07:02	9 nM	0.68

Table.1: *Detected HLA class 1 glycopeptides containing extended O-GlcNAc moieties. Glycan extensions are given by their nominal masses: 0 = unmodified, 203 = HexNAc, 365 = HexNAc(1)Hex(1), 511 = HexNAc(1)Hex(1)Fuc(1), 656 = HexNAc(1)Hex(1)NeuAc(1), 730 = HexNAc(2)Hex(2), 1021 = HexNAc(2)Hex(2)NeuAc(1). Site of modification is indicated by g = glycosylation, p = phosphorylation, m = methylation and d = dimethylation.*

O-GlcNAc and O-GalNAc initiated HLA-class I peptides

For 28 of the 32 identified glycosylated peptides (Supplementary Table.1), we initially interpreted the O-linked 'core' saccharide as O-GlcNAc. We based our first hypothesis on protein subcellular localization and previously identified O-GlcNAc proteins and modification sites in the dbOGap. (22) In fact, the 14 O-GlcNAc-ylated source proteins reside in the nucleus and/or

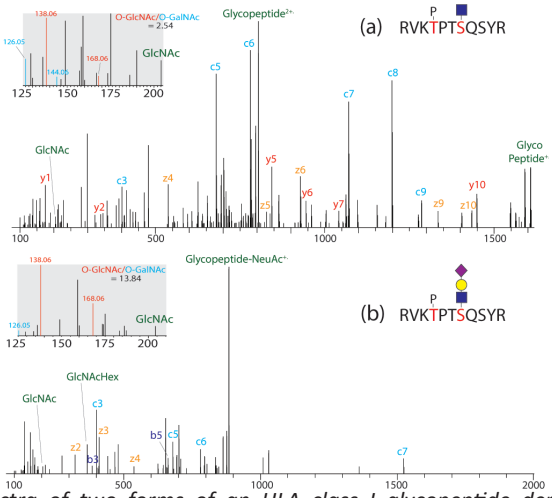


Figure.1: *EThcD spectra of two forms of an HLA class I glycopeptide derived from ZN281, a known nuclear protein, carrying (a) GlcNAc (203 Da) and (b) GlcNAc(1)Hex(1)NeuAc(1) (656 Da) O-linked to Thr345. To decipher the saccharide identity of the O-linked saccharide moiety, the ratio of the abundance of the (m/z 138 + m/z 168) to (m/z 126 + m/z 144) oxonium ions are taken. Ratio values >1 are indicative of an O-GlcNAc core, whereas values <1 indicate an O-GalNAc core. (21) The insert shows the prevalence for the diagnostic O-GlcNAc fragment ions.*

cytosol, and are involved in DNA or RNA binding and other functions (Supplementary Table.1). Interestingly, 7 of the 14 source proteins have been previously reported as carrying O-GlcNAc, 4 at the identical sites observed here. Moreover, of the 14 O-GlcNAc-ylated peptides observed, all rank highly by the YingOYang computational O-GlcNAc predictor (Table.1). (24) If most of the modified peptides seemed to have an O-GlcNAc 'core' we observed that the glycopeptide PSSGLGV(+HexNAc)TKQDLGPVPM was derived from the golgi/membrane-associated HLA class II histocompatibility antigen invariant chain, and is therefore more likely to carry O-GalNAc than O-GlcNAc. Processing and presentation of extracellular O-glycosylation have been extensively studied for the tumor-associated mucin MUC1 in in vitro models and clinical studies, but these studies (25, 26) have not identified naturally occurring mucin-type HLA class I peptides. Membrane glycoproteins can recycle in endolysosomal compartments from which they may gain access to the classical HLA class I pathway through escape into the cytosol; this is a known cross-presentation route. (7) Alternatively, processing of membrane proteins without glycan removal

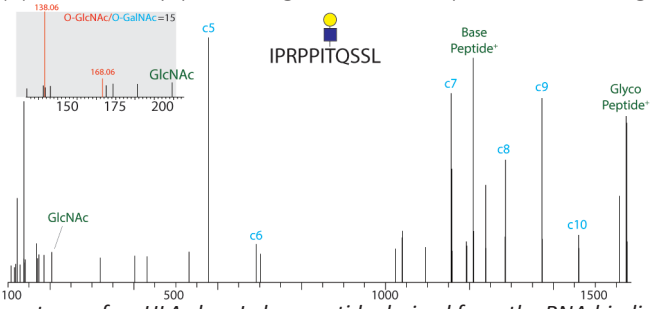


Figure.2: *EThcD spectrum of an HLA class I glycopeptide derived from the RNA-binding protein 27, a known O-GlcNAc-ylated protein. Annotation of c5-c10 ions enabled the exact localization of GlcNAc(1) Hex(1) (365 Da) on Thr386. The insert highlights the fragment ions diagnostic of an O-GlcNAc core.*

can likely occur in the endolysosomal compartment, already evidenced in dendritic cells by the detection of MUC1 glycosylated peptides (27) presented by I-Ab molecules, the murine counterpart of HLA class II. A fraction of cell surface HLA-class I molecules recycles through the endolysosomal compartment where reloading with peptidic cargo may occur. (28)

.

N-glycan initiated HLA-class I peptide

The glycopeptide KPAPPFgNVTV, derived from the membrane-associated Interleukin-21 receptor, and differently from all the other glycans identified, it was found to be modified at Asn125 (Supplementary Table.1). The glycan is HexNAc(1)Hex(1)Fuc(1), which we interpret to be fucosylated core, a common N-glycan truncation, extended in the Golgi by β1,4 galactosyltransferase. N-glycosylated proteins can enter the classical HLA class I loading route by retrograde transport from the ER into the cytosol, a pathway for degradation of misfolded glycoproteins. This process however is linked to the activity of cytosolic peptide-N-glycanase, cleaving between Asn and the core and deamidating the asparagine. Deamidated asparagines are regularly found on HLA class I peptides, (29) including in our dataset. (18) It is possible that the glycanase failed to remove the truncated N-glycan, or alternatively, the peptide with a truncated N-glycan trafficked to the ER from the endolysosome. N-glycans can survive glycanase activity in the endolysosomal environment, as N-glycopeptides presented by HLA class II molecules have been observed. (30)

Elongation of O-GlcNAc on HLA class I glycopeptides

In our hypothesis most of the observed glycopeptides have an O-GlcNAc "core" which was further extended by Gal, or optionally by N-acetyl lactosamine, and optionally capped by NeuAc (N-acetyl neuraminic acid). Figure.3 illustrates our interpretation regarding the formation of these modifications on HLA class I-bound peptides using the classical class I antigen presentation pathway.(6) O-GlcNAc starts on a cytosolic protein, survives degradation by the proteasome and translocation into the endoplasmic reticulum (ER), and is then loaded onto an HLA class I molecule and extended by glycosyltransferases, starting with β1, 4 galactosyltransferase, in the Golgi. Galactosyltransferase extension is a standard method to de-

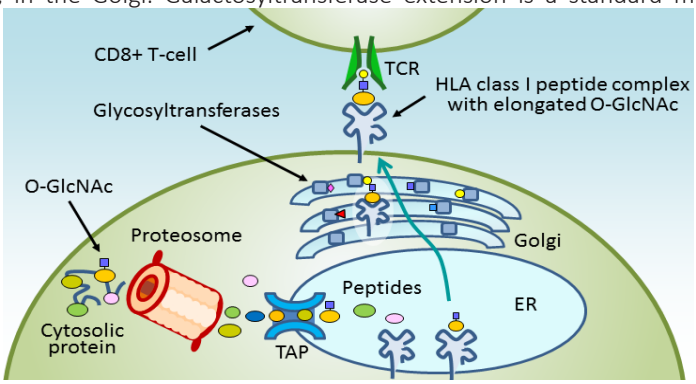


Figure.3: *Elongation of O-GlcNAc on HLA class I glycopeptides. Cytosolic proteins carrying O-GlcNAc are degraded in the immune proteasome, leaving the O-GlcNAc intact. Glycopeptides are translocated via the transporter associated with antigen presentation (TAP) into the endoplasmic reticulum (ER) and loaded onto HLA class I molecules. The loaded glycopeptide travels through the Golgi, where the O-GlcNAc is elongated by glycosyltransferases, on the way to the plasma membrane for presentation to the T cell receptor (TCR) on CD8+ T cells.*

tect O-GlcNAc, (31) and single N-linked GlcNAc's in the ER of engineered cells almost uniformly elongate to GlcNAc-Gal-NeuAc in the Golgi. (5) To our knowledge, the findings reported here are the first observations of O-linked GlcNAc extended *in vivo.*

Distinguishing O-GlcNAc modifications from O-GalNAc modifications by MS/MS fragmentation patterns

In order to prove that the identified O-glycosylated peptides had either an O-GlcNAc or O-GalNAc 'core', we decided to follow a recent approach which documented that O-GlcNAc and O-GalNAc glycopeptides can be distinguished based on their energy-resolved gas phase fragmentation profiles. (21) To decipher the saccharide identity of the O-linked saccharide moiety, the ratio of the abundance of the (m/z 138 + m/z 168) to (m/z 126 + m/z 144) fragment ions is taken upon EThcD spectra, termed the GlcNAc/GalNAc ratio. According to the criteria set out by Halim et al. , (21) GlcNAc/GalNAc ratio values >1 are indicative of an O-GlcNAc core, whereas GlcNAc/GalNAc values <1 indicate an O-GalNAc core. Figure.4 shows examples of three O-GlcNAc core and three O-GalNAc core peptides that were assigned based on their gas phase EThcD fragmentation.

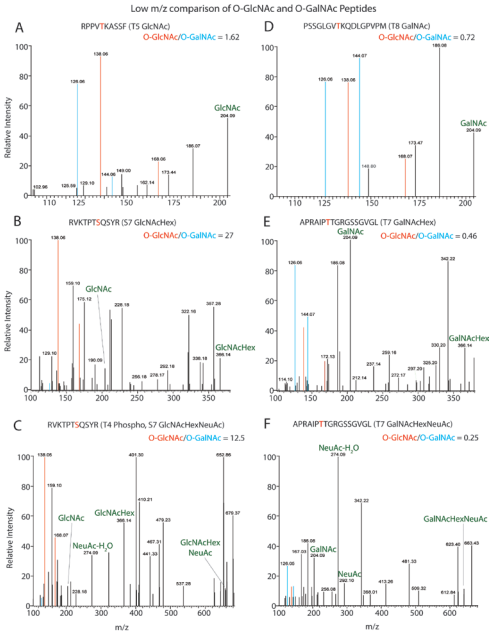


Figure.4: *O-GlcNAc/O-GalNAc determination. The released saccharide moieties upon glycopeptide EThcD fragmentation are labelled in green: GlcNAc/GalNAc (204.09 m/z), GlcNAc/GalNAc-Hex (366.14 m/z), NeuAc (292.10 m/z) and GlcNAc/GalNAc-HexNeuAc (657.24 m/z). The relative intensities of the HexNAc-2H2O (168.06 m/z) and HexNAc-CH6O3¬ (138.06 m/z) ions (red) and the HexNAc-C2H6O3 (126.06 m/z) and HexNAc-C2H4O2 (144.06 m/z) ions (blue) were used to calculate the O-GlcNAc/O-GalNAc ratio, following the procedure outlined in ref. (21)*

Probing O-GlcNAc-ylation in vitro

To determine whether the observed O-GlcNAc core glycopeptides are putative substrates for the enzyme O-GlcNAc transferase (OGT), 11 peptides (Supplementary Table.3) were synthesized and allowed to undergo O-GlcNAcylation in vitro. In the case of the observed HLA class I O-GlcNAcylated peptides, all the O-GlcNAcylated serine/threonine res-

idue matched that of the identified HLA class I glycopeptides, with the GlcNAc/GalNAc ratios upon EThcD fragmentation being, in all cases, consistent with O-GlcNAc modification (Supplementary Table.3). Figure.5 shows three examples of EThcD spectra of the in vitro O-GlcNAcylated peptides. The c ions clearly highlight the location of the O-GlcNAc modification, and the ions in the low m/z region are diagnostic of O-GlcNAc. Moreover, importantly the low m/z region of the MS/MS spectrum of the in vitro O-GlcNAcylated peptide RPPVTKASSF shown (Figure.5d) mirrors precisely that of the in vivo O-GlcNAcylated HLA class I peptide observed (Figure.5a), further corroborating the O-GlcNAc nature of the peptides detected in vivo. In addition, the 2 synthetic peptides that we hypothesized to be O-GalNAcylated in vivo were not relevantly O-GlcNAcylated in vitro at the O-GalNAcylated threonine residue identified in the HLA class I glycopeptides (Figure.6).

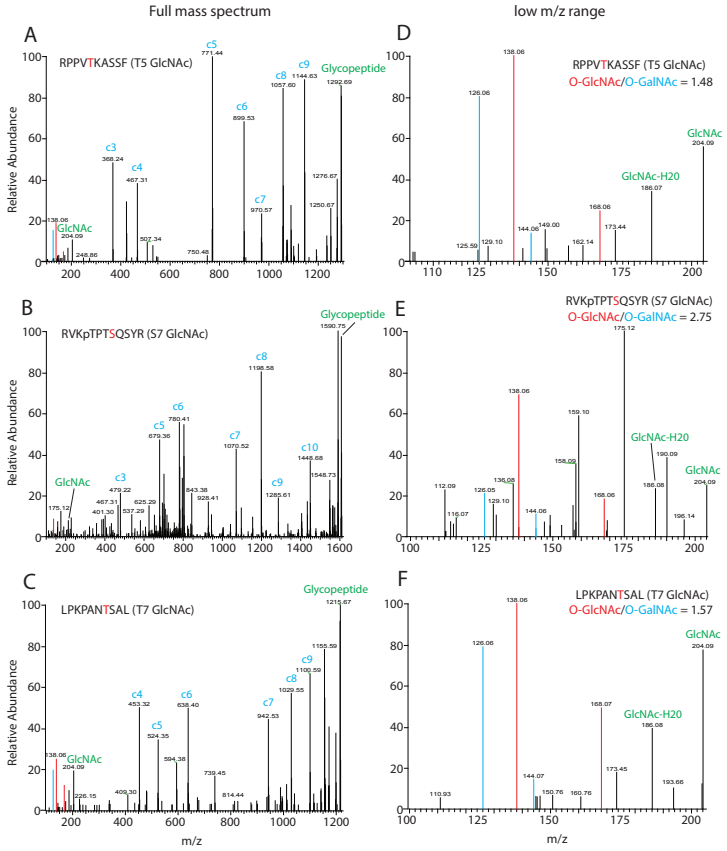


Figure.5: *EThcD spectra of the in vitro O-GlcNAc-ylated synthetic peptides RPPVTKASSF (a, d), RVKpTPTSQSYR (b, e) and LPKPANTSAL (c, f). The c ions show the O-GlcNAc location in each case (a, b, c). The relative intensities of the HexNAc-2H2O (168.06 m/z) and HexNAc-CH6O3¬ (138.06 m/z) ions (red) and the HexNAc-C2H6O3 (126.06 m/z) and HexNAc-C2H4O2 (144.06 m/z) ions (blue) were used to calculate the O-GlcNAc/O-GalNAc ratio.*

O-GlcNAc group is not directly involved in binding to the HLA class I groove

To further prove our O-GlcNAc elongation model, starting from a crystal structure (32) of a HLA-B*07-peptide complex (3VCL), we modeled the HLA-B*07 molecule

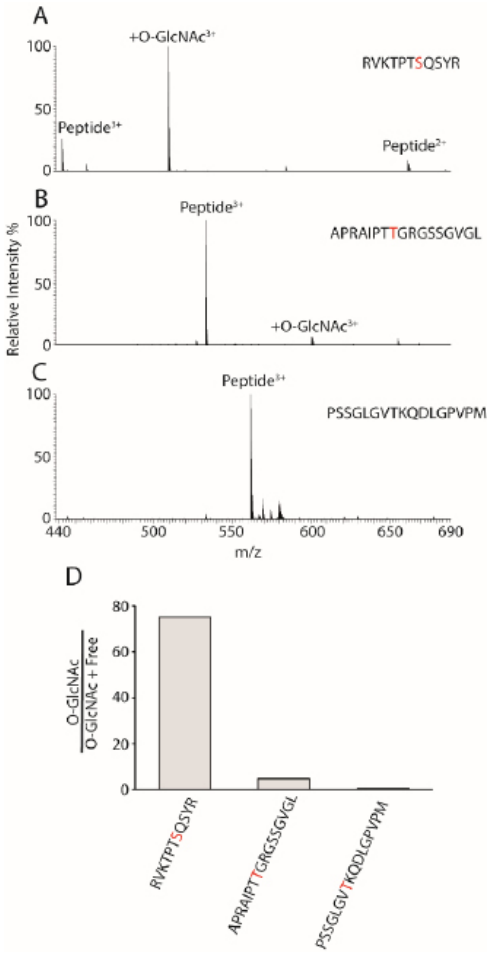


Figure.6: *Mass spectra of the synthetic peptides RPPVTKASSF (A), APRAIPTTGRGSSGVGL (B) and PSSGLGVTKQDLGPVPM (C) after 16 h of incubation with O-GlcNAc transferase in vitro. The O-GlcNAcylated residue as determined by EThcD are highlighted in red. The relative intensities of the O-GlcNAcylated peptides are expressed as a percentage of O-GlcNAcylated+free peptides (D). Note that although <5 % of the APRAIPTTGRGSSGVGL peptide was O-GlcNAcylated, this site (Thr8) is different to the O-GalNAcylated site (Thr7) observed on the MHC class I peptides detected in vivo. A small percentage (<0.5 %) of the PSSGLGVTKQDLGPVPM peptide was O-GlcNAcylated in vitro, however, this low stoichiometry is unlikely to have in vivo relevance.*

complexed to GlcNAc-ylated RPPVTKASSF using the information-driven docking program HADDOCK. (11) Importantly, we observed that the modified threonine was solvent exposed, and not directly involved in binding to the HLA class I molecule groove (Figure 7 and Supplementary Tables.4), which suggests that an O-GlcNAc at this site would be accessible to glycosyltransferases in the Golgi apparatus.Similar modeling observations were made for other glycosylated peptides (see Supplementary Tables.4).

Positional and allele preference of O-GlcNAc-ylated HLA-class I peptides

The mono- or oligo O-GlcNAc moieties were found exclusively at the peptides' center residues (P3-P7), with a preference for P5 (7 out of 15 glycosylation sites). As opposed to the

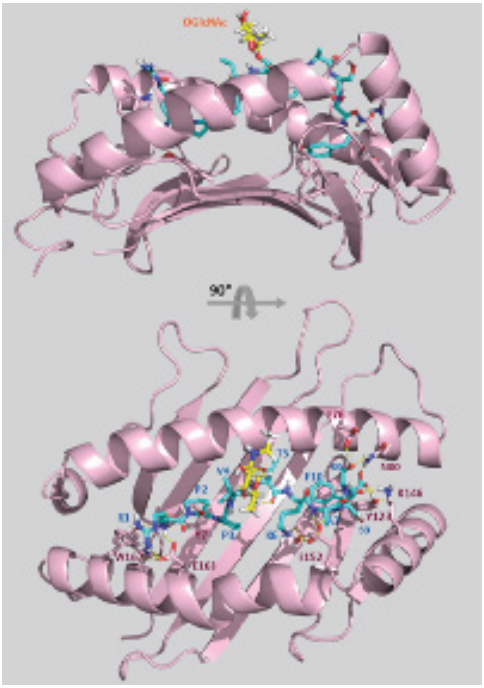


Figure.7: *HADDOCK model of HLA-B*07 complexed with RPPVT(+GlcNAc)KASSF. GlcNAc (yellow) is surface exposed and likely to be accessible to glycosyltransferases in the Golgi.*

previously reported set of unmodified peptides, (18) a significant preference (p-value < 10-5) was found for the full set of glycopeptides to be bound to the HLA-B*07 allele (15 out of 16), rather than to any of the other alleles present in the GR lymphoblastoid cell line. Upon further inspection of the O-GlcNAc-ylated peptide sequences, there is resemblance to motifs recognized by various kinases and preference for nearby proline residues (33) especially at the -2 and -3 positions, so the common proline anchor at position P2 in HLA-B*07-presented peptides may lead to a preference for peptides harboring centrally the O-GlcNAc group. Indeed, proline at the -3 position is consistent with the consensus sequence for O-GlcNAc (34) giving additional evidence for an O-GlcNAc core saccharide over the alternative O-GalNAc core whereby proline is typically found at the -1 position. (35)

4. CONCLUSIONS

The glycans described here are likely to affect immune recognition. In mice it has been already reported that adding centrally located oligosaccharides to the murine H-2Db class I peptides can increase their immunogenicity. (36) Other structurally unusual HLA class I peptides, such as long viral peptides, centrally bulging from the HLA class I binding groove due to their non-canonical lengths, are able to engage TCRs and induce immunodominant responses. (37–39) Thus the extended-O-GlcNAc HLA class I peptides found in this study may constitute a functional category of T cell epitopes and play a role in sensing metabolic processes regulated by O-GlcNAc-ylation, including processes that may be dysregulated in diseases such as cancer and diabetes. (40)

5. ACKNOWLEDGMENT

We kindly acknowledge dr. H. Meiring (IntraVacc, Bilthoven) for SCX fractionation of samples, and the laboratory of dr. H. Ovaa (Nederlands Kanker Instituut) for providing the synthetic analogues and the O-GlcNAc transferase. This work was partly supported by the project Proteins At Work (project 184.032.201), a program of the Netherlands Proteomics Centre financed by the Netherlands Organization for Scientific Research (NWO) as part of the National Roadmap Large-scale Research Facilities of the Netherlands, and by the projects Immunoproteomics and Correlates of Protection, financed by the Dutch Government. Protein Metrics gratefully acknowledges support from NIH grants GM100634 and GM103362. This work was also supported by the Institute for Chemical Immunology, an NWO Gravitation project funded by the Ministry of Education, Culture and Science of the Netherlands

6. REFERENCES

1. Parmiani, G., Castelli, C., Dalerba, P., Mortarini, R., Rivoltini, L., Marincola, F. M., and Anichini, A. (2002) Cancer immunotherapy with peptide-based vaccines: what have we achieved? Where are we going? J. Natl. Cancer Inst. 94, 805–18
2. Couzin-Frankel, J. (2013) Breakthrough of the year 2013. Cancer immunotherapy. Science 342, 1432–3
3. Cobbold, M., Peña, H. D. La, Norris, A., Polefrone, J., Qian, J., Michelle, a, Cummings, K., Penny, S., Turner, J. E., Cottine, J., and Jennifer, G. (2014) Memory-Like Immunity in Leukemia. 5,
4. Haurum, J. S., Høier, I. B., Arsequell, G., Neisig, a, Valencia, G., Zeuthen, J., Neefjes, J., and Elliott, T. (1999) Presentation of cytosolic glycosylated peptides by human class I major histocompatibility complex molecules in vivo. J. Exp. Med. 190, 145–150
5. The O-GlcNAc Modification- PubMed- NCBI
6. Yewdell, J. W., Reits, E., and Neefjes, J. (2003) Making sense of mass destruction: quantitating MHC class I antigen presentation. Nat. Rev. Immunol. 3, 952–61
7. Neefjes, J., and Sadaka, C. (2012) Into the intracellular logistics of cross-presentation. Front. Immunol. 3, 31
8. Karosiene, E., Rasmussen, M., Blicher, T., Lund, O., Buus, S., and Nielsen, M. (2013) NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. Immunogenetics 65, 711–724
9. de Vries, S. J., van Dijk, M., and Bonvin, A. M. J. J. (2010) The HADDOCK web server for data-driven biomolecular docking. Nat. Protoc. 5, 883–97
10. Trellet, M., Melquiond, A. S. J., and Bonvin, A. M. J. J. (2013) A unified conformational selection and induced fit approach to protein-peptide docking. PLoS One 8, e58769
11. Dominguez, C., Boelens, R., and Bonvin, A. M. J. J. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J. Am. Chem. Soc. 125, 1731–7
12. de Vries, S. J., van Dijk, A. D. J., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, T., and Bonvin, A. M. J. J. (2007) HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. Proteins 69, 726–33
13. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. D. Biol. Crystallogr. 54, 905–21
14. Jorgensen, W. L., and Tirado-Rives, J. (1988) The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. J. Am. Chem. Soc. 110, 1657–1666
15. {Schrödinger, L. (2010) The {PyMOL} Molecular Graphics System, Version~1.3r1.
16. Hubbard SJ, Thornton JM (1993) "NACCESS", Computer Program, Department of Biochemistry and Molecular Biology. University College London.- Open Access Library
17. Petersen, J., Wurzbacher, S. J., Williamson, N. a, Ramarathinam, S. H., Reid, H. H., Nair, A. K. N., Zhao, A. Y., Nastovska, R., Rudge, G., Rossjohn, J., and Purcell, A. W. (2009) Phosphorylated self-peptides alter human leukocyte antigen class I-restricted antigen presentation and generate tumor-specific epitopes. Proc. Natl. Acad. Sci. U. S. A. 106, 2776–81

