Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Developing mass spectrometry towards applications in clinical proteomics

Improved sample preparation techniques and mass spectrometric methods for unbiased identification of proteome from clinical samples

Nagarjuna Nagaraj

aus

Thanjavur, India

2010

Erklärung

Diese Dissertation wurde im Sinne von § 13 Abs. 3 der Promotionsordnung vom 29. Januar 1998 von Herrn Prof. Dr. Matthias Mann betreut.

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbstständig ohne unerlaubte Hilfe erarbeitet.

München, am ……

……..……

(Nagarjuna Nagaraj)

Dissertation eingereicht am ……

1. Gutachter: Prof. Dr. Matthias Mann

2. Gutachter: Prof.Dr. Jacek R Wisniewski

Mündliche Prüfung am 05.Ocktober 2010

Contents

1.0 Introduction

Proteome versus proteins

Each organism contains a single genome in virtually all its cells and thus the difference between the numerous cell types within the same organisms are attributed mainly to genetic imprinting, epigenetics, transcriptional and post- transcriptional processes. Thus studying all the protein molecules present in the living system (termed proteomics) at any given time and physiological condition will logically provide insights into the mechanism of life to an unprecedented depth. Historically when proteins were studied 'one at a time' achieving a systems view of the cell or organelle was a daunting and laborious endeavor to undertake. Further when proteins are studied individually the results could be occasionally bewildering owing to the crosstalk between the signaling pathways and nodes[1, 2]. Though there are successful structural biology technologies on a small scale and techniques like yeast two hybrid[3] and phage display system[4] to characterize the protein- protein interaction at a larger scale they do not provide a system wide view apart from establishing the interaction between groups of proteins. Further the false positive and false negative rates are difficult to estimate and the information is less quantitative (if at all) to construct a stoichiometric protein complex[5].

Mass spectrometry based proteomics

Mass Spectrometry (MS) - one the most sensitive analytical techniques - has played an enormous role in the development of proteomics to its existing capabilities. The advent of ESI and MALDI techniques initially led to mass spectrometry being used for individual protein sequencing and identification of gel bands thus gradually replacing the Edman degradation method. At first, proteomics was associated with 2-dimensional gel electrophoresis techniques (2DE)[6] but 2DE has serious limitations[7]. When protein and peptide separation techniques like liquid chromatography were combined with mass spectrometry[8] , complex protein and peptide

mixtures were successfully analyzed leading to the emergence of MS-based proteomics[9-13]. Complementary to advances in separation strategies, development of new mass spectrometers especially the FT ICR[14] and Orbitrap[8, 15-18] hybrid instruments, have facilitated routine large scale high accuracy and high resolution MS.

Mass spectrometry is not quantitative by itself and strategies have been developed to obtain relative quantification between conditions either using non-radioactive isotopes or the spectral information itself[19-21]. Proteins or peptides are labeled with light and different heavy non-radioactive isotopes which are identical in terms of biochemical properties including the ionization efficiency and differ only by mass and thus can be distinguished in the mass spectrum[22-25]. Labeling can be performed by (1) the metabolic labeling of cells using ^{15}N isotopes[26, 27]or by stable isotope labeling of amino acids in cell culture (SILAC)[28, 29] and (2) chemical labeling of peptides like isotope coded affinity tags (ICAT)[30], Hys Tag[31], dimethyl labeling[32] and isobaric tags iTRAQ[33]. Recently, Geiger *et al* have demonstrated the applicability of SILAC based quantification to clinical samples[34], which could potentially lead to a new paradigm of clinical proteomics. In a label-free format quantification can be performed from the extracted ion chromatogram (XIC) for the peptides ('label free quantification'). Currently a wide range of algorithms are available for label free quantification making this an attractive strategy especially for clinical samples and samples for which labeling is generally not feasible.

With concurrent maturation in many facets, MS based quantitative proteomics is now set to become an indispensable tool in cell and systems biology. It is now possible to identify complete or near complete proteomes of eukaryotic cells within a reasonable amount of measuring time[35, 36]. Quantitative proteomics in its current state is routinely employed in studying overall expression changes, cell signaling networks[37], classification of cell types[38], post translational modification(PTM) (phosphorylation[39, 40], acetylation[41], ubiquitination[42] etc), protein-protein interactions[43-46], organelle specific localization[47-49] and protein turnover rates.

MS proteomics for clinical applications: hype and hope

As explained above, MS based proteomics is now a common and powerful tool in cell biological experiments. However one of the ultimate goals of proteomics is to transfer the technology to clinical applications. In clinical chemistry a few proteins are monitored based on assays or ELISA methods for diagnostic and prognostic purposes. These proteins are only a minute subset of a plethora available in the sample (plasma, serum or urine[50, 51]). This warrants the application of proteomics, which can handle hundreds if not thousands of proteins simultaneously. Thus, the field of clinical proteomics provides opportunity to identify and monitor new disease markers in body fluids, cells and tissues that can be used in diagnosis at a very early stage of disease, stratification of patients for specific treatment and therapeutic monitoring[52, 53]. The feature of quantitative analysis of thousands of proteins simultaneously and the inherent sensitivity of mass spectrometry implies a huge promise for application in clinical diagnosis.

The potential of MS has generated hopes of finding new biomarkers for clinical diagnosis of many diseases including cancer especially at very early stage of disease progression. However to date there is still no successful protein biomarker developed from mass spectrometric methods that is validated and verified - so far turning the early claims into mere hype[54]. Several contributing reasons include the application of premature technology, the inherent difficulties in analyzing body fluid samples that were thought to be an ideal source of samples, the need to measure a large number of samples with extensive fractionation and initial lack of reliable identification[55] and of robust quantification methods. Several strategies were tried to address these problems – especially the dynamic range challenge of plasma proteins[56, 57] - with varying degrees of success. For example, the dynamic range of plasma, serum and urine was reduced by applying antibody based depletion methods. However reports suggest that important low abundant proteins like cytokines may also be lost concomitantly during the depletion process[58].

Thesis work

Given the challenges of clinical proteomics, it is clear that the technology biomarker identification from body fluids needs to be drastically improved, which is the topic of my thesis[59]. Sample preparation methods, mass spectrometric analysis and downstream data analysis constitute the different modules of the clinical proteomics platform. As I will show development in these modules including speed, accuracy and resolution of the mass spectrometers, unbiased sample handling strategies and the availability of sophisticated algorithms to perform label free quantification has greatly advanced progress towards 'real' MS based clinical proteomics.

In my thesis, I have worked on different aspects including sample preparation and fragmentation techniques that could potentially improve the technology for clinical proteomics. In particular I have applied these improvements together with state of the art LC-MS/MS technologies to study the normal human urinary proteome.

The first two projects are closely related and deal with improved sample preparation methods. The resulting sample is unbiased, cleaner and thus leads to more efficient sample fractionation and higher MS/MS identification success rates. The third project investigates the feasibility of a new, high accuracy fragmentation method for the analysis of phosphopeptides. This 'HCD' fragmentation method was thought to be relatively slow and less sensitive than existing low resolution methods, however, my work shows that this is not so and that the high resolution method is superior in all aspects. This opens up for interesting applications in large scale clinical studies of the phosphorylation status of tumor samples. The last project deals with application of our sample preparation and LC-MS/MS platform to study the human urinary proteome. The project involves development of robust and high-throughput method for identification and quantification of more than 600 proteins in human urine. For the first time, we accurately determine inter and intra-individual variations in comparison to the technical variation in the normal urinary proteome. This information will be crucial in determining the suitability of proteins as urinary biomarkers for any disease.

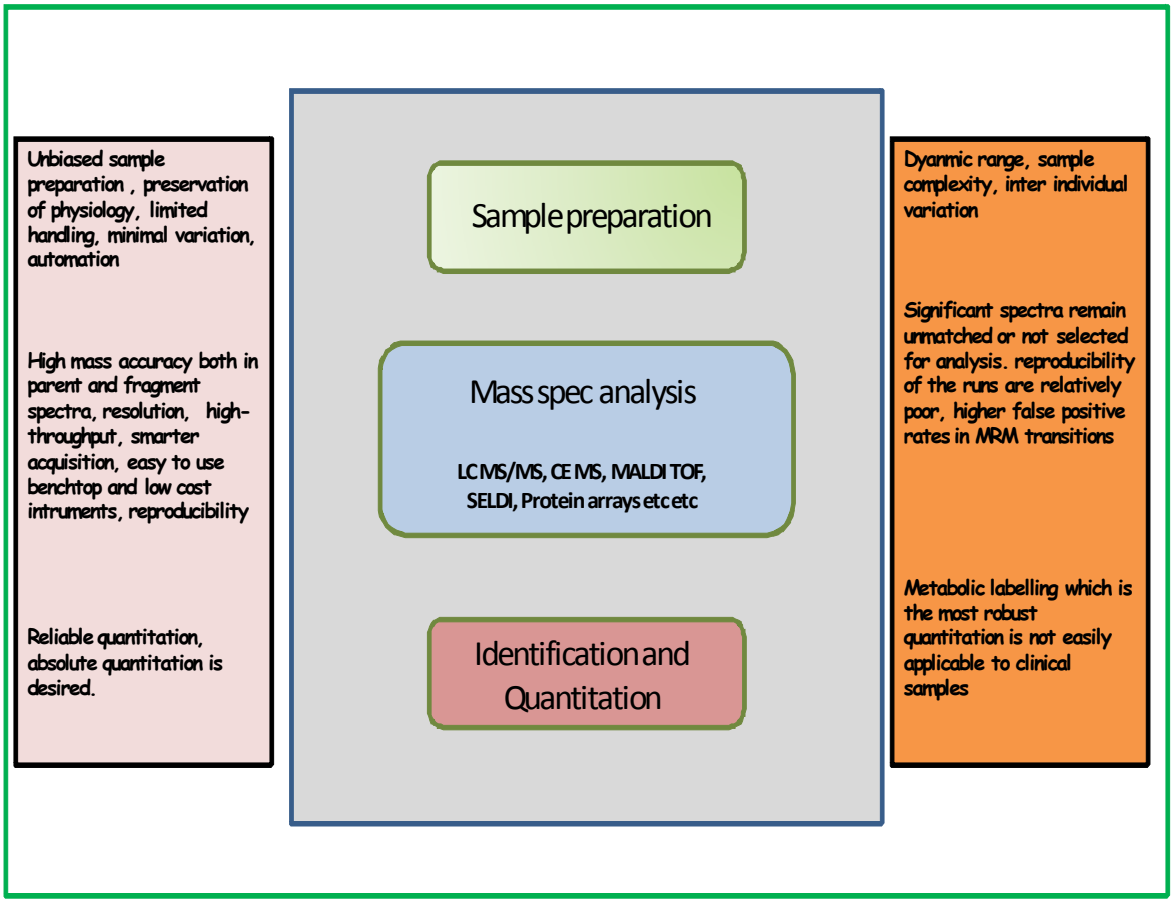


Figure 1. Modules for a clinical proteomics platform (middle panel). The leftmost panel specifies some of the desired attributes for the modules and the current issues to be solved either are listed in the right panel.

1.1 Introduction to sample preparation and fractionation for liquid chromatography- mass spectrometry

1.1.1 Challenges in proteome sample preparation in comparison to other omics technologies

In the post-genomic era, global profiling of cells and tissues to determine changes at the transcriptome level has become routine practice in order to gain insight in to the process or phenomenon under study. However as this profiling process shifts from transcriptome to proteome there is a paradigm shift in the level of complexity and diversity regarding the nature of the sample and also the sample handling. For example, the transcriptomic approaches deals with mRNA populations that have minor variation in terms of biochemical properties and its diversity is observed only in the sequence and length of the mRNAs. In contrast proteins are very diverse in terms of biochemical properties like amino acid sequence, length, three dimensional structures and allosteric conformation, solubility and biological properties like cellular localization and post translational modification. However this problem itself provides a potential solution since these differences and complexity of the proteins can be exploited to fractionate them thereby simplifying the proteome and hence the analysis.

Abundance of proteins in a sample can span up to 12 orders of magnitude[60] whereas the severity of this problem in mRNA and other high throughput profiling methods is less pronounced. Sample preparation constitutes the very first step in proteomic experiments and therefore determines the overall success of proteomics be it in large or small scale studies. The ideal sample preparation technique would not involve any 'sample handling'. However owing to the enormous diversity and complexity of the protein mixture, separation techniques are indispensable.

The key factors to be considered when designing any sample preparation method include maintaining the state of the proteome and avoiding artifacts like dephosphorylation or proteolysis, minimizing the bias against specific class of proteins such as membrane proteins, keeping the number of fractions to be analyzed in the mass spectrometer to a minimum and

maintaining the possibility of automation for the sample preparation to reduce the technical variation. Many proteomic analyses rely on protein level fractionation that can be based on the biochemical properties or biological characteristics of the proteins.

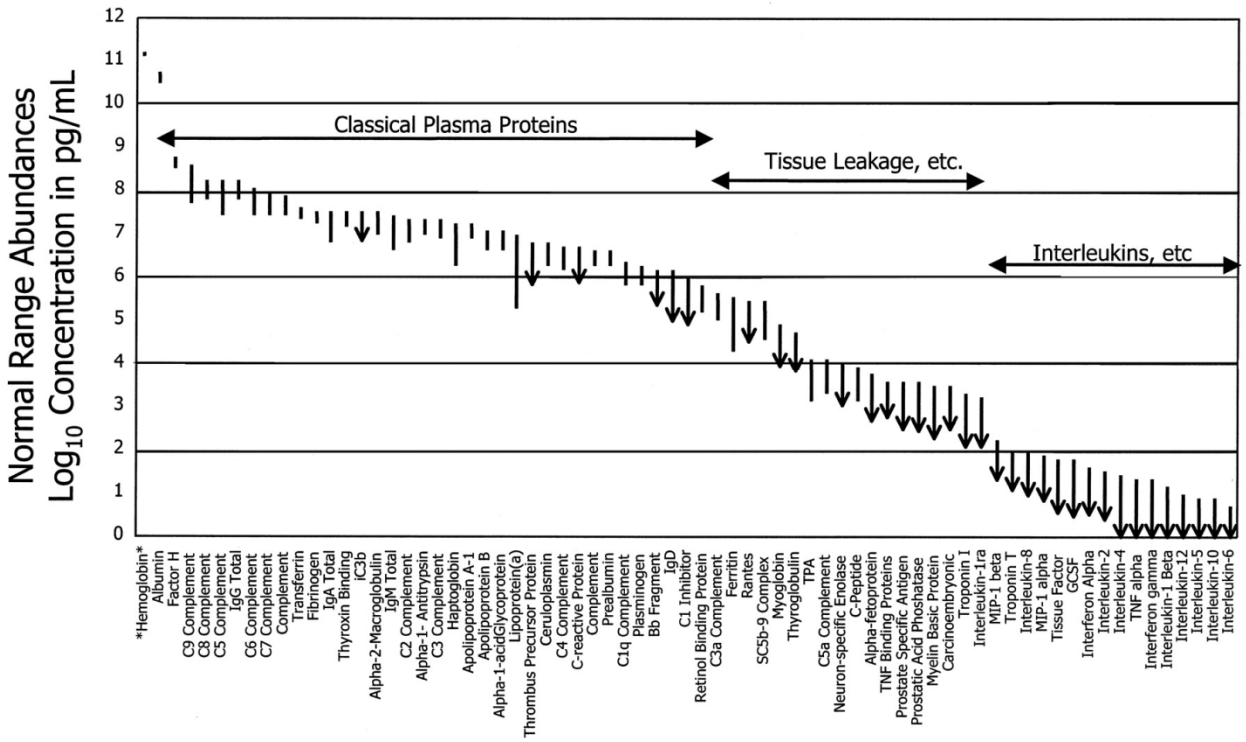


Figure 1.1.1 Dynamic range of the plasma proteome. This is one of the popular plots demonstrating the abundance range spanned by plasma proteins. The presence of highly abundant proteins may preclude the analysis low abundance proteins that are potential biomarkers. This 'dynamic range problem' is particularly severe in clinical proteomics. Adopted from ref[60]

1.1.2 Protein fractionation based on biochemical properties

Reversed phase high performance liquid chromatography (HPLC) is by far the most commonly used fractionation technique in MS-based proteomics. The online coupling of reversed phased chromatography to mass spectrometer by an electrospray interface is now commonly called LC MS or LC MS/MS. The inherently acidic nature of separation is beneficial for peptide ionization in the positive electrospray mode. Even when MALDI is applied for

ionizing the peptides before mass spectrometry, HPLC can be used for fractionation in an offline mode where samples are spotted onto the MALDI source plate.

Separation based on protein size by SDS-PAGE is also a commonly used strategy to fractionate samples. The proteins are usually loaded on the SDS gel and the gel is cut into several slices followed by in-gel digestion of the proteins using an endoproteinase, which is most commonly trypsin[61, 62]. The peptides are then extracted and separated on a reverse phase column before online electrospray in front of the mass spectrometer. This mode of sample preparation and analysis is sometimes termed GeLC MS and has been widely adopted for proteomic studies.

Proteins or peptides can also be separated based on their isoelectric point by isoelectric focusing by immobilized pH gradient strips or tubes[63]. Separation based both on size and isoelectric point is the principle of 2D gel electrophoresis. In this technique proteins are separated in the first dimension by isoelectric focusing followed by separation based on size in the second dimension and it was believed to be a promising strategy for high resolution separation of proteins. However 2D gel technology has many technical issues and limitations including bias against membrane proteins, very large and small proteins etc. Furthermore, although the 2D gel can give rise to many thousands of protein spots, it actually leads to identification of only a few hundred proteins, many fewer than LC MS/MS based approaches. Finally, the 2D gel process is time consuming, difficult to automate and replicate runs are cumbersome. Though proteomics was initially associated with 2D gels they have been almost entirely been replaced by LC MS/MS based proteomics.

Proteins and especially peptides can also be fractionated by ion exchange chromatography. Ion exchange is coupled to reversed phase separation in an orthogonal approach to fractionate the complex proteomes in large scale analysis of the peptides resulting from in solution digestion of the proteome. In one approach, strong cation exchange and reversed phase chromatography were orthogonally coupled online in order to achieve extensive identification of these peptides. This approach was termed multi-dimensional protein identification technology (MuDPIT) or shotgun proteomics[64].

1.1.3 Enrichment based on biological properties

Proteome samples can also be fractionated based on the biological properties of the proteins. This way of separating proteins serves two purposes; firstly, the sample complexity of the proteome mixture is drastically reduced secondly it leads to the enrichment of the proteins of biological interest that may otherwise be non-detectable despite reasonable fractionation. The biological characteristics that can be exploited for such fractionation are cellular localization (nuclear, cytoplasmic, membrane, mitochondrial, synaptic cleft etc)[65]; specific protein- protein interaction[43, 44]; DNA/RNA – protein interaction[66, 67]; enrichment for specific post-translational modifications such as phosphorylation[68-71], acetylation[41], glycosylation[72], SUMOlyation[73], redox state of the cellular environment or protein to name a few.

1.1.4 Sample preparation for systems / cell biology versus clinical proteomics

Proper combination of these separation techniques yields high resolution fractionation and thus deeper coverage of the proteome. However with each fractionation step, the number of 'fragmentsamples' that arise from a single fraction multiplies and thus increases the mass spectrometric measurement time. In clinical proteomics a large number of fractions is not desired considering the number of samples to be analyzed. Furthermore, multiple fractionations would in turn require much starting material to obtain reasonable peptide amounts in each of the final fractions that are to be measured in the mass spectrometer. While the sample is not limiting in case of cell lines and animal model tissues, fractionation may not be a good option for precious and limited sample material like clinical biopsies, bio-fluids obtained by invasive methods like cerebrospinal fluid (CSF). However limited fractionation can still be performed and the process can be scaled down significantly. This miniaturization is not only desirable for clinical samples but also for general proteomic applications. Peptide pre-fractionation based on ion exchange[74] and reversed phase separation (for example StageTips)[75, 76] can be performed in a pipette tip and this is now routinely applied in general proteomic experiments.

In a clinical perspective apart from body fluids, other major sources of sample typically include tissue banks where samples are preserved as formalin fixed and paraffin-embedded

(FFPE) or frozen specimens. According to recent reports, it is possible to access the proteome and its PTMs in either of these states using new sample preparation techniques[77, 78].

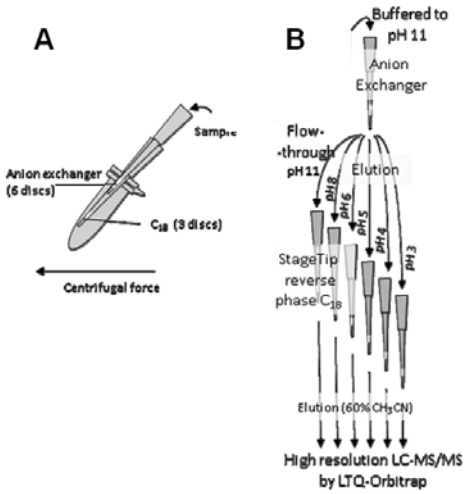


Figure 1.1.4 Miniaturization of ion exchange separation of peptides. Peptides obtained from protease digestion are fractionated on an ion-exchanger by precise pH elution. The fractions eluted from the anion exchanger are readily cleaned up in StageTips for LC-MS/MS. A. The set up for coupling anion exchange to reversed phase fractionation. B. Resolution can be improved by increasing the number of pH elution steps. The process can be scaled down to 20 µg peptides and the whole procedure takes less than two hours. Modified from ref[74].

1.2 Introduction to mass spectrometry based proteomics

1.2.0 History and introduction to mass spectrometers

Mass spectrometry is an analytical technology that is used in a variety of fields, including medicine, life sciences, pharmaceutical sciences, organic chemistry and physics[79]. Having such a wide range of applications, the rapidly evolving technique of mass spectrometry has its roots in one of the key experiments of JJ Thomson where he studied the movement of electrons (negatively charged cathode ray particles) and other charged particles in electromagnetic fields[80]. In his experiments, J. J. Thompson demonstrated that the movement of these charged particles in vacuum under the influence of electric and magnetic fields is dependent upon the mass to charge ratio denoted as m/z. Mass spectrometers were commercially available from as early as 1943 and the early reports explaining the principles of time of flight (TOF) and ion cyclotron resonance (ICR) mass spectrometry were published in 1946 and 1948 respectively[81, 82]. The ICR mass spectrometry requires strong magnets and this issue was alleviated when Paul described the use of quadrupole and ion traps as mass analyzers[83, 84] and hence these ion traps are also called Paul traps. Later the idea of tandem mass spectrometry evolved which resulted in the birth of fragmentation mass spectrometry (MS/MS) a hallmark in unambiguous structure determination[85]. This fragmentation or MS/MS turned out to be at the heart of MS-based proteomics for obtaining peptide sequences essential for determining the protein identity.

Any mass spectrometer consists of three basic units (1) source of ions, (2) mass analyzer and (3) detector. The mass spectrometer also has an inlet system for the ion source, a data system, vacuum system and control electronics. The ion source transfers the analyte into the gaseous state and also ionizes the analyte so that it can be manipulated inside the mass spectrometer. From the ion source the charged analyte molecules flow into the mass analyzer facilitated by pressure differences and by a series of electric potential difference. Inside the mass analyzer the charged analyte are separated and analyzed by their motion based on their characteristic charge to mass ratio under the influence of magnetic and/or electric fields. These mass analyzers can differ in the principle of separation and analysis of ions and are of various forms such as magnetic sector, time-of-flight, quadrupole, ion trap and Fourier transform ion

cyclotron resonance analyzer. The detector measures the ion signal and amplifies it to improve the signal and sensitivity of the instrument. Data acquired in the mass analyzer that are recorded by the detection system are usually represented as a spectrum. The mass to charge ratio of the analyte (denoted as m/z) has units of Thompson (Th) and the relative intensity of each species is represented in the y axis. In an over-simplistic representation, with respect to a proteomic context, the m/z value yields the identity of a protein and the intensity value translates to the relative abundance of a protein in the system under study. In the following sections the key aspects of mass spectrometry and proteomics in general are discussed briefly.

1.2.1 Ionization methods

The first step in a mass spectrometric process is to generate charged analyte in the vapor state. Ionization of the analyte is crucial as it is easy and practical to control the movement of charged particles in electric and magnetic fields and to detect them. If neutral particles were to be analyzed this should be under the influence of gravitational fields, for example, and for monitoring such minor differences between different species the analyzer might need to be several kilometers or even longer to achieve mediocre resolution.

Methods like electron ionization, chemical and photo ionization, atom bombardment ionization were the commonly used techniques in mass spectrometry several decades ago. However these methods are too energetic for the large and fragile biomolecules that can decompose during the ionization process leading to less informative spectra. The huge gap between the potential of mass spectrometry and its application to biomolecules and life sciences remained until the development of two soft ionization techniques electrospray ionization (ESI) and Matrix assisted Laser desorption ionization (MALDI) that ionizes the peptides and other biomolecules in a gentle way. For this breakthrough with ESI and MALDI, John B Fenn and Koichi Tanaka were awarded a share of the Nobel Prize in chemistry 2002.

1.2.1.1 Electrospray Ionization

The concept of electrospray ionization (ESI) was put forth by Malcom Dole and development of electrospray for mass spectrometry was pioneered by John Fenn[86]. It is now a technique routinely used in proteomics though the mechanism of electrospray ionization itself is not well understood. In proteomics work flows, ESI is usually coupled to peptide separation techniques like liquid chromatography or capillary electrophoresis. In electrospray the sample flows through a capillary with its end maintained at high voltage up to 5 kV. The solution phase when reaching the tip of the column forms a cone shaped structure owing to its surface tension combined with the forces of the electric field. This structure called Taylor cone eventually results in formation of small droplets that in turn rapidly evaporate. As the charged droplets become smaller the increased surface charge density causes the ions to repel each other and the droplet bursts into smaller droplets. As the solvent completely evaporates the peptide ions fly into the mass spectrometer as gas phase ions that can be easily analyzed. The typical flow rate in electrospray used to be from 2- 10 µl/min. Further improvements in ESI were achieved through the nano ESI developed by Matthias Wilm and Matthias Mann where they demonstrated that the flow rate in the ESI capillary can be reduced up to 20 nl/min while increasing the sensitivity to the attomole range[87]. In their set up the spray needle was as narrow as few microns and the sample was loaded directly close to the spray needle without any pump delivery systems (‘nanoelectrospray’). Electrospray ionization usually results in multiply charged ions and thus provides information rich MS/MS spectra for peptide sequencing.

1.2.1.2 Matrix assisted laser desorption ionization (MALDI)

A precursor to the MALDI technique was discovered by Kiochi Tanaka in Japan and Karras and Hillenkamp independently developed MALDI in Germany. This breakthrough facilitated the analysis of molecules of more than 200 kDa with very high sensitivity. MALDI is usually coupled to time of flight analyzers and given the method of sample preparation for MALDI the technique is not coupled to liquid chromatography in an ‘online’ fashion. For MALDI, the analyte is mixed with an organic matrix material and dried on the plate. The mixing

of analyte and sample needs optimization in order to avoid segregation of sample and matrix during co-crystallization and to achieve good signals for the sample in the spectrum. The matrix and sample when hit with a laser beam of very high irradiance (up to 10^6 W/cm^2) desorb and ionize simultaneously. The matrix plays a key role by receiving the laser energy and transferring it to the sample, permitting desorption of even large and fragile molecules. Unlike ESI, singly charged ions are generated most prominently in MALDI. Since the laser is irradiated in a pulsed manner MALDI was initially coupled mainly to TOF analyzers, which anyway operate in a pulsed manner. Modifications like orthogonal acceleration have facilitated the use of trapping mass analyzers with MALDI. Since MALDI mostly generates singly charged species, the fragmentation process is not as efficient and often fails to provide sufficient peaks to sequence or identify a peptide. Thus MALDI has mostly been used in peptide mass fingerprinting where the peptides are not fragmented for identification. For complex mixture analysis ESI generally outperforms MALDI.

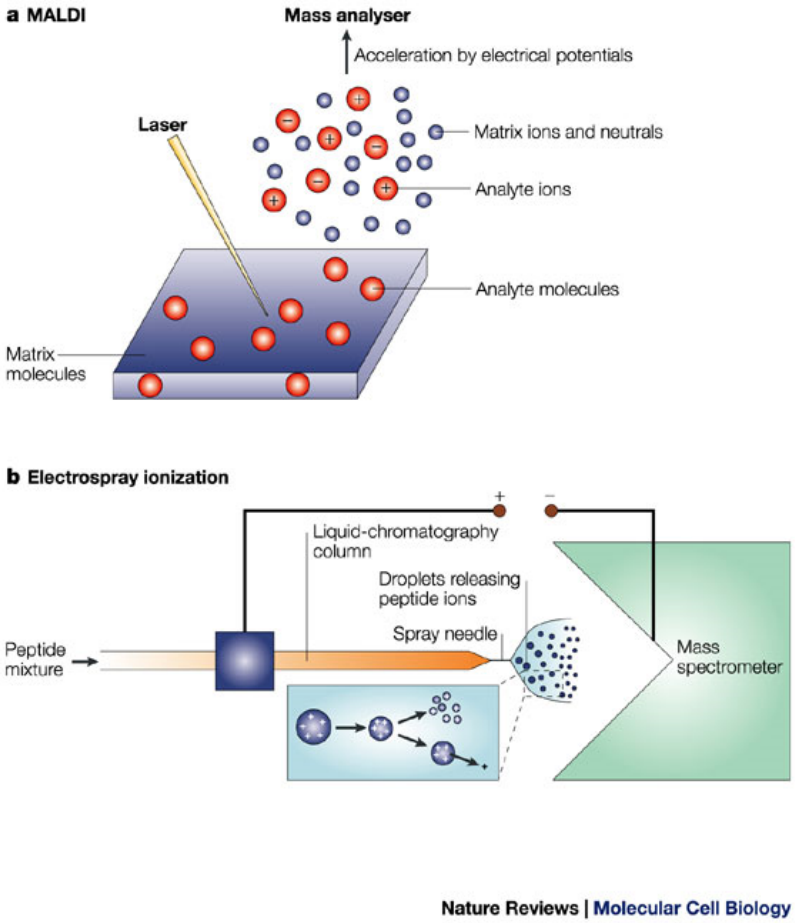


Figure 1.2.1 Soft ionization methods for biological mass spectrometry. MALDI generates singly charged ions and ESI usually results in multiply charged ions. Adopted from ref[9]

1.2.2 Different types of mass analyzers

Mass analyzers are of diverse types and can be classified and grouped based on several properties. Each mass analyzer type separates the ions by their mass to charge ratio (m/z) employing different principles of ion motion in electric and magnetic fields. Based on the scanning mode the analyzers can be classified into two types. In scanning analyzers, ions of different m/z are analyzed as the instrument allows ions of selected m/z to pass through at any given time. The scanning analyzers include magnetic sector and quadrupole instruments. Paul ion traps can also scan out the ions in an m/z dependent manner. Other mass analyzers like those of the time of flight (TOF) type can allow all ions to pass through at once. The analyzer can also be classified into beam type and trapping instruments based on how they store the ions and continuous and pulsed analyzers based on the scanning capabilities and high and low energy analyzers based on their fragmentation features.

The key parameters of mass analyzer include mass range, mass accuracy and resolution, sensitivity, dynamic range, speed and fragmentation capabilities. In order to achieve best results, two mass analyzers are often combined in tandem mass spectrometry. For example, ion trap analyzers have good sensitivity and speed and can be coupled to Orbitrap mass analyzers (see below) which have very high accuracy and dynamic range however somewhat lower speed compared to ion trap. In this configuration, a high accuracy mass spectrum of the precursor peptide ion is acquired in the Orbitrap analyzer at a very high resolution and simultaneously the fragmentation spectra of up to ten peptide ions are acquired in the ion trap with a fast duty cycle. This combination of analyzers is one of the most commonly applied ones in proteomic studies.

1.2.2.1 Magnetic sector analyzer

Magnetic sector analyzers are among the oldest types of mass analyzers. The magnetic sector analyzers can be single or double focusing. The single focusing instruments employ a magnetic field and for the double focusing instrument, electrostatic field in addition to magnetic field are employed. Sector instruments are known for the high energy collision induced

dissociation regimes that fragment peptides not only along the backbone but also at side chains – in principle allowing differentiation of isobaric amino acids such as isoleucine and leucine.

1.2.2.2 Time of flight (TOF) analyzer

As the name suggests, in the TOF analyzer different ions are distinguished based on the time required to traverse a particular field free path. The TOF analyzer is one of the oldest and simplest set ups. The experiments regarding TOF concepts were started as early as 1932[88]. William E Stephens described the first mass spectrometer built on the TOF principle in 1946[89] Starting from very moderate mass accuracy and resolution its performance has continuously increased over the decades and it is still one of most commonly used analyzers for proteomic applications. Because of the long field free path, most of the commercial TOF instruments are easily recognized by the presence of long tubes. This implies that in order to have long "mean free path" for the ions, the neutral particles should be removed and this necessitates relatively expensive vacuum systems.

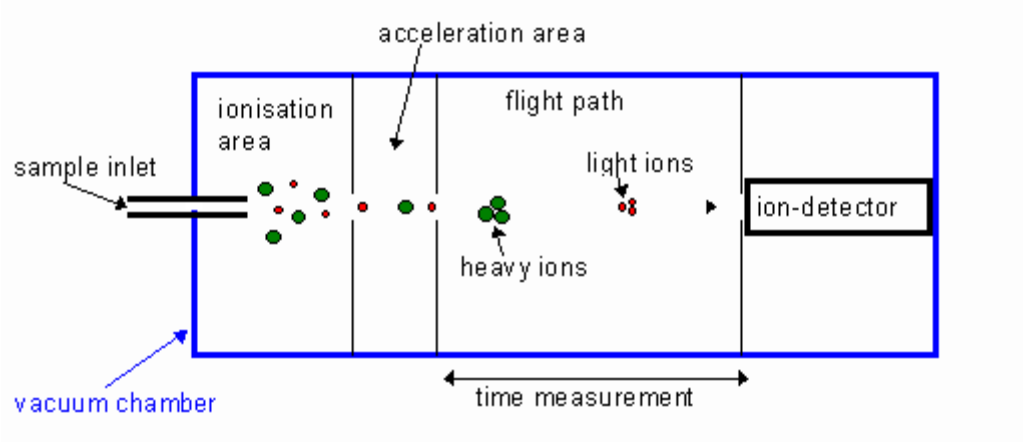
The ions are accelerated by a potential difference in the acceleration area and then traverse the drift region. Since the initial kinetic energy is same for all the ions, the velocity at the start of the drift region is inversely related to the mass of the ions. The following equation denotes the kinetic energy from the potential energy difference that is accelerating the ions. The mass of ion of charge q is denoted by m and the potential energy difference V

$$K.E = \frac{(mv^2)}{2} = q.V$$

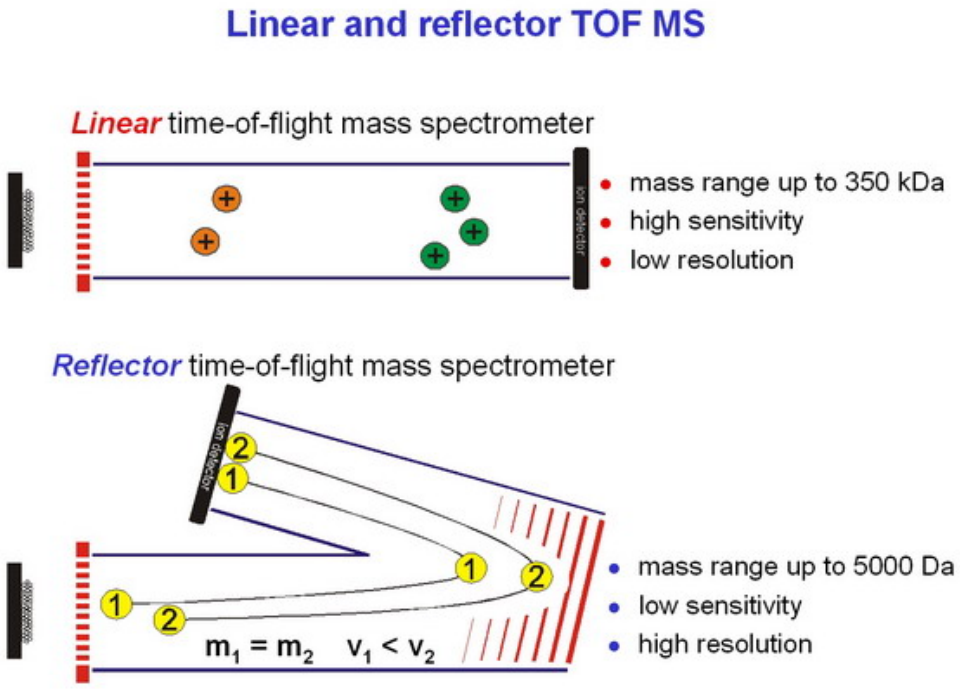
Replacing the velocity (v) with distance and time, the equation can be re-written as

$$t = \frac{d}{\sqrt{2V}} \sqrt{\left(\frac{m}{q}\right)}$$

Where *d* is the distance of the flight path.



Further improvements have been made in the TOF technology with reflectron technology in order to increase the resolving power. Most of the commercial instrument now uses the reflectron TOF; however in the reflectron TOF the mass range is not as broad as in the linear



TOF analyzer.

http://www.anagnostec.eu/uploads/pics/20060127133154maldiprinzip02_600.jpg

1.2.2.3 Transmission quadrupole

Quadrupole, 3-D ion trap and the linear ion trap are based on slight variations of common principles that use the stability of the ion trajectory in oscillating electric fields. These analyzers measure the mass to charge ratio by stability of the trajectory unlike the TOF and FT ICR, which use velocity and frequency of oscillation, respectively. A quadrupole analyzer consists of four parallel rods usually of an oval geometry with opposing rods of the same polarity and adjacent rods of opposite polarity. Application of a time varying field to each set creates a quasi harmonic potential well. The rods should ideally have a hyperbolic cross-section and the alternating electric field generated focuses the ions to the centre of the quadrupole analyzer. The principle of the quadrupole was first described in the early 1950s by Paul and Steinwedel[90] and it later turned out to be one of the most successful commercial mass spectrometers[91].

The trajectory of ions is controlled by the combined effect of radio frequency oscillating quadrupole field and the constant DC potential applied to the parallel rods. Ions are initially accelerated in the vacuum along the length axis of the quadrupole (z axis) and inside the quadrupole by the electric fields in the x and y direction as well, which plays a key role in focusing the ion beam A positively charged ion inside the quadrupole will be attracted towards the rod of negative polarity. However the potential changes with a radio frequency, thus changing the direction of the ion motion, which eventually results in the ions being focused along the z axis. At any given time, only a small subset of ions with a narrow window of m/z values has a stable trajectory in the quadrupole for the given set of parameters. These parameters when plotted present a few regions (called stability areas) in the graph (called stability diagram) that allow the ions not to hit the rods where they would become discharged. In order to obtain a full spectrum, these parameters needs to be swept through a range of values to scan ions of different mass to charge ratios while maintaining a constant ratio between the RF and DC potentials. The stability of any ion in the quadrupole is given by a set of equation of motion known as Mathieu's equations.

The quadrupole mass analyzer requires that the time for an ion to cross the analyzer is shorter than the time used to switch from one mass to the other and that the ions stay long enough inside the quadrupole to encounter at least few oscillations of the rod potentials. Apart

from being a mass analyzer, the quadrupole can serve many other purposes in a mass spectrometer[92]. For example, the quadrupole may just be used as a collision cell for tandem mass spectrometry, or as a transmission device or as an ion guide to transport ions from one part of the mass spectrometer to the next and rather than serving as a mass filter to select for ions of specific mass to charge.

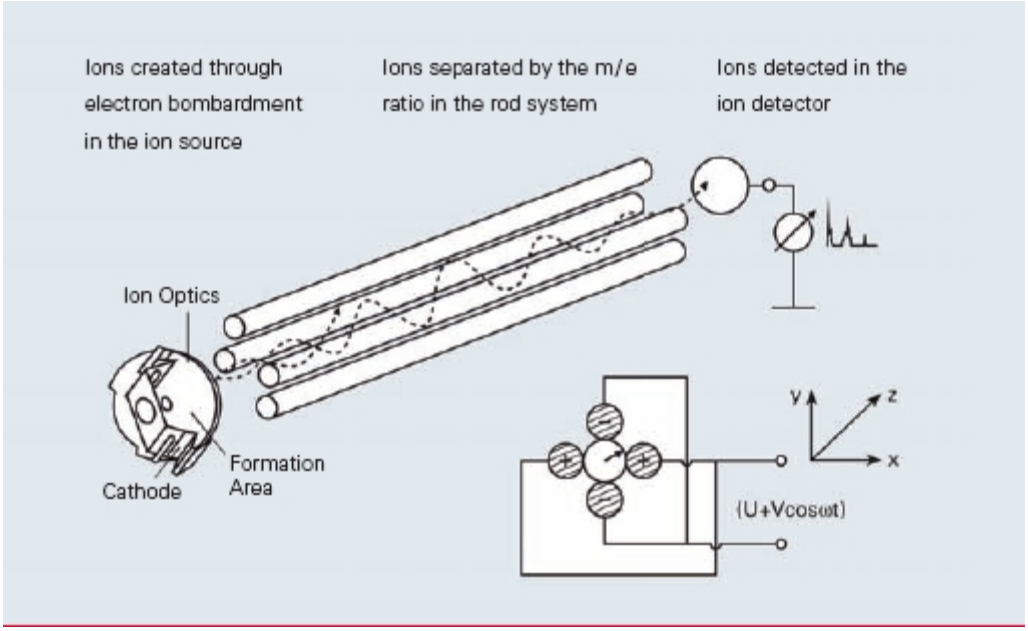


Figure 1.2.2.3 A quadrupole mass filter. The ions that are stable for the given set of voltages are focused in the x, y direction towards the center of the quadrupole and moves along the z-axis. Adopted from[79]

1.2.2.4 Quadrupole ion trap analyzer

The quadrupole ion trap is a by-product of quadrupole technology pioneered by Wolfgang Paul. Quadrupole ion traps can be either 3D or linear ion traps. These ion traps are called quadrupole ion traps in order to distinguish them from the magnetic ion traps which include FT ICR and Orbitrap analyzers. The 1989 Nobel Prize in physics was shared between Wolfgang Paul and Hans Georg Dehmelt, for their contributions on ion trapping techniques through the development of quadrupole and magnetic ion traps. Since the ion trap works based on the Paul's principles, they are also called Paul traps.

In ion trap instruments the electric field is three dimensional affecting the ion movement in all directions. This confine the ions within the analyzer whereas in the quadrupole analyzer the electric field is two dimensional acting on the x, y directions and the ions can travel in the z-direction along the axis of the analyzer. The movement of ions both in 3D ion trap and linear ion trap are also governed by the Mathieu equations. The linear ion trap can store more ions compared to the 3D version[93] and is quite commonly used in tandem mass spectrometry together with very high accuracy and high resolution mass analyzers like the Orbitrap and FT ICR.

