Department of Medical Biochemistry and Biophysics

Karolinska Institutet, Stockholm, Sweden

PROTEOME STUDIES, WITH EMPHASIS ON THE KIDNEY GLOMERULUS

Sam Tryggvason

Stockholm 2012

Published by Karolinska institutet
Box 200, SE-171 77 Stockholm, Sweden

ISBN 978-91-7457-696-2

ABSTRACT

Proteomics aims to link genes to protein expression and function. Proteomics is a very versatile and technology-dependent discipline and the development of mass spectrometry-based proteomics has advanced the understanding of the complex and dynamic nature of proteomes. Proteomics does not have a “one-size-fits-all” strategy to answer all biological questions. New technical approaches are needed to overcome the hurdles, e.g. issues related to complexity and dynamic ranges of the samples, high throughput and post-translational modifications, in the field of proteomics. Due to recent technological development, proteomics has become complementary to genomics and transcriptomics in systems biology and can provide unique contributions.

In this thesis a novel microfluidic compact disc based system was tested in use of high throughput sample preparation of digested peptide samples prior to matrix assisted laser desorption/ionization mass spectrometry (Paper I). A novel microfluidic electrocapture technology able to immobilize charged biomolecules in an electric field along a hydrodynamic flow was further developed to establish its ability to separate peptides of protein digests (Paper III). The same technology was later also used in analysis of glomerular membrane proteins as a part of a two-dimensional fractionation system in combination with liquid chromatography mass spectrometry (Paper IV).

The kidney glomerulus functions as a specialized blood filter unit that forms the primary urine. The filtration barrier is composed of a three-layer barrier, which consists of endothelial cells, the glomerular basement membrane and the slit diaphragm between epithelial podocyte foot processes. The glomeruli are sensitive to damage and glomerular diseases underlie the majority of diseases leading to chronic kidney disease and renal failure. Even though the molecular composition of the glomerulus is quite well known, the pathomechanisms of glomerular diseases are still poorly unknown. The mouse glomerular proteome was now studied with two-dimensional gel electrophoresis in combination with mass spectrometry (Paper II). Finally a meta-analysis of currently available proteomic and transcriptomic glomerular expression data in normal and disease states was performed to explore the usability of expression profiling in the diagnosis, prognosis and prediction of glomerular diseases (Paper V).

LIST OF PUBLICATIONS

I. Hirschberg D., **Tryggvason S.**, Gustafsson M., Bergman T., Swedenborg J., Hedin U. and Jörnvall H. Identification of endothelial proteins by MALDI-MS using a compact disc microfluidic system. Protein J. 2004 May 23(4):263-71.

II. **Tryggvason S.***, Nukui M.*, Oddsson Á., Tryggvason K. and Jörnvall H. Glomerulus proteome analysis with two-dimensional gel electrophoresis and mass spectrometry. Cell Mol Life Sci. 2007 Dec 64(24):3317-35.

III. Astorga-Wells J., Vollmer S., **Tryggvason S.**, Bergman T. and Jörnvall H. Microfluidic electrocapture for separation of peptides. Anal Chem. 2005 Nov 15;77(22):7131-6.

IV. Astorga-Wells J.*, **Tryggvason S.***, Vollmer S., Alvelius G., Palmberg C. and Jörnvall H. Membrane protein identifications by mass spectrometry using electrocapture-based as part of a two-dimensional fractionation. Anal Biochem. 2008 Oct 1;381(1):33-42.

V. **Tryggvason S.**, Nukui M., Norlin J., Haraldsson B., Betsholtz C., Jörnvall H., Tryggvason K. and He L. Expression pathways and networks in glomerular disease – a meta-analysis. Manuscript submitted.

* These authors contributed equally

TABLE OF CONTENTS

LIST OF SELECTED ABBREVIATIONS

2-DE	2-dimensional gel electrophoresis
ADR	Adriamycin
Ambic	Ammonium bicarbonate
CD	Compact disc
CE	Capillary electrophoresis
CHAPS	3-[(3-cholamido propyl)-dimethylammonio]-1-propane sulfonate
CKD	Chronic kidney disease
DN	Diabetic nephropathy
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EC	Electrocapture
ESI	Electrospray ionization
ESI-MS/MS	Electrospray ionization tandem mass spectrometry
ESRD	End-stage renal disease
FSGS	Focal segmental glomerulosclerosis
GBM	Glomerular basement membrane
GO	Gene ontology
HPLC	High performance liquid chromatography
IEF	Isoelectric focusing
IPA	Ingenuity pathway analysis
IPI	International protein index
IPG	Immobilized pH gradient
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LPS	Lipopolysaccharide
m/z	Mass-to-charge-ratio
MALDI	Matrix assisted laser desorption ionization
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
pI	Isoelectric point
PMF	Peptide mass fingerprinting
PFF	Peptide fragmentation fingerprinting

PTM	Post-translational modification
Q	Quadrupole
Q-TOF	Quadrupole-time-of-flight
RF	Radio frequency
RP	Reverse phase
RPLC	Reverse phase liquid chromatography
RT-PCR	Reverse transcriptase-polymerase chain reaction
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TOF	Time-of-flight

Three and one letter codes for the 20 naturally occurring amino acids and their monoisotopic residue masses in Da.

Alanine	Ala	A	71.037
Arginine	Arg	R	156.101
Asparagine	Asn	N	114.043
Aspartic acid	Asp	D	115.027
Cysteine	Cys	C	103.009
Glutamic acid	Glu	E	129.043
Glutamine	Gln	Q	128.059
Glycine	Gly	G	57.021
Histidine	His	H	137.057
Isoleucine	Ile	I	113.084
Leucine	Leu	L	113.084
Lysine	Lys	K	128.095
Methionine	Met	M	131.040
Phenylalanine	Phe	F	147.068
Proline	Pro	P	97.053
Serine	Ser	S	87.032
Threonine	Thr	T	101.048
Tryptophan	Trp	W	186.079
Tyrosine	Tyr	T	163.063
Valine	Val	V	99.068

INTRODUCTION: REVIEW OF LITERATURE

Proteomics

Marc Wilkins was the first to describe the term proteome as the entire complement of proteins expressed by a genome in a limited entity[1]. Proteomics refers to complex-scale studies of protein systems, particularly the protein content of cells, tissues and organs, as well as their structures and functions. Proteins are vital parts of living organisms, as they are the main components of the physiological pathways of cells and biological structures[2]. Proteomics is instrumental in the discovery of biomarkers, which can have significance in diagnosis and prognosis of a disease, as well as monitoring the efficacy of a therapy[3,4].

The study of the proteome *i.e.* proteomics, differs from protein analysis in that it is the characterization of the complete repertoire of individual protein species that comprise the proteome rather than focusing on a single protein. Proteomics aims to identify not only all the proteins in any given cell or tissue, but also the set of all protein isoforms and modifications, the interactions between them, the structural description of proteins and their higher-order complexes, and for that matter almost everything post-genomic[1].

The development of mass spectrometry (MS) has been the most critical for the rapid advance of proteomics. Together with fractionation methods, they provide the technology to measure and identify peptides and proteins in large scale with high sensitivity[5-7]. Proteomics is more complicated than genomics, mostly because the proteome consists of tens of thousands of proteins, the concentrations of which differ from cell to cell, while the genome is the same in all cells. Furthermore, the proteome constantly changes through post-translational modifications (PTMs) and biochemical interactions of proteins with other proteins, the genome and the environment[8,9]. Proteomic analyses are important because they provide a much more complete understanding of an organism than genomics. Proteomics is today considered an integral part of biological research together with genomics and transcriptomics. Proteomics provides a functional link between expressed genes and phenotypic outcomes.

Proteomics is a difficult field and still encounters many problems in comparison with genomics and transcriptomics. The number of genes in a cell is constant in comparison with the proteins. The proteome is not a predetermined static entity. It is in constant flux, which is influenced by environmental conditions, cellular function, developmental status and

extracellular challenge. Even when a cell is in steady state, the protein complement is constantly changing, with new proteins being synthesized and older proteins being degraded and recycled in a way that does not harm the cellular equilibrium[10]. DNA and mRNA can be amplified easily using specific methods. Studying mRNA has its advantages. Reverse transcriptase polymerase chain reaction (RT-PCR) can be used to amplify mRNA. For qualitative analysis real-time RT-PCR is used. Detection and analysis of mRNAs are quite straightforward with the use of hybridization techniques. However the levels of mRNA in a cell do not always reflect the levels of corresponding proteins[11-13]. Furthermore, studying only mRNA levels cannot validate PTMs and functional analysis of proteins.

Proteins are a group of heterogeneous molecules with variable properties in weight, hydrophobicity and charge. The analysis of a full proteome poses inevitable challenges due to the limitation and variability of sample material, post-translational modifications, splice variants, sample stability and degradation, variety of their dynamic range, transient protein associations and dependence on cell type or physiological state. Success in a proteomic study is enabled and confined by the biological system, the study plan, available technology. To overcome these significant challenges in proteomics, new analytical strategies have been developed in which mass spectrometry (MS) is the central element with increased performance and versatility of the instrumentation[14,15].

Mass spectrometry

Mass spectrometry (MS) is a well-known analytical tool for the measurement of molecular weight of a sample or distinguishing molecules by their mass-to-charge ratios. Mass spectrometers are now routinely used in both industry and academia for various purposes such as drug discovery, diagnostics and bio-analyses. A mass spectrometer is an analytical device that determines the molecular weight of chemical compounds by separating molecular ions in a gas phase according to mass-to-charge ratio (m/z). Mass spectrometers have three basic components: an ion source, a mass analyzer and a detector. In the ion source, the ions are transferred into the gas phase as positive or negative ions and accelerated in an electric field. The m/z of the ions is determined in the mass analyzer. The detector registers the amount of ions at each m/z value. An optional pre-fractionation part is sometimes coupled for enhanced performance. In schemes for tandem MS analysis, dual mass analyzers, that are separated by a

collision cell for fragmentation of target molecules, are usually employed, (Figure 1). Matrix-assisted laser desorption ionization (MALDI)[16,17] and Electrospray ionization (ESI)[18] are the most common ionization techniques. The developers of these two techniques received the Nobel Prize in chemistry 2002. MALDI and ESI are so called soft ionization techniques meaning they are compatible with large and fragile biomolecules.

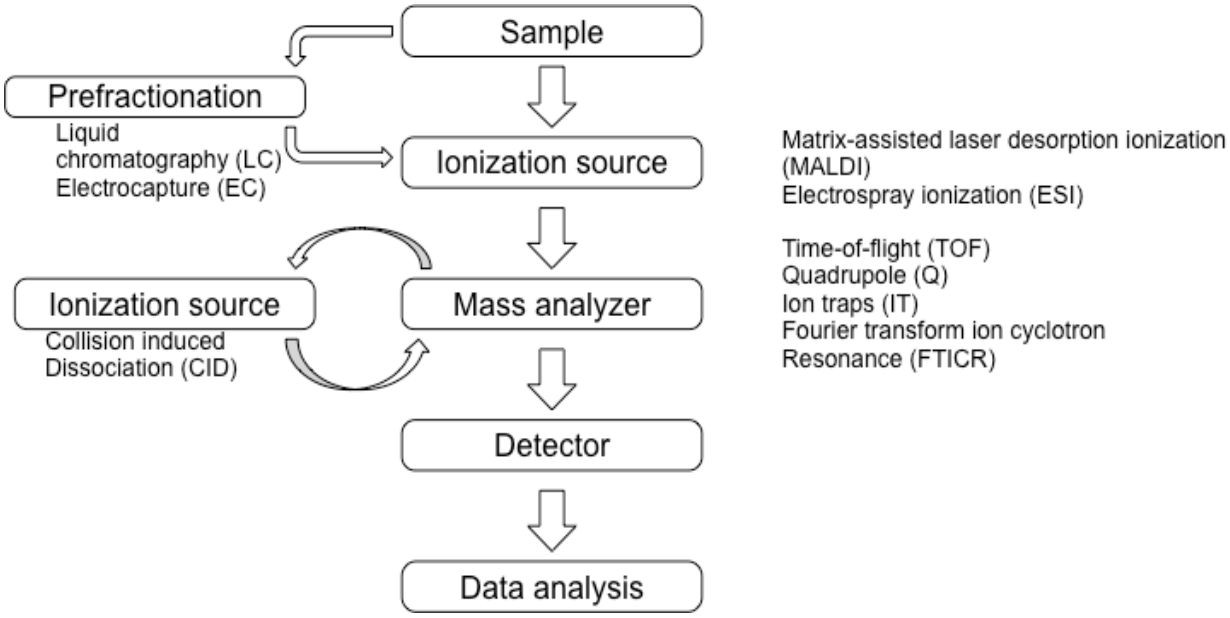


Figure 1. A simplified workflow for a mass spectrometry (MS) analysis in proteomics. A protein/peptide sample can be analyzed directly with MS if the sample is relative simple. For more complex samples, prefractionation methods can be added to increase the number of identification. Depending on the method of choice, the masses or the sequences of the peptides can be analyzed.

In the beginning of the 21st century, mass spectrometry was used in three major fields. It was used to identify proteins and biomolecules. It was also the method of choice to detect and characterize post-translational modifications. Finally, it was the preferred technique for characterization and quality control of recombinant proteins and macromolecules in both academia and biotechnological industry[19]. Since then, driven by the need to identify, characterize, and quantify proteins at ever increasing sensitivity and in ever more complex samples, a wide range of MS-based analytical approaches and experimental strategies have emerged.

The idea of shotgun proteomics evolved from the fact that the new large-scale analysis methods were needed to overcome the disadvantages of 2-DE. The term shotgun proteomics involves the analysis of a complex mixture of proteins digested into peptides, which are separated by at least a two-dimensional chromatography based separation system and eluted directly into a tandem mass spectrometer in an automated fashion. The data are analyzed by powerful computational systems[20-22]. It is also possible to do quantitative and comparative pattern analyses with mass spectrometry[23].

Ionization techniques

Matrix-assisted laser desorption ionization (MALDI)

Matrix-assisted laser desorption ionization (MALDI) was developed at the end of the 1980s[16,17] and is today widely used in mass spectrometers. MALDI ionization is based on the utilization of a crystalline matrix, which is able to absorb the energy of a pulsed laser beam and to mediate the generation of peptide ions. Small organic molecules such as α-cyano-4-hydroxy-cinnamic acid and 3,5-methoxy-4-hydroxy-cinnamic acid are commonly used matrices. MALDI is compatible with relatively simple peptide mixtures, proteins or lipids. Advantages with MALDI are that it is sensitive and can tolerate high concentrates of salts and other impurities. MALDI instruments are also robust and easy to use at low cost. MALDI can further be used for *in situ* profiling and imaging of proteins in tissues[24]. MALDI is usually coupled to a time-of-flight (TOF) analyzer. MALDI-TOF instruments (Figure 2) have become an analytical technique in proteomics for the identification of proteins separated by 2-DE [25,26].