18. Mommen, G. P. M., Frese, C. K., Meiring, H. D., van Gaans-van den Brink, J., de Jong, A. P. J. M., van Els, C. a C. M., and Heck, A. J. R. (2014) Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proc. Natl. Acad. Sci. U. S. A. 111, 4507–12
19. Frese, C. K., Altelaar, A. F. M., Hennrich, M. L., Nolting, D., Zeller, M., Griep-Raming, J., Heck, A. J. R., and Mohammed, S. (2011) Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. J. Proteome Res. 10, 2377–88
20. Frese, C. K., Altelaar, A. F. M., van den Toorn, H., Nolting, D., Griep-Raming, J., Heck, A. J. R., and Mohammed, S. (2012) Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. Anal. Chem. 84, 9668–9673
21. Halim, A., Westerlind, U., Pett, C., Schorlemer, M., Rüetschi, U., Brinkmalm, G., Sihlbom, C., Lengqvist, J., Larson, G., and Nilsson, J. (2014) Assignment of saccharide identities through analysis of oxonium ion fragmentation profiles in LC-MS/MS of glycopeptides. J. Proteome Res. 13, 6024–32
22. Wang, J., Torii, M., Liu, H., Hart, G. W., and Hu, Z.-Z. (2011) dbOGAP- an integrated bioinformatics resource for protein O-GlcNAcylation. BMC Bioinformatics 12, 91
23. Wang, Z., Udeshi, N. D., Slawson, C., Compton, P. D., Sakabe, K., Cheung, W. D., Shabanowitz, J., Hunt, D. F., and Hart, G. W. (2010) Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis. Sci. Signal. 3, ra2
24. Kastrup, I. B., Stevanovic, S., Arsequell, G., Valencia, G., Zeuthen, J., Rammensee, H. G., Elliott, T., and Haurum, J. S. (2000) Lectin purified human class I MHC-derived peptides: evidence for presentation of glycopeptides in vivo. Tissue Antigens 56, 129–35
25. Wolfert, M. A., and Boons, G.-J. (2013) Adaptive immune activation: glycosylation does matter. Nat. Chem. Biol. 9, 776–84
26. Roulois, D., Grégoire, M., and Fonteneau, J.-F. (2013) MUC1-specific cytotoxic T lymphocytes in cancer therapy: induction and challenge. Biomed Res. Int. 2013, 871936
27. Vlad, A. M., Muller, S., Cudic, M., Paulsen, H., Otvos, L., Hanisch, F.-G., and Finn, O. J. (2002) Complex carbohydrates are not removed during processing of glycoproteins by dendritic cells: processing of tumor antigen MUC1 glycopeptides for presentation to major histocompatibility complex class II-restricted T cells. J. Exp. Med. 196, 1435–46
28. Grommé, M., Uytdehaag, F. G., Janssen, H., Calafat, J., van Binnendijk, R. S., Kenter, M. J., Tulp, A., Verwoerd, D., and Neefjes, J. (1999) Recycling MHC class I molecules and endosomal peptide loading. Proc. Natl. Acad. Sci. U. S. A. 96, 10326–31
29. Skipper, J. C., Hendrickson, R. C., Gulden, P. H., Brichard, V., Van Pel, A., Chen, Y., Shabanowitz, J., Wolfel, T., Slingluff, C. L., Boon, T., Hunt, D. F., and Engelhard, V. H. (1996) 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. 183, 527–34
30. Dengjel, J., Schoor, O., Fischer, R., Reich, M., Kraus, M., Müller, M., Kreymborg, K., Altenberend, F., Brandenburg, J., Kalbacher, H., Brock, R., Driessen, C., Rammensee, H.-G., and Stevanovic, S. (2005) Autophagy promotes MHC class II presentation of peptides from intracellular source proteins. Proc. Natl. Acad. Sci. U. S. A. 102, 7922–7927
31. Meuris, L., Santens, F., Elson, G., Festjens, N., Boone, M., Dos Santos, A., Devos, S., Rousseau, F., Plets, E., Houthuys, E., Malinge, P., Magistrelli, G., Cons, L., Chatel, L., Devreese, B., and Callewaert, N. (2014) GlycoDelete engineering of mammalian cells simplifies N-glycosylation of recombinant proteins. Nat. Biotechnol. 32, 485–9
32. Brennan, R. M., Petersen, J., Neller, M. A., Miles, J. J., Burrows, J. M., Smith, C., McCluskey, J., Khanna, R., Rossjohn, J., and Burrows, S. R. (2012) The impact of a large and frequent deletion in the human TCR β locus on antiviral immunity. J. Immunol. 188, 2742–8
33. Kötzler, M. P., and Withers, S. G. (2016) Proteolytic Cleavage Driven by Glycosylation. J. Biol. Chem. 291, 429–34
34. Gupta, R., and Brunak, S. (2002) Prediction of glycosylation across the human proteome and the correlation to protein function. Pac. Symp. Biocomput., 310–22
35. Gerken, T. A., Ten Hagen, K. G., and Jamison, O. (2008) Conservation of peptide acceptor preferences between Drosophila and mammalian polypeptide-GalNAc transferase ortholog pairs. Glycobiology 18, 861–70
36. Abdel-Motal, U. M., Berg, L., Rosén, A., Bengtsson, M., Thorpe, C. J., Kihlberg, J., Dahmén, J., Magnusson, G., Karlsson, K. A., and Jondal, M. (1996) Immunization with glycosylated Kb-binding peptides generates carbohydrate-specific, unrestricted cytotoxic T cells. Eur. J. Immunol. 26, 544–51
37. Tynan, F. E., Borg, N. A., Miles, J. J., Beddoe, T., El-Hassen, D., Silins, S. L., van Zuylen, W. J. M., Purcell, A. W., Kjer-Nielsen, L., McCluskey, J., Burrows, S. R., and Rossjohn, J. (2005) High resolution structures of highly bulged viral epitopes bound to major histocompatibility complex class I. Implications for T-cell receptor engagement and T-cell immunodominance. J. Biol. Chem. 280, 23900–9
38. Tynan, F. E., Burrows, S. R., Buckle, A. M., Clements, C. S., Borg, N. A., Miles, J. J., Beddoe, T., Whisstock, J. C.,

Wilce, M. C., Silins, S. L., Burrows, J. M., Kjer-Nielsen, L., Kostenko, L., Purcell, A. W., McCluskey, J., and Rossjohn, J. (2005) T cell receptor recognition of a "super-bulged" major histocompatibility complex class I-bound peptide. Nat. Immunol. 6, 1114–22
39. Ebert, L. M., Liu, Y. C., Clements, C. S., Robson, N. C., Jackson, H. M., Markby, J. L., Dimopoulos, N., Tan, B. S., Luescher, I. F., Davis, I. D., Rossjohn, J., Cebon, J., Purcell, A. W., and Chen, W. (2009) A long, naturally presented immunodominant epitope from NY-ESO-1 tumor antigen: implications for cancer vaccine design. Cancer Res. 69, 1046–54
40. Slawson, C., Copeland, R. J., and Hart, G. W. (2010) O-GlcNAc signaling: a metabolic link between diabetes and cancer? Trends Biochem. Sci. 35, 547–55

7. SUPPLEMENTARY

Peptide (length)	Glycans	Core Glycan	Protein	HLA Allele, Predicted affinity (IC50)	Comment
LPKPANgTSAL (10)	365	O-GlcNAc	RFX7_HUMAN DNA-binding protein RFX7	B*07:02, 17 nM	Nuclear
RPPVgTKASSF (10)	203, 656	O-GlcNAc	R3HD2_HUMAN R3H domain-containing protein 2	B*07:02, 27 nM	Nuclear
RVKpTPTgSQSY (10)	0, 203, 365, 730, 1021	O-GlcNAc	ZN281_HUMAN Zinc finger protein 281	A*03:01, 284 nM	Known O-GlcNAc site
RVKpTPTgSQSYR (11)	0, 203, 365, 656, 730, 1021	O-GlcNAc	ZN281_HUMAN Zinc finger protein 281	A*03:01, 445 nM	Known O-GlcNAc site
APVgSPSSQKL (10)	656	O-GlcNAc	TE2IP_HUMAN Telomeric repeat-binding factor 2-interacting protein 1	B*07:02, 70 nM	Known O-GlcNAc protein
IPdRPPIgTQSSL (10)	203, 365, 656, 730, 1021	O-GlcNAc	RBM27_HUMAN RNA-binding protein 27	B*07:02, 9 nM	Known O-GlcNAc site
APVgSSKSSL (9)	203	O-GlcNAc	HIPK1_HUMAN Homeodomain-interacting protein kinase 1	B*07:02, 10 nM	Known O-GlcNAc protein
APFgSCRTEL (9)	203, 365	O-GlcNAc	ICAM3_HUMAN Intercellular adhesion molecule 3	B*07:02, 11 nM	Extracellular membrane protein
RPgTPRGITL (9)	0, 203	O-GlcNAc	SNX20_HUMAN Sorting nexin-20	B*07:02, 8 nM	Nuclear, Cytosolic, and Membrane
VPEVgTKPSL (9)	203	O-GlcNAc	TNR6B_HUMAN Trinucleotide repeat-containing gene 6B protein	B*07:02, 57 nM	Cytoplasm
IPAVgTRSTI (9)	0, 203	O-GlcNAc	LAP2_HUMAN Protein LAP2	B*07:02, 12 nM	Nuclear, Cytosolic, and Membrane
IVQAgTRTSL (9)	203	O-GlcNAc	MCAF1_HUMAN Activating transcription factor 7 interacting protein	B*07:02, 35 nM	Nuclear
IPTgSARSML (9)	203	O-GlcNAc	P66B_HUMAN Transcriptional repressor P66-beta	B*07:02, 11 nM	Known O-GlcNAc protein
IPVSgSHNSL (9)	203	O-GlcNAc	MEF2C_HUMAN Myocyte-specific enhancer factor 2C	B*07:02, 12 nM	Nuclear and cytoplasmic
TPASgSRAQTL (10)	203	O-GlcNAc	SPTB2_HUMAN Spectrin beta chain	B*07:02, 21 nM	Known
KPAPPFgNVTV (10)	511	N-HexNAc	IL21R_HUMAN Interleukin-21 receptor	B*07:02, 45 nM	Known N-glycosylation site
APRAIPgTTGRGSSGVGL (17)	0, 365, 656	O-GalNAc	MCM3_HUMAN DNA replication factor MCM3	N/A	Nuclear
PSSGLGVgTKQDLGPVPM (17)	0, 203	O-GalNAc	CD74_HUMAN HLA class II histocompatibility antigen invariant chain	N/A	Golgi, Membrane

Supplementary Table.1: *Overview of HLA Class I glycopeptides detected in the GR lymphoblastoid cell line. The first six entries are also shown in Table 1 of the main article. Glycan extensions are given by their nominal masses: 0 = unmodified, 203 = HexNAc, 365 = HexNAc(1) Hex(1), 511 = HexNAc(1)Hex(1)Fuc(1), 656 = HexNAc(1)Hex(1)NeuAc(1), 730 = HexNAc(2) Hex(2), 1021 = HexNAc(2)Hex(2)NeuAc(1). Site of modification is indicated by g = glycosylation, p = phosphorylation, m = methylation and d = dimethylation. Fifteen out of eighteen detected glycopeptide sequences are predicted to be strong binders to the B07:02 allele.*

Peptide (length)	Glycans	Protein	GlcNAc/GalNAc Ratio	Ying-O-Yang
RPPVgTKASSF (10)	203	R3HD2_HUMAN R3H domain-containing protein 2	1.62	0.83
RVKTPTgSQSYR (11)	365	ZN281_HUMAN Zinc finger protein 281	27	0.69
RVKpTPTgSQSYR (11)	656	ZN281_HUMAN Zinc finger protein 281	12.5	0.7
PSSGLGVgTKQDLGPVPM (17)	203	CD74_HUMAN HLA class II histocompatibility antigen invariant chain	0.72	0.32
APRAIPgTTGRGSSGVGL (17)	365	MCM3_HUMAN DNA replication licensing factor MCM3	0.46	0.52
APRAIPgTTGRGSSGVGL (17)	656	MCM3_HUMAN DNA replication licensing factor MCM3	0.25	0.52

Supplementary Table.2: *Selected example peptides for MS/MS fragmentation to determine the nature of the O-linked saccharide moiety. Site of modification is indicated by g = glycosylation, p = phosphorylation. In all cases the determined O-GlcNAc/GalNAc ratio was crucial for correct glycopeptide identification*

Synthetic Peptide (length)	Protein	GlcNAc/GalNAc Ratio
LPKPANgTSAL (10)	RFX7_HUMAN DNA-binding protein RFX7	1.57
RPPVgTKASSF (10)	R3HD2_HUMAN R3H domain-containing protein 2	1.48
RVKTPTgSQSYR (11)	ZN281_HUMAN Zinc finger protein 281	3.12
RVKpTPTgSQSYR (11)	ZN281_HUMAN Zinc finger protein 281	2.75
APVgSSKSSL (9)	HIPK1_HUMAN Homeodomain-interacting protein kinase 1	2.25
VPEVgTKPSL (9)	TNR6B_HUMAN Trinucleotide repeat-containing gene 6B protein	1.71
IPAVgTRSTI (9)	LAP2_HUMAN Protein LAP2	1.47
IPVSgSHNSL (9)	MEF2C_HUMAN Myocyte-specific enhancer factor 2C	2.65
TPASgSRAQTL (10)	SPTB2_HUMAN Spectrin beta chain	2.16
APRAIPgTTGRGSSGVGL (17)	MCM3_HUMAN DNA replication licensing factor MCM3	N/A
PSSGLGVgTKQDLGPVPM (17)	HG2A_HUMAN HLA class II histocompatibility antigen invariant chain	N/A

Supplementary Table.3: *Data on in vitro O-GlcNAcylated synthetic peptides. Site of modification in vivo is indicated by g = glycosylation. Phosphorylation modifications made synthetically are indicated by p = phosphorylation. All synthetic peptides were subjected to O-GlcNAcylated in vitro by using the O-GlcNAc transferase enzyme. N/A = not applicable.*

Residue number	Residue short name	Residue name	Relative Side chain accessibility	Standard deviation
1	L	LEU	0.9	1.2
2	P	PRO	1.0	0.9
3	K	LYS	26.1	17.4
4	P	PRO	34.1	8.1
5	A	ALA	56.2	24.5
6	N	ASN	40.4	38.5
7	T	THR	24.9	21.1
8	S	SER	19.1	17.3
9	A	ALA	17.8	16.0
10	L	LEU	0.9	0.7

Supplementary Table.4.1: *Data on the peptide LPKPANTSAL*

Residue number	Residue short name	Residue name	Relative Side chain accessibility	Standard deviation
1	R	ARG	1.1	1.4
2	P	PRO	2.3	2.3
3	P	PRO	11.3	11.1
4	V	VAL	15.3	6.1
5	T	THR	46.8	9.0
6	K	LYS	55.5	16.1
7	A	ALA	19.4	11.8
8	S	SER	17.5	12.4
9	S	SER	27.6	26.3
10	F	PHE	6.4	3.8

Supplementary Table.4.2: *Data on the peptide RPPVTKASSF*

Residue number	Residue short name	Residue name	Relative Side chain accessibility	Standard deviation
1	R	ARG	20.6	3.8
2	V	VAL	2.2	1.4
3	K	LYS	21.3	13.5
4	T	THR	39.8	7.4
5	P	PRO	18.4	5.1
6	T	THR	55.7	14.4
7	S	SER	43.6	35.6
8	Q	GLN	43.7	25.7
9	S	SER	7.3	4.0
10	Y	TYR	2.9	2.3

Supplementary Table.4.3: *Data on the peptide RVKTPTSQSY*

Residue number	Residue short name	Residue name	Relative Side chain accessibility	Standard deviation
1	R	ARG	13.6	1.2
2	V	VAL	1.5	0.9
3	K	LYS	18.2	1.8
4	T	THR	35.9	2.5
5	P	PRO	25.3	2.2
6	T	THR	99.1	2.8
7	S	SER	22.7	6.5
8	Q	GLN	106.6	8.6
9	S	SER	21.0	5.7
10	Y	TYR	10.2	3.2
11	R	ARG	0.6	0.2

Supplementary Table.4.4 *Data on the peptide RVKTPTSQSYR*

Residue number	Residue short name	Residue name	Relative Side chain accessibility	Standard deviation
1	A	ALA	0.5	0.9
2	P	PRO	0.2	0.2
3	V	VAL	15.4	2.4
4	S	SER	18.4	1.3
5	P	PRO	25.2	3.9
6	S	SER	63.7	6.6
7	S	SER	0.9	0.5
8	Q	GLN	27.3	2.0
9	K	LYS	1.4	1.2
10	L	LEU	0.0	0.0

Supplementary Table.4.5 *Data on the peptide APVSPSSQKL*

Residue number	Residue short name	Residue name	Relative Side chain accessibility	Standard deviation
1	I	ILE	0.8	0.5
2	P	PRO	0.9	0.4
3	R	ARG	21.3	1.5
4	P	PRO	55.6	2.0
5	P	PRO	28.6	10.5
6	I	ILE	18.2	6.9
7	T	THR	32.3	4.6
8	Q	GLN	18.2	10.6
9	S	SER	6.4	4.2
10	S	SER	0	0
11	L	LEU	6.1	1.3

Supplementary Table.4.6 *Data on the peptide IPRPPITQSSL*

CHAPTER 6

A molecular basis for the presentation of phosphorylated peptides by HLA-B antigens

Adan Alpizara[a], Fabio Marino[b,c], Antonio Ramos-Fernandezd[d], Manuel Lombardia[a], Anita Jeko[b,c], Florencio Pazos[e], Alberto Paradela[a], César Santiago[f], Albert J. R. Heck[b,c] and Miguel Marcilla[a]

[a] Proteomics Unit, Spanish National Biotechnology Centre (CSIC), Darwin 3, 28049, Madrid, Spain.
[b] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Science Faculty, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands.
[c] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, the Netherlands.
[d] Proteobotics SL, Spanish National Biotechnology Centre (CSIC), Darwin 3, 28049, Madrid, Spain.
[e] Computational Systems Biology Group, Spanish National Biotechnology Centre (CSIC), Darwin 3, 28049, Madrid, Spain.
[f] Macromolecular X-ray Crystallography Unit, Spanish National Biotechnology Centre (CSIC), Darwin 3, 28049, Madrid, Spain.

[1] Equal contributors

Manuscript under revision

ABSTRACT

As aberrant phosphorylation is a hallmark of tumor cells, the display of tumor-specific phosphopeptides by Human Leukocyte Antigen (HLA) class I molecules can be exploited in the treatment of cancer by T-cell-based immunotherapy. Yet, the characterization and prediction of HLA-I phospholigands is challenging as the molecular determinants of the presentation of such post-translationally modified (PTM) peptides are not fully understood. Here, we employed a peptidomic workflow to identify 256 unique phosphorylated ligands associated with HLA-B*40, -B*27, -B*39 or -B*07. Remarkably, these phosphopeptides showed similar molecular features. Besides the specific anchor motifs imposed by the binding groove of each allotype, the predominance of phosphorylation at peptide position 4 (P4) became strikingly evident, as was the enrichment of basic residues at P1. To determine the structural basis of this observation, we carried out a series of peptide binding assays and solved the crystal structures of HLA-B*40 in complex with a phosphorylated ligand or its non-phosphorylated counterpart. Overall, our data provide a clear explanation to the common motif found in the phosphopeptidomes associated to different HLA-B molecules. The high prevalence of phosphorylation at P4 is dictated by the presence of the conserved residue Arg62 in the heavy chain, a structural feature shared by most HLA-B alleles. In contrast, the preference for basic residues at P1 is allotype-dependent and might be linked to the structure of the A pocket. This molecular understanding of the presentation of phosphopeptides by HLA-B molecules can help in predicting tumor-specific neo-antigens that arise from aberrant phosphorylation in cancer cells.

1. INTRODUCTION

HLA class I molecules display at the cell surface peptide ligands derived from the degradation of endogenous proteins and present them to cytotoxic T lymphocytes (CTLs). In this way, virally infected or tumor cells can be specifically recognized by the immune system leading to the activation of CTLS that exert their cytotoxic effect on the antigen presenting cell. Classical HLA-I molecules –encoded by genes in the HLA-A,-B and-C loci– encompass over 7000 different proteins derived from more than 10,000 alleles (1). The peptide repertoire associated to a particular class I allotype includes several thousand ligands with defined structural motifs that allow their binding to the class I molecule. These peptides can also harbor post-translational modifications (PTMs) (2). In particular, phosphorylated HLA-I ligands have lately received attention, being proposed as potential targets for cancer immunotherapy on the basis that aberrant phosphorylation is a hallmark of tumor cells (3, 4). In this context, it is known that CTLs can recognize and respond specifically to phosphorylated HLA-I epitopes (5, 6). Despite their potential relevance for cancer immunotherapy, the number of HLA class I-bound phosphopeptides described so far is limited and phosphopeptidomic studies have been performed for a few allotypes only. For instance, Zarling et al. identified 15 phospholigands bound to different class I molecules, among them 6 presented by HLA-B*27 and 6 by HLA-B*07 (5). The same group also described a total of 36 phosphopeptides displayed by HLA-A*02 (3). Meyer et al. characterized 11 phosphorylated ligands, 8 of them associated to HLA-B*07 (4). Finally, Cobbold et al. identified 10 and 85 phosphopeptides restricted, respectively, by HLA-A*02 and HLA-B*35 (7). Here, following up on our previous study on the phospholigandome of HLA-B*40 (8), we

combined a phosphopeptide enrichment strategy with high-resolution LC-MS/MS analysis employing the relatively new Electron Transfer/Higher-Energy Collision Dissociation (EThcD) fragmentation scheme (9) to expand the known repertoire of B*40-bound phosphopeptides and to investigate the presentation of phosphorylated ligands by HLA-B*39,-B*27 and-B*07. We earlier showed that EThcD is especially suited for the characterization of the peptidomes bound to HLA-I (10) and HLA-II molecules (11) and their PTMs (12). Using this workflow, we were able to generate a resource of over 260 phosphorylated HLA-I ligands. These sets of phosphopeptides showed remarkable similarities, displaying a substantial enrichment of phosphorylation at P4 and a high frequency of basic residues at P1. Moreover, the phospholigands reported in previous studies involving HLA-A*02 (3) and HLA-B*07 (7) share the same generic features. We further used biochemical binding assays and X-ray crystallography to investigate the interaction of the ligands phosphorylated at P4 with the HLA-B*40 binding groove. Overall, our data suggest a common structural mechanism to explain the molecular features observed in all the HLA-B-bound phosphopeptidomes studied so far, and provide us with new rules for predicting phosphorylated HLA epitopes.

2. EXPERIMETAL SECTION

Cell lines and Monoclonal Antibodies

HMy2.C1R (C1R) is a human lymphoid cell line with very low expression of HLA-C*04:01 and HLA-B*35:03. The stable C1R transfectants expressing HLA-B*40:02 (C1R-B*40) and HLA-B*39:01 (C1R-B*39) have been described previously (8, 17). GR is a human lymphoblastoid cell line expressing HLA-A*01,-A*03,-B*07,-B*27,-C*02 and-C*07. The monoclonal antibody W6/32 is an IgG2a specific for a monomorphic HLA class I determinant.

Isolation of the HLA-I ligandomes and phosphopetide enrichment

The peptidomes bound to HLA-B*40 and –B*39 were purified from C1R-B*39 or C1R-B*40 cells and subjected to phosphopeptide enrichment exactly as described elsewhere (8). The isolation of the HLA-I-bound peptidome from the GR cell line was carried out as described by Mommen et al (10).

Trypsin digestion and phosphopeptide enrichment for full proteome and phosphoproteome analysis

C1R-B*40 cells were lysed in 8M urea, 50 mM ammonium bicarbonate (pH 8.0) supplemented with a cocktail of phosphatase (PhosStop, Roche) and protease (Complete mini EDTA free, Roche) inhibitors. After centrifugation, the soluble fraction was recovered and proteins were reduced with 2 mM DTT at 56°C for 25 minutes and alkylated with 4 mM iodacetamide at room temperature for 30 minutes in the dark. A 1:75 ratio of Lys-C was used for a digestion at 37°C for 4h. Then, the sample was diluted 4 times in 50 mM ammonium bicarbonate, trypsin was added at a 1:100 ratio and digestion was left to proceed overnight at 37°C. The phosphopeptide enrichment by Ti4+-IMAC was performed as established by Zhou et al. (30).

LC-MS/MS Analysis

The B*40-associated peptidome and phosphopeptidome were analyzed by LC-MS/MS using a Thermo Scientific EASY-nLC 1000 (Thermo Fisher Scientific) coupled to an ETD-enabled LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific). MS spectra were

acquired in the Orbitrap analyzer at a resolution of 60,000 (FWHM). The 10 most abundant precursor ions –excluding unknown and 1+ charge states– were selected for EThcD fragmentation. Daughter ions were detected in the Orbitrap at a resolution of 15,000 (FWHM). The HLA-I ligandome of the GR cell line was analyzed in a similar way using an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific). The HLA-B*39-bound peptidome and phosphopeptidome were analyzed in a nano-LC Ultra HPLC (Eksigent) coupled online with a 5600 triple TOF mass spectrometer (AB Sciex) as previously described (8). The tryptic peptides derived from the proteome of the C1R-B*40 cell line were analyzed in a Thermo Scientific EASY-nLC 1000 (Thermo Fisher Scientific) coupled online to an ETD-enabled Orbitrap Fusion mass spectrometer. The phosphopeptide enriched fractions were analyzed with two different methods, either only EThcD with an Orbitrap readout at a resolution of 15,000 (FWHM) or a decision tree procedure carried out similarly to Frese et al. (31) with the instrument switching between HCD and ETciD and an ion trap readout.