In contrast to most other mass spectrometers, the ion traps are operated at relatively high pressure in the range of 10^{-2} Pa. The high pressure is maintained inside the trap by a constant flow of gas (nitrogen or helium). The gas acts like a cushion and slows down the fast moving ions contributing a dampening effect. The gas helps in improving the trapping efficiency and mass resolution. The damping gas also has a role to play in activation and fragmentation of ions in collision induced dissociation (CID). In such a hybrid fashion in the linear ion trap Orbitrap instrument the ion trap is usually employed for fragmentation and fragment mass spectrum acquisition while the other analyzer simultaneously acquires a full mass spectrum with very high resolution and accuracy.

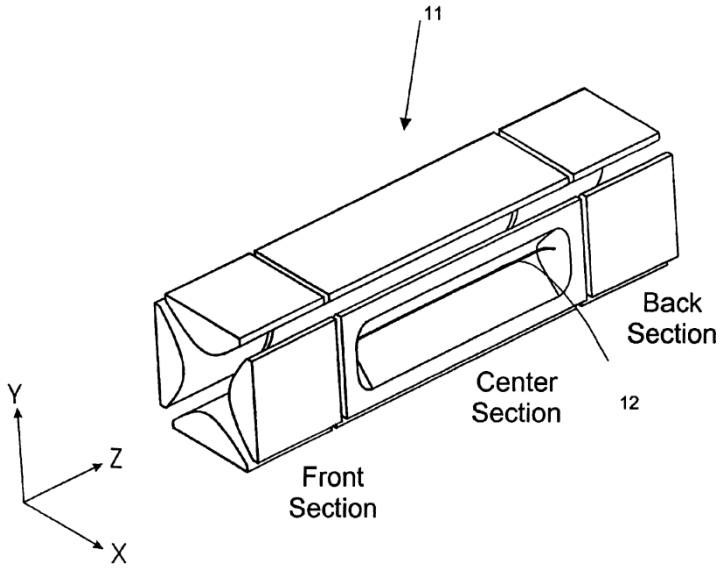


Figure 1.2.2.4 Linear ion trap mass analyzer. The center section has a slit that facilitates the axial ejection of ions. Modified from ref[94]

1.2.2.5 The Orbitrap analyzer

The Orbitrap analyzer is one of the latest developments in mass spectrometry and was invented by Alexander Makarov and colleagues[18, 95]. The Orbitrap derives its basic design from the trap device described by K. H. Kingdon in early 1920s that was named Kingdon trap. Interestingly as Makarov pointed out in the lecture at International Mass Spectrometry Conference, IMSC 2009, Bremen, the construction of the Orbitrap faced so many daunting challenges that the project appeared doomed many times. One such case was the problem of efficient ion injection into the Orbitrap, which was solved by the introduction of a 'C-trap' or a curved linear trap device. The C trap 'squeezes' the ions in terms of space and time and shoots them into the Orbitrap through the z-lens perpendicular to the rectangular section plane of the Orbitrap analyzer.

The Kingdon trap, from which the Orbitrap derives its fundamental principles, employs only electrostatic field by applying an electric potential between an outer cylindrical electrode and an inner thin wire which acts as the central electrode. In the Orbitrap analyzer, in contrast,

the outer electrode is barrel shaped while the inner electrode is spindle shaped. As a result, the space between the two electrodes is not constant along the z-axis (length of the Orbitrap), implying that the electric field is weakest in the middle where the space between the two electrodes is largest. When the ions are injected in packages from the C-trap, they enter a circular motion owing to the interaction of centripetal and centrifugal forces generated from the tangential movement and the electric field between the electrodes. This electric field has two forms of heterogeneity. First the field strength is varying along the z- axis in opposite direction from the middle of the Orbitrap analyzer and second, the direction of electric field vectors from different points along the z-axis is not parallel to each other. This inhomogeneity results in the mass dependent oscillation of ions along the z axis simultaneous to the circular motion around the central electrode. This oscillation is the measure of the mass of ions in the field between the electrodes which is detected as image current by the electrically isolated sections of the outer barrel electrodes. This frequency of oscillation is independent of energy and spatial spread of ions.

Since the oscillating frequency is a direct measure of the mass of the ions and independent of the energy, the Orbitrap mass analyzer boasts a very high resolving power given that the frequency can be measured with very high precision. The Orbitrap analyzer has significantly higher ion trapping capacity compared to the quadrupole ion trap and the FT ICR instruments, and therefore much higher space charge tolerance. The Orbitrap mass analyzer has a very low mass deviation of routinely less than 3 ppm. However this mass accuracy of the Orbitrap requires very high vacuum as collisions with background molecules can cause dephasing of ions and thus deterioration of the mass accuracy and resolution. For this reason, ion activation by collision to neutral gas molecules is not generally possible in Orbitrap analyzers. The Orbitrap analyzer has comparative resolution to FT ICR without the need for cooling any superconducting magnets. In routine practice, the mass accuracy of the Orbitrap analyzer is further improved to sub ppm level by real time calibration with ions present in ambient air (lock masses)[96] which are frequently present in the spectra throughout the chromatography gradient.

$$\omega = \sqrt{\left(\frac{k}{\frac{m}{z}}\right)}$$

The mass dependent frequencies of ion motions are given by the above equation, where, ω is the frequency in rad/s and $k$is an instrumental constant.

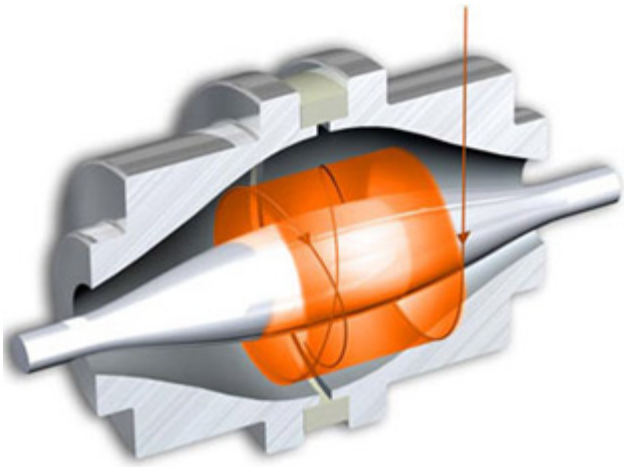


Figure 1.2.2.5 (a) Cross-sectional view of the Orbitrap. The ions move both along the axis and around the central electrode shaded in orange. The outer barrel electrode is split into two electrically isolated halves to detect the image current. Adopted from ref[97]

Orbitrap in a hybrid instrument

The Orbitrap analyzer is commercially available from Thermo Fisher Scientific in a tandem configuration coupled to a linear ion trap instrument. Since the Orbitrap has the single function of detection (in principle, the Orbitrap is nothing but an expensive detector!!) it cannot be used as a standalone device. It requires the C-trap for ion storage and injection.

In the hybrid configuration the full scans are usually acquired in the Orbitrap analyzer while simultaneously abundant peptide ions are selected, isolated, fragmented and analyzed in the ion trap. For the full scans, the ions are first accumulated in the ion trap and then axially ejected into the C-trap and subsequently into the Orbitrap. While the Orbitrap is acquiring the MS transients, the ion trap is programmed to perform several CID fragmentation events and to scan out the resulting peptide fragments to the multipliers by lateral ejection. In the latest version of the instrument (LTQ-Orbitrap Velos) the linear ion trap actually consists of two traps. This dual cell

ion trap is maintained at differential pressure so that the high pressure ion isolation and activation and low pressure scanning are performed[98].

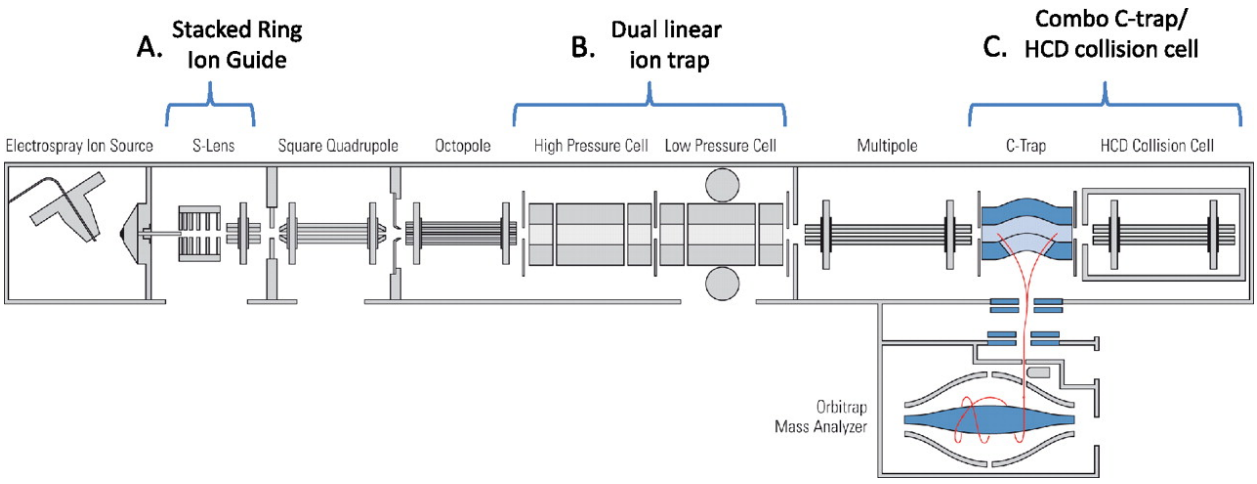


Figure 1.2.2.5 (b) The hybrid mass spectrometer configuration of the LTQ-Orbitrap Velos. The dual cell linear ion trap is designed for fast CID fragmentation scans while a high resolution full scan is acquired it the Orbitrap. Adopted from ref[98].

1.2.3 Different proteomic approaches and modes of operation

1.2.3.1 Top down and bottom up proteomics

In an MS-based proteomic experiment, the proteins can either be delivered to the mass spectrometer in an intact form or in the form of digested peptides produced by adding a proteolytic enzyme. Top down proteomics involves the analysis of whole proteins in the mass spectrometer and is a relatively young and immature field compared to bottom up proteomics where the peptides rather than proteins are introduced into the mass spectrometer.

Top down proteomics has the potential advantage that the entire sequence of protein is presented for analysis. This could enable distinguishing isoforms of proteins and to characterize the post translation modifications (PTMs) directly on the protein. However the top down approach suffers from many limitations. Firstly the ions generated are multiply charged resulting in highly complex MS/MS spectra to be deconvoluted. This implies that only high mass accuracy and high resolution instruments like FT ICR and Orbitrap analyzers can handle the this complexity, and these instruments are very expensive. Further in order to perform protein sequencing, the fragmentation techniques that are amenable for top down approach like electron capture/transfer dissociation (ECD/ETD) can be less efficient than CID of peptides. Moreover the fragmentation behavior of proteins is less understood compared to peptides. Separation techniques that are commonly employed before MS analysis to reduce sample complexity are challenging for top down MS because insoluble proteins are difficult to handle. Therefore top down proteomics is generally not used in a high throughput manner and seldom on proteins larger than 50 kDa.

In contrast, bottom up proteomics is a widely applied approach in variety of applications starting from simple mixtures to complex total cell and tissue lysates. The complex mixtures can be separated using different techniques including reversed phase, ion exchange chromatography, isoelectric focusing and others. For peptide sequence identification, the peptide ions are isolated in the mass analyzer, fragmented and the fragmentation spectra are usually searched against a database containing the theoretical fragmentation spectra. Unlike the top down approach, bottom up proteomics can be carried out in many different instrument configurations. The most

commonly used analyzers for peptide fragmentation includes quadrupole and ion trap analyzer, where the peptides are usually fragmented by collision induced dissociation (CID). The poor resolution capabilities of ion trap are well compensated by the high speed and sensitivity of fragmentation. The major advantages of bottom up proteomics include the possibility of automation of separation techniques prior to mass spectrometric analysis (for eg., reversed phased chromatography), tailor made software and instrumentation available and robust quantification techniques well suited to this approach. One of the major problems in bottom up proteomics is assigning the identified peptides back to proteins. In many cases since only a part of the protein sequence is covered by the identified peptides, protein isoforms become indistinguishable. This makes analysis difficult for proteins whose isoforms have different and roles and different cellular localization. For the same reasons some of the crucial PTMs might be missed in single experiments or they may be entirely undetectable because they are located in unfavorable sequence contexts for the proteases employed.

1.2.3.2 Tandem mass spectrometry and Ion fragmentation in bottom up proteomics

As mentioned above peptides are fragmented in tandem mass spectrometry to decipher the peptide sequence. Tandem mass spectrometry can be performed in two ways namely tandem in space mass spectrometry and tandem in time mass spectrometry. As the name suggests, the tandem in space mass spectrometry involves isolation of peptide ion in one analyzer followed by activation in the second analyzer and finally detection in the third analyzer. Typical examples include the TOF-TOF and triple quadrupole configurations. By its nature, in space separation places a limit on the number of MS/MS events that can be sequentially performed as for each MS/MS event additional analyzers would be required. Furthermore the transmission efficiency will keep decreasing with increasing numbers of analyzers.

Tandem in time separation involves isolation, activation and detection of ions in the same analyzer however in a sequential manner. Tandem in time mass spectrometry is typically performed in ion trap and FT ICR instruments. For in time separation typically up to 6-7 MS/MS cycles can in principle be performed. However, as the fragmentation cycles increase the size of the ion population becomes smaller and smaller, eventually making analysis impossible. In the Paul type ion trap analyzers the ions have to be ejected to be detected and hence can be observed

only once whereas in FT ICR, the fragments are analyzed non destructively and thus can be observed continuously through the cycle.

In tandem mass spectrometry the generated fragments ideally constitute a ladder similar to the ladder generated in DNA sequencing, which can be read from high mass region to low mass region of spectra and vice versa with different ion series. The types of fragment ions observed in tandem mass spectrometry are influenced by peptide sequence, amount of energy used, how the energy is transferred, charge state, the instrumentation used for fragmentation among other factors. The peptides can be fragmented in several different places apart from its peptide bond (CO=NH) making the phenomenon complex. A common nomenclature for the fragment ions was proposed and it is still in general use[99] (shown in the figure below). The a, b and c ions retain a net positive charge on the N-terminal part of the peptide whereas the x, y and z ions retain the charge in the C-terminal part of the peptide. The nomenclature can be further extended for the cleavage at other bonds but this in not shown here.

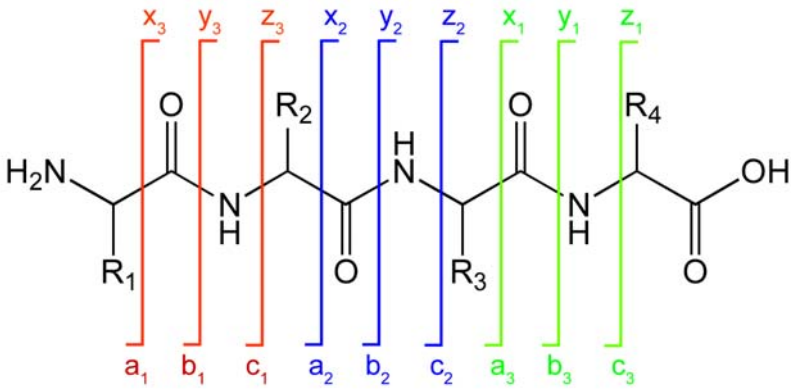


Figure 1.2.3.1 Peptide fragment ions nomenclature. (Adopted from Wikipedia)

1.2.4 Ion activation and dissociation methods for tandem mass spectrometry

The peptides mass alone will not suffice to determine the identity of the peptide and activation and fragmentation of the peptides is therefore essential[100]. The fragment peaks or product ions ideally allow reading out the amino acid sequence of the peptide. Ion activation methods involve increasing the internal energy of ions resulting in fragmentation of these ions to yield structural information. There are many different dissociation techniques several of which require specific instrumental configurations. Collision induced dissociation (CID), surface induced dissociation (SID) dissociation from absorption of electromagnetic radiation, electron capture dissociation and electron transfer dissociation are examples of dissociation methods. The absorption of electromagnetic radiation can be subclassified into ultraviolet photo dissociation (PD), infrared multiphoton dissociation (IRMPD) and blackbody induced radiative dissociation (BIRD).

Collision induced dissociation

Collision induced dissociation is by far the most commonly used dissociation method that is used in peptide tandem mass spectrometry. It is one of the oldest techniques and used from the late 1960s. CID involves two steps namely collisional activation and unimolecular dissociation[101]. In the collisional activation step, the selected peptide ions are collided against inert gas molecules like helium, nitrogen or argon and thus vibrationally excited. If during this process a fraction of the peptides have internal energy in excess of that needed to break bonds this will lead to a fragmentation process[102]. The kinetic energy that is converted to internal energy depends on the mass of the collision partners, for example helium imparts less energy per collision than argon. Fragmentation of peptides by CID usually cleaves them at the CO=NH bond and hence b and y ions are predominant in the spectra.

Depending on the instrumental configuration, CID can be performed either at high energy or low energy mode. Higher energy CID usually occurs in TOF and magnetic sector instruments with translational energies up to 10 keV and is characterized by very short activation times. The lower energy CID mode employs energy less than 200 eV in quadrupole instruments. In the ion

traps, multiple, low-energy collisions occur over relatively long activation times - typically 30 ms in the linear ion trap instruments. In the lower energy CID mode in quadrupoles, using larger gas molecules may be preferred as the amount of kinetic energy transferred is higher with larger gas molecules. Furthermore, when the gas molecules are larger like Xe, the cross sectional area of the gas atom is larger and thus the probability of collision is higher. The ladder of b and y ions obtained by CID fragmentation is usually sufficient in order to identify the peptide sequence by matching to a theoretical spectrum. In low energy CID fragmentation in ion traps, both b and y ions are observed prominently whereas in beam type instruments the b ions tend to fragment further resulting in y ion dominated spectrum.

Electron capture dissociation

ECD is an alternative fragmentation technique to CID which is based on the reaction of low energy electrons with multiply protonated peptides[103]. The electrons are captured by multiply-protonated peptides, and as a result the peptides undergo partial charge neutralization generating radical species in an excited state. The radicals within a very short time period undergo bond cleavage producing mostly c and z fragment ions[104]. The cleavage is very bond specific owing to the presence of radicals and disulphide bonds and halogen bonds dissociate with the highest rates[105]. The retention of labile PTMs is superior in ECD compared to the conventional MS/MS with CID[106, 107]. However, the loss of signal owing to charge neutralization in ECD makes may make it less attractive in large scale tandem mass spectrometry. In general ECD is thought to give complementary data to the conventional CID, for example in de novo sequencing, PTM and disulphide bond mapping and in top down mass spectrometry[108].

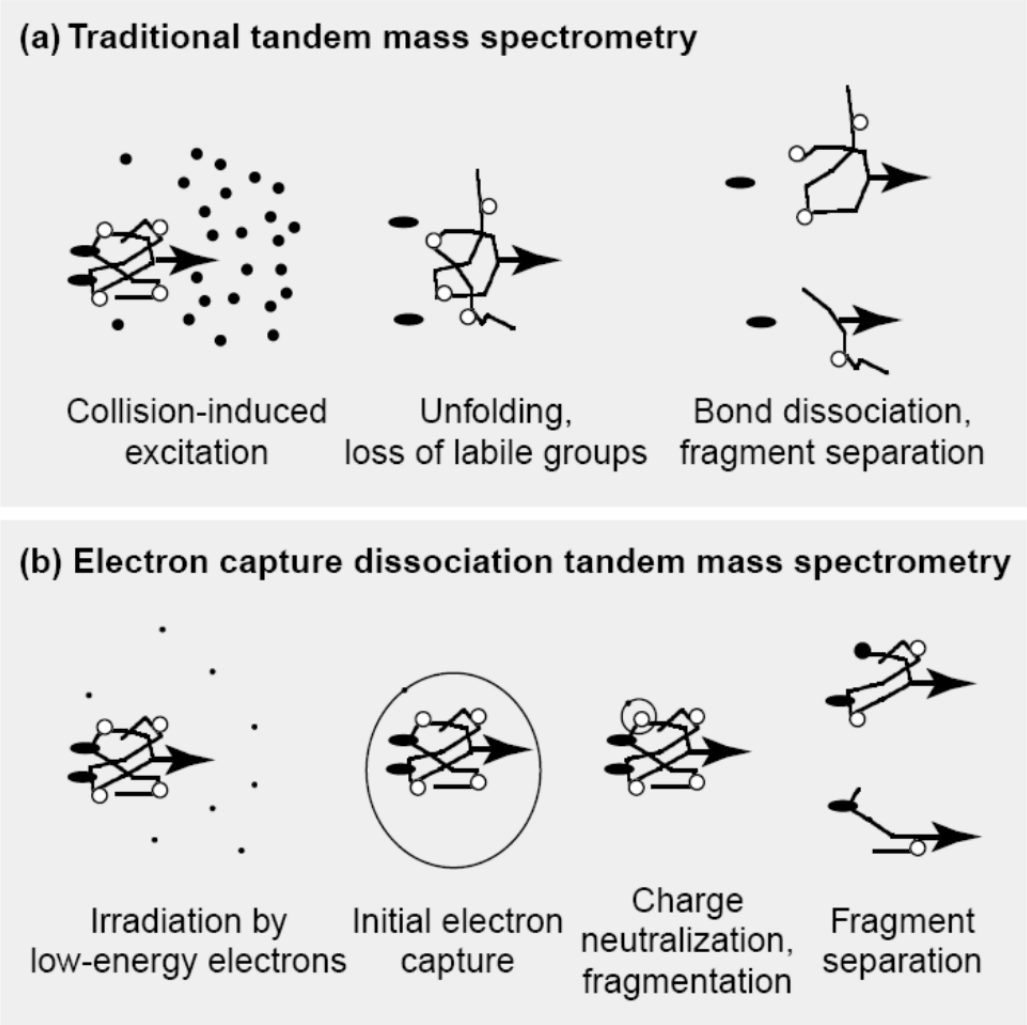


Figure 1.2.4 (a) Comparative illustration of tandem mass spectrometry by CID and ECD. Adopted from[103]

Electron transfer dissociation

Although ECD provides reliable and efficient mapping of PTMs in peptides it is not feasible to mix the peptides with low energy electron in a majority of trapping mass spectrometers that lack magnetic fields. The ion trap that uses radio frequency oscillating electric fields can trap thermal electrons only for few microseconds which is not sufficient for an efficient reaction to occur[109]. Further ECD often required averaging of spectra acquired over minutes precluding its use in large scale tandem mass spectrometry experiments[110]. Electron Transfer Dissociation (ETD) is similar to ECD, but fluoranthene radical anions supply electrons

for the reaction, making ETD feasible in quadrupole ion traps. It was developed in the Hunt lab in 2004[109]. ETD has also been proposed to tackle the problem of highly charged peptides and proteins and hold promise in top down and PTM analysis.

Higher energy collisional dissociation (HCD)

It is well known that CID fragmentation in an ion trap configuration can result in the loss of labile PTMs like phosphorylations of serine and threonine residues. However, numerous advances like improved scan functionality[111, 112], injection of externally formed ions[113, 114], extended mass range capabilities, use of helium buffer gas, the control of number of ions in the trap called "automatic gain control" propelled the popularity and wide-spread use of ion trap based tandem mass spectrometry in proteomic research. The loss of phosphorylation and any similar modification from the peptide during CID in ion trap was elegantly circumvented by the application of pseudo-MS^3 which is otherwise known as 'multi-stage activation'. However one more limitation of the ion trap fragmentation includes the 'one-third rule', where the low mass fragment ions are not retained in the ion-trap. This is a severe limitation for quantitative proteomics that employ low molecular mass reported ions that are isotopically labeled like in iTRAQ quantification.

Recently higher energy collisional dissociation was introduced in which the ions are accumulated and fragmented either in the C-trap or in the adjacent collision cell present in the LTQ-Orbitrap instrument configuration[115]. The fragment ions are then sent into the Orbitrap analyzer for detection. The main implications of this technique are first, that the fragment mass accuracy is at the ppm level facilitating stringent database searches and *de novo* sequencing. Second, the collision energy is higher than the energy used for ion trap CID resulting in a 'triple quadrupole like' fragmentation. Together this means that there is no loss of low molecular weight reporter ions, and the fragmentation spectra are dominated by y ions because the b ions undergo further fragmentation and only low molecular weight b ions are observed in the spectrum.

HCD (or CID with detection in the Orbitrap analyzer) need about tenfold more ions than CID in the ion trap. In the new LTQ-Orbitrap Velos, this fragmentation mode is feasible on a large scale because of greatly improved ion current of the instrument[98].

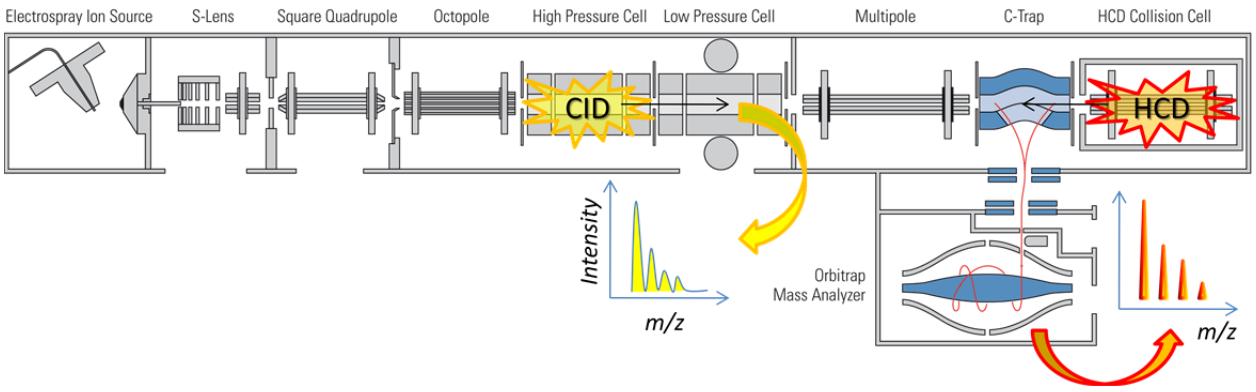


Figure 1.2.4 (b) Schematic representation of HCD fragmentation in a LTQ-Orbitrap Velos instrument.

1.3 MS proteomics for urinary biomarker discovery platform

1.3.1. Urine as a source of biomarkers

One of the major functions of body fluids is to serve as a means of transport of chemicals, enzymes, metabolic waste to the appropriate destinations. Thus these body fluids like plasma, cerebrospinal fluid (CSF) or urine reflect the physiological status of the living system and consequently have been being exploited for clinical examination and disease diagnosis for decades.

The ready availability of urine by non-invasive collection makes it an attractive source for biomarker discovery. This is in contrast to other body fluids such as cerebrospinal fluid (CSF) or even blood. Intuitively, urine could be an ideal source of biomarker discovery with high relevance to patho-physiological conditions of urogenital and associated proximal systems like renal failure, bladder cancer, and prostate cancer. Even attempts to find urinary biomarkers for conditions that are not directly linked to kidney and proximal organs are not scarce in the literature. Furthermore, urine samples from bio banks are readily available for pathological conditions along with associated patient history. Thus the potential advantages of urinary proteomic biomarker discovery and the already existing annotated urine sample collection has made the proteome analysis of urine samples a very attractive proposition.

Urine is produced in the kidney as part of maintaining whole body homeostasis. The kidney has a major role in removing metabolic waste from the body by filtering the blood plasma. It maintains the homeostasis of the body by regulating the water and electrolyte content and maintaining the acid base equilibrium of the body. More than 150 L of plasma are filtered in the kidney corresponding to a rate of 125 mL/min and on average 1.5 L of urine is produced per day. The entire process of urine formation is carried out by nephrons or malpighian bodies that are the fundamental units of the kidney; each kidney is made up of up to 1.3 million nephrons. These nephrons consist of a capsule like structure called glomerulus or Bowman's capsule and a

tubular structure called the renal tubules. The entire process of urine formation occurs in three steps (i) glomerular filtration (ii) tubular reabsorption and (iii) tubular secretion following which the urine is collected and stored in the bladder until excretion. Because of this extensive filtration and reabsorption process the protein content in the final urine is much diluted compared to plasma. Approximately 150 mg of proteins are excreted through urine in a normal individual per day. Increased protein content in the urine is termed proteinuria; a physiological condition indicating the malfunctioning of kidney in glomerular filtration or in the tubular reabsorption process.

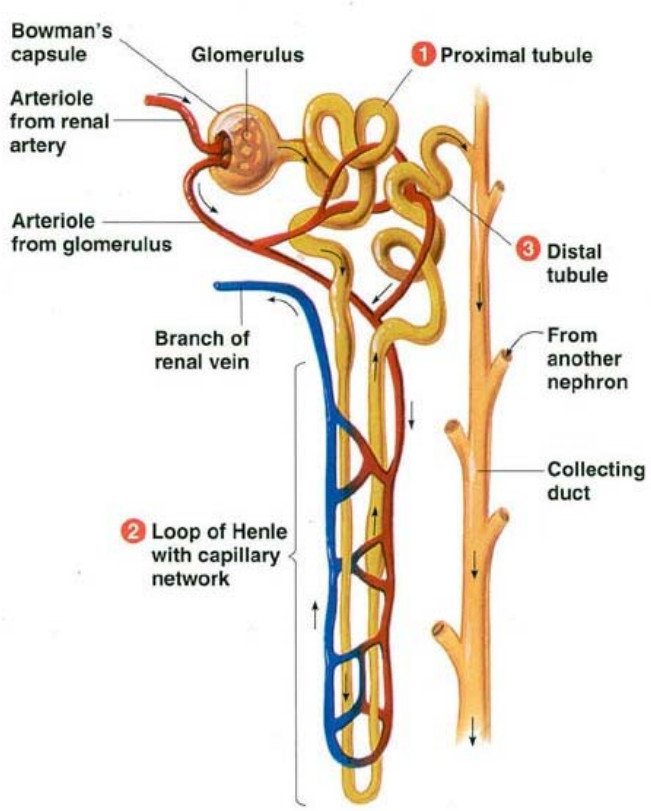


Figure 1.3.1 Diagram of nephron the fundamental unit of kidney involved in urine production. Adopted from ref[116]

1.3.2 Urinary proteome map

Second only to plasma, urine is a source of diagnostic molecules (biomarkers) for a wide range of diseases[116] therefore urine should be an attractive source for biomarker discovery[117-119]. Though serum/plasma is a preferred source for biomarker discovery the dynamic range problem makes analysis very difficult. Since the majority of the plasma proteins are filtered out, urine should be a simpler mixture to handle compared to plasma. It is thought that potential biomarkers relevant not only to kidney, proximal and urogenital tract but also to distal organs like brain and lung are present in urine. In spite of numerous attempts in identifying the proteins in urinary proteome[120-123], until recently the number of proteins in urine appeared low at under 300[118, 124]. One of the studies on urinary exosome, membrane bound vesicles present in this body fluid, reported the presence of 1000 proteins[125]. The first large scale urinary proteome map employing high resolution mass spectrometry, reported the presence of more than 1, 500 proteins in urine[126]. It was observed that a large proportion of urinary proteins are membrane proteins. A recent study employing extensive fractionation techniques in addition to employing similar high resolution mass spectrometry has even reported more than 2,300 proteins in urine. Together all these studies indicate that at least 3,000 proteins in urine. Furthermore, the urinary proteome consists of significant population of proteins that have a molecular weight larger than 40 kDa. This was unexpected as glomerular filtration of the kidney was thought to removed all proteins above this weight. The presence of many disease related proteins in urine further accentuates the potential of finding new biomarkers in urine.

1.3.3 Source of proteins in urine and relevance of urinary biomarkers to different patho-physiological states

The main sources of proteins in urine include plasma, cells and membrane components from cell debris from epithelial lining in urine. The proteins derived from plasma are soluble reportedly constituting nearly 50% of the protein source. The proteins of the cell debris are usually sediment at low speed centrifugation and in contrast to those derived from exosomes that sediment only under very high centrifugation speeds. A study by Zhou et al reported that the

exosomes contribute only about 5% of the total urinary protein content[127]. Current literature suggest that the urinary proteome can aid in the investigation of non-cancerous urogenital conditions, cancer pathology of urogenital system and non urogenital conditions, as described below.

Urogenital disease: non-cancerous

One of the main interests of urinary biomarker discovery is in the area of kidney transplantation. Avoidance of complications in kidney transplantation like acute rejection could play a pivotal role in the survival of renal transplant patients and in the organ resource management and should in principle be possible by urinary proteomics[128, 129]. SELDI is a low resolution from of MALDI using a linear TOF instrument without sequencing capability, and this somewhat controversial technology has been applied to urinary proteomics. In three independent reports, potential protein biomarkers have been described that could predict acute allograft rejection in kidney transplant patients[130-132]. Notably, the three studies came up with different sets of biomarkers presumably due to the different chip surfaces used in the SELDI technique and the different instrumental settings used. The potential biomarkers identified in these studies were not tested against separate patient cohorts for validation. Limitations of the SEDI method are further described below.

Capillary electrophoresis (CE) has also been used to identify a peptide signature that could differentiate patients with evidence of different levels of rejections and patients with no evidence of rejection[133].

Chronic kidney disease (CKD) is a condition where kidney function deteriorates over time primarily due to previous conditions like diabetes and high blood pressure[134]. Detection of a reliable and sensitive biomarker pattern for prediction of CKD is of prime importance given the wide prevalence of the condition, and complications that occur in other organs because of CKD. Few studies have been undertaken in proteomic biomarker discovery for CKD and they are dominated by SELDI and CE platforms[135, 136]. However, these studies are not completely global (or 'unbiased') as only a subset of proteins binds to the SELDI surface and neither SELDI nor CE are ideal techniques for analysis of polypeptides of more than 20 kDa. Preliminary studies

aimed at finding biomarkers for diabetic nephropathy and obstructive nephropathy have also been undertaken[135, 137-140].

Urogenital cancers

The most common examples of urogenital cancer include renal cell carcinoma (RCC), bladder cancer and prostate cancer. Biomarker discovery studies for RCC have been carried out by SELDI TOF technology however no significant marker discovery has been reported. Bladder cancer is one of the top 5 malignancies in the USA. There are only few proteomics studies of this disease and they have only lead to a list of candidate biomarkers that have not been validated.

Non-urogenital disease

In addition, reports have identified markers for non-urogenital diseases related to distal organs. Clinical follow up profiling of patients after allogeneic hematopoietic stem cell transplantation revealed a pattern that is currently being followed up in a larger population [141]. Zimmerli et al identified peptide patterns that may identify coronary artery disease (CAD)[142]. This peptide patterns was reported to be more than 90% specific and sensitive.

1.3.4 Evaluation of technology applied for biomarker discovery

In summary, different technological platforms have been used for urinary proteome biomarker discovery. Most investigations used capillary electrophoresis, SELDI and a few used low resolution LC-MS/MS[135, 143-150]. However, to my knowledge, the candidate biomarkers that came out of such studies have are not been validated. The 2 DE technique used in urinary biomarker discovery has many limitations as discussed earlier. Low resolution LC-MS/MS approaches cannot quantify large numbers of proteins and peptides owing to the overlap of co-eluting peptide species in complex mixtures. The SELDI technology is in principle a high throughput technique suitable for analysis of large number of samples. However SELDI based techniques lack robustness in terms of quantification and reproducibility between sample plates. Furthermore, SELDI similar to protein microarrays has limited depth of coverage.

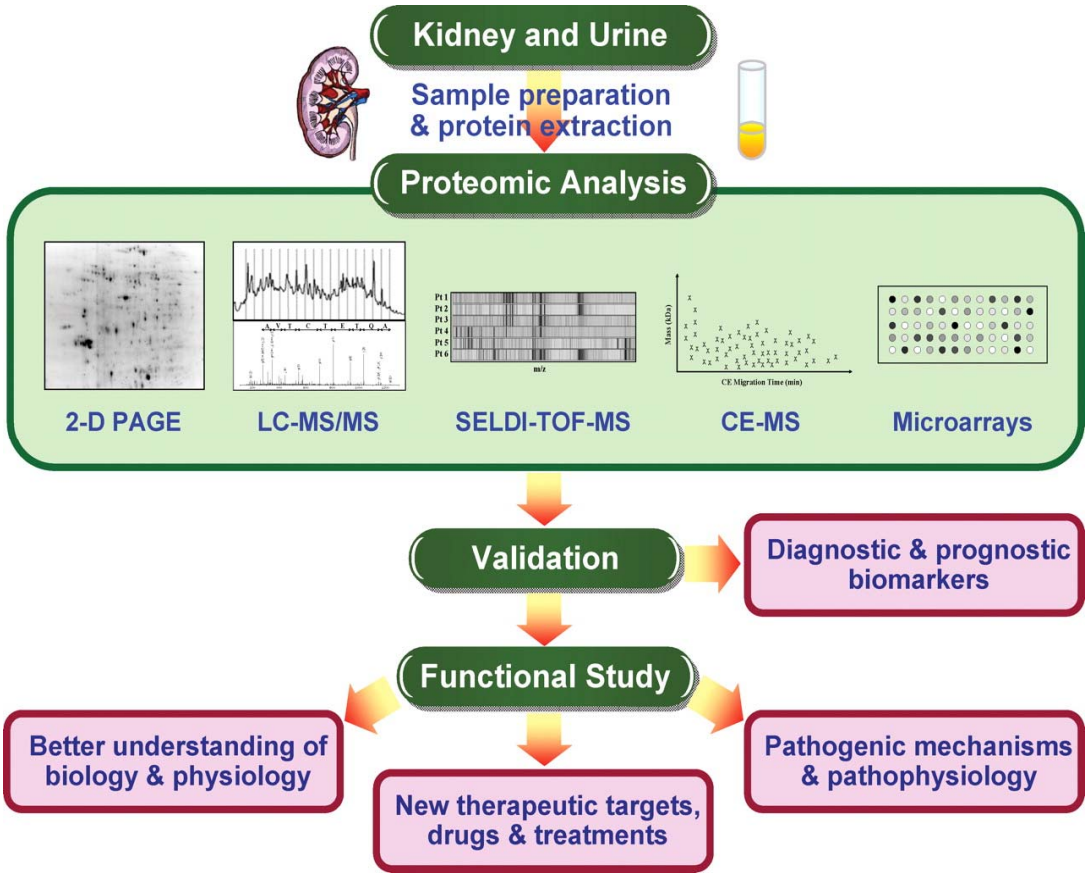


Figure 1.3.2 Technological platforms applied to urinary biomarker discovery. Adopted from ref[151]. Only one of these techniques (LC MS/MS) is a high accuracy, quantitative biomarker platform.

Capillary electrophoresis has dominated the platform for early biomarker discovery in urine. CE is a peptidomic technology and lacks essential information on most full length proteins. Furthermore, like SELDI, most of the CE platform based biomarker studies did not involve MS/MS and therefore need a second round of LC-MS/MS experiments to identify the sequence of the candidate peptide biomarkers.

2. Projects / Manuscripts

1. Analysis of detergent solubilized membrane proteome by LC-MS/MS
2. Universal sample preparation method for proteome analysis
3. Large scale phosphoproteomics using HCD
4. Variation of normal human urinary proteome

2.1 A detergent based method for efficient analysis of membrane proteome

Plasma membrane proteins play a crucial role in relaying signal information from the exterior into the cell and vice versa and proteins present in organellar membranes have analogous functions. Membrane proteins constitute a significant proportion of the membrane structure[152]. Their analysis by mass spectrometry is of prime importance in clinical proteomics, given the association of cancer phenotypes with alterations in membrane proteins. However, studying membrane proteins is technically challenging even in bottom up proteomics, a main difficulty being the solubilizing the membrane proteins for enzymatic digestion. Unfortunately, while detergents are the best solubilizing agents for membrane proteins, they are notorious contaminants in LC-MS/MS experiments. The detergent from the sample is therefore usually removed by SDS-PAGE separation followed by in-gel digestion. In this project, we show that detergents can be used to solubilize the membrane proteins but later removed by urea displacement prior to in-solution protease digestion. By performing in-solution rather than in-gel digestion it was possible to identify several hundred membrane proteins in a high throughput manner. After detergent removal and by digesting in solution we obtained almost twice the number of proteins as were identified by the conventional in-gel method with concomitant improvements in the sequence coverage of the proteins.

This work is published as a research article in the "Journal of proteome research" as follows.

research articles Journal of proteome research

Detergent-Based but Gel-Free Method Allows Identification of Several Hundred Membrane Proteins in Single LC-MS Runs

Nagarjuna Nagaraj,[#] Aiping Lu,[#] Matthias Mann,* and Jacek R. Wiśniewski*

Department of Proteomics and Signal Transduction, Max-Planck Institute for Biochemistry, Martinsried, Germany

Received June 6, 2008

Detergents are indispensable solubilizing agents in the purification and analysis of membrane proteins. For mass spectrometric identification of proteins, it is essential that detergents are removed prior to analysis, necessitating an in-gel digestion step. Here, we report a procedure that allows use of detergents and in-solution digestion of proteins. Crude membrane preparations from mouse brain were solubilized with Triton X-100, CHAPS, or SDS, and the detergents were depleted from the membrane proteins using a desalting column equilibrated with 8 M urea. Following digestion with endoproteinase Lys-C, the resulting peptides were analyzed by LC-MS/MS on Linear ion trap-Orbitrap instrument. Applying stringent identification criteria, in single-LC-MS-runs, 1059 ± 108 proteins, including 797 ± 43 membrane proteins, were mapped from mouse brain. The identified proteins represented a broad spectrum of neurotransmitter receptors and other ion channels. The general applicability of the method is demonstrated by profiling of membrane proteins from four other mouse organs. Single-run analyses of eye, liver, spleen, and skeletal muscle allowed identification of 522 ± 9, 610 ± 7, 777 ± 8, and 307 ± 7 membrane proteins. Our results demonstrate that membrane proteins can be analyzed as efficiently as soluble proteins.

Keywords: Membrane proteomics • integral membrane proteins • detergent removal • brain • liver • eye • spleen • muscle • LTQ-Orbitrap

Introduction

The use of detergents in biochemical research ranges from standard procedures, such as SDS-PAGE or pull-down experiments, to complex, specialized applications, such as extraction of integral membrane receptors consisting of multiple subunits. In the field of membrane biochemistry, detergents are indispensable tools for solubilization and fractionation of membrane proteins. However, detergents, even in small concentrations, dominate mass spectra and preclude peptide or protein analysis. Thus, in mass spectrometry (MS)-based proteomics, detergents have to be efficiently and thoroughly removed from proteins or peptides prior to analysis, but this is not an easy task. Different methods have been described for separation of proteins from detergents including gel filtration, ion-exchange and hydrophobic adsorption chromatography, density gradient centrifugation, dialysis, ultra filtration, phase partition, and precipitation (for a review see ref 1). However, they have not become popular in mass spectrometry because of their inability to completely remove the detergents.[2] Moreover, these methods can lead to substantial sample losses as they have been designed to deal with relatively high protein amounts; thus, their applicability to proteomics is limited.