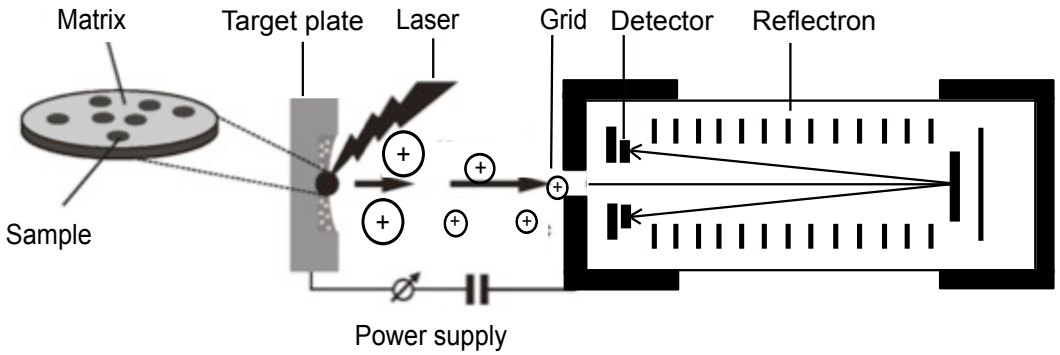


Figure 2. General principle of matrix-assisted laser desorption ionization with a time-of-flight analyzer. The peptide mixture of a protein is crystallized on a target plate with matrix, which absorbs the energy of the laser and mediates the generation of peptide ions. Ions then pass through a grid to the time-of-flight (TOF)

analyzer. A general principle is that smaller ions will reach the detector first. A reflectron in the TOF-analyzer increases the resolution of the mass spectrometer.

Electrospray ionization (ESI)

Electrospray ionization (ESI) (Figure 3) is the most widely used technique for the analysis of samples in liquid form[27]. Because it ionizes molecules directly from the liquid phase, it is compatible with traditional chromatographic separation techniques widely used in analytical chemistry. Ions released by electrospray are very stable. The ionization process is unlimited in mass[28]. These characteristics, paired with its very high ionization efficiency, are the basis for the wide use of electrospray ion sources[29].

Electrospray is an atmospheric ionization method that produces small charged droplets from the liquid medium under the influence of an electric field. In a usual electrospray, a flow of liquid is passed through a thin conducting capillary needle at high electric voltage. The potential difference is applied between the needle and the counter electrode. The analytes exist as ions in the liquid phase and the applied potential will create an accumulation of charged ions at the tip of the needle. Either the positive or the negative ions migrate to the end of the capillary, depending on polarity of the applied electric field. The high density of positively charged ions at the tip leads to the formation of a Taylor cone[30] due to repulsive forces between the positive ions. Under a sufficiently high electric field, the forces become stronger than the surface tension at the tip of the cone and a liquid mist is ejected. The mists break up in small highly charged droplets when moving towards the counter electrode[31]. Both a potential and a pressure gradient will direct the droplets towards the inlet of the mass spectrometer and a counter current gas flow facilitates solvent vaporization and prevents further ion cluster formation. As the solvent continues to evaporate, the size of the droplets will decrease until they reach a diameter-size in the nanometer range[32,33]. ESI mass spectrometers are coupled with different sets of quadrupole (Q) and time-of-flight (TOF) mass analyzers.

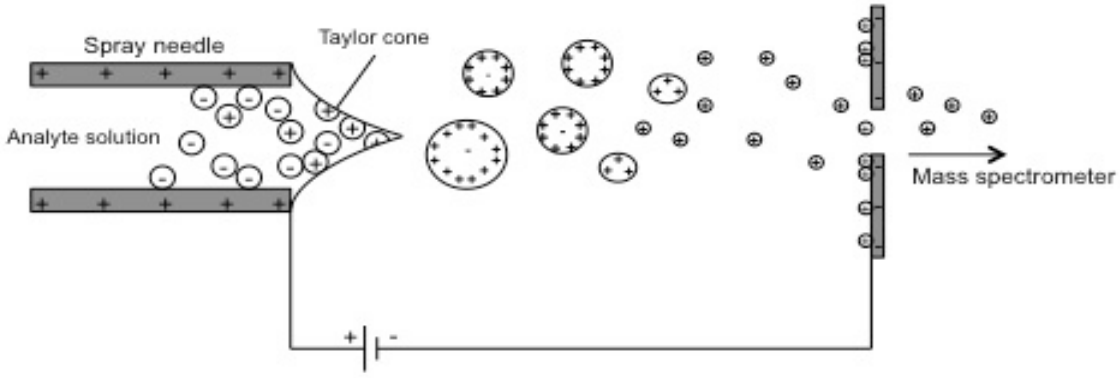


Figure 3. Representation of the electrospray ionization (ESI) process. A positive potential is applied to the capillary needle, causing positive ions in the solution to migrate towards the tip of the droplet and negative ions tend to move toward the wall of the needle. Destabilization of the meniscus leads to the formation of a Taylor cone emitting positively charged droplets into the atmosphere. Gas-phase ions are formed from the droplets as the solvent evaporates when they are accelerated towards the entrance of the mass spectrometer.

Mass analyzers

Mass analyzers separate the ions according to the m/z ratios. There are many types of mass analyzers, time-of-flight (TOF), quadrupole (Q), many types of ion traps and fourier transform ion cyclotron resonance (FTICR) devices, which can be coupled together in different manners with MALDI or ESI sources to influence the sensitivity, mass accuracy, mass range and resolution of the instrument. Only TOF and Q analyzers are discussed here.

Time-of-flight (TOF) analyzer

The TOF analyzers are usually interfaced with MALDI to perform pulsed analysis of individual samples. The analyzer measures the time of flight of the ions across the tube to the detector. The ions will be separated in the analyzer according to their m/z ratios, thus light ions will fly faster than heavy ions when they carry the same number of charges. To increase the resolution of the TOF analyzer a reflectron was introduced in 1973[34]. The reflectron is an electrostatic device inside the TOF tube, which reflects the ions 180 degrees from the original direction. The reflectron is used to compensate the differences of kinetic energy of same m/z ions. The reflectron serves to increase the amount of time (t) ions need to reach the detector while reducing the differences of kinetic energy of the same m/z ions, thereby reducing the temporal distribution Δt. Since resolution is defined by the mass of a peak divided by the

width of a peak or m/Δm (or t/Δt since m is related to t), increasing t and decreasing Δt results in higher resolution. (Figure 2).

Quadrupole (Q) analyzer

A quadrupole mass analyzer[35] contains four rods that work in pairs. Each pair is connected to a direct current (DC) voltage with a superimposed oscillating radio frequency (RF). One pair has a positive voltage while the other pair is negatively charged at a time. Specific DC/RF ratios let ions of certain m/z ratios through. Too low m/z values make the ions oscillate in the RF field with increasing amplitude until they hit a rod and are discharged. Ions with too high m/z values will also hit the rods and will be discharged. By varying the DC/RF ratio, the mass analyzer thus scans an m/z interval. By locking the DC/RF ratio only ions with a limited m/z value may pass the analyzer. An important feature of the quadrupole is that it stably guides and focuses ions, making them suitable for operation as an ion transmission guide and as a collision cell. This makes the quadrupoles excellent mass filters and the technique is used in tandem mass spectrometry. Quadrupoles are today used in many different setups and combinations of analyzers in mass spectrometers.

Clinical proteomics and mass spectrometry

Proteomics has an important role in clinical research. To understand what is really occurring in cells, tissues or organs in disease states, it would be important to study the expression profiles of proteins, because they are the functional units and regulators of cells and organs. Mass spectrometry has gradually become the primary choice of tool in clinical proteomics research. Mass spectrometers are sensitive (femto- or attomole sensitivity)[36] and can handle complex peptide mixtures. Mass spectrometers can be coupled online with various liquid phase prefractionation methods to reduce the sample loss and increase the rate of identification of proteins. MS-based proteomics has been used to identify and quantify almost the whole yeast proteome in a single run[37,38]. This same approach can be used with clinical samples. MS can identify peptides and proteins in normal cells or tissues, and do relative comparisons of protein expression with disease states using tags. In this manner it is possible to see the expression levels of proteins, what cell functions they represent, and which signaling pathways that are affected in diseases. It is possible to study post-translational

modifications in disease states and get knowledge about activation, deactivation or re-localization of proteins and how PTMs regulate biological functions and processes. Proteomics has now been accepted as an important part of systems biology. Studies on proteomic expression profiles and pathways might contribute to detection of novel drug targets or biomarkers, and may also explain why some patients react to some treatment while others do not.

Sample preparation techniques

Proper sample preparation is crucial in proteomics. Out of the genome of approximately 20,500 genes, tissue cells may express more than 10,000 genes and therefore the number of proteins expressed in cells and tissues is very large. The dynamic range is the ratio between the concentration of the least abundant and the most abundant protein in a complex mixture. The dynamic range of plasma has been measured to be more than 10 orders of magnitude in concentration difference between albumin and the rarest proteins clinically detected[39]. Two-dimensional gels can separate roughly up to 2,000 protein spots. It is usually the high abundant proteins that will be visualized in a crude sample. Large, hydrophobic and alkaline proteins (with pI >10) are difficult to detect. To be able to detect low abundant proteins an appropriate sample preparation method has to be used[40-42].

Several properties of the sample have to be taken into account before it can be prepared. First the biological origin of the sample has a significant role in how to process the sample. Cell culture and cell component studies are quite straightforward. Cell cultures contain homogeneous sub-populations that produce much material for further analysis. Biological fluid-samples can easily be obtained in standard clinical procedures. Despite this, it is important to standardize the whole procedure for sample handling from the sampling to the analysis. In contrast to cell cultures, fluid-samples are often non-homogeneous mixtures of different cell types and soluble proteins from the organism. Therefore, separation of cellular and soluble proteins is often required. Tissue samples obtained from surgery are challenging because the samples usually contain different cell types and blood. Samples need to be handled in a standardized and correct manner to remove the impurities in such a mode that the proteins remain intact. A delay in sample handling may compromise the integrity of the sample.

Depending on the localization, solubility, activity and quantity of a protein or proteins of interest, different purification, preparation and analysis approaches are needed, because a certain strategy for one set of proteins is not always compatible for another one. Samples are prepared differently to study cytosolic and membrane proteins. In projects that focus on the expression and identification of proteins, samples are treated with chaotropic agents, detergents and protease inhibitors to denature and inactivate the proteases. When protein activities are studied, it is instead important to keep the proteins intact, properly folded and as active as possible during the experiment. This limits the extraction methods available and there is a risk for continuous changes in the integrity of the sample. To be able to study low-abundant proteins, prefractionations must be carried out using chromatographic and microfluidic methods.

There are two main approaches to separate complex protein samples prior to mass spectrometric analysis: gel-based and gel-free separation methods. Gel-based methods include one- (1-DE) and two-dimensional gel electrophoresis (2-DE). One of the most significant breakthroughs in proteomics was the mass spectrometric compatibility of gel-separated proteins. Gel-separated proteins are often analyzed with matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS)[43]. In this approach separated proteins are prepared manually for offline mass spectrometric analysis. High-pressure liquid chromatography (HPLC) and other liquid chromatographic methods are the most commonly used gel-free methods. Coupling of HPLC to electrospray tandem mass spectrometry (ESI-MS/MS) has gained wide acceptance for gel-free proteome analysis. It has been used to study low abundance proteins and membrane proteins[44,45]. Compared to 2-DE, LC-based methods are more efficient for analysis of small proteins and peptides as well as highly hydrophobic proteins[46]. A new technology to separate is a microfluidic electrocapture to separate proteins and peptides according to electric and hydrodynamic forces on the proteins. A major advantage with the gel-free approaches is that protein or peptide samples can be coupled on-line with electrospray ionization mass spectrometry (ESI-MS), and therefore avoid sample loss during manual sample handling.

Two-dimensional gel electrophoresis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is a well-established method for the separation of proteins on the basis of the molecular weights[47,48]. Isoelectric focusing (IEF) is used to separate proteins according to their charge (isoelectric point, pI = pH at which the net charge of the protein/peptide is zero)[49]. Kenrick and Margolis were the first to introduce a method which combined the two methods[50]. O'Farrell developed the method resulting in high resolution 2-DE[51] (Figure 4).

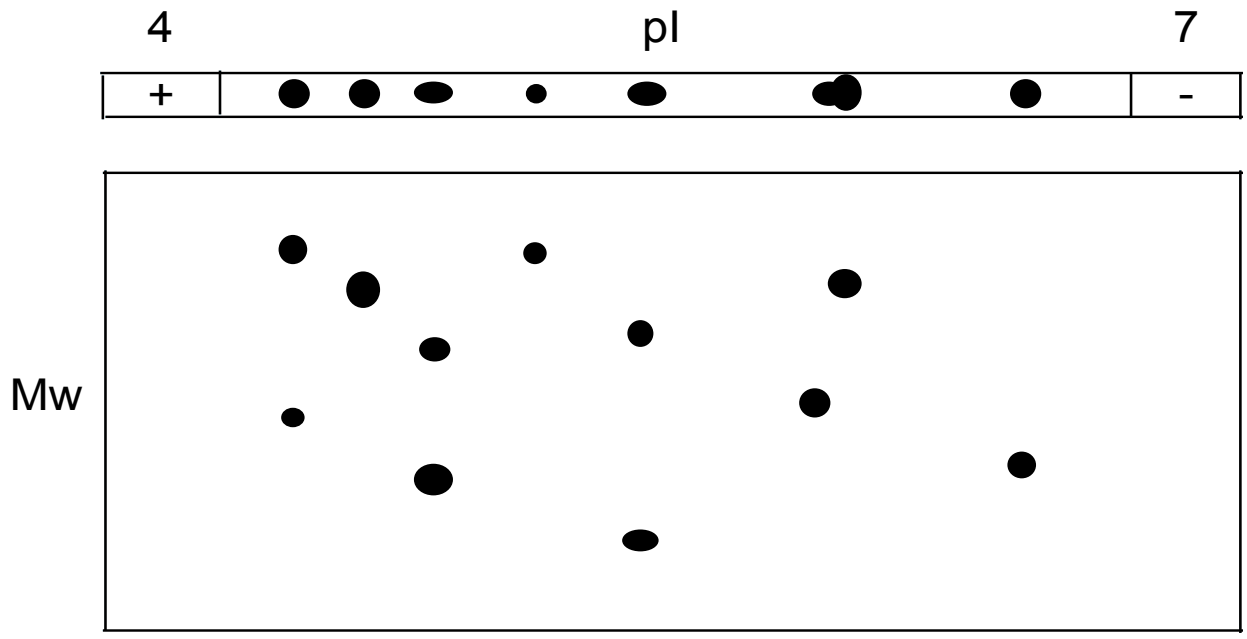


Figure 4. The general principle of 2-DE. Proteins are separated first according to their isoelectric point (pI) on an immobilized pH gradient strip. Proteins are then separated according to their molecular weight (Mw). As a result, each spot on the two-dimensional gel represents one protein.