MS/MS Ion Search and Peptide Identification

Peptide identification was carried out as described elsewhere (32). LC-MS/MS data were searched against a target-decoy database containing the Homo sapiens complete proteome set (Uniprot) and their corresponding reverse sequences. Five search engines were used: MASCOT, OMSSA, X!Tandem2, X!Tandem2 with k-score plugin and Myrimatch. Each engine-specific scoring scheme was converted to a common probability-based scale as previously described (33). The data corresponding to HLA-I-bound peptidomes were searched with no enzyme restriction, an MS tolerance of 0.01 Da and an MS/MS tolerance of 0.02 Da. The following variable modifications were considered: oxidation of methionine, protein N-terminal acetylation, pyro-Glu formation from N-terminal Gln or Glu and phosphorylation of Ser, Thr and Tyr. For the analysis of the phosphoproteome of C1R-B*40 cells, trypsin was selected as enzyme allowing up to 2 missed cleavages. The variable modifications specified were: oxidation of methionine, protein N-terminal acetylation, pyro-Glu formation from N-terminal Gln or Glu and phosphorylation of Ser, Thr and Tyr. MS tolerance was set to 10 ppm. MS/MS tolerance was set to 0.05 Da (EThcD) or 0.6 Da (ETciD and HCD). Identifications were filtered at a FDR ≤ 1% at the peptide level. For the identification of HLA-B*39-bound phosphopeptides, the spectra matching phosphorylated sequences with His or Arg at P2 were manually inspected. Those that showed signals that could derive from the neutral loss of the phosphate group were compared with the MS2 spectra of the corresponding synthetic peptides.

Peptide Synthesis and binding assays

Peptides were synthesized using standard F-moc chemistry and purified by reversed phase chromatography. Those peptides intended for binding assays were quantified by amino acid analysis. The peptide binding assays were carried out exactly as described elsewhere (8).

Protein expression and structure determination

The cDNA coding for residues 1-274 of the HLA-B*40:02 heavy chain was cloned in the pET-22b vector (Novagen) and transformed into the E. coli strain Rosetta-gami(DE3)pLys. A cDNA encoding µ2m cloned in the pET-30a vector (Novagen) was kindly provided by Prof. James McCluskey (University of Melbourne) and was transformed in the E. coli strain BL21(DE3)pLys. Protein expression was induced with 0.5 mM IPTG when the cultures, in LB

medium, reached an OD600 = 0.6. Bacteria were harvested 4 (µ2m) or 5 h (heavy chain) after induction and the inclusion bodies were purified as described elsewhere (34). Refolding of the trimeric complex was carried out similarly as described by Reid et al (35) with 15 mg of heavy chain, 20 mg of µ2m and 10 mg of peptide (molar ratio 1:4:20). The mixture was concentrated and fractionated by gel filtration. The fractions corresponding to the properly folded complexes were pooled and concentrated up to 15-20 mg/ml. The crystallization of HLA-B*40 in complex with the peptides REF(pS)KEPEL or REFSKEPEL was performed by vapor diffusion in sitting drop. To that end, 150 nl of the trimeric complexes were mixed with 150 nl of every crystallization solution included in the Pi-Minimal kit (Jena Bioscience). Crystals of both complexes were observed under the following conditions: 8% ethylene glycol, 20% PEG 5000, 150 mM TRIS, pH 8.0 (REF(pS)KEPEL) and 70 mM potassium thiocyanate, 23% PEG 8000, 150 mM TRIS, pH 8.5 (REFSKEPEL). Crystals were collected and frozen in liquid nitrogen using 20% ethylene glycol (REF(pS)KEPEL) or 20% glycerol (REFSKEPEL) as cryo-protector. X-ray diffraction experiments were carried out using the BL13-XALOC beamline of the ALBA synchrotron. The recorded images were integrated with XDS and scaled with AIMLESS and the resulting data were processed with PHASER to solve the structures by molecular replacement using the crystal structure of HLA-B*41:04 (PDB: 3NL5) as template. When a solution was found, the amino acid sequence was manually replaced using COOT and the structure was refined with PHENIX.

3. RESULTS

In-depth characterization of the HLA-B*40-bound peptidome and phosphopeptidome

As a follow-up to our previous work (8), we set out to identify as many HLA-B*40 phospholigands as possible. B*40 was immunopurified from C1R cells stably transfected with this allele and, following peptide elution and phosphopeptide enrichment, its associated peptidome and phosphopeptidome were analyzed by LC-MS/MS employing EThcD as peptide fragmentation. Database searching on the peptidome data allowed the identification of 7375 unique peptides at a FDR < 1%. Of them, 70 (1%), 214 (3%) and 6674 (91%) were classified as B*35, C*04 or B*40 ligands, respectively, according to the binding motif reported for these molecules (8, 13-15). Only 417 sequences (6%) could not be confidently assigned to any allotype. Following analysis of the phosphopeptide-enriched fraction of the B*40 ligandome, a total of 113 phosphorylated sequences with Glu, Asp or pSer at P2 –the canonical B*40 binding motif (8)– could be identified. To evaluate the reliability of these identifications, we synthesized 26 phosphopeptides and subjected those to analysis by EThcD. The spectra of the B*40-bound phosphopeptides and their synthetic counterparts showed excellent correlation allowing the unambiguous confirmation of the sequences (Supplementary Figure.1). We extended the catalog of the here-identified B*40-associated phosphopeptides, including those from a previously published dataset of 85 phosphorylated ligands identified by us using CID-MS (8). The final compilation comprises 148 unique phosphopeptides derived from 136 proteins. Next, we queried this list of sequences for specific molecular characteristics. Several remarkable observations emerged when the position of the phosphorylated residue was considered. First, phosphorylation was found at P4 in 102 ligands (68%, Figure.1A). Of them, 65 (63%) carried Pro at P5 and, in the remaining 37 sequences (27%) P5 was enriched for hydrophobic amino acids (i.e. Leu, Phe, Val, Ile and Met) (Figure.1B). Moreover, 57 (56%) of the HLA-B*40 ligands phosphorylated at P4 had also a basic residue,

primarily Arg, at P1. Performing a similar analysis on the full non-modified ligandome it became quite apparent that all these sequence features were unique for the HLA-B*40-bound phosphopeptides (Figure.1C). The sequence motif exhibited by the non-modified peptides was primarily defined by the two main anchor positions of HLA-B*40, (i.e. Glu or Asp at P2 and Met, Phe or aliphatic residues at the C-terminus). Beside the predominance of phosphorylation at P4, we also detected, in agreement with our previous report (8), 7 ligands harboring phosphorylation at P2 instead of Glu or Asp, hinting at the fact that pSer could mimic and replace these residues at this major anchor position.

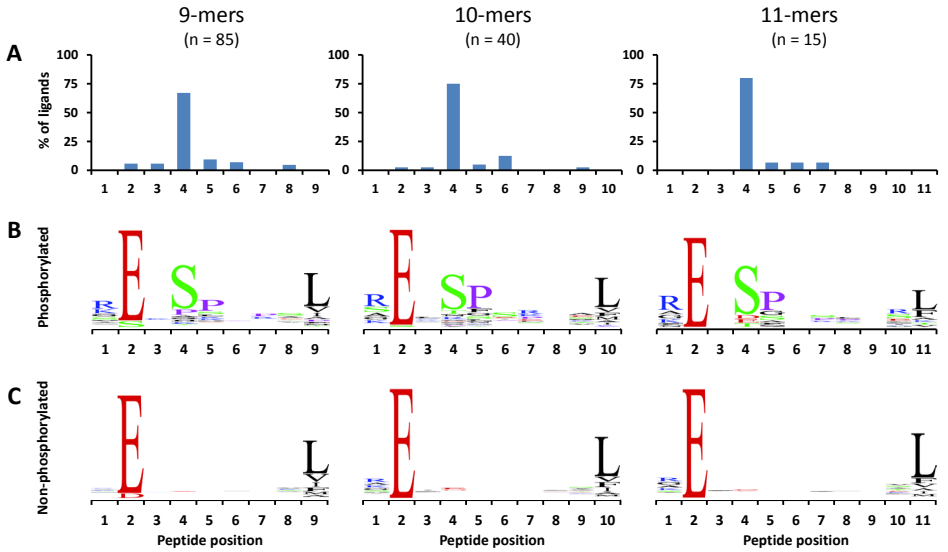


Figure.1: *Sequence analysis of the phosphorylated and the non-phosphorylated peptides associated to HLA-B*40. Only the data corresponding to peptides of 9 to 11 residues long are shown. (A) Frequency distribution of phosphorylation among the identified HLA-B*40 phospholigands (blue) and the in silico predicted HLA-B*40 binders (red). (B and C) Sequence logos of the phosphopeptides (B) and the non-modified peptides (C) identified in the HLA-B*40 ligandome.*

Influence of the cell kinome in the shaping of the B*40-bound phosphopeptidome

Next, to test if the bias for phosphorylated residues at P4 was the result of a kinase specific recognition motif we analyzed the full phosphoproteome of the C1R-B*40 cells. In this comprehensive study we detected a total of 10,153 unique phosphopeptides (FDR ≤ 1%) corresponding to 2369 proteins. Then, we used this information to predict in silico 2017 sequences of 9 to 11 amino acids that matched the canonical B*40 binding motif (i.e. Glu at P2 and Leu, Ile, Val, Met or Phe at the C-terminus). No enrichment of pSer or pThr at P4 or any other position was observed among the predicted binders (Figure.1A). Moreover, when the sequences phosphorylated at P4 were considered, only 50 out of 284 (18%) carried basic residues at P1. In contrast, Pro was somewhat overrepresented at P5 (67 sequences, 24%). This suggests that, while the Arg1/Lys1 motif probably reflects a structural constraint imposed by the B*40 binding groove the high frequency of Pro5 most likely reflects the activity of specific kinases.

Effects of phosphorylation and Arg at P1 on the peptide binding affinity to HLA-B*40

A possible explanation for the observed high frequency of phosphorylation at P4 is that

the phosphate moiety at this position increases the stability of the HLA-peptide complex. To address this point, 7 endogenous phosphorylated ligands and their non-phosphorylated counterparts were synthesized and assayed for binding to HLA-B*40. Most peptides displayed IC50 values in the high nanomolar or low micromolar range (0.3-3.8 µM), corresponding to strong and medium affinity binders (Figure 2A). We also identified a weak binder (REASPSRLS) with an IC50 of 9.9 µM and 14.6 µM for the non-phosporylated and the phosphorylated species, respectively. In every tested pair of peptides, the IC50 of the phosphorylated ligand was slightly higher, revealing that phosphorylation at P4 has a somewhat negative effect on the binding affinity to B*40. Given that phosphorylation at P4 certainly does not enhance binding, we wondered whether this recurrent feature was the result of a negative selection operating on other peptide positions. Therefore, we evaluated the binding of a set of poly-Gly analogues displaying the B*40 binding motif and a single phosphoserine residue at different positions. As shown in figure 2B, the highest binding efficiencies were observed when phosphoserine was placed at P4 (0.5 µM) and P8 (0.6 µM) while phosphorylation at P5, P6 and P7 led to a moderate decrease of affinity (1.9 µM, 1.8 µM, 1.1 µM, respectively). Finally, phosphorylation at P3 caused a marked increase of the IC50 (25.1 µM) reflecting a considerable negative effect on binding. To assess the effect of a positively charged residue at P1, we tested a second set of poly-Gly analogs harboring Arg at the peptide N-terminus. Regardless of the phosphorylated position, the presence of this residue at P1 enhanced the binding affinity of each assayed peptide (Figure 2B), partly explaining the increased frequency of this motif in the B*40-associated phosphopeptidome.

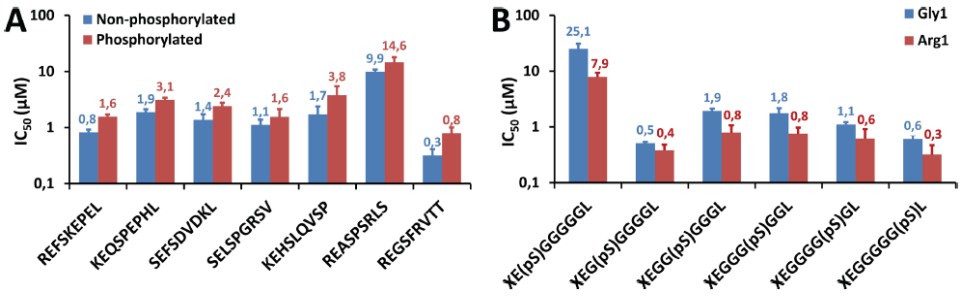


Figure.2: *Peptide binding assays to HLA-B*40. (A) Binding affinities –expressed as IC50 values– of seven phospholigands associated to HLA-B*40 and their non-phosphorylated counterparts. In all cases, the pair of peptides exhibit alike affinities although the affinity of the phosphopeptides is somewhat lower. (B) Binding affinities of 12 phosphorylated poly-Gly analogs with Gly (blue) or Arg (red) at P1.*

Structures of HLA-B*40 in complex with an endogenous phospholigand or its non-phosphorylated counterpart

We next set out to crystallize HLA-B*40 –a molecule whose crystal structure had not been reported before– in complex with a natural phosphorylated ligand (REF(pS)KEPEL) and its non-phosphorylated counterpart (REFSKEPEL). The two structures were solved and refined at a resolution of 1.5 and 1.8 Å, respectively. The overall fold of both complexes was almost identical (RMSD = 0.128 Å), with both ligands adopting the typical extended conformation and the peptide N- and C-termini anchored in the binding cleft (Figure.3A-B). As expected, P2-Glu and P9-Leu were accommodated in the B and F pockets, respectively. P6-Glu was oriented towards the floor of the binding groove whereas P1-Arg, P4-Ser/P4-pSer, P5-Lys and P8-Glu were pointed upwards, mak-

ing these side chains potentially accessible for interaction with the TCR. Both ligands adopted a very similar conformation, although we observed minor differences in the orientation of residues at P1, P5 and P8. Both P1-Arg and P8-Glu showed two alternative conformations in the phosphorylated peptide (Figure.3B) inducing some changes in the positions of residues Arg62, Glu76 and Glu163 of the heavy chain (Figure.3C). Regarding P5-Lys, the weak electron density around its side chain is probably indicative of a high flexibility. Nevertheless, in the peptide REF(pS)KEPEL, the ε-amino group of this residue tended to be oriented towards P4-pSer, likely influenced by the negative charge of the phosphate group (Figure.3B). In the phosphorylated ligand, the phosphate moiety of P4-pSer interacted directly with Arg62 and via two ordered water molecules with Arg62 and Glu163 on the heavy chain and P1-Arg on the peptide (Figure.4A-B). In contrast, in the non-phosphorylated peptide, P4-Ser was not involved in any significant interaction with residues of the heavy chain but took part in a network of ordered water molecules, indirectly linking its side chain to that of Arg62 and Glu163 (Figure.4C-D). Finally, in both complexes, the side chain of P1-Arg was stabilized in the A pocket through three main interactions: 1) π-π stacking of its guanidinium group with that of Arg62, 2) hydrophobic interactions of the aliphatic part of its side chain with the indole ring of Trp167 and 3) a salt bridge formed between its guanidinium group and the carboxyl group of Glu163 (Figure.4). This mode of binding provides at least a partial explanation to the increased affinity for HLA-B*40 of the phosphopeptides with Arg at P1.

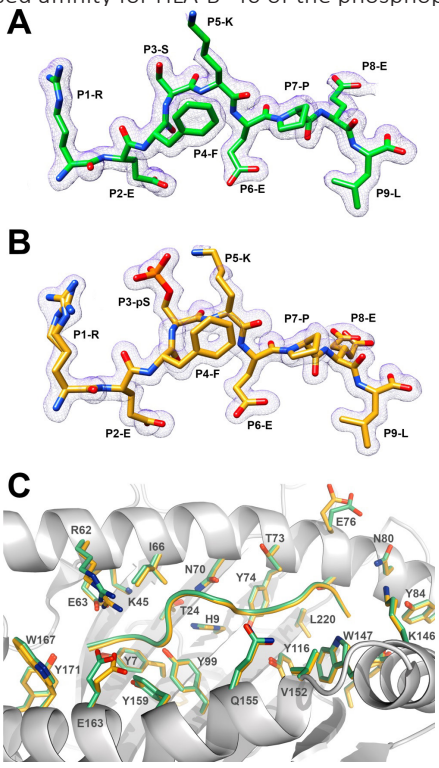


Figure.3: *Crystal structures of HLA-B*40 in complex with the peptides REF(pS)KEPEL and REFSKEPEL. (A and B) Structure and electron density map (2Fo-Fc) at 1σ (blue mesh) of the peptides REFSKEPEL (A) and REF(pS)KEPEL (B) anchored to the B*40 binding cleft. (C) Superposition of the binding grooves of both complexes with the peptides REFSKEPEL (green) and REF(pS)KEPEL (yellow) displayed in ribbon style.*

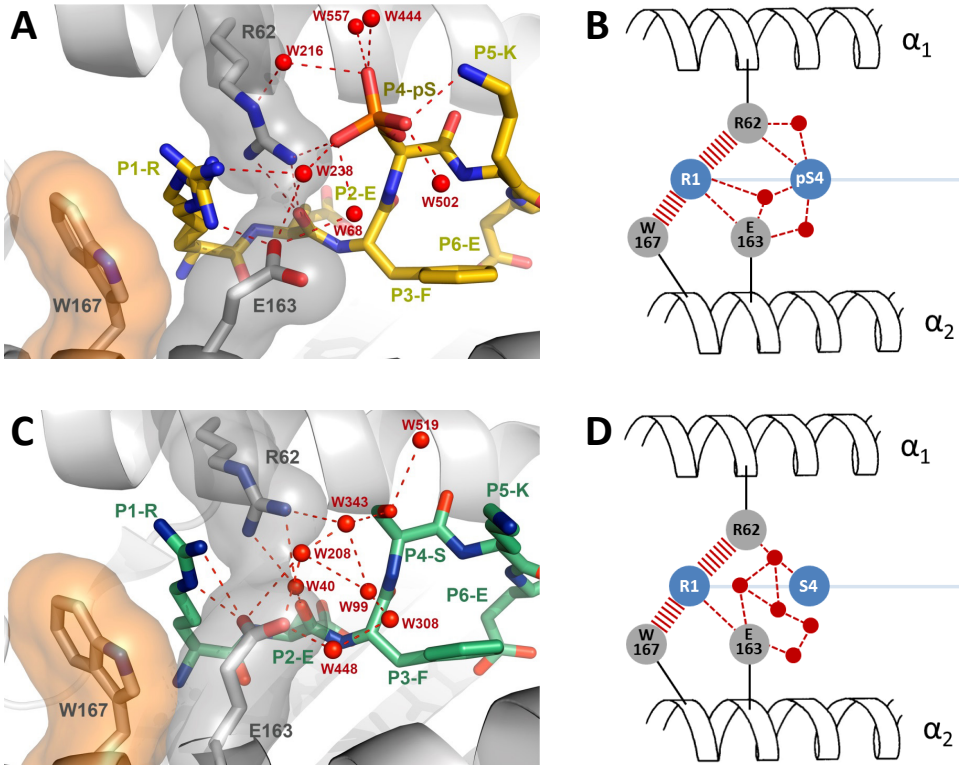


Figure.4. *Overview of the interactions of residues at P1 and P4 of the peptides REF(pS)KEPEL (A and B) and REFSKEPEL (C and D) with the binding groove of HLA-B*40. (A) Interactions involving residues P1-R and P4-pSer in the complex B*40-REF(pS)KEPEL. The carbon atoms of the peptide ligand are depicted in yellow. (B) Schematic representation of the interactions displayed in panel A. (C) Interactions involving residues P1-R and P4-pSer in the complex B*40-REFSKEPEL. The carbon atoms of the peptide ligand are depicted in green. (D) Schematic representation of the interactions displayed in panel C.*

Preferred Phosphorylation at P4 is a landmark of more HLA-B alleles

Notably, inspecting the scarce and scattered available data from the literature, the strong preference of HLA-B*40 for peptides phosphorylated at P4 can be also observed in other HLA-B allotypes like B*27 (5) or B*07 (7) (Table.1). We above showed that Arg62 is a key residue in the stabilization of the phosphate moiety at P4 in HLA-B*40. Since it is conserved in most HLA-B alleles, we hypothesized that this could be the basis of a common phosphorylation motif. To test this hypothesis, we extended our analysis to the phosphopeptidomes associated to HLA-B*39,-B*07 and-B*27, expressed in different cell lines. First, we isolated the peptidome and phosphopeptidome displayed by HLA-B*39 from C1R cells stably transfected with this allele as described for HLA-B*40 and analyzed them by LC-MS/MS. A total of 1128 unique peptides of 8 to 13 residues long were identified at a FDR ≤ 1% at the peptide level. Of them, 953 (84%) had His or Arg at P2, the canonical binding motif of this allele (16, 17) while 14 (1%) and 53 (5%) matched respectively the binding motifs of HLA-B*35 (13) and HLA-C*04 (14). The remaining 108 sequences (10%) could not be confidently assigned to any allotype. Nonetheless, we found 55 peptides (51%) in this latter group that had Gln at P2, suggesting that HLA-B*39 can also bind and present ligands with this suboptimal anchor motif.

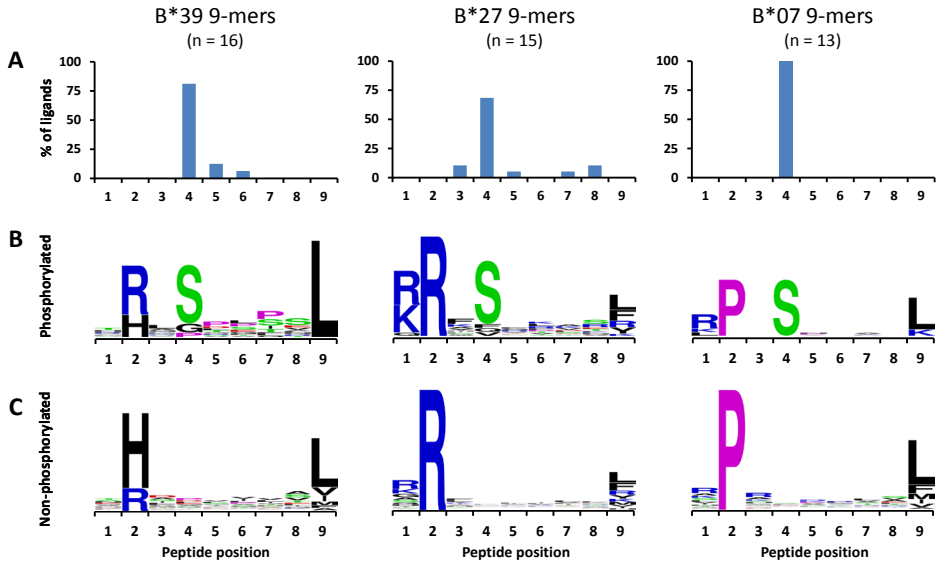


Figure.5: *Sequence analysis of the phosphorylated and the non-phosphorylated peptides associated to HLA-B*39 (left column), HLA-B*27 (middle column) or HLA-B*07 (right column). For clarity, only the data corresponding to 9-mer peptides are shown. For other peptide lengths, see Supplementary Figure.3 to 5. (A) Frequency distribution of phosphorylation at each peptide position in the phospholigandomes bound to these allotypes, revealing the preference for phosphorylation at P4. (B and C) Sequence logos of the phosphopeptides (B) and the non-modified peptides (C) identified in the ligandomes of these HLA-B molecules.*

The analysis by LC-MS/MS of the phosphopeptide-enriched fractions allowed the characterization of 24 phosphorylated ligands that matched the B*39 binding motif. Since data acquisition was carried out using CID –a fragmentation scheme not particularly well-suited for the identification of phosphopeptides– some of the sequences were confirmed by fragmentation of the corresponding synthetic peptide (Supplementary Figure.2). In this set of sequences, phosphorylation was found at P4 in 18 peptides (75%). Distinct from the B*40-bound phosphopeptidome, only 3 phospholigands (17%) carried positively charged residues at the N-terminus. This latter number was, however, significantly higher than the frequency of basic residues at P1 in the non-phosphorylated ligandome where less than 1% of the peptides had Arg or Lys at this position (Figure.5). Finally, we characterized the peptidome and phosphopeptidome displayed by HLA-B*27 and-B*07 in the lymphoblastoid cell line GR. To that end, the HLA-I ligandome of this cell line was fractionated by strong cation exchange (SCX) chromatography prior to analysis by LC-MS/MS with EThcD fragmentation. A total of 11,041 peptides were identified at a FDR ≤ 1% of which 10,462 (95%) were 8 to 13 residues long. Of them, 4153 (40%) and 2523 (24%) matched the binding motifs of HLA-B*27 (18, 19) and HLA-B*07 (20), respectively. Among the B*07 ligands, we could identify 32 phosphorylated peptides of which 19 (59%) displayed their phosphorylation at P4, with 13 of those (68%) harboring a basic amino acid at the N-terminus (Table.1 and Figure.5). For HLA-B*27, we were able to identify 52 phospholigands, including 37 (65%) that were phosphorylated at P4. In this case, the preference for basic residues at P1 was even more apparent since 34 (92%) of the peptides phosphorylated at P4 displayed positively charged residues at P1.