To circumvent the difficulties with removal of detergents, alternative approaches avoiding the use of detergents have been proposed. For example, 60% methanol[3,4] was used for the solubilization of membranes and the extracted proteins were digested with trypsin. In another approach, membranes were solubilized with 90% formic acid and the proteins were chemically cleaved with cyanogen bromide.[5] In addition, digestion of membrane proteins directly in a suspension of fractions enriched in membranes has been described. Wu et al. used proteinase K at high pH to digest protein chains protruding from the membrane bilayer.[6] Using a related concept, we analyzed mouse brain membrane proteins by digesting purified plasma membranes in 4 M urea with endoproteinase Lys-C.[7] We further applied this 'solid-phase digestion' strategy in protein profiling[8] and comparative, semiquantitative mapping of plasma membrane proteins between distinct regions of mouse brain.[9,10]

Despite these developments, detergents are preferred due to their strength in membrane solubilization and are widely used in sample preparation for subsequent proteomic analysis. Unfortunately, so far the only method to efficiently remove detergents once they were introduced involved in-gel digestion after SDS-PAGE or, alternatively, incorporation of detergent-containing protein lysates into a polyacrylamide matrix without electrophoresis.[11]

In this work, we present a novel procedure for detergent removal and digestion of membrane proteins, and compare it with an in-gel protein immobilization and digestion procedure.

* Correspondence address: Jacek R. Wiśniewski and Matthias Mann, Department of Proteomics and Signal Transduction, Max-Planck Institute for Biochemistry, Am Klopferspitz 18, D-82152 Martinsried near Munich, Germany, E-mail, (J.R.W.) jwisniew@biochem.mpg.de, (M.M.) mmann@biochem.mpg.de; fax, +49 89 8578 2219.

[#] These authors contributed equally.

Published on Web 10/08/2008

10.1021/pr800412j CCC: $40.75

Detergent Removal

We show that membrane proteins solubilized with Triton X-100, CHAPS, or SDS can be efficiently separated from the detergents by 'desalting' on a column equilibrated with 8 M urea. In this protocol, the proteins are digested with endoproteinase Lys-C and the resulting peptides are bound to a C_{18} membrane and analyzed by LC-MS/MS. Our method results in an almost 2-fold increased protein identification in comparison to the in-gel based approach. We demonstrate that our procedure is useful for profiling of membrane proteins from various tissues including mouse brain, liver, spleen, eye, and muscle tissues.

Materials and Methods

Membrane Preparation and Protein Solubilization. Frozen mouse brain, liver, spleen, leg muscle and eye were purchased from Pel-freez Biologicals, Rogers, AR. Membrane preparation was carried out as described previously.[8] Briefly, 20 mg of tissue was homogenized in 1 mL of high salt buffer (2 M NaCl, 10 mM HEPES-NaOH, pH 7.4, 1 mM EDTA) using IKA Ultra Turbax blender at maximum speed for 20 s. The suspension was centrifuged at 16 000*g*, at 4 °C for 15 min. The resulting pellet was re-extracted twice with carbonate buffer (0.1 M Na_2CO_3, pH 11.3, 1 mM EDTA) as above. After incubation for 30 min, the pellet was washed with urea buffer (4 M urea, 100 mM NaCl, 10 mM HEPES/NaOH, pH 7.4, and 1 mM EDTA). Following urea wash, the pellet was solubilized in 50 μL of 100 mM sodium phosphate buffer, pH 7.0, containing either 2% (w/v) SDS, 0.5% (v/v) Triton X-100, or 3.5% (w/v) CHAPS.

Removal of Detergents and Protein Digestion with Endoproteinase Lys-C. Detergents were removed on a HiTrap desalting column (5 mL, Amersham Biosciences, Uppsala, Sweden). The column was equilibrated with 8 M urea, 25 mM Tris-HCl, pH 8.0. The proteins were digested with 0.5 μg of endoproteinase Lys-C from Wako (Richmond, VA) at 25 °C overnight. Digestion was terminated by addition of 1% (v/v) trifluoroacetic acid. The digested peptide mixture was purified and stored in C_{18} StageTips as described.[12] Usually, 5% of the digestion mixture was loaded on a StageTip containing two membrane plugs.

In-Gel Digestion. Detergent solubilized membrane preparations were mixed with sample buffer, loaded on NuPAGE 4−12% Bis-Tris gel (Invitrogen, Carlsbad, CA), and stacked in the gel by electrophoresis at 50 V for 15 min. The gel was stained with Coomassie blue staining kit (Invitrogen), and entire lanes (usually 0.5 cm in length) were excised and in-gel digested as described.[13,14]

Mass Spectrometric Analysis. Protein digests were analyzed by online capillary LC-MS/MS. The LC-MS/MS setup was similar to that described before.[15] Briefly, samples were separated on an in-house made 15 cm reversed phase capillary emitter column (inner diameter 75 μm, 3 μm ReproSil-Pur C18-AQ media (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany)) using 120 min gradients and an Agilent 1100 nanoflow system (Agilent Technologies, Palo Alto, CA) or 90 min gradients using a Proxeon EASY-nLC (Proxeon Biosystems, Odense, Denmark). The LC setup was connected to an LTQ-Orbitrap (Thermo Fisher, Bremen, Germany) equipped with a nanoelectrospray ion source (Proxeon Biosystems). The mass spectrometers were operated in data-dependent mode. Survey MS scans were acquired in the orbitrap with the resolution set to a value of 60 000. Up to 5 most intense ions per scan were fragmented and analyzed in the linear ion trap. For accurate mass measurements, the lock-mass option was employed.[16]

Peak List Generation, Database Searching and Validation. The raw files were processed with MaxQuant, an in-house developed software suite (version 1.0.6.3).[17,18] The peak list files were searched against decoy IPI-mouse database version 3.24 containing both forward and reversed protein sequences, by the MASCOT search engine.[19] The initial parent and fragment ion maximum mass deviation[20] were set to 7 ppm and 0.5 Th, respectively. The search included variable modifications of oxidation of methionine and protein N-terminal acetylation. Peptides with at least seven amino acids were considered for identification and proteins with two or more peptides (at least one of them unique to the protein sequence) were considered valid hits. The false discovery rate for both the peptides and proteins were set a threshold value of 0.01. All proteins identified in this study are listed in Supplementary Table 1 in Supporting Information.

Bioinformatics Analysis. Gene ontology analysis of the identified proteins were performed using the Protein Center platform (Proxeon Biosystems, Odense, Denmark) primarily for cellular component analysis and membrane and transmembrane domain annotations.

Results and Discussion

Previously, we have described a detergent-free method for proteomic analysis of membrane proteins.[7] In that method, membranes are directly digested with endoproteinase Lys-C and the released peptides are analyzed using one-[7,9] or two-dimensional LC-MS/MS.[8,10,21] Even though this method is a powerful tool for mapping of membrane proteins, it has some limitations. Proteolytic digestion is performed on only partially denatured proteins (4 M urea); therefore, the yield of peptides is restricted by accessibility of the cleavage sites to proteases. Moreover, this method cannot be combined with chromatography techniques for separation of membrane proteins before digestion. We wished to develop a method that can be coupled with chromatographic separation like size-exclusion chromatography for in-depth analysis of membrane proteome of tissue samples. The use of detergents for extraction of membrane proteins would circumvent the above-mentioned limitations and, in addition, offer the option of stepwise extraction of membrane proteins which potentially can be used for selective protein solubilization and fractionation of membrane proteins.[22−24]

Removal of Detergents. The major goal of this work was to establish a simple, effective, and robust method for removal of detergents from solubilized membranes, such that mass spectrometric analysis would not be affected. For this purpose, crude membrane preparations from mouse brain were extracted with three different detergents including SDS, Triton X-100, and CHAPS, which are representative of anionic, nonionic, and zwitterionic detergents, respectively. The detergents were used in relatively high concentrations, exceeding their critical micellar concentration (CMC) values several-fold. The membrane preparations were solubilized with 3.5% (w/v) CHAPS, 2% (w/v) SDS, and 0.5% (v/v) Triton X-100. Since size exclusion chromatography has been reported to be highly effective in detergent removal,[25] we considered the use of gel filtration in our experiments. To separate detergents from proteins and to dissociate micelles, while keeping membrane proteins in solution, we employed the strongly chaotropic reagent urea at 8 M concentration. In the presence of 8 M urea, micelles dissociate while membrane proteins stay in solution. Importantly, the detergent migrates into the gel filtration matrix, while proteins elute in the void volume. Thus, when

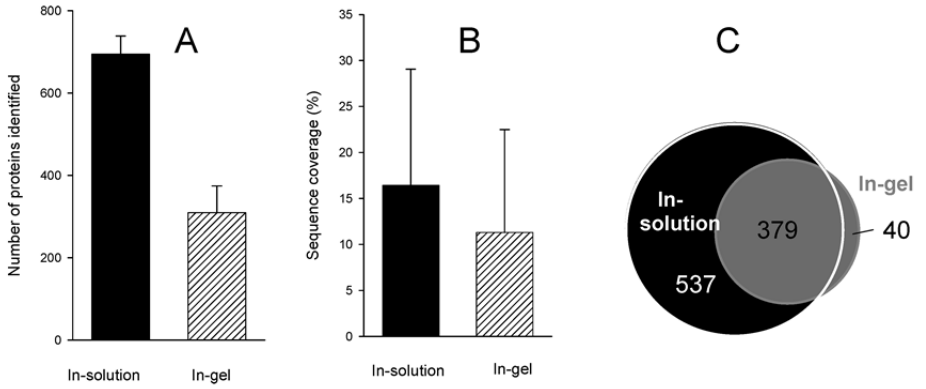


Figure 1. Comparison of membrane protein identification efficiency (A) and sequence coverage (B) using in-solution and in-gel methods. Brain membranes were extracted with 2% SDS. Five aliquots from the same extract were digested either in solution or in-gel. C, Venn diagram comparing the numbers of identifications achieved with both digestion methods.

the gel filtration columns equilibrated with 8 M urea were used, it was possible to efficiently separate proteins from the detergent.

After digestion with Lys-C, and removal of urea on StageTips, the peptides were analyzed by LC-MS/MS. We did not observe any detergent contamination in any of our experiments, demonstrating the efficiency of the depletion. Results were similar for all three tested detergents. Overall, 56.2 ± 2.5% of the identified proteins from the brain membrane preparation contained predicted transmembrane domains. This is an even higher proportion than in our previously reported method for profiling of membrane proteomes where the proteins were digested directly from membranes in a suspension.[8] In that method, about 40% of the identified proteins contained predicted transmembrane segments.[8−10] Note that not all membrane proteins contain transmembrane domains; therefore, the proportion of membrane proteins in our preparations is even higher (nearly 80%; see below).

To demonstrate the efficiency of the protein digestion in urea, we compared it to the in-gel digestion procedure. Five samples for each procedure were prepared using aliquots from the same membrane preparation. We found that the number of proteins identified by the in-solution method was more than twice that identified by the in-gel method (Figure 1A). Furthermore, the sequence coverage of identified proteins was a third higher in the new method (Figure 1B). The total number of membrane proteins identified in the 10 runs was 956. Of those, 537 and 40 proteins were exclusively identified in the in-solution and the in-gel method, respectively (Figure 1C). The proteins that were identified only by the in-solution method do not show any obvious physio-chemical difference compared to proteins identified by in-gel method. These proteins were likely not identified by the in-gel method due to a combination of less efficient peptide extraction by the in-gel method and stochastic 'picking' of peptide peaks for sequencing. However, we observed that we were able to recover more large proteins by the in-solution method. Using our method, we identified 27 proteins consisting of more than 1500 residues, whereas only one protein of this size was found using the in-gel approach. Thus, the results achieved using the new method encompass, rather than complement, the in-gel method.

Comparison to Published Membrane Proteome Analyses. The membrane proteome has been one of the most difficult challenges for 2D gel electrophoresis and usually very few or no transmembrane proteins are reported with that technology. MS-based proteomics has been more successful, but so far, researchers had to reduce protein complexity by applying different protein and peptide fractionation strategies that can facilitate the identification of less abundant proteins. However, multidimensional separation also generates large numbers of fractions that have to be analyzed individually which requires extensive measurement time. For rapid screening of tissue specimens such as clinical biopsy material, high-throughput methods are required. As described above, the combination of the previously developed method for extraction of membranes[8] with the here described detergent-based solubilization of membrane proteins resulted in identification of 530 proteins with predicted transmembrane domains from a mouse brain sample. To assess the relative usefulness of our method, we compared our results from single MS runs to other membrane proteome analyses in the recent literature, which employed extensive fractionation. In terms of identification of membrane proteins and the analysis time required, our method shows significant advantages. For example, a recently reported 3-D-strategy for analysis of membrane proteins allowed identification of only 125 proteins membrane proteins in mouse brain.[26] Other approaches for studying membrane proteins from various sources including *Corynebacterium glutamicum* and human platelet membranes have been carried out. Fischer et al., characterized the membrane proteome of two strains of *C. glutamicum* and reported 326 integral membrane proteins involving multiple fractionation steps and extensive mass spectrometric measuring time.[27] Moebius et al. in their platelets study identified less than 300 proteins with approximately 30% membrane proteins.[28] More recently, improved protocols for analysis of enriched membrane proteins have been published. Analysis of membrane fraction from HeLa cells using a phase transfer surfactant-aided digestion procedure resulted in identification of 764 membrane proteins (53% of total identified proteins) in 12 cumulative LC-MS/MS runs.[29] In another study, methanol was used to improve the efficiency of tryptic digestion which allowed identification of a total of 690 integral membrane proteins in 72 LC-MS/MS runs.[30]

Identification of Brain-Specific Proteins. A single run on the LTQ-Orbitrap instrument identified approximately 800 proteins. More than 70% of these were membrane proteins and 59% had predicted transmembrane domains (Figure 2C). The combination of our sample preparation procedure with high

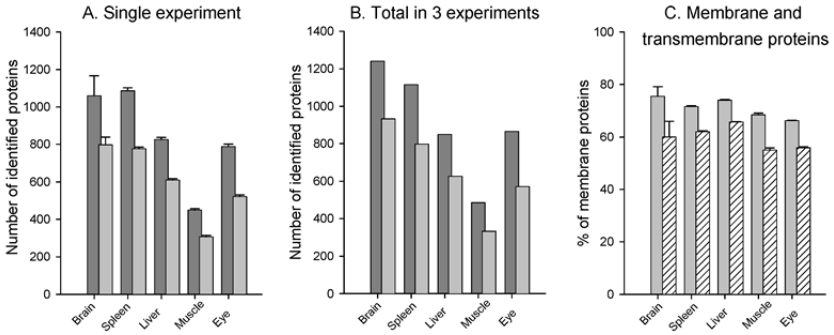


Figure 2. Identification of proteins from membrane preparations of mouse brain, spleen, liver, skeletal muscle, and eye. (A) single experiment; (B) cumulative result from 3 independent experiments; (C) content of membrane annotated proteins and proteins with predictable transmembrane segments. *Dark bars*, total protein number; *light bars*, membrane annotated proteins, *striped bars*, proteins with predictable transmembrane segments.

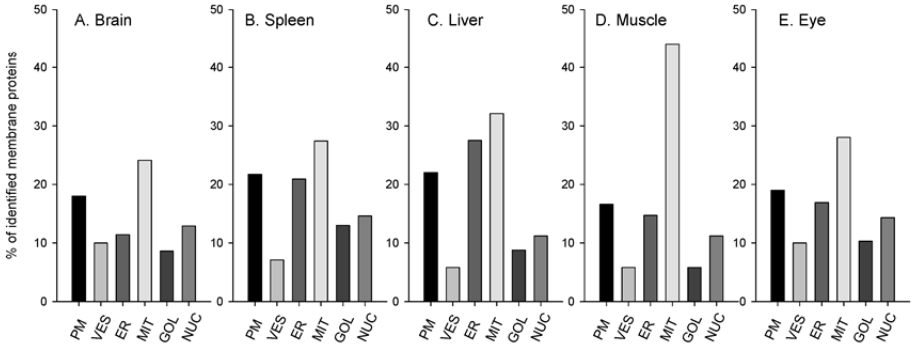


Figure 3. Subcellular localization of membrane proteins from mouse brain (A), spleen (B), liver (C), skeletal muscle (D), and eye (E). *GOL*, Golgi apparatus; *PM*, plasma membrane; *VES*, cytoplasmic vesicles; *ER*, endoplasmatic reticulum; *MIT*, mitochondria, *N*, nucleus.

resolution MS analysis yielded a large number of brain-specific proteins. These include neurotransmitter receptors such as glutamate and GABA receptors and also ion channels such as sodium and potassium channel proteins (Supplementary Table 2 in Supporting Information). The identified glutamate receptors represent ionotropic-AMPA, NMDA, metabotropic, and GluRdelta-2 receptors. Four subunits of $GABA_A$ receptor and two subunits of $GABA_B$ receptors were identified. A complete set of subunits of the voltage-dependent calcium channel was also mapped, including the channel subunit Cacna1a (and its isoform Cacna1e) as well the auxiliary subunits alpha2/delta, beta, and gamma. The channel subunit is a 281 kDa polypeptide with 23 predicted transmembrane helices. Similar in size and number of transmembrane domains is the identified sodium channel protein Scn1a (230 kDa, 23 transmembrane segments). Identification of such large proteins is an important advantage of our method.

Application of the Method to Other Tissues. Having demonstrated the efficiency of our method with mouse brain tissue, we wanted to show its applicability to a wide range of tissues. We selected liver, spleen, eye, and skeletal muscle which represent widely different tissue types. Each sample was analyzed in triplicate (Figure 3). We observed that the number of membrane proteins identified was dependent on the properties of the organs and ranged from 797 $\pm$ 43 in the brain to 307 $\pm$ 7 in leg muscle (Figure 2A). With respect to the total number of proteins identified, the percentage membrane proteins ranged from 66% in eye to 78% in brain (Figure 2C). The comparatively low number of proteins identified in leg muscle was due to the highly abundant titin, which has a molecular weight of several MDa and takes up most of the sequencing time in the mass spectrometer. Only small differences in the number of identifications were observed between each of the three runs, and therefore, the cumulative number of identified proteins for each tissue was only slightly higher compared to a single run which emphasizes the reproducibility of our method (Figure 2A,B).

Analysis of subcellular location of identified membrane proteins using Gene Ontology revealed distinct origin of membranes, which may reflect abundance of different organelles in the analyzed tissues (Figure 3). Mitochondrial membranes appear to be the most abundant in all analyzed samples ranging from 24% in brain to 44% in skeletal muscle. The high content of mitochondrial proteins in leg muscle reflects the fact that muscles are extremely rich in mitochondria for ATP production. The percentage of proteins annotated as extracellular or cell surface, which are mainly plasma membrane proteins, was very similar between the tissues and at 17−22%. Liver and spleen membranes contained the highest percentage of proteins annotated as endoplasmatic reticulum (Figure 3C,B). The abundance of endoplasmatic reticulum in liver is related to the high level of protein synthesis including major plasma proteins such as albumin. Compared to other tissues, brain and eye have the highest content of cytoplasmic vesicle proteins (Figure 3A,E), which reflects the importance of vesicular transport of neurotransmitters in nerve tissue.

In most cases, purification of organelles is not an easy task, in particular, when only minute amounts of frozen tissue are available. Our results demonstrate that relatively high numbers of membrane proteins belonging to various organelles can be profiled without extensive fractionation, simplifying protein quantitation.

Conclusions

Detergents are powerful agents for solubilization of biological membranes and allow separation of membrane proteins using chromatographic methods such as size exclusion and ion-exchange chromatography. Development of methods for mass spectrometry-based proteomics of biological membranes is currently a subject of many investigations. As detergents are almost indispensable reagents in membrane biochemistry, the majority of relevant studies involve in-gel digestion to remove detergent prior to mass spectrometric analysis. In this work we introduced a simple and highly reproducible method for membrane proteomics that allows use of detergents. Moreover, we showed that the gel-free analysis of membrane proteins yields more than twice the number of protein identifications compared to in-gel digestion. Since our method offers a fast and reproducible means for analysis of membrane proteins, we believe that it may be suitable for high-throughput applications.

Acknowledgment. We thank Dr. Alexandre Zougman for the helpful advice in mass spectrometric analysis and critical reading of the manuscript.

Supporting Information Available: Tables of all proteins identified in this study and selection of neurotransmitter receptors and ion channels identified in mouse brain in two single-run experiments. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Furth, A. J. Removing unbound detergent from hydrophobic proteins. *Anal. Biochem.* **1980**, *109* (2), 207–15.
(2) Speers, A. E.; Wu, C. C. Proteomics of integral membrane proteins-theory and application. *Chem. Rev.* **2007**, *107* (8), 3687–714.
(3) Blonder, J.; Conrads, T. P.; Veenstra, T. D. Characterization and quantitation of membrane proteomes using multidimensional MS-based proteomic technologies. *Expert Rev. Proteomics* **2004**, *1* (2), 153–63.
(4) Blonder, J.; Xiao, Z.; Veenstra, T. D. Proteomic profiling of differentiating osteoblasts. *Expert Rev. Proteomics* **2006**, *3* (5), 483–96.
(5) Washburn, M. P.; Wolters, D.; Yates, J. R., III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19* (3), 242–7.
(6) Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Yates, J. R. 3rd, A method for the comprehensive proteomic analysis of membrane proteins. *Nat. Biotechnol.* **2003**, *21* (5), 532–8.
(7) Olsen, J. V.; Andersen, J. R.; Nielsen, P. A.; Nielsen, M. L.; Figeys, D.; Mann, M.; Wisniewski, J. R. HysTag-a novel proteomic quantification tool applied to differential display analysis of membrane proteins from distinct areas of mouse brain. *Mol. Cell. Proteomics* **2004**, *3* (1), 82–92.
(8) Nielsen, P. A.; Olsen, J. V.; Podtelejnikov, A. V.; Andersen, J. R.; Mann, M.; Wisniewski, J. R. Proteomic mapping of brain plasma membrane proteins. *Mol. Cell. Proteomics* **2005**, *4* (4), 402–8.
(9) Olsen, J. V.; Nielsen, P. A.; Andersen, J. R.; Mann, M.; Wisniewski, J. R. Quantitative proteomic profiling of membrane proteins from the mouse brain cortex, hippocampus, and cerebellum using the HysTag reagent: Mapping of neurotransmitter receptors and ion channels. *Brain Res.* **2007**, *1134* (1), 95–106.
(10) Le Bihan, T.; Goh, T.; Stewart, I. I.; Salter, A. M.; Bukhman, Y. V.; Dharsee, M.; Ewing, R.; Wisniewski, J. R. Differential analysis of membrane proteins in mouse fore- and hindbrain using a label-free approach. *J. Proteome Res.* **2006**, *5* (10), 2701–10.
(11) Lu, X.; Zhu, H. Tube-gel digestion: a novel proteomic approach for high throughput analysis of membrane proteins. *Mol. Cell. Proteomics* **2005**, *4* (12), 1948–58.
(12) Rappsilber, J.; Ishihama, Y.; Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **2003**, *75* (3), 663–70.
(13) Shevchenko, A.; Tomas, H.; Havlis, J.; Olsen, J. V.; Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2006**, *1* (6), 2856–60.
(14) Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **1996**, *379* (6564), 466–9.
(15) Olsen, J. V.; Mann, M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (37), 13417–22.
(16) Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **2005**, *4* (12), 2010–21.
(17) Cox, J.; Mann, M. Is proteomics the new genomics. *Cell* **2007**, *130* (3), 395–8.
(18) Graumann, J.; Hubner, N. C.; Kim, J. B.; Ko, K.; Moser, M.; Kumar, C.; Cox, J.; Scholer, H.; Mann, M. Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. *Mol. Cell. Proteomics* **2008**, *7* (4), 672–83.
(19) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20* (18), 3551–67.
(20) Zubarev, R.; Mann, M. On the proper use of mass accuracy in proteomics. *Mol. Cell. Proteomics* **2007**, *6* (3), 377–81.
(21) Kristensen, D. B.; Brond, J. C.; Nielsen, P. A.; Andersen, J. R.; Sorensen, O. T.; Jorgensen, V.; Budin, K.; Matthiesen, J.; Veno, P.; Jespersen, H. M.; Ahrens, C. H.; Schandorff, S.; Ruhoff, P. T.; Wisniewski, J. R.; Bennett, K. L.; Podtelejnikov, A. V. Experimental Peptide Identification Repository (EPIR): an integrated peptide-centric platform for validation and mining of tandem mass spectrometry data. *Mol. Cell. Proteomics* **2004**, *3* (10), 1023–38.
(22) Rabilloud, T.; Gianazza, E.; Catto, N.; Righetti, P. G. Amidosulfobetaines, a family of detergents with improved solubilization properties: application for isoelectric focusing under denaturing conditions. *Anal. Biochem.* **1990**, *185* (1), 94–102.
(23) Churchward, M. A.; Butt, R. H.; Lang, J. C.; Hsu, K. K.; Coorssen, J. R. Enhanced detergent extraction for analysis of membrane proteomes by two-dimensional gel electrophoresis. *Proteome Sci.* **2005**, *3* (1), 5.
(24) Chevallet, M.; Santoni, V.; Poinas, A.; Rouquie, D.; Fuchs, A.; Kieffer, S.; Rossignol, M.; Lunardi, J.; Garin, J.; Rabilloud, T. New zwitterionic detergents improve the analysis of membrane proteins by two-dimensional electrophoresis. *Electrophoresis* **1998**, *19* (11), 1901–9.
(25) Amons, R.; Schrier, P. I. Removal of sodium dodecyl sulfate from proteins and peptides by gel filtration. *Anal. Biochem.* **1981**, *116* (2), 439–43.
(26) Lohaus, C.; Nolte, A.; Bluggel, M.; Scheer, C.; Klose, J.; Gobom, J.; Schuler, A.; Wiebringhaus, T.; Meyer, H. E.; Marcus, K. Multidimensional chromatography: a powerful tool for the analysis of membrane proteins in mouse brain. *J. Proteome Res.* **2007**, *6* (1), 105–13.
(27) Fischer, F.; Wolters, D.; Rogner, M.; Poetsch, A. Toward the complete membrane proteome: high coverage of integral membrane proteins through transmembrane peptide detection. *Mol. Cell. Proteomics* **2006**, *5* (3), 444–53.
(28) Moebius, J.; Zahedi, R. P.; Lewandrowski, U.; Berger, C.; Walter, U.; Sickmann, A. The human platelet membrane proteome reveals several new potential membrane proteins. *Mol. Cell. Proteomics* **2005**, *4* (11), 1754–61.
(29) Masuda, T.; Tomita, M.; Ishihama, Y. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J. Proteome Res.* **2008**, *7* (2), 731–40.
(30) Chick, J. M.; Haynes, P. A.; Molloy, M. P.; Bjellqvist, B.; Baker, M. S.; Len, A. C. Characterization of the rat liver membrane proteome using peptide immobilized pH gradient isoelectric focusing. *J. Proteome Res.* **2008**, *7* (3), 1036–45.

PR800412J

2.2 Filter aided sample preparation method an universal method for proteome samples

Sample preparation is the first and crucial step in large scale proteomic experiments. In a system wide approach, when the expression pattern of all proteins expressed in the cell or tissue is investigated, owing to the physiochemical properties of different proteins, 'difficult' protein classes can be underrepresented. This project involved development of an efficient sample preparation strategy that combines the advantages of digestion in solution with those of digestion in gel. The cell or tissue lysate is efficiently solublized in up to 5% SDS, and followed by detergent exchange by urea and by reduction alkylation and enzyme digestion all occurring in a single filter unit that serves as a 'chemical reactor'. The detergent is replaced by urea based on a similar principle of detergent removal method described before. This method is called Filter-aided sample preparation (FASP) and it has excellent peptide recovery (an attribute of in-solution digestion) and facilitates the use of strong and ionic detergents (so far an exclusive attribute of in-gel digestion). Additionally, peptides are in a cleaner state resulting in higher identification rates of the mass spectra. The absence of interfering substances increases the resolution of pre-fractionation techniques like OFFGEL electrophoresis. Application of this method led to identification of close to 3,000 proteins in single LC-MS/MS runs from cell lines, mouse brain and liver tissues. The unbiased nature of the method is demonstrated by the high percentage of membrane proteins among the identified proteins. When combined with OFFGEL electrophoresis, FASP resulted in identification of more than 7,000 proteins within 2 days. This remains the largest proteome data set for such a short measurement time.

The manuscript is published in the journal Nature methods as follows.

Universal sample preparation method for proteome analysis

Jacek R Wiśniewski, Alexandre Zougman, Nagarjuna Nagaraj & Matthias Mann

We describe a method, filter-aided sample preparation (FASP), which combines the advantages of in-gel and in-solution digestion for mass spectrometry–based proteomics. We completely solubilized the proteome in sodium dodecyl sulfate, which we then exchanged by urea on a standard filtration device. Peptides eluted after digestion on the filter were pure, allowing single-run analyses of organelles and an unprecedented depth of proteome coverage.

There are two major strategies for converting proteins extracted from biological material to peptides suitable for mass spectrometry (MS)-based proteome analysis. The first involves solubilization of proteins with detergents, separation of proteins by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis and digestion of the gel-trapped proteins ('in-gel' digestion)[1]. The second method is detergent-free, comprising protein extraction with strong chaotropic reagents such as urea and thiourea, protein precipitation and digestion of proteins under denaturing conditions ('in-solution' digestion). This second approach is frequently followed by two-dimensional peptide separation, for example, in the 'MudPit' strategy[2]. Advantages of in-gel digestion include its robustness against impurities, which interfere with digestion, but the gel may prevent peptide recovery and the method cannot easily be automated. In-solution digestion is more readily automatable and minimizes sample handling, but the proteome may be incompletely solubilized, and digestion may be impeded by interfering substances.

SDS is the reagent of choice for total solubilization of cells and tissues, and is routinely used in biochemical studies. Unfortunately, detergents, even in small concentrations, can preclude enzymatic digestion and dominate mass spectra owing to their ready ionizability and their great abundance compared to individual peptides. Therefore, depletion of SDS is a prerequisite for efficient mass-spectrometric analysis in proteomics. Because in-solution removal of SDS has been thought to be impossible, various alternative approaches have been developed for analyzing membrane proteomes. Early attempts involved membrane solubilization with formic acid[2], organic solvents[3,4] or digestion of the protein chains protruding from the membrane bilayer of nonsolubilized membranes[5–8]. We had recently discovered that membrane proteins can be fully depleted from detergents by gel filtration in 8 M urea such that they can then be analyzed as efficiently as soluble proteins[9]. Using this observation as a starting point, we sought to develop a method that combines strong detergents for universal solubilization with a means to efficiently 'clean up' the proteome before digestion and obtain purified peptides after digestion while avoiding the disadvantages of the gel format.

We reasoned that a common ultrafiltration device could be used for detergent removal to enable subsequent proteome analysis. We describe a method, filter-aided sample preparation (FASP), in which the sample is solubilized in 4% SDS, then retained and concentrated into microliter volumes in an ultrafiltration device (Online Methods). The filter unit then acts as a 'proteomic reactor' for detergent removal, buffer exchange, chemical modification and protein digestion. The four critical steps of the FASP method are: (i) depletion of detrimental low-molecular-weight components in urea-containing buffer, (ii) carboamidomethylation of thiols, (iii) digestion of proteins and (iv) elution of peptides (**Fig. 1**). Notably, during peptide elution, the filter retains high-molecular-weight substances that would otherwise interfere with subsequent peptide separation.

As the key feature of the method is the ability of the filter membrane to retain high-molecular-weight substances (proteins and DNA) and to allow through low-molecular–weight substances (impurities and digested peptides), selecting a filter with the desired separation properties is essential. We tested filters with relative molecular mass (M_r) cut-offs of 3,000 (3k filter) and 10,000 (10k filter). Note that the manufacturer determined these cutoffs with folded rather than detergent-denatured proteins. We performed all MS analyses by electrospray liquid chromatography–tandem MS (LC-MS/MS) using a linear ion trap–orbitrap instrument (LTQ-Orbitrap) essentially as described previously[10,11] (Online Methods).

We first compared the distribution of molecular weights of the identified proteins using either a 3k or 10k filter. We observed no substantial differences in the number of proteins identified per molecular-weight interval down to the 5–10 kDa bin (**Supplementary Fig. 1a,c** online). Next, we compared the efficiency of peptide elution using either a 3k or 10k filter. The number of identified tryptic peptides with a molecular weight above about 1,500 Da was much reduced for the 3k filter compared to the 10k filter. The number of peptides larger than 1,500 Da decreased gradually with increasing size, and peptides with masses over 2,500 Da were almost completely retained by the 3k filter

Department of Proteomics and Signal Transduction, Max Planck Institute for Biochemistry, Martinsried, Germany. Correspondence should be addressed to J.R.W. (jwisniew@biochem.mpg.de) or M.M. (mmann@biochem.mpg.de).

RECEIVED 25 NOVEMBER 2008; ACCEPTED 11 MARCH 2009; PUBLISHED ONLINE 19 APRIL 2009; DOI:10.1038/NMETH.1322

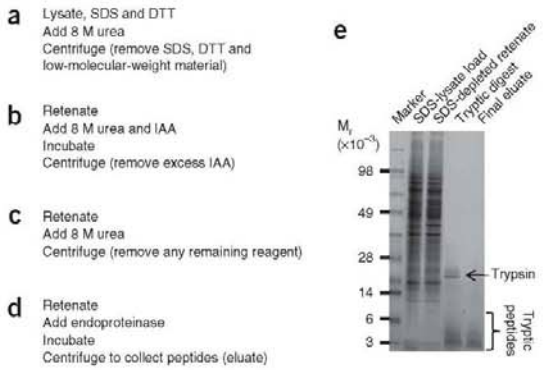


Figure 1 | Filter-aided sample preparation (FASP) for MS-based proteomic analysis. (**a**) Cell or tissue lysates can be prepared in the presence of high concentrations of detergent. Disulfide bridges are reduced with dithiothreitol (DTT). Detergent micelles and protein detergent complexes are dissociated in the presence of 8 M urea. The detergent, DTT and other low-molecular-weight components are removed by ultrafiltration (Microcon units) facilitated by centrifugation. (**b**) Thiols are carboxyamidomethylated with iodoacetamide (IAA) and excess reagent is removed by ultrafiltration. (**c**) In repeated washes with 8 M urea any remaining detergent is depleted from the proteins. (**d**) The protein suspension is digested with endoproteinase, and the resulting peptides are collected as a filtrate. High-molecular-weight molecules including the endoprotease are retained on the filter. When nuclei or total cell lysates are processed in the units, DNA is retained on the filter. (**e**) SDS polyacrylamide gel electrophoresis analysis of total cell lysate, SDS-depleted and alkylated proteins, tryptic digest and eluted peptides. (Note that FASP does not involve any gel separation.)

(**Supplementary Fig. 1b,d**). As the 10k filter efficiently retained small proteins (5–10 kDa) and efficiently released peptides up to 5,000 Da (**Supplementary Fig. 1b,d**), we used it as the standard in the subsequent experiments.

To test the efficiency and range of applicability of the method, we processed and analyzed various amounts of bovine serum albumin (BSA) protein standard and total HeLa cell lysates. As judged by UV-light absorption and LC-MS/MS analysis of BSA peptides, we determined that FASP resulted in very high yield over at least three orders of magnitude of protein abundance (**Supplementary Fig. 2** online). Analysis of different numbers of HeLa cells, down to a few thousand cells, showed no substantial decrease in the number of identified peptides and proteins (**Supplementary Fig. 3** online).

In common with frequently used in-solution digestion protocols, FASP uses high concentrations of urea. In solution, a small fraction of urea decomposes to cyanic acid, which reacts with side chains of lysine and arginine and N-terminal amino groups to form carbamylated residues. As decomposition of urea is facilitated by high temperature, we performed all steps in FASP at room temperature (18–22 °C) and carried out the centrifugation steps at constant 20 °C. Under these conditions, less than 0.5% of identified peptides carried carbamylated arginine or lysine residues, a similar proportion as observed in our previous in-solution experiments (data not shown).

Next, we tested FASP on samples including mouse liver and brain tissues in addition to cultured cells. Preparation of tissue lysates was extremely simple, consisting only of tissue homogenization in the presence of SDS and subsequent application of an aliquot of this homogenate to the membrane reactor, taking less than 10 min. Notably, the presence of SDS efficiently inactivated detrimental enzymatic functions such as protease and phosphatase activity. In single-run analyses with 4-h gradients, we identified 1,800–2,200 proteins, with 99% confidence and at least two identified peptides per protein using the MaxQuant algorithms[12]. When we added proteins identified with one peptide, the number of identified proteins increased to 2,200–2,700 proteins (**Fig. 2a** and **Supplementary Table 1** online). In comparison, in our recent characterization of the liver proteome, using extensive cytosolic and membrane fractionation with analysis of 20 in-gel slices, we identified 2,210 proteins in total[13]. In the FASP datasets, 75–80% of fragmentation events resulted in the identification of the peptide in a database. Such high identification rates had been previously only observed for stable isotope labeling with amino acids in culture (SILAC) pairs[12], suggesting that the high purity of eluted FASP peptides minimized fragmentation events associated with chemical noise, which cannot lead to peptide identifications. Gene Ontology analysis revealed 42% (HeLa total cell lysate) and 52% (brain tissue) proteins matching to the membrane category (**Supplementary Table 1** and **Supplementary Fig. 4** online). This high percentage of membrane proteins indicated the absence of bias against hydrophobic proteins compared to soluble proteins. We also observed better sequence coverage for membrane proteins via the FASP preparation method than by the standard in-solution digestion method (**Supplementary Fig. 5** online).

We had previously reported the identification of 22,905 peptides and 3,979 proteins from HeLa cells by combining peptide isoelectric focusing in the 'OFFGEL' fractionator (Agilent) with 12 peptide fractions and conventional in-solution digestion[14]. In comparison, using FASP we identified 40,582 unique peptides corresponding to 7,093 proteins from HeLa cells (**Supplementary Table 1**). To our knowledge, this is the largest reported proteome in any single experiment. The measurements took only 2 d, showing that deep proteome coverage is possible within a reasonable analysis time. In comparison, comparable in-depth measurement of the embryonic stem cell proteome took more than three weeks of measurement time[15].

We next used Gene Ontology analysis to investigate whether the FASP-prepared proteome was biased for proteins from any compartments or protein classes. As a reference set for the expressed genome in HeLa cells, we used the Affymetrix GeneChip Human Genome U133 Plus 2.0 dataset, which detected 23,348 probes that have at least two positive absorbance calls in three experiments corresponding to 10,937 genes[16]. For all genes whose messages were detected on the chip, we determined the proportion of Gene Ontology cellular compartments as well as biological functions. We then compared them to the same categories in the FASP-based proteome measurements. None of the categories were considerably different, demonstrating that the FASP preparation method is universal in that it does not lead to preferential extraction of proteins from specific cellular compartments or with specific functions (**Fig. 2b–d**).

As expected, compared to the protein coverage of the previous HeLa cell experiment, low-abundance protein classes were represented more extensively using the FASP method. For example, the percentage of proteins that were Gene Ontology–annotated for transcription, signal transduction and receptor activity increased by 20–30%. This was paralleled by a corresponding decrease in the percentage of proteins annotated for metabolic and catalytic

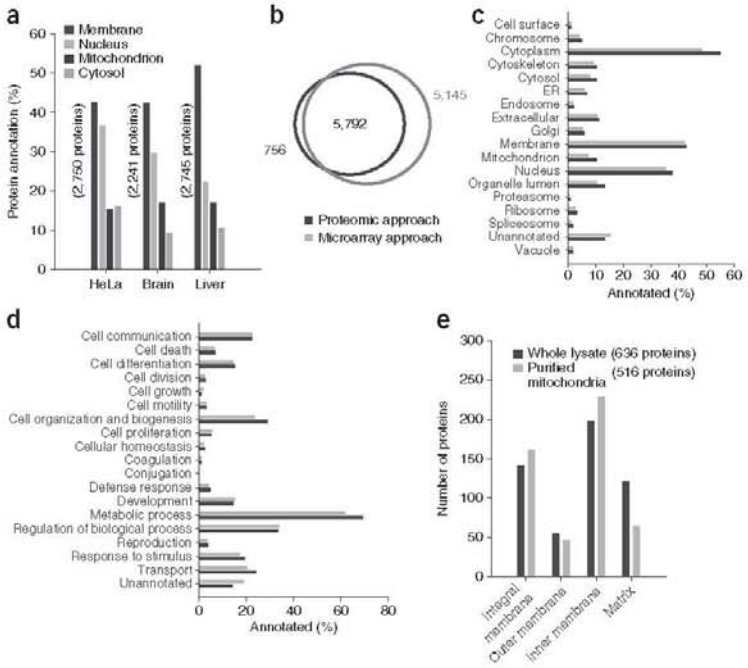


Figure 2 | FASP-based proteomic analysis of SDS lysates. (**a**) Single-run analysis of total lysates of HeLa cells, mouse brain and mouse liver processed in 10k filter units using two-step LysC and trypsin digestion. Bars show the percentage of proteins with the indicated Gene Ontology annotations. Total numbers of proteins identified per run are indicated in parentheses. (**b**) Venn diagram shows the overlap of genes identified by the FASP-based proteomic and microarray approaches. Note that a subset of identified genes cannot be matched to Affymetrix identifiers. (**c**,**d**) Comparison of Gene Ontology annotations for cell component (**c**) and biological processes (**d**) show that proteome and mRNA data are in concordance. (**e**) Single-run analysis of mouse liver mitochondria compared to mitochondrial proteins identified in 12 isoelectric focusing fractions of 'whole lysate' of HeLa cells. Bars show the number of proteins with the indicated Gene Ontology annotations. Total numbers of mitochondrial proteins identified in each experiment are indicated in parentheses.

function (**Supplementary Fig. 4**). We identified more than 90% of the proteins involved in the oxidative phosphorylation pathway, assembly of the ribosome, RNA polymerase and the polymerase II transcriptional machinery (**Supplementary Figs. 6–8** online). Considering that some of these proteins were cell type–specific and cell stage–specific and therefore were not expressed in all conditions, our data had very high coverage.

Before establishing the FASP protocol, we often separated proteomes into soluble and pellet fractions to achieve uniform representation of the proteome. These pellet fractions led to particularly poorly focused peptides in isoelectric focusing, with many peptides in three or more fractions, presumably because of contamination by nucleic acids, which are highly charged. With FASP, there is only a single proteome fraction, and focusing of all peptides was improved considerably (we detected 82% of peptides only in a single well, and 14% were focused into two wells; **Supplementary Fig. 9** online).

One major advantage of the FASP over the 'in-gel' and 'in-solution' approaches is its ability to accommodate a wide range of digestion conditions. We observed specific digestion for five different endoproteases (**Supplementary Table 2** and **Supplementary Fig. 10** online) and efficient digestion even without urea (**Supplementary Table 3** online).

Last, we investigated whether FASP would allow extensive proteomic analysis of isolated organelles without any fractionation. Indeed, with two single runs of a mouse liver fraction enriched in mitochondria, we identified 516 proteins annotated to this organelle (**Fig. 2e**) with a very high coverage of the core machinery of oxidative phosphorylation (**Supplementary Fig. 11** online). For this experiment, we used a filter unit with a relative molecular mass cut-off of 30,000 (30k filter) as the FASP reactor. This device shortened preparation time by a factor of three (2 h) and did not prevent identification of very small proteins.