2-DE is a powerful and widely used method for separation of complex protein mixtures extracted from cells, tissues or other biological samples. The proteins are first separated by IEF according to their pIs using immobilized pH gradient strips and then according to their molecular weights using SDS-PAGE. After the 2-D gel electrophoresis, the proteins are visualized. Ideally each protein spot on a gel would represent an individual protein. Different forms of the same proteins with different post-translational modifications are also separated. Gel spots are commonly stained with Coomassie brilliant blue (CBB) or silver. Organic solvent-based CBB-R staining has a limit detection limit in the order of 0.1 μg while the

water-based CBB-G is approximately four times more sensitive[52]. Silver stains can reach down to approximately 1 ng[53]. Silver staining methods have developed drastically during the past decade and are today compatible with mass spectrometry[54]. Approximately 2,000 proteins can be separated simultaneously with a single 2-DE and information of the pI of the protein, the apparent molecular weight, post-translational modifications and the amount of proteins of each protein can be obtained[55,56].

Two-dimensional electrophoresis is a nice method for protein separation, because it is highly reproducible, has capability to readily visualize post-translationally modified proteins as they tend to appear as distinct rows of spots, and because it can be used for quantitative expression profiling. But it also has its drawbacks. 2-DE has a limited dynamic range for detection. Housekeeping proteins and highly abundant proteins drench the low-abundant proteins in the 2-DE[55,57,58]. To improve the detection of the low abundant proteins, crude protein mixtures should be prefractionated to study the sub-proteomes. Proteins can be fractionated according to their chemical properties, post-translational modifications and subcellular location[59,60]. Another major drawback of 2-DE is the poor resolution of membrane and large hydrophobic proteins since most of them are not soluble in aqueous solution and thus do not enter the pH strips during the first dimension[61]. Other drawbacks include large amount of sample handling[56], difficulty to identify proteins with high pI values[57], high molecular weight[62] and hydrophobic properties[63].

Typically, 2-DE analysis combined with MALDI-MS mass fingerprinting results in the identification of 300-500 different proteins with many post-translational modifications and isoforms[64,65]. This represents only approximately 5-10% of a given cellular and organelle proteome. Therefore, sensitive liquid chromatographic separation methods need to be used, additionally, to explore the entire proteome. Despite the drawbacks, 2-DE is still a very widely used tool in the proteomic research to identify novel proteins and especially to do relative quantitative analyses between healthy and disease state tissue samples. This approach is e.g. extensively used in tumor proteomics[66].

High performance liquid chromatography

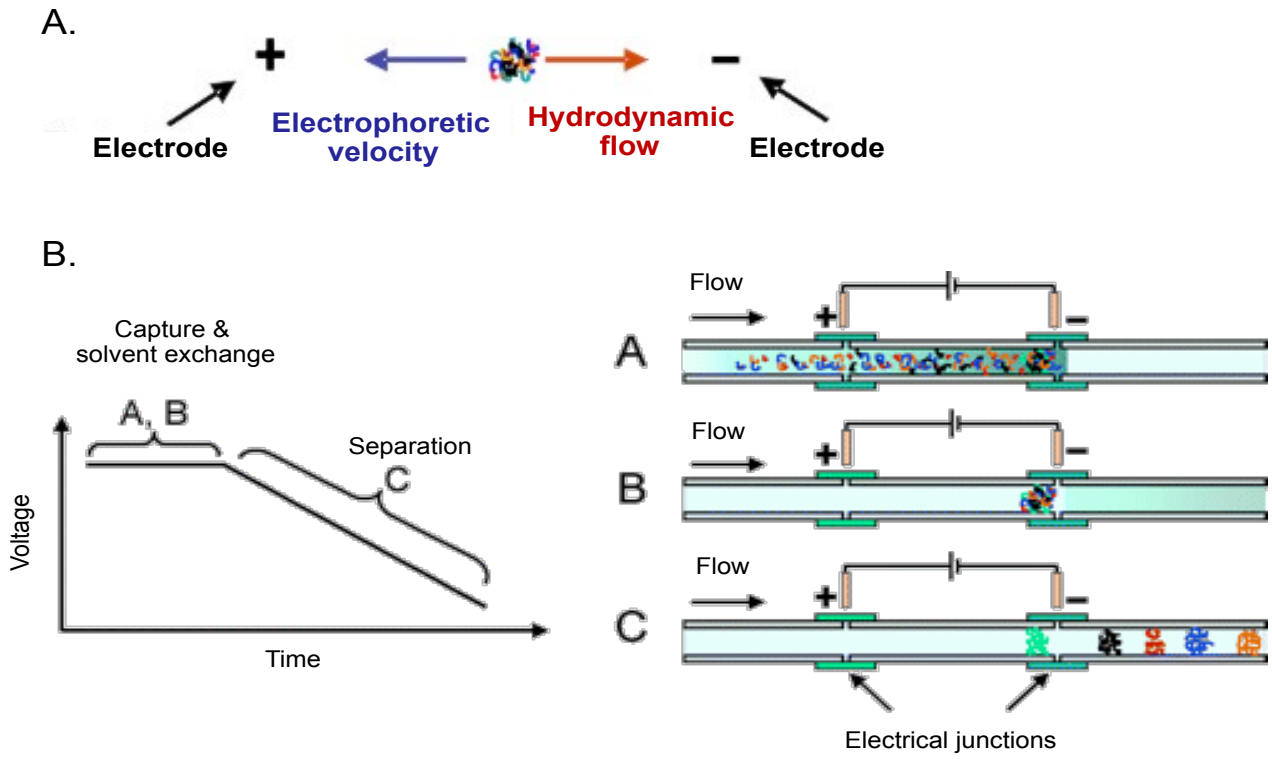
High performance liquid chromatography (HPLC) separates biological molecules based on their unique characteristics and interactions with the stationary phase of the

chromatography column. In MS-based proteomics, HPLC can be directly coupled to mass spectrometry instruments. Typically reverse phase (RP), ion exchange (IEX), hydrophilic-interaction chromatography (HILIC) and affinity chromatography are coupled with MS[67].

Reverse phase liquid chromatography (RPLC) separates peptides based on their hydrophobicity. It is probably the most widely used method of liquid chromatography. The significant advantages of RPLC are that the buffers used are compatible with electrospray ionization and that the peptide sample is desalted at the same time[21,68]. Given high resolution, efficiency, reproducibility, and mobile phase compatibility with ESI, the analytical RPLC performs the final separation of peptides prior to mass analysis[21]. The use of longer columns with smaller particle sizes and ultra high pressure has been shown to further improve the resolution, sensitivity and analysis time[69-71].

Microfluidic electrocapture technology

Microfluidics deals with the behavior, control and manipulation of fluids at the micro- or nanoscale. It involves and has its origin in many different fields. Early microanalytical methods were traced to several methods, e.g. gas chromatography (GC), liquid chromatography (LC) and capillary electrophoresis (CE)[72,73]. Microfluidics provides a multifunctional and high-throughput tool for the analysis of proteins and other biomolecules[74,75]. Microfluidic electrocapture (Figure 5) is a novel technique introduced in the early 2000 that is used in the analysis of proteins and other biomolecules by mass spectrometry[76]. It involves the use of an electrocapture cell integrated into a platform with a syringe pump, microinjector and power supply for electroimmobilization and manipulation of charged molecules. The technique has the ability to trap charged species such as proteins, peptides and DNA in a microfluidic chamber and manipulate them in different manners. It is capable of simultaneously performing several steps in protein analysis e.g. reduction, alkylation, digestion, concentration, cleanup, fractionation and separation of biological samples[75,77,78]. It has been rapidly developed in recent years, particularly in terms of its applications to proteomics, consolidating the advantages of both microfluidic and sorbent-free electric field gradient systems. Major advantages are that it can handle hydrophilic and membrane proteins and can be connected on-line with mass spectrometry to minimize the sample handling.



Modified from Astorga-Wells *et al*[78]

Figure 5. General principle of the microfluidic electrocapture technique. The electrocapture cell consists of two electrodes – the upstream electrode positively charged and the downstream electrode negatively charged. As charged molecules e.g. peptides are injected with a constant and continuous flow into the electrocapture cell, positively charged molecules tend to move towards the cathode (-) and negatively charged molecules towards the anode (+). **A.** Immobilization is achieved through the counteracting effects of hydrodynamic and electric forces. When the hydrodynamic forces of the charged molecules are greater than the electric forces, the molecules are eluted. **B.** By gradually reducing the potential of the electrocapture cell, charged molecules can be separated in solution phase compatible for subsequent mass spectrometric analyses.

Microfluidic compact disc based technology

The microfluidic compact disc based technology (Gyros AB) for protein assays was tested in our laboratory during the early phase of my thesis. The technology constituted a novel approach for a high throughput processing of protein digests integrated with MALDI-MS. Unfortunately, the company behind the development of this concept later quit supplying this technology and it was therefore not extensively used. The CD consisted of 96 individual microstructures with individual sample inlets, stationary phases depending on the particular application and a target area. A reverse phase chromatography (RPC) column stationary phase was used for ordinary protein identifications and an immobilized metal affinity (IMAC)

chromatography column for studies of phosphopeptides. The CD is divided into 6 segments with a common inlet for buffers and reagents. The CD technology uses centrifugal forces to move reagents and samples through the microstructures[79-81]. In this thesis this technology was only used in paper I and partly paper II for sample preparation of protein digests (after 2-DE) for MALDI-MS analysis.

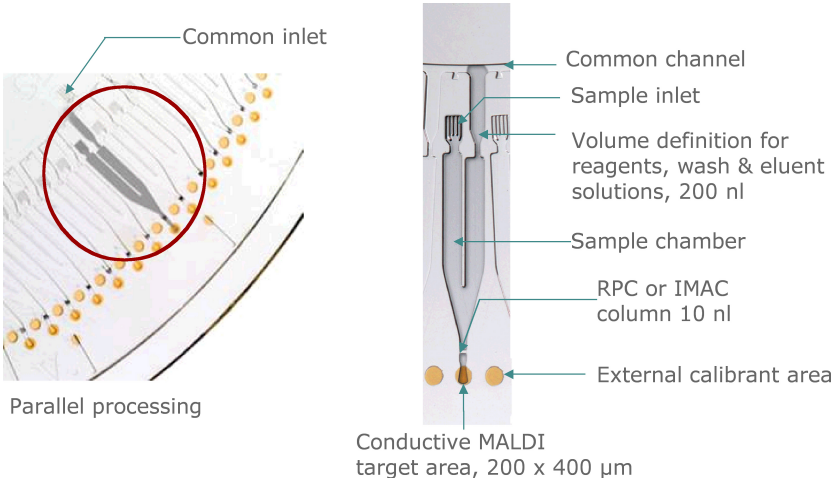


Figure 6. General principle of the microfluidic CD system. In the commercial discs previously available, there were 96 microstructures divided into 6 segments. Each segment had 16 microstructures with a common reagent inlet and a separate sample inlet. A sample of 1-5 µl was loaded to the sample inlets. After spinning the CDs, the peptides were bound to the reverse phase column. The peptides were washed and eluted with a matrix to the target area. The CDs were then cut and applied to a MALDI target plate. External calibrant areas beside the MALDI target areas could be used during the MALDI-MS analysis if internal calibration could not be used.

Bioinformatics in Proteomics

The present high-throughput screening technologies in genomics, transcriptomics and proteomics generate large amounts of data. Computational analysis, *i.e.* bioinformatics, is a key requirement for the study of large data sets. In this area of research, systems biology, a whole system rather than its components, is being studied[82]. The focus in proteomics has changed from small-scale projects towards large-scale studies and from qualitative expression patterns towards dynamics and turnover in different cell types and under different conditions. The emphasis is on the study of the global patterns in a cell, tissue or an organism, such as protein-protein interactions and protein-nucleotide interaction networks, metabolic pathways, signaling cascades networks, and gene and protein expression patterns[83]. To be able to extract the significant biological information from the large data sets,

powerful computational tools are needed. Many of the bioinformatics strategies in proteomics can be adapted from the genomics, but there are also some unique features in proteomics, for example in the identification of proteins and in the study of PTMs. Systems biology can help us to understand the functions of cells, organs and organisms, as well as the pathogeneses of disease.

Protein identification

The main goal of MS-based proteomics is accurate identification and characterization of as many proteins as possible in the samples. A sample can be one protein or a mixture of proteins. Peptide mass fingerprinting (PMF) and peptide fragmentation fingerprinting (PFF) are the two main MS-based approaches for the identification of proteins.

A peptide mass fingerprint is obtained of a protein that has been enzymatically degraded with an amino acid-specific protease such as trypsin. The sets of masses, typically obtained by MALDI-TOF for isolated proteins, are then compared to the theoretically expected tryptical fragment masses for each entry in the databases (Swiss-Prot[84], NCBI[85], TrEMBL[86]) using search engines such as ProteinProspector[87] and Mascot[88]. The proteins can be ranked according to the number of peptide matches. More sophisticated scoring algorithms take also into account the mass accuracy, the percentage of protein sequence covered, the rarity i.e. uniqueness of the peptide mass, the peptide mass specificity of a certain species, and calculate a level of confidence for the match[89,90]. The accuracy of the peptide masses affects also the specificity of the search. With MALDI-TOF mass spectrometers, mass accuracy of <0.1 Da can be achieved.

Complex mixtures of peptides are analyzed by LC-MS/MS-based systems, which create first a PMF pattern of the peptides identified and then isolate them one by one and fragment them to create the sequence, an MS/MS spectrum. Theoretical MS/MS spectra are computed from the theoretical peptide sequences and correlated with the experimental MS/MS spectra to find the most similar. There are many available databases e.g. Mascot[88], Phenyx[91] and SEQUEST[92], that can analyze and validate the complex data obtained from the peptide mixtures.

The search engines use different algorithms for the validation of protein identifications to overcome factors that might interfere with the identification process, such as the presence of

contaminants, imprecise precursor masses, spectra derived from proteins not included in the database, unexpected PTMs and sequencing errors. Many search engines use statistical tools to calculate different scores, probabilities and false discovery rates for the identification. The algorithms are complex and take into account a number of parameters such as input parameters, number of peptides and amino acid sequences found/identified, the quality of the amino acid sequences, number of unique peptides and use of target-decoy databases[93]. Target databases consist of true proteins with their peptides and sequencing. Decoy databases are constructed from reversed amino acid sequences, which do not occur in nature[94].

Normalization of data

The initial steps in the interpretation of the long lists of identified proteins is to normalize the protein names to specific identifiers because different databases may have different names and the names may change in the updated versions. Uniprot[95], Ensembl[96] and international protein index (IPI)[97] identifiers are most commonly used. There are several tools[98,99] to do it, so one does not have to do it manually. There are also tools, which can be used to compare proteomic results to related experiments performed by other groups in the past[100,101].

Gene Ontology

To describe a gene or protein and its functions by a definite vocabulary, Gene Ontology (GO)[102] was created. In this way, laboratories all around the world can categorize their results in a similar manner. In GO, genes and proteins are categorized in three GO-categories: biological process, molecular function and cellular component. The Database for Annotation, Visualization and Integrated Discovery (DAVID)[103] is a popular and free internet-based program used for GO-annotation of large datasets. However, it is important to be aware of that all genes and proteins are not annotated in GO-terms.