All this new data reiterates that the increased phosphorylation at P4 is a general feature of the phosphopeptidomes displayed by different HLA-B allotypes, despite the diversity of their anchoring motifs. Additionally, even though the preference for a basic residue at P1 in the ligands phosphorylated at P4 is common to all the alleles studied so far, the magnitude of this preference is certainly allotype dependent.

Allotype	Phosphorylated Ligands	P4 phosphorylation	P4 phosphorylation + basic residue at P1	Preferred motif	Reference
	6	5 (83%)	5 (100%[a])	-	-5
HLA-B*07	8	5 (63%)	5 (100% [a])	-	-4
	85	62 (73%)	43 (69% [a])	-	-7
	32	19 (59%)	13 (68%[a])	(R/K)PRpS	This study
HLA-B*27	6	6 (100%)	5 (83% [a])	-	-5
	52	37 (65%)	34 (92%[a])	RRXpS	This study
HLA-B*40	149	102 (68%)	57 (56% [a])	REXpS(L/F/M) and XEXpSP	This study
HLA-B*39	24	18 (75%)	3 (17% [a])	X(R/H)XpS	This study

Table.1: *Molecular features of the phoshopeptidomes associated to different HLA-B molecules.*

4. DISCUSSION

T-cell-based immunotherapy is an emerging new strategy for cancer treatment, especially in late-stage diseases or those that do not respond to conventional therapies (21). This approach relies critically on the identification of tumor-specific MHC-restricted antigens capable of triggering an effective antitumor response. Neo-antigens arising from tumor-specific mutations are arguably the most obvious targets for these personalized therapies, as they are uniquely associated to tumor development. Nevertheless, since kinase and phosphatase function is often deregulated in tumor cells (22), it has been suggested that phosphorylated antigens could also be ideal candidates for immunotherapy, as their higher abundance may be cancer-related (3, 4). Supporting this hypothesis, Zarling et al. showed that CD8+ T-cells directed against two phosphorylated HLA-I epitopes were able to reduce tumor growth in vivo (23). Yet, the characterization of MHC class I-bound phospholigands remains challenging due to their often sub-stoichiometric levels compared to the non-modified ligandome and the difficulties associated with their characterization by most common LC-MS/MS workflows. In an effort to improve the identification of HLA-I phospholigands, we combined a rather novel peptide fragmentation technique (EThcD), phosphopeptide enrichment and advanced database searching to identify about 150 unique phosphorylated peptides presented by HLA-B*40 on the surface of C1R cells. Intriguingly, the phosphopeptidome bound to HLA-B*40 shared some molecular features with those associated to other previously studied MHC class I molecules: phosphorylation occurred mainly at P4 and was frequently accompanied by the presence of a basic residue at P1 (3-5, 7). This prompted us to study the molecular mechanism underlying the presentation of phosphopeptides by this allotype. The binding assays described in our study proved that phosphorylation at P4 has a small negative effect on binding affinity to B*40. Although this result seemingly conflicts with the high occurrence of phosphorylation at this position, the experiments performed with poly-Gly analogues demonstrated that phosphorylation at P4 correlates with higher binding affinity to B*40 when compared with other peptide positions. From a structural point of view, this effect can be explained by the direct contact of the phosphate moiety of P4-Ser with the side chain of Arg62 and

its water-mediated interaction with the carboxyl group of Glu163. This mode of stabilizing the phosphate group at P4 shares some similarities with that described for HLA-A*02, where residues Arg65 and/or Lys66 –also localized in the α1 helix but absent in HLA-B molecules– play a key role in the interaction with the phosphorylated residue (24, 25). The binding assays performed here also account for the high frequency of basic residues at the N-terminus in the B*40-bound phosphopeptidome since the presence of a positively charged residue at this position enhances complex stability, compensating the negative effect on binding caused by phosphorylation at P4. This is probably due to the set of interactions established by the basic residue at P1 with residues of the A pocket. In the two crystal structures presented here, P1-Arg is stabilized through contacts with Arg62, Glu163 and Trp167. Finally, in the phosphorylated ligand, residues P1-Arg and P4-pSer are linked through a water molecule and by a long-distance interaction mediated by residues Arg62 and Glu163, contributing to the overall stability of the complex. Based on these results, we argued that the conservation of Arg62 in the vast majority of HLA-B allotypes could be the basis for their common preference for peptides phosphorylated at P4. The subsequent analysis of the phosphopeptidomes associated to HLA-B*07, HLA-B*27 and HLA-B*39 –which maintain the Arg62 residue– seems to support this hypothesis as they were all found to be enriched in peptides with this molecular feature. Furthermore, the enrichment for Arg or Lys at P1 among the peptides phosphorylated at P4 was also observed in all the HLA-B alleles studied here. Nevertheless, the magnitude of this effect varied greatly among allotypes and correlated with the frequency of basic residues at P1 in the unmodified ligandomes. For instance, in HLA-B*27, the frequency of positively charged residues at P1 among the peptides phosphorylated at P4 was above 90% while it was close to 50% in the unmodified peptidome. In contrast, in B*39, only 17% of the ligands phosphorylated at P4 had positively charged residues at the N-terminus while this frequency was below 1% in the non-phophorylated peptide set. This latter fact might be related to the presence of Thr163 instead of Glu163 in this molecule that would make the A pocket less prone to accommodate positively charged side chains. By generating the largest data set of HLA-I-associated phosphopeptides so far, from four different allotypes, we could conclude that they share a strong preference for peptides phosphorylated at P4. This may become a key parameter in predicting tumor antigens arising from aberrant phosphorylation. In addition, our data also highlights that each HLA-B allotype has its own binding preferences that may have an effect on the molecular features of the presented phosphopeptides. Thefore, if the peptide binding motif of an MHC-I molecule aligns with the phosphorylation motif of a particular kinase, this allotype may become an indicator of the aberrant activity of this kinase. To illustrate our reasoning, nearly all the phosphorylated peptides presented by HLA-B*27 confirm to the RRXpS motif (Table.1), which is a clear hallmark of basophilic kinases such as PKA and PKC (26). In contrast, HLA-B*40 displays a set of phosphorylated ligands that can be separated in the following sequence motifs: REXpS(L/F/M) and XEXSpP. The former resembles the well-known Plk1 kinase motif (27), whereas the latter is informative for proline-directed kinases (28). For B*07 ligands, the prevalent phosphorylation motif is (R/K)PXpS, which loosely confirms to the substrate motif of CDKL5 (29). Therefore, our data seem to predict that individuals expressing specific HLA allotypes may be more prone to present phosphorylated peptides following aberrant function of specific kinases, providing another reason for further research into personalized therapies.

5. ACKNOWLEDGEMENTS

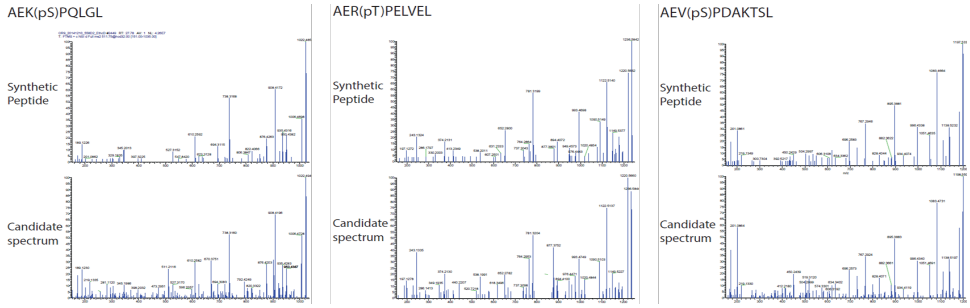
The authors thank Prof. James McCluskey (University of Melbourne) for providing the β2m cDNA clone and Dr. Francisco Gavilanes (Universidad Complutense, Madrid) for his expert help in peptide quantification. The X-ray diffraction experiments were performed at XALOC beamline at ALBA Synchrotron with the collaboration of ALBA staff. This research was performed within the framework of PRIME-XS, grant number 262067, funded by the European Union 7th Framework Program. The Proteomics Unit of the Spanish National Biotechnology Centre belongs to ProteoRed (PRB2-ISCIII) and is supported by the FIS grant PT13/0001. FM and AJRH are supported by the project Proteins At Work, a program of the Netherlands Proteomics Centre financed by the Netherlands Organization for Scientific Research (NWO) as part of the National Roadmap Large-scale Research Facilities of the Netherlands (project number 184.032.201). Additionally, AJRH received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 686547 (MSMed) and from the NWO Gravity Program Institute for Chemical Immunology. AA was funded by the JAE-Pre 2011 program of the Spanish National Research Council (CSIC).

6. REFERENCES

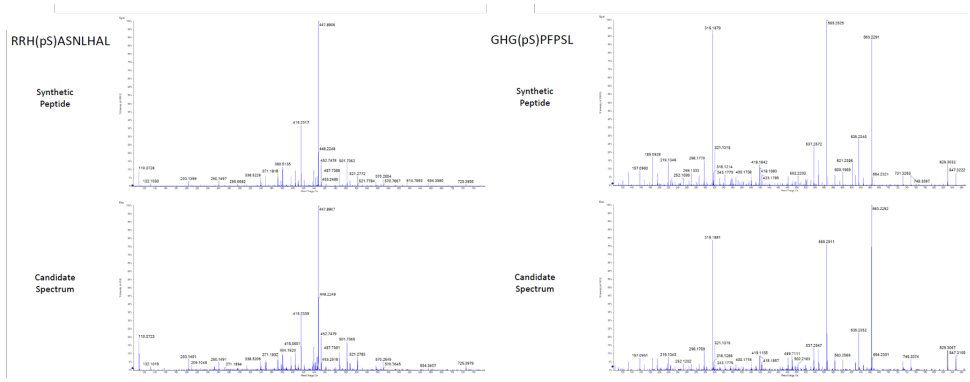
1.Robinson J, Soormally AR, Hayhurst JD, & Marsh SG (2016) The IPD-IMGT/HLA Database- New developments in reporting HLA variation. Human immunology.
2.Engelhard VH, Altrich-Vanlith M, Ostankovitch M, & Zarling AL (2006) Post-translational modifications of naturally processed MHC-binding epitopes. Current opinion in immunology 18(1):92-97.
3.Zarling AL, et al. (2006) Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. Proc Natl Acad Sci U S A 103(40):14889-14894.
4.Meyer VS, et al. (2009) Identification of natural MHC class II presented phosphopeptides and tumor-derived MHC class I phospholigands. J Proteome Res 8(7):3666-3674.
5.Zarling AL, et al. (2000) Phosphorylated peptides are naturally processed and presented by major histocompatibility complex class I molecules in vivo. The Journal of experimental medicine 192(12):1755-1762.
6.Andersen MH, et al. (1999) Phosphorylated peptides can be transported by TAP molecules, presented by class I MHC molecules, and recognized by phosphopeptide-specific CTL. Journal of immunology 163(7):3812-3818.
7.Cobbold M, et al. (2013) MHC class I-associated phosphopeptides are the targets of memory-like immunity in leukemia. Science translational medicine 5(203):203ra125.
8.Marcilla M, et al. (2014) Increased diversity of the HLA-B40 ligandome by the presentation of peptides phosphorylated at their main anchor residue. Mol Cell Proteomics 13(2):462-474.
9.Frese CK, et al. (2012) Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. Anal Chem 84(22):9668-9673.
10.Mommen GP, et al. (2014) Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proc Natl Acad Sci U S A 111(12):4507-4512.
11.Mommen GP, et al. (2016) Sampling from the proteome to the HLA-DR ligandome proceeds via high specificity. Mol Cell Proteomics.
12.Marino F, et al. (2015) Extended O-GlcNAc on HLA Class-I-Bound Peptides. Journal of the American Chemical Society 137(34):10922-10925.
13.Steinle A, et al. (1996) Motif of HLA-B*3503 peptide ligands. Immunogenetics 43(1-2):105-107.
14.Buchsbaum S, et al. (2003) Large-scale analysis of HLA peptides presented by HLA-Cw4. Immunogenetics 55(3):172-176.
15.Schittenhelm RB, Dudek NL, Croft NP, Ramarathinam SH, & Purcell AW (2014) A comprehensive analysis of constitutive naturally processed and presented HLA-C*04:01 (Cw4)-specific peptides. Tissue antigens 83(3):174-179.
16.Falk K, et al. (1995) Peptide motifs of HLA-B38 and B39 molecules. Immunogenetics 41(2-3):162-164.
17.Yague J, et al. (1999) The South Amerindian allotype HLA-B*3909 has the largest known similarity in peptide specificity and common natural ligands with HLA-B27. Tissue antigens 53(3):227-236.
18.Lopez de Castro JA, et al. (2004) HLA-B27: a registry of constitutive peptide ligands. Tissue antigens 63(5):424-445.

19.Ben Dror L, Barnea E, Beer I, Mann M, & Admon A (2010) The HLA-B*2705 peptidome. Arthritis and rheumatism 62(2):420-429.
20.Huczko EL, et al. (1993) Characteristics of endogenous peptides eluted from the class I MHC molecule HLA-B7 determined by mass spectrometry and computer modeling. Journal of immunology 151(5):2572-2587.
21.Wang M, Yin B, Wang HY, & Wang RF (2014) Current advances in T-cell-based cancer immunotherapy. Immunotherapy 6(12):1265-1278.
22.Blume-Jensen P & Hunter T (2001) Oncogenic kinase signalling. Nature 411(6835):355-365.
23.Zarling AL, et al. (2014) MHC-restricted phosphopeptides from insulin receptor substrate-2 and CDC25b offer broad-based immunotherapeutic agents for cancer. Cancer research 74(23):6784-6795.
24.Mohammed F, et al. (2008) Phosphorylation-dependent interaction between antigenic peptides and MHC class I: a molecular basis for the presentation of transformed self. Nature immunology 9(11):1236-1243.
25.Petersen J, et al. (2009) Phosphorylated self-peptides alter human leukocyte antigen class I-restricted antigen presentation and generate tumor-specific epitopes. Proc Natl Acad Sci U S A 106(8):2776-2781.
26.Amanchy R, et al. (2007) A curated compendium of phosphorylation motifs. Nat Biotechnol 25(3):285-286.
27.Nakajima H, Toyoshima-Morimoto F, Taniguchi E, & Nishida E (2003) Identification of a consensus motif for Plk (Polo-like kinase) phosphorylation reveals Myt1 as a Plk1 substrate. J Biol Chem 278(28):25277-25280.
28.Weigel NL & Moore NL (2007) Kinases and protein phosphorylation as regulators of steroid hormone action. Nuclear receptor signaling 5:e005.
29.Katayama S, Sueyoshi N, & Kameshita I (2015) Critical Determinants of Substrate Recognition by Cyclin-Dependent Kinase-like 5 (CDKL5). Biochemistry 54(19):2975-2987.
30.Zhou H, et al. (2013) Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. Nat Protoc 8(3):461-480.
31.Frese CK, et al. (2011) Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. J Proteome Res 10(5):2377-2388.
32.Marcilla M, et al. (2016) Comparative Analysis of the Endogenous Peptidomes Displayed by HLA-B*27 and Mamu-B*08: Two MHC Class I Alleles Associated with Elite Control of HIV/SIV Infection. J Proteome Res.
33.Ramos-Fernandez A, Paradela A, Navajas R, & Albar JP (2008) Generalized method for probability-based peptide and protein identification from tandem mass spectrometry data and sequence database searching. Mol Cell Proteomics 7(9):1748-1754.
34.Jimenez D, Roda-Navarro P, Springer TA, & Casasnovas JM (2005) Contribution of N-linked glycans to the conformation and function of intercellular adhesion molecules (ICAMs). J Biol Chem 280(7):5854-5861.
35.Reid SW, et al. (1996) Production and crystallization of MHC class I B allele single peptide complexes. FEBS Lett 383(1-2):119-123.

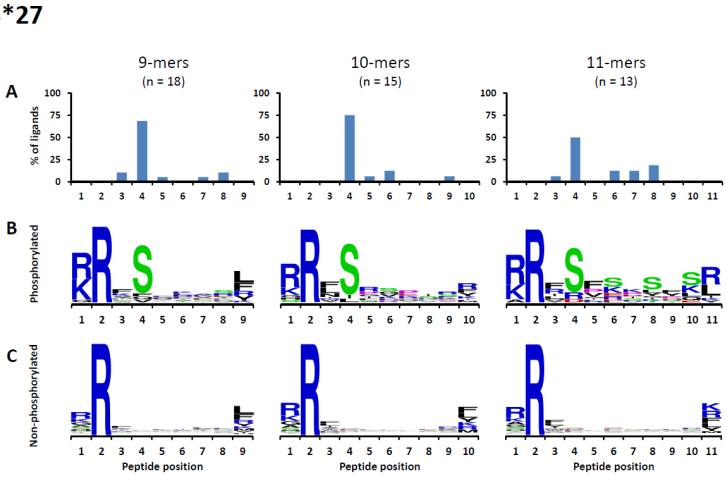
7. SUPPLEMENTARY



Supplementary Figure.1: *EThcD spectra of the endogenous unmodified HLA*B40 peptides (top). For comparison the EThcD spectrum of the synthetic peptide analogue is given (bottom).*

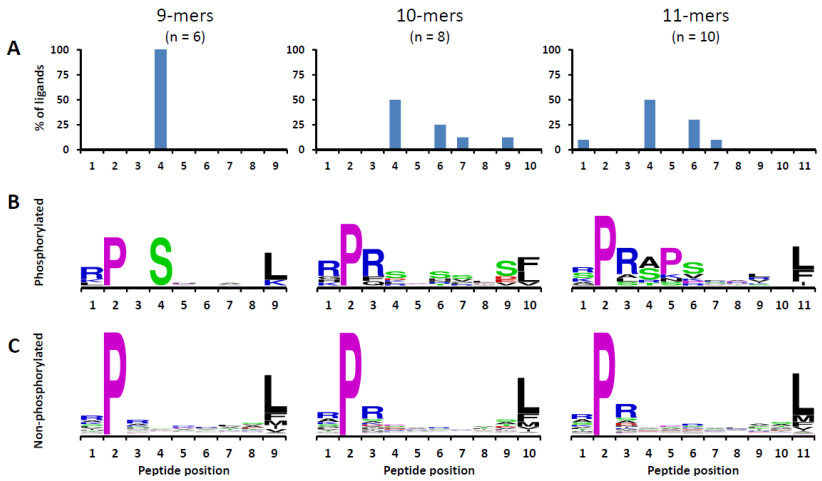


Supplementary Figure.2: *CID spectra of the endogenous unmodified HLA*B39 peptides (top). For comparison the CID spectrum of the synthetic peptide analogue is given (bottom).*



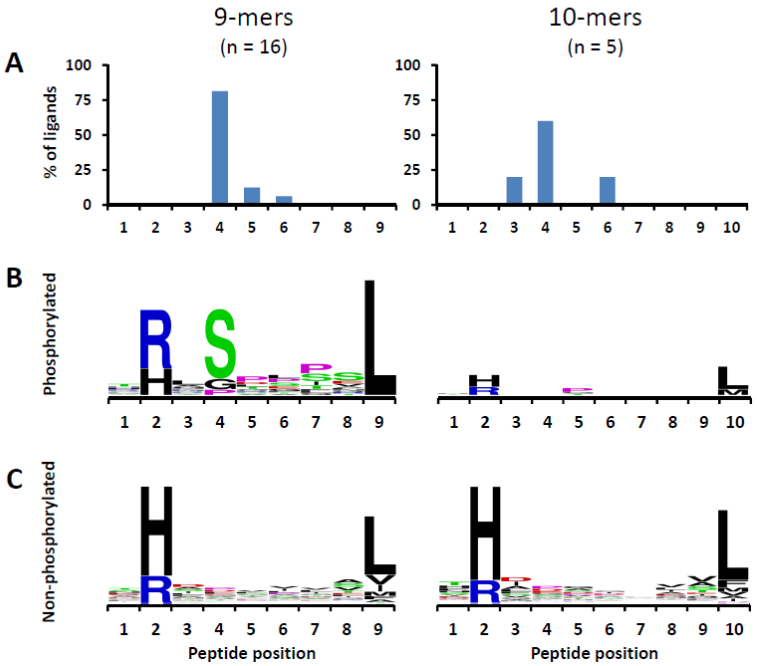
Supplementary Figure.3: *(A) Frequency distribution of phosphorylation at each peptide position in the phospholigandomes bound to the B*27 allotype, revealing the preference for phosphorylation at P4. (B and C) Sequence logos of the phosphopeptides (B) and the non-modified peptides (C) identified in the ligandomes of the HLA-B*27 molecules.*

B*07



Supplementary Figure.4: *(A) Frequency distribution of phosphorylation at each peptide position in the phospholigandomes bound to the B*07 allotype, revealing the preference for phosphorylation at P4. (B and C) Sequence logos of the phosphopeptides (B) and the non-modified peptides (C) identified in the ligandomes of the HLA-B*07 molecules.*

B*39



Supplementary Figure.5: *(A) Frequency distribution of phosphorylation at each peptide position in the phospholigandomes bound to the B*39 allotype, revealing the preference for phosphorylation at P4. (B and C) Sequence logos of the phosphopeptides (B) and the non-modified peptides (C) identified in the ligandomes of the HLA-B*39 molecules.*

CHAPTER 7

Sampling from the proteome to the HLA-DR ligandome proceeds via high specificity

Geert P.M. Mommen[1,2,*], Fabio Marino[2,3,*], Hugo D. Meiring[1], Martien C.M. Poelen[4], Jacqueline A.M. van Gaans-van den Brink[4], Shabaz Mohammed[2,3,5,6], Albert J.R. Heck[2,3], and Cécile A.C.M. van Els[4]

[1]Institute for Translational Vaccinology, P.O. Box 450, 3720 AL Bilthoven, the Netherlands
[2]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Science Faculty, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands.
[3]Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, the Netherlands.
[4]Centre for Infectious Disease Control, National Institute for Public Health and the Environment, P.O. Box 1, 3720 AL Bilthoven, the Netherlands
[5]Chemistry Research Laboratory, Department of Chemistry, University of Oxford, Mansfield Road, OX13TA, Oxford, United Kingdom.
[6]Department of Biochemistry, University of Oxford, South Parks Road, OX1 3QU, Oxford, United Kingdom

* Both authors contributed equally

Molecular Cellular Proteomics: January 13, 2016

ABSTRACT

Comprehensive analysis of the complex nature of the Human Leukocyte Antigen (HLA) class II ligandome is of utmost importance to understand the basis for CD4+ T cell mediated immunity and tolerance. Here, we implemented important improvements in the analysis of the repertoire of HLA-DR-presented peptides, using hybrid mass spectrometry-based peptide fragmentation techniques on a ligandome sample isolated from matured human monocyte-derived dendritic cells (DC). The reported data set constitutes nearly 14 thousand unique high-confident peptides, i.e. the largest single inventory of human DC derived HLA-DR ligands to date. From a technical viewpoint the most prominent finding is that no single peptide fragmentation technique could elucidate the majority of HLA-DR ligands, due to the wide range of physical chemical properties displayed by the HLA-DR ligandome. Our in-depth profiling allowed us to reveal a strikingly poor correlation between the source proteins identified in the HLA class II ligandome and the DC cellular proteome. Important selective sieving from the sampled proteome to the ligandome, was evidenced by specificity in the sequences of the core regions both at their N- and C- termini, hence not only reflecting binding motifs but also dominant protease activity associated to the endolysosomal compartments. Moreover, we demonstrate that the HLA-DR ligandome reflects a surface representation of cell-compartments specific for biological events linked to the maturation of monocytes into antigen presenting cells. Our results present new perspectives into the complex nature of the HLA class II system and will aid future immunological studies in characterizing the full breadth of potential CD4+ T cell epitopes relevant in health and disease.

1. INTRODUCTION

Human Leukocyte Antigen (HLA) class II molecules on professional antigen presenting cells such as dendritic cells (DC) expose peptide fragments derived from exogenous and endogenous proteins to be screened by CD4+ T cells (1, 2). The activation and recruitment of CD4+ T cells recognizing disease-related peptide antigens is critical for the development of efficient anti-pathogen or anti-tumor immunity. Furthermore, the presentation of self-peptides and their interaction with CD4+ T cells is essential to maintain immunological tolerance and homeostasis (3). Knowledge of the nature of HLA class II-presented peptides on DC is of great importance to understand the rules of antigen processing and peptide binding motifs (4), while the identity of disease-related antigens may provide new knowledge on immunogenicity and leads for the development of vaccines and immunotherapy (5, 6). Mass spectrometry (MS) has proven effective for the analysis HLA class II-presented peptides (4, 7, 8). MS-based ligandome studies have demonstrated that HLA class II molecules predominantly present peptides derived from exogenous proteins that entered the cells by endocytosis and endogenous proteins that are associated with the endo-lysosomal compartments (4). Yet proteins residing in the cytosol, nucleus or mitochondria can also be presented by HLA class II molecules, primarily through autophagy (9–11). Multiple studies have mapped the HLA class II ligandome of antigen presenting cells in the context of infectious pathogens (12), autoimmune diseases (13–17) or cancer (14, 18, 19), or those that are essential for self-tolerance in the human thymus (3, 20). Notwithstanding these efforts, and certainly not in line with the extensive knowledge on the HLA class I ligandome (21), the nature of the HLA class II-presented peptide repertoire and

particular its relationship to the cellular source proteome remains poorly understood. To advance our knowledge on the HLA-DR ligandome on activated DC without having to deal with limitations in cell yield from peripheral human blood (12, 21, 22) or tissue isolates (3), we explored the use of MUTZ-3 cells. This cell line has been used as a model of human monocyte-derived DCs. MUTZ-3 cells can be matured to act as antigen presenting cells and express then high levels of HLA class II molecules, and can be propagated in vitro to large cell densities (23–25). We also evaluated the performance of complementary and hybrid MS fragmentation techniques Electron-Transfer Dissociation (ETD), Electron-Transfer/higher-energy collision Dissociation (EThcD) (26) and Higher-energy Collision Dissociation (HCD) to sequence and identify the HLA class II ligandome. Together this workflow allowed for the identification of an unprecedented large set of about 14 thousand unique peptide sequences presented by DC derived HLA-DR molecules, providing an in-depth view of the complexity of the HLA class II ligandome, revealing underlying features of antigen processing and surface-presentation to CD4+ T cells.