The FASP method allowed processing of total SDS lysates of essentially any class of protein from biological material of any origin, thus solving the long-standing problem of efficient and unbiased solubilization of all cellular proteins irrespective of their subcellular location. In particular, FASP enables digestion of membrane proteins under conditions previously applied only to soluble proteins. With larger volume filter units, FASP also allowed handling of milligram amounts of protein.

The identification of more than 2,000 proteins in single runs using only 1–2 μg of material opens up interesting applications for proteomics, especially as the entire sample workflow is very streamlined. In organelle analysis, for example, this depth of analysis may already be sufficient: it is at least an order of magnitude greater in sensitivity and number of identified proteins than widely used proteome techniques such as two-dimensional gel analysis. For in-depth analysis of complex, mammalian proteomes FASP could be a crucial enabling sample preparation technology.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemethods/.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank S. Ren for bioinformatic support. This work was supported by the Max Planck Society for the Advancement of Science, High-throughput Epigenetic Regulatory Organization In Chromatin (HEROIC), Role of Ubiquitin and Ubiquitin-like Modifiers in Cellular Regulation (RUBICON) 6th Framework grants of the European Commission and the Munich Center for Integrated Protein Science (CIPSM).

Published online at http://www.nature.com/naturemethods/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. *Anal. Chem.* **68**, 850–858 (1996).
2. Washburn, M.P., Wolters, D. & Yates, J.R., III *Nat. Biotechnol.* **19**, 242–247 (2001).

3. Blonder, J., Xiao, Z. & Veenstra, T.D. *Expert Rev. Proteomics* **3**, 483–496 (2006).
4. Blonder, J., Conrads, T.P. & Veenstra, T.D. *Expert Rev. Proteomics* **1**, 153–163 (2004).
5. Wu, C.C., MacCoss, M.J., Howell, K.E. & Yates, J.R., III *Nat. Biotechnol.* **21**, 532–538 (2003).
6. Olsen, J.V. *et al. Mol. Cell. Proteomics* **3**, 82–92 (2004).
7. Nielsen, P.A. *et al. Mol. Cell. Proteomics* **4**, 402–408 (2005).
8. Le Bihan, T. *et al. J. Proteome Res.* **5**, 2701–2710 (2006).
9. Nagaraj, N., Lu, A., Mann, M. & Wisniewski, J.R. *J. Proteome Res.* **7**, 5028–5032 (2008).
10. Olsen, J.V. *et al. Mol. Cell. Proteomics* **4**, 2010–2021 (2005).
11. Rappsilber, J., Ishihama, Y. & Mann, M. *Anal. Chem.* **75**, 663–670 (2003).
12. Cox, J. & Mann, M. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
13. Shi, R. *et al. J. Proteome Res.* **6**, 2963–2972 (2007).
14. Hubner, N.C., Ren, S. & Mann, M. *Proteomics* **8**, 4862–4872 (2008).
15. Graumann, J. *et al. Mol. Cell. Proteomics* **7**, 672–683 (2008).
16. Mense, S.M. *et al. Physiol. Genomics* **25**, 435–449 (2006).

ONLINE METHODS

Tissue and cell solubilization. We homogenized 50-mg pieces of frozen mouse liver or brain in 0.40 ml of 0.1 M Tris-HCl, pH 7.6 using an Ultra Turbax blender (IKA) at maximum speed (approximately 25,000 r.p.m.) at 4 °C for 30 s. Then, 50 µl aliquots of 20% SDS and 1 M DTT were added to the homogenate and the mixture was incubated for 3 min at 95 °C. Frozen aliquots of 5×10^7 HeLa cells were lysed in 0.5 ml of 4% SDS and 0.1 M DTT in 0.1 M Tris-HCl, pH 7.6 at room temperature and briefly sonicated to reduce viscosity of the lysate. BSA was denatured and reduced in 4% SDS and 0.1 M DTT in 0.1 M Tris-HCl, pH 7.6 at 95 °C for 3 min. Membrane and cytosolic fractions were prepared from HeLa cells and mouse brains as described previously[9].

Filter-aided sample preparation (FASP). Aliquots of lysates corresponding to 1 mg wet tissue (0.1 mg protein) or 2×10^5 HeLa cells (0.13 mg of protein) were mixed with 200 µl of 8 M urea in Microcon devices YM-10 or YM-3 (Millipore). The device was centrifuged at 14,000*g* at 20 °C for 40 min. All following centrifugation steps were performed applying the same conditions allowing maximal concentration. The concentrate was diluted with 200 µl of 8 M urea in 0.1 M Tris-HCl, pH 8.5 and the device was centrifuged. Subsequently, 100 µl of 0.05 M iodoacetamide in 8 M urea in 0.1 M Tris-HCl, pH 8.5 were added to the concentrate followed by centrifugation. The resulting concentrate was diluted with 100 µl 8 M urea in 0.1 M Tris-HCl, pH 7.9 and concentrated again. This step was repeated 2 times, and the concentrate was subjected to proteolytic digestion (**Supplementary Table 4** online) The digests were collected by centrifugation, and the filter device was rinsed with 50 µl 0.5 M NaCl and centrifuged. Detailed instructions for performing FASP are described in the **Supplementary Protocol** online. The combined filtrates were desalted on MILI-SPE Extraction disk cartridge (C_{18}-SD); 7 mm per 3 ml (Millipore).

The peptide content was estimated by UV light spectral density at 280 nm using an extinctions coefficient of 1.1 of 0.1% (g l^{-1}) solution that was calculated on the basis of the frequency of tryptophan and tyrosine (the main UV light–absorbing amino acids at 280 nm) in vertebrate proteins[17].

Isoelectric focusing of peptides. We separated 0.05 mg of peptides into 12 fractions on the 3100 OFFGEL Fractionator (Agilent Technologies) as described previously. The immobilized pH gradient strips (IPG strips) from GE Healthcare (Immobiline DryStrip pH 3–10, 13 cm) were rehydrated with 20 µl per well of isoelectric focusing buffer containing 5% glycerol and 50-fold diluted IPG buffer pH 3–10 (GE Healthcare) for 20 min. Peptides were dissolved in 1.68 ml of the isoelectric focusing buffer and 0.14 ml of the solution were loaded into to each well. Mineral oil was added to the ends to prevent the drying of the filter wicks wetted with the buffer. Focusing was performed at 20 °C with maximum values of 4,500 V and 200 mW. The limiting maximum current was set to 50 µA. Focusing was carried out for a target of 20 kVh. The focused peptides were acidified by adding 20 µl of acidic mixture (0.5% acetic acid, 1% TFA and 2% acetonitrile) before desalting and LC-MS/MS analysis.

Preparation and FASP of mitochondria. Frozen mouse liver was homogenized in a motor-driven glass-Teflon Potter-Elvehjem homogenizer at a 1:10 ratio of tissue to homogenization buffer (0.3 M sucrose, 10 mM MOPS-NaOH, 1 mM EDTA). The cell debris and nuclei were removed by centrifugation at 1,000*g* for 10 min. Then, the supernatant was collected and centrifuged at 16,000*g* for 15 min. The mitochondrial pellet was washed once by resuspending in the homogenization buffer and pelleting at 16,000*g* for 15 min. Mitochondrial pellet was lysed in 0.5 ml of 4% SDS and 0.1 M DTT in 0.1 M Tris-HCl, pH 7.6 at room temperature. The lysates were processed with the FASP method as described above but using 30k filtration units (Microcon; Millipore).

Mass spectrometric analysis. The digests were purified and stored in C_{18} StageTips as described[11]. Usually up to 10 µg peptide mixture was loaded on a StageTip containing two membrane plugs. Approximately a half of the sample was applied to the high-performance liquid chromatography column in each experiment. Peptide mixtures were analyzed by online capillary LC-MS/MS. The LC-MS/MS setup was similar to that described before[18]. Briefly, samples were separated on an in-house made 15 cm reversed-phase capillary emitter column (inner diameter 75 µm, 3 µm ReproSil-Pur C18-AQ medium; Dr. Maisch GmbH) using 240 min (cell and tissue lysates) or 60 min (BSA standard) gradients and analyzed using the LTQ-Orbitrap instrument (Thermo Fisher Scientific). Survey MS scans were acquired in the orbitrap with 60,000 resolution. For accurate mass measurements, the lock-mass option was used[10]. Up to 10 most intense ions in each full MS scan were fragmented and analyzed in the LTQ.

Peak list generation, database searching and validation. Raw MS files were processed with MaxQuant, an in-house developed software suite[12]. Peak list files were searched against decoy International Protein Index mouse database version 3.46 containing both forward and reverse protein sequences by the MASCOT search engine[19]. Initial parent and fragment ion maximum mass deviation[20] were set to 7 p.p.m. and 0.5 Da, respectively. The search included variable modifications of oxidation of methionine and protein N-terminal acetylation. Peptides with at least six amino acids were considered for identification. The false discovery rate for both peptides and proteins were set at 0.01. All peptides and proteins identified in this study are listed with posterior error probability values in **Supplementary Data 1–7** online.

Bioinformatics analysis. Gene ontology analysis of the identified proteins was performed using the Protein Center platform (Proxeon Biosystems).

17. Zhuang, Y., Ma, F., Li-Ling, J., Xu, X. & Li, Y. *Mol. Biol. Evol.* **20**, 1978–1985 (2003).
18. Olsen, J.V. & Mann, M. *Proc. Natl. Acad. Sci. USA* **101**, 13417–13422 (2004).
19. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. *Electrophoresis* **20**, 3551–3567 (1999).
20. Zubarev, R. & Mann, M. *Mol. Cell. Proteomics* **6**, 377–381 (2007).

2.3 Large scale phospho-proteomics by higher energy collision dissociation

Phosphorylation of proteins is one of the key post translation modifications and is regularly studied by MS based proteomics. Phosphorylation on serine and threonine residues is labile while phosphorylated tyrosine is relatively stable. Over the years, the ion trap based tandem mass spectrometry has gained in popularity over the quadrupole mass spectrometer. Thus most of phosphoproteomics analysis is performed in a low energy CID regime in ion trap instruments. Advantages of the ion trap include high sensitivity, fast scan speeds and parallel operation with the Orbitrap or FT instrument. However, since the phospho group is lost in the CID fragmentation (neutral loss), the analysis includes an additional fragmentation step called pseudo MS^3 or multi-stage activation (MSA). This multi-stage activation results in complex spectra and further the low molecular weight cut-off precludes low molecular weight reporter ion analysis. In contrast, HCD fragmentation yields high resolution spectra potentially at the cost of lower sensitivity. In this project, we investigated the feasibility of large scale phosphoproteomics by HCD collision and detection of fragment ions in the Orbitrap analyzer. We observed that phosphoproteomics by HCD fragmentation is superior to CID fragmentation. We identified more phosphosites by HCD fragmentation in spite of low throughput and lack of parallel operation in the LTQ-Orbitrap tandem mass spectrometer. As a result of this project, in our laboratory we now routinely analyze phosphopeptides by HCD fragmentation.

A manuscript has been submitted to the Journal of proteome research. It has gotten positive reviews and a revision is being prepared.

In revision as a "Technical Note" in the Journal of Proteome Research

Feasibility of large scale phosphoproteomics with HCD fragmentation

Nagarjuna Nagaraj[1, 3], Rochelle C. J. D'Souza[1, 3], Juergen Cox[1], Jesper V. Olsen[2] and Matthias Mann[1]

[1]Department of Proteomics and Signal Transduction, Max-Planck Institute for Biochemistry, 82152 Martinsried, Germany.

[2]Novo Nordisk Foundation Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3b, 2200 Copenhagen, Denmark.

[3] both authors contributed equally

Corresponding author: mmann@biochem.mpg.de

Keywords:

Phospho-proteomics, HCD, mass spectrometry LTQ-Orbitrap Velos, mass accuracy

Running Title: HCD based large scale phospho proteomics

Abbreviations: MS, mass spectrometry; MS/MS, tandem mass spectrometry; HCD, Higher energy Collisional Dissociation, CID, Collisional Induced Dissociation, PTM, Post-Translational Modification, HPLC, high performance liquid chromatography; IPI, international protein index; FDR, false discovery rate

Abstract

Mass spectrometry (MS)-based proteomics now enables the analysis of thousands of phosphorylation sites in single projects. Among a wide range of analytical approaches, the combination of high resolution MS scans in an Orbitrap analyzer with low resolution MS/MS scans in a linear ion trap has proven to be particularly successful ('high-low' strategy). Here we investigate if the improved sensitivity of higher energy collisional dissociation (HCD) on an LTQ-Orbitrap Velos instrument allows a 'high-high' strategy. A high resolution MS scan was followed by up to 10 HCD MS/MS scans and we achieved cycle times of about 3 s making the method compatible with chromatographic time scales. Fragment mass accuracy increased about 50-fold compared to the 'high-low' strategy. Unexpectedly, the HCD approach mapped up to 16,000 total phosphorylation sites in one day's measuring time – the same or better than the standard high-low strategy. Reducing the target values from a standard of 30,000 to 5,000 ions did not severely affect identification rates but did decrease identification and localization scores for phosphorylation sites. We conclude that HCD in the new configuration is now a viable method for large-scale phosphoproteome analysis alongside CID and electron capture / transfer dissociation (ECD/ETD).

INTRODUCTION

Global mapping and localization of post-translational modifications (PTMs) such as phosphorylation is crucial for understanding the activity of the cell. Phosphorylation acts as a molecular switch in various signaling pathways and plays a pivotal role in many biological processes[1]. Mass spectrometry (MS)-based proteomics has emerged as a powerful technique for studying PTMs[2-4]. Among many different versions of MS-based phosphoproteomics, hybrid instruments with quadrupole and time of flight analyzers (quadrupole – TOF) or with two different types of ion traps have gained popularity in the last decade[5]. In particular, the combination of high mass accuracy for the precursor ion and low mass accuracy for the fragment ions from linear ion trap – Orbitrap instruments (LTQ-Orbitrap) is a widely applied instrumental configuration. Employing this 'high-low' strategy in large scale phosphoproteomic approaches has led to the identification and quantification of several thousand phosphosites in single projects[6-8]. However, analyzing phosphopeptides by Collision Induced Dissociation (CID) in the ion trap (resonant excitation mediated collision) results in significant neutral loss for phospho-serine (pS) and phospho-threonine (pT) containing peptides and this can require multiple activation steps to efficiently fragment them[9, 10]. Furthermore, in ion trap fragmentation the 'one third rule' (loss of low mass ions depending on the fragmentation q value) [11] precludes the analysis of low molecular weight reporter ions that are very informative for example in the case of phospho-tyrosine (pY) ions[12, 13]. A different class of fragmentation techniques, electron capture dissociation (ECD)[14] or electron transfer dissociation (ETD) [15], complements CID, particularly for labile phosphopeptides[16].

Higher energy C-trap fragmentation (HCD) is an additional fragmentation technique that can be used on the LTQ-Orbitrap[17]. HCD is performed by injecting peptide ions into a collision cell and by analyzing fragments at high resolution and mass accuracy in the Orbitrap analyzer. HCD fragmentation is similar to the fragmentation in triple-quadrupole or quadrupole – TOF instruments and it overcomes the problem of low mass cut-off of ion trap fragmentation. In addition, high accuracy at both the precursor mass and fragment levels (a 'high-high' strategy) should be very desirable as it would dramatically improve the quality of fragmentation spectra. For example, charge states can easily be distinguished in high resolution spectra and fragments are less likely to be assigned to the wrong peptide sequence. High-high strategies based on HCD

fragmentation have been adopted for proteomic and phosphoproteomics studies but they have previously been restricted to specialized applications because of lower sensitivity compared to CID, resulting in extended injection times and a slower duty cycle[18 19, 20]. The newly developed LTQ Orbitrap Velos features an S-lens, improving ion current into the instrument by at least ten-fold, as well as a more efficient HCD cell[21]. This improved sensitivity could make routine HCD measurements of phosphopeptides feasible within the short MS and MS/MS duty cycles required in large-scale phosphoproteomics.

Here we investigate the feasibility of routine large scale phosphoproteomics by HCD fragmentation on a LTQ Orbitrap Velos instrument. We analyze key parameters such as sensitivity and fill times, cycle times for MS/MS experiments, identification success rates and depth of phosphoproteome coverage that are of prime importance for designing a large scale phosphoproteome study. Comparison with high-low strategy showed a similar or superior performance.

EXPERIMENTAL PROCEDURES

Cell culture and peptide preparation

HeLa S3 cells (ATCC)) were cultured in roller bottles in RPMI 1640 (Gibco) supplemented with 10% fetal bovine serum (Invitrogen) and 1% penicillin/streptomycin (Invitrogen). On reaching sufficient confluence in suspension, the cells were centrifuged at 1000 rpm. The HeLa cell pellet was lysed in a buffer of 4% SDS and 100 mM DTT in 100 mM tris-HCl pH 7.5. The lysate was processed by the FASP method[22]. Briefly, the lysate was sonicated and heated in the SDS buffer to ensure complete homogenization and denaturation. The protein concentration was measured and lysate was loaded onto 15 ml Amicon filter units (10 kDa MWCO) (Millipore) and washed with Tris buffer (UA buffer) containing 8 M urea to remove SDS[23]. Proteins were alkylated with 50 mM iodoacetamide in urea buffer and incubated for 20 minutes followed by removal of excess iodoacetamide by multiple washes with urea buffer. After reduction and alkylation the proteins were equilibrated in 20 mM ammonium bicarbonate and digested with trypsin (Promega) in a protein to enzyme ratio of 100:1 at 37 ^{o}C overnight. After digestion the peptides were eluted by centrifugation with an additional elution with 50 µl of water. Elution with water

avoids the desalting step for further processing of the peptides. Peptide yields for the first large-scale HCD experiment were 4 mg and 2.5 mg for the second experiment.

Fractionation of peptides by Strong Cation Exchange (SCX) chromatography

Peptides were concentrated into a volume of seven ml, the pH adjusted to 2.7 and then adjusted to 10 ml with 100% ACN. The peptides were then separated by strong cationic exchange chromatography (SCX) as described[6, 7]. The peptide mixture was loaded onto a cation exchanger column equilibrated with 30% ACN containing 5 mM KH_2PO_4. The flow-through which predominantly contains multiply phosphorylated peptides was collected. The peptides bound to the column were eluted in an increasing salt gradient with buffer containing 5 mM KH_2PO_4 and 150 mM KCl. The fractions generated by SCX were then pooled to 7 fractions based on UV absorbance. The flow-through from the SCX column was also used as one fraction.

Enrichment of phosphopeptide by TiO_2 beads

The flow through and the seven SCX fractions were subjected to TiO_2 enrichment with 3 consecutive incubations for the flow through and one incubation each for the remaining 7 fractions as described . First the UV absorbance of the fractions was measured and they were incubated with TiO_2 (MZ-Analysentechnik, Germany) with a peptide to bead ratio of 1:2 to 1:8 [24]. Before mixing with the fractions the TiO_2 beads were re-suspended in 30 mg/ml solution of dihydrobenzoic acid (Sigma) to prevent non-specific binding. Next the phospho-peptide bound beads were washed with 30% ACN and 3% TFA twice followed by two washes with 75% ACN and 0.3% TFA. The phospho-peptides were then eluted under basic conditions using 25% ammonium hydroxide and ACN. Finally, the eluted phospho-peptides were loaded on C_{18} StageTips[25].

Reverse phase chromatography and mass spectrometry

Peptides were separated in a 15 cm column (75 μm inner diameter) packed in-house with 3 μm C_{18} beads (Reprosil-AQ Pur, Dr. Maisch) on a Proxeon EASY-nLC system (Proxeon Biosystems, Odense, Denmark) using a binary gradient provided by buffer A (0.5 % acetic acid) and buffer B (0.5% acetic acid and 80% ACN). The peptides (4 μl) were loaded directly without any trapping column with buffer A at a flow rate of 500 nl/min. Elution was carried out at a flow rate of 250 nl/min, with a linear gradient from 10% to 35% buffer B in 95 minutes followed by 50% B for 15 minutes. At the end of the gradient the column was washed with 90% B and equilibrated with 5% B for 10 min. The LC system was directly coupled in-line with a LTQ-Orbitrap Velos instrument (Thermo Fisher Scientific) via the Proxeon Biosystems nanoelectrospray source. The source was operated at 2.1-2.25 kV, with no sheath gas flow, with the ion transfer tube at 200 ^{o}C.

The mass spectrometer was programmed to acquire in a data dependent mode. For the high-high strategy, survey scans were acquired in the Orbitrap mass analyzer with resolution 30,000 at m/z 400 with lock mass option enabled for the 445.120025 ion[26]. However the target lock mass abundance was set to 0% instead of 5-10% in order to save the injection time for the lock mass. For the full scans, 1E6 ions were accumulated within a maximum injection time of 250 ms in the C trap and detected in the Orbitrap analyzer. The ten most intense ions with charge states ≥ 2 were sequentially isolated (signal threshold of 10,000) to a target value of 3E4 or 4E4 with a maximum injection time of 150 ms and fragmented by HCD in the collision cell (normalized collision energy of 40%) and detected in the Orbitrap analyzer at 7,500 resolution. For the high–low strategy, full scans were acquired in the Orbitrap analyzer at 60,000 resolution as parallel acquisition is enabled in the high-low mode. Up to the 20 most intense peaks with charge state ≥ 2 were selected for sequencing (signal threshold of 1000) to a target value of 5,000 with a maximum injection time of 25 ms and fragmented in the ion trap by collisional induced dissociation with normalized collision energy of 35%, activation q=0.25 and activation time of 10 ms. For CID ‘wideband activation’ and ‘multi-stage activation’ options were enabled with the appropriate neutral loss mass list for singly, doubly and triply phosphorylated peptides. The fragmentation spectra were acquired in the ion trap at normal scan rate by lateral ejection and recorded by the dynode-multiplier system.

For all sequencing events dynamic exclusion was enabled to minimize repeated sequencing. Peaks selected for fragmentation more than once within 30 s were excluded from selection (10 ppm window) for 60 s.

Data processing and analysis

The raw data acquired were processed with the MaxQuant software version and processed as per the standard workflow. Since the HCD spectra were acquired in profile mode, de-isotoping was performed similar to the survey MS scans to obtain singly charged peak lists. Peaks lists generated from the 'quant' module were searched against IPI Human version 3.46 database using the Mascot search engine version 2.2 (Matrix Science, UK) with initial precursor mass tolerance of 7 ppm and fragment mass deviation of 0.02 Da for the 'high-high' strategy. For the 'high-low' strategy these values were 7 ppm and 0.5 Th, respectively. The search included cysteine carbamidomethylation as a fixed modification and N-acetylation of protein and oxidation of methionine as variable modifications. Up to two missed cleavages were allowed for protease digestion and peptide had to be fully tryptic. The 'identify' module in MaxQuant was used to filter identifications at 1% False Discovery Rate (FDR) at three levels namely, site, peptide and protein. As such there is no fixed cut-off score threshold but instead spectra are accepted until the 1% FDR rate is reached[27]. Only peptides with minimum six amino acid length were considered for identification. The identified phosphosites are listed in Supplementary Table 1 and annotated spectra can be visualized as described at the end of the document.

RESULTS AND DISCUSSION

Although the phosphoproteome is very complex, phosphopeptides only constitute a small minority of all peptides after enzymatic degradation of mammalian cell lysates. We digested 20 mg of HeLa cell lysate using the FASP method[22] and divided the resulting peptides into two parts. Both were separated by SCX into 10 fractions (three flow-through and seven SCX fractions). Each fraction was enriched by TiO_2 beads using DHB[28], providing a rich and diverse source of phosphopeptides. To minimize the variations owing to sample processing, the enriched fractions were pooled and for each injection half of the samples were used. Peptides were loaded onto a reverse phase column and separated by 140 min chromatographic runs (100 min gradient time). For mass spectrometric analysis the eluent of the column was electrosprayed into the LTQ Orbitrap Velos instrument. Making use of the more than ten-fold increased ion current and more efficient HCD of this instrument[21], we devised a top10 method consisting of an MS scan with 30,000 resolution at 400 m/z (0.5 s scan) in the Orbitrap analyzer, followed by up to 10 MS/MS scans at 7,500 resolution (0.95 s) also in the Orbitrap analyzer. Total measurement time for phosphoproteome analysis of HeLa cells was one day. In this experiment, through analysis by MaxQuant[27] we identified 16,559 distinct phosphorylation sites with 99% confidence (1% FDR). Of all identified peptides, 76% were phosphorylated. For the 9,668 Class I sites mean localization probability[7] was higher than 0.997 (Supplementary Table 1).

The experiment was then repeated with the other half of the sample but with the difference that CID and a top20 method was used. Due to parallel operation, MS resolution was set to 60,000 at 400 m/z (1 s scan). In this 'high-low' mode and within one day of measuring time we identified 11,893 sites (9,016 Class I sites). For accessing overlap of identifications we also repeated the HCD experiment with another preparation of HeLa cell lysate.

Proportions of pS/pT/pY were similar between HCD and CID (Supplementary Table 1). However, HCD was somewhat more efficient at identifying doubly and more highly phosphorylated peptides than CID (singly:doubly:higher phosphorylated peptides were 51.4%:43.4%:5.2% for HCD whereas they were 64.6%:32.6%:2.8% for CID). Apart from these three large-scale data sets, several other data sets were acquired with parameters described below.

General features of phosphopeptides in high- high strategy

For phosphopeptide detection in HCD mode, peptides are fragmented in the collision cell and the fragment ions are sent back to the C-trap from where they are injected into the Orbitrap analyzer for detection. In CID mode, in contrast, phosphopeptides are isolated and fragmented by 'pseudo MS^3', meaning that both the precursor mass and the mass of potential dominant fragments due to neutral loss of one or more phospho groups are excited together. Furthermore, potential water loss fragments of the precursor are also excited. The resulting fragment spectrum is recorded in the ion linear ion trap at normal scan speed (16,700 Th/s).

From the large-scale experiments described above, we inspected dozens of identified spectra to obtain a qualitative view of the differences between the HCD and the CID spectra. Figure 1 shows representative examples visualizing these differences. The most striking distinction is the mass accuracy of the fragments, which generally deviated 0.1 to 0.3 Da from calculated values for ion trap measurements, whereas they deviated only a few ppm for the Orbitrap measurements expected. Figure 1A shows the HCD spectrum of the doubly phosphorylated peptide PIpSPSPpSAILER, which features the typical characteristics of extensive sequence coverage by y ions and a few low mass b ions (Figure 1A). The a_2b_2 ion pair is among the intense peaks. In the corresponding spectrum of this peptide for CID fragmentation acquired in the ion trap (Figure 1B), the a_2 ion is not recorded owing to the low mass cut-off and the b_2 ion is less than 2% of the base peak intensity and has a mass deviation of 0.104 Da. The noise level in the spectrum is high compared to HCD spectra, which consequently have larger dynamic range (Note that exact noise levels are difficult to compare because the Orbitrap data system employs noise filtering). For peptides of typical length we observe similar amino acid coverage in both CID and HCD. CID spectra have more b-type ions compared to the HCD. This is because b-ions are less stable and fragment further in HCD[21, 29].

For the longer peptides the high resolution of the HCD method allowed unambiguous assignment of charge states of multiply charged fragment ions. Due to the many possible fragmentation pathways, charge states and the low resolution, the CID spectra tend to become crowded and unambiguous peak assignment becomes difficult. This is illustrated by the singly phosphorylated

peptide AAAAAALSGAGTPPAGGGAGGGAGGGGpSPPGGWAVAR, the MS/MS spectra of which appear much cleaner in HCD than in CID. The dynamic range of the HCD spectrum was much higher than in CID as exemplified with one of the fragments annotated in the HCD but not the CID spectrum (Figure 1C and D). Together this led to better amino acid sequence coverage of 84% for HCD as compared to 53% in CID. Of the total ion current in the MS/MS spectra 78% corresponds to annotated peaks in HCD versus 44% in CID. As a consequence the Mascot score is 115.8 for the HCD and 34.4 for the CID spectrum. Furthermore, the Mascot search engine does not take account of mass accuracies better than +- 0.25 Da in the score. If this would be done, the database identification score of the HCD Orbitrap spectra in general would be several fold higher than the scores for CID linear ion trap spectra. Note, however, that the high mass accuracy is still incorporated into database searching, because decoy hits at high mass accuracy are less likely, which has the effect of lowering the score for statistically significant database hits.

High mass accuracy of fragment ions

To assess the overall mass accuracy of in the HCD experiments we plotted the mass deviation between calculated and measured fragments for all identified peptides (1.6 million data points; Figure 2A,C). Almost all the fragment peaks were identified within 20 ppm with 95% of the peaks falling within a 12.5 ppm window. In contrast the fragment mass deviations resulting from ion trap detection were spread much more widely and 95% of the peaks fall within 542 ppm (Figure 2B). Thus HCD mass accuracy was about 50-fold improved over CID with detection in the ion trap.

While fragment mass accuracy in HCD was much higher than in CID, it was nevertheless much lower than the sub-ppm mass accuracy that can be achieved for precursors with Orbitrap measurements[27]. This can mainly be attributed to the fact that precursor ions are measured several times across the elution peak and with higher resolution. To test if the low abundance of phosphopeptides and their fragments influences the mass accuracy, we plotted measured mass deviations against fragment peak intensity (Figure 2C). Mass accuracy is indeed dependent on intensity. However, as the Orbitrap analyzer has a large dynamic range also for mass

measurement accuracy[30], mass deviations do not generally exceed 15 ppm even for very weak peaks.

Cycle time for the top 10 HCD method

One of the obstacles to using high resolution MS/MS for phosphopeptide analysis at a large scale has been their very low abundance, which led to long fill times for peptide fragmentation. In complex peptide mixtures that are separated by HPLC the resulting MS and MS/MS cycle times were previously too long compared to LC peak widths. Our top10 method with 30,000 resolution for MS and 7,500 for MS/MS uses 2.5 s measurement time in the Orbitrap analyzer[21] and should therefore be compatible with the LC time scale. However, fill times to reach the desired target values have to be added to the scan time.

To investigate the actual cycle times in our experiment, we classified all cycles into 10 groups according to number of sequencing events. As this number increased from 1 through 10 the cycle times increased (Figure 3). Maximum fill time for MS/MS was set to 150 ms; therefore the maximum total time for filling of the C-trap and for acquiring all transient should be 4 s. The median cycle time for the full 10 MS/MS events was 3.4 s (the longest cycle time was 4.24s), indicating that relatively fast cycles were achieved even in this case.

With our parameters, only 12% of scan cycles with MS/MS events had all 10 fragmentation events (Supplementary Figure 1A). Many such cycles had only one MS/MS event (35%) and the median number was two. We expected that this number would vary as a function of elution time. Indeed, the median number of MS/MS events per cycle was zero until the first peptides eluted from the column at 20 min, then rose to 10 at 42 min where the density of phosphopeptides was highest and gradually dropped towards zero at the end of the gradient (Supplementary Figure 1B). This indicates that the top10 method was a good choice under these conditions because at all points in the gradient there were sufficient MS/MS events to target the peptides recognized by the instrument data system. However, this does not imply that all peaks were chosen for fragmentation. While the number of MS/MS events peaks at 40-45 minutes the total ion current (TIC) was spread evenly across the gradient for the first but not for later fractions (Supplementary Figure 1C).

The data system had been programmed to accumulate 30,000 ions or – if calculated injection times were longer than 150 ms – to accumulate for that time interval ('maxing out'). We found that due to the low intensity of the phosphopeptides overall a substantial proportion of all MS/MS events reached the 150 ms maximum injection time (Figure 4A). This is in contrast to complex mixture analysis of unmodified peptides on the same instrument, for which we previously determined a mean injection time 8.7 ms with the same HCD target values[21]. Thus sensitivity remains a critical parameter in phosphoproteome analysis; much more so than in the analysis of unmodified peptides. Fortunately, under-filling of the Orbitrap analyzer has no deleterious effects on the mass accuracy apart from reducing the signal of the fragments.

Given the very large number of phosphopeptide identifications of our experiment, we suspected that identification rates were high despite the under-filling. In total, 38.8 % of all MS/MS events led to peptide identifications and 84% of these were phosphopeptides. We first determined the identification percentage of the SCX flow-through fractions, which have more concentrated phosphopeptide populations and therefore had a median injection time of only 118 ms. These fractions indeed had somewhat higher identification rates (47.8%) than non-flowthrough fractions (29.9%). This clearly indicated that increased peptide intensity was beneficial but also that it was possible to identify phosphopeptides with fewer than 30,000 accumulated ions.

Next, we divided the MS/MS spectra into those for which the injection time was within 150 ms and those for which injection time maxed out at 150 ms. Approximately 27.83% of the maxed out fragmentation spectra were still assigned to a peptide sequence and these were binned into four quantile based on the total ion signal in the MSMS spectra (Figure 4B). Interestingly, the apparent total ion signal varied over more than two orders of magnitude and the identification rates were not reduced significantly at least for the two most intense quartiles. These have similar identification rates as compared to the least intense quantile of the spectra that did reach their target values within 150 ms. For the second least intense quantile the identification rate went down to 30% and for the least intense quantile it dropped to less than 15%. This suggested that far less than 30,000 ions were still sufficient for spectra identification in many cases. To investigate this phenomenon in more detail we performed experiments with different target values for each run.

Phosphopeptide identification rates at different target values for MS/MS

To test the identification success rate at defined lower target values we enriched the HeLa lysates for phospho-peptides by TiO_2 beads without any prior separation by SCX. Peptides were separated and analyzed with a 90 min reversed phase LC MS/MS method (70 min gradient). The target value of 30,000 ions served as a reference value for comparison. In duplicate experiments, we decreased target values to 15,000, 10,000 and 5,000 ions. Because the phosphopeptides were concentrated into one fraction median fill times ranged from 6.3 ms for the 30,000 target value to 1.2 ms for the 5,000 target value. Accordingly, only a few percent of the spectra maxed out. The identification rate of MS/MS spectra for the reference target value run was 66%, which is comparable to overall HCD identification rates for non-phosphopeptides in our experience. Remarkably, the number of identified phosphosites dropped only slightly from 1,661 for the reference run at the 30,000 target value to an average of 1,332 for the runs with 5,000 target value (about a 20% reduction in both the number of identifications and the MS/MS identification percentage). For the 10,000 ion target value, the reduction was less than 10% (Table 1). However, the quality of the tandem mass spectra was affected as the target values were reduced. For all different target values and runs 752 sites were identified in common. Mean Mascot or PTM scores[7, 31] reduced to 90% of the reference value for these sites already at the 15,000 ion target value and further reduced to about half at the 5,000 ion target value.

In each of the identified phosphopeptides the site of phosphorylation was assigned and the quality of this assignment quantified by the localization probability, which is the probability value for a site to be phosphorylated among other possible sites in the same peptide[7]. While the Mascot or PTM scores are a measure of quality of identification of phosphopeptides, localization probability is a quality measure of how well the site of phosphorylation can be pinpointed with the available information in the fragmentation spectrum. The median localization probability was larger than 0.99 for the reference run at the 30,000 target value. It did not change appreciably at the 15,000 target value but then declined to about 0.9 for the run with 5,000 target value. This implies that at the lower target values fragmentation peaks that are important to pinpoint the site of phosphorylation start to be lost. This apparently happens before the quality of the spectrum deteriorates such that the peptide cannot be identified with high confidence any more. Cycle

times did not improve markedly between the reference and the low target values (median of 2.71 s vs. 2.67 s).

These findings explain why the overall number and the identification percentage were very high in our large-scale experiment. Despite the fact that many spectra did not reach their target values, their reliable identification was generally not impaired even though the localization confidence dropped for a subset of the spectra.

Together, the above results suggest that top5 to top10 methods with about 30,000 target values and about 150 ms maximum injection times are a good choice for HCD phosphopeptide experiments. Values larger than 10 MS/MS events per cycle or much larger than 150 ms injection times risk extending cycle times to more than 4 s in complex mixtures. In less complex mixtures, however, a large maximum injection time at the expense of the number of MS/MS events may be advantageous. In any case, many phosphopeptides are confidently identified with one tenth or even fewer ions than the target value of 30,000. As a practical point, we have found it important to regularly clean the S-lens as otherwise ion currents can drop drastically (presumably due to charging effects) leading to very long fill times.

Comparison of HCD and CID for large-scale phosphoproteomics

Next we analyzed the second half of the sample on the same machine and with the same methods, except that we fragmented phosphopeptides by CID with analysis in the ion trap instead of HCD and detection in the Orbitrap analyzer. We used a top20 method with a target value of 5,000 ions. Distribution of MS/MS events actually performed by the instrument was similar to the HCD case and there were roughly similar number of full scans (HCD: 91,245; CID: 87,338) and MS/MS scans (HCD 129,468; CID 145,776). In particular, the instrument only performed 10 or more CID scans in 1% of the cases and in most cases only fragmented none or one precursor ion (Supplementary Figure 2).

Interestingly, both fragmentation methods identified similar number of Class I sites (9,668 for HCD and 9,016 for CID) but HCD identified more phosphorylation sites overall (16,559 vs. 11,893). We therefore checked the quality of identification in both methods. For all

phosphorylation states of peptides with Class I sites, the high-high strategy yields better identification quality (Supplementary Figure 3A,B) as the Mascot scores are 30% higher as compared to the high-low strategy (Figure 6) . As mentioned above, this is despite the fact that the Mascot score does not increase with high mass accuracy. For the non-Class I phosphorylation sites the mean peptide Mascot score is still higher in HCD compared to CID but the localization probabilities are not significantly different. Further the CID and HCD runs have near identical peptide intensity distribution indicating that HCD detection in Orbitrap analyzer is not biased against low abundant peptides compared to CID measurement in ion trap (Supplementary Figure 4).

We next compared both data sets with our previous phospho-proteomic study of EGF signaling in HeLa cells where we identified 5,849 Class I sites in a high-low strategy using CID[7]. Overlap with the HCD data was even higher than that with the CID data as we covered 2,985 (51.0%) and 2,408 (41.2%) sites, respectively. Even though the experimental situations are not directly comparable because one experiment employed growth factor stimulation, these results suggest that HCD does not miss specific subsets of phosphosites previously identified in large-scale phosphoproteomics studies.

To address the question of overlap in phosphosites between CID and HCD in more detail we classified phosphopeptides into four quantiles by intensity. As shown in Figure 7 for the most intense quantile, more than 80% of the sites identified in the CID experiment were also identified by HCD measurement whereas in the least intense quantile the overlap is only 35%. When we compared the overlap between two large-scale HCD experiments we found the same result. This demonstrates that the degree of overlap is given by the probability of the peptide to be 'picked for sequencing' by the instrument rather than any difference between the fragmentation methods.

We also combined the results of the two large-scale HCD experiments and this resulted in 13,529 Class I sites which covered all but 2,691 sites of the CID experiment. We were not able to identify any attribute that distinguishes these peptides from the other peptides. Together our results demonstrate that HCD is capable of identifying phosphopeptides to at least the same depths as CID and suggest that it does not discriminate against any particular phosphopeptide populations.

CONCLUSION AND OUTLOOK

Our experiments show that the data quality of HCD makes it an attractive method for phosphopeptide characterization. In contrast to its implementation on previous LTQ-Orbitrap instruments, HCD in conjunction with the S-lens and higher efficiency HCD on the Velos instrument now make it a viable method for large-scale phosphoproteomics. Interestingly despite the somewhat slower measurement cycle of our "high-high" strategy HCD showed similar or even better performance compared to CID. Furthermore, there was no indication of subpopulations of phosphopeptides that were preferentially better suited to CID than to HCD. Therefore, there appears to be no reason against using HCD in a routine manner on the Velos instrument.

Advantages of HCD with the Orbitrap analyzer detection compared to CID with ion trap detection include ready determination of charge states, assignment of neutral loss ions and the presence of reporter ions such as the 216.041 ion for phosphotyrosine. HCD spectra generally have higher dynamic range and are less noisy. Conversely, CID spectra often retain high mass b-ions that are absent in HCD because they fragment further.

The nominal number of ions requested for HCD spectra was 30,000 as opposed to 5,000 for CID. This should have given an advantage to CID for the preferential detection of low abundance phosphopeptides. However, we did not observe such a trend and precursor peptide intensity distributions were identical between HCD and CID. Indeed, phosphopeptides were still efficiently identified with as little as 5,000 ion target values (even though identification and localization scores started to deteriorate). Nevertheless, we found that it is important to load a relatively high amount of phosphopeptides as this increased the percentage of MS/MS spectra identified and the total number of detected phosphosites. Therefore, improvement of sensitivity for phosphopeptide analysis remains an important research and development goal.

Here we have not systematically compared CID with fragment detection in the Orbitrap analyzer to HCD. That approach should also gain from the sensitivity improvement due to the S-lens and should offer many of the advantages of HCD that are related to the excellent fragment mass accuracy. Likewise, comparison of HCD to ETD with low or high resolution[32] would be of

interest on the Velos platform. Another important area for future study is the scoring and localization of phosphogroups using algorithms specifically designed for high mass accuracy data. Database search scores should increase several fold for high resolution MS/MS spectra once this high mass accuracy is taken into account. Additionally, we expect that HCD will further improve relative to CID if MS/MS speed, which is currently only half as fast for HCD as for CID in the ion trap, could be increased.

Acknowledgement

We thank other members of the department for proteomics and signal transduction, for sharing insights. This work was partially supported by PROSPECTS, a 7th Framework grant by the European Directorate (grant agreement HEALTH-F4-2008-201648/PROSPECTS) and by the Max Planck Society for the advancement of Science.

Table 1 Measurement of phosphosites with different target values for MS/MS fragmentation.

Ion Target value	Number of sites identified	Median mascot score	Median PTM score	Median localization probability
30,000	1,661	41.91	114.15	0.997
15,000	1,457	35.69	96.33	0.993
15,000	1,548	39.02	96.33	0.995
10,000	1,461	30.33	80.46	0.978
10,000	1,516	32.01	81.30	0.985
5,000	1,361	22.15	57.67	0.903
5,000	1,304	21.59	56.73	0.912

FIGURE LEGENDS

Figure 1. Comparison of CID and HCD fragment spectra. A. HCD spectrum of the double phosphorylated peptide 'PISphPSPSphAILER'. The peptide has near complete coverage by y ions and low mass b ions. The characteristic a_2,b_2 ion pair is clearly visible (low mass inset) and all peaks are clearly isotope resolved (high mass inset). Mass deviations between calculated and measured fragment masses are given in parentheses. **B**. CID spectrum for the same peptide. Y-ion coverage is less but the spectrum shows high mass b-ions absent from the HCD spectrum. **C.** HCD spectrum of the longer peptide AAAAAALSGAGTPPAGGGAGGGAGGGGpSPPGGWAVAR and the corresponding CID spectrum in **D.** Note that HCD but not CID allows confident charge state assignment of the fragments as seen in the inset for the y_{26}^{2+} ion. The y_6 ion is not present in CID because it was fragmented by pseudo MS^3 of the neutral phospho-loss of the precursor.