Pathway analysis and interaction networks

A biological pathway consists of a series of well-defined protein interactions that lead to a specific biological outcome. With pathway analysis it is possible to identify the most relevant signaling and metabolic pathways of the experimental data. Pathways are manually curated

from the published data. There are still unknown or poorly studied proteins, that are not included in any pathways. This must be taken into account during the analysis of the data. Kyoto Encyclopedia of Genes and Genomes (KEGG)[104] and Ingenuity Pathway Analysis (IPA) (www.ingenuity.com) provide good models for hundreds of pathways.

Interaction networks are based on known relationships of proteins in databases. Protein-protein interactions (PPIs) comprise both physical and functional interactions. Physical interactions include for example the formation of multiprotein complexes, such as the actin cytoskeleton network. Functional interactions include for example sharing of a common substrate in a pathway or inhibition, or activation of each other at the transcriptional level. The Human Protein Reference Database (HPRD)[105] contains approximately 40,000 manually curated human PPIs.

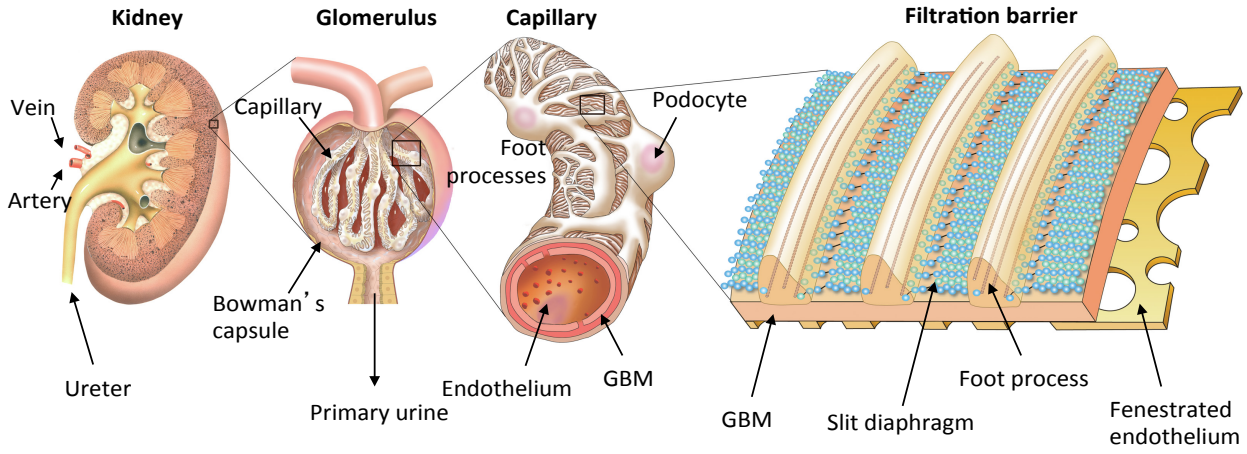
There are several tools that can be used to visualize interactions and networks. As mentioned above, KEGG and IPA databases are used for the study of pathways. IPA software can also construct networks. Another software to construct software is Cytoscape[106]. With the help of these bioinformatics tools it is possible to validate large data sets, i.e. expressed genes/proteins in normal cells or disease models by concretely seeing which components and pathways are most relevant for the experimental data. The results may also imply novel connections to pathways and processes and may be helpful in the discovery of novel biomarkers. It is important to appreciate that the tools do not give definite answers, but instead help scientists to develop testable hypotheses, which need to be evaluated in the laboratory.

Kidney filtration system

Kidney glomerulus

The kidney glomerulus is the filtration unit of the kidney (Figure 7). It consists of a tuft of capillaries surrounded by a Bowman's capsule at the proximal end of the nephron. It functions as a high-capacity molecular sieve through which plasma is filtered into the Bowman's space as primary urine. The glomerular filtration barrier is composed of glomerular fenestrated endothelial cells, a basement membrane (GBM) and an epithelial podocyte cell monolayer[107,108]. The cell body of a podocyte lies in the urinary space and

extends primary, secondary and finally fine terminal foot processes that cover the outer surface of the basement membrane in an interdigitating manner. Between the fine foot processes remains a narrow 40 nm wide slit. Adjacent foot processes are interconnected with a cell-cell junction referred to as slit diaphragm, that comprises the ultimate filter[109,110].



Modified from Tryggvason and Wartiovaara[111]

Figure 7. The general structure of the glomerular filtration barrier. There are approximately 1,000,000 glomeruli in a human kidney. A glomerulus consists of tufts of small capillaries. The capillary filter contains three layers: a fenestrated endothelium, a glomerular basement membrane (GBM) and epithelial podocytes. The podocytes have foot processes that cover the GBM. Each foot process links to the adjacent foot process through a specialized intercellular junction, the slit diaphragm, and supports the GBM through an integrin-linked adhesion mechanism. This structure serves to create maximal filtration space between cells while at the same time supporting and maintaining the GBM.

There is evidence from biochemical and physiological analyses of the glomerular filter barrier, suggesting that the glomerular filter barrier functions in a size- and charge-selective manner, but the mechanisms of the regulation of the plasma filtration are still not completely understood[107,112].

The kidney filter is a unique structure. The slit diaphragm has been demonstrated to be composed of a uniformly organized network of convoluted strands in a high-resolution electron tomography analysis[109]. Several glomerular specific proteins associated with the diaphragm (nephrin[113,114], CD2AP[115], Neph1[114], podocin[116]) have been identified and the abnormal function or absence of any of these proteins leads to proteinuria and the progression of renal failure[117-119].

Glomerular diseases and expression signatures

The glomerular filter is sensitive to damage which often leads to an irreversible disease process progressing to chronic kidney disease and renal failure. Glomerular diseases are currently classified according to etiological, clinical, morphological and immunopathological findings. Glomerular diseases can be primary, where the pathology is found in the kidney and any systemic feature is a direct consequence of the glomerular dysfunction (e.g. minimal change disease, membranous glomerular nephropathy, focal segmental glomerulosclerosis (FSGS), membranoproliferative glomerulonephritis and IgA nephropathy). In secondary glomerular diseases or injuries, the glomerulus is affected because of systemic (e.g. systemic lupus erythematosus, hypertension), vascular (vasculitis) or metabolic disorders (diabetes). There are also many known inherited glomerular diseases (e.g. Alport syndrome[120-124], Congenital nephrosis of Finnish type[124], Pierson's disease[125,126], steroid resistant nephrotic syndrome[127-130], FSGS[131-133] , familial diffuse mesangial sclerosis[134], Galloway-Mowat syndrome[135], Denys-Drash syndrome[136]).

According to clinical classifications, glomerular diseases are divided in asymptomatic urinary abnormalities, acute nephritic syndrome, rapidly progressive glomerulonephritis, nephrotic syndrome and chronic glomerulonephritis. Pathological classification involved the glomerular involvement (>50% diffuse, <50% focal), capillary involvement (>50 % general, <50% segmental), cell involvement (proliferative, non-proliferative), cell damage (necrosis, foot process effacement), mesangial involvement (expansion, depots) and site of immune complex and complement deposits (Table 1-2). Although glomerular diseases are classified based on morphology, the classification does not provide much information about etiology and the morphological diagnosis does not usually provide proper directions for therapy.

Table 1. Classification of acquired kidney diseases. This table illustrates a simplified classification of glomerular diseases based on pathology, together with the manifesting syndrome, etiology, morphological changes and deposits characteristic for the respective diseases.

Disease	Syndrome	Etiology	Morphological changes	Deposits
Diabetic glomerulosclerosis	Nephrotic syndrome	Type I, II DM	Focal segmental sclerosis, foot process effacement, mesangial expansion	-
Proliferative glomerulonephritis (GN)				
Mesangial proliferative GN	Nephritic syndrome	SLE, vasculitis, bacterial endocarditis, hepatitis B or C, HIV	Mesangial expansion	immune-complexes in the mesangium
Membranoproliferative GN	Nephrotic/nephritic syndrome		GBM thickening, double contour capillary wall, mesangial expansion	immune-complexes and complement in the subendothelial space and GBM
Focal proliferative GN (IgA)	Nephritic syndrome	unknown	GBM thickening	IgA in the mesangium
Diffuse proliferative GN	Nephritic syndrome	Autoimmune diseases, post-streptococcal GN	Obliteration of the capillary loops and sclerosis	Immune-complexes in mesangium, GBM, subendothelial and subepithelial locations
Crescentic GN (rapidly progressive GN)	Nephritic syndrome	Autoimmune diseases, post-infectious GN	cellular proliferation of parietal epithelial cells and Inflammatory cell infiltrates in the Bowman's capsule	ANCA or IgG in GBM
Non-proliferative glomerulonephritis				
Minimal change disease	Nephrotic syndrome		Effacement of podocyte foot processes	-
Focal segmental glomerulosclerosis	Nephrotic syndrome		Focal segmental sclerosis, mesangial expansion	-
Membranous nephropathy	Nephrotic syndrome		GBM thickening	Immune complexes in subepithelial layer
Hereditary nephritis				
Alport's syndrome	Nephritic syndrome	Mutations in COL4A3, COL4A4 or COL4A5	Thickening, thinning and splitting of GBM	-
Chronic glomerulonephritis			Extensive tubular atrophy, interstitial fibrosis, mononuclear cell infiltration, vascular changes	May be present

Table 2. Inherited glomerular diseases and their clinical manifestations.

Disease	Gene	Protein	Trait	Clinical manifestation
Alport's syndrome	*COL4A3, COL4A4 or COL4A5*	Type IV collagen	Mainly autosomal recessive, X-linked (COL4A5)	Nephritic syndrome at adolescence or early adulthood. Thickening, thinning and splitting of GBM
Pierson's Disease	*LAMB2*	Laminin β2	Autosomal recessive	Lethal nephrotic syndrome at birth, diffuse mesangial sclerosis
Congenital nephrotic syndrome of the Finnish type (CNF)	*NPHS1*	Nephrin	Autosomal recessive	Nephrotic syndrome after birth, foot process effacement and absence of slit diaphragm.
Steroid-resistant nephrotic syndrome	*NPHS2*	Podocin	Autosomal recessive	Usually onset at early childhood, steroid resistant, rapid progression, Nephrotic syndrome, FSGS
Focal segmental glomerulosclerosis	*ACTN4*	α-actinin 4	Autosomal dominant	Nephrotic syndrome at adolescence, slow progression
Focal segmental glomerulosclerosis	*TRPC6*	Transient receptor potential cation channel 6	Autosomal dominant	Nephrotic syndrome at adulthood
Focal segmental glomerulosclerosis	*CD2AP*	CD2-associated protein	Autosomal recessive	Nephropathy at infancy, rapid progresion
Familial diffuse mesangial sclerosis	*PLCE1*	Phospholipase C, epsilon	Autosomal recessive	Nephrotic syndrome at early childhood
Galloway-Mowat syndrome	*??*			Nephrotic syndrome at infancy of age, FS
Denys-Drash syndrome	*WT1*	Wilms tumor oncogene 1	Autosomal recessive	Nephropathy at infancy, rapid progression, diffuse mesagial sclerosis

Glomerulonephritis encompasses a wide range of immune-mediated disorders that cause inflammation in the glomerulus. The pathomechanisms of glomerulonephritides are complex and not understood. First, many types of insults can initiate glomerulonephritis e.g. endogenous processes such as autoimmune disorders, cancers, structural abnormalities of glomeruli, and exogenous factors such as infectious organisms, toxins and drugs. Secondly, there are also individual differences in the susceptibility to glomerulonephritis. Thirdly, there are complex interactions between different inflammatory mediators (antibodies, chemokines, complement, growth factors) and cells (macrophages, T-cells) occurring at the same time. A final cause of complexity is the progressive nature of many of the kidney diseases, with factors such as hypertension and proteinuria promoting the continuing glomerular damage.

Podocyte injury and effacement of the foot processes results in excessive proteinuria, and therefore podocytes are frequently the primary targets of interest in studies on glomerular diseases. Alport syndrome, a genetic disorder of type IV collagen in the basal lamina, only results in modest proteinuria. It is believed that the development of proteinuria with systematic diseases is probably a multistep process. The endothelium is first damaged to produce more components of the GBM as the negatively charged glycocalyx also disrupts and loses its ability to repulse the negatively charged proteins. The organization of actin skeleton of the podocytes is then slowly disrupted and podocyte injury and foot process effacement occurs, which results in proteinuria. Even though the structure of the glomerular filter is quite well known, the molecular understanding of the pathomechanisms that lead to proteinuria are still poorly understood and therefore we are still lacking specific therapies for kidney diseases.

In the clinic, diagnosis of the glomerular disease is important because the outcome and treatment differs with different types. Many cases of glomerulonephritis result in mild asymptomatic illness that is not recognized by the patients. They might have microscopic hematuria or subnephrotic range proteinuria with normal renal function. Signs of nephritic syndrome include recent onset of hematuria, proteinuria, renal impairment, and salt and water retention causing hypertension. The characteristics for nephrotic syndrome are proteinuria (>3.5g/1.73m^2/24h), hypoalbuminemia, edema and hyperlipidemia. Chronic glomerulonephritis includes persistent proteinuria with or without hematuria and slowly progressive impairment of renal function. Kidney disease diagnostics involve laboratory tests

and biopsy. Typical laboratory diagnostics of renal diseases include: blood pressure, S/P-creatinine, S/P-albumin, urine analysis, dU-albumin, cystatin-C and S/P-urea.

Renal function and the stage of chronic kidney disease are determined by the glomerular filtration rate (GFR). It describes the flow rate of filtered fluid through the kidney. The Cockcroft-Gault equation is the most commonly used formula used to calculated the GFR when serum/plasma creatinine is known. GFR for a healthy person is 100-140 ml/min. Chronic kidney disease patients with GFR < 15 ml/min are subjected to kidney replacement therapy.

Glomerular expression profiling, both with regard to protein and RNA expression are limited. No glomerular proteome analyses existed at the beginning of the present project, and only limited amount of information had been reported on glomerular transcriptomes (Paper V).

AIMS OF THE THESIS

The aims of my thesis have evolved during my time as a PhD student. The initial goal was limited to characterize the glomerular proteomes in healthy and diseased mice to provide new knowledge about the protein composition of glomeruli and pathomechanisms in glomerular diseases.

As my studies were elongated in time by my parallel work with my medical education, the aims were extended to get a more general view of proteomics and to include an insight into new technical methodologies in the field, still having the emphasis on the kidney glomerulus. Paper II was a straightforward proteomic study of the glomerular proteome. Papers I, III and IV involved the participation in the development of new technologies in proteomics and finally I addressed my interest on the importance of proteome and transcriptome studies of glomerular disease from the clinical point of view (Paper V).