2. EXPERIMENTAL SECTION

Ethics

Institutional principles of RIVM relating to the use of material and data obtained from human subjects, including prototype cell lines obtained via cell bank catalogues, are in agreement with the guidelines expressed for Good Clinical Practice expressed in the Declaration of Helsinki.

Cell culturing, lysis of cells and isolation of HLA class II-associated peptides

The deposited MUTZ-3 cell line, a human HLA-DR10, -DR11, -DR52 (HLA-DRB1*10, HLA-DRB1*11, HLA-DRB3*01) positive acute myelo-monocytic leukemia serving as a dendritic cell model (kindly provided by Dr. R. Scheeper, VU University Medical Center, Amsterdam), was grown under maintenance conditions in roller bottles in α-Minimum Essential Medium (Gibco), supplemented 20% heat-inactivated FBS (Hyclone), 100 U/ml penicillin, 100 ug/ml streptomycin (Gibco), 2 mM L-glutamine (Gibco), and 25 U/ml GM-CSF (Prepotech)1. MUTZ-3 cells were induced into an immature DC state by a 5-day exposure to 1000 U/ml GM-CSF (100 ng/ml), 1000 U/ml IL-4 (20 ng/ml) and 2.5 ng/ml TNF-α. Immature MUTZ-3 DC were matured by increasing the concentration of TNFa to 75 ng/ml for 20 hr. During this maturation phase BCG antigens were present as an antigenic pulse (kindly provided by Camille Locht, Institut Pasteur de Lille, France). The mature state was verified based on the expression of DC-associated maturation markers CD40, CD80, CD83, and CD86 by flowcytometry (data not shown). The large bulk of 1.2x 109 cells stimulated MUTZ-3 was washed in ice cold PBS and snap frozen before lysing and solubilizing of cell membrane proteins with Nonidet P40 containing IP lysisbuffer (Thermo Scientific). After removal of the non-solubilized fraction using ultracentrifugation, HLA class II molecules were immunoprecipitated from the MUTZ-3 cell lysate using the HLA-DR-specific monoclonal antibody L243. An aliquot of the MUTZ-3 cell lysate after HLA-DR pull down was used for proteomics. HLA class II molecules and associated peptides were eluted with 10% acetic acid and peptides were collected by passage over a 10-kDa mw cutoff membrane and concentrated using vacuum centrifugation.

MUTZ-3 cell lysate digestion

The MUTZ-3 cell lysate was diluted in 2 M urea, 50 mM ammonium bicarbonate containing one tablet of EDTA-free protease inhibitor mixture (Sigma) and one tablet of

PhosSTOP phosphatase inhibitor mixture (Roche). Cysteine residues were reduced and alkylated using 200 mM dithiotreitol (Sigma) and 200 mM iodoacetamide (Sigma). The proteins were digested with Lys-C (Roche Diagnostics) at an enzyme: protein ratio of 1:75 for 4 h at 37 °C. Two times diluted samples were digested with trypsin (Roche Diagnostics) overnight at 37°C at an enzyme: protein ratio of 1:100. Peptide mixtures were desalted using a 1-cc Sep Pack C18 columns (Waters) according manufacture's protocol.

Fractionation of HLA-DR ligands and tryptic peptides

HLA-DR eluted peptides were fractionated by strong cation exchange (SCX) chromatography (27). The system consist of a Hypercarb™ trapping column (200 µm I.D., 5 mm, 7 µm particle size, Thermo Fisher) and SCX analytical column (200 µm I.D., 12 cm, polysulfoethyl aspartamide, 5 µm, Poly LC). The peptides were separated by a linear salt gradient ramping to 500 mM KCl in 0.1M HOAc and 35% acetonitrile at a column flow rate of 2 µl/min. A total number of 26 fractions (2 min per fraction) were collected, dried down using a vacuum centrifuge and reconstituted. Based on the LC-MS/MS signal intensities of pre-analyzed sample aliquots the 10 most informative fractions were selected for analysis. Tryptic peptides from the MUTZ-3 digest were fractioned by SCX using a ZorbaxBioSCX-Series II column (0.8 mm I.D., 50 mm, 3.5 µm particle size, Agilent Technologies). A multistep gradient up to 500 mM Nacl in 0.05% formic acid 20% acetonitrile was used to separate the tryptic peptides (28). Fractions were pooled based on their UV signal intensity to a total of 10 fractions.

Reversed phase liquid chromatography and mass spectrometry

For the HLA ligands, each individual SCX fraction was analyzed in triplicate 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) with either EThcD, HCD or ETD fragmentation . The system comprises an in-house packed 20 mm x 100 µm ID trapping column (Reprosil C18, 3 µm, Dr Maisch, Ammerbuch, Germany) and a 50 cm x 50 µm ID analytical column (Poroshell 120 EC C18, 2.7 µm, Agilent Technologies) heated to 40 °C. The gradient for the separation linearly ranged from 7% to 30% of solvent B in 90 min at a flow rate of 100 nl/min. The column effluent was directly electro-sprayed into the MS using a gold-coated fused silica tapered tip of ~5 µm I.D. 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, HCD or ETD fragmentation (isolation width of 1.5 Th) at an abundance threshold of 500 counts. Fragment ions were detected in the Orbitrap analyzer at 15,000 resolution (FWHM). The automatic gain control (AGC) target in MS/MS was set to 3 × 105 for EThcD, 7 × 104 for HCD, and 1 × 105 ETD. The maximum ion accumulation time for MS scans was set to 250 ms and for MS/MS scans to 1500 ms. For EThcD, modified instrument firmware was used to allow all-ion HCD fragmentation after an initial ETD. The HCD normalized collision energy was set to 32%. The ETD reaction time was set to 50 ms and supplemental activation and charge dependent activation time was enabled. Precursor ions with unknown and +1 charge states were excluded from MS/MS analysis. Dynamic exclusion was enabled (exclusion size list 500) with a repeat count of 1 and an exclusion duration of 60 s. The SCX fractions of the tryptic digested MUTZ-3 cells were analyzed by LC-MS/MS using an Agilent 1290 Infinity System (Agilent Technologies, Waldbronn, DE) modified for nanoflow LC (passive split) connected to a TripleTOF analyzer (AB Sciex). Peptides were eluted using a similar trapping and analytical column system and LC gradient conditions as

described above. A voltage of 2.7 kV was applied to the needle. The survey scan was from 375 to 1250 m/z and the high resolution mode was utilized, reaching a resolution of up to 40,000. Tandem mass spectra were acquired in high sensitivity mode with a resolution of 20,000. The 20 most intense precursors were selected for subsequent fragmentation using an information dependent acquisition, with a minimum acquisition time of 50 ms.

Data analysis

The raw files collected from the TripleTOF were first recalibrated based on five background ions with m/z values of 391.2847, 445.12003, 51913882, 593.15761, 667.17640. The calibrated raw files were converted to mgf by the AB Sciex MS Data Converter (version 1.3 beta) program before analysis with Proteome Discoverer 1.4. RAW files acquired with the Orbitrap Elite were directly analyzed with Proteome Discoverer 1.4 software package (Thermo Fisher Scientific, Bremen, Germany) using default settings unless otherwise stated. For the EThcD and ETD spectra the non-fragment filter was added with the following settings: the precursor peak was removed within a 1 Da window, charged reduced precursors and neutral loss peaks were removed within a 0.5 Da window. MS/MS scans were searched against the human Uniprot database (2012, 20,205 entries) using the SEQUEST HT mode (Proteome Discoverer 1.4, Thermo Fisher Scientific). Precursor ion and MS/MS tolerances were set to 3 ppm and 0.02 Da, respectively. In SEQUEST, spectrum matching was set to one for c and z ions for ETD data, b and y ions for HCD and b, y, c and z for EThcD. The data were searched with no enzyme specificity, methylation and dimethylation (R, K), acetylation (N-terminus, K), cysteinylation (C), deamidation (N, Q, R) and oxidation (M) set as variable modifications. The allowed peptide length was set between 6 and 30 amino acids, the typical length distribution of HLA class II peptides. An additional search was performed with no enzyme specificity and phosphorylation (S, T, Y), deamidation (N) and oxidation (M) set as dynamic modification. PTM assignments were validated manually. The data sets were searched (separately) against the full reversed database, and the Percolator software was used to re-score and filter the peptide-to-spectrum matches (PSM) to a < 1% false discovery rate (FDR) (29). SEQUEST searching combined with Percolator is particularly useful for the analysis of EThcD data and boosts the performance of HLA ligandome identification while maintaining stringency, as validated elsewhere (30). The peptide identification list was additionally filtered for Xcorr score ≥ 1.5. The final re-filtered peptide identification list was used as the HLA-DR ligandome for further analysis. TripleTOF data files were analyzed using identical settings unless otherwise stated. Precursor ion tolerance was set to 20 ppm and the MS/MS tolerance to 0.15 Da. In SEQUEST, spectrum matching was set to 1 for y and b ions. The data were searched with specificity for trypsin and enabling 2 miss cleavages. Oxidation (M), N-terminal acetylation, phosphorylation (S, T and Y), methylation (R, K), dimethylation (R, K) were set as dynamic modification and carbamidomethylation (C) was set as a static modification. The amount of HLA-DR peptides presented at the cell surface expressed as copy number per cell were estimated based on the MS intensities provided by proteome discoverer and known amounts of the synthetic peptides angiotensin-III and oxytocin which were spiked in each fraction prior to LC-MS analysis. CELLO2GO (31) was used for protein subcellular localization prediction. The peptide binding affinities and the 9 a.a. binding core for HLA-DR10, HLA-DR11 and HLA-DR52 were predicted using the NetMHpan-3.0 algorithm (32). Peptides with a moderate to high binding affinity (IC) < 1000 nM were considered as potential binder for a particular allele. The GibbsCluster-1.0 algorithm (33)

was for simultaneous alignment and clustering of complete date set of HLA-DR-associated peptides. Gibbs clustering was performed using default settings (33), with preference of hydrophobic amino acids at position P1, the number of clusters set to 1-4, λ set to 0.8. Sequence logo's we created using IceLogo (34), with p-values set to 0.005. Following accepted procedures the protein/peptide abundances were estimated based on their spectral counts normalized to the protein length (35). The mass spectrometry data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (36) with the dataset identifier PXD002951.

3. RESULTS

Identification of naturally presented HLA-DR-peptides

HLA-DR-peptides were isolated from human matured MUTZ-3 DCs and pre-fractionated by strong cation exchange chromatography prior to analysis by reversed phase LC-MS/MS, employing the complementary peptide fragmen-

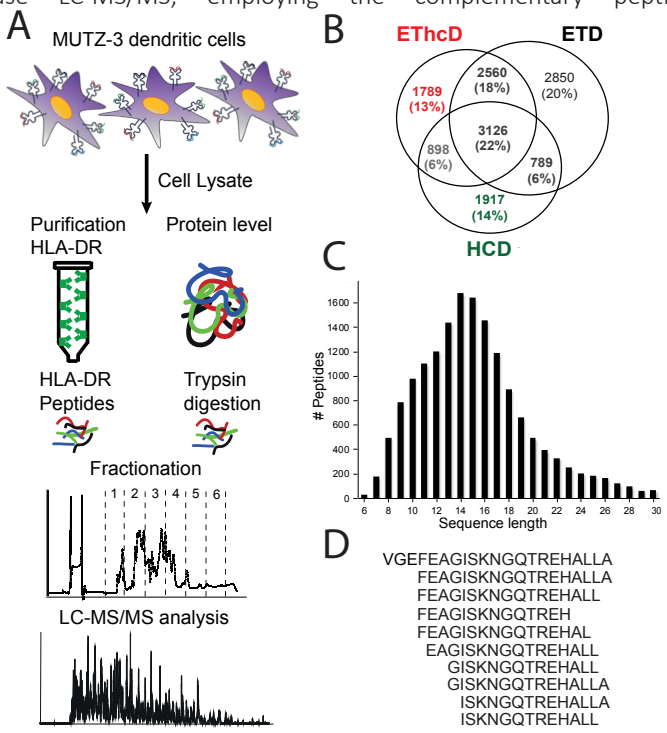


Figure.1: *HLA-DR ligandome analysis and characteristics. a) Workflow for the analysis of the HLA-DR-presented ligandome and proteome of MUTZ-3 DC. For the ligandome, following affinity purification of HLA-DR-peptide complexes, peptides were released by acid elution, fractionated by strong cation exchange (SCX) and analyzed by LC-MS/MS using ETD, EThCD and HCD fragmentation. Global proteome analysis was performed following trypsin digestion of lysed MUTZ-3 cells, using SCX fractionation prior to LC-MS/MS analysis. b) Venn diagram displaying the overlap and unique contribution to the identified HLA-DR ligandome by EThcD, ETD and HCD. c) Peptide length distribution in the combined HLA-DR ligandome (n = 13,918). d) Illustrative nested set of HLA-DR-associated peptides derived from Elongation factor 1-alpha 1, consisting of 10 peptide length variants with a consensus core binding motif but varying extended N- and C-termini.*

tation methods HCD, ETD, and EThcD (Figure.1a). Table.1 summarizes the global outcome of the HLA-DR peptide identification results (Supplementary data). Although the highest number of MS/MS events were acquired by HCD, it showed a substantial lower identification rate than ETD and EThcD. In our work by using ETD the largest number of unique peptides could be identified (9,431), followed by EThcD (8,254) and finally HCD (6,725). The combination of the three techniques was essential to expand the ligandome identification results, since only 53% of all peptides were identified by two or more fragmentation techniques (Figure.1b). The combined results yielded 13,918 unique HLA-DR-peptides originating from just under two thousand (1,980) source proteins (Table.1). To check whether peptide identification depended on the peptide's physicochemical properties, we next evaluated the performance of HCD, ETD, and EThcD in terms of peptide length and charge state. Supplementary Figure.1 shows that there were moderate differences in the physicochemical properties of the peptides identified by the applied peptide fragmentation techniques. EThcD displayed a preference for smaller (6 – 11 amino acids in length) and doubly charged peptides, while ETD displayed a bias for longer peptides with higher charge state. The inventory of 13,918 HLA-DR peptides included peptides that contained oxidized methionine, deamidated asparagine and glutamine, most likely modifications that are associated with sample preparation (613 modified peptides in total). We identified an additional fraction of 916 unique peptides bearing a variety of post translational modifications, including cysteinylation, N-terminal and lysine acetylation, serine and threonine phosphorylation and (di)methylation on arginine and lysine residues. When compared to some of the most extensive previous HLA class II studies in literature, our dataset represents a substantial improvement of the number of HLA-DR peptides identified (Supplementary Table.1). Despite the limited overlap between peptides identified, likely due to different HLA-DR backgrounds, our study covered a large fraction (~50%) of the previous identified source proteins and extensively increased the list of protein antigens presented by HLA class II molecules.

Global characteristics of the HLA-DR ligandome

The 13,918 HLA-DR peptides identified in the total dataset ranged in length from 6 up to 30 amino acids (Figure.1c), with a preference for 14-17 residues (50%) (8). The vast majority of peptides (84%) corresponded to so-called nested sets, having a common binding core extended by flanking N- and/or C-terminal residues (Table.1, Figure.1d). In total, 2,060 nested sets were observed, an average of approximately 1 nested set per protein (Table.1). The majority of nested sets were represented by 2 or 3 peptides but some consisted of more than 60 variants. The NetMHCIIpan-3.0 algorithm (32) was used to predict the 9 amino acids core sequence of the peptides that facilitates binding to the expressed HLA-DR10, HLA-DR11 and HLA-DR52 molecules. A total of 10,393 binding cores were predicted to have a moderate to strong affinity to the expressed HLA-DR molecules (< 1000 nM). These predicted binding cores were derived from only 5,795 unique peptides, indicating that for many peptide multiple core sequences were predicted to bind to different HLA-DR molecules and no single HLA-DR presenting molecule could be assigned. From the 10,393 predicted binding cores, only 1,877 ones were predicted to bind to a single HLA-DR molecule, while the remaining ones displayed unambiguity and/or redundancy having a moderate to strong predicted affinity to multiple HLA-DR molecules. These results likely indicate that the NetMHCIIpan3.0 software lacks specificity to discriminate in binding between the three investigated different HLA-DR molecules. Therefore, we decided to also make use of the Gibbs alignment and clustering algorithm (Gibbs-Cluster-1.0) (33) to extract the bind-

ing motifs directly from the complete set of HLA-DR ligands. We clustered the complete set of peptides into four groups, one for each investigated HLA-DR molecule and one non-specific cluster. The Gibbs-Cluster-1.0 could assign to the different HLA-DR clusters a total of 8,569 unique peptides, without overlap between the generated peptide groups. Figure.2 shows three sequence logo's generated by the Gibbs-Cluster-1.0 analysis (Figure.2 d, e, f), two of which (Figure.2 d, f) agree reasonably well with the sequence logo's of the NetMHCIIpan3.0 predictions for HLA-DR10 and HLA-DR52 (respectively Figure.2 a, c). The sequence logo's of the predicted binding motifs exhibited marked differences between HLA-DR10, HLA-DR11 and HLA-DR52 (Figure.2) at the P1, P4, P6 and P9 anchor residues. All three expressed HLA-DR molecules had a clear selectivity for both aliphatic and aromatic hydrophobic residues such as F, Y, L and I at one or more of the anchor

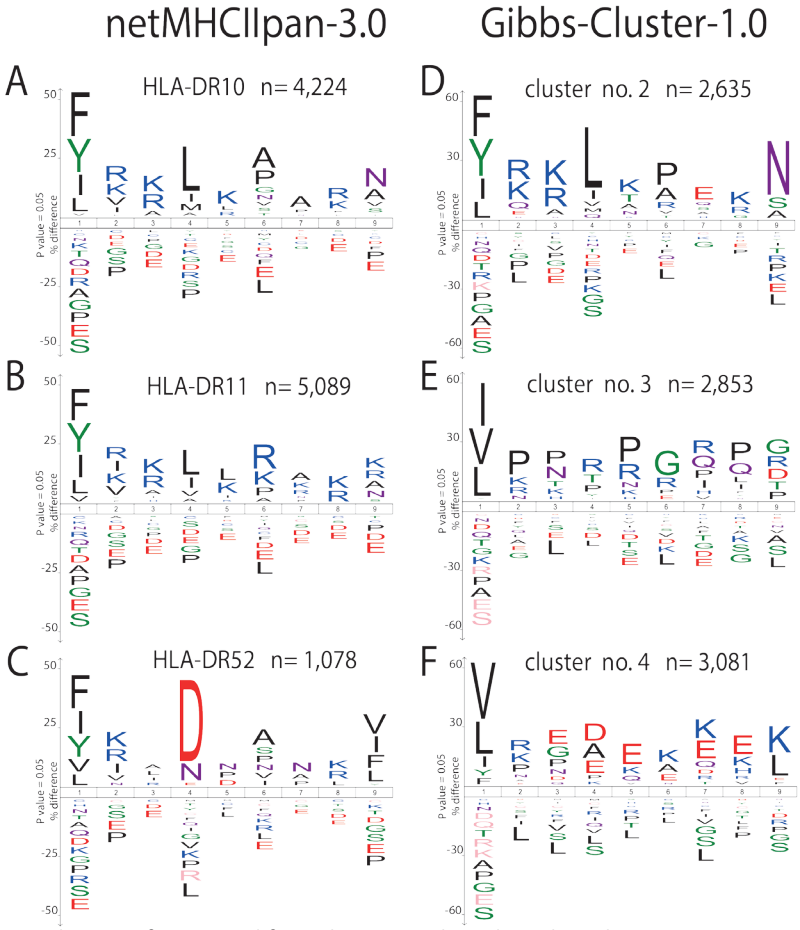


Figure.2: *Binding motifs extracted from the HLA-DR ligandome by either NetMHCIIpan3.0. and by unsupervised clustering using Gibbs-Cluster-1.0. Sequence logo's are shown for the 9 amino acid binding motif enabling binding to the expressed HLA-DR molecules. Anchor residues located on position P1, P4, P6 and P9 are known to be important for binding. (a-c) consensus binding motifs from the complete set of HLA-DR ligands for HLA-DR10 (a), HLA-DR11 (b) and HLA-DR52 (c) as extracted by using NetMHCIIpan3.0. (d-f) Unsupervised analysis by alignment and clustering analysis (using Gibbs-Cluster-1.0) reveals unique HLA-DR binding motifs. The inserts display the number of predicted 9 amino acid binding cores found that contribute to the shown sequence logo's.*

positions. Although aspartic acid (D) was found exclusively enriched at anchor position P4 for HLA-DR52, the Gibbs cluster analysis indicated that the negatively charged glutamic acid (E) was preferred on P3, P4 and P5 as well. For all three HLA-DR molecules we found that the positively charged residues K and R were moderately enriched at the positions P2, P3, P5 and P8. This observation was confirmed by the Gibbs cluster analysis.

Cleavage specificity of proteases involved in HLA-DR ligand processing

To determine which endo-lysosomal proteases may be involved in the generation of the HLA-DR ligands, we analyzed the distribution of amino acids at the cleavage sites at both the N-terminus and the C-terminus of the ligands. Figure.3 shows the over- and under-represented amino acids flanking the cleavage site both at the N-terminus (left panel) and the C-terminus (right panel). For both termini, cleavage preferentially occurs between hydrophobic and acidic residues. The high preference for L, F and D at P1 of the cleavage site hints to a dominant role of the proteases Cathepsin D or Cathepsin E (37, 38). In our subsequent global proteome analysis of MUTZ-3 cells (Figure.1a) a prominent expression of Cathepsin D compared to other proteases involved in the endo-lysosomal degradation, could be established (Supplementary Table.2). The proteomic analysis also revealed the presence of several other proteases possibly involved in the ligand processing, such as Cathepsin B, H, S, Z, albeit generally detected at lower abundance than Cathepsin D. Therefore, Cathepsin D is likely the most prominent protease involved in the processing of MHC II molecules in MUTZ-3 cells. The specificity profiles also showed amino acid residues that were under-represented at the cleavage site (Figure.3, lower parts of logo's). The small amino acids proline and serine were identified as poor substrates for the enzymes involved in antigen processing. Although proline residues were generally under-represented surrounding the cleavage sites, we found enrichment of proline residues close to the N-terminal end of HLA-DR ligands. Enrichment of proline residues located penultimate to the N-terminal cleavage has been linked to the blocking of N-terminal trimming in HLA class II peptides (39). This observation, linked to the identification of Aminopeptidase-N in both the global proteome and the HLA-DR ligandome, hints at a possible role for this enzyme in HLA class II processing in MUTZ-3 (Supplementary Table.2, Figure.4 d).

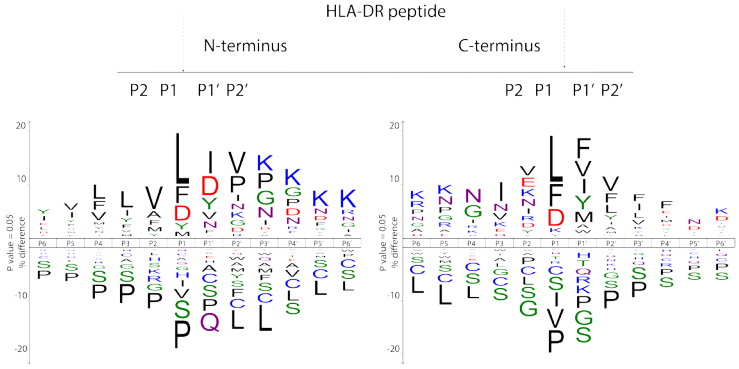


Figure.3: *Observed sequence specificity in the proteolytic processing of HLA-DR peptides. Both the N-terminal and C-terminal cleavage specificity observed in HLA-DR ligand processing are shown. The sequence logo's depict six amino acids represented as P6…P1 and six amino acids as P1'…P6', which are located at the N- (left) and C-terminal (right) scissile sites of HLA-DR-associated peptides, respectively. The residues that are statistically over-represented are shown on the upper part of the IceLogo, while under-represented at the lower part (95% confidence level).*

High dominance in the HLA-DR binding groove

The median number of HLA-DR-peptides identified per source protein was 2, but varied widely ranging from a single peptide up to 800 unique peptides (Table.1). The latter high number was found to be largely due to very extensive nested set formation. To obtain a (semi) quantitative impression of the HLA-DR sampling process, we calculated the peptide copy number per cell using known amounts of two synthetic peptides as reference which were spiked in each sample fraction prior to LC-MS analysis. These calculated copy numbers may however not accurately reflect the actual presentation levels since we do not account for losses during sample preparation and variation in MS detection (e.g. ionization efficiency). We estimated that HLA-DR peptides were presented on MUTZ-3 cells with an median of 2 copies/cell with a range between <1 and 962 copies/cell, indicating a dynamic range of four orders of magnitude in the display of potential individual CD4+ T cell epitopes (Table.1). The surface presentation levels of the source proteins were determined by summing the copy per cell numbers of their representative peptides. The cumulative peptide copy number per protein per cell ranged from <1 – 7,450 (5 orders of magnitude) (Table.1), which follows directly from the high diversity in number of peptides presented per protein (i.e. nested sets). Based on these copy numbers, we estimated that the 100 most frequently sampled proteins already occupy ~80% of the HLA-DR molecules, while the remaining 1,881 proteins account for only ~20% of the quantitative HLA-DR ligandome.