Figure 2. Distribution of fragment-ion mass deviations. Mass deviation of fragment ions from calculated values in ppm is plotted as a histogram. HCD data are in blue (**A)** and CID data in red **(B)**. **C**. Mass deviations for 1.8 million fragments of identified phosphopeptides. The two-dimensional density distribution of mass deviation and number of fragment peaks is shown as a function of the fragment peak intensity.

Figure 3. Duty cycle for different number of MS/MS events. The box plots show the distribution of cycle times for scan cycles with different numbers of MS/MS events per cycle. Cycles with no MSMS events are not shown. Even when 10 peptides were sequenced in one cycle, total cycle time is within the chromatographic width of most phosphopeptides. Unlike standard complex peptide mixtures of non-modified peptides the phosphopeptides are not abundant and, this is reflected in the distribution of cycle times using up to 150 ms fill times per MS/MS event. The black bars in the plot represent the median and the circles represent the outliers

Figure 4. Injection times and MS/MS spectra identification rates. **A**. Density distribution of injection time for MS/MS events for scan cycles. **B**. Distribution of total ion signal for all the MS/MS spectra that maxed out the injection time. The distribution is divided into four quantiles. **C**. Identification rates in each quantiles of the histogram are shown. The two most intense quantiles have identification rates of about 40%. The identification rate drops drastically only for the two least intense quantiles of the distribution, which have one or two orders of magnitude fewer ions.

Figure 5. Identification and localization scores as a function of different target values. A. Percentage reduction in the number and quality of identification with respect to the reference ion target value used (30,000). While peptides are still efficiently identified, identification score decreases below a target value of 10,000. **B.** In tandem with the identification scores median localization probability is considerably decreased below 10,000 target value.

Figure 6. Comparison of HCD and CID database identification scores. The histogram of Mascot scores for all Class I sites of HCD (blue) and CID (red) shows a higher average and mean value for HCD. Very high identification scores were only achieved with HCD even though Mascot scores are insensitive to mass accuracy.

Figure 7. Overlap of phosphosites identified with HCD and CID. **A.** All phosphorylation sites were divided into four quantiles based on total intensity of the precursor phosphopeptide ion. 85% of the peptides in the most intense quantile identified in CID were also identified by HCD, whereas only 34% of the ones in the lowest quantile overlapped. **B.** Overlap determined as in A but between the two HCD experiments.

Supplementary figures

Figure S1. **MS/MS events over the gradient for HCD. A.** Bar plot showing the numbers of MS/MS scan events in every duty cycle with at least one fragmentation event. Only 12% of scans have the full 10 fragmentation events. **B**. Distribution of MS/MS events over the gradient. Most sequencing events took place between 20 and 80 minutes. **C.** Distribution of Total Ion Current (TIC) over the gradient.

Figure S1. MS/MS events over the gradient for CID. A. Bar plot showing the numbers of MS/MS scan events in every duty cycle with at least one fragmentation event. Only 1% of scans have more than 10 fragmentation events even though a top20 method was used. **B**. Distribution of MS/MS events over the gradient. Most sequencing events took place between 20 and 80 minutes. **C.** Distribution of Total Ion Current (TIC) over the gradient.

Figure S3. Localization probabilities for phosphorylation sites in HCD. A. For all phosphorylation states HCD has a better Mascot score than CID. **B.** For all phosphorylation states HCD has a better localization probability than CID.

Figure S4. Density plot of phosphopeptide precursors. Intensities of peptides from HCD (blue) and CID (red) are the same, demonstrating that HCD is not biased to high abundance peptides as compared to CID.

REFERENCES

1. Cohen, P., The regulation of protein function by multisite phosphorylation - a 25 year update. *Trends in Biochemical Sciences* **2000,** 25, (12), 596-601.

2. Mann, M.; Jensen, O. N., Proteomic analysis of post-translational modifications. *Nature Biotechnology* **2003,** 21, (3), 255-261.

3. Witze, E. S.; Old, W. M.; Resing, K. A.; Ahn, N. G., Mapping protein post-translational modifications with mass spectrometry. *Nat Methods* **2007,** 4, (10), 798-806.

4. Grimsrud, P. A.; Swaney, D. L.; Wenger, C. D.; Beauchene, N. A.; Coon, J. J., Phosphoproteomics for the Masses. *Acs Chemical Biology* **2010,** 5, (1), 105-119.

5. Nita-Lazar, A.; Saito-Benz, H.; White, F. M., Quantitative phosphoproteomics by mass spectrometry: past, present, and future. *Proteomics* **2008,** 8, (21), 4433-43.

6. Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P., Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* **2004,** 101, (33), 12130-5.

7. Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M., Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **2006,** 127, (3), 635-648.

8. Olsen, J. V.; Vermeulen, M.; Santamaria, A.; Kumar, C.; Miller, M. L.; Jensen, L. J.; Gnad, F.; Cox, J.; Jensen, T. S.; Nigg, E. A.; Brunak, S.; Mann, M., Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci Signal* **2010,** 3, (104), ra3.

9. Schroeder, M. J.; Shabanowitz, J.; Schwartz, J. C.; Hunt, D. F.; Coon, J. J., A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. *Anal Chem* **2004,** 76, (13), 3590-8.

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

11. Louris, J. N.; Cooks, R. G.; Syka, J. E. P.; Kelley, P. E.; Stafford, G. C.; Todd, J. F. J., INSTRUMENTATION, APPLICATIONS, AND ENERGY DEPOSITION IN QUADRUPOLE ION-TRAP TANDEM MASS-SPECTROMETRY. *Analytical Chemistry* **1987,** 59, (13), 1677-1685.

12. Annan, R. S.; Huddleston, M. J.; Verma, R.; Deshaies, R. J.; Carr, S. A., A multidimensional electrospray MS-based approach to phosphopeptide mapping. *Anal Chem* **2001,** 73, (3), 393-404.

13. Steen, H.; Kuster, B.; Fernandez, M.; Pandey, A.; Mann, M., Tyrosine phosphorylation mapping of the epidermal growth factor receptor signaling pathway. *J Biol Chem* **2002,** 277, (2), 1031-9.

14. Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W., Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem* **2000,** 72, (3), 563-73.

15. Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* **2004,** 101, (26), 9528-33.

16. Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A., Electron capture dissociation of singly and multiply phosphorylated peptides. *Rapid Commun Mass Spectrom* **2000,** 14, (19), 1793-800.

17. Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M., Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods* **2007,** 4, (9), 709-712.

18. Zhang, Y.; Askenazi, M.; Jiang, J.; Luckey, C. J.; Griffin, J. D.; Marto, J. A., A robust error model for iTRAQ quantification reveals divergent signaling between oncogenic FLT3 mutants in acute myeloid leukemia. *Mol Cell Proteomics* 9, (5), 780-90.

19. Zougman, A.; Pilch, B.; Podtelejnikov, A.; Kiehntopf, M.; Schnabel, C.; Kurnar, C.; Mann, M., Integrated analysis of the cerebrospinal fluid peptidome and proteome. *Journal of Proteome Research* **2008,** 7, (1), 386-399.

20. Zhang, Y.; Ficarro, S. B.; Li, S. J.; Marto, J. A., Optimized Orbitrap HCD for Quantitative Analysis of Phosphopeptides. *Journal of the American Society for Mass Spectrometry* **2009,** 20, (8), 1425-1434.

21. Olsen, J. V.; Schwartz, J. C.; Griep-Raming, J.; Nielsen, M. L.; Damoc, E.; Denisov, E.; Lange, O.; Remes, P.; Taylor, D.; Splendore, M.; Wouters, E. R.; Senko, M.; Makarov, A.; Mann, M.; Horning, S., A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. *Molecular & Cellular Proteomics* **2009,** 8, (12), 2759-2769.

22. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nature Methods* **2009,** 6, (5), 359-U60.

23. Nagaraj, N.; Lu, A.; Mann, M.; Wisniewski, J. R., Detergent-based but gel-free method allows identification of several hundred membrane proteins in single LC-MS runs. *J Proteome Res* **2008,** 7, (11), 5028-32.

24. Li QR, N. Z., Tang JS, Nie S, Zeng R, Effect of peptide-to-TiO2 beads ratio on phosphopeptide enrichment selectivity. *Journal of Proteome Research* **2009,** 8, (11), 5375-81.

25. Rappsilber, J.; Ishihama, Y.; Mann, M., Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **2003,** 75, (3), 663-70.

26. Olsen, J. V.; de Godoy, L. M. F.; Li, G. Q.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M., Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Molecular & Cellular Proteomics* **2005,** 4, (12), 2010-2021.

27. Cox, J.; Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* **2008,** 26, (12), 1367-1372.

28. Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jorgensen, T. J., Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics* **2005,** 4, (7), 873-86.

29. Sleno, L.; Volmer, D. A., Ion activation methods for tandem mass spectrometry. *J Mass Spectrom* **2004,** 39, (10), 1091-112.

30. Makarov, A.; Denisov, E.; Lange, O.; Horning, S., Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. *J Am Soc Mass Spectrom* **2006,** 17, (7), 977-82.

31. Olsen, J. V.; Mann, M., Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** 101, (37), 13417-13422.

32. Wenger, C. D.; McAlister, G. C.; Xia, Q.; Coon, J. J., Sub-part-per-million precursor and product mass accuracy for high-throughput proteomics on an ETD-enabled Orbitrap mass spectrometer. *Mol Cell Proteomics* **2010**, in press.

Supplementary material

Supplementary Table 1_A:

Information about all phosphopeptides from the large scale experiment HCD1

Supplementary Table 1_B:

Information about all phosphopeptides from the large scale experiment CID1

Supplementary Table 1_C:

Information about all phosphopeptides from the large scale experiment HCD2

Viewing annotated spectra for any phosphorylation site:

Annotated phosphorylation spectra can be viewed at TRANCHE (www.proteomecommons.org) using the hash key below. For any phosphorylation site of interest, note the raw file name and scan number from Supplementary Table 1 and open the file with this information as the concatenated name. An annotated MS/MS spectrum in png format will appear (png files can be

viewed in almost all imaging applications, for example Adobe Illustrator or Microsoft picture manager).

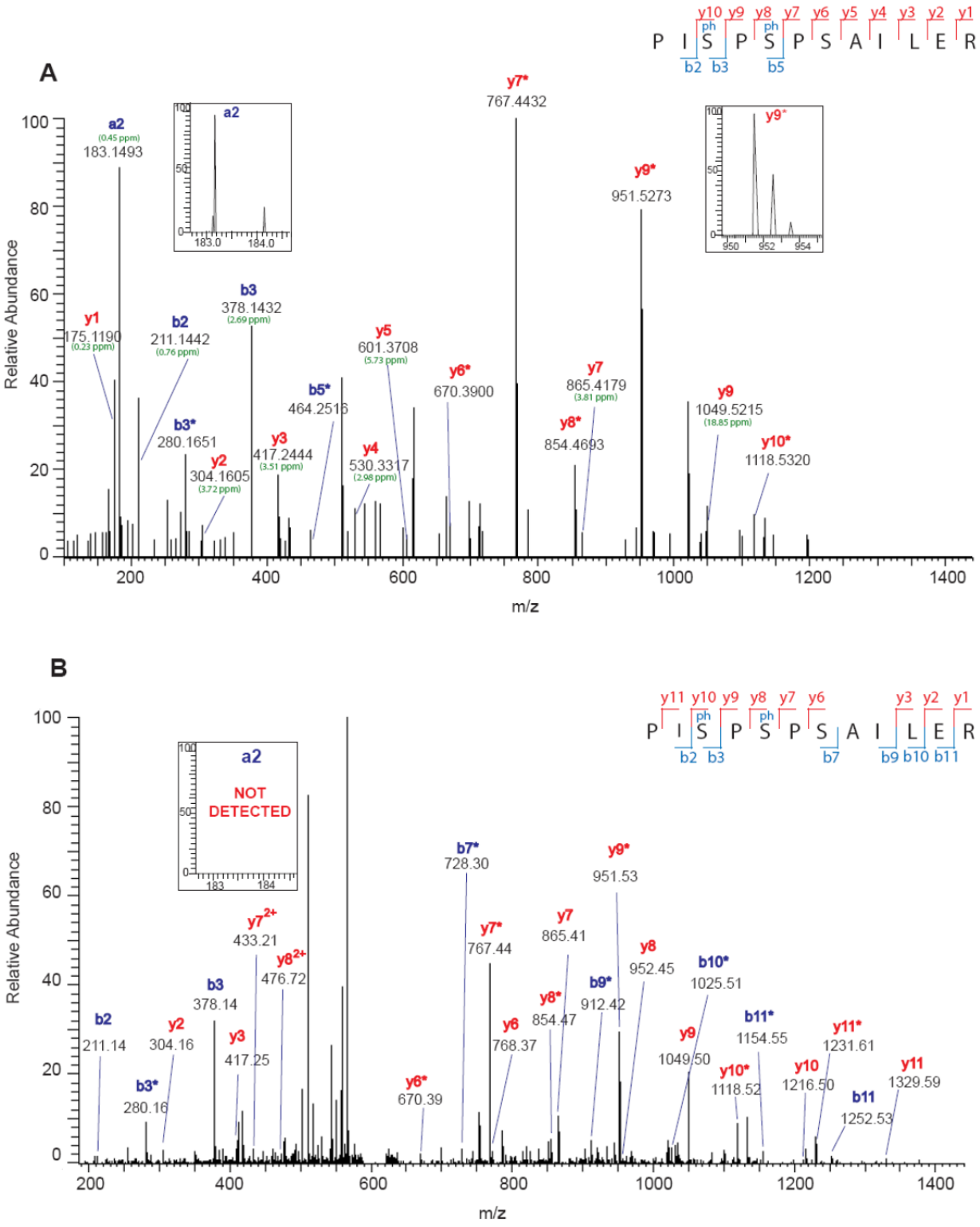
Accession codes for raw mass spectrometry files:

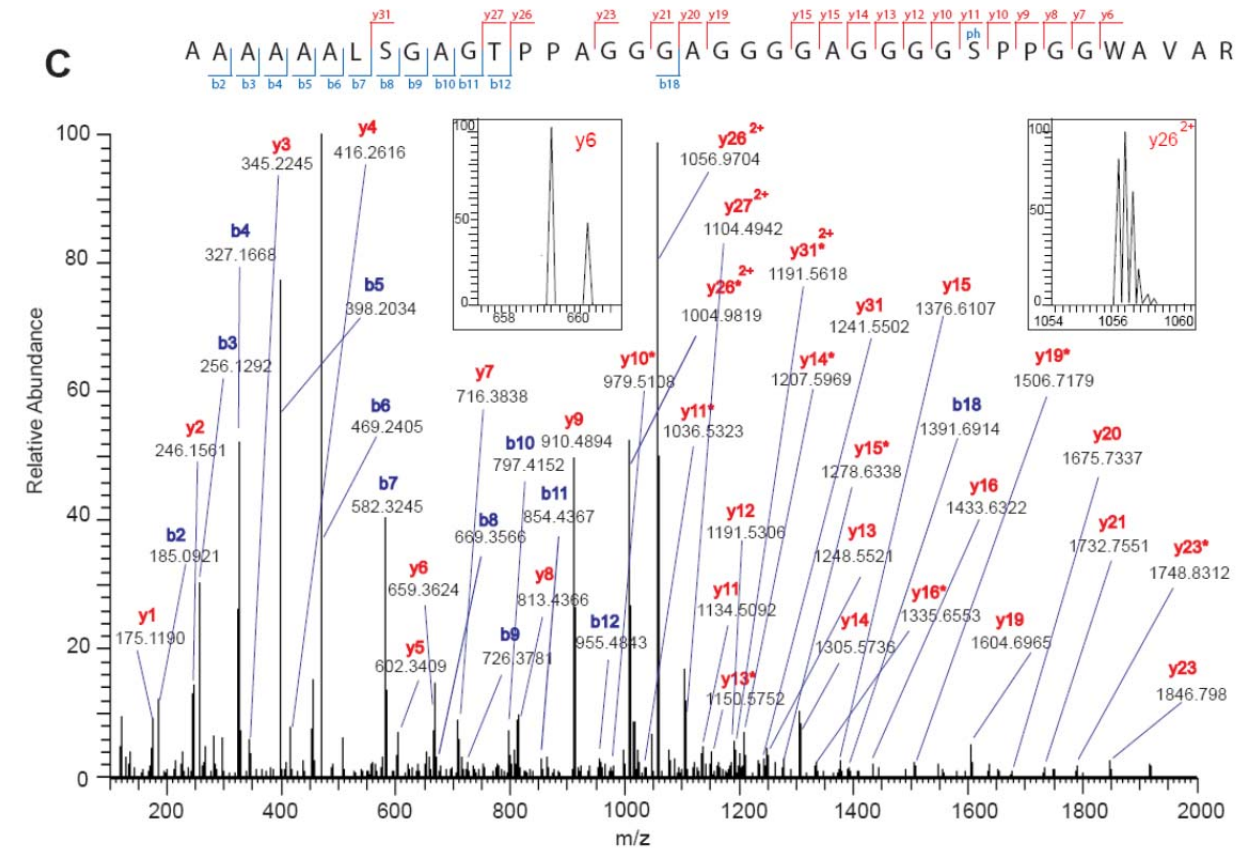
The data associated with this manuscript may be downloaded from ProteomeCommons.org Tranche using the following hash: **/Gyf6Csx8Xlx8aUTof4/OcFDVdL3TDb6J4UPLceZSTXL2kdZr9OUb5j6NdIduK6+ehqHJ3Td9GZSQTKaDtUM4/gMsNYAAAAAAAATLQ==**

For the annotated phospho HCD spectra (more than 30,000):

w/ouOjs+THet/KE1R91+8O3zh/4xQ2X8I5AvTslHFC3yvcC9tF/5qIOJhDDhkYNuLXfLwCObF1tvbav4Y5eG7hLPZe8AAAAAAFKvRQ==

Figure 1


A
P I S P S P S A I L E R
y10 y9 y8 y7 y6 y5 y4 y3 y2 y1
ph ph
b2 b3 b5
a2
183.1493
y7*
767.4432
y9*
951.5273
b3
378.1432
y1
175.1190
b2
211.1442
y5
601.3708
y6*
670.3900
y7
865.4179
b5*
464.2516
y9
1049.5215
b3*
280.1651
y3
417.2444
y8*
854.4693
y4
530.3317
y2
304.1605
y10*
1118.5320
Relative Abundance
m/z
B
P I S P S P S A I L E R
y11 y10 y9 y8 y7 y6 y3 y2 y1
ph ph
b2 b3 b7 b9 b10 b11
a2
NOT
DETECTED
b7*
728.30
y9*
951.53
y7
865.41
y7*
767.44
y8
952.45
b10*
1025.51
b9*
912.42
y8*
854.47
b11*
1154.55
y11*
1231.61
y6
768.37
y9
1049.50
y10
1216.50
y11
1329.59
y10*
1118.52
b11
1252.53
y7 2+
433.21
y8 2+
476.72
b3
378.14
b2
211.14
y2
304.16
y3
417.25
b3*
280.16
y6*
670.39
Relative Abundance
m/z


C
A A A A A A L S G A G T P P A G G G A G G G G A G G G G S P P G G W A V A R
Relative Abundance
m/z
y6
y26 2+
y3 345.2245
y4 416.2616
y26 2+ 1056.9704
y27 2+ 1104.4942
y31* 2+ 1191.5618
y26* 2+ 1004.9819
y15 1376.6107
y31 1241.5502
b4 327.1668
b5 398.2034
b3 256.1292
y10* 979.5108
y14* 1207.5969
y19* 1506.7179
y7 716.3838
b18 1391.6914
b6 469.2405
y11* 1036.5323
y9 910.4894
y20 1675.7337
y2 246.1561
b10 797.4152
y15* 1278.6338
b11 854.4367
y16 1433.6322
b7 582.3245
y12 1191.5306
y21 1732.7551
y23* 1748.8312
b8 669.3566
y13 1248.5521
b2 185.0921
y6 659.3624
y8 813.4366
y11 1134.5092
y16* 1335.6553
y19 1604.6965
y1 175.1190
b12 955.4843
y14 1305.5736
y5 602.3409
b9 726.3781
y13* 1150.5752
y23 1846.798

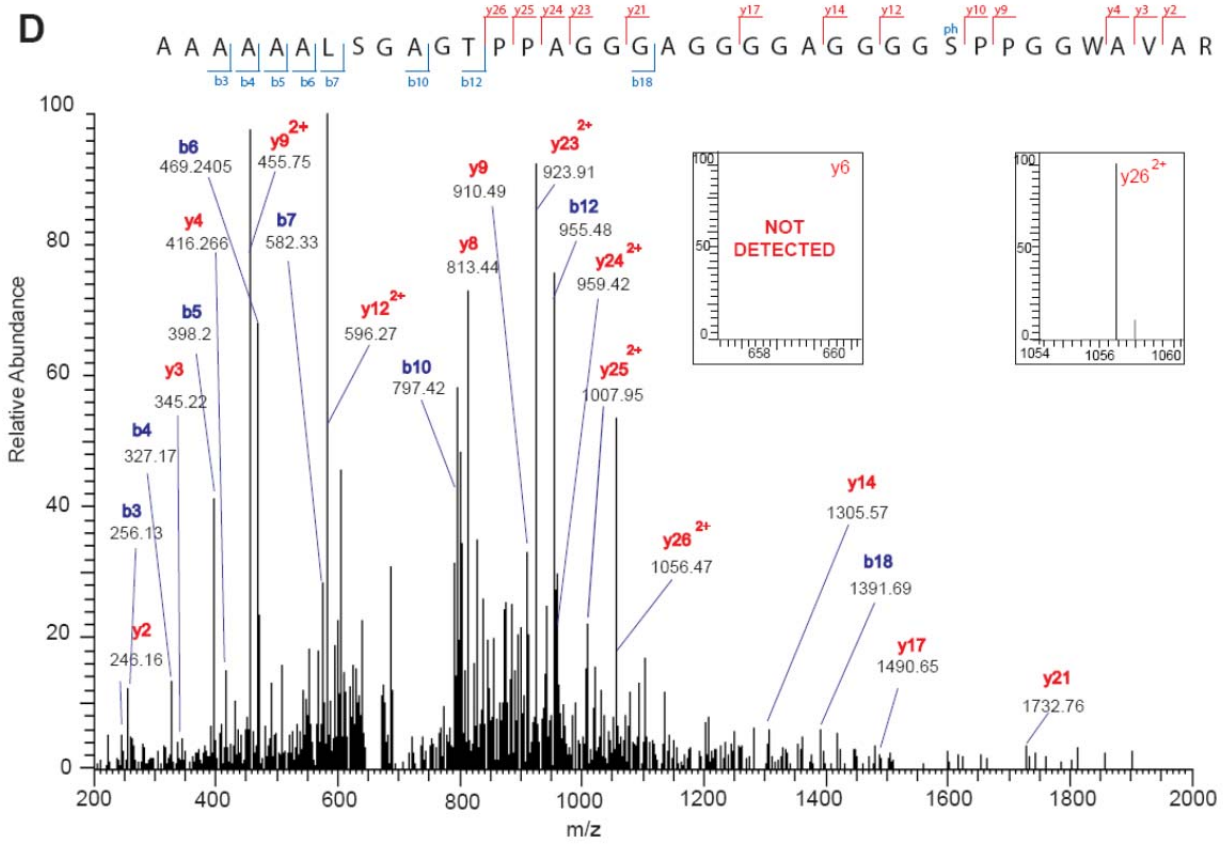

D
A A A A A A L S G A G T P P A G G G A G G G G A G G G G S P P G G W A V A R
Relative Abundance
m/z
y6
NOT DETECTED
y26 2+
b6 469.2405
y9 2+ 455.75
y23 2+ 923.91
y9 910.49
b12 955.48
y4 416.266
b7 582.33
y8 813.44
y24 2+ 959.42
y12 2+ 596.27
b5 398.2
y25 2+ 1007.95
y3 345.22
b10 797.42
b4 327.17
y14 1305.57
b3 256.13
y26 2+ 1056.47
b18 1391.69
y2 246.16
y17 1490.65
y21 1732.76

Figure 2

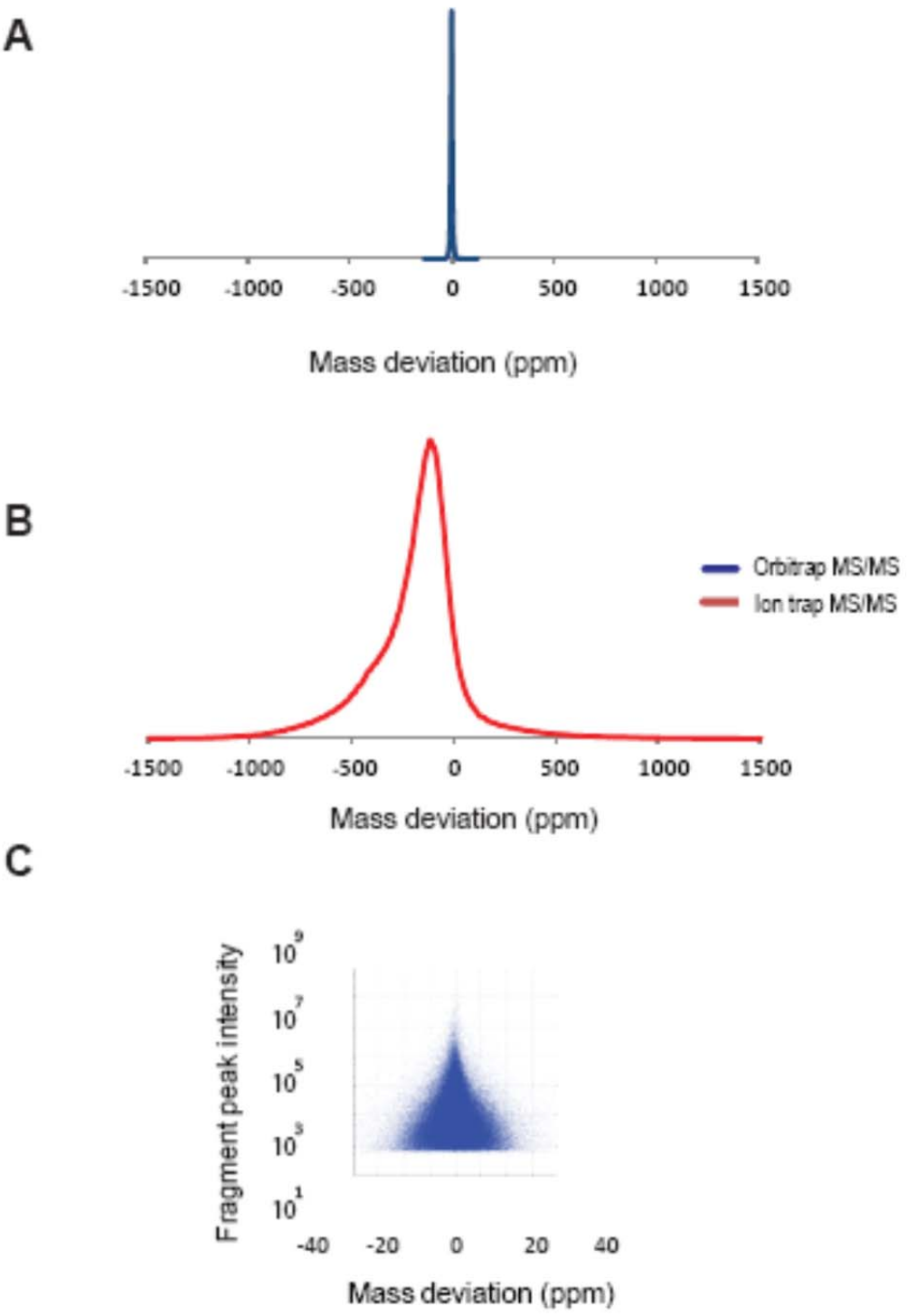

A
-1500
-1000
-500
0
500
1000
1500
Mass deviation (ppm)
B
Orbitrap MS/MS
Ion trap MS/MS
-1500
-1000
-500
0
500
1000
1500
Mass deviation (ppm)
C
Fragment peak intensity
10^9
10^7
10^5
10^3
10^1
-40
-20
0
20
40
Mass deviation (ppm)

Figure 3

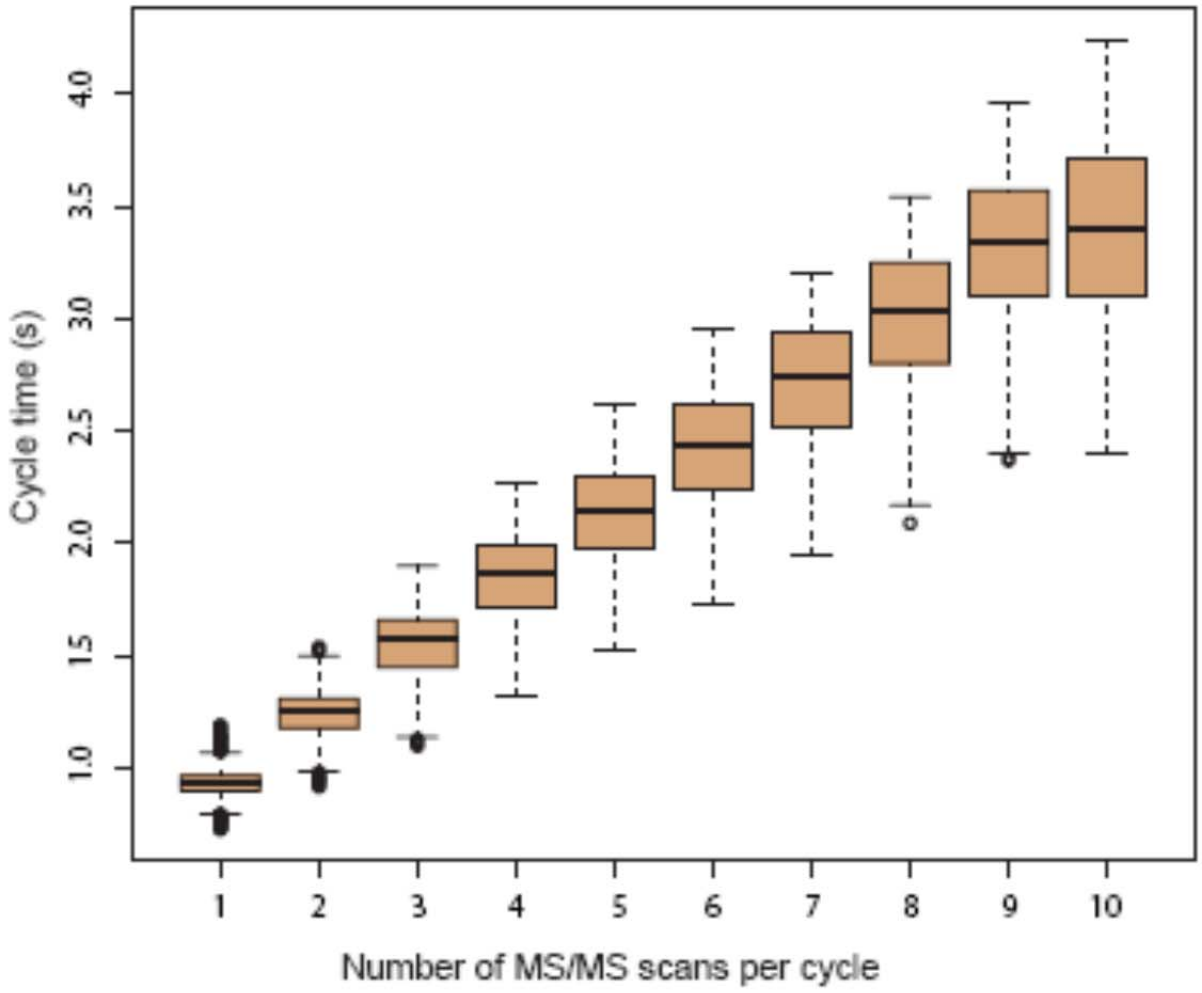

Cycle time (s)
1.0
1.5
2.0
2.5
3.0
3.5
4.0
1
2
3
4
5
6
7
8
9
10
Number of MS/MS scans per cycle

Figure 4

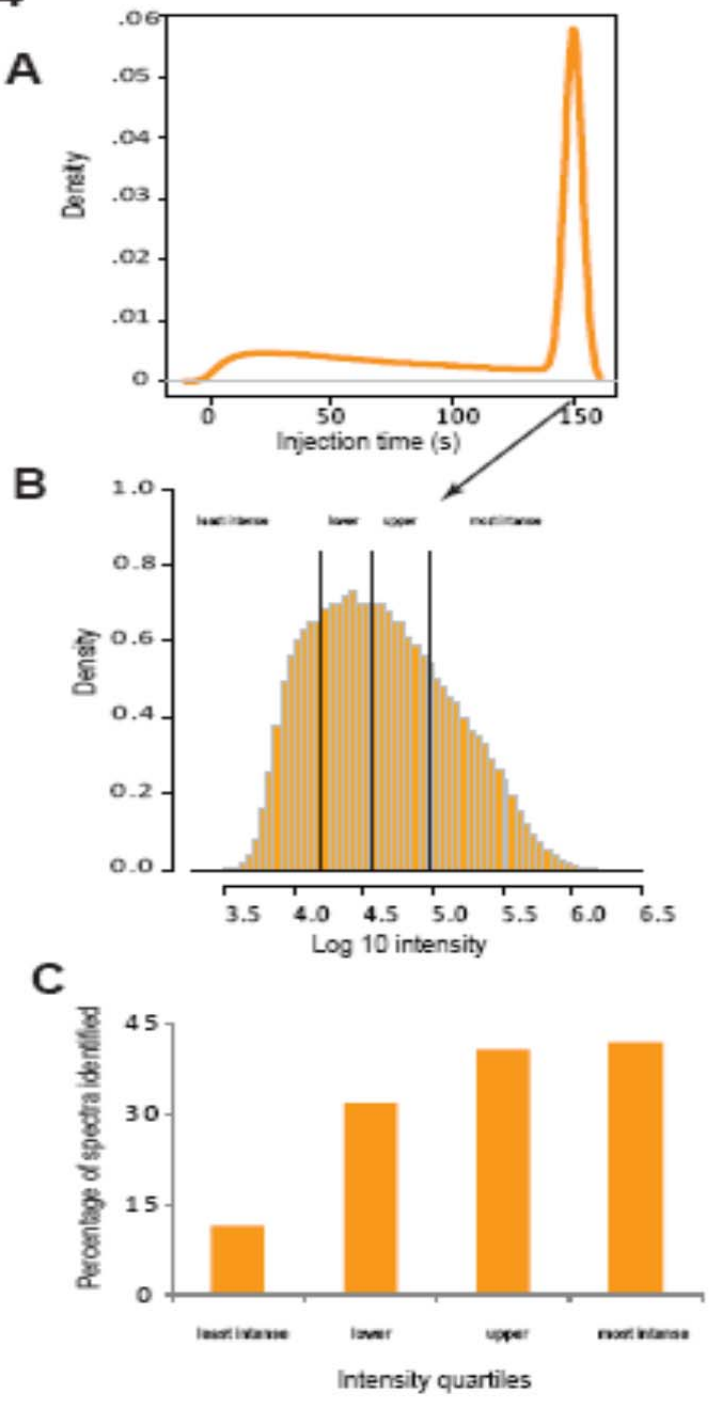
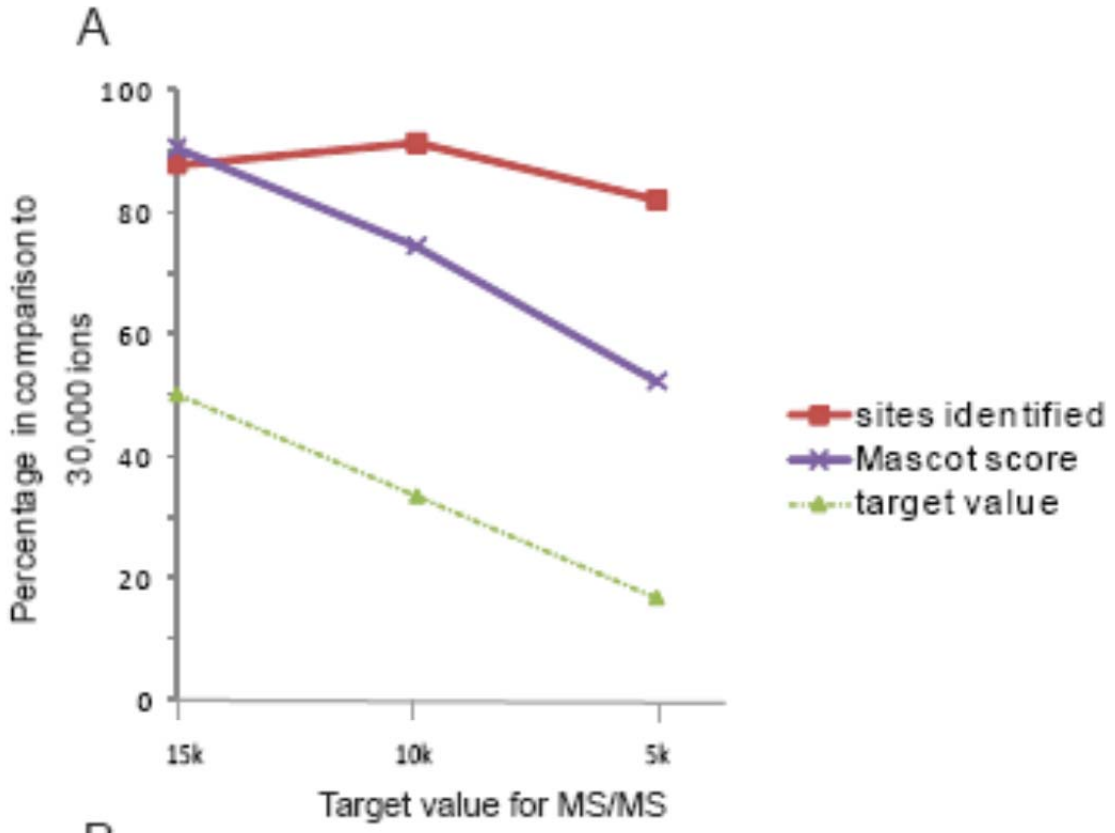

A
.06
.05
.04
.03
.02
.01
0
Density
0
50
100
150
Injection time (s)
B
1.0
0.8
0.6
0.4
0.2
0.0
Density
3.5
4.0
4.5
5.0
5.5
6.0
6.5
Log 10 intensity
C
45
30
15
0
Percentage of spectra identified
Intensity quartiles

Figure 5


A
100
80
60
40
20
0
Percentage in comparison to 30,000 ions
15k
10k
5k
Target value for MS/MS
sites identified
Mascot score
target value

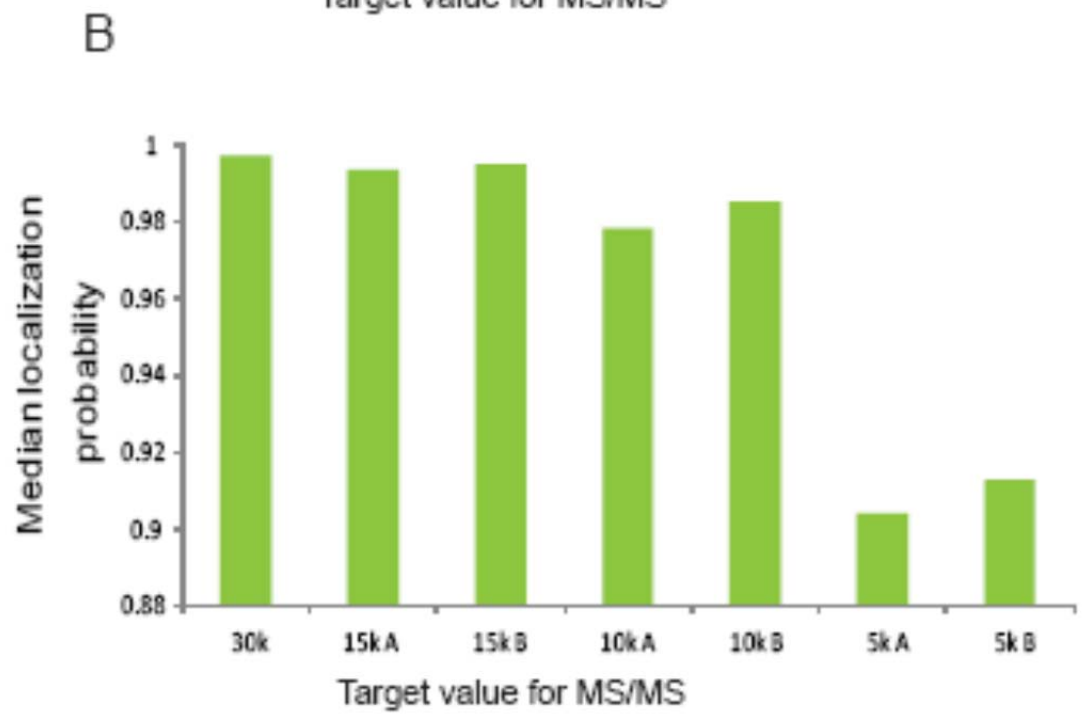

B
1
0.98
0.96
0.94
0.92
0.9
0.88
Median localization probability
30k
15k A
15k B
10k A
10k B
5k A
5k B
Target value for MS/MS