Specific aims of the study have been:

- Characterization of endothelial proteins in human saphenous vein endothelial cells using a novel high throughput CD-based technology for sample preparation prior to MALDI-MS (Paper I)

- Characterization of mouse glomerular proteins by 2-D gel electrophoresis and mass spectrometry (Paper II)

- Development of a novel microfluidic electrocapture technology for separation of peptides (Paper III)

- Characterization of glomerular membrane proteins using electrocapture and mass spectrometry, a novel two-dimensional system for proteomic analysis (Paper IV)

- Meta-analysis of current data on pathways and networks in glomerular disease (Paper V)

MATERIALS AND METHODS

This chapter includes a brief overview of the methods used in this thesis. Detailed description of the materials and methods can be found in the corresponding papers.

Sample Preparation

Isolation of glomeruli (Paper II, IV)

Kidney glomeruli were isolated from mice according to Takemoto et al[137]. Briefly, mice were anesthetized by an intraperitoneal injection of 20 µL/g of 2% (w:v) Avertin [100% stock (5 g 2,2,2-tribromoethyl in 5 ml tertiary amyl alcohol) diluted in sterile H_2O]. The mice were perfused with HBSS containing 8 x 10^7 magnetic Dynabeads® (Dynal) through the left ventricle via vena cava superior. The kidneys were removed, minced into small pieces and digested with DNase I for 30 min at 37 °C. The digested tissue was pressed though a 100 µm cell strainer and washed two times. Finally the glomeruli were collected with a magnetic particle collector (Dynal) and washed. The glomeruli were counted and the purity was controlled under a light microscope.

Sample preparation for 2-DE (Paper I, II)

Human vena saphena endothelial cells (Paper I)

Human endothelial cells from vena saphena magna (Paper I) were isolated during coronary bypass operations and then grown under laboratory conditions. Protein samples were prepared according to the BioRad ReadyPrep Kit. Isolated human vein endothelial cells were suspended in 50 mM Tris-buffer and disrupted with 6x10s sonication bursts. The sample was centrifuged for 15 min at 13,000g and the supernatant was collected. The pellet was dissolved in 8 M urea, 4% CHAPS, 40 mM Tris, 0.2% w/v Biolyte 3/10 and 2 mM tributylphosphine. Protein concentrations were measured using the Bradford assay[138]. The protein sample of the supernatant was used for the 2-DE analysis.

Mouse glomeruli (Paper II)

Isolated glomeruli (Paper II) were suspended in water, freeze-thawed four times in liquid nitrogen and freeze-dried under vacuum. Dried samples were solubilized in a lysis buffer containing 9M urea, 65 mM DTT, 2% CHAPS, 0.5% NP-40, 35 mM Tris, protease inhibitor cocktail tablet (Roche) in a shaker for 4 h. In this manner, almost all proteins were dissolved in the same sample.

Sample Preparation for EC-based studies

Protein stock solution (Paper III)

Protein stock solutions (Paper III) of standard proteins were prepared at concentration of 1-5 mg/ml of 10 proteins. Carbamidomethylation of cysteine residues was carried out by incubation of 25 µl of protein solution with 5 µl of 45 mM DTT at 50 °C for 15 min and subsequent incubation with 5 µl of 100 mM IAA at room temperature for 15 min. After alkylation, reagents were diluted by addition of Tris-HCl (64 µl of 40 mM, pH 8), and digested overnight at 37 °C by addition of 1 µl of trypsin (1 µg/µl in 50 mM acetic acid). Resulting digests were stored at –20 °C.

Membrane proteins from mouse kidney (Paper III)

Tissue from mouse kidney was washed in 500 µl of 12.5 mM NH_4HCO_3 and incubated on ice for 30 min. The sample was then treated by five freeze/thaw cycles in liquid nitrogen, sonicated 4 times for 10s and centrifuged for 5 min at 2,000 g, after which the supernatant was discarded. The pellet was resuspended in 500 µl of 12.5 mM NH_4HCO_3 and centrifuged at 6,000 g for 10 min. The supernatant was transferred to an ultracentrifuge tube and centrifuged at 80,000 g for 30 min (Beckman Ti 70.1). The supernatant was discarded and the pellet dissolved in 25 mM NH_4HCO_3. The resulting sample was ultracentrifuged at 100,000 g for 1 h. The supernatant was discarded and the pellet suspended in 1.5 ml of 100 mM Na_2CO_3, pH 11.3 and centrifuged at 13,000 rpm at 4 °C. The pellet was solubilized in 50 mM NH_4HCO_3, pH 7.9, 1% CHAPS. For digestion, 1 µl of trypsin (1 µg/µl in 50 mM acetic acid) was added to 100 µl of sample.

Membrane proteins from mouse glomeruli (Paper IV)

Isolated glomeruli were dissolved in 30 mM HEPES (pH 7.8) with protease inhibitors (Roche). Glomeruli were disrupted by 6 cycles of freeze/thawing in liquid nitrogen. The sample was centrifuged at 13,000g. The pellet was dissolved in 65 mM HEPES, 2% CHAPS and protease inhibitor tablet (Roche). The proteins were reduced in 10 mM DTT ad 56 °C, alkylated in 35 mM IAA and digested with trypsin (0.03 µg/µl) overnight.

Two-dimensional gel electrophoresis (Paper I, II)

In paper I, IEF was carried out on 13 cm IPG strips (Amersham Biosciences). The strips were rehydrated overnight. A sample of 10 µg was loaded for Sypro Ruby® staining and 100 µg for CBB staining. It was carried out with Multiphor II (Amersham Biosciences) for a total of 57,400 Vh. The SDS-PAGE was performed on 10 % gels and stained.

In paper II, IEF was performed on 17 cm IPG strips (BioRad) with pI 4-7 and 3-10. The strips were actively rehydrated for 12 h. 75 µg was loaded for silver stained gels and 500 µg for CBB stained gels. The focusing was carried out in an IEF cell (BioRad) with a total of 35,000 Vh for pI 4-7 strips and 51,000 Vh for pI 3-10 strips. The second dimension was performed on 8-16% gradient SDS-PAGE gels with Ettan™ DALTSix.

In-gel digestion (Paper I, II)

In-gel digestion was performed in an automated MassPrep station (Micromass/Waters). Gel pieces were destained twice in 100 µL 50 mM Ambic / 50% (v/v) acetonitrile at 40°C for 10 min. Proteins were reduced with 10mM DTT in 100 mM Ambic for 30 min and alkylated with 55 mM iodoacetamide (IAA) in 100 mM Ambic for 20 min. Proteins were digested with trypsin (12 ng/µL solution in 50 mM Ambic) for 4.5 h at 40°C. The peptides were extracted with 30 µL 5% formic acid / 2% acetonitrile followed by 24 µL 2.5% formic acid / 50% acetonitrile.

Desalting and concentration of peptide samples prior to MS (Paper I, II)

Peptide samples were desalted and concentrated either manually using ZipTips (Millipore) or with the CD-based microfluidic technology (Gyros AB) according to similar procedures. The C_{18}-RPC column was activated with 50% AcN/0.1% TFA solution, ZipTips were then equilibrated with 0.1% TFA (the CD-based system skipped this step). 10 μL of sample was loaded with the ZipTips and 1 μL with the CD-system. Peptides were wash with 0.1% TFA. Finally peptides were eluted with 5-10 μL 75% AcN/0.1% TFA. Aliquots of 1 μL sample was mixed with 1:1 (v/v) of saturated α-cyano-hydroxy-cinnamic acid matrix in in 60% AcN/0.1% TFA. In the CD-based method the peptides were eluted with 50% AcN/0.1% TFA mixed with the matrix.

Mass Spectrometry

MALDI-TOF Voyager DE-Pro (paper I, II, III)

MALDI-TOF Voyager DE-Pro (Applied Biosystems) and MALDI-Q-TOF Ultima (Waters) was used for PMF. The mass spectra were obtained using a Voyager DE-PRO MALDI-TOF mass spectrometer (Applied Biosystems) operated in the positive ion mode. The reflector mode was used at 20 kV accelerating voltage, 74.5 % grid voltage, 0.005 % guide wire and 200 ns delayed extraction. Calibration was performed internally with the tryptic peptides of the samples (Paper I, II). In paper III the calibration was performed with Sequazyme Peptide Mass Standard (Applied Biosystems) with the α-cyano matrix. Identification of proteins was performed using the ProteinProspector MS-Fit and Mascot search engines.

MALDI-Q-TOF Ultima (paper II)

The MALDI-Q-TOF Ultima (Micromass) instrument was calibrated with a polyethylene glycol standard between 80–2500 *m/z*, and was operated in single reflector mode and in the positive ion mode. The MS software, MassLynx Version 3.4, was used to create a peak list, and a macro MaxEnt3 (800-3000 Da, single charged ions, peak width 0.1 Da) was applied for each list before submission into the Mascot search engine.

LC-API Q-TOF Ultima (Paper II, IV)

An LC-API Q-TOF Ultima tandem mass spectrometer (Waters) was utilized for the analysis of amino acid sequences of samples that could not be identified with MALDI-MS. Prior to LC-separation, digests were desalted with an Atlantis dC18 5-µm NanoEase Trap column using 20 µL of 0.1% formic acid. Peptides were then separated with an analytical column (Waters Atlantis C_{18}, 3 µm, 100 Å, 75 µm i.d x 15 cm) and the peptides were eluted with a linear gradient of 5-80% acetonitrile in 0.1% formic acid (Solvent A: 5% AcN/0.1% TFA, Solvent B 95% AcN/0.1% TFA) for 42 min at 200 nL/min (Paper II) and for 45 min at 200 nL/min (Paper IV). Peptides were introduced into the mass spectrometer using a PicoTip Emitter (SilicaTip, New Objective) and data-dependent acquisition (DDA) was employed over a mass range of 300-1600 m/z (Paper II) and 300-2,000 m/z (Paper IV). Data analysis was performed using ProteinLynx Global SERVER 2.2.3 software (PLGS 2.2.3, Waters) and MassLynx 4.1 peptide sequence software. Mascot (Paper II) and Phenyx (Paper IV) search engines were used for protein identification.

Microfluidic electrocapture (Paper III, IV)

The microfluidic electrocapture was manufactured using a piece of poly(ether-ether-ketone) (PEEK) tubing (512 µm o.d., 127 µm i.d, Upchurch Scientific). Two small openings 2.6 cm apart were made and covered with a conductive tubular cation-selective poly(tetrafluoroethylenesulfonate) membrane (Permapure inc) with dimesions (330 µm i.d. and 610 µm o.d.). The junctions were placed into separate electrode chambers made from 500 µL Eppendorf tubes (Eppendorf) and filled with 40 mM Tris-HCl (pH 8). Electrodes of platinum wire were placed into the tubes. The anode was connected to a high voltage power supply and the cathode was grounded. A syringe pump (Harard Apparatus) equipped with a 100 µL gas-tight syringe (Hamilton), that was filled with 40 mM Tris-HCl (pH 8) provided a continuous hydrodynamic flow of 0.2 µL/min. The syringe pump was connected with a microinjector with a 1 µL (Paper III) or 4 µL (Paper IV) internal loop. Using this system the sample manipulation was minimized.

In paper III, the anode was situated on the downstream and the cathode on the upstream gap of the cell. The initial voltage of the electrocapture was 139 V/cm, and the flow rate was 0.2 µL/min. After sample injection, the peptides were captured for 45 min (Paper III). The voltage

was the reduced with 5-10V every minute, the samples were collected and analyzed with MALDI-TOF Voyager DE-Pro. For a single protein of 200 fmol phosphorylase b, the initial voltage was 98 V/cm and was reduced 9-15 volts in every step.

In paper IV, the anode was situated on the upstream and the cathode on the downstream gap of the cell. By this way, the EC captured best negatively charged molecules. The initial voltage was set to 200 V, the continuous flow rate at 0.2 μL/min. Peptides were captured for 40 min and then the potential was reduced by 10 V every 30 min. Eluted fractions were collected for 15 min, followed by a 15 min washing period. 3 μL of 0.3% formic acid was added before the analysis of the samples with the LC-MS/MS system.

Meta-analysis of glomerular proteomes and transcriptome studies (Paper V)

The aim was to compile all reported proteomic and transcriptomic data for healthy glomeruli, and glomeruli isolated from animal models and human patients with different glomerular diseases, in order to get insight into expression signatures. It can be anticipated that different types of glomerular disease processes exhibit disease specific expression patterns, so one may be able to get new understanding of the pathomechanisms of different diseases and dentify possible biomarker candidates. All data were obtained from previously published articles and public databases. Results of three glomerular proteome studies were compiled and the comparison of the glomerular disease transcriptome analyses was done for six different disease models involving both rodent models and human patients. The annotations of the proteins of the three proteome-studies were updated using Uniprot (www.uniprot.org) and the international protein index (IPI) databases (www.ebi.ac.uk/IPI/IPIhuman.html). The Affymetrix microarray data was updated with the latest version of NetAffx (http://www.affymetrix.com/analysis/index.affx). The comparative analyses were performed using Microsoft Excel. The glomerular proteome was compared with a glomerulus-enriched transcriptome[139] to show the current state of large-scale proteome analyses in glomerular research. The differentially regulated genes of the transcriptome analyses were compared with each other in order to elucidate the possible similarities between the diseases.

The data were analyzed with the IPA version 9.0 (the Ingenuity® Systems, www.ingenuity.com). IPA is a widely utilized comprehensive database and software based on Ingenuity Pathway Knowledge Base (IPKB)[140] involving, e.g. KEGG, GO and EntrezGene. The significance of the canonical pathways is evaluated by IPA in two ways: 1) A ratio of the number of molecules from the data set that map to the pathway divided by the total number of molecules that map to the canonical pathway is displayed. 2) Fisher's exact test was used to calculate a p-value determining the probability that the association between the genes in the dataset and the canonical pathway is explained by chance alone. The networks are ranked using a score, which is the negative logarithm of a p-value. The p-value indicates the likelihood of the proteins of interest being found together by random chance in a common network.

RESULTS AND DISCUSSION

Identification of endothelial proteins by MALDI-MS using a compact disc microfluidic system (Paper I)

The goal in this project was to identify proteins of cultured endothelial cells and to test a novel high-throughput technology, the CD-based microfluidic system, used for sample preparation prior to MS. This project consisted of three parts: 1) Identification of proteins from CBB-stained gels, 2) comparison of use of conventional C_{18} ZipTips and the CD-based RPC system and finally, 3) testing the CD-based technology for sample preparation of low concentration protein digests from Sypro Ruby stained gels.