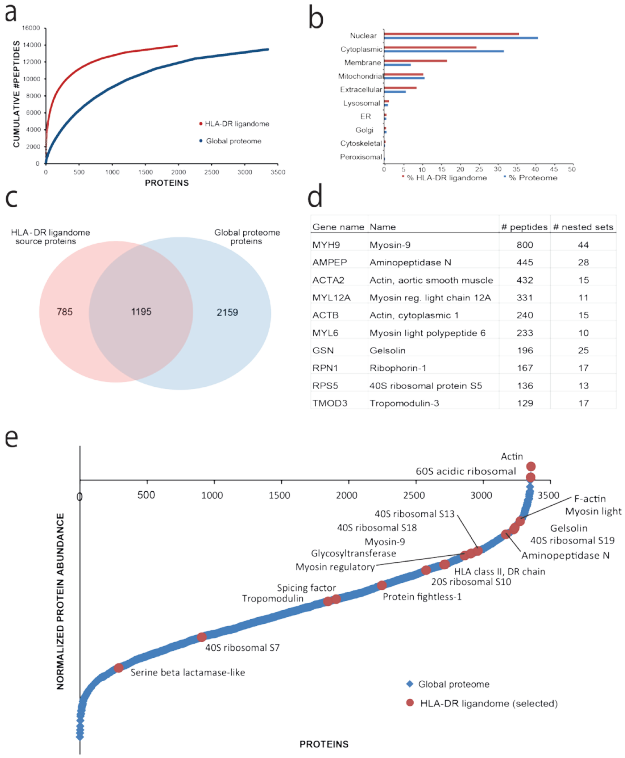
Data set	ETD	HCD	EThcD	Total
Total # of MS/MS spectra	54,059	61,129	46,351	161,539
peptide-to-spectrum-matches, PSMs (<1% FDR)	14,100	10,694	12,755	37,549
identification rate (%)	26	17	28	23
unique HLA-DR peptides	9,431	6,725	8,254	13,918
unique source proteins	1,516	1,314	1,280	1,980
Characteristics of the HLA-DR ligandome				
'nested sets'	1,472	1,066	1,237	2,060
peptides presented in nested sets	7,624	5,382	6,998	11,278
median (and range) of peptide copy numbers per cell	2 (<1 – 596)	1 (<1 – 382)	3 (<1 – 962)	2 (<1 – 962)
median (and range) of unique HLA-DR peptides per source protein	2 (1 – 612)	2 (1 – 445)	2 (1 – 598)	2 (1 - 800)
median (and range) of nested sets per source protein	1 (0 - 38)	1 (0 – 32)	1 (0 – 36)	1 (0 – 44)
median (and range) of cumulative peptide copy number per source protein per cell	4 (<1 – 4,085)	2 (<1 – 1,958)	6 (<1 – 6,610)	3 (<1 – 7,450)

Table.1: *Comprehensive HLA-DR ligandome analysis of mature MUTZ-3 DC by mass spectrometry using the complementary peptide fragmentation techniques ETD, HCD and EThcD.*

Selective sampling of the cellular proteome by HLA-DR

The most frequently HLA-DR-sampled proteins included actins, myosins, aminopeptidase N and the HLA class II β chain itself (Figure.4 d, e). These proteins are known to be involved in cytoskeletal organization and antigen processing and presentation during maturation of monocytes into dendritic cells after stimulation (40, 41). We used PANTHER (42) for Gene Ontology (GO) classification of the HLA-DR ligandome, whereby the concomitantly analyzed MUTZ-3 cellular proteome was used as a reference set (Figure.4 b).

Analysis of the GO terms of the source proteins of the HLA-DR ligandome against the background of the global MUTZ-3 proteome (rather than against the full human proteome) revealed an MHC class II pathway-specific signature in GO annotation for stimulated DCs. Compared to the cellular proteome, the HLA-DR ligandome was enriched in proteins involved in cell adhesion, immune system processes and system development (Supplementary Table.3). CELLO2GO (31) was used to predict protein cellular localization of the source proteins detected in the HLA-DR ligandome and the global cellular proteome (Figure.4 b). A major fraction of the ligandome source proteins originated from the nuclear compartment (36%), followed by the cytoplasm (25%) and the plasma membrane (17%). Compared to the MUTZ-3 global proteome, proteins contributing to the ligandome were relatively enriched in membrane associated proteins and proteins from the extracellular matrix, while proteins localized in the nucleus and cytoplasm were underrepresented. Finally, we investigated the relation between the HLA-DR ligandome and the global MUTZ-3 proteome with respect to the identified proteins and their associated number of peptides.



Gene name	Name	# peptides	# nested sets
MYH9	Myosin-9	800	44
AMPEP	Aminopeptidase N	445	28
ACTA2	Actin, aortic smooth muscle	432	15
MYL12A	Myosin reg. light chain 12A	331	11
ACTB	Actin, cytoplasmic 1	240	15
MYL6	Myosin light polypeptide 6	233	10
GSN	Gelsolin	196	25
RPN1	Ribophorin-1	167	17
RPS5	40S ribosomal protein S5	136	13
TMOD3	Tropomodulin-3	129	17

Figure.4: *Source proteome of the HLA-DR ligandome. (a) The cumulative number of unique peptides as function of the number of proteins identified by global proteome (blue) and the HLA-DR ligandome (red) analysis. Proteins are ranked from highest to the lowest number of representative peptides. (b) The CELLO2GO algorithm was used to predict the subcellular localization of source proteins obtained by HLA-DR ligandome (blue bars) and global cellular proteome (red bars) analysis. (c) Venn diagram showing the overlap between source-proteins identified in the HLA-DR ligandome and the full proteome. (d) Source proteins most frequently sampled by HLA-DR with the number of observed unique HLA-DR peptides and number of nested sets. e) Dynamic range plot of the global cellular proteome of MUTZ-3 (blue dots). The assigned proteins (red dots) were the 19 out of 25 most abundantly expressed proteins in the HLA-DR ligandome (6 proteins HLA-DR-sampled proteins were not identified by global proteome analysis).*

Global proteome analysis provided abundance estimates for about 3,300 proteins (Figure.4 a). Only 1,195 proteins (35%) were also identified as source proteins in the HLA-DR ligandome (Figure.4 c). Moreover, we observed that certain source proteins that were represented by a high number of HLA-DR peptides were absent in the global cellular analysis of MUTZ-3 cells, and vice versa, which could point to a rather weak correlation between expression levels of the HLA-DR ligandome and global proteome. To illustrate this, we determined the correlation between the proteins most frequently sampled by HLA-DR and their cellular expression levels (global proteome analysis) using a recently described peptide count approach (41). Figure.4 e shows that the most abundantly observed proteins in the HLA-DR ligandome are distributed over the full dynamic range plot of the source proteome.

4. DISCUSSION

Natural processing and Human Leukocyte Antigen (HLA) class II presentation of extracellular and intracellular proteins are key yet still unpredictable functions of activated dendritic cells (DC) that steer and regulate adaptive immune responses. Advances in understanding the outcome of these processes require comprehensive analysis of endogenously processed and surface presented peptide repertoires, demanding cutting edge technology (43) To expand the mass spectrometry-based identification of HLA class II presented peptides we combined the (hybrid) peptide fragmentation techniques, HCD, ETD, and EThcD, on a large scale preparation of HLA-DR peptides derived from an in vitro matured human prototype monocyte-derived DC cell line, MUTZ-3. Cumulatively, the exceptionally high number of 13,918 unique identified HLA-DR peptides, a significant improvement in the number of ligands compared to previous endeavors (18, 22, 44, 45), gave unprecedented insight into various underlying features of selectivity in HLA-DR sampling of the ligandome. MS-based identification of HLA class II-associated peptides by non-standard ETD or EThcD based sequencing was until now little explored. Our data revealed an improved rate of peptide identification for techniques involving electron transfer dissociation over classical collision induced dissociation for HLA-DR peptides, confirming previous reports on the analysis of endogenous peptides (46, 47). HLA-DR-presented peptides were highly variable in nature, especially in their wide distribution in peptide length and charge state. ETD had a slight preferential bias for longer and/or higher charged HLA-DR-peptides (48, 49). Earlier, using identical non-standard MS technology in an HLA class I peptide inventory (50) we identified 12,199 unique HLA class I peptides originating from 5,603 proteins in a human B lymphoblastoid cell line (i.e. on average ~2 peptides per protein). In our present comprehensive survey addressing the HLA class II pathway we identified 13,918 HLA-DR ligands derived from only 1,980 proteins (i.e. on average ~7 peptides per protein). This much more sparse source proteome, together with the observation of many nested sets of length variants of the same epitope region, suggests that the overall HLA-DR epitope landscape available for surveillance by circulating CD4+ T cell populations is less diverse than the HLA class I ligandome inspected by CD8+ T cells. The source proteins sampled by HLA-DR molecules reflected only a minor part (~35% overlap) of the global proteome identified in the MUTZ-3 cell lysate. Vice versa, some most frequently sampled proteins in the HLA-DR ligandome were absent from the detectable MUTZ-3 proteome, which rather suggests a poor correlation between expression levels in the global proteome and protein representation in the HLA-DR ligandome. For the intrinsically different HLA class I antigen presentation route, multiple reports have established that

it is difficult to predict ligand copy numbers from overall protein or RNA levels (21, 51–53), perhaps with the exception of proteins with high turn-over (35). Therefore, most studies including ours substantiate the selective sampling of proteome cargo for both HLA class I and class II molecules on their way to the cell surface. Indeed, GO analysis revealed that the surface-presented HLA-DR ligandome of MUTZ-3 DCs was biased for proteins involved in cell mobility, cytoskeletal re-organization, HLA class II processing and presentation, functions strongly linked to the maturation of monocytes into antigen presenting cells (40, 41). The MUTZ-3 DCs in our study were matured via TNFα, an endogenous 'danger' signal that can induce the transition from an antigen capturing DC to an antigen presenting DC (54). Relative enrichment was also found for proteins associated with the membrane and extracellular matrix, the main contributors to the lysosomal degradation and HLA class II presentation pathway (55). Direct or indirect evidence for the dominance of membrane and extracellular proteins as source proteins for MHC class II ligandomes has been implied before (20, 56–58), in line with the mechanisms of endocytosis, lysosomal trafficking and processing, preceding MHC class II loading with peptides. Even while being underrepresented relative to the global MUTZ-3 proteome, proteins associated with the nuclear compartment or cytoplasm made up to 50% of the source proteins of the HLA-DR ligandome. We interpret this to be the outcome of autophagy, a stress response process shuttling large fractions of intracellular proteins into the lysosomal compartment promoting their presentation by HLA class II (59). Earlier autophagy has been implied to be essential for HLA-DR loading with intracellular material relevant for CD4+ T cells to monitor cellular virus infection (60, 61), transformation (62–64) or stress (65), or to shape their self-tolerance (10, 66, 67). The observation that the most abundantly sampled source proteins occupied with their peptide sequences already ~80% of the HLA-DR molecules, further indicates that in the antigen selection process certain proteins have strong advantage of being presented over others. The dynamic range of HLA-DR ligandome ranged from less than a single peptide copy per protein to extremes where cumulatively almost 7,500 peptide copies per protein were detected per cell. Strong peptide hierarchies have been seen in the HLA class I ligandome (68) and human and murine MHC class II ligandomes of DC (69), and are essential for the recruitment and outcome of T cell responses (70, 71). Our study substantiated the biochemical evidence that, apart from the cell biology of a particular type of antigen presenting cell in bringing together proteome and enzyme content in the MHC class II pathway, physical-chemical properties of protein sequences themselves play a dominant role in the high selectivity of proteome representation in HLA-DR. First, peptide sequences have to be generated but not destroyed by the proteolytic activity in a complex network of endo-lysosomal compartments of a particular cell (71, 72). Analysis of proteolytic sites located at the peptide N- and C-terminus revealed that processing of HLA-DR ligands mainly involved cleavage between hydrophobic and/or acidic residues, corresponding to the substrate specificity of Cathepsin D (37, 38). Cathepsin D is known to be highly expressed is lysosomes, but how exactly this enzyme contributes to HLA class II peptide processing and presentation remains ambivalent as it can both generate and destroy T cell epitopes. Whether the here-observed dominant role of Cathepsin D in antigen processing is specifically associated with the TNFα maturation of the DCs needs further research. In another study Cathepsin S was reported to be more involved in the thymic HLA-DR peptidome (20). Second, the created ligands need to meet any of the binding motifs of the HLA-DR allomorphs expressed by the cell in order to be presented. We complemented NetMHCII-pan3.0 predictions, that may struggle with allocating 9-mer binding cores in large sets

of length variants, with Gibbs alignment and clustering analysis (Gibbs-Cluster-1.0) (33). With this latter algorithm, binding motifs were unbiasedly extracted from the complete set of HLA-DR-peptides, indicating the strong selectivity of the expressed HLA-DR molecules to bind peptides mainly through hydrophobic and aliphatic anchor residues. We also found an enrichment of positively charged residues (K, R) on non-anchor positions (P2, P3, P5 and P8). Selectivity for arginine on the P8 positions of HLA-DR10-associated peptides has been reported previously (45). Positively charged amino acids flanking the anchor residues could play an important role in the interaction of the HLA-DR-peptide complex with the T-cell receptor. Novellino et al. (73) showed that a melanoma-specific T-cell clone that interacts with the HLA-DR10-restricted peptide YFAAELPPRN lost recognition when arginine was substituted by glycine. Likely, the here ETD and EThcD have contributed to the unambiguous assignment of these positively charged amino acids as these fragmentation techniques are specifically beneficial for peptides harboring internal basic residues (49, 74). In conclusion, complementary peptide fragmentation techniques allow the in depth profiling of human DC derived HLA class II-associated peptides. Our results, based on a single MUTZ-3 HLA-DR ligandome, pave the way for further exploration of HLA class II peptide repertoires under different maturation and antigenic conditions. Exposure of peptides at a wide dynamic range of abundances yet with a relatively limited diversity through strong selective mechanisms seem hallmarks of this pathway. Whether these here-observed characteristics of antigen processing and presentation are widely applicable to the HLA class II pathway in human DCs needs further investigation of various DC models and primary cell cultures. The challenge in future studies will be to quantitatively and qualitatively characterize the full breadth of potential CD4+ T cell epitopes relevant in health and disease, especially if hidden as minor specificities among the larger self ligandome.

5. ACKNOWLEDGEMENTS

This work was partly supported by the project Proteins At Work (project 184.032.201), a program of the Netherlands Proteomics Centre financed by the Netherlands Organization for Scientific Research (NWO) as part of the National Roadmap Large-scale Research Facilities of the Netherlands, and by the strategic research projects Immunoproteomics and Correlates of Protection, financed by the Dutch Ministry of Health. This work was also supported by the Institute for Chemical Immunology, an NWO Gravitation project funded by the Ministry of Education, Culture and Science of the Netherlands. The NWO is also kindly acknowledged for financial support with the VIDI grant for Shabaz Mohammed (700.10.429).

6. REFERENCES

1. Neefjes, J., Jongsma, M. L., Paul, P., and Bakke, O. (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol 11, 823–836
2. Roche, P. A., and Furuta, K. (2015) The ins and outs of MHC class II-mediated antigen processing and presentation. Nat Rev Immunol 15, 203–216
3. Adamopoulou, E., Tenzer, S., Hillen, N., Klug, P., Rota, I. a, Tietz, S., Gebhardt, M., Stevanovic, S., Schild, H., Tolosa, E., Melms, A., and Stoeckle, C. (2013) Exploring the MHC-peptide matrix of central tolerance in the human thymus. Nat. Commun. 4, 2039
4. Suri, A., Lovitch, S. B., and Unanue, E. R. (2006) The wide diversity and complexity of peptides bound to class II MHC molecules. Curr Opin Immunol 18, 70–77
5. Ovsyannikova, I. G., Johnson, K. L., Bergen, H. R., and Poland, G. A. (2007) Mass spectrometry and peptide-based

vaccine development. Clin Pharmacol Ther 82, 644–652
6. Thibodeau, J., Bourgeois-Daigneault, M. C., and Lapointe, R. (2012) Targeting the MHC Class II antigen presentation pathway in cancer immunotherapy. Oncoimmunology 1, 908–916
7. Purcell, A. W. (2004) Isolation and characterization of naturally processed MHC-bound peptides from the surface of antigen-presenting cells. Methods Mol Biol 251, 291–306
8. Lippolis, J. D., White, F. M., Marto, J. A., Luckey, C. J., Bullock, T. N., Shabanowitz, J., Hunt, D. F., and Engelhard, V. H. (2002) Analysis of MHC class II antigen processing by quantitation of peptides that constitute nested sets. J Immunol 169, 5089–5097
9. Dengjel, J., Schoor, O., Fischer, R., Reich, M., Kraus, M., Müller, M., Kreymborg, K., Altenberend, F., Brandenburg, J., Kalbacher, H., Brock, R., Driessen, C., Rammensee, H.-G., and Stevanovic, S. (2005) Autophagy promotes MHC class II presentation of peptides from intracellular source proteins. Proc. Natl. Acad. Sci. U. S. A. 102, 7922–7927
10. Münz, C. (2012) Antigen processing for MHC class II presentation via autophagy. Front. Immunol. 3, 1–6
11. Deretic, V., Saitoh, T., and Akira, S. (2013) Autophagy in infection, inflammation and immunity. Nat Rev Immunol 13, 722–737
12. Stenger, R. M., Meiring, H. D., Kuipers, B., Poelen, M., van Gaans-van den Brink, J. A., Boog, C. J., de Jong, A. P., and van Els, C. A. (2014) Bordetella pertussis proteins dominating the major histocompatibility complex class II-presented epitope repertoire in human monocyte-derived dendritic cells. Clin Vaccine Immunol 21, 641–650
13. Bergseng, E., Dørum, S., Arntzen, M. Ø., Nielsen, M., Nygård, S., Buus, S., de Souza, G. a., and Sollid, L. M. (2014) Different binding motifs of the celiac disease-associated HLA molecules DQ2.5, DQ2.2, and DQ7.5 revealed by relative quantitative proteomics of endogenous peptide repertoires. Immunogenetics 67, 73–84
14. Depontieu, F. R., Qian, J., Zarling, A. L., McMiller, T. L., Salay, T. M., Norris, A., English, A. M., Shabanowitz, J., Engelhard, V. H., Hunt, D. F., and Topalian, S. L. (2009) Identification of tumor-associated, MHC class II-restricted phosphopeptides as targets for immunotherapy. Proc Natl Acad Sci U S A 106, 12073–12078
15. Benham, H., Nel, H. J., Law, S. C., Mehdi, A. M., Street, S., Ramnoruth, N., Pahau, H., Lee, B. T., Ng, J., G Brunck, M. E., Hyde, C., Trouw, L. A., Dudek, N. L., Purcell, A. W., O'Sullivan, B. J., Connolly, J. E., Paul, S. K., Lê Cao, K. A., and Thomas, R. (2015) Citrullinated peptide dendritic cell immunotherapy in HLA risk genotype-positive rheumatoid arthritis patients. Sci Transl Med 7, 290ra87
16. Fissolo, N., Haag, S., de Graaf, K. L., Drews, O., Stevanovic, S., Rammensee, H. G., and Weissert, R. (2009) Naturally presented peptides on major histocompatibility complex I and II molecules eluted from central nervous system of multiple sclerosis patients. Mol Cell Proteomics 8, 2090–2101
17. Hill, J. a, Southwood, S., Sette, A., Jevnikar, A. M., Bell, D. a, and Cairns, E. (2003) Cutting edge: the conversion of arginine to citrulline allows for a high-affinity peptide interaction with the rheumatoid arthritis-associated HLA-DRB1*0401 MHC class II molecule. J. Immunol. 171, 538–541
18. Chornoguz, O., Gapeev, a., O'Neill, M. C., and Ostrand-Rosenberg, S. (2012) Major Histocompatibility Complex Class II+ Invariant Chain Negative Breast Cancer Cells Present Unique Peptides That Activate Tumor-specific T Cells From Breast Cancer Patients. Mol. Cell. Proteomics, 1457–1467
19. Dengjel, J., Decker, P., Schoor, O., Altenberend, F., Weinschenk, T., Rammensee, H. G., and Stevanovic, S. (2004) Identification of a naturally processed cyclin D1 T-helper epitope by a novel combination of HLA class II targeting and differential mass spectrometry. Eur J Immunol 34, 3644–3651
20. Collado, J. A., Alvarez, I., Ciudad, M. T., Espinosa, G., Canals, F., Pujol-Borrell, R., Carrascal, M., Abian, J., and Jaraquemada, D. (2013) Composition of the HLA-DR-associated human thymus peptidome. Eur J Immunol 43, 2273–2282
21. Granados, D. P., Laumont, C. M., Thibault, P., and Perreault, C. (2015) The nature of self for T cells—a systems-level perspective. Curr. Opin. Immunol. 34, 1–8
22. van Haren, S. D., Herczenik, E., ten Brinke, A., Mertens, K., Voorberg, J., and Meijer, A. B. (2011) HLA-DR-presented peptide repertoires derived from human monocyte-derived dendritic cells pulsed with blood coagulation factor VIII. Mol Cell Proteomics 10, M110.002246
23. Masterson, A. J., Sombroek, C. C., de Gruijl, T. D., Graus, Y. M. F., van der Vliet, H. J. J., Lougheed, S. M., van den Eertwegh, A. J. M., Pinedo, H. M., and Scheper, R. J. (2002) MUTZ-3, a human cell line model for the cytokine-induced differentiation of dendritic cells from CD34+precursors. Blood 100 , 701–703
24. Hoefnagel, M. H., Vermeulen, J. P., Scheper, R. J., and Vandebriel, R. J. (2011) Response of MUTZ-3 dendritic cells to the different components of the Haemophilus influenzae type B conjugate vaccine: towards an in vitro assay for vaccine immunogenicity. Vaccine 29, 5114–5121
25. Hoonakker, M. E., Verhagen, L. M., Hendriksen, C. F., van Els, C. A., Vandebriel, R. J., Sloots, A., and Han, W. G. (2015) In vitro innate immune cell based models to assess whole cell Bordetella pertussis vaccine quality: a proof of principle. Biologicals 43, 100–109
26. Frese, C. K., Altelaar, A. F., van den Toorn, H., Nolting, D., Griep-Raming, J., Heck, A. J., and Mohammed, S. (2012) Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. Anal Chem 84, 9668–9673