Figure 6

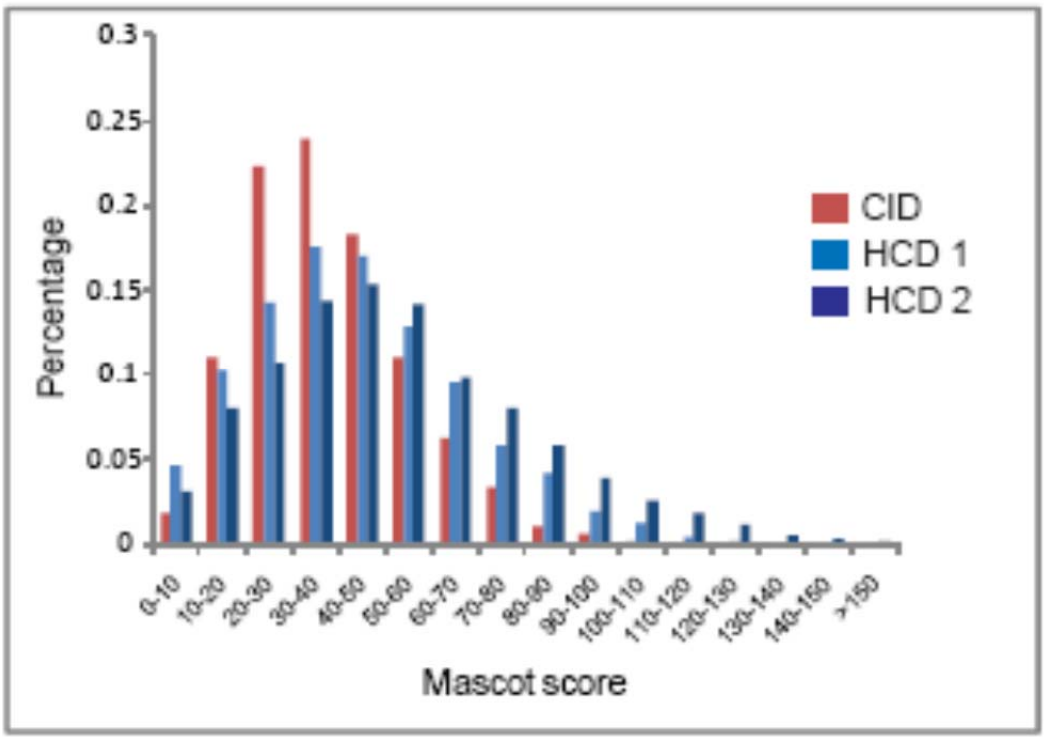
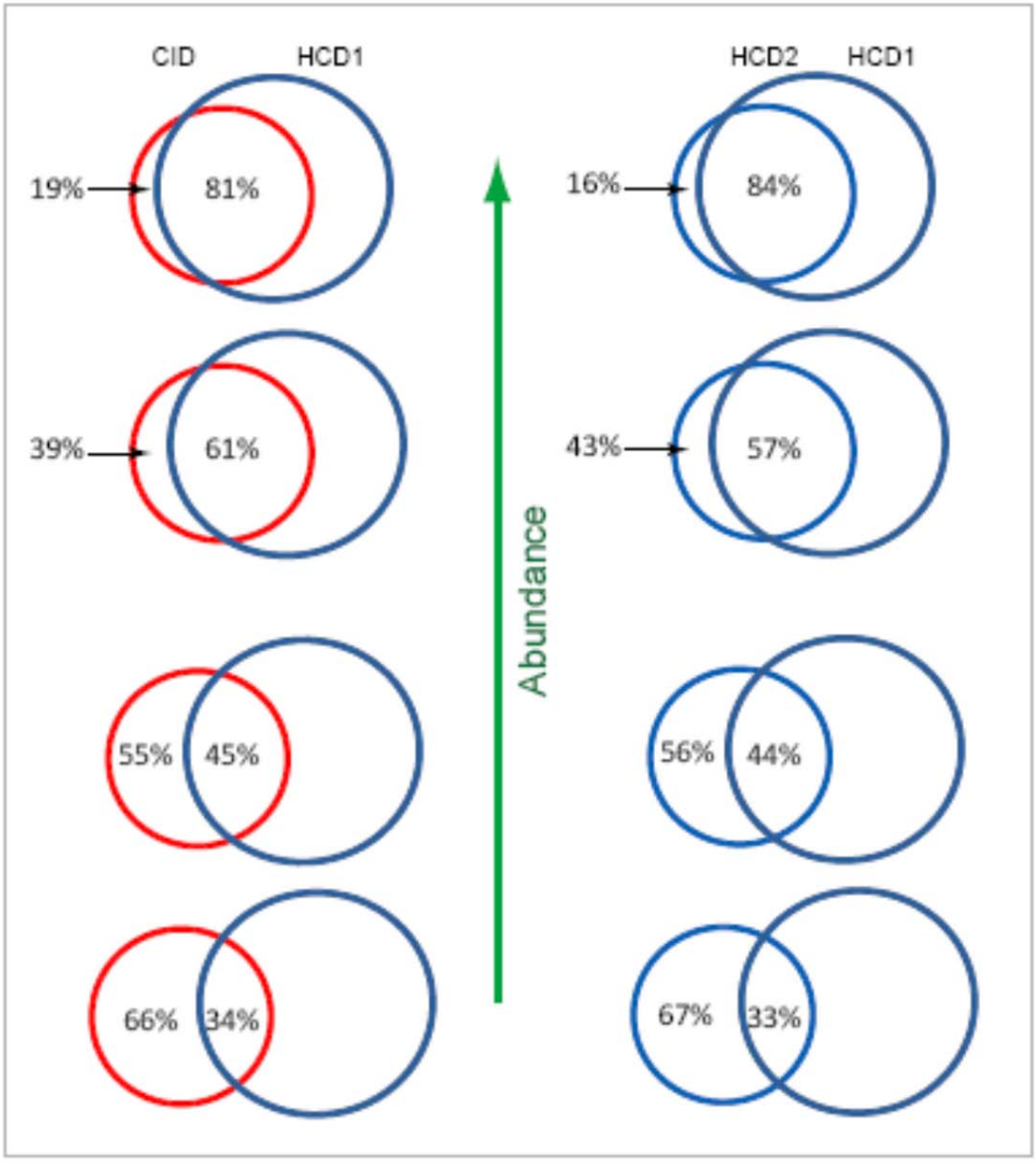
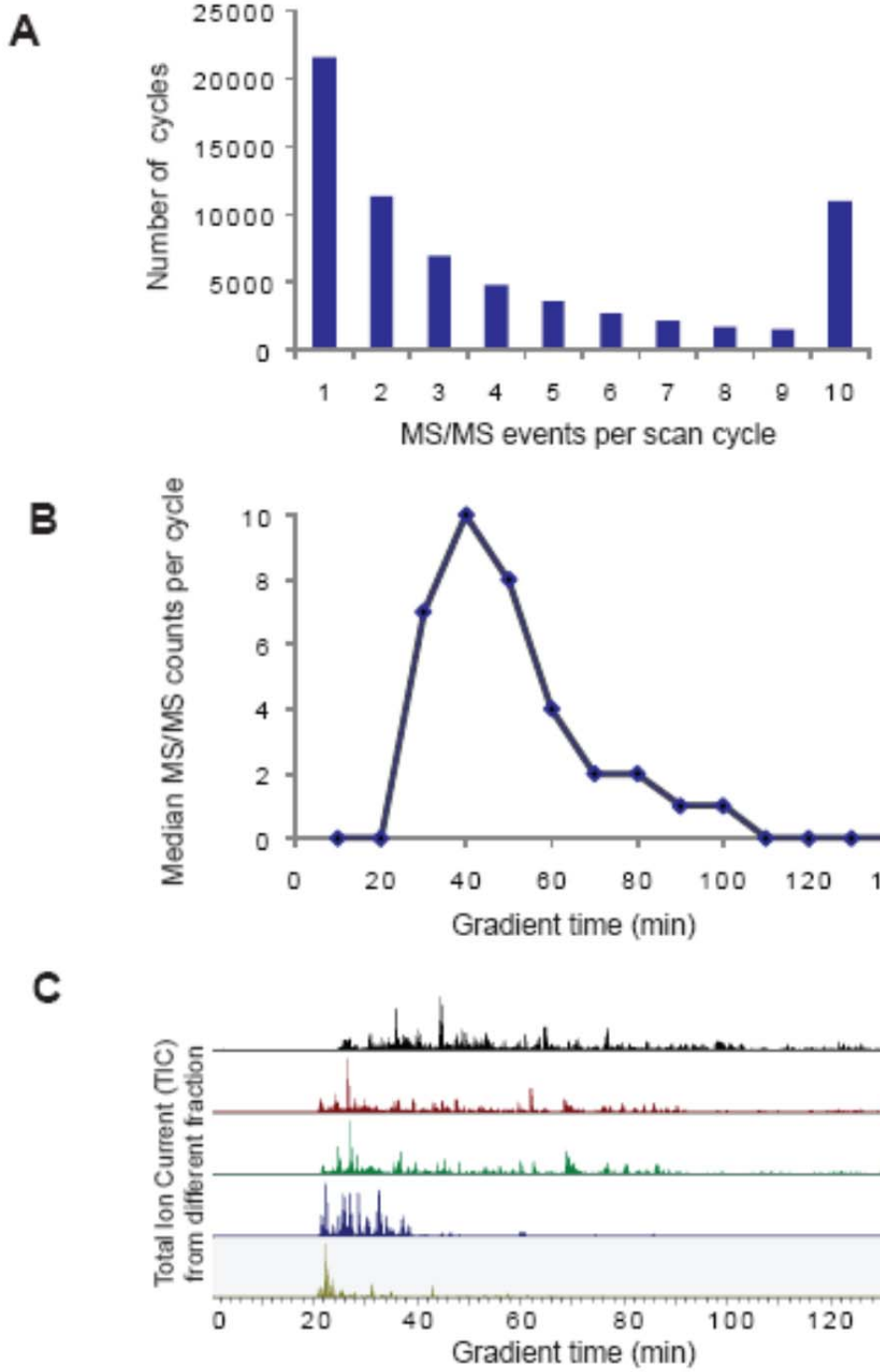

0.3
0.25
0.2
0.15
0.1
0.05
0
Percentage
CID
HCD 1
HCD 2
0-10
10-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100
100-110
110-120
120-130
130-140
140-150
>150
Mascot score

Figure 7

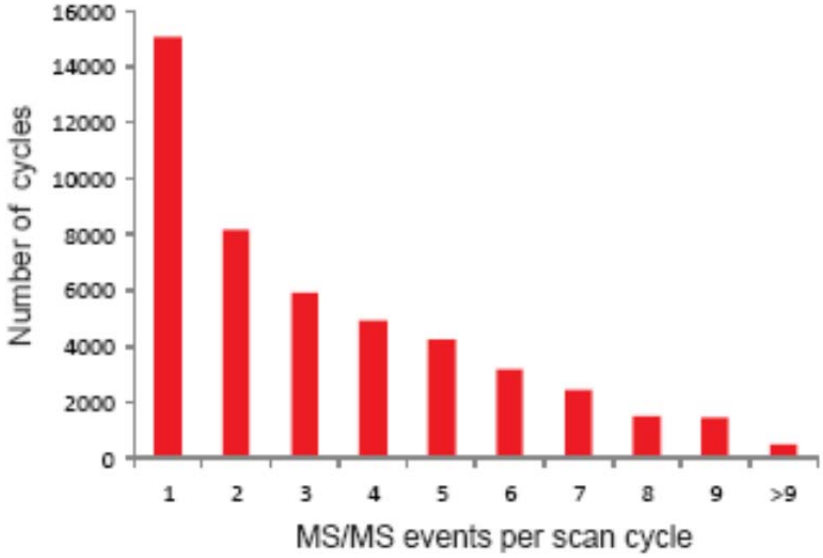

CID
HCD1
19%
81%
39%
61%
55%
45%
66%
34%
Abundance
HCD2
HCD1
16%
84%
43%
57%
56%
44%
67%
33%

Supplementary figure 1

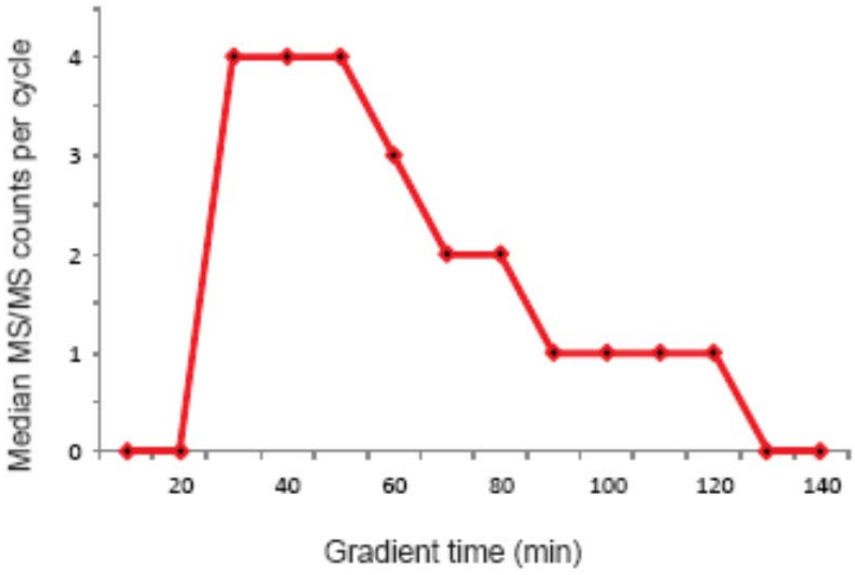


Supplementary figure 2

A

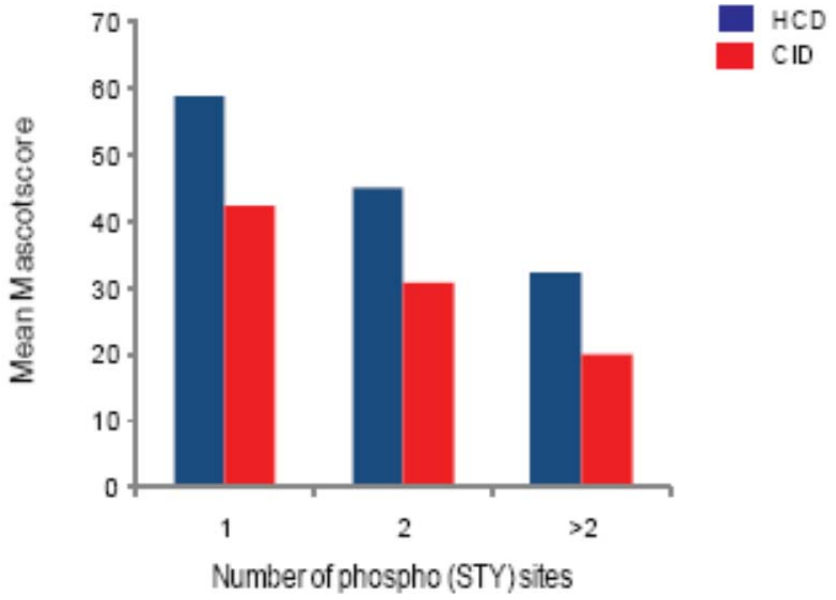


B

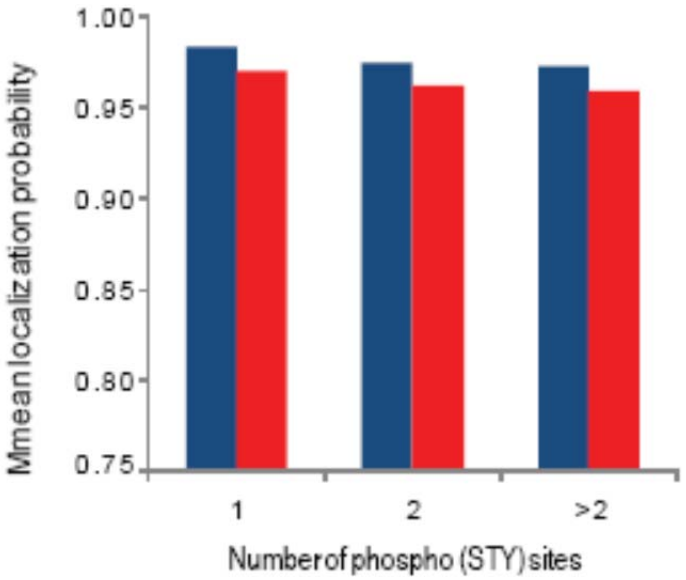


Supplementary Figure 3

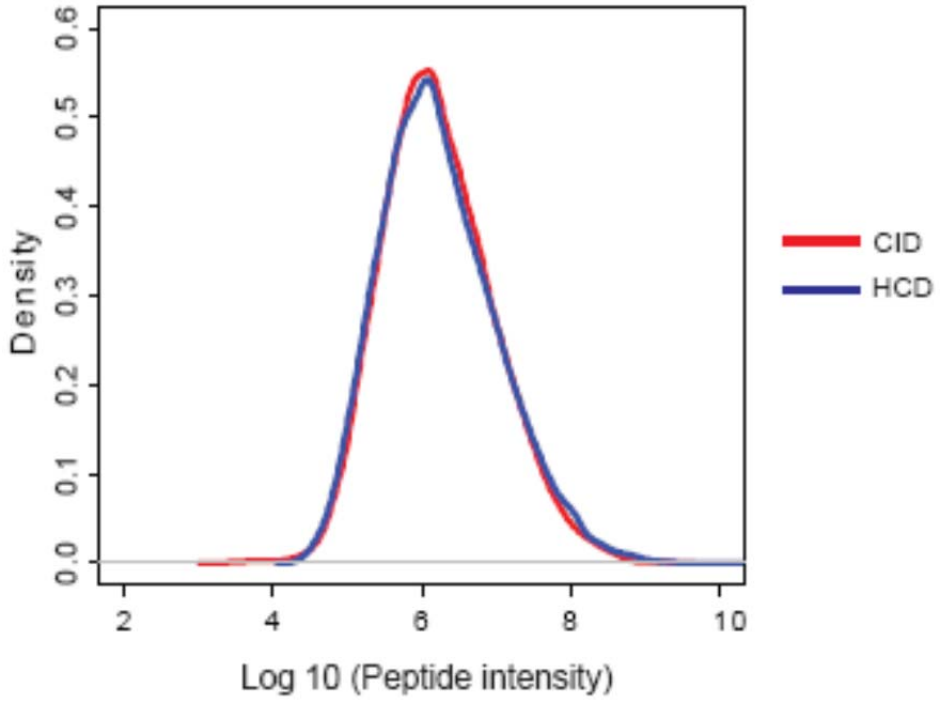
A



B



Supplementary figure 4



2.4 LC-MS/MS based platform for urinary proteomics and investigation of variation of normal human urinary proteome

Urine is a readily available non-invasive source for clinical biomarker discovery. Large scale proteome mapping studies have shown that urine contains more than 3,000 proteins and thus it is rich in information for biomarker discovery[125, 126, 153]. These studies involved pooling different samples and extensive fractionation in order to achieve greater depth of proteome coverage. However, a proteomics based biomarker discovery platform requires analyzing large number of samples and thus fractionation is not an optimal choice. Furthermore, before comparing the urinary proteome of normal versus patient samples, it is important to investigate the extent of variation of urinary proteome among normal individuals, and also to study the variation of a single person's urinary proteome over time. A previous study reported on the extent of random variation of the urinary proteome but it analyzed less than 50 proteins[154]. In this project we show identification of 600 to 800 proteins in single run LC-MS/MS without any pre-fractionation techniques and we quantify of about 600 proteins across seven individuals using label free quantification. We observe that label-free quantification is highly reproducible and accurate and that the urinary proteome varies significantly within the same person over three consecutive days. This has to be taken in to account in large scale biomarker studies.

A manuscript on this project has been submitted to the Journal of proteome research as follows

Quantitative analysis of the intra and inter individual variability of the normal urinary proteome

Nagarjuna Nagaraj and Matthias Mann*

Department for Proteomics and Signal Transduction at the Max-Planck Institute of Biochemistry, 82152 Martinsried, Germany

Running title: Variability of the normal urinary proteome

Keywords: Urinary proteome, label free quantification, technical and biological variation, absolute quantitation

*Corresponding author:

Matthias Mann

Department of Proteomics and Signal Transduction

Max-Planck Institute for Biochemistry

Am Klopferspitz 18

D-82152 Martinsried

Germany

Email: mmann@biochem.mpg.de

Fax: +49 89 8578 2219

Abstract

Urine is a readily and non-invasively obtainable body fluid. Mass spectrometry (MS)-based proteomics has shown that urine contains thousands of proteins. It is a potential source of biomarkers for diseases of proximal and distal tissues but it is thought to be more variable than the more commonly used plasma. By LC MS/MS analysis on an LTQ-Orbitrap without pre-fractionation we characterized the urinary proteome of seven normal human donors over three consecutive days. Label free quantification of triplicate single runs covered the urinary proteome to a depth of more than 600 proteins. The median coefficient of variation (cv) of technical replicates was 0.18. Inter-day variability was markedly higher with a cv of 0.48 and the overall variation of the urinary proteome between individuals was 0.66. Thus technical variability in our data was 7.5% whereas intra-personal variability contributed 45.5% and inter-personal variability contributed 47.1% to total variability. Determination of the normal fluctuation of individual urinary proteins should be useful in establishing significance thresholds in biomarker studies. Our data also allowed definition of a common and abundant set of 500 proteins that are readily detectable in all studied individuals. This core urinary proteome has a high proportion of secreted, membrane and relatively high molecular weight proteins.

Introduction

Human urine is one of the major body fluids and its use in medical diagnostics is second only to plasma[1]. It should therefore be an attractive source for proteomics based biomarker discovery[2-4]. Urinary proteomics has been pursued for a number of years, first with techniques like 2D gel electrophoresis[5, 6] and SELDI[7, 8]. However, coverage of the proteome was very limited and robust quantitation was lacking. As an alternative to the proteome, the peptidome has also been analyzed early on, often by capillary electrophoresis coupled to MS or MS/MS instruments[9-11].

During the last few years great technological advances have occurred in proteomics, especially in workflows based on liquid chromatography – tandem mass spectrometry (LC-MS/MS). MS-based proteomics is now routinely applied to the large-scale analysis of cellular proteomes[12-15]. It has also led to the identification of a large number of proteins in the urinary proteome[16, 17]. Employing a range of fractionation techniques and high resolution MS we previously reported more than 1,500 proteins in this body fluid[18] and Steen and co-workers identified 2,362 proteins – the largest urinary proteome to date[19]. Apart from showing that many disease-related proteins can be identified in urine, these studies uncovered a great proportion of membrane and high molecular weight proteins. This was unexpected given the 40 kDa cutoff of the glomerular filtration apparatus in the kidney but may be explained by the presence of extracellular vesicles – termed exosomes. In fact, dedicated exosome urinary proteomics has resulted in the identification of more than 1000 proteins in this fraction alone[20, 21].

The ultimate goal of urinary proteomics is the identification of biomarkers. Efforts so far have mainly focused on diseases of proximal tissues, such as bladder, prostate and kidney. Potential biomarkers have been reported for bladder cancer, rejection of kidney transplants and renal failure in diabetic nephropathy (for examples see refs.[22-30]). However, the obstacles to establish true biomarkers are daunting and most candidates proposed so far are not likely to be specific to the disease under investigation. Clearly the proteomics workflows employed need to be further developed in sensitivity, throughput and quantitative accuracy. This is also the conclusion of large community efforts to study body fluids such as human plasma[31, 32].

A central question to be addressed in urinary proteomics is the variability of the proteome. Because most of the proteins do not have a functional role in this fluid, there is no physiological need for precise homeostatic control. As a result, variability of the urinary proteome is thought to be much larger than in plasma. The quantitative degree of urinary proteome variability - or any other body fluid - has, however, not been established. Lee et al. investigated the dynamic urinary proteome during rat development[33]. They found that the presence of some proteins appeared to be associated with specific developmental stages. Sun et al used low resolution shotgun proteomics to identify proteins in urine collected from a number of individuals during the course of a single day and between days[34]. Only very few proteins were identified in common between the individuals and quantification was performed by spectral counting, limiting the conclusions that could be drawn from the data.

Studying the variability of a body fluid proteome, in common with biomarker studies, requires establishing protocols with reasonable depth of proteome coverage, while limiting overall analysis time. Preparation of urinary samples is not trivial because proteins are dilute and interfering substances are present[35-38]. Furthermore, to achieve good proteome coverage in urinary biomarker studies, additional fractionation steps upstream of LC MS/MS like ion exchange have so far been required. Here we set out to establish a robust workflow for analysis of the urinary proteome that employs single run LC MS/MS analysis and label-free quantification. We then applied this platform to determine the technical, daily and inter individual variations of the urinary proteome. Finally, we used our results to define and investigate the core urinary proteome in normal individuals.

Experimental Procedures

Urine collection and concentration

Second morning urine was collected from healthy donors for three consecutive days. The samples were immediately centrifuged at 5,000 g for 15 min at 4^{o} C to remove cell debris. The supernatant was stored at -20^{o}C until usage. For each donor approximately 50 ml of urine was concentrated by ultrafiltration. Urine was sequentially passed through a filtration membrane of 3

kDa cut-off at 2,500 g and at 4^{o} C. The final concentrate was washed twice with 50 mM Tris buffer, pH 8.0 to remove excess salt occasionally present in urine. Final protein concentration was estimated by the Bradford method.

Protein digestion and peptide preparation

The urinary proteome was digested in-solution in three ways: direct digestion, after acetone precipitation and after a combination of acetone precipitation and supplemental heat treatment. The direct in-solution digestion was performed as described in Shi et al[39]. Briefly, an initial reaction volume of 100 µl was obtained by adding 100 µg of protein to Tris buffer 50 mM, pH 8.0 containing 8 M final concentration of urea. The proteins were reduced with 10 mM DTT for 30 min and alkylated with 50 mM iodoacetamide. The alkylated proteins were digested with 2 µg of endoproteinase Lys-C for two hours. Following the initial digestion the reaction mixture was diluted to 2 M urea with 10mM Tris, pH 8.0. To the diluted sample an additional 2 µg of endoproteinase Lys-C or alternatively trypsin was added to the incubated at room temperature for overnight digestion. For acetone precipitation, the urine concentrates from ultra-filtration were mixed with four volumes of cold acetone and incubated at -20^{o} C for two hours followed by centrifugation at 13, 000 g for 10 min. The protein precipitate thus obtained was digested directly as above or heat treated before enzyme digestion. Heat treatment was carried out by re-suspending the protein precipitate in 20 µl of 10 mM Tris and heating at 95 ^{o}C for 5 min.

After overnight digestion, enzymatic activity was stopped by adding trifluoroacetic acid (TFA) to a final concentration of 0.3%. The resulting peptides were purified in a C_{18} StageTip as described[40].

LC MS/MS analysis

The StageTip purified peptides were eluted to a final volume of 4 µl. Liquid chromatography was performed with an Agilent 1100 system (Agilent Technologies, Palo Alto,

CA). Peptides were separated on a 15 cm fused silica column packed in-house with the reverse phase material ReproSil-Pur C_{18}–AQ, 3 μm resin (Dr. Maisch, Ammerbuch-Entringen, Germany). The reversed-phase separation was carried out by a 120 minute gradient from 2% to 80% of 80% (v/v) CH_3CN, 0.5% (v/v) acetic acid at flow rate of 250 nl/min unless otherwise noted.

The LC system was coupled to the LTQ-Orbitrap XL (Thermo Fischer Scientific) via a Proxeon nanoelectrospray ionization system (Proxeon Biosystems, now Thermo Fischer Scientific). The spray voltage was set to 2.1-2.35 kV and the heated capillary temperature was set at 175 ^{o}C. The mass spectrometer was operated in a data dependent fashion. Full scans were acquired in the Orbitrap analyzer with 60,000 resolution at 400 m/z. In order to achieve very high mass accuracy the lock mass option was enabled[41] and the real time internal calibration was performed using the 445.120025 ion. In data dependent scan events, up to the 5 most abundant ions were isolated (isolation width 2 Th), fragmented by collision induced dissociation (CID) in the linear ion trap with normalized collision energy of 35 and activation time of 30 ms. Fragment ions were subsequently ejected laterally for detection by the electron multipliers. For CID fragmentation, the wide band activation option was enabled and dynamic exclusion was used to minimize the extent of repeat sequencing of the peptides. Singly charged peptides were excluded from sequencing throughout the run.

Data Analysis

The MS data were analyzed using the MaxQuant software[42, 43], version 1.0.12.36. For MS/MS peak list file construction, up to top 8 peaks per 100 Da window were included for data base searching. The generated peak lists were searched against the decoy version of the international protein index (IPI) human database (version 3.68) with an initial precursor mass window of 7 ppm and a fragment mass window of 0.5 Th. Carbamidomethylation of cysteines was set as fixed modification and protein N-terminal acetylation and methionine oxidation were set as variable modifications for the database search. The cut-off false discovery rate for proteins and peptides were set to 0.01 and peptides with minimum 7 amino acids were considered for

identification. Additionally, we required at least two peptide identification per protein, of which at least one peptide had to be unique to the protein group.

Label free analysis was performed in MaxQuant[44] for protein quantitation. This included quantification of peptides recognized on the basis of mass and retention time but identified in other LC MS/MS runs ("match between the runs" option in MaxQuant). Only proteins that had a label free quantification intensity in at least 32 (50%) of the 63 samples were included for clustering analysis.

Statistical analysis

Pearson correlation and Spearman rank correlation were used for hierarchical clustering and rank ordering the quantitative results of the proteomic experiments respectively. (The Spearman rank correlation measures the degree of rank consensus between two ordered lists.) Percentages of variances were calculated from the median coefficient of variation (cv), which is the standard deviation divided by the mean of a measurement. For visualization of the cluster, the Perseus program in the MaxQuant environment and Genesis software[45] were used.

Results

An LC MS/MS method without pre-fractionation suitable for the urinary proteome

In our previous study, we had aimed at identifying as many proteins as possible in pooled urine[18]. Here we needed to analyze a relatively large number of different urinary proteomes. We therefore developed a single run LC MS/MS method that would enable identification of a reasonable number of urinary proteins. Mass spectrometric analysis was performed on a linear ion trap Orbitrap analyzer (LTQ-Orbitrap XL) at a precursor resolution of 60,000. Concurrently, peptides were fragmented by CID in the linear ion trap with a 'top5' method (Experimental Procedures). Data analysis was done in MaxQuant[42], resulting in sub-ppm average absolute mass

deviations and peptide identification rates of about 33%, somewhat lower than typically observed for cellular lysates under otherwise similar conditions.

Standard in-solution digestion conditions as normally employed in our laboratory at first resulted in the identification of only about 100 proteins in single runs (Figure 1A, Supplementary Table 1). As the same protocol routinely leads to the detection of more than 1,000 proteins in cell lysates, we suspected that digestion was inhibited by small molecular weight compounds in urine or by abundant protease inhibitors. Indeed, when using acetone for denaturing and precipitation, we detected close to 300 proteins. This number increased slightly when additionally heating the sample to 95 ^{o}C for five minutes (Figure 1A).

For greatest simplicity, these initial experiments were performed with endoprotease LysC digestion alone, which is the first digestion step in our digestion protocols. This is then followed by trypsin digestion under less denaturing conditions to produce smaller and more easily identified peptides. However, when we coupled acetone precipitation with LysC and subsequent trypsin digestion, the number of identifications increased substantially to 462+-21 urinary proteins in single runs (Figure 1A). The final workflow for analysis of urinary proteomes in this study is shown in Suppl. Figure 1.

The single LC MS/MS runs contain information on many more peptides and proteins than apparent from the above numbers. First our identification criteria were very conservative and they excluded single peptide identifications of proteins, even those with greater 99% confidence for protein identification. Additionally, many peptides that are clearly present are not 'picked for sequencing' by the mass spectrometer in single runs. In our experiments the total number of identifications increased to 624 in triplicate single runs of the same sample, even when keeping the requirement of two independent peptides for protein identification, because different peptides were fragmented and identified (Figure 1B). This number would then be the expected depth of proteome coverage of a single donor sample analyzed in triplicate in a biomarker study. Measurement time for the triplicate analysis was slightly more than six hours and total sample consumption was 6 µg (one urinary sample typically yields 1 mg of protein). In quadruple runs 654 proteins were identified (Figure 1B).

The very high mass accuracy achieved by the Orbitrap analyzer in conjunction with MaxQuant analysis facilitates accurate matching of peptide identities between LC MS/MS runs, even when the peptide was not sequenced in a given run. Using 'match between runs' of the quadruple analysis, we obtained an average of 633 protein identifications in each single run (Figure 1C).

The variability of the normal human proteome

To investigate intra and inter individual changes in normal urinary proteomes, we obtained second morning urine from a total of seven persons on three consecutive days. These five males and two females of different ethnicities were apparently healthy and their ages were between 25 to 35 years, a range where normal kidney function is expected.

We performed triplicate analysis of each of the 21 samples according to the protocol described above, resulting in a total of 63 LC MS/MS runs of 140 min each and a total measurement time of slightly less than one week. One additional 'random test' sample was obtained from one of the donors on a different day. The resulting 64 raw mass spectrometric files were analyzed together in MaxQuant using common criteria. The false positive rate for identifications as determined from a reverse database was less than 1%. This analysis resulted in the identification of 808 proteins from 4,871 sequence unique peptides. Each protein was identified by a mean of 6 peptides and a median of 4 peptides. Average absolute peptide mass deviation was 0.9 ppm for the entire data set. Raw MS data files as well as the MaxQuant peptide and protein evidence files are uploaded to the Tranche repository.

Rank order statistics is very robust and frequently applied in functional genomics. For each of the 64 measured proteomes, we ordered proteins by summed peptide intensity from most abundant to least abundant. To obtain a first general overview of the urinary proteome measurements we performed pair-wise rank order correlation between the protein intensities. All Spearman rank order coefficients were above 0.66, the intra-individual average was 0.92 and between technical replicates it was 0.97. For visualization, we color-coded the rank order coefficients. Figure 2A clearly shows seven blocks, corresponding to the seven individuals. These blocks are further divided into three sub-blocks, indicating the triplicate measurements of

samples obtained on three days (zoom in Figure 2A). The high rank coefficients imply that the data are sufficiently precise to capture differences in urinary proteomes within the samples of one person and between different persons.

Next, we directly quantified proteins levels by label-free methods between all 64 LC-MS/MS runs. This analysis makes use of much more of the raw data and should therefore be much more accurate than the rank order statistics described above. The total intensity values of all proteins in all samples were determined by MaxQuant, converted to z-scores and subjected to unsupervised 2-dimensional hierarchical clustering (Experimental Procedures). As shown in Figure 2B, all triplicate measurements of the same proteome sample clustered together, followed by clustering of each of the individual urinary proteomes. This demonstrates that in our data set technical variation was less than intra individual variation (urinary samples obtained on different days). Moreover, the urinary proteomes of different subjects were clearly distinguishable from each other by our proteomic workflow. To further confirm this finding, we tested where a 'random' proteome sample, obtained on a different day from one of the individuals, would align with respect to the other samples. Indeed, this single run clustered along with the other samples of the same donor (black arrow in Figure 2B).

We then determined the coefficients of variation (cv) for each of the three levels of sample variation. Within the 21 technical triplicate repeat measurements, the average cv was 0.18 (Figure 3A). This is an excellent value, considering that it was achieved by a label-free method on hundreds of proteins in a body fluid proteome. The variation of the proteome of each of the seven individuals over three days was 0.48. This figure includes the technical as well as the intra-individual variation. However, because it is significantly higher than the technical variability, the day to day variation of the urinary proteome must be the major contributor to this cv value. The inter individual cv values were 0.66, again significantly larger than either the technical or the intra-individual variations (Figure 3A). Variances (squares of standard deviation) are additive and because the mean is same at all three levels, allows estimating the quantitative contribution of each level. In our data, the contribution of technical variability to the total variability is only 7.45 % = $0.18^2/0.66^2$, the contribution of day to day variation is 45.45 % = $(0.48^2-0.18^2)/0.66^2$ and the contribution of the differences between the individuals is 47.1 % =

$(0.66^2 - 0.48^2)/0.66^2$. We found no further obvious relation between the individual proteomes; for example they did not cluster by gender.

On a protein by protein basis, total variability ranged from 0.24 to 3 (disregarding three outliers). The apparent variability of a protein did not strongly depend on its abundance as might have been expected (Figure 3B). This finding suggests that these variabilities capture physiological rather than technical aspects. Inspection of protein names did not point to obvious associations of cv values with specific molecular characteristics.

The core urinary proteome

Measurement of the normal urinary proteome of seven different individuals should allow not only the determination of the variability of the urinary proteome but also of its commonality. When we inspected the proteins in the 23 urinary proteomes, we found that a large subset was identified in all of them. These proteins represent the common and most easily identifiable proteins of urine and we term this subset the 'core urinary proteome'. Note that this is an operational definition because the core urinary proteome will expand as better instrumentation becomes available (see also below).

The total number of proteins identified with two peptides in urine was 808 as noted above and remarkably, a full 587 of them were found in each of the participant's urinary proteomes on each day (Figure 4A). Only a small number of proteins were found in all except one person on every day. Almost no proteins appeared to be unique to a single person - meaning it was only identified in one person on each day (28 of 808 proteins).

We and others have found that the added peptide intensity of each protein correlates well with the total protein amount[14, 46]. By this measure the absolute protein abundances of the core urinary proteome varied over almost 5 orders of magnitude, whereas apparent protein abundances in the total proteome measured here varied over six orders of magnitude. There are a few high abundance proteins whereas the majority clusters within a factor ten around an average value and a small fraction appears to have much lower abundance (Figure 4B). The most

abundant proteins are equally present in the core and entire protein set but some proteins of medium abundance are not found in the core proteome (Figure 4B).

In concordance with the above observations, a plot of the contribution of each protein to the cumulative proteome fraction revealed relatively few dominant proteins (Figure 5). Human serum albumin (HSA) is the most abundant protein in the core urinary proteome, accounting for 25% of its mass. The 20 most abundant proteins each contribute on average a few percent and together they make up about 2/3rd to the core proteome (Table 1). Almost all of them are secreted glycoproteins, with the exception of the plasma protein HSA. Ubiquitin was the other exception and the only intracellular protein (15th most abundant protein; 1.5%).

Zinc-alpha-2-glycoprotein (AZGP1) was the 4th most abundant protein but according to the UniProt database its existence has so far only been inferred by homology. Another interesting protein is epidermal growth factor (EGF) a small protein much studied in cell signaling and cancer, which appeared to be the 9th most abundant protein in urine. We have previously found EGF to be difficult to detect in an equimolar mixture of 50 proteins[47], which suggests that its levels in urine are even higher than estimated here. Urine is a known source of EGF, and growth factors are reported to be secreted by glomerular podocytes to support the integrity of the endothelium.

Table 1 also lists the individual cv values for the 20 most abundant proteins. They are between 0.2 to 1, roughly the same range as that of the core urinary proteome, showing that high protein abundance per se is not strongly correlated with a low degree of variability.

The 100 most abundant proteins account for 88% and the 200 most abundant for 94.7% of the core urinary proteome, which itself makes up 92% of the total urinary protein mass measured in our experiments. The lower abundant 50% of all core urinary proteins together only account for 2.6% of its total mass.

We also compared the absolute quantities of proteins in urine with those in a cell line proteome. In HeLa cells, more than 1500 proteins can readily be detected in triplicate 2h gradient runs of total cell lysate. Their quantitative distributions are less extreme, with fewer very high abundance proteins and a delayed falloff to very low apparent protein intensities. The observed

dynamic range is somewhat lower than in the total urinary proteome, presumably because there are more proteins of medium abundance (Supplementary Figure 2A, B).

Gene Ontology (GO) analysis[48] of the core urinary proteome showed a preponderance of extracellular and membrane compartments, in agreement with our previous findings[18]. Intracellular and especially nuclear compartments were underrepresented compared to the total proteome (Figure 6A). The overlap of the core urinary proteome with our previous data set was 87% (Figure 6B). The size distributions were very comparable, with roughly half of all proteins of molecular weights above the glomerular cut-off of roughly 40 kDa (Figure 4C). This also held true with considering the quantitative contribution of the different size classes of proteins found in urine. Note that our analysis does not preclude that only fragments of proteins were present. However, previous analyses employed 1D gel electrophoresis, where high molecular weight proteins were found at the expected locations[18 19].

Discussion and Outlook

Here we have described a proteomic workflow that allows the label-free quantification of more than 500 proteins in single LC MS/MS runs. Sample preparation only involving acetone precipitation and in-solution digest but no depletion of abundant proteins was found to be adequate for this depth of proteome coverage. The absence of pre-fractionation, in particular, means that relatively large numbers of proteome samples can be analyzed. This is in contrast to previous in-depth measurements, which required pooling of samples and extensive fractionation. We used this ability to measure the proteomes of a number of normal individuals at different time points allowing an assessment of both the variability and the commonality of the normal human urinary proteome.

By label-free quantification we found that differences between technical replicates contributed less than 8% to the total variability. This was substantially less than the contribution of either intra-individual (45%) or inter-individual variation (47%) to total variance. These values suggest that label free quantification may be of sufficient accuracy for the discovery of urinary biomarkers. They also suggest that analysis time may be best allocated to single measurements of proteomes obtained at different time points, rather than repeat measurements of

pooled samples from an individual. This conclusion agrees with current practice in microarray measurements, which are now considered sufficiently precise that usually only biological and not technical replicates are performed[49].

Label free quantification of the normal proteomes was also used to cluster the samples. Except two LC-MS/MS runs, the triplicate proteome measurements clustered together and then these triplicate measurements clustered together according to the donor. This further demonstrates that label-free quantification of urinary proteomes by high resolution MS/MS is robust and that it captures characteristics of individual body fluid proteomes.

Based on the diverse normal proteome measurements, we defined a 'core urinary' proteome containing all those proteins that were measured in each proteome sample of each person. This core proteome contained close to 600 proteins and accounted for 92% of total proteome mass compared to all quantified urinary proteins. The 20 most abundant proteins of the urinary proteome contribute about 2/3rd of the core proteome mass. Almost all of its members are secreted glycoproteins. In agreement with our and other high resolution measurements of the urinary proteome, we found a large proportion of membrane proteins, relatively high molecular weight proteins, and a corresponding underrepresentation of intracellular and especially nuclear proteins. The apparent dynamic range of the urinary core proteome spans almost five orders of magnitude, which is somewhat more than the apparent dynamic range of a cell line proteome measured under similar conditions. This is likely due to the fact that the quantitative distribution of urinary proteins is more extreme than that of cell lines.

Although we have been able to quantify a relatively large number in a short time, it would still be desirable to further extend the depth of analysis for potential urinary proteome biomarker studies. To investigate if this is possible, we implemented a number of improvements made in our general proteomics work-flow after this study had been completed. Indeed, when employing the Filter Aided Sample Preparation (FASP) method[50] in conjunction with longer LC gradients (4h instead of 2h), we increased the number of identifications in triplicate single runs to 811 and in quadruple runs to 836 proteins (Supplementary Table 3). We believe that quantification of about 1000 proteins with 2h gradients should be possible in the future, given ongoing improvements in instrumentation and computational methods. Thus it appears feasible to conduct medium scale urinary proteomics efforts on the basis of high resolution LC MS/MS

technology. For example, measuring the urinary proteome of 500 cases and 500 controls with the above technology would theoretically involve total instrument time of 100 days and correspondingly less if several mass spectrometers were used. This seems a reasonable instrument commitment compared to the resources needed for other parts of a biomarker study such as patient acquisition and downstream validation and assay development.

Data of the type acquired here could also readily be obtained for other body fluids. We suggest that this would be useful for several reasons. Definition of a core body fluid proteome and its individual variability on a protein by protein basis would help in deciding how distinctive a putative biomarker is likely to be. If the candidate is present and highly variable in the normal population then it would need to be highly upregulated in a specific condition to be a useful marker. Conversely, a putative biomarker may be promising for further study if is not a member of the core body fluid proteome but if it has a rank position in the case group clearly within the core body fluid proteome.

In summary, the technology of high resolution LC MS/MS employing single runs appears to be approaching sufficient maturity to contemplate biomarker studies on the urinary proteome. Measurement of the composition and variability of the normal urinary proteome will greatly help in interpreting the results of such studies.

Acknowledgements

We thank Alexandre Zougman and Jacek R Wisniewski for help and fruitful discussions and Juergen Cox greatly helped with statistical analysis. This work was partially supported by PROSPECTS, a 7th Framework grant by the European Directorate (grant agreement HEALTH-F4-2008-201648/PROSPECTS) and by the Max Planck Society for the advancement of Science.