Cytosolic proteins extracted from human vena saphena magna endothelial cells were separated with 2-DE using a pI gradient of 4-7. From the total of 297 protein spots excised, 116 different proteins were identified, meaning that several of the proteins were represented by more than one spot. Many of the proteins identified had been reported before, but we could identify several interesting previously unreported proteins. Eukaryotic translation initiation factor 5A (eIF5A) is the only cellular protein that contains the unusual amino acid hypusine, that is formed by a series of PTMs of a lysine residue. Eif5A is an mRNA-binding protein and is believed to promote translation and the formation of the first peptide bond in protein synthesis[141]. The protein is involved in cell growth, differentiation and p53 regulated apoptosis[142]. There are two different isoforms of eIF5A, which have a sequence identity of 84%. EIF5A-1 is expressed equally in all tissues, but EIF5A-2 has some tissue specificity[143]. EIF5A has been shown to be up-regulated in many cancer cells and angiogenesis[144]. Chen et al showed that lung adenocarcinoma patients with elevated levels of eIF5A have decreased survival tendency[145]. Therefore EIF5A could be used as a prognostic marker of certain cancers.

Forty-eight of the 297 proteins were selected for the comparison analysis between the C_{18} ZipTips and the RPC in the CD-based system. Sequence coverage between three different experiments (manual ZipTiping, MassPrep station ZipTiping, duplicate samples for the CD-based method to control the reproducibility) was combined and compared. The duplicates of the CD-treated samples showed high reproducibility with a sequence coverage difference of 8 ± 8%.

The sequence coverage was better in 32 of the spots prepared with the CD-based technology, whereas in 7 samples the ZipTips resulted in better coverage and in 5 cases the manual ZipTips yielded a higher coverage. Three proteins were not identified and in one spot there was major differences between the methods. Overall the differences between the methods were not that significant, but manual ZipTiping is very time-consuming and therefore technologies to speed up this step are needed.

Finally the sensitivity of the method analyzed by running 2-DE loaded with only 10 µg of protein and stained with Sypro Ruby was tested. Sixteen spots were excised and after digestion, prepared either with manual ZipTiping or the CD technology. For the CD preparation sample volumes of 1 µL, 5 µL and 10 µL were used. The CD system yielded a result for 12 samples (sequence coverage ≥ 20%), the ZipTips only for five samples. The coverage was better in 11 of the samples prepared with the CD-system. The 5 µL sample improved the sequence coverage significantly compared with the 1 µL, but the 5 µL and 10 µL experiments did not show any differences. Therefore Sypro Ruby staining is compatible with subsequent routine MS identification.

Glomerulus proteome analysis with two-dimensional gel electrophoresis and mass spectrometry (Paper II)

In Paper II, we describe analysis of the proteome of healthy mouse glomeruli using 2-DE and MS. The purity of the isolated glomeruli was controlled with a light microscope. The glomeruli appeared to be highly pure and devoid of Bowman's capsule. The number of glomeruli obtained from one mouse was approximately 20,000. Ten mice yielded approximately 1 mg of protein.

A total of 768 protein spots were excised from three gels: Two CBB stained gels with pI 4-7 and 3-10 and a silver stained gel with pI 4-7. The PMF analysis was performed with MALDI-TOF MS for each spot, using a minimum of four matching peptides and sequence coverage of 12%. Unidentified proteins were analyzed with LC-MS/MS. A total of 414 protein spots could be identified, which represented 232 different proteins. A total of 72 proteins were identified only by silver staining.

Most of the identified proteins are ubiquitous, high-abundant housekeeping proteins whereas

several are cytoskeletal structural proteins highly expressed in podocytes (actin, F- actin, vimentin, α-tubulin, β-tubulin), and consistent with the human glomerular proteome analysis by Yoshida *et al*[146]. In their study from 1,559 spots, 347 protein spots representing 212 proteins were identified. They were not able to identify any proteins expressed specifically in the kidney podocyte slit diaphragm, but we could identify α-actinin-4 and nephrin, which are known to play important roles in the glomerular filtration barrier and its diseases[131,147].

Surprisingly, only 53 out of the 232 identified proteins had been detected in other glomerular proteome studies[146,148]. This result could be due to many factors, the quality of tissue preparation, sample handling and the choice of methods and technology used. Altogether, our result demonstrates the value of repeated proteome analyses using different methodologies.

Comparison of our results with previously conducted proteome studies of endothelial [149] and mesangial cells[150] revealed that these proteome studies detect many of the housekeeping proteins shared between similar tissues. Interestingly we noted that an independent glomerular-specific cDNA library analysis[151] could not identify all of these proteins. This shows that multiple approaches are indispensable to complement each other in final interpretations of functional relationships.

Miyamoto *et al*. analyzed the human glomerular proteome[152] using 1-D (SDS-PAGE) and 2-D (solution-phase IEF in combination with SDS-PAGE) prefractionation prior to the shotgun analysis with LC–MS/MS. They could identify 2,966 distinct proteins. This fact proves that prefrationation of the sample before application to MS analyses is crucial. Due to drawbacks of 2-DE such as low dynamic range, large amount of sample handling and poor resolution of membrane, hydrophobic and basic proteins, novel methods to overcome these drawbacks are needed.

Microfluidic electrocapture for separation of peptides (Paper III)

As mentioned before, separations are needed to reduce the complexity of biological samples. Liquid phase separation techniques are attractive because of their compatibility with MS analysis. The theoretical possibility to couple the liquid phase separation techniques online with MS is fascinating and important to decrease the steps of sample handling. In this study, the potential of the electrocapture device as a separation tool for peptides before MALDI-TOF-MS analysis was investigated. The separation process with the electrocapture device is described in Figure 5. The tryptic peptide mixture of 10 proteins (1 μL) was injected into the system at a flow rate of 0.2 μL/min, while applying an electric field of 139 V/cm. Peptides were captured for 45 min whereupon the electric field was reduced stepwise and each fraction analyzed with MALDI-TOF-MS.

In the MS-analysis we could see different profiles of peptides indicating a successful fractionation of the peptides. Positively charged peptides with pI 9.8-11.8 were observed in the first fractions, at high electric field strengths. On the contrary negatively charged peptides with pI 3.8-7.8 were eluted at lower electric field strengths (Table 1). The prefractionation with EC improved the identification rate by 14 peptides compared with a crude sample analyzed with MS (Figure 8). The ionization by MALDI is a competitive process, where molecules at high concentrations suppress the ionization of molecules at lower concentrations.

In this study, we also explored the possibility to carry out a combined online cleanup and separation of a peptide mixture obtained by a shotgun digestion of membrane proteins of rat kidney. Detergents, such as CHAPS, are used to solubilize membrane and hydrophilic proteins. But CHAPS has to be removed from the peptide digests because it interferes with the ionization by MALDI-MS. In the experiment, the peptide digest was captured for 1 h in the cell and washed extensively. We found that the presence of acetonitrile was critical for the removal of CHAPS, probably due to the necessity to exchange the detergent attached to the hydrophobic peptides for other molecules of low polarity. After the cleanup and separation, a nice separation of peptides was obtained. In this study we showed a novel strategy for peptide separation and online sample cleanup of a complex biological sample. This study demonstrates the great potential of the microfluidic technique for the analysis of protein samples.

Table 3. Molecular properties of the identified tryptic peptides separated with electrocapture and identified with MALDI-MS. Peptides positively charged with high pI were eluted first at higher electric fields while negatively charged peptides were eluted at lower electric fields.

Electric field (V/cm)	Peptide sequence	m/z	pI	q/(M2/3)
130	TGPNLHGLFGR (cytochrome c)	1168.62	11.8	+9.0 × 10-3
92	AVPYPQR (casein)	830.45	9.8	+11 × 10-3
	SHHWGGYGK (carbonic anhydrase)	1013	9.8	+9.9 × 10-3
	KTGQAPGFTYTDANK (cytochrome c)	1598.76	9.8	+9.1 × 10-3
	SISIVGSYVGNR (ADH)	1251.67	9.8	+8.6 × 10-3
67	HNGPEHWHK (carbonic anhydrase)	1141.53	8	0
	DYYFALAHTVR (phosphorylase b)	1355.63	7.8	−4.1 × 10-4
	HGLDNYR (lysozyme)	874.4	7.8	−5.5 × 10-4
	IVSDGNGMNAWVAWR (lysozyme)	1675.8	6.7	−7.1 × 10-4
	FESNFNTQATNR (lysozyme)	1428.65	6.9	−7.9 × 10-4
	DGPLTGTYR (carbonic anhydrase)	979.48	6.8	−10 × 10-4
	GTDVQAWIR (lysozyme)	1045.54	6.7	−19 × 10-4
62	LHVDPENFR (hemoglobin)	1126.5	5.1	−92 × 10-4
	SIGGEVFIDFTK (ADH)	1312.68	4	−92 × 10-4
	EALDFFAR (ADH)	968.5	4	−120 × 10-4
	VEADIAGHGQEVLIR (myoglobin)	1606.85	4.3	−150 × 10-4
39	VNVDEVGGEALER (hemoglobin)	1386.44	3.8	−233 × 10-4

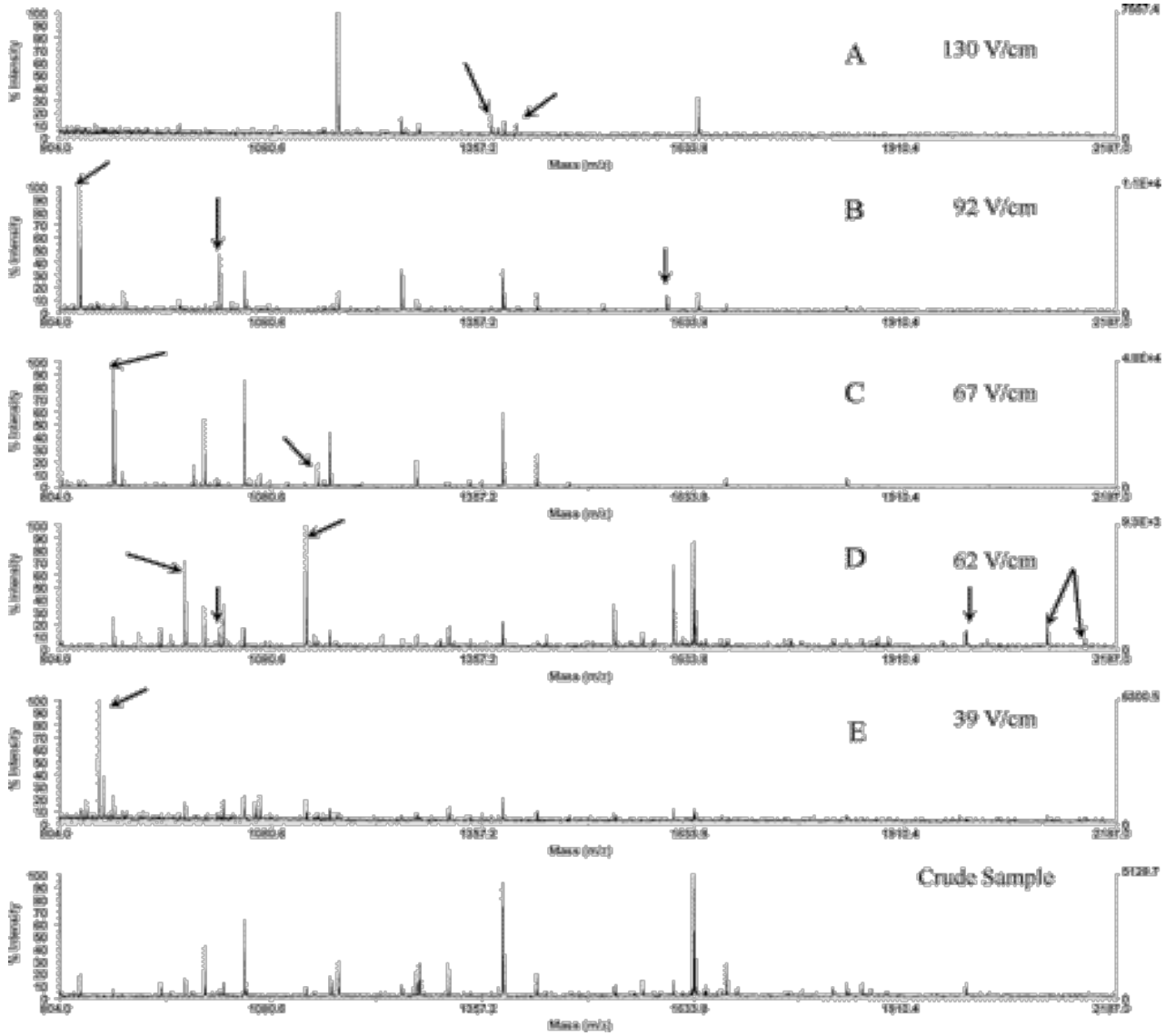


Figure 8. MALDI-MS spectra of tryptic peptides separated by electrocapture. The figure illustrates the MALDI-MS spectra of tryptic peptides of 10 proteins after prefractionation with microfluidic electrocapture. Arrows in the electrocapturegrams indicate the peptides, which were not detected by MALDI-MS in the crude sample.

Membrane protein identifications by mass spectrometry using electrocapture-based separation as part of a 2-D fractionation system (Paper IV)

In the previous paper, the EC device was shown to be applicaple for the separation of complex mixtures of peptides. In this study, we tested whether the EC system could also be integrated into a 2-D system. We, therefore, designed such a two-dimensional liquid-phase separation technique possible to use for membrane protein identification by mass spectrometry. The first dimension involved the separation of tryptic peptides with EC. The fractions were collected and injected to the RPLC-MS/MS system for the second dimension separation and analysis of the peptides. This method was controlled with a 1-D system of LC-MS/MS, injecting 3 µL of membrane protein digest. For the 2-D system, 4 µL of each fraction was injected into the LC-MS/MS system. The initial voltage of the EC was 200 V, the capture time was 40 min and the potential was reduced at steps of 10 V. Fractions were collected for 15 min and subsequently washed for 15 min before the next step.

Most of the peptides were eluted between the first (200 V) and fourth (160 V) fractions (Table 2). 74 % of the peptides were eluted in a single fraction and 17% were eluted in the adjacent fractions (Table 3). Most of the peptides, which were eluted in more than one fraction, were present in higher concentrations. We could see in the MS spectrum that the signal intensity of the peptide was higher in one fraction and lower in adjacent fractions (Figure 9), an effect that can be seen in ion exchange or RPLC[153]. There were also differences between the physicochemical properties i.e. the titration curves of peptides between different fractions. Titration curves provide useful information about the electrophoretic properties of the peptides present in the EC fractions. The electrophoretic properties can be translated to a parameter predicting the elution voltage for a particular peptide and therefore contribute to the validation of the MS/MS-predicted sequences. During EC-based separations, peptides elute when the electrophoretic velocity is smaller than the velocity of the flow. The electrophoretic velocity can be formulated $V_e = \mu \times E$, where µ is electrophoretic mobility and E is the potential of the electric field. In addition, µ is directly related to $q/M^{2/3}$, where q is the charge and M is the molecular weight of the peptide[154]. Theoretically, this means that peptides with larger $q/M^{2/3}$ will be eluted at lower electric fields and vice versa. This value could thus serve as a novel parameter to verify the sequence retrieved from the MS/MS database search.