27. Meiring, H. D., Soethout, E. C., Poelen, M. C. M., Mooibroek, D., Hoogerbrugge, R., Timmermans, H., Boog, C. J., Heck, A. J. R., de Jong, A. P. J. M., and van Els, C. a C. M. (2006) 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 5, 902–913
28. Meiring, H. D., Soethout, E. C., de Jong, A. P., and van Els, C. A. (2007) Targeted identification of infection-related HLA class I-presented epitopes by stable isotope tagging of epitopes (SITE). Curr Protoc Immunol Chapter 16, Unit 16.3
29. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., and MacCoss, M. J. (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods 4, 923–925
30. Mommen, G. P., Frese, C. K., Meiring, H. D., van Gaans-van den Brink, J., de Jong, A. P., van Els, C. A., and Heck, A. J. (2014) Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proc Natl Acad Sci U S A 111, 4507–4512
31. Yu, C. S., Cheng, C. W., Su, W. C., Chang, K. C., Huang, S. W., Hwang, J. K., and Lu, C. H. (2014) CELLO2GO: a web server for protein subCELlular LOcalization prediction with functional gene ontology annotation. PLoS One 9, e99368
32. Karosiene, E., Rasmussen, M., Blicher, T., Lund, O., Buus, S., and Nielsen, M. (2013) NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. Immunogenetics 65, 711–724
33. Andreatta, M., Lund, O., and Nielsen, M. (2013) Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach. Bioinformatics 29, 8–14
34. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009) Improved visualization of protein consensus sequences by iceLogo. Nat Methods 6, 786–787
35. Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L. J., and Mann, M. (2015) Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. Mol Cell Proteomics 14, 658–673
36. Vizcaíno, J. A., Côté, R. G., Csordas, A., Dianes, J. A., Fabregat, A., Foster, J. M., Griss, J., Alpi, E., Birim, M., Contell, J., O'Kelly, G., Schoenegger, A., Ovelleiro, D., Pérez-Riverol, Y., Reisinger, F., Ríos, D., Wang, R., and Hermjakob, H. (2013) The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res 41, D1063–9
37. Impens, F., Colaert, N., Helsens, K., Ghesquière, B., Timmerman, E., De Bock, P. J., Chain, B. M., Vandekerckhove, J., and Gevaert, K. (2010) A quantitative proteomics design for systematic identification of protease cleavage events. Mol Cell Proteomics 9, 2327–2333
38. Sun, H., Lou, X., Shan, Q., Zhang, J., Zhu, X., Wang, Y., Xie, Y., Xu, N., and Liu, S. (2013) Proteolytic characteristics of cathepsin D related to the recognition and cleavage of its target proteins. PLoS One 8, e65733
39. Nelson, C. A., Vidavsky, I., Viner, N. J., Gross, M. L., and Unanue, E. R. (1997) Amino-terminal trimming of peptides for presentation on major histocompatibility complex class II molecules. Proc Natl Acad Sci U S A 94, 628–633
40. Larsen, S. L., Pedersen, L. O., Buus, S., and Stryhn, A. (1996) T cell responses affected by aminopeptidase N (CD13)-mediated trimming of major histocompatibility complex class II-bound peptides. J Exp Med 184, 183–189
41. West, M. A., Wallin, R. P., Matthews, S. P., Svensson, H. G., Zaru, R., Ljunggren, H. G., Prescott, A. R., and Watts, C. (2004) Enhanced dendritic cell antigen capture via toll-like receptor-induced actin remodeling. Science (80-.). 305, 1153–1157
42. Mi, H., Muruganujan, A., Casagrande, J. T., and Thomas, P. D. (2013) Large-scale gene function analysis with the PANTHER classification system. Nat Protoc 8, 1551–1566
43. Caron, E., Espona, L., Kowalewski, D. J., Schuster, H., Ternette, N., Alpízar, A., Schittenhelm, R. B., Ramarathinam, S. H., Lindestam Arlehamn, C. S., Chiek Koh, C., Gillet, L. C., Rabsteyn, A., Navarro, P., Kim, S., Lam, H., Sturm, T., Marcilla, M., Sette, A., Campbell, D. S., Deutsch, E. W., Moritz, R. L., Purcell, A. W., Rammensee, H.-G., Stevanovic, S., and Aebersold, R. (2015) An open-source computational and data resource to analyze digital maps of immunopeptidomes. Elife 4, 1–17
44. Seward, R. J., Drouin, E. E., Steere, A. C., and Costello, C. E. (2011) Peptides presented by HLA-DR molecules in synovia of patients with rheumatoid arthritis or antibiotic-refractory Lyme arthritis. Mol. Cell. Proteomics 10, M110.002477
45. Alvarez, I., Collado, J., Daura, X., Colomé, N., Rodríguez-García, M., Gallart, T., Canals, F., and Jaraquemada, D. (2008) The rheumatoid arthritis-associated allele HLA-DR10 (DRB1*1001) shares part of its repertoire with HLA-DR1 (DRB1*0101) and HLA-DR4 (DRB*0401). Arthritis Rheum. 58, 1630–1639
46. Frese, C. K., Boender, A. J., Mohammed, S., Heck, A. J., Adan, R. A., and Altelaar, A. F. (2013) Profiling of diet-induced neuropeptide changes in rat brain by quantitative mass spectrometry. Anal Chem 85, 4594–4604
47. Sasaki, K., Osaki, T., and Minamino, N. (2013) Large-scale identification of endogenous secretory peptides using electron transfer dissociation mass spectrometry. Mol. Cell. Proteomics 12, 700–9
48. Frese, C. K., Altelaar, A. F., Hennrich, M. L., Nolting, D., Zeller, M., Griep-Raming, J., Heck, A. J., and Mohammed,

S. (2011) Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. J Proteome Res 10, 2377–2388
49. Mommen, G. P. M., Frese, C. K., Meiring, H. D., van Gaans-van den Brink, J., de Jong, A. P. J. M., van Els, C. a C. M., and Heck, A. J. R. (2014) Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proc. Natl. Acad. Sci. U. S. A. 111, 4507–12
50. Marino, F., Bern, M., Mommen, G. P. M., Leney, A. C., van Gaans-van den Brink, J. a. M., Bonvin, A. M. J. J., Becker, C., van Els, C. a. C. M., and Heck, A. J. R. (2015) Extended O-GlcNAc on HLA Class-I-Bound Peptides. J. Am. Chem. Soc., 150819112655001
51. Milner, E., Barnea, E., Beer, I., and Admon, A. (2006) The Turnover Kinetics of Major Histocompatibility Complex Peptides of Human Cancer Cells. Mol. Cell. Proteomics, 357–365
52. Weinzierl, A. O., Lemmel, C., Schoor, O., Muller, M., Kruger, T., Wernet, D., Hennenlotter, J., Stenzl, A., Klingel, K., Rammensee, H.-G., and Stevanovic, S. (2006) Distorted Relation between mRNA Copy Number and Corresponding Major Histocompatibility Complex Ligand Density on the Cell Surface. Mol. Cell. Proteomics 6, 102–113
53. Fortier, M.-H., Caron, É., Hardy, M.-P., Voisin, G., Lemieux, S., Perreault, C., and Thibault, P. (2008) The MHC class I peptide repertoire is molded by the transcriptome. J. Exp. Med. 205, 595–610
54. Blanco, P., Palucka, A. K., Pascual, V., and Banchereau, J. (2008) Dendritic cells and cytokines in human inflammatory and autoimmune diseases. Cytokine Growth Factor Rev 19, 41–52
55. Watts, C. (2012) The endosome-lysosome pathway and information generation in the immune system. Biochim. Biophys. Acta- Proteins Proteomics 1824, 14–21
56. Chicz, R. M., Urban, R. G., Gorga, J. C., Vignali, D. A., Lane, W. S., and Strominger, J. L. (1993) Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. J Exp Med 178, 27–47
57. Rotzschke, O., and Falk, K. (1994) Origin, structure and motifs of naturally processed MHC class II ligands. Curr Opin Immunol 6, 45–51.
58. Juncker, A. S., Larsen, M. V, Weinhold, N., Nielsen, M., Brunak, S., and Lund, O. (2009) Systematic characterisation of cellular localisation and expression profiles of proteins containing MHC ligands. PLoS One 4, e7448
59. Schmid, D., Dengjel, J., Schoor, O., Stevanovic, S., and Münz, C. (2006) Autophagy in innate and adaptive immunity against intracellular pathogens. J Mol Med 84, 194–202
60. Jaraquemada, D., Marti, M., and Long, E. O. (1990) An endogenous processing pathway in vaccinia virus-infected cells for presentation of cytoplasmic antigens to class II-restricted T cells. J Exp Med 172, 947–954
61. Jacobson, S., Sekaly, R. P., Jacobson, C. L., McFarland, H. F., and Long, E. O. (1989) HLA class II-restricted presentation of cytoplasmic measles virus antigens to cytotoxic T cells. J Virol 63, 1756–1762
62. Li, Y., Wang, L. X., Yang, G., Hao, F., Urba, W. J., and Hu, H. M. (2008) Efficient cross-presentation depends on autophagy in tumor cells. Cancer Res 68, 6889–6895
63. Li, Y., Wang, L. X., Pang, P., Cui, Z., Aung, S., Haley, D., Fox, B. A., Urba, W. J., and Hu, H. M. (2011) Tumor-derived autophagosome vaccine: mechanism of cross-presentation and therapeutic efficacy. Clin Cancer Res 17, 7047–7057
64. Wang, R. F., Wang, X., Atwood, A. C., Topalian, S. L., and Rosenberg, S. A. (1999) Cloning genes encoding MHC class II-restricted antigens: mutated CDC27 as a tumor antigen. Science (80-.). 284, 1351–1354
65. Dengjel, J., Schoor, O., Fischer, R., Reich, M., Kraus, M., Müller, M., Kreymborg, K., Altenberend, F., Brandenburg, J., Kalbacher, H., Brock, R., Driessen, C., Rammensee, H. G., and Stevanovic, S. (2005) Autophagy promotes MHC class II presentation of peptides from intracellular source proteins. Proc Natl Acad Sci U S A 102, 7922–7927
66. Münz, C. (2009) Enhancing immunity through autophagy. Annu. Rev. Immunol. 27, 423–449
67. Crotzer, V. L., and Blum, J. S. (2009) Autophagy and its role in MHC-mediated antigen presentation. J Immunol 182, 3335–3341
68. Herberts, C. a., Meiring, H. M., Van Gaans-van den Brink, J. a M., Van der Heeft, E., Poelen, M. C. M., Boog, C. J. P., De Jong, A. P. J. M., and Van Els, C. a C. M. (2003) Dynamics of measles virus protein expression are reflected in the MHC class I epitope display. Mol. Immunol. 39, 567–575
69. Bozzacco, L., Yu, H., Zebroski, H. A., Dengjel, J., Deng, H., Mojsov, S., and Steinman, R. M. (2011) Mass spectrometry analysis and quantitation of peptides presented on the MHC II molecules of mouse spleen dendritic cells. J Proteome Res 10, 5016–5030
70. Yewdell, J. W. (2006) Confronting complexity: real-world immunodominance in antiviral CD8+ T cell responses. Immunity 25, 533–543
71. Blum, J. S., Wearsch, P. a, and Cresswell, P. (2013) Pathways of Antigen Processing
72. Villadangos, J. A., Schnorrer, P., and Wilson, N. S. (2005) Control of MHC class II antigen presentation in dendritic cells: a balance between creative and destructive forces. Immunol Rev 207, 191–205
73. Novellino, L., Renkvist, N., Rini, F., Mazzocchi, A., Rivoltini, L., Greco, A., Deho, P., Squarcina, P., Robbins, P. F., Parmiani, G., and Castelli, C. (2003) Identification of a mutated receptor-like protein tyrosine phosphatase kappa as a novel, class II HLA-restricted melanoma antigen. J. Immunol. 170, 6363–6370
74. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) Peptide and protein sequence

analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A 101, 9528–9533

7. SUPPLEMENTARY

Cells	Number of cells	HLA typing	# HLA-DR ligands (overlap with this study)	# source proteins (overlap with this study)	References
Awells (B cells)	1x10E9 -5.7x10E10	HLA-DRB1*0401, -DRB4*0101	404 (11)	123 (62)	Dengjel J, et al, 2005, Proc Natl Acad Sci U S A. 102; 7922-7
Tumor tissue (kidney) 2 donors	1x10E6	HLA-DRB1*11	452[1] (27)	144[1] (58)	Dengjel J, et al 2006, Clin Cancer Res. 12; 4163-70
		HLA-DRB1*15			
T cell, B cell clone from 1 donor	1x10E7-1x10E8	HLA-DRB1*1301, -DRB1*1501, -DRB3*0202, -DRB5*0101	537[1] (2)	327[1] (110)	Costantino CM, et al, 2012, PLoS One.7(1):e29805
Thymic DCs 3 donors2	0.3-6.6x10E7	HLA-DRB1*04, -DRB1*12, -DRB3, -DRB4	175[1] (20)	76[1] (50)	Adamopoulou E, et al, 2013, Nature comm. 4:2039
		HLA-DRB1*03, -DRB1*11			
		HLA-DRB1*0103, -DRB1*15, -DRB5			
MUTZ-3	1.2x10E9	HLA-DRB1*10, -B1*11, -DRB3*01	13,918	1,980	This study

Supplementary Table.1: *Comparison of our data to previously reported HLA-DR-associated ligandomes.*

[1] Published peptide and protein identification data of multiple donors/cell types were combined

[2] Data is shown for the donors were HLA-DR peptides were listed

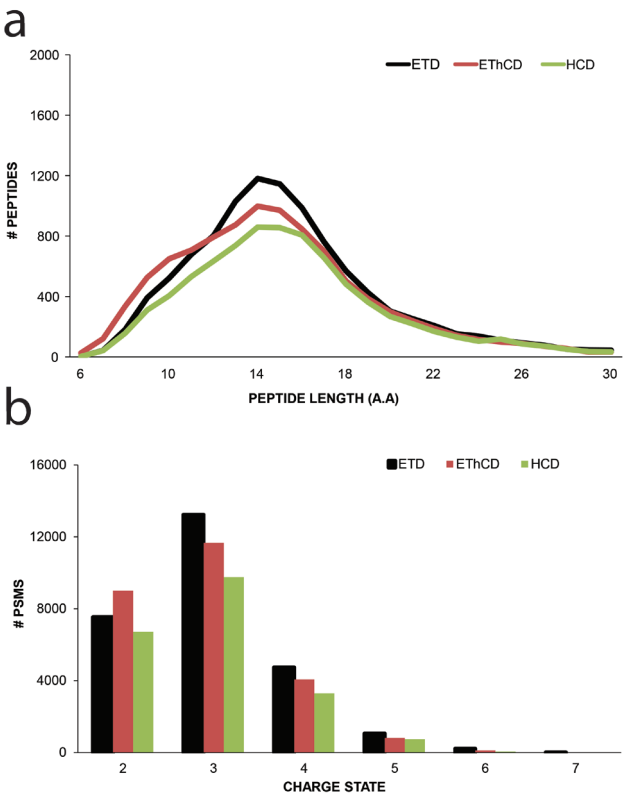
Accession	Description	Coverage (%)	Unique Peptides	PSMs[1]	Protein length	Log_{10} spectral count
P15144	Aminopeptidase N	29.8	26	156	967	-0.81
P07339	Cathepsin D	19.7	7	23	412	-1.25
P09668	Pro-cathepsin H	7.8	2	20	335	-1.26
P55786	Puromycin-sensitive aminopeptidase	15.6	13	48	919	-1.3
P42785	Lysosomal Pro-X carboxypeptidase	5.2	2	16	496	-1.49
Q9UHL4	Dipeptidyl peptidase 2	16.1	6	14	492	-1.54
Q9UBR2	Cathepsin Z	7.6	2	9	303	-1.57
P07858	Cathepsin B	5.3	2	6	339	-1.83
P25774	Cathepsin S	16.3	4	6	331	-1.91
O14773	Tripeptidyl-peptidase 1	9.4	4	5	563	-2.05

Supplementary Table.2: *List of proteases associated with lysosomal compartments identified by global proteome analysis of MUTZ-3 cell lysate*

[1] PSMs, peptide-to-spectrum matches

Biological Process	P-value
cell-cell adhesion	6.42E-07
system development	6.68E-07
cell adhesion	4.92E-05
biological adhesion	4.92E-05
metabolic process	6.55E-05
immune system process	3.06E-04
system process	5.72E-04
mesoderm development	6.78E-04
cell-matrix adhesion	8.09E-04
multicellular organismal process	1.26E-03
single-multicellular organism process	1.26E-03
cellular amino acid metabolic process	4.83E-03
ion transport	8.19E-03
biological regulation	1.28E-02
primary metabolic process	1.39E-02
fertilization	3.11E-02
sensory perception	4.77E-02

Supplementary Table.3: *Biological processes enriched in the HLA-DR ligandome when analyzed against the MUTZ-3 global proteome.*



Supplementary Figure.1: *Peptide length and charge state distribution. a) Shown are the peptide length distributions of the HLA-DR peptides identified by ETD (black), EThCD (red) and HCD (green). b) Charge state distributions of the identified HLA-DR peptides identified by the employed fragmentation techniques.*

CHAPTER 8

SUMMARY

SAMENVATTING

CURRICULUM VITAE

OUTLOOK

PUBLICATIONS

ACKNOWLEDGEMENTS

SUMMARY

The work, described in this thesis, has been aimed at the improvement of proteomic workflows; from the sample preparation, to the optimization of state-of-art chromatographical separations and alternative MS fragmentation techniques. An important focus of my thesis is represented by the analysis of peptides harboring various post translational modifications (PTMs) such as glycosylation, phosphorylation, methylation and di-methylation. PTM modified peptides often require customized approaches in order to improve their detection. To reach these goals, I show how the improvement of MS and chromatographical methods can benefit PTMs identification. Furthermore I show how these advancements can benefit research areas, namely to enhance the detection of endogenous peptides eluted from HLA class I and II molecules, which present exotic characteristics when compared to 'standard' tryptic peptides obtained in proteomic workflows. In chapter 2 I discuss the development and implementation of an online two-dimensional SCX/RP UHPLC system coupled to a fast sequencing and high resolution MS. In this work I initially characterized the one-dimensional UHPLC-MS/MS workflow. We then asked the question whether a fast and sensitive online 2D SCX-RP UHPLC-MS/MS workflow could compete with a 1D super long gradient analyses, using total analysis time versus proteome coverage and sample usage as benchmark parameters. In this work I demonstrate the limitation of the 1D LC-MS strategy, which plateaued of at about 4400 unique proteins and required approx. 8 hours, while the 2D workflow could reach 6000 proteins identified in only 7h, whereby the increase in proteome coverage could even be extended with longer analysis times. In chapter 3 I explore a method that is based on two sequential steps of SCX chromatography combined with differential stable isotope labeling, to define kinase consensus motifs with high accuracy. I thereby make use of proteolytic cell lysates as a source for peptide-substrate libraries that are incubated with the kinase of interest. I demonstrate the value of this approach by evaluating the motifs of two very distinct kinases: the cAMP regulated protein kinase A (PKA) and the human monopolar spindle 1 (Mps1) kinase, also known as TTK. PKA is a well-studied basophilic kinase while Mps1, has been less well characterized. Here we show that Mps1 is an acidophilic kinase with a striking tendency for phosphorylation of Thr. The final outcomes of our work are high-definition kinase consensus motifs for PKA and Mps1. We conclude that our generic method can be implemented for any kinase. In chapter 4 I further demonstrate that recent developments in mass spectrometry (MS)-based sequencing technology can expand the detectable peptide repertoire to an unprecedented depth, revealing unique features in the antigen presentation machinery. Here, I analyzed a vast repertoire of HLA class I eluted peptides, through which we find that they can harbor so far in the immunological field little explored PTMs such as arginine methylation and di-methylation. Alterations in post-translational modification are a recognized hallmark of diseases, therefore this kind of PTMs potentially provide a unique source of Human Leukocyte Antigen (HLA) class I-presented peptide antigens that can potentially uniquely stimulate the immune response. Proceeding further on this theme in chapter 5 I report unexpected mass spectrometric observations of glycosylated human leukocyte antigen (HLA) class I-bound peptides. Complemented by molecular modeling, in vitro enzymatic assays, and oxonium ion patterns, we propose that the observed O-linked glycans carrying up to five monosaccharides are extended O-GlcNAc rather than GalNAc-initiated O-glycans. A cytosolic O-GlcNAc modification is typically terminal and does not extend to produce a poly-

saccharide, but O-GlcNAc on an HLA peptide may present a special case because the loaded HLA class I complex traffics through the endoplasmic reticulum and Golgi apparatus on its way to the cell membrane, and can hence be exposed to glycosyltransferases that further modify the initiator O-GlcNAc group. Certain HLA class I peptides with centrally located oligosaccharides have been shown to be immunogenic and therefore the here reported antigens may also represent novel targets for immune-surveillance. In chapter 6 we charted the phosphopeptides bound to four distinct HLA-B molecules and observed that they shared a preference for peptides phosphorylated at position 4. We investigated the structural basis for this observation concluding that the observed phosphorylation motif may be common to most HLA-B molecules. The presentation of cancer related neo-antigens can trigger a specific response to destroy transformed cells. Since deregulated phosphorylation is a hallmark of malignant transformation, tumor-specific phospholigands could be prime targets for cancer immunotherapy. Thus, the knowledge added by our work provides a base for the improved prediction and identification of phosphorylated neo-antigens, as potentially used for cancer immunotherapy.

In the last chapter I present a mass spectrometry (MS)-based analysis of HLA-DR-associated self-peptides presented by the human monocyte-derived dendritic cell line MUTZ-3. The complementary peptide fragmentation techniques ETD, EThcD and HCD were employed to improve the identification and coverage of HLA-DR-associated peptides. Evaluation of the proteolytic cleavage sites of HLA-DR-associated peptides revealed a dominant role of the lysosomal enzyme Cathepsin D in antigen processing. Additional global proteomics analysis of the MUTZ-3 cell enabled a comparison of the composition of the cellular proteome and the surface-presented HLA-DR ligandome, revealing marked specificity in ligand presentation. The data suggest selection of proteins in the HLA class II antigen processing and presentation pathway, related to the route of antigen delivery, diversity in proteolytic activity and strict selection by HLA-DR molecules. Hence, MS-based peptide sequencing provides an unbiased picture of the landscape of cell surface-presented HLA class II-associated peptides, as visible for T-cells.

SAMENVATTING

Het werk, dat wordt beschreven in deze thesis, heeft als doel het verbeteren de verschillende stappen in de Proteomics werkwijze; van monster voorbereiding, naar de optimalisatie van geavanceerde chromatografische scheidingsmethoden en alternatieve MS fragmentatie technieken. Een belangrijke focus in deze thesis is gelegd op de analyse van peptide met verschillende post translationele modificaties (PTMs), zoals glycosylering, fosforylering, methylering en di-methylering. Om de detectie van PTM's te verbeteren is vaak een specifieke aanpak per modificatie nodig. Om deze verbeteringen te behalen, laat ik hier zien hoe verbeteringen in MS- en chromatografische technieken een voordeel kunnen zijn voor PTM identificatie. Verder laat ik zien hoe deze verbeteringen specifieke onderzoeksgebieden kunnen ondersteunen, voornamelijk bij de detectie van endogene peptide van HLA klasse I en II moleculen. Deze molecule hebben bijzondere karakteristieken in vergelijking met een 'standaard' proteomics tryptisch digest. In hoofdstuk 2 bespreek ik de ontwikkeling en implementatie van een online tweedimensionaal SCX/RP UHPLC systeem, gekoppeld aan een snelle sequencing en hoge resolutie MS. Als eerste beschrijf ik hier de karakterisering van een één-dimensionale UHPLC MS/MS techniek. Hierna vroegen we ons af of een snelle en sensitieve online 2D SCX-RP UHPLC-MS/MS techniek wellicht kon concurreren met een 1D extra lange gradiënt, waarbij de totale analyse tijd tegen de dekkingsgraad van het proteoom en het monster verbruik als referentiepunt zijn genomen. Hierbij laat ik de tekortkomingen van de 1D LC/MS strategie zien, dit bereikte zijn plateau bij de identificatie van ±4400 unieke eiwitten gedurende ongeveer 8 uur aan analyse. Terwijl de 2D analyse 6000 eiwit identificaties kon halen in slechts 7uur, waarbij de totale dekking van het proteoom nog vergroot kon worden door de analyse tijd te verlengen. In hoofdstuk 3 is een methode onderzocht waarbij 2 opeenvolgende stappen van SCX chromatografie worden gecombineerd met stabiel isotoop labelen, om op accurate wijze kinase consensus motieven te definiëren. Daarbij maak ik gebruik van proteolytische cel lysaten als bron voor een peptide-substraat database, welke met het de kinase van interesse zijn geïncubeerd. Ik laat de meerwaarde van deze aanpak zien, door op deze wijze de twee motieven van twee zeer verschillende kinases te evalueren: het cAMp regulerend eiwit kinase A (PKA) en het human monopolare spindle (Mps1) kinase, ook bekend als TTK. PKA is een goed onderzocht basofiel kinase, terwijl MPS1 minder goed gekarakteriseerd is. Hier laten we zien dat Mps1 een acidofiel kinase is met een erg sterke voorkeur voor Thr fosforylatie. Uiteindelijk laten we de wel gedefinieerde kinase motieven voor zowel PKA als Mps1 zien. We concluderen hier dan ook, dat onze methode potentieel bij elk kinase kan worden ingezet. In hoofdstuk 4 laat ik verder zien dat recente ontwikkelingen in massa spectrometrie gebaseerde sequencing technieken leiden tot de uitbreiding van het totale repertoire van te detecteren peptide, waarbij unieke eigenschappen van de antigen presenterende machine kunnen worden aangetoond. Ik heb hier een groot repertoire aan HLA klasse I eluerende peptiden geanalyseerd, waarbij we, de tot zover in het immunologische veld, weinig bestudeerde PTM's zoals arginine methylering en di-methylering laten zien. Veranderingen in PTM's zijn een kenmerk van ziekten, daarom zijn deze PTM's een unieke bron van humane leukocyten antigen (HLA) klasse I presenterend peptide antigenen, die potentieel de immuun respons op een unieke manier kunnen stimuleren. Om verder op dit onderwerp door te gaan rapporteer ik in hoofdstuk 5 onverwachte massa spectrometrie observaties van geglycosyleerde HLA klasse I gebonden peptiden. We stellen, op basis van deze observaties, aangevuld met moleculaire modelering, in vitro

enzymatisch assays en oxonium ion patronen, dat de geobserveerde O-gelinkte glycans die tot vijf monosacchariden dragen verlengde O-GlcNAc modificaties zijn in plaats van GalNAc-geiniteerde O-Glycans. Een cytosolische O-GlcNAc modificatie is typisch een eindpunt modificatie en verlengd dan niet verder tot een polysaccaride. Echter O-GlcNAc op een HLA peptide vormt wellicht een uitzondering op deze regel, omdat het gevulde HLA klasse I complex door het endoplasmatisch reticulum en het Golgi apparaat reist op weg naar het celmembraan en daarbij kan het worden bloot gesteld aan glycosyltransferases, die de O-GlcNAc group verder kunnen modificeren. Van bepaalde HLA klasse I peptide met centraal gelokaliseerde oligosachariden is aangetoond dat ze immunogeen zijn en wellicht zijn de hier getoonde antigenen ook een nieuwe aangrijpingspunt voor immuun-surveillance. In hoofdstuk 6 beschrijven we fosfopeptiden die binden aan de vier specifieke HLA-B moleculen, waarbij we ontdekte dat ze een preferentie delen voor een fosforylering op positie 4. We hebben de structurele basis voor deze observatie verder onderzocht en hebben geconcludeerd dat deze fosforylering door de meest HLA-B moleculen gedeeld wordt. De presentatie van kanker gerelateerde neo-antigenen kan een specifieke respons opwekken om deze gemuteerde cellen te vernietigen. Aangezien gedereguleerde fosforylering een kenmerk is van maligne veranderingen, kunnen tumor-specifieke glycans een belangrijk aangrijpingspunt zijn voor kanker-immunotherapie. Onze resultaten kunnen hierdoor een basis zijn voor verbeterde voorspelling en identificatie van gefosforyleerde neo-antigenen en potentieel gebruikt worden voor kanker immuuntherapie. In het laatste hoofdstuk laat ik een massa spectrometrie gebaseerde analyse van HLA-DR-geassocieerde “self” peptiden zien, welke gepresenteerd worden door de humane monocyten-gedreven dendritische cellijn MUTZ-3. Om de identificatie en totale dekking van HLA-DR-geassocieerde peptide te verbeteren zijn de complementerende peptide fragmentatie technieken ETD, EThcD en HCD gebruikt. Evaluatie van de proteolytische splitising plaats van HLA-DR-geassocieerde peptiden liet een dominante rol voor het lysosomale enzym Cathepsin D in antigen verwerking zien. Door additionele Proteomics analyse van de MUTZ-3 cellijn kon er een vergelijking gemaakt worden tussen de samenstelling van het cellulaire proteoom en het aan de oppervlakte gepresenteerde HLA-DR ligandome, waarbij duidelijke specificiteit zichtbaar was voor de ligand presentatie. De data suggereert een selectie van eiwitten in de HLA klasse II antigen verwerkings- en presentatie route, gerelateerd aan de antigen afleveringsroute, een diversiteit aan proteolytische activiteit en strikte selectie door HLA-DR moleculen. Met andere woorden, MS gebaseerde peptide sequencing geeft een onbevooroordeeld beeld van de aan de cel oppervlakte gepresenteerde HLA-klasse II geassocieerde peptiden, zoals aanwezig in T-cellen.