References

1. Brunzel, N., *Fundamentals of Urine & Body Fluid Analysis*. 2004; p 480.

2. Pisitkun, T.; Johnstone, R.; Knepper, M. A., Discovery of urinary biomarkers. *Mol Cell Proteomics* **2006,** 5, (10), 1760-71.

3. Decramer, S.; de Peredo, A. G.; Breuil, B.; Mischak, H.; Monsarrat, B.; Bascands, J. L.; Schanstra, J. P., Urine in Clinical Proteomics. *Molecular & Cellular Proteomics* **2008,** 7, (10), 1850-1862.

4. Thongboonkerd, V., Recent progress in urinary proteomics. *Proteomics Clinical Applications* **2007,** 1, (8), 780-791.

5. Zerefos, P. G.; Vougas, K.; Dimitraki, P.; Kossida, S.; Petrolekas, A.; Stravodimos, K.; Giannopoulos, A.; Fountoulakis, M.; Vlahou, A., Characterization of the human urine proteome by preparative electrophoresis in combination with 2-DE. *Proteomics* **2006,** 6, (15), 4346-55.

6. Candiano, G.; Santucci, L.; Petretto, A.; Bruschi, M.; Dimuccio, V.; Urbani, A.; Bagnasco, S.; Ghiggeri, G. M., 2D-electrophoresis and the urine proteome map: Where do we stand? *Journal of Proteomics* **2010,** 73, (5), 829-844.

7. Voshol, H.; Brendlen, N.; Muller, D.; Inverardi, B.; Augustin, A.; Pally, C.; Wieczorek, G.; Morris, R. E.; Raulf, F.; van Oostrum, J., Evaluation of biomarker discovery approaches to detect protein biomarkers of acute renal allograft rejection. *J Proteome Res* **2005,** 4, (4), 1192-9.

8. Ward, D. G.; Nyangoma, S.; Joy, H.; Hamilton, E.; Wei, W.; Tselepis, C.; Steven, N.; Wakelam, M. J.; Johnson, P. J.; Ismail, T.; Martin, A., Proteomic profiling of urine for the detection of colon cancer. *Proteome Sci* **2008,** 6, 19.

9. Zurbig, P.; Renfrow, M. B.; Schiffer, E.; Novak, J.; Walden, M.; Wittke, S.; Just, I.; Pelzing, M.; Neususs, C.; Theodorescu, D.; Root, K. E.; Ross, M. M.; Mischak, H., Biomarker discovery by CE-MS enables sequence analysis via MS/MS with platform-independent separation. *Electrophoresis* **2006,** 27, (11), 2111-25.

10. Theodorescu, D.; Mischak, H., Mass spectrometry based proteomics in urine biomarker discovery. *World J Urol* **2007,** 25, (5), 435-43.

11. Coon, J. J.; Zurbig, P.; Dakna, M.; Dominicza, A. F.; Decramer, S.; Fliser, D.; Frommberger, M.; Golovko, I.; Good, D. M.; Herget-Rosenthal, S.; Jankowski, J.; Julian, B. A.; Kellmann, M.; Kolch, W.; Massy, Z.; Novak, J.; Rossing, K.; Schanstra, J. P.; Schiffer, E.; Theodorescu, D.; Vanholder, R.; Weissinger, E. M.; Mischak, H.; Schmitt-Kopplin, P., CE-MS analysis of the human urinary proteome for biomarker discovery and disease diagnostics. *Proteomics Clinical Applications* **2008,** 2, (7-8), 964-973.

12. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** 422, (6928), 198-207.

13. Cravatt, B. F.; Simon, G. M.; Yates, J. R., 3rd, The biological impact of mass-spectrometry-based proteomics. *Nature* **2007,** 450, (7172), 991-1000.

14. de Godoy, L. M.; Olsen, J. V.; Cox, J.; Nielsen, M. L.; Hubner, N. C.; Frohlich, F.; Walther, T. C.; Mann, M., Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **2008,** 455, (7217), 1251-4.

15. Choudhary, C.; Mann, M., Decoding signalling networks by mass spectrometry-based proteomics. *Nat Rev Mol Cell Biol* **2010,** 11, (6), 427-39.

16. Chen, Y. T.; Tsao, C. Y.; Li, J. M.; Tsai, C. Y.; Chiu, S. F.; Tseng, T. L., Large-scale protein identification of human urine proteome by multi-dimensional LC and MS/MS. *Proteomics Clinical Applications* **2007,** 1, (6), 577-587.

17. Castagna, A.; Cecconi, D.; Sennels, L.; Rappsilber, J.; Guerrier, L.; Fortis, F.; Boschetti, E.; Lomas, L.; Righetti, P. G., Exploring the hidden human urinary proteome via ligand library beads. *J Proteome Res* **2005,** 4, (6), 1917-30.

18. Adachi, J.; Kumar, C.; Zhang, Y.; Olsen, J. V.; Mann, M., The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome Biol* **2006,** 7, (9), R80.

19. Kentsis, A.; Monigatti, F.; Dorff, K.; Campagne, F.; Bachur, R.; Steen, H., Urine proteomics for profiling of human disease using high accuracy mass spectrometry. *Proteomics Clinical Applications* **2009,** 3, (9), 1052-1061.

20. Pisitkun, T.; Shen, R. F.; Knepper, M. A., Identification and proteomic profiling of exosomes in human urine. *Proc Natl Acad Sci U S A* **2004,** 101, (36), 13368-73.

21. Gonzales, P. A.; Pisitkun, T.; Hoffert, J. D.; Tchapyjnikov, D.; Star, R. A.; Kleta, R.; Wang, N. S.; Knepper, M. A., Large-scale proteomics and phosphoproteomics of urinary exosomes. *J Am Soc Nephrol* **2009,** 20, (2), 363-79.

22. Theodorescu, D.; Fliser, D.; Wittke, S.; Mischak, H.; Krebs, R.; Walden, M.; Ross, M.; Eltze, E.; Bettendorf, O.; Wulfing, C.; Semjonow, A., Pilot study of capillary electrophoresis coupled to mass spectrometry as a tool to define potential prostate cancer biomarkers in urine. *Electrophoresis* **2005,** 26, (14), 2797-808.

23. Haubitz, M.; Wittke, S.; Weissinger, E. M.; Walden, M.; Rupprecht, H. D.; Floege, J.; Haller, H.; Mischak, H., Urine protein patterns can serve as diagnostic tools in patients with IgA nephropathy. *Kidney Int* **2005,** 67, (6), 2313-20.

24. Theodorescu, D.; Wittke, S.; Ross, M. M.; Walden, M.; Conaway, M.; Just, I.; Mischak, H.; Frierson, H. F., Discovery and validation of new protein biomarkers for urothelial cancer: a prospective analysis. *Lancet Oncol* **2006,** 7, (3), 230-40.

25. Kreunin, P.; Zhao, J.; Rosser, C.; Urquidi, V.; Lubman, D. M.; Goodison, S., Bladder cancer associated glycoprotein signatures revealed by urinary proteomic profiling. *Journal of Proteome Research* **2007,** 6, (7), 2631-2639.

26. Quintana, L. F.; Campistol, J. M.; Alcolea, M. P.; Banon-Maneus, E.; Sole-Gonzalez, A.; Cutillas, P. R., Application of Label-free Quantitative Peptidomics for the Identification of Urinary Biomarkers of Kidney Chronic Allograft Dysfunction. *Molecular & Cellular Proteomics* **2009,** 8, (7), 1658-1673.

27. Schlatzer, D. M.; Dazard, J. E.; Dharsee, M.; Ewing, R. M.; Ilchenko, S.; Stewart, I.; Christ, G.; Chance, M. R., Urinary Protein Profiles in a Rat Model for Diabetic Complications. *Molecular & Cellular Proteomics* **2009,** 8, (9), 2145-2158.

28. Ling, X. B.; Sigdel, T. K.; Lau, K.; Ying, L. H.; Lau, I.; Schilling, J.; Sarwal, M. M., Integrative Urinary Peptidomics in Renal Transplantation Identifies Biomarkers for Acute Rejection. *Journal of the American Society of Nephrology* **2010,** 21, (4), 646-653.

29. Kentsis, A.; Lin, Y. Y.; Kurek, K.; Calicchio, M.; Wang, Y. Y.; Monigatti, F.; Campagne, F.; Lee, R.; Horwitz, B.; Steen, H.; Bachur, R., Discovery and Validation of Urine Markers of Acute Pediatric Appendicitis Using High-Accuracy Mass Spectrometry. *Annals of Emergency Medicine* **2010,** 55, (1), 62-70.

30. Sigdel, T. K.; Kaushal, A.; Gritsenko, M.; Norbeck, A. D.; Qian, W. J.; Xiao, W. Z.; Camp, D. G.; Smith, R. D.; Sarwal, M. M., Shotgun proteomics identifies proteins specific for acute renal transplant rejection. *Proteomics Clinical Applications* **2010,** 4, (1), 32-47.

31. Omenn, G. S.; States, D. J.; Adamski, M.; Blackwell, T. W.; Menon, R.; Hermjakob, H.; Apweiler, R.; Haab, B. B.; Simpson, R. J.; Eddes, J. S.; Kapp, E. A.; Moritz, R. L.; Chan, D. W.; Rai, A. J.; Admon, A.; Aebersold, R.; Eng, J.; Hancock, W. S.; Hefta, S. A.; Meyer, H.; Paik, Y. K.; Yoo, J. S.; Ping, P.; Pounds, J.; Adkins, J.; Qian, X.; Wang, R.; Wasinger, V.; Wu, C. Y.; Zhao, X.; Zeng, R.; Archakov, A.; Tsugita, A.; Beer, I.; Pandey, A.; Pisano, M.; Andrews, P.; Tammen, H.; Speicher, D. W.; Hanash, S. M., Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* **2005,** 5, (13), 3226-45.

32. States, D. J.; Omenn, G. S.; Blackwell, T. W.; Fermin, D.; Eng, J.; Speicher, D. W.; Hanash, S. M., Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nat Biotechnol* **2006,** 24, (3), 333-8.

33. Lee, R. S.; Monigatti, F.; Lutchman, M.; Patterson, T.; Budnik, B.; Steen, J. A. J.; Freeman, M. R.; Steen, H., Temporal variations of the postnatal rat urinary proteome as a reflection of systemic maturation. *Proteomics* **2008,** 8, (5), 1097-1112.

34. Sun, W.; Chen, Y.; Li, F. X.; Zhang, L.; Yang, R. F.; Zhang, Z.; Zheng, D. X.; Gao, Y. H., Dynamic urinary proteomic analysis reveals stable proteins to be potential biomarkers. *Proteomics Clinical Applications* **2009,** 3, (3), 370-382.

35. Thongboonkerd, V., Practical points in urinary proteomics. *Journal of Proteome Research* **2007,** 6, (10), 3881-3890.

36. Lee, R. S.; Monigatti, F.; Briscoe, A. C.; Waldon, Z.; Freernan, M. R.; Steen, H., Optimizing sample handling for urinary proteomics. *Journal of Proteome Research* **2008,** 7, (9), 4022-4030.

37. Thomas, C. E.; Sexton, W.; Benson, K.; Sutphen, R.; Koomen, J., Urine Collection and Processing for Protein Biomarker Discovery and Quantification. *Cancer Epidemiology Biomarkers & Prevention* **2010,** 19, (4), 953-959.

38. Loftheim, H.; Nguyen, T. D.; Malerod, H.; Lundanes, E.; Asberg, A.; Reubsaet, L., 2-D hydrophilic interaction liquid chromatography-RP separation in urinary proteomics - Minimizing variability through improved downstream workflow compatibility. *Journal of Separation Science* **2010,** 33, (6-7), 864-872.

39. Shi, R.; Kumar, C.; Zougman, A.; Zhang, Y.; Podtelejnikov, A.; Cox, J.; Wisniewski, J. R.; Mann, M., Analysis of the mouse liver proteome using advanced mass spectrometry. *J Proteome Res* **2007,** 6, (8), 2963-72.

40. Rappsilber, J.; Ishihama, Y.; Mann, M., Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **2003,** 75, (3), 663-70.

41. Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M., Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* **2005,** 4, (12), 2010-21.

42. Cox, J.; Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **2008,** 26, (12), 1367-72.

43. Cox, J.; Matic, I.; Hilger, M.; Nagaraj, N.; Selbach, M.; Olsen, J. V.; Mann, M., A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat Protoc* **2009,** 4, (5), 698-705.

44. Luber, C. A.; Cox, J.; Lauterbach, H.; Fancke, B.; Selbach, M.; Tschopp, J.; Akira, S.; Wiegand, M.; Hochrein, H.; O'Keeffe, M.; Mann, M., Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. *Immunity* **2010,** 32, (2), 279-89.

45. Sturn, A.; Quackenbush, J.; Trajanoski, Z., Genesis: cluster analysis of microarray data. *Bioinformatics* **2002,** 18, (1), 207-8.

46. Malmstrom, J.; Beck, M.; Schmidt, A.; Lange, V.; Deutsch, E. W.; Aebersold, R., Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature* **2009,** 460, (7256), 762-5.

47. Geiger, T.; Cox, J.; Mann, M., Proteomics on an Orbitrap benchtop mass spectrometer using all ion fragmentation. *Mol Cell Proteomics* **2010**.

48. Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock,

G., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **2000,** 25, (1), 25-9.

49. Allison, D. B.; Cui, X.; Page, G. P.; Sabripour, M., Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet* **2006,** 7, (1), 55-65.

50. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat Methods* **2009,** 6, (5), 359-62.

Abundance Rank	Protein Name	Gene name	CV	Protein amount %	Cumulative amount %
1	Serum albumin	ALB	0.41	25.45	25.5
2	Kininogen-1	KNG1	0.38	4.33	29
3	Prosaposin	PSAP	0.62	2.99	33
4	Zinc-alpha-2-glycoprotein	AZGP1	1.01	2.77	35.5
5	Apolipoprotein D	APOD	0.31	2.73	38.3
6	Ig kappa chainC region	IGKC	0.46	2.73	41.0
7	Alpha-1-microglobulin	AMBP	0.63	2.67	43.7
8	Alpha-1-antitrypsin	SERPINA1	0.30	2.34	46.03
9	Pro-epidermal growth factor	EGF	0.47	2.07	48.11
10	Prostaglandin D2 synthase 21kDa (Brain)	PTGDS	0.45	1.92	50
11	Osteopontin	SPP1	0.2	1.80	51.8

			8		
12	Serotransferrin	TF	0.69	1.7	53.5
13	Pancreatic alpha-amylase	AMY2A	0.91	1.7	55.24
14	Inter-alpha-trypsin inhibitor heavy chain H	ITIH4	0.28	1.6	56.9
15	Ubiquitin C	UBC	0.22	1.4	58.3
16	Uromodulin	UMOD	0.34	1.3	59.6
17	CD59 glycoprotein	CD59	0.48	1.27	60.8
18	Plasma protease C1 inhibitor	SERPING1	0.26	1.2	62.02
19	Polymeric immunoglobulin receptor	PIGR	0.38	1.1	63.1
20	cDNA FLJ55606, highly similar to Alpha-2-HS-glycoprotein	AHSG	0.67	0.96	64.1

Figure legends`

Figure 1. Proteome coverage with different sample preparation and replicate runs. **A.** The number of proteins identified by the different sample preparation methods are shown. Error bars as determined by quadruple runs. **B.** The cumulative increase in the number of proteins identified from two, three and four replicate runs of the same sample by acetone precipitation and trypsin digestion. **C.** Number of proteins identified in each of the replicate runs with (green) and without (brown) matching between the runs. Matching between the runs uses the high mass accuracy to transfer identifications to previously unidentified peptides from different LC-MS/MS runs where they were fragmented and identified.

Figure 2. Rank order correlation and hierarchical clustering. A Matrix of Spearman rank correlation coefficients for each LC-MS/MS run against every other run. The matrix is color coded as shown in the figure. The left inset shows the zoom of one particular individual indicating the high rank correlation for the technical replicates within the same proteome sample. **B**. Hierarchical clustering of 64 LC-MS/MS runs shows that the technical triplicates are much closer than the intra-individual and inter-individual samples. A random sample clustered together with the correct donor, as shown by the arrow.

Figure 3. Distribution of coefficient of variation. A. The distribution of co-efficient for technical, intra-individual and inter-individual variation is shown in the box plot. The cv for technical variation was much lower than the intra and inter individual variation. **B.** A plot of co-efficient of variation against the logarithmic summed peptide intensity of the proteins show that there is no strong correlation between the abundance level of the protein and the extent of variation.

Figure 4. The core proteome. A. The core proteome was defined as those proteins that were detected on all three days in all seven individuals. The core proteome constitutes the majority of the total proteome data set (64%). **B**. The dynamic range of the core proteome (orange) in comparison to the total proteome (red). The logarithmic summed peptide intensity is used as a measure of the abundance of the proteins. The core proteome has a dynamic range of about 5 orders of magnitude.

Figure 5. Cumulative abundance of the core proteins. Contribution of each of the proteins to total core urinary proteome mass. Human serum albumin, kininogen, epidermal growth factor and uromodulin are a few of the top 20 most abundant proteins that contribute to more than 60% of the proteome (see also Table 1).

Figure 6. Gene ontology analysis of core proteome. A. GO annotations for the proteins in the core proteome agree with the previous observations that the urinary proteome contains a large proportion of membrane proteins. Nuclear and cytoplasmic proteins are under-represented in the urinary proteome. The GO annotation distribution is similar for the core proteome, the total protein set and the large-scale urinary proteome dataset (Adachi et al). **B**. The core proteome is almost entirely contained within the Adachi et al data set. **C**. The size distribution of the core proteome agrees with previously published datasets and indicates many proteins above the 40 kDa glomerular filtration cut-off.

Supplementary Figure 1. Sample preparation and digestion for urinary proteomics by LC-MS/MS. The urine protein sample is concentrated by ultra-filtration and processed in the three different ways. The peptides obtained from overnight digestion are purified by C_{18} Stage Tips and loaded injected for LC-MS/MS analysis.

Supplementary Figure 2. Comparison of dynamic range of urine and cell line proteome. A. The logarithmic summed peptide intensity of urine and HeLa cell line proteome is shown. The proteome of HeLa was obtained by three technical replicates of whole cell lysate peptides in similar 2 hour gradients in a LTQ-Orbitrap instrument. B. The cumulative proteome fraction of urine contributed by the most abundant proteins is much higher than that of the HeLa proteome.

Supplementary materials

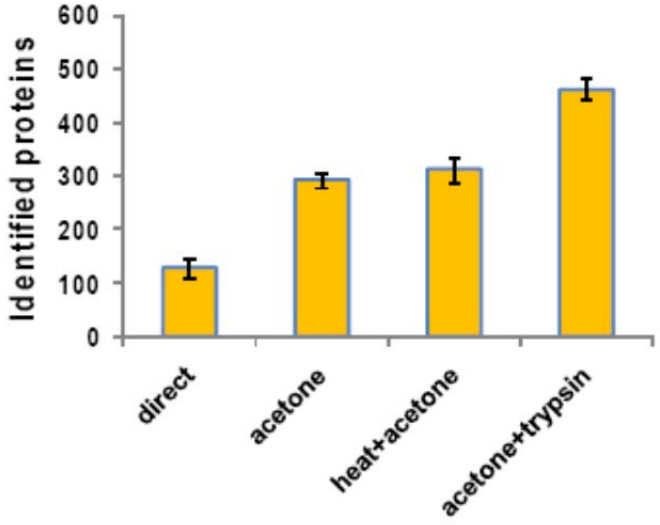
Supplementary Table 1. List of proteins and peptides identified by different sample preparation methods and enzyme digestion.

Supplementary Table 2. List of proteins in the core proteome and their cumulative proteome fraction

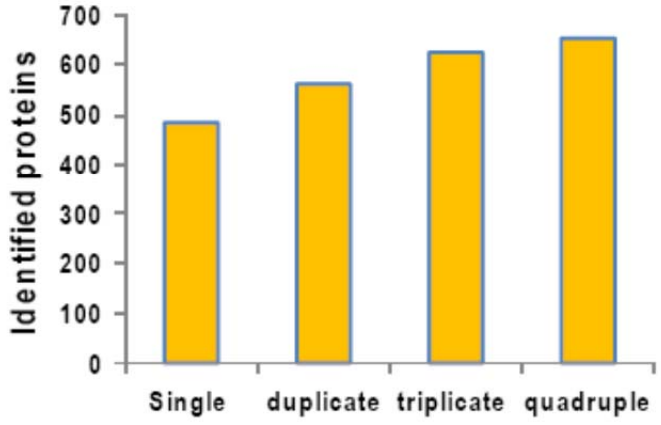
Supplementary Table 3. List of proteins identified with FASP and long gradient methods.

Figure 1

A


Identified proteins
600
500
400
300
200
100
0
direct
acetone
heat+acetone
acetone+trypsin

B


Identified proteins
700
600
500
400
300
200
100
0
Single
duplicate
triplicate
quadruple

C

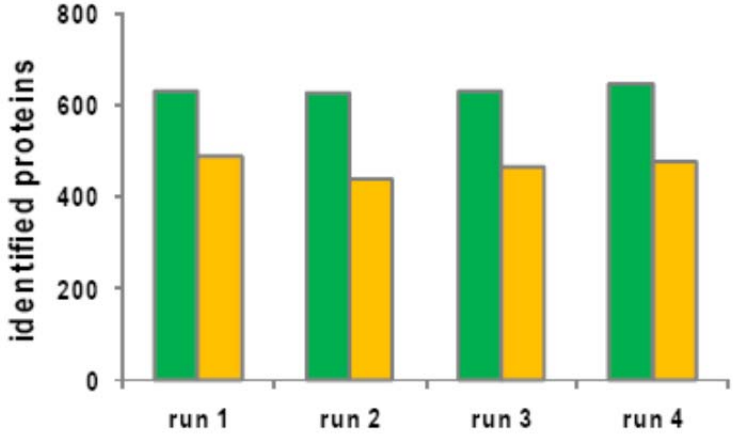
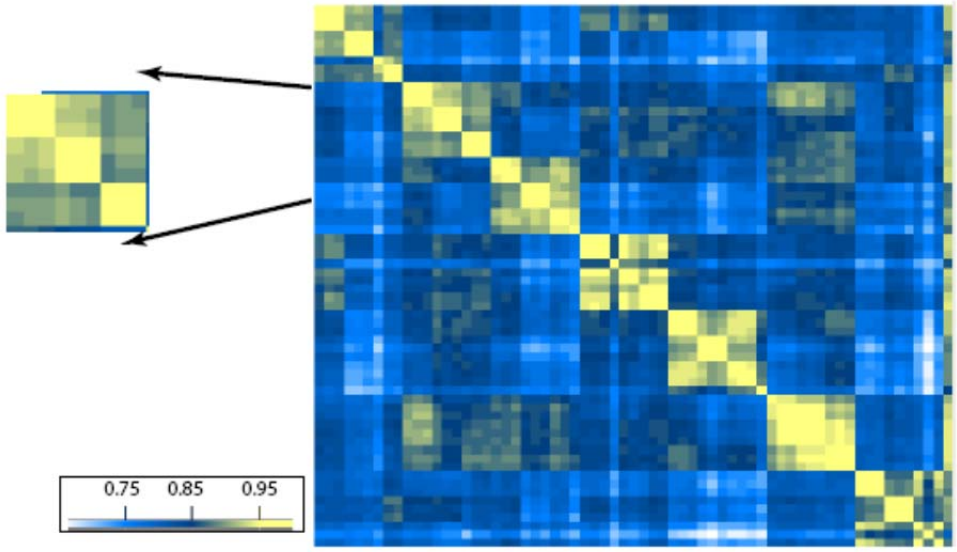

identified proteins
800
600
400
200
0
run 1
run 2
run 3
run 4

Figure 2

A



B

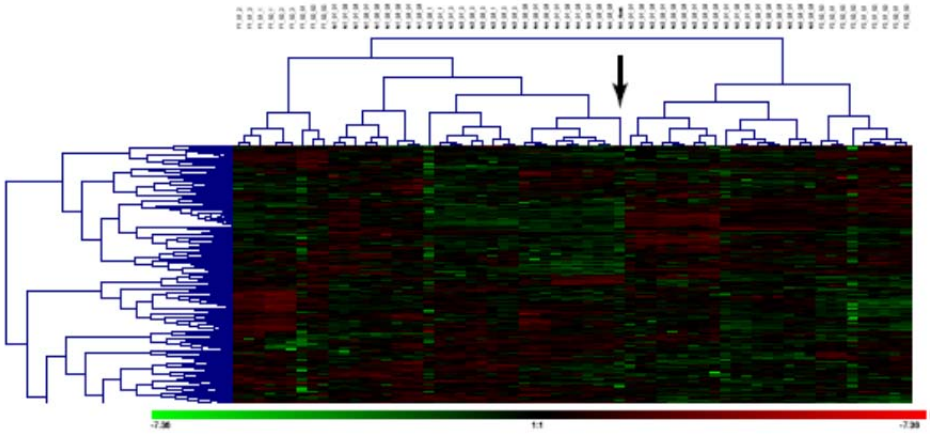
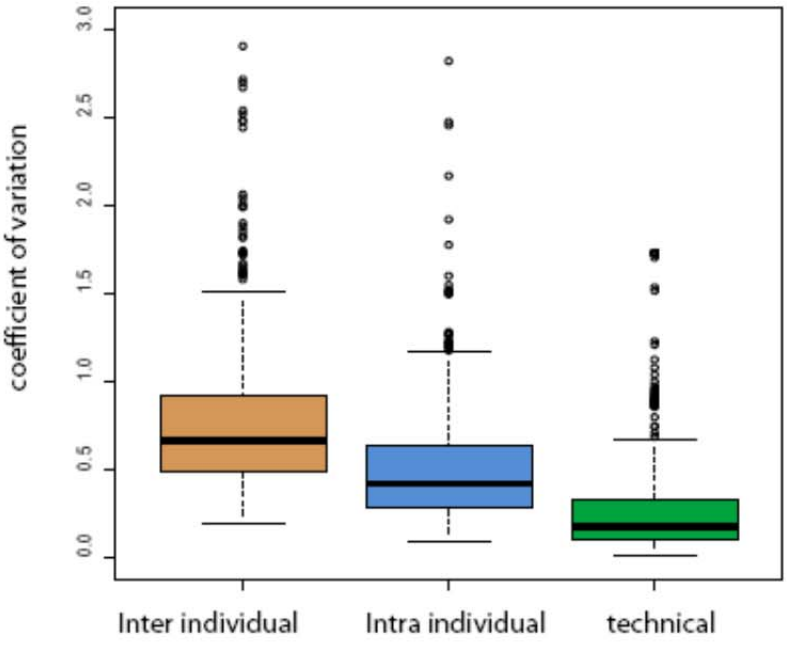


Figure 3

A


coefficient of variation
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Inter individual
Intra individual
technical

B

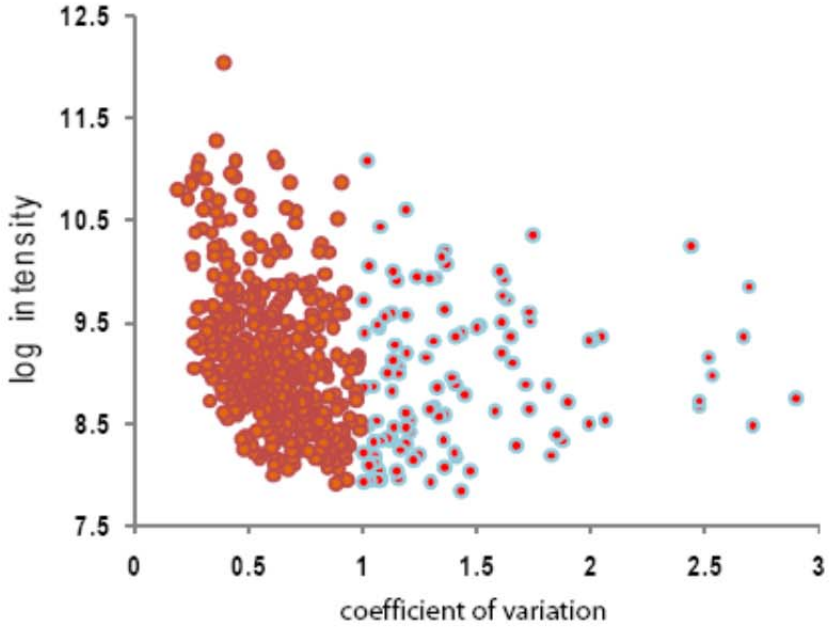
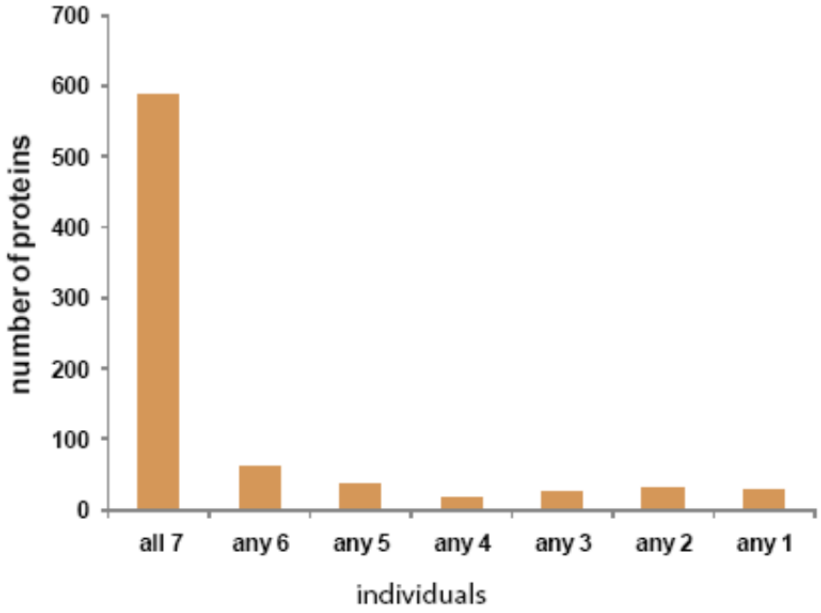

log intensity
12.5
11.5
10.5
9.5
8.5
7.5
0
0.5
1
1.5
2
2.5
3
coefficient of variation

Figure 4

A


700
600
500
400
300
200
100
0
number of proteins
all 7
any 6
any 5
any 4
any 3
any 2
any 1
individuals

B

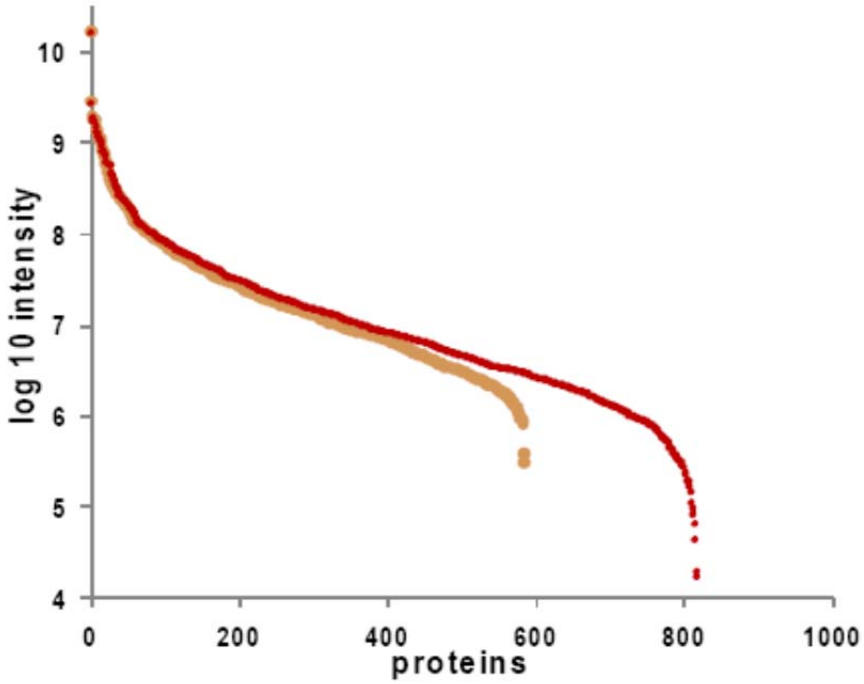

10
9
8
7
6
5
4
log 10 intensity
0
200
400
600
800
1000
proteins

Figure 5

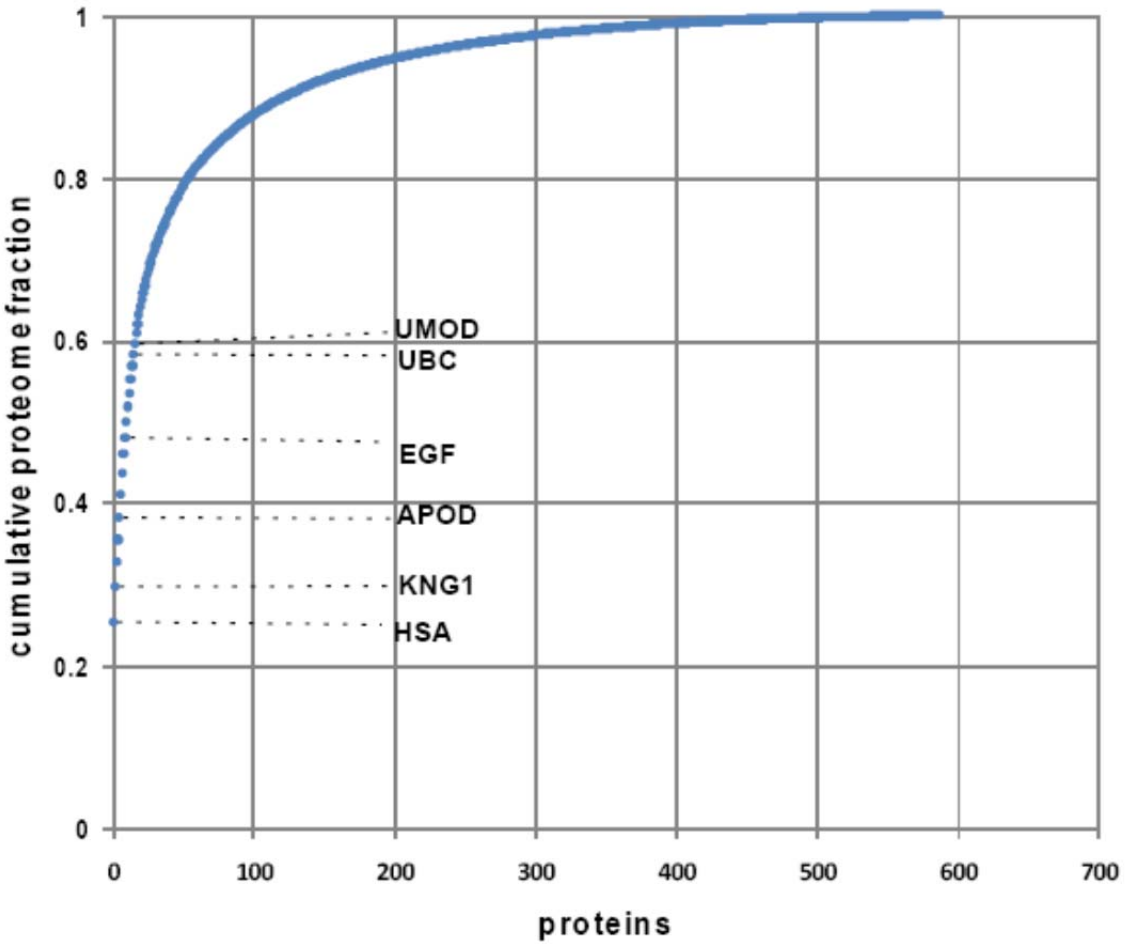
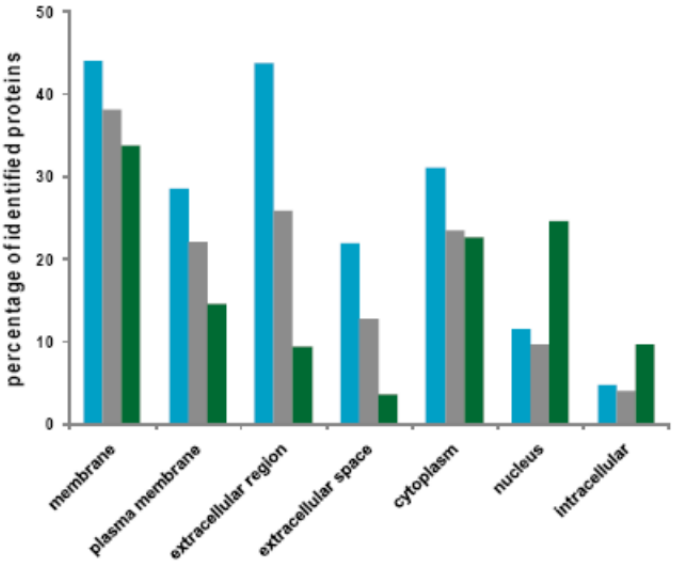
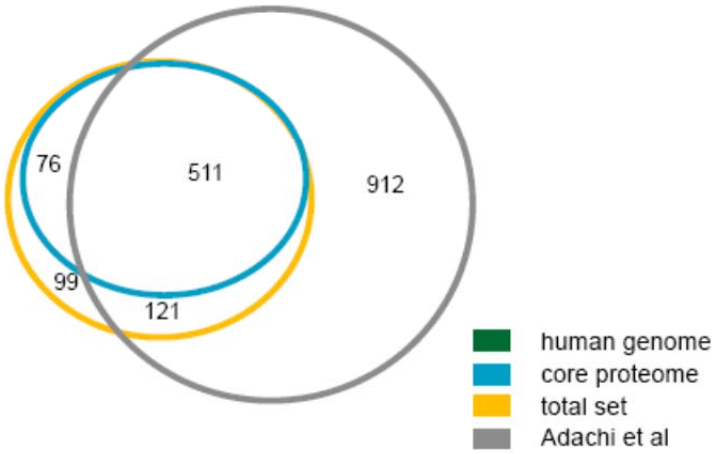

1
0.8
0.6
0.4
0.2
0
cumulative proteome fraction
UMOD
UBC
EGF
APOD
KNG1
HSA
0
100
200
300
400
500
600
700
proteins

Figure 6

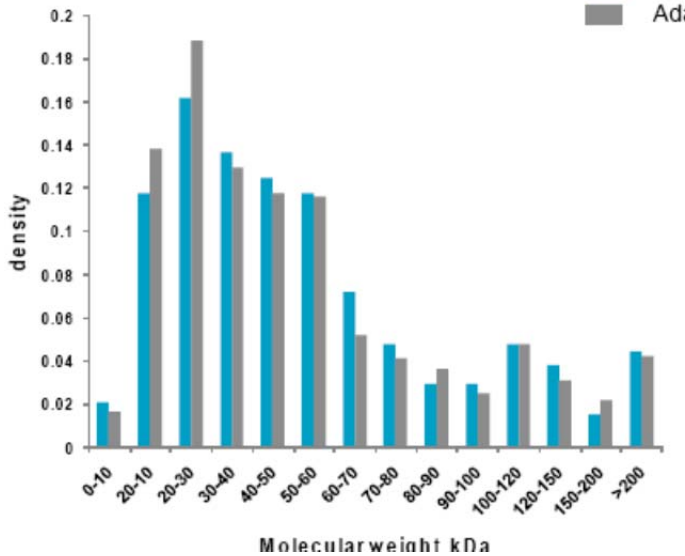
A


percentage of identified proteins
50
40
30
20
10
0
membrane
plasma membrane
extracellular region
extracellular space
cytoplasm
nucleus
intracellular

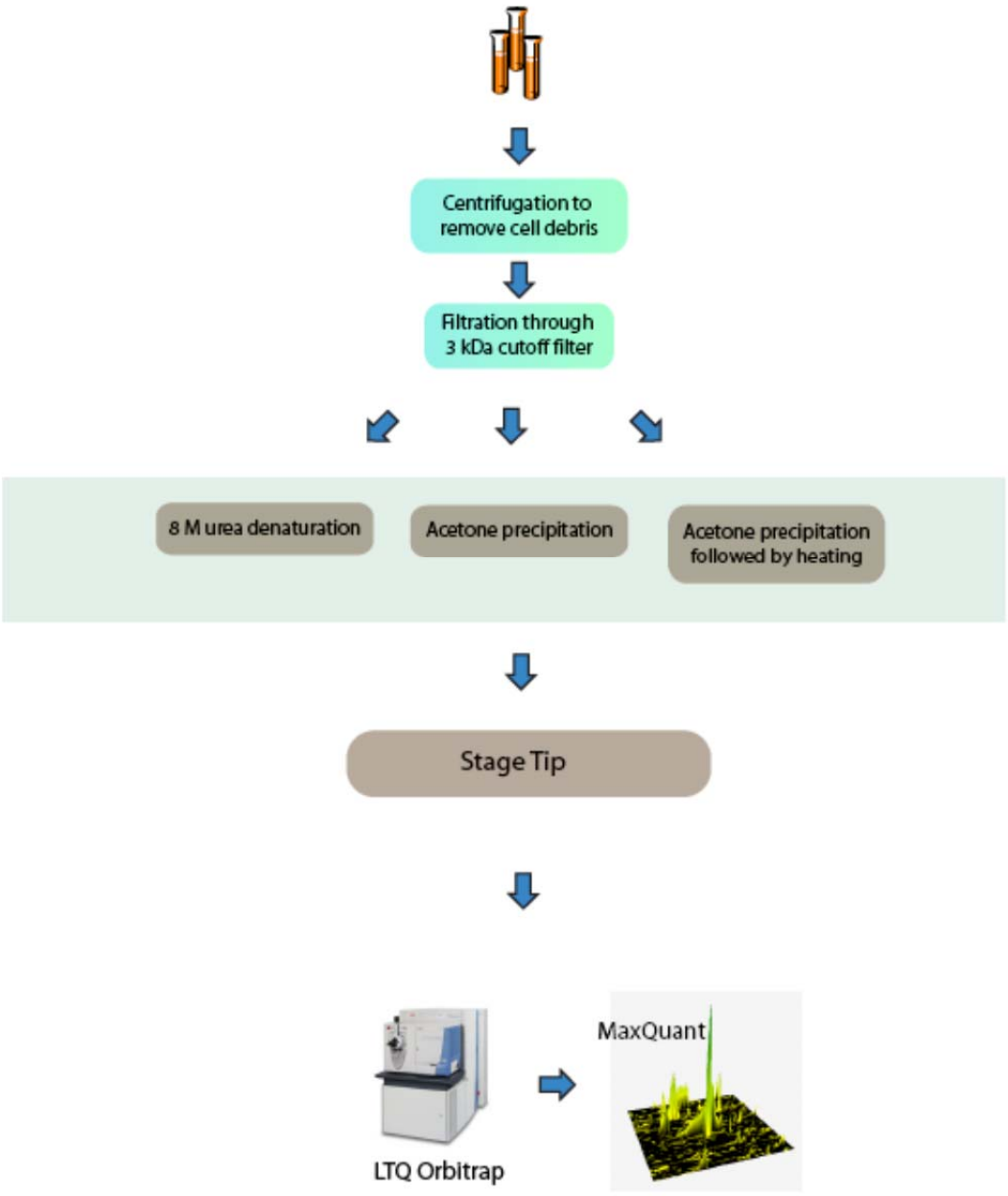
B


76
511
912
99
121
human genome
core proteome
total set
Adachi et al

C

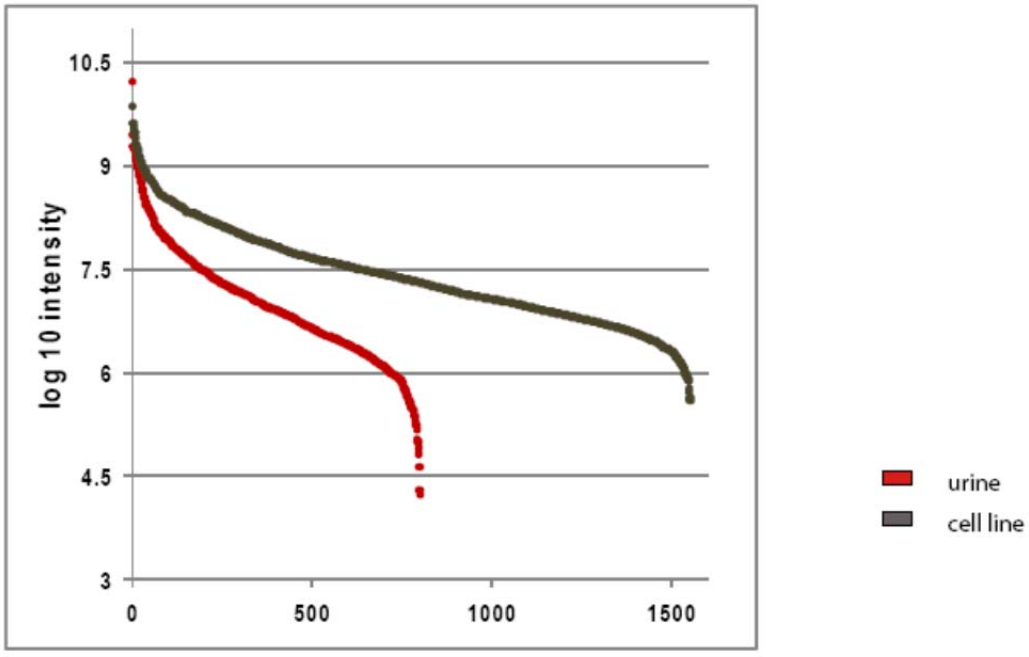

density
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0-10
20-10
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100
100-120
120-150
150-200
>200
Molecular weight kDa

Supplementary figure 1

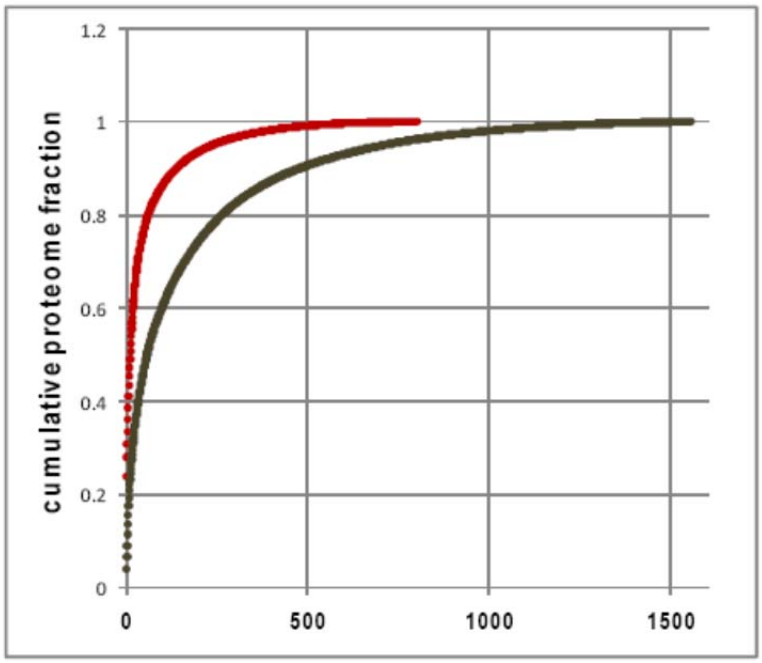


Supplementary figure 2

A



B



Conclusion and outlook

The interdisciplinary nature of scientific research is crucial for the advancement of mankind and application of mass spectrometry and fluorescence spectroscopy to biological sciences are among the best examples for this happening. Without such techniques, identification of protein inside a cell, localization of a protein etc would be a too difficult problem to solve requiring a large consortium of scientist. Mass spectrometry based proteomics in its current state has finally enabled the investigation of proteomic phenotype of a cell/tissue in a "systems-wide' perspective.