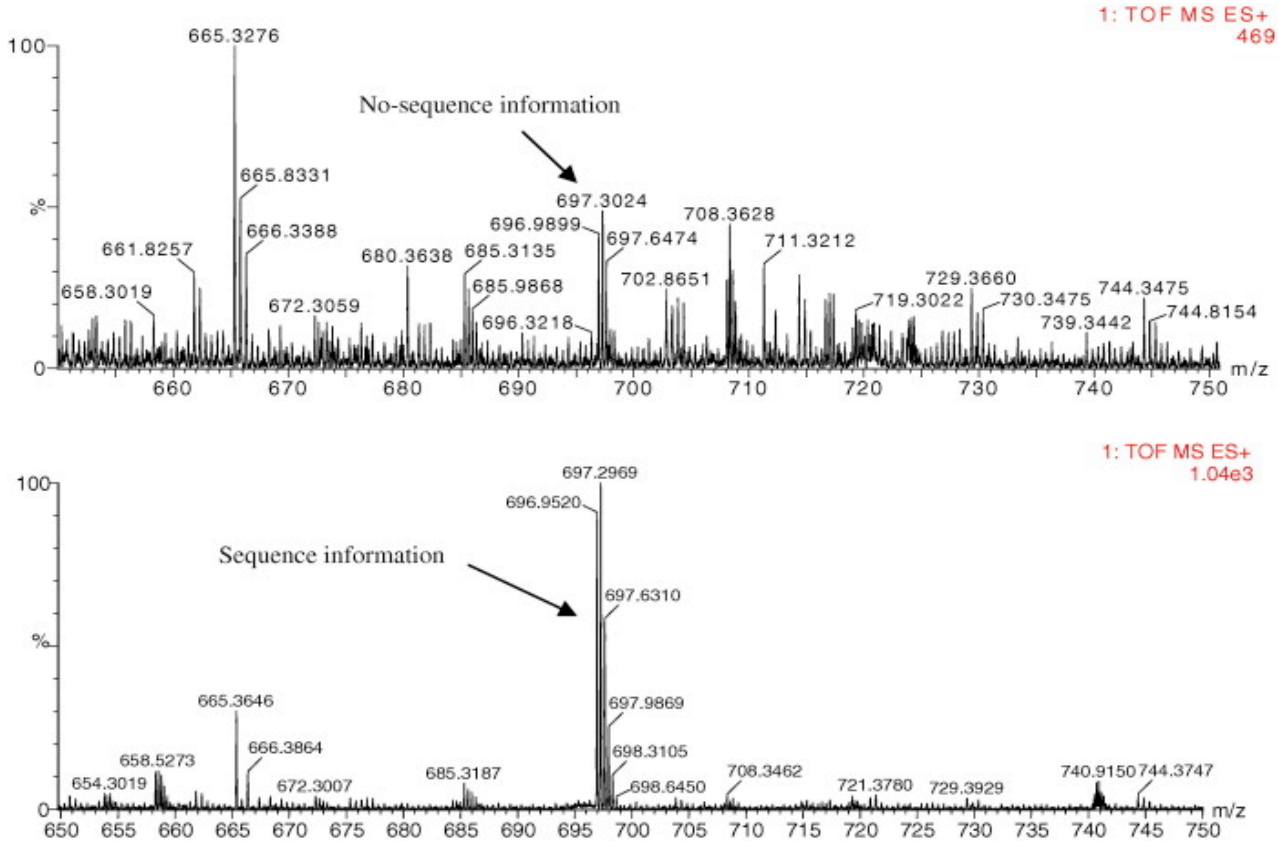


Figure 9. RPLC-MS spectra of peptides derived from 1-D and 2-D separations. The prefractionation of the complex peptide mixture with the electrocapture, improves significantly the detection of peptides.

The peptide sample analyzed only with the LC-MS/MS system resulted in the detection of 129 peptides. With the 2-D approach of EC and LC-MS/MS, we could identify 322 peptides. It was interesting to see that 240 of the peptides were only identified with the 2-D system. Although there are clear advantages of coupling an EC-based separation prior to an LC separation, some peptides were observed and sequenced only in the unfractionated sample. A similar effect has been reported for the combination of ion-exchange and RP-LC[155]. Taking these points into account, it is not surprising that the highest number of identified proteins was obtained by combining all of the identified peptides form the 1-D and 2-D separations. This combination also increases the sequence coverage of several proteins.

Table 4. Distribution of peptides in electrocapture fractions. The overall distribution of all peptides detected by MS/MS in different electrocapture fractions.

Electrocapture voltage	Total number of peptides observed	Peptides observed in this fraction only	Peptides also observed in other fractions
200	144	139	5
190	46	30	16
180	66	20	46
170	83	22	61
160	70	10	60
150	35	6	29
140	23	0	23
130	18	3	15
0	18	9	9

Table 5. Cross-occurrence of the distinct peptides between the different electrocapture fractions. The results showed that a majority of the identified peptides were identified from one of two EC fractions. This suggests that the EC technique has a good resolution power for the separation of peptides.

Spread of peptides into voltage gradient fractions	Number of peptides	Percentage (total = 322)
One fraction	239	74
Two fractions	43	13
Three fractions	16	5
Four fractions	13	4
Five fractions	8	2
Six fractions	5	2
Seven fractions	2	1
Eight fractions	0	0
Nine fractions	0	0

From the total of 369 distinct peptides, 102 distinct proteins could be identified (Figure 10). All of the proteins could be identified in the 2-D system. Only 34 proteins could be identified in the 1-D system. Even though the sample was washed extensively during the 1-D approach to remove the excess of detergent, the poor identity rate could be caused by the detergent, which suppresses the ionization rate.

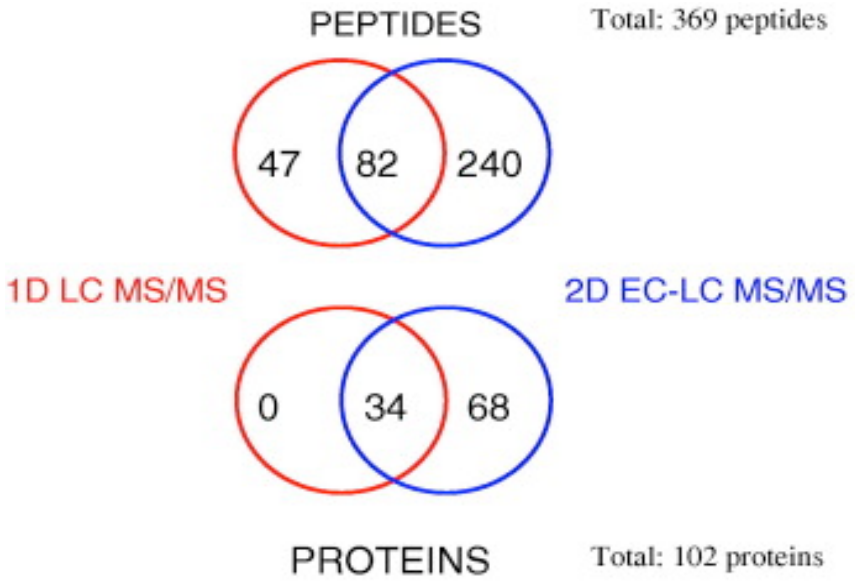


Figure 10. Number of peptides and proteins identified in the 1-D and 2-D approaches. 2-D separation technique using elecrocapture as a prefractionation increases the number of idenfied peptides and proteins. Even though many peptides were only identified with the 1-D approach, it did not have an effect on the total number of proteins identified.

The differences in identification between the 1-D and 2-D approaches follows from three aspects: First, although some peptides were present in both samples, good MS/MS data were obtained only in the fractionated sample, because the signal intensity of the precursor was higher that that in the unfractionated sample (Figure 9). Second, peptides were observed in both samples and with the same intensity but were not selected for MS/MS because of oversampling. Third, some peptides were observed and sequenced only in the fractionated sample.

As a conclusion, EC-based fractionation system can be used as a prefractionation method prior to an LC-MS/MS system to increase the detection rate of peptides in a complex mixture. The 2-D approach increases the coverage and the number of proteins identified. In addition, it removed considerable amount of detergent to improve the MS performance. Improvements to this 2-D approach are envisioned by means of an online connection of the EC separation with the LC-MS/MS step. This will decrease the handling of the sample and sample losses. In conclusion, this work demonstrates that microfluidic devices integrated into standard proteomic technologies can be used to facilitate the analysis of biological samples.

Expression pathways and networks in glomerular disease – a meta-analysis (Paper V)

The aim of this study was to compile all reported glomerular proteome and transcriptome analyses in order to obtain a comprehensive picture of field. Such an analysis allows comparison of expression signatures, both protein and transcriptome expression, of normal glomeruli and in animal models and human disease, and thus insight into mechanistic pathways and possible biomarker candidates.

Proteome analysis

From the PubMed database, a total of five glomerular proteome studies were found where the number of identified proteins was over 100. All the studies involved healthy glomeruli. Two of the studies were performed by our laboratory[156,157], two by a Japanese group[146,152] and the fifth by Mathias Mann's group[158]. The Japanese group had studied human glomeruli, the rest of the studies were performed on mouse glomeruli. Tryggvason et al[157] and Yoshida et al[146] used a 2-DE MALDI-MS approach and could identify over 200 distinct protein in their studies. Both studies were performed in or around 2005 , Yoshida et al published their data in 2005, we reported our results in 2007, but had initiated our studies in 2004.. Miyamoto et al[152] reported later a study where they could identify 6,686 proteins, representing 2,713 distinct genes using an in-solution phase isoelectric focusing as a prefractionation method combined with LC-MS/MS. All of the proteins identified by Yoshida et al were identified by Miyamoto et al. Waanders et al[158] could identify over 2,500 distinct proteins in 2009 from only 50 mouse glomeruli using a LC-MS/MS system. So far, there are not any reported glomerular proteome studies on glomerular diseases. The increase in number of identified proteins shows that the proteomics undergoes rapid development. However, it should be emphasized that older methods should not be forgotten, on the contrary the different approaches complemented each other.

In the five glomerular studies (S1), a total of 3,993 proteins representing distinct genes were identified (Figure 11). The study by Yoshida et al is not represented in Figure 11, because all of the proteins identified in that study were also identified by the same group later in the study by Miyamoto et al[152]. The subcellular location analysis (S2) according to the GO annotation revealed that approximately 50% of the proteins so far identified are cytosolic.

Membrane proteins represent 20% of the proteins. Cytosolic proteins are usually more abundant and hydrophilic, thus more suitable for proteomic analysis. Large and low abundant membrane proteins identified presents only a small percentage of the total number. The proteome of a cell is believed to contain tens of thousands of proteins - this applies also to the glomerulus. Thus, the total number of hitherto identified glomerular proteins represents only a proportion of the glomerular proteome.

Of the 3,993 glomerular proteins identified by MS, 436 (11%) are glomerulus-specific when compared with the list of glomerular-enriched proteins according He et al [139], corresponding to 30% of the total number of 1,407 highly glomerulus-specific genes. A commercial Affymetrix Mouse Genome 2.0 Array (microarray chip) contains 1,218 (87% glomerulus-enriched genes). This clearly indicates that the conventional large-scale proteomic tools are still not as sensitive as transcriptome tools to study glomerular disease. The low number of glomerular protein identifications is mainly due to the complexity of samples and sample preparation, and the technical sensitivity issues, which are affected by the large dynamic range of protein concentrations. However, new methods and technologies to deal with these problems are emerging rapidly, which raises hopes that protein profiles complementing transcriptome profiles will soon be available and more importantly give information about the possible PTMs of the proteins involved in glomerular diseases.

Transcriptome analysis

Glomerulus enriched transcriptome expression profiles have thus far only been reported for a few mouse or human glomerular disorders: (i) Reversible murine LPS-induced nephrosis analyzed 24 h after administration of LPS[159]; (ii) Lupus nephritis mice[160]; (iii) Irreversible murine Adriamycin (ADR) -induced nephrosis, a model for FSGS, analyzed 4, 7 and 14 days after induction of glomerular injury[161]; (iv) Murine db/db mice, a spontaneous model for type 2 diabetes and nephropathy, analyzed 2 and 6 months after birth[162]; (v) Human diabetic nephropathy (DN)[163] and; (vi) Human glomerular FSGS and minimal change disease (MCD)[164]. Proteinuria is a common feature for all these diseases that vary with respect to molecular pathophysiological mechanisms. Although there is extensive variability between approaches and methodological procedures of most of these studies, the results provide support for the existence of a variety of molecular mechanisms one can visualize by expression profiling and that may have potential for development of novel diagnostic methods for glomerular disease.

The transcriptomes of the seven disease models included a total of 3,966 differentially expressed genes of which 3,362 (85%) were differentially expressed in one disease only. This indicates a marked difference in gene expression in the different glomerular disease. The main problem with this kind of comparison is the low number of individual studies and the diversity of the disease and its stages of progression. Genes up- and downregulated in most of the diseases are showed in Figure 12. Many of the down-regulated genes that were involved in more than four diseases were glomerulus-specific. NEBL (nebulette), an actin binding protein involved in the organization of the cytoskeletal actin filament network and THSD7A (thrombospondin type I, domain 7A) were downregulated in five diseases. The lack of NEBL might reflect irreversible disruption of podocin actin cytoskeleton. Endothelial THSD7A interacts with extracellular laminin, fibronectin, fibrinogen and type IV collagen suggesting that the endothelium is also intact and the organization between the mesangium and capillary tufts is disruptured. NEBL and THSD7A are thus downregulated in both acute and chronic disease. These genes or proteins are not well studied, were not linked to any canonical pathway in IPA, and are involved in only few interactions according to the Ingenuity Knowledgebase or the Human Protein Reference Database. These genes/proteins should be studied more closely as they might be potential biomarker candidates for acute and chronic kidney injury. These findings also imply that basic research is still needed to gain more basic information about the functions and interactions of separate genes and proteins to be able perform informative systems biology analyses.

Most of the genes upregulated in the disease models play roles in cell growth, apoptosis and inflammation. This correlates well with the notion that constant inflammation and structural changes in the glomeruli lead to sclerosis in CKD[165]. *CDKN1A,* which encodes cyclin dependent kinase inhibitor A, is the only glomerulus specific gene, which is upregulated in at least four disease models. CDKN1A is an acute phase protein, which promotes apoptosis during acute kidney injury. Closer study of CDKN1A related pathways and networks might reveal novel knowledge of the initial pathomechanisms of different glomerular diseases.

Pathway and network analysis in glomerular disease

Global expression analyses and tools such as IPA enable us to visualize the results and show which canonical molecular pathways and networks that are active at different stages of disease progression and recovery. Canonical pathways and functional networks for the

differentially expressed genes in the disease models were compared with IPA (S5-S8). The diseases compared in this meta-analysis have different etiology, are at different stages of progression, have different pathophysiological mechanisms, and have different morphological changes in the glomeruli. This can be seen also in the comparisons of pathways and networks that are affected in the disease models. Clearly, one can state that inflammation and fibrosis responses are common to all disease models, but the important items to know are the pathways involved in the triggering and early stages of the disease. With the available data, more detailed comparisons of all the disease models at the same time is not possible due to the different nature of etiology, stages of the diseases, and low number of individuals studied.