Curriculum Vitae

In July 2011, I earned my five years master degree in Chemistry and Pharmaceutical Technologies from the University of Palermo, Italy. During my last year of university I also participated to an exchange program in collaboration with the Hungarian Academy of Sciences under the supervision of Prof. Karoly Vekey, where I undertook my master research thesis in mass spectrometry-based proteomics. Since December 2011 I have been a PhD student in a 4-year program at the Biomolecular Mass Spectrometry and Proteomics Group at the University of Utrecht under the supervision of Prof. Albert J.R. Heck and Prof. Shabaz Mohammed. In the first two years of PhD my activities were mainly focused on the development of single and multidimensional chromatographical separations and MS methods for improving the detection of post-translational modified (PTM) and unmodified peptides. In the last period of my PhD I have been applying state-of-art technologies, which we successfully improved, in order to enhance the identification of HLA class I and II-bound peptide enabling the analysis of even lower abundant HLA peptide harboring a variety of post-translational modifications

OUTLOOK

Introduction

Over the last decade, advances in sample preparation and LC-MS instrumentation have made MS-based proteomics the method of choice for the investigation of proteomes. (1–5) MS-based proteomics typically relies on a 'bottom-up' approach, where proteins are isolated from biological samples of interest, enzymatically cleaved into smaller peptides and analyzed by liquid chromatography coupled to mass spectrometry (LC-MS). Over the last decade, bottom-up proteomics has made major leaps forward through parallel advances in every segment of the proteomics workflows, including preparation, LC-MS instrumentation and database search algorithms. (6)

Trypsin-based bottom-up proteomics

During all these developments protein digestion has been largely monopolized by trypsin, which has been addressed as the 'gold standard' protease for proteomics experiments because of its high cleavage specificity, tolerance to a large range of conditions as well as generating peptides that are very suitable for MS analysis. (7) This strict adherence to the conventional bottom-up trypsin-centric approach in the proteomics community can be also largely attributed to its relatively affordable cost, efficiency and robustness making it a high-throughput strategy. However, the use of such workflows need to be constantly and critically scrutinized in order to achieve progress. A disadvantage of tryptic digests is accessibility is restricted to a segment of each protein sequence.(8) In total fairness for some proteomics experiments 'complete' protein coverage is not necessary, but the consequence of this limited sequence information leads often to not being able to distinguish between protein isoforms and also to identify all the PTMs residing on proteins. A seeming advantage of trypsin is the generation of relatively small peptides (0.7 kDa to 3 kDa) which seem to be ideal for current standard LC-MS/MS and search algorithm-based identification methods. However, by having a closer look at the tryptic peptides generated in-silico from the human proteome, the majority of all tryptic peptides are shorter than 6 residues. (8) These peptides are often too small for successful binding to stationary phases commonly used in proteomics (RP) (9) and MS sequencing, (10) thus are not observed. Shorter peptides also rarely pass rigid filtering to reach an acceptable level of the estimated number of false-positive identifications. The fact that the majority of the generated protein fragments are discarded further limits attempts to observe and characterize 'complete' proteomes,(11, 12). The frequent cleavages by trypsin, often hundreds a peptides per protein, imposes a heavy burden on separations prior to MS detection. Intuitively, this aspect further limit the in-depth characterization of proteomes and even more so when considering PTMs. A partial solution to obtain more in-depth proteomics analysis can be achieved by using multidimensional separation schemes. Off-line coupling of multiple dimensions is, nowadays, the most common but this configuration comes at the cost of higher sample consumption. Robust on-line setups might be a valid alternative for clinical applications since they offer the possibility to reduce sample consumption and the overall analysis time. (2) Due to the tryptic peptides general features (i.e. mainly presence of two basic residues in C- and N-term of the peptide) MS based sequencing through (a number of forms of) collision induced dissociation (CID) has been efficiently employed in the characterization of proteomes. However, several PTMs have proven

problematic when subjected to collision-based fragmentation techniques. In some cases PTMs can hinder cleavage by trypsin consequently leading to missed cleaved peptides which can exacerbate sequencing by CID. (13) For instance, it is known that trypsin often does not cleave efficiently when phosphorylated Ser and Thr are closely located to Arg or Lys residues, due to the presence of the negative charge of the phospho moiety. (14)

Use of alternative proteases for more comprehensive proteomics data sets

Another often used way to increase the proteome coverage in bottom-up approaches is the use of alternative proteases. Enzymes such as Asp-N, GluC, LysC, chymotrypsin and several others, generate peptides that cover distinct parts of the proteome. Using multiple enzymes can also prevent introduction of biases.(15) Products of enzymes other than trypsin might not be very suitable for CID due to their chemical physical properties such as the presence of multiple internal basic residues in the peptide sequences. (16–19) On the other hand peptides with a high density of charge enable efficiently usage of alternative fragmentation methods such as ETD and ECD. (20, 21)

Middle-down the next proteomics avenue

Intriguingly, some of the above mentioned enzymes have the advantage over trypsin to generate longer peptides and has given rise to an emerging sub-field so-called middle-down proteomics. (22) This approach, where substantially longer peptides (ideally above 5 kDa) are generated, (23) might help in decreasing the complexity of protein digests, allow deeper proteome coverage and achieve further valuable sequence information. (24) For instance, longer peptides increase the probability to observe multiple PTMs at once, further facilitating investigation of PTM crosstalk. (25) Middle-down proteomics is not yet often used in a high-throughput fashion because they pose distinct analytical challenges. Consistency and specificity of these enzymes can be an issue and there is still no enzyme that robustly generates peptides within the optimal range. (8) These longer peptides require adjustment of chromatographic separations, features such as the stationary phase pore size need special attention. High mass accuracy and resolution in combination with ETD or ECD fragmentation would be also beneficial for the sequencing of such peptides. (23) Ultimately, data base search strategies, which have been mainly build for tryptic data sets, still represent a serious limitation for such kinds of alternative workflows.

MS-based immunopeptidomics the next challenge for proteomics workflows

The challenges associated with the accurate identification of peptides are also prescient for endogenously processed and HLA class I and II-bound peptides. Several classes of PTMs harboring the associated HLA peptides have been discovered, (26) whose identification pose further analytical challenges. Although the repertoire of HLA-presented peptides (including the PTM ligands) serves as a valuable source of information for the development of vaccines and cancer immunotherapy, (27, 28) their identity remains largely unexplored due to several technical limitations. So far even in the most extensive studies a small fraction of the associated peptides can be identified by MS-based proteomics.(29) HLA-peptide complexes are typically purified by immunoprecipitation using HLA-specific antibodies for class I and II. In a recent study it has been calculated that the overall work-up efficiency is about 5% for cultured cell systems, (30) therefore more purification methods with much lower sample losses are still needed. In particular for HLA-peptide complexes presented on cells of small organelles or tumor biopsies because for these types of samples it is extremely dif-

ficult to enrich sufficient peptide material for LC-MS analysis. (31) Moreover, the identification of HLA-presented peptides by tandem MS is complicated because the spectra are often of insufficient quality for high confident assignment of the peptide sequence. To address this problem, recently Electron Transfer / Higher Energy Dissociation (EThcD) (32) has been introduced for the analysis of HLA class I peptides, which can be very beneficial also in the study of ligands harboring PTMs. Development in fragmentation of HLA class II peptides has been neglected somewhat. Here, the challenge is the higher diversity of HLA class II bound peptides when compared to HLA class I ligands. Thus, new analytical approaches are needed to expand the depth of analysis for this class of ligandomes. Database searching of MS/MS data is also not straightforward because the HLA peptides are endogenously processed by multi-step digestion. Therefore, subsequent database search needs to be performed without enzyme specificity, which dramatically enlarges the search space and increases the number of potential false positive identifications. Moreover, if more than a few modifications at a time are considered, the search time increases exponentially alongside a drop in sensitivity due to the required p-value corrections. Sophisticated bioinformatics tools emerge constantly that are able to improve the reliability in HLA-peptide identifications. These methods, for example, combine in silico the information of multiple fragmentation techniques to generate more informative MS/MS spectra (33)or rescoring the database results using a machine learning method to improve confidence in peptide assignments.(34)

Conclusion

In conclusion, future technical developments in chromatography and combination of approaches using multiple enzymes and multiple fragmentation techniques will further maximize proteome coverage and provide valuable insight into biology and thus extend the ability to profile PTMs. These improvements will need to be explored together with improved bioinformatics tools. Moreover, in the field of immunopeptidomics the exploration of alternative workflows for immunopurification, and the use of dedicated, separation, fragmentation schemes and bioinformatics tools will greatly expand the knowledge on this class of endogenous peptides which are a key source for the development of vaccines and cancer immunotherapy.

REFERENCES

1. Wiśniewski, J. R., Duś, K., and Mann, M. (2013) Proteomic workflow for analysis of archival formalin-fixed and paraffin-embedded clinical samples to a depth of 10000 proteins. Proteomics Clin Appl 7, 225–233
2. Di Palma, S., Hennrich, M. L., Heck, A. J. R., and Mohammed, S. (2012) Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. J. Proteomics 75, 3791–3813
3. Hebert, A. S., Richards, A. L., Bailey, D. J., Ulbrich, A., Coughlin, E. E., Westphall, M. S., and Coon, J. J. (2014) The one hour yeast proteome. Mol Cell Proteomics 13, 339–347
4. Yates, J. R., Mohammed, S., and Heck, A. J. R. (2014) Phosphoproteomics. Anal. Chem. 86, 1313
5. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014) Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. Nat. Methods 11, 319–24
6. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C., and Yates, J. R. (2013) Protein analysis by shotgun/bottom-up proteomics. Chem Rev 113, 2343–2394
7. Raijmakers, R., Neerincx, P., Mohammed, S., and Heck, A. J. R. (2010) Cleavage specificities of the brother and sister proteases Lys-C and Lys-N. Chem. Commun. (Camb). 46, 8827–9
8. Tsiatsiani, L., and Heck, A. J. R. (2015) Proteomics beyond trypsin. FEBS J. 282, 2612–26
9. Mant, C. T., Burke, T. W., Black, J. A., and Hodges, R. S. (1988) Effect of peptide chain length on peptide retention behaviour in reversed-phase chromatography. J. Chromatogr. 458, 193–205
10. Swaney, D. L., Wenger, C. D., and Coon, J. J. (2010) NIH Public Access. Digestion 9, 1323–1329

11. Zhou, H., Di Palma, S., Preisinger, C., Peng, M., Polat, A. N., Heck, A. J., and Mohammed, S. (2012) Towards a comprehensive characterization of a human cancer cell phosphoproteome. J Proteome Res,
12. Lundby, A., Secher, A., Lage, K., Nordsborg, N. B., Dmytriyev, A., Lundby, C., and Olsen, J. V (2012) Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues. Nat. Commun. 3, 876
13. Tabb, D. L., Huang, Y., Wysocki, V. H., and Yates, J. R. (2004) Influence of basic residue content on fragment ion peak intensities in low-energy collision-induced dissociation spectra of peptides. Anal Chem 76, 1243–1248
14. Molina, H., Horn, D. M., Tang, N., Mathivanan, S., and Pandey, A. (2007) Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. Proc. Natl. Acad. Sci. U. S. A. 104, 2199–2204
15. Peng, M., Taouatas, N., Cappadona, S., van Breukelen, B., Mohammed, S., Scholten, A., and Heck, A. J. (2012) Protease bias in absolute protein quantitation. Nat Methods 9, 524–525
16. Zubarev, R. A. (2004) Electron-capture dissociation tandem mass spectrometry. Curr. Opin. Biotechnol. 15, 12–16
17. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A 101, 9528–9533
18. Frese, C. K., Altelaar, A. F. M., Hennrich, M. L., Nolting, D., Zeller, M., Griep-Raming, J., Heck, A. J. R., and Mohammed, S. (2011) Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. J. Proteome Res. 10, 2377–88
19. Wiesner, J., Premsler, T., and Sickmann, A. (2008) Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. Proteomics 8, 4466–4483
20. Gauci, S., Helbig, A. O., Slijper, M., Krijgsveld, J., Heck, A. J., and Mohammed, S. (2009) Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. Anal Chem 81, 4493–4501
21. Quan, L., and Liu, M. (2013) CID , ETD and HCD Fragmentation to Study Protein Post-Translational Modifications. Mod. Chem. Appl. 1, 1–2
22. Moradian, A., Kalli, A., Sweredoski, M. J., and Hess, S. (2014) The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. Proteomics 14, 489–497
23. Laskay, Ü. A., Lobas, A. A., Srzentić, K., Gorshkov, M. V, and Tsybin, Y. O. (2013) Proteome digestion specificity analysis for rational design of extended bottom-up and middle-down proteomics experiments. J. Proteome Res. 12, 5558–69
24. Zubarev, R. A. (2013) The challenge of the proteome dynamic range and its implications for in-depth proteomics. Proteomics 13, 723–6
25. van Noort, V., Seebacher, J., Bader, S., Mohammed, S., Vonkova, I., Betts, M. J., Kühner, S., Kumar, R., Maier, T., O'Flaherty, M., Rybin, V., Schmeisky, A., Yus, E., Stülke, J., Serrano, L., Russell, R. B., Heck, A. J. R., Bork, P., and Gavin, A.-C. (2012) Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium. Mol. Syst. Biol. 8, 571
26. Engelhard, V. H., Altrich-Vanlith, M., Ostankovitch, M., and Zarling, A. L. (2006) Post-translational modifications of naturally processed MHC-binding epitopes. Curr. Opin. Immunol. 18, 92–97
27. Ovsyannikova, I. G., Johnson, K. L., Bergen, H. R., and Poland, G. A. (2007) Mass spectrometry and peptide-based vaccine development. Clin Pharmacol Ther 82, 644–652
28. Admon, A., Barnea, E., and Ziv, T. (2003) Tumor antigens and proteomics from the point of view of the major histocompatibility complex peptides. Mol. Cell. Proteomics 2, 388–98
29. Hillen, N., and Stevanovic, S. (2006) Contribution of mass spectrometry-based proteomics to immunology. Expert Rev Proteomics 3, 653–664
30. Hassan, C., Kester, M. G. D., de Ru, A. H., Hombrink, P., Drijfhout, J. W., Nijveen, H., Leunissen, J. a M., Heemskerk, M. H. M., Falkenburg, J. H. F., and van Veelen, P. a (2013) The human leukocyte antigen-presented ligandome of B lymphocytes. Mol. Cell. Proteomics 12, 1829–43
31. Adamopoulou, E., Tenzer, S., Hillen, N., Klug, P., Rota, I. a, Tietz, S., Gebhardt, M., Stevanovic, S., Schild, H., Tolosa, E., Melms, A., and Stoeckle, C. (2013) Exploring the MHC-peptide matrix of central tolerance in the human thymus. Nat. Commun. 4, 2039
32. Mommen, G. P. M., Frese, C. K., Meiring, H. D., van Gaans-van den Brink, J., de Jong, A. P. J. M., van Els, C. a C. M., and Heck, A. J. R. (2014) Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proc. Natl. Acad. Sci. U. S. A. 111, 4507–12
33. Hayakawa, E., Menschaert, G., De Bock, P.-J., Luyten, W., Gevaert, K., Baggerman, G., and Schoofs, L. (2013) Improving the identification rate of endogenous peptides using electron transfer dissociation and collision-induced dissociation. J. Proteome Res. 12, 5410–21
34. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., and MacCoss, M. J. (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods 4, 923–925

PUBLICATIONS

Arginine (di)methylated Human Leukocyte Antigen class I peptides are favorably presented by HLA-B*07. Fabio Marino, Geert P.M. Mommen, Anita Jeko, Hugo D. Meiring, Jacqueline A.M. van Gaans-van den Brink, Cécile A.C.M. van Els, Albert J. R. Heck. (Manuscript under revision)

A molecular basis for the presentation of phosphorylated peptides by HLA-B molecules. Adan Alpizar, Fabio Marino, Antonio Ramos-Fernandez, Manuel Lombardia, Anita Jeko, Florencio Pazos, Alberto Paradela, Cesar Santiago, Albert J. R. Heck and Miguel Marcilla. (Manuscript under revision)

The dark side of the HLA ligandome is populated by proteasome-generated spliced peptides. Juliane Liepe, Fabio Marino , Daniel E. Bunting, John Sidney, Alessandro Sette, Michael P.H. Stumpf, Peter M. Kloetzel, Albert J.R. Heck, Michele Mishto (Manuscript under revision)

Implementation of UV-photodissociation on a benchtop Q-Exactive mass spectrometer and its application to phospho proteomics. Kyle Fort, Andrey Dyachenko, Clement Potel, Eleonora Corradini, Fabio Marino, Arjan Barendregt, Alexander Makarov, Richard Scheltema, Albert J. Heck. (Analytical Chemistry)

Sampling from the proteome to the HLA-DR ligandome proceeds via high specificity. Geert P.M. Mommen, Fabio Marino, Hugo D. Meiring, Martien C.M. Poelen, Jacqueline A.M. van Gaans-van den Brink, Shabaz Mohammed, Albert J.R. Heck, and Cécile A.C.M. van Els. (Molecular & Cellular Proteomics).

Rapid Analyses of Proteomes and Interactomes Using an Integrated Solid-Phase Extraction-Liquid Chromatography-MS/MS System. Nadine A. Binai, Fabio Marino, Peter Soendergaard, Nicolai Bache, Shabaz Mohammed, Albert J. Heck. (Journal Proteome Research).

Extended O-GlcNAc on HLA class-I-bound peptides. Fabio Marino#, Marshall Bern#, Geert P.M. Mommen, Alexandre M. Bonvin, Cecile A.C.M. van Else, Christopher Becker, Albert J. Heck. (JACS Communication).

Simultaneous assessment of kinetic, site-specific, and structural aspects of enzymatic protein phosphorylation. Michiel van de Waterbeemd, Philip Lossl, Violette Gautier, Fabio Marino, Masami Yamashita, Elena Conti, Arjen Scholten, Albert J. Heck. (Angewandte Chemie)

Characterization and usage of the EASY-spray technology as part of an online 2D SCX-RP ultra-high pressure system. Fabio Marino, Alba Cristobal, Nadine A. Binai, Nicolai

Bache, Albert J. Heck, Shabaz Mohammed (Analyst).

Universal quantitative kinase assay based on diagonal SCX chromatography and stable isotope dimethyl labelling provides high-definition kinase consensus motifs for PKA and human Mps1. Marco L. Hennrich, Fabio Marino, Vincent Groenewold, Geert Kops, Albert J. Heck, Shabaz Mohammed (Journal Proteome Research).

Digestion protocol for small protein amounts for nano-HPLC-MS (MS) analysis. Lilla Turiak, Oliver Ozohanics, Fabio Marino, Laszlo Drahos, Karoly Vekey (Journal Proteomics).

ACKNOWLEDGEMENTS

Before beginning with my acknowledgements, I want to start with a self-reflection. One day one collegue (Aneika) told me 'sometimes is important to look back, remember what you have achieved and how far you have come'. I think for me this is the right moment to remember how hard it was to come to the Netherlands and get my PhD, coming from a such beautiful island, which unfortunately is still so problematic. Today is my day, I will get what I have been fighting for so long, today is not the time to look to the future, it is the right moment to enjoy the present and feel proud.

It seems retoric but for me the support of my family has been everything. You know me, I am not a person of too many words but even with all my might I could't have been here today without you. You have not been simply my support, you are everything a son and a brother can desire. You are a great combination between family and friends. We might be far away but our bounds will not change. I want also to thank the rest of my family, you are not simple relatives, you are extra brothers and sisters, fathers and mothers. Thanks for taking care of me, my sister, my father and mother.

My story in the Netherlands starts more than 4 years ago, thanks to an invitation from Shabaz. I still like to laugh and think about the interview, a poor guy coming from an island facing for the first time the Manchester accent. I could at best understand half of the words you pronounced and I even had to answer randomly. I really thought you would not chose me for the PhD position, but something convinced you and that was my lucky day. Of course we had hard times but I like to remember only what you have done for me and all the time you spent teaching me. Thanks, I wish yo uall the best in Oxford.
Once Shabaz left another chapter of my PhD started. Albert, your guidance has been quite different from Shabaz's. Neverthless, you help me a lot in growing up as a scientist with your being honest about my defects. You also taught me how to think more broadly in science and thanks for believing in me when I asked you to change research topic.

Thanks to all my collaborators because you imporved me as a scientist. Thanks to Cecile who patiently commented several manuscripts and contribute to my knowledge of HLA class I and II pathways. Geert, I started working in the HLA field with you and we shared many works. You are a great guy and it was mazing to work together. I wish you all the best in Oxford with your family. Thanks to Michele and Juliane who share my same passion and energy for science. We worked for more than 2 years together and it was a great exchange of knowledge. Miguel and Adan, thanks for your enthusiasm and professionality.

I am going already to apology if I do not mention some of you in these pages but I hope the great moments spent together and the moemories of the past years will make up for my lack of words.
Albiola thanks for having been a great support, for spending so much time together and cheer me up in every occasion, i will miss you in Switzerland my friend.
Nicolas thanks for having shortened my life with alcohol and bouncing in concerts. I know I will see you very soon so I am not going to be too sentimental.
Lucrece I think you are by far the person who heard the most my whining, do not worry I have not replaced you, we will catch up once we meet again.

My italians, I should probalby write a book if I want to thank our amazing colony. You are the ones who could probably understand me the most, not only because we are from the same country but because we share similar stories and destinies. I wish us to find our happiness and stability but without compromising our dreams that we have been fighting so much for.
Renske, we shared 4 great years together. Thanks for being always sweet with me. I will miss seeing you nearly every morning at 8.45 in the corridory. I cross my fingers for your PhD, I hope to see you soon entering the cerimony room.
Marco, thanks for all your patience and all your teachings. Without you at the beginning of my PhD I would have been lost.
Nadine, we shared 4 years and now for both is the moment to start a new life. Despite you being a quiet person, we had a lot of fun together (I still think that the schnapps made by your father was sometimes a bit heavier than what was in the label).
Anita, it was fun to assist you in the past year. I hope I wasn't too much of a bother for you. Thanks for all the nice moments and discussions. I wish you to find your happiness!
Martin, Henk and Mao , you have been great room mates. we had a lot of fun together and you helped me a lot with my bioinformatic problems for dummies.
Anja, Arjan, Michiel, Kyle, Vojta and Philip, buddies thanks for all the super fun Fridays and for all the 'normal' conversations we had. I will not forget you.
Ana, Violette, Lianisky (mommy), Clement, Celine, Esther, Jing, David, Suzy, Sem, Oleg, Thierry, Matina, Aneika, Andrea, Pierre, Miao, Teck, Fiona, Andrey, Bas, Richard, Maarten, Barbara, Linsey, Wei, Yang, Guambo, Fan, Saar, Anna H., Charlotte, Sander, Sibel, Tomislav. and to all the former members..I hope I did not forget anybody. Thanks for all the moments together.
A special thank to all the technicians (Soenita, Mirjam, Harm, Anja, Arjan, Luc and Jamila) for all your help and patience with me.
How to forget Corine, our angel. Thanks for your help in nearly anything! We really could not survive without you.

Now, last but not least. Natasja, we met during my PhD and we have been together for several years. You are not just a girlfirend, you are my friend, support and with you I feel constantly balanced and loved despite all my defects. I want to particularly thank you because you really tried to understand my passion for science, even if sometimes I understand that it goes beyond logic. Thanks for accepting to move from your country and leave your security because of me and my dreams. I hope I will ever be able to make up to you.