During my PhD thesis, I have been very lucky to be involved in different themes of MS based proteomics as a main driver or as a collaborator. Sample preparation techniques and separation techniques are ever evolving and the most favorite technique/method that I would love to see is a development that can tackle the dynamic range problem in plasma and other samples. This could cause a dramatic change in the field even in terms of analysis time, cost etc. The continuous development of mass spectrometric technology is fascinating and the recent development of bench top high resolution mass spectrometers clearly suggests that mass spectrometry will be a routine tool that every biologist will be able to use in general laboratories.

With the parallel improvements in technology and sample preparation, I believe that real clinical proteomic is very near. Increasing number of reports using LC-MS/MS based proteomics applied to clinical samples indicate the rapid development in the field. Recent reports suggest preservation of proteome and its PTMs in the frozen samples and FFPE samples will open a window of opportunity to investigate the proteome and phosphoproteome of tumor samples that may prove vital in the field of medical oncology.

Though spending four years in one of the best proteomics laboratories in the world, there are several subjects like peptide pull down and immuno-precipitation coupled to mass spectrometry in which I lack practical experience. In the near future I look forward to make further progress in understanding mass spectrometry and related principles and applying them to clinical and biological questions that are plentiful in our environment.

5. Acknowledgments

I thank my parents and sister for their immense support and constant prayers for my well being.

I am grateful to Matthias Mann for (1) giving me an excellent opportunity, (2) his unbelievable patience and I sincerely thank him for providing such a great atmosphere.

I want to thank Jacek Wisniewski and Alexander Zougman for their support especially during my ‘childhood’ in the Mann lab.

I want to thank Hans-Joerg Schaeffer and Maximilliane Reif from IMPRS for taking care of me during the early days in Germany taking care of most of the administrative affairs.

I should thank Peter Bandilla for his immense help in several mass spectrometry related questions and mathematical and engineering drawing questions. I thank him for his knowledge sharing and ever helping nature. Johannes Graumann, for his discussions and feedback on many figures, biology, etc, etc, etc. I express my thanks to Jürgen Cox for his excellent software and answers with the statistical questions.

Dr. K. Sankaran, Dr. Meenakshi Sundaram and Late. Dr.V.Murugan from Anna University without whose support I would not have gotten the opportunity to come to such a wonderful working atmosphere. I want to thank Dr. Helena Neveleinen and Stephen C Brown at the Macquarie University, Sydney.

Theresa Schneider, Alison Dalfavo and Tine Klitmoeller were helpful and would like to thank them for their administrative support.

A special thanks to Aarathi Balijepalli for her support.

Krishna, Kirti, Ranga, Rochelle, Anoope, Venky, Juhi, Jyoti, Satish, Pratiba, and the rest of the ‘Indian mafia’ that took care of fun, food and ‘drinking’.

I want to thank Mara, Vanessa, and Frank for the good working environment. Especially thanks to Mara who made work as thrilling and as intense as a soccer game.

Ivan Matic, Mario Oroshi, Cuiping Pan and the lunch troop for putting up with my stupid/smart jokes and questions.

6. Curriculum vitae

Nagarjuna Nagaraj

Personal Information

Date of birth:	29 December 1982
Place of birth:	Thanjavur
Familial status:	Unmarried
Gender:	Male
Nationality:	India

Academic Qualification

2006 – Till date	PhD student in the Mann Lab, Max Planck Institute for Biochemistry, Martinsried, Germany
2005-2006	Master of Biotechnology, Macquarie University, Sydney, Australia
2000-2004	Bachelor of Technology, Allagappa College of Technology, Anna University, Chennai, India

Research Experience

Proteomics	The main projects involved improving the membrane protein coverage in LC-MS/MS set up, applicability of HCD for large-scale phosphoproteomics in LTQ-Orbitrap Velos instrument and application of LC-MS/MS based platform for developments in clinical proteomics
Molecular biology	Expression of proteins with C and N terminal tags for assessing the solubility of proteins as a part of large –scale structural genomics pipeline
Biochemistry	Purification of IgG molecules from goat blood for constructing a mini bio-sensor.

List of publications and manuscripts

1. Wisniewski, J. R.; Nagaraj, N.; Zougman, A.; Gnad, F.; Mann, M., Brain Phosphoproteome Obtained by a FASP-Based Method Reveals Plasma Membrane Protein Topology. *J Proteome Res*.
2. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat Methods* **2009,** 6, (5), 359-62.
3. Cox, J.; Matic, I.; Hilger, M.; Nagaraj, N.; Selbach, M.; Olsen, J. V.; Mann, M., A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat Protoc* **2009,** 4, (5), 698-705.
4. Nagaraj, N.; Lu, A.; Mann, M.; Wisniewski, J. R., Detergent-based but gel-free method allows identification of several hundred membrane proteins in single LC-MS runs. *J Proteome Res* **2008,** 7, (11), 5028-32.

Manuscripts under preparation

1. Nagarjuna Nagaraj[1, 3], Rochelle C. J. D'Souza[1, 3], Juergen Cox[1], Jesper. V. Olsen [2] and Matthias Mann[1]. Large scale phosphoproteomics with HCD fragmentation. *J Proteome Res* in revision.
2. Nagarjuna Nagaraj and Matthias Mann. Quantitative analysis of the intra and inter individual variability of the normal urinary proteome. *J Proteome Res* submitted.
3. Mara Monetti, Nagarjuna Nagaraj and Matthias Mann**.** Quantification of the liver phosphoproteome response to insulin stimulation reveals novel regulated sites. *Nat Biotechnol* submitted.
4. Jürgen Cox, Christian Luber, Nagarjuna Nagaraj and Matthias Mann. Delayed normalization and maximal peptide ratio pairing for proteome-wide label-free quantitation. *Nat Methods* in revision

7. References

1. Taniguchi, C. M.; Emanuelli, B.; Kahn, C. R., Critical nodes in signalling pathways: insights into insulin action. *Nat Rev Mol Cell Biol* **2006,** 7, (2), 85-96.

2. Gossage, L.; Eisen, T., Targeting multiple kinase pathways: a change in paradigm. *Clin Cancer Res* **2010**
16, (7), 1973-8.

3. Luban, J.; Goff, S. P., The yeast two-hybrid system for studying protein-protein interactions. *Curr Opin Biotechnol* **1995,** 6, (1), 59-64.

4. Smith, G. P., Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **1985,** 228, (4705), 1315-7.

5. Deane, C. M.; Salwinski, L.; Xenarios, I.; Eisenberg, D., Protein interactions: two methods for assessment of the reliability of high throughput observations. *Mol Cell Proteomics* **2002,** 1, (5), 349-56.

6. Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J. C.; Yan, J. X.; Gooley, A. A.; Hughes, G.; Humphery-Smith, I.; Williams, K. L.; Hochstrasser, D. F., From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* **1996,** 14, (1), 61-5.

7. Corthals, G. L.; Wasinger, V. C.; Hochstrasser, D. F.; Sanchez, J. C., The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis* **2000,** 21, (6), 1104-15.

8. Makarov, A.; Scigelova, M., Coupling liquid chromatography to Orbitrap mass spectrometry. *J Chromatogr A* **2010,** 1217, (25), 3938-45.

9. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** 422, (6928), 198-207.

10. Gstaiger, M.; Aebersold, R., Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nat Rev Genet* **2009,** 10, (9), 617-27.

11. Yates, J. R.; Ruse, C. I.; Nakorchevsky, A., Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng* **2009,** 11, 49-79.

12. Domon, B.; Aebersold, R., Mass spectrometry and protein analysis. *Science* **2006,** 312, (5771), 212-7.

13. Kocher, T.; Superti-Furga, G., Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat Methods* **2007,** 4, (10), 807-15.

14. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S., Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev* **1998,** 17, (1), 1-35.

15. Makarov, A.; Denisov, E.; Lange, O.; Horning, S., Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. *J Am Soc Mass Spectrom* **2006,** 17, (7), 977-82.

16. Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S., Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem* **2006,** 78, (7), 2113-20.

17. Makarov, A.; Denisov, E., Dynamics of ions of intact proteins in the Orbitrap mass analyzer. *J Am Soc Mass Spectrom* **2009,** 20, (8), 1486-95.

18. Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R., The Orbitrap: a new mass spectrometer. *J Mass Spectrom* **2005,** 40, (4), 430-43.

19. Wilm, M., Quantitative proteomics in biological research. *Proteomics* **2009,** 9, (20), 4590-605.

20. Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* **2007,** 389, (4), 1017-31.

21. Schulze, W. X.; Usadel, B., Quantitation in mass-spectrometry-based proteomics. *Annu Rev Plant Biol* **2010,** 61, 491-516.

22. Ong, S. E.; Mann, M., Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* **2005,** 1, (5), 252-62.

23. Steen, H.; Pandey, A., Proteomics goes quantitative: measuring protein abundance. *Trends Biotechnol* **2002,** 20, (9), 361-4.

24. Yan, W.; Chen, S. S., Mass spectrometry-based quantitative proteomic profiling. *Brief Funct Genomic Proteomic* **2005,** 4, (1), 27-38.

25. Ong, S. E.; Foster, L. J.; Mann, M., Mass spectrometric-based approaches in quantitative proteomics. *Methods* **2003,** 29, (2), 124-30.

26. Lahm, H. W.; Langen, H., Mass spectrometry: a tool for the identification of proteins separated by gels. *Electrophoresis* **2000,** 21, (11), 2105-14.

27. Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T., Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A* **1999,** 96, (12), 6591-6.

28. Mann, M., Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Biol* **2006,** 7, (12), 952-8.

29. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **2002,** 1, (5), 376-86.

30. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **1999,** 17, (10), 994-9.

31. Olsen, J. V.; Andersen, J. R.; Nielsen, P. A.; Nielsen, M. L.; Figeys, D.; Mann, M.; Wisniewski, J. R., HysTag--a novel proteomic quantification tool applied to differential display analysis of membrane proteins from distinct areas of mouse brain. *Mol Cell Proteomics* **2004,** 3, (1), 82-92.

32. Hsu, J. L.; Huang, S. Y.; Chow, N. H.; Chen, S. H., Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem* **2003,** 75, (24), 6843-52.

33. Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **2004,** 3, (12), 1154-69.

34. Geiger, T.; Cox, J.; Ostasiewicz, P.; Wisniewski, J. R.; Mann, M., Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat Methods* **2010,** 7, (5), 383-5.

35. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat Methods* **2009,** 6, (5), 359-62.

36. de Godoy, L. M.; Olsen, J. V.; Cox, J.; Nielsen, M. L.; Hubner, N. C.; Frohlich, F.; Walther, T. C.; Mann, M., Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **2008,** 455, (7217), 1251-4.

37. Choudhary, C.; Mann, M., Decoding signalling networks by mass spectrometry-based proteomics. *Nat Rev Mol Cell Biol* **2010,** 11, (6), 427-39.

38. Luber, C. A.; Cox, J.; Lauterbach, H.; Fancke, B.; Selbach, M.; Tschopp, J.; Akira, S.; Wiegand, M.; Hochrein, H.; O'Keeffe, M.; Mann, M., Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. *Immunity* 32, (2), 279-89.

39. Olsen, J. V.; Vermeulen, M.; Santamaria, A.; Kumar, C.; Miller, M. L.; Jensen, L. J.; Gnad, F.; Cox, J.; Jensen, T. S.; Nigg, E. A.; Brunak, S.; Mann, M., Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci Signal* **2010,** 3, (104), ra3.

40. Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M., Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **2006,** 127, (3), 635-48.

41. Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M., Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **2009,** 325, (5942), 834-40.

42. Xu, G.; Paige, J. S.; Jaffrey, S. R., Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nat Biotechnol* **2010**.

43. Hubner, N. C.; Bird, A. W.; Cox, J.; Splettstoesser, B.; Bandilla, P.; Poser, I.; Hyman, A.; Mann, M., Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *J Cell Biol* **2010,** 189, (4), 739-54.

44. Vermeulen, M.; Hubner, N. C.; Mann, M., High confidence determination of specific protein-protein interactions using quantitative mass spectrometry. *Curr Opin Biotechnol* **2008,** 19, (4), 331-7.

45. Butter, F.; Kappei, D.; Buchholz, F.; Vermeulen, M.; Mann, M., A domesticated transposon mediates the effects of a single-nucleotide polymorphism responsible for enhanced muscle growth. *EMBO Rep* **2010,** 11, (4), 305-11.

46. Bonaldi, T.; Straub, T.; Cox, J.; Kumar, C.; Becker, P. B.; Mann, M., Combined use of RNAi and quantitative proteomics to study gene function in Drosophila. *Mol Cell* **2008,** 31, (5), 762-72.

47. Foster, L. J.; de Hoog, C. L.; Zhang, Y.; Zhang, Y.; Xie, X.; Mootha, V. K.; Mann, M., A mammalian organelle map by protein correlation profiling. *Cell* **2006,** 125, (1), 187-99.

48. Au, C. E.; Bell, A. W.; Gilchrist, A.; Hiding, J.; Nilsson, T.; Bergeron, J. J., Organellar proteomics to create the cell map. *Curr Opin Cell Biol* **2007,** 19, (4), 376-85.

49. Yates, J. R., 3rd; Gilchrist, A.; Howell, K. E.; Bergeron, J. J., Proteomics of organelles and large cellular structures. *Nat Rev Mol Cell Biol* **2005,** 6, (9), 702-14.

50. Liotta, L. A.; Ferrari, M.; Petricoin, E., Clinical proteomics: written in blood. *Nature* **2003,** 425, (6961), 905.

51. Araujo, R. P.; Petricoin, E. F.; Liotta, L. A., Critical dependence of blood-borne biomarker concentrations on the half-lives of their carrier proteins. *J Theor Biol* **2008,** 253, (3), 616-22.

52. Apweiler, R.; Aslanidis, C.; Deufel, T.; Gerstner, A.; Hansen, J.; Hochstrasser, D.; Kellner, R.; Kubicek, M.; Lottspeich, F.; Maser, E.; Mewes, H. W.; Meyer, H. E.; Mullner, S.; Mutter, W.; Neumaier, M.; Nollau, P.; Nothwang, H. G.; Ponten, F.; Radbruch, A.; Reinert, K.; Rothe, G.; Stockinger, H.; Tarnok, A.; Taussig, M. J.; Thiel, A.; Thiery, J.; Ueffing, M.; Valet, G.; Vandekerckhove, J.; Verhuven, W.; Wagener, C.; Wagner, O.; Schmitz, G., Approaching clinical proteomics: current state and future fields of application in fluid proteomics. *Clin Chem Lab Med* **2009,** 47, (6), 724-44.

53. Apweiler, R.; Aslanidis, C.; Deufel, T.; Gerstner, A.; Hansen, J.; Hochstrasser, D.; Kellner, R.; Kubicek, M.; Lottspeich, F.; Maser, E.; Mewes, H. W.; Meyer, H. E.; Mullner, S.; Mutter, W.; Neumaier, M.; Nollau, P.; Nothwang, H. G.; Ponten, F.; Radbruch, A.; Reinert, K.; Rothe, G.; Stockinger, H.; Tarnok, A.; Taussig, M. J.; Thiel, A.; Thiery, J.; Ueffing, M.; Valet, G.; Vandekerckhove, J.; Wagener, C.; Wagner, O.; Schmitz, G., Approaching clinical proteomics: current state and future fields of application in cellular proteomics. *Cytometry A* **2009,** 75, (10), 816-32.

54. Rifai, N.; Gillette, M. A.; Carr, S. A., Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol* **2006,** 24, (8), 971-83.

55. States, D. J.; Omenn, G. S.; Blackwell, T. W.; Fermin, D.; Eng, J.; Speicher, D. W.; Hanash, S. M., Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nat Biotechnol* **2006,** 24, (3), 333-8.

56. Omenn, G. S., Strategies for plasma proteomic profiling of cancers. *Proteomics* **2006,** 6, (20), 5662-73.

57. Barker, P. E.; Wagner, P. D.; Stein, S. E.; Bunk, D. M.; Srivastava, S.; Omenn, G. S., Standards for plasma and serum proteomics in early cancer detection: a needs assessment report from the national institute of standards and technology--National Cancer Institute Standards,

Methods, Assays, Reagents and Technologies Workshop, August 18-19, 2005. *Clin Chem* **2006,** 52, (9), 1669-74.

58. Granger, J.; Siddiqui, J.; Copeland, S.; Remick, D., Albumin depletion of human plasma also removes low abundance proteins including the cytokines. *Proteomics* **2005,** 5, (18), 4713-8.

59. Good, D. M.; Thongboonkerd, V.; Novak, J.; Bascands, J. L.; Schanstra, J. P.; Coon, J. J.; Dominiczak, A.; Mischak, H., Body fluid proteomics for biomarker discovery: lessons from the past hold the key to success in the future. *J Proteome Res* **2007,** 6, (12), 4549-55.

60. Anderson, N. L.; Anderson, N. G., The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* **2002,** 1, (11), 845-67.

61. Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M., Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* **1996,** 68, (5), 850-8.

62. Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M., Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **1996,** 379, (6564), 466-9.

63. Hubner, N. C.; Ren, S.; Mann, M., Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. *Proteomics* **2008,** 8, (23-24), 4862-72.

64. Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **2001,** 19, (3), 242-7.

65. Forner, F.; Kumar, C.; Luber, C. A.; Fromme, T.; Klingenspor, M.; Mann, M., Proteome differences between brown and white fat mitochondria reveal specialized metabolic functions. *Cell Metab* **2009,** 10, (4), 324-35.

66. Butter, F.; Scheibe, M.; Morl, M.; Mann, M., Unbiased RNA-protein interaction screen by quantitative proteomics. *Proc Natl Acad Sci U S A* **2009,** 106, (26), 10626-31.

67. Mittler, G.; Butter, F.; Mann, M., A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. *Genome Res* **2009,** 19, (2), 284-93.

68. Grimsrud, P. A.; Swaney, D. L.; Wenger, C. D.; Beauchene, N. A.; Coon, J. J., Phosphoproteomics for the masses. *ACS Chem Biol* **2010,** 5, (1), 105-19.

69. Macek, B.; Mann, M.; Olsen, J. V., Global and site-specific quantitative phosphoproteomics: principles and applications. *Annu Rev Pharmacol Toxicol* **2009,** 49, 199-221.

70. Thingholm, T. E.; Jensen, O. N.; Larsen, M. R., Analytical strategies for phosphoproteomics. *Proteomics* **2009,** 9, (6), 1451-68.

71. Zhong, J.; Molina, H.; Pandey, A., Phosphoproteomics. *Curr Protoc Protein Sci* **2007,** Chapter 24, Unit 24 4.

72. Zielinska, D. F.; Gnad, F.; Wisniewski, J. R.; Mann, M., Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **2010,** 141, (5), 897-907.

73. Matic, I.; van Hagen, M.; Schimmel, J.; Macek, B.; Ogg, S. C.; Tatham, M. H.; Hay, R. T.; Lamond, A. I.; Mann, M.; Vertegaal, A. C., In vivo identification of human small ubiquitin-like modifier polymerization sites by high accuracy mass spectrometry and an in vitro to in vivo strategy. *Mol Cell Proteomics* **2008,** 7, (1), 132-44.

74. Wisniewski, J. R.; Zougman, A.; Mann, M., Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J Proteome Res* **2009,** 8, (12), 5674-8.

75. Ishihama, Y.; Rappsilber, J.; Mann, M., Modular stop and go extraction tips with stacked disks for parallel and multidimensional Peptide fractionation in proteomics. *J Proteome Res* **2006,** 5, (4), 988-94.

76. Rappsilber, J.; Ishihama, Y.; Mann, M., Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **2003,** 75, (3), 663-70.

77. Ostasiewicz, P.; Zielinska, D. F.; Mann, M.; Wisniewski, J. R., Proteome, phosphoproteome, and N-glycoproteome are quantitatively preserved in formalin-fixed paraffin-embedded tissue and analyzable by high-resolution mass spectrometry. *J Proteome Res* 9, (7), 3688-700.

78. Zanivan, S.; Gnad, F.; Wickstrom, S. A.; Geiger, T.; Macek, B.; Cox, J.; Fassler, R.; Mann, M., Solid tumor proteome and phosphoproteome analysis by high resolution mass spectrometry. *J Proteome Res* **2008,** 7, (12), 5314-26.

79. Dass, C., *Fundamentals of contemporary mass spectrometry*

John Wiley and Sons: 2007; p 585.

80. Thomson, J. J., Cathode Rays. *Philosophical magazine* **1897,** 43, (293).

81. Stephens, W. E., A pulsed mass spectrometer with time dispersion *Physical reviews* **1946,** 69, 691.

82. H. Sommer, H. A. T. a. J. A. H., Measurement of e/m by cyclotron resonance. *Physical reviews* **1951,** 82, 697-702.

83. Steinwedel, W. P. a. H., A new mass spectrometer without magnetic field. *Zeitschrift für Naturforschung* **1953,** 8a, 448-450.

84. W. Paul, P. R. a. O. Z., The electric mass filter as mass spectrometer and isotope separator. *Zeitschrift für Physik* **1958,** 152, 143-182.

85. Jennings, K. R., Collision-induced decompositions of aromatic molecular ions. *International Journal of Mass Spectrometry and Ion Physics* **1968,** 1, 227-235.

86. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989,** 246, (4926), 64-71.

87. Wilm, M.; Mann, M., Analytical properties of the nanoelectrospray ion source. *Anal Chem* **1996,** 68, (1), 1-8.

88. Smythe, W. R.; Mattauch, J., A new mass spectrometer. *Physical Review* **1932,** 40, (3), 0429-0433.

89. Stephens, W. E., A PULSED MASS SPECTROMETER WITH TIME DISPERSION. *Physical Review* **1946,** 69, (11-1), 691-691.

90. Paul, W.; Steinwedel, H., *EIN NEUES MASSENSPEKTROMETER OHNE MAGNETFELD. *Zeitschrift Fur Naturforschung Section a-a Journal of Physical Sciences* **1953,** 8, (7), 448-450.

91. Finnigan, R. E., Quadrupole Mass Spectrometers - from Development to Commercialization. *Analytical Chemistry* **1994,** 66, (19), A969-A975.

92. Dawson, P. H., Quadrupole Mass Analyzers - Performance, Design and Some Recent Applications. *Mass Spectrometry Reviews* **1986,** 5, (1), 1-37.

93. Schwartz, J. C.; Jardine, I., Quadrupole ion trap mass spectrometry. *Methods Enzymol* **1996,** 270, 552-86.

94. Stroobant, E. d. H. a. V., *Mass Spectrometry Principles and Applications*. 3 ed.; John Wiley & Sons Ltd.: 2007.

95. Perry, R. H.; Cooks, R. G.; Noll, R. J., Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrom Rev* **2008,** 27, (6), 661-99.

96. Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M., Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* **2005,** 4, (12), 2010-21.

97. Scigelova, M.; Makarov, A., Orbitrap mass analyzer--overview and applications in proteomics. *Proteomics* **2006,** 6 Suppl 2, 16-21.

98. Olsen, J. V.; Schwartz, J. C.; Griep-Raming, J.; Nielsen, M. L.; Damoc, E.; Denisov, E.; Lange, O.; Remes, P.; Taylor, D.; Splendore, M.; Wouters, E. R.; Senko, M.; Makarov, A.; Mann, M.; Horning, S., A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol Cell Proteomics* **2009,** 8, (12), 2759-69.

99. Roepstorff, P.; Fohlman, J., Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* **1984,** 11, (11), 601.

100. Sleno, L.; Volmer, D. A., Ion activation methods for tandem mass spectrometry. *J Mass Spectrom* **2004,** 39, (10), 1091-112.

101. McLuckey, S. A., PRINCIPLES OF COLLISIONAL ACTIVATION IN ANALYTICAL MASS-SPECTROMETRY. *Journal of the American Society for Mass Spectrometry* **1992,** 3, (6), 599-614.

102. Hayes, R. N.; Gross, M. L., Collision-Induced Dissociation. *Methods in Enzymology* **1990,** 193, 237-263.

103. Zubarev, R. A., Electron-capture dissociation tandem mass spectrometry. *Current Opinion in Biotechnology* **2004,** 15, (1), 12-16.

104. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W., Electron capture dissociation of multiply charged protein cations. A nonergodic process. *Journal of the American Chemical Society* **1998,** 120, (13), 3265-3266.

105. Zubarev, R. A.; Kruger, N. A.; Fridriksson, E. K.; Lewis, M. A.; Horn, D. M.; Carpenter, B. K.; McLafferty, F. W., Electron capture dissociation of gaseous multiply-charged proteins is favored at disulfide bonds and other sites of high hydrogen atom affinity. *Journal of the American Chemical Society* **1999,** 121, (12), 2857-2862.

106. Kelleher, R. L.; Zubarev, R. A.; Bush, K.; Furie, B.; Furie, B. C.; McLafferty, F. W.; Walsh, C. T., Localization of labile posttranslational modifications by electron capture

dissociation: The case of gamma-carboxyglutamic acid. *Analytical Chemistry* **1999,** 71, (19), 4250-4253.

107. Mirgorodskaya, E.; Roepstorff, P.; Zubarev, R. A., Localization of O-glycosylation sites in peptides by electron capture dissociation in a fourier transform mass spectrometer. *Analytical Chemistry* **1999,** 71, (20), 4431-4436.

108. Ge, Y.; Lawhorn, B. G.; ElNaggar, M.; Strauss, E.; Park, J. H.; Begley, T. P.; McLafferty, F. W., Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *Journal of the American Chemical Society* **2002,** 124, (4), 672-678.

109. Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* **2004,** 101, (26), 9528-33.

110. Udeshi, N. D.; Shabanowitz, J.; Hunt, D. F.; Rose, K. L., Analysis of proteins and peptides on a chromatographic timescale by electron-transfer dissociation MS. *Febs J* **2007,** 274, (24), 6269-76.

111. Schwartz, J. C.; Syka, J. E. P.; Jardine, I., High-Resolution on a Quadrupole Ion Trap Mass-Spectrometer. *Journal of the American Society for Mass Spectrometry* **1991,** 2, (3), 198-204.

112. Williams, J. D.; Cox, K. A.; Cooks, R. G.; Kaiser, R. E.; Schwartz, J. C., High Mass-Resolution Using a Quadrupole Ion-Trap Mass-Spectrometer. *Rapid Communications in Mass Spectrometry* **1991,** 5, (7), 327-329.

113. Louris, J. N.; Amy, J. W.; Ridley, T. Y.; Cooks, R. G., Injection of Ions into a Quadrupole Ion Trap Mass-Spectrometer. *International Journal of Mass Spectrometry and Ion Processes* **1989,** 88, (2-3), 97-111.

114. Vanberkel, G. J.; Glish, G. L.; McLuckey, S. A., Electrospray Ionization Combined with Ion Trap Mass-Spectrometry. *Analytical Chemistry* **1990,** 62, (13), 1284-1295.

115. Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M., Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* **2007,** 4, (9), 709-12.

116. Brunzel, N. A., *Fundamentals of Urine and Body Fluid Analysis*. 2 ed.; SAUNDERS: Philadelphia, 2004.

117. Decramer, S.; Gonzalez de Peredo, A.; Breuil, B.; Mischak, H.; Monsarrat, B.; Bascands, J. L.; Schanstra, J. P., Urine in clinical proteomics. *Mol Cell Proteomics* **2008,** 7, (10), 1850-62.

118. Pisitkun, T.; Johnstone, R.; Knepper, M. A., Discovery of urinary biomarkers. *Mol Cell Proteomics* **2006,** 5, (10), 1760-71.

119. Thongboonkerd, V., Recent progress in urinary proteomics. *Proteomics Clinical Applications* **2007,** 1, (8), 780-791.

120. Candiano, G.; Santucci, L.; Petretto, A.; Bruschi, M.; Dimuccio, V.; Urbani, A.; Bagnasco, S.; Ghiggeri, G. M., 2D-electrophoresis and the urine proteome map: Where do we stand? *Journal of Proteomics* **2010,** 73, (5), 829-844.

121. Zerefos, P. G.; Vougas, K.; Dimitraki, P.; Kossida, S.; Petrolekas, A.; Stravodimos, K.; Giannopoulos, A.; Fountoulakis, M.; Vlahou, A., Characterization of the human urine proteome by preparative electrophoresis in combination with 2-DE. *Proteomics* **2006,** 6, (15), 4346-55.

122. Voshol, H.; Brendlen, N.; Muller, D.; Inverardi, B.; Augustin, A.; Pally, C.; Wieczorek, G.; Morris, R. E.; Raulf, F.; van Oostrum, J., Evaluation of biomarker discovery approaches to detect protein biomarkers of acute renal allograft rejection. *J Proteome Res* **2005,** 4, (4), 1192-9.

123. Ward, D. G.; Nyangoma, S.; Joy, H.; Hamilton, E.; Wei, W.; Tselepis, C.; Steven, N.; Wakelam, M. J.; Johnson, P. J.; Ismail, T.; Martin, A., Proteomic profiling of urine for the detection of colon cancer. *Proteome Sci* **2008,** 6, 19.

124. Pisitkun, T.; Shen, R. F.; Knepper, M. A., Identification and proteomic profiling of exosomes in human urine. *Proc Natl Acad Sci U S A* **2004,** 101, (36), 13368-73.

125. Gonzales, P. A.; Pisitkun, T.; Hoffert, J. D.; Tchapyjnikov, D.; Star, R. A.; Kleta, R.; Wang, N. S.; Knepper, M. A., Large-scale proteomics and phosphoproteomics of urinary exosomes. *J Am Soc Nephrol* **2009,** 20, (2), 363-79.

126. Adachi, J.; Kumar, C.; Zhang, Y.; Olsen, J. V.; Mann, M., The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome Biol* **2006,** 7, (9), R80.

127. Zhou, H.; Yuen, P. S.; Pisitkun, T.; Gonzales, P. A.; Yasuda, H.; Dear, J. W.; Gross, P.; Knepper, M. A.; Star, R. A., Collection, storage, preservation, and normalization of human urinary exosomes for biomarker discovery. *Kidney Int* **2006,** 69, (8), 1471-6.

128. Gwinner, W., Renal transplant rejection markers. *World J Urol* **2007,** 25, (5), 445-55.

129. Steinhoff, J.; Buhner, U.; Preuss, R.; Sack, K., C-reactive protein and alpha 2 macroglobulin in urine as markers of renal transplant rejection. *Transplant Proc* **1994,** 26, (3), 1768.

130. Clarke, W.; Silverman, B. C.; Zhang, Z.; Chan, D. W.; Klein, A. S.; Molmenti, E. P., Characterization of renal allograft rejection by urinary proteomic analysis. *Ann Surg* **2003,** 237, (5), 660-4; discussion 664-5.

131. O'Riordan, E.; Orlova, T. N.; Mei, J. J.; Butt, K.; Chander, P. M.; Rahman, S.; Mya, M.; Hu, R.; Momin, J.; Eng, E. W.; Hampel, D. J.; Hartman, B.; Kretzler, M.; Delaney, V.; Goligorsky, M. S., Bioinformatic analysis of the urine proteome of acute allograft rejection. *J Am Soc Nephrol* **2004,** 15, (12), 3240-8.

132. Schaub, S.; Rush, D.; Wilkins, J.; Gibson, I. W.; Weiler, T.; Sangster, K.; Nicolle, L.; Karpinski, M.; Jeffery, J.; Nickerson, P., Proteomic-based detection of urine proteins associated with acute renal allograft rejection. *J Am Soc Nephrol* **2004,** 15, (1), 219-27.

133. Wittke, S.; Haubitz, M.; Walden, M.; Rohde, F.; Schwarz, A.; Mengel, M.; Mischak, H.; Haller, H.; Gwinner, W., Detection of acute tubulointerstitial rejection by proteomic analysis of urinary samples in renal transplant recipients. *Am J Transplant* **2005,** 5, (10), 2479-88.

134. Meguid El Nahas, A.; Bello, A. K., Chronic kidney disease: the global challenge. *Lancet* **2005,** 365, (9456), 331-40.

135. Haubitz, M.; Wittke, S.; Weissinger, E. M.; Walden, M.; Rupprecht, H. D.; Floege, J.; Haller, H.; Mischak, H., Urine protein patterns can serve as diagnostic tools in patients with IgA nephropathy. *Kidney Int* **2005,** 67, (6), 2313-20.

136. Neuhoff, N.; Kaiser, T.; Wittke, S.; Krebs, R.; Pitt, A.; Burchard, A.; Sundmacher, A.; Schlegelberger, B.; Kolch, W.; Mischak, H., Mass spectrometry for the detection of differentially expressed proteins: a comparison of surface-enhanced laser desorption/ionization and capillary electrophoresis/mass spectrometry. *Rapid Commun Mass Spectrom* **2004,** 18, (2), 149-56.

137. Decramer, S.; Bascands, J. L.; Schanstra, J. P., Non-invasive markers of ureteropelvic junction obstruction. *World J Urol* **2007,** 25, (5), 457-65.

138. Decramer, S.; Wittke, S.; Mischak, H.; Zurbig, P.; Walden, M.; Bouissou, F.; Bascands, J. L.; Schanstra, J. P., Predicting the clinical outcome of congenital unilateral ureteropelvic junction obstruction in newborn by urinary proteome analysis. *Nat Med* **2006,** 12, (4), 398-400.

139. Meier, M.; Kaiser, T.; Herrmann, A.; Knueppel, S.; Hillmann, M.; Koester, P.; Danne, T.; Haller, H.; Fliser, D.; Mischak, H., Identification of urinary protein pattern in type 1 diabetic adolescents with early diabetic nephropathy by a novel combined proteome analysis. *J Diabetes Complications* **2005,** 19, (4), 223-32.

140. Rossing, K.; Mischak, H.; Dakna, M.; Zurbig, P.; Novak, J.; Julian, B. A.; Good, D. M.; Coon, J. J.; Tarnow, L.; Rossing, P., Urinary proteomics in diabetes and CKD. *J Am Soc Nephrol* **2008,** 19, (7), 1283-90.

141. Weissinger, E. M.; Schiffer, E.; Hertenstein, B.; Ferrara, J. L.; Holler, E.; Stadler, M.; Kolb, H. J.; Zander, A.; Zurbig, P.; Kellmann, M.; Ganser, A., Proteomic patterns predict acute graft-versus-host disease after allogeneic hematopoietic stem cell transplantation. *Blood* **2007,** 109, (12), 5511-9.

142. Zimmerli, L. U.; Schiffer, E.; Zurbig, P.; Good, D. M.; Kellmann, M.; Mouls, L.; Pitt, A. R.; Coon, J. J.; Schmieder, R. E.; Peter, K. H.; Mischak, H.; Kolch, W.; Delles, C.; Dominiczak, A. F., Urinary proteomic biomarkers in coronary artery disease. *Mol Cell Proteomics* **2008,** 7, (2), 290-8.

143. Theodorescu, D.; Fliser, D.; Wittke, S.; Mischak, H.; Krebs, R.; Walden, M.; Ross, M.; Eltze, E.; Bettendorf, O.; Wulfing, C.; Semjonow, A., Pilot study of capillary electrophoresis coupled to mass spectrometry as a tool to define potential prostate cancer biomarkers in urine. *Electrophoresis* **2005,** 26, (14), 2797-808.

144. Theodorescu, D.; Wittke, S.; Ross, M. M.; Walden, M.; Conaway, M.; Just, I.; Mischak, H.; Frierson, H. F., Discovery and validation of new protein biomarkers for urothelial cancer: a prospective analysis. *Lancet Oncol* **2006,** 7, (3), 230-40.

145. Kreunin, P.; Zhao, J.; Rosser, C.; Urquidi, V.; Lubman, D. M.; Goodison, S., Bladder cancer associated glycoprotein signatures revealed by urinary proteomic profiling. *Journal of Proteome Research* **2007,** 6, (7), 2631-2639.

146. Quintana, L. F.; Campistol, J. M.; Alcolea, M. P.; Banon-Maneus, E.; Sole-Gonzalez, A.; Cutillas, P. R., Application of Label-free Quantitative Peptidomics for the Identification of Urinary Biomarkers of Kidney Chronic Allograft Dysfunction. *Molecular & Cellular Proteomics* **2009,** 8, (7), 1658-1673.

147. Ling, X. B.; Sigdel, T. K.; Lau, K.; Ying, L. H.; Lau, I.; Schilling, J.; Sarwal, M. M., Integrative Urinary Peptidomics in Renal Transplantation Identifies Biomarkers for Acute Rejection. *Journal of the American Society of Nephrology* **2010,** 21, (4), 646-653.

148. Schlatzer, D. M.; Dazard, J. E.; Dharsee, M.; Ewing, R. M.; Ilchenko, S.; Stewart, I.; Christ, G.; Chance, M. R., Urinary Protein Profiles in a Rat Model for Diabetic Complications. *Molecular & Cellular Proteomics* **2009,** 8, (9), 2145-2158.

149. Kentsis, A.; Lin, Y. Y.; Kurek, K.; Calicchio, M.; Wang, Y. Y.; Monigatti, F.; Campagne, F.; Lee, R.; Horwitz, B.; Steen, H.; Bachur, R., Discovery and Validation of Urine Markers of Acute Pediatric Appendicitis Using High-Accuracy Mass Spectrometry. *Annals of Emergency Medicine* **2010,** 55, (1), 62-70.

150. Sigdel, T. K.; Kaushal, A.; Gritsenko, M.; Norbeck, A. D.; Qian, W. J.; Xiao, W. Z.; Camp, D. G.; Smith, R. D.; Sarwal, M. M., Shotgun proteomics identifies proteins specific for acute renal transplant rejection. *Proteomics Clinical Applications* **2010,** 4, (1), 32-47.

151. Thongboonkerd, V., Current status of renal and urinary proteomics: ready for routine clinical application? *Nephrol Dial Transplant* 25, (1), 11-6.

152. Guidotti, G., Membrane proteins. *Annu Rev Biochem* **1972,** 41, 731-52.

153. Kentsis, A.; Monigatti, F.; Dorff, K.; Campagne, F.; Bachur, R.; Steen, H., Urine proteomics for profiling of human disease using high accuracy mass spectrometry. *Proteomics Clinical Applications* **2009,** 3, (9), 1052-1061.

154. Sun, W.; Chen, Y.; Li, F. X.; Zhang, L.; Yang, R. F.; Zhang, Z.; Zheng, D. X.; Gao, Y. H., Dynamic urinary proteomic analysis reveals stable proteins to be potential biomarkers. *Proteomics Clinical Applications* **2009,** 3, (3), 370-382.