The db/db mouse is a well-established model to study diabetic nephropathy. There was therefore a particular interest to compare the expression profiles in human diabetic nephropathy and db/db mice, as diabetic nephropathy is the commonest cause of human end-stage renal disease, and whether db/db mice serve as a good model for type 2 diabetes. At 2 months of age, eighty percent of the differentially expressed genes were upregulated. The db/db mice exhibited acute inflammatory signs in the glomeruli and development of fibrosis and angiogenesis signaling. At age 6 months, which closely resembles clinical stage 3 of diabetic nephropathy in humans, the acute inflammation has subdued. However an innate immune response prevails in the db/db mice, and the sugar metabolism pathways are down-regulated, which suggests that the ability to metabolize sugar is diminishing. Sixty three percent of the differentially expressed genes were still upregulated. Woroniecka et al[163] did a transcriptome study for type 2 diabetes nephropathy patients with stage 3-4 nephropathy, showing proteinuria and advanced glomerulosclerosis. Only thirty percent of the differentially expressed genes were upregulated. Pathway analysis showed a decrease in integrin and actin cytoskeleton signaling, and an increase in leukocyte extravasation and fibrosis. At the gene level, only 19 genes were similarly expressed between db/db mouse glomeruli at 6 months and the human DN glomeruli. IPA did not match these genes to any reasonable pathway or network. Despite the differences between the transcriptomes of db/db and human DN glomeruli, we still cannot state that the db/db mouse is not a good model for diabetic nephropathy. This was the first time a comparison was made between gene expression patterns conducted in different laboratories. The high p-values in the data of 6-month-old db/db mice indicate significant differences between the individual mice. It may also be so that the stages of mice and human do not match. Recently, Hudkins et al reported that BTBR ob/ob mice develop a rapidly progressive diabetic nephropathy more similar to the clinical and

morphological changes found in humans[166]. There is however no glomerular expression pattern published from that model yet.

Expression patterns, pathway and network analyses are more suitable for the analysis of disease progression. During the progression of a disease, the cells undergo a series of stages. Each stage in the cell can be associated with a specific transcriptome or proteome profile. With a detailed analysis of the transcriptome or proteome expression changes during disease progression, the pathomechanical pathways and networks can be determined. With the help of expression profile databases, one may then be able to accurately identify the disease profiles of patients, determine the progression status of the disease and assign a personalized therapy for patients. This will of course need extensive expression databases and the knowledge of the pathomechanisms of the disease. The expression profiles may not only be used to study the progression of the disease, but also follow up and monitor the effects of therapies and the recover process. Figure 13 illustrates the concept of expression profiling during disease progression.

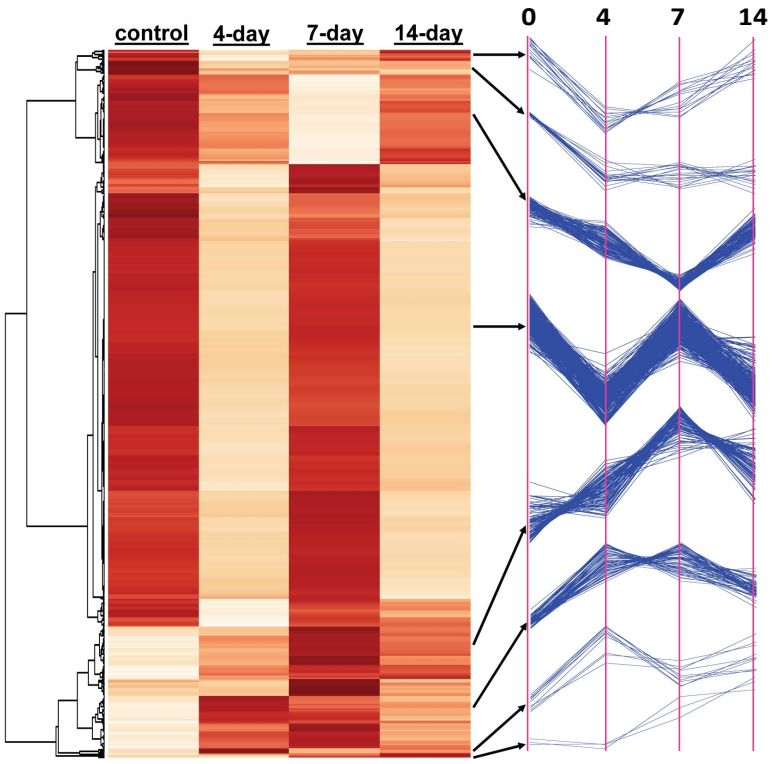


Figure 13. Illustration of transcriptome profile clustering during disease progression in ADR-induced nephrosis. The glomerular transcriptome was analyzed at three different time points during disease progression of Adriamycin-induced nephrosis in mice[161]. Day 0, prior to administration of ADR, day 4, when proteinuria had developed but abnormal histology was not apparent by light microscopy, day 7, when proteinuria was at maximum and electron microscopy revealed foot process effacement without GBM changes and day 14, when the proteinuria had diminished, but histological analyses showed FSGS changes in about 5-8% of the glomeruli. Such expression signatures could be important and useful for the development of new diagnostic methods identifying molecular pathways and expression signatures for different types of early and advanced glomerular disease.

The analysis methods and graphics for visualizing the results are still too complicated for routine clinical nephrology, but the development of software is rapid and the procedures become more and more user-friendly and clear. The analysis of transcriptome and proteome profiling actually apply better to glomeruli than to samples from many other tissues due to the easy access of pure glomeruli. We are also convinced that global gene expression profiles of isolated glomeruli will make a significant impact on future diagnosis and therapy of glomerular disease. The currently available research results on glomerular RNA and protein expression profiling are amazingly limited in number, but this is likely to change in the near future.

GENERAL CONCLUSIONS AND FUTURE PERSPECTIVES

Proteomic analyses include a wide variety of methods in many areas of applications. The selection of methods is of great importance for the outcome of a proteomic study. Large-scale proteome analyses require analytical tools, which can handle complex protein mixtures of tens of thousands of proteins.

Two-dimensional gel electrophoresis is a widely used separation method of complex protein mixtures. The method can separate thousands of proteins, is highly reproducible and can readily distinguish different PTMs. However, the drawbacks of 2-DE have facilitated the development of proteomics towards MS-based proteomics. 2-DE-based proteomics is quite complex and includes a lot of sample handling and steps before MS analysis of protein digests. Technological automation such as the microfluidic CD-based system for sample cleanup and concentration were needed to speed up the process maintaining sensitivity. The CD-based system was found to be reproducible, fast and sensitive and could replace the manual step of desalting and sample application to a MALDI target plate. The identification of mouse glomerulus proteins by 2-DE provided some new knowledge about the glomerular proteome even though the number of identified proteins remained quite low. To overcome the need of extensive sample handling in 2-DE, new technologies are developed, especially liquid phase microfluidic systems, which could be coupled directly with ESI-MS. The electrocapture technology is a relatively young technology, but has further promise. It is based on the capability to capture charged molecules in a hydrodynamic flow under an electric field. The device could perform several proteomic applications such as sample concentration, cleanup, reduction and alkylation of proteins. In this thesis, we showed that the technology could be used to separate peptide or protein mixtures and that the device could be used as a prefractionation method to identify glomerular membrane bound proteins with an LC-MS/MS system.

Proteomics have developed much during the time of my PhD studies. The focus in proteomic has changed from small-scale projects towards large-scale studies and from qualitative expression patterns towards dynamics and turnover in different cell types and under different conditions. The emphasis is on the study of the global patterns in a cell, tissue or an organism, such as protein-protein interactions and protein-nucleotide interaction networks, metabolic pathways, signaling cascades networks, and gene and protein expression patterns. Even though proteomic tools and approaches have developed much, proteomic tools still

cannot completely handle the complexity of human cells. With current transcriptome profiling methods, it is possible to analyze the whole transcriptome from a single cell. Together transcriptome and proteome expression profiling can complement each other to elucidate pathomechanisms and cascades in different diseases. With regard to the still poorly understood glomerular diseases, RNA and protein expression signatures provide a promising approach for elucidation of the mechanistic pathways that are active at the early stages of disease. Glomerular diseases render themselves particularly well to such methods as pure glomeruli can be obtained from kidney biopsies. Global expression profiles from isolated glomeruli can have a significant impact on future diagnosis and therapy of glomerular disease.

ACKNOWLEDGEMENTS

Finally I have come so far that I can sincerely thank all the people involved in making this thesis possible. The road to completion has been complicated and could not have been made without your help.

Professor Hans Jörnvall – my supervisor. Thank you for giving me the chance to be part of your lab. I sincerely thank you for believing in me and supporting me all these years. Without your knowledge and optimism this would not have been possible. I hope you can now start really enjoying your retirement days, which you certainly have deserved.

Professor Tomas Bergman – my supervisor during my first summer jobs. Thank you for introducing me to the lab facilities and giving me private lectures in the field of proteomics, for your kindness and for sharing your knowledge whenever there were problems.

Professor Karl Tryggvason – Thank you for providing me a good project and for sharing all your knowledge in the field of kidney research.

Dr. Daniel Hirschberg – Thank you for the time we shared together in the lab, for teaching me the practical things and for helping me to get started in the lab, and of course for all the serious chats we had during that time.

Dr. Juan Astorga-Wells – Thank you for teaching me the secrets of the microfluidic electrocapture device and helping me out during the projects.

Gunvor Alvelius – a mass spectrometry specialist and my teacher. I am very grateful for your help and effort with the mass spectrometry studies. With your excellence, knowledge and kindness this thesis was possible. **Carina Palmberg** – our second mass spec specialist – for your contributions during my research projects.

Ingegerd Nylander – For taking care of all the administration with care and patience.

Dr. Masatoshi Nukui and Ásmundur Oddsson – Thank you for your effort in paper II. Ási, for sharing the office during my last year, for the chats and the Matrix poker evenings.

Dr. **Liqun He** – Thank you for introducing me to bioinformatics, with the great help, understanding and incredibly fast replies to my emails during the last project.

All my co-authors who are not previously mentioned – **Magnus Gustafsson, Jesper Swedenborg, Ulf Hedin, Jenny Norlin, Börje Haraldsson** – Thank you for sharing your knowledge and all the effort in the different projects.

Professor Christer Betsholtz – Thank you for the participation and use of data in the last project. Wish you all the best in the future with new challenges.

My former roommates at Chemistry I – Thank you **Essam Refai, Susanne Vollmer, Theres Jägerbrink** and **Emma Lindahl** for the great company, discussions and relaxed atmosphere. Essam especially for the kindness, sharing and caring. Susanne for the help with the initial experiments with the electrocapture.

The present and former members of Chemistry I. **Birgitta Agerberth**, **Andreas Almlén**, **Peter Bergman, Gudmundur Bergsson, Andreas Cederlund, Ella Cederlund, Charlotte Filling, Timo Gemoll, Magnus Gustafsson, Mikko Hellgren, Waltteri Hosia, Malin Hult, Jan-Olov Höög, Jan Johansson, Michael Landreh, Jing Li, Monica Lindh, Ylva Kai-Larsen, Ermias Melles, Åke Norberg, Erik Nordling, Udo Oppermann, Anna Päiviö, Jawed Shafqat, Naeem Shafqat, Jan Sjövall, Claudia Staab, Marie Ståhlberg, Johnny Söderlund, Maria Tollin, Jan Wiberg, Xiaoqiu Wu.** Thanks.

The present members of Matrix Biology: **Lotta Andersson, Anna Domogatskaya, Bing He, Timo Janhukainen, Kan Katayama, Mark Lal, Ann-Sofi Nilsson, Juha Ojala, Jaakko Patrakka, Ljubica Perisic, Timo Pikkarainen, Elisabeth, Raschberger, Sergey Rodin, Patricia Rodriquez, Berit Rylander, Zhijie Xiao, Anne-May Österholm.** Thanks.

Thanks to my great schoolmates, friends and colleagues from Tartu. **Mathias Bäck, Kim Kalima and Matias Tenovuo.** Six years in medical school together with you guys was incredible, a lot of fun and the time just flew by.

Mom and Dad – Thanks for all the support throughout my life and for really pushing me to get this thesis done. **Kristian** – for being my best big brother and **Janne** for being my best little brother, always fun to be with you guys.

And finally I would like to thank my great home support – **Pauliina and Emma.** Thanks for having the strength and patience during the final struggle with the thesis. Now it's done and I'm coming home. Love you two.

REFERENCES

1. Wilkins, M.R.*, et al.* From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* **14**, 61-65 (1996).
2. Tyers, M. & Mann, M. From genomics to proteomics. *Nature* **422**, 193-197 (2003).
3. Hanash, S. Disease proteomics. *Nature* **422**, 226-232 (2003).
4. Gupta, P. & Lee, K.H. Genomics and proteomics in process development: opportunities and challenges. *Trends Biotechnol* **25**, 324-330 (2007).
5. Larsen, M.R. & Roepstorff, P. Mass spectrometric identification of proteins and characterization of their post-translational modifications in proteome analysis. *Fresenius J Anal Chem* **366**, 677-690 (2000).
6. Roepstorff, P. Mass spectrometry in protein studies from genome to function. *Curr Opin Biotechnol* **8**, 6-13 (1997).
7. Yates, J.R., 3rd. Database searching using mass spectrometry data. *Electrophoresis* **19**, 893-900 (1998).
8. Mallick, P. & Kuster, B. Proteomics: a pragmatic perspective. *Nat Biotechnol* **28**, 695-709 (2010).
9. Yates, J.R., Ruse, C.I. & Nakorchevsky, A. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng* **11**, 49-79 (2009).
10. Doherty, M.K. & Beynon, R.J. Protein turnover on the scale of the proteome. *Expert review of proteomics* **3**, 97-110 (2006).
11. Gygi, S.P., Rochon, Y., Franza, B.R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**, 1720-1730 (1999).
12. Griffin, T.J.*, et al.* Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae. *Mol Cell Proteomics* **1**, 323-333 (2002).
13. Mehra, A., Lee, K.H. & Hatzimanikatis, V. Insights into the relation between mRNA and protein expression patterns: I. Theoretical considerations. *Biotechnol Bioeng* **84**, 822-833 (2003).
14. Walther, T.C. & Mann, M. Mass spectrometry-based proteomics in cell biology. *The Journal of cell biology* **190**, 491-500 (2010).
15. Gstaiger, M. & Aebersold, R. Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nat Rev Genet* **10**, 617-627 (2009).
16. Tanaka, K., Waki, H. Ido, Y., Akita, S. Yoshida, Y. and Yoshida, T. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of flight mass spectrometry. *Rapid Commun Mass Spectrom* **2**, 3 (1988).
17. Karas, M. & Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical chemistry* **60**, 2299-2301 (1988).
18. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. & Whitehouse, C.M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71 (1989).
19. Mann, M., Hendrickson, R.C. & Pandey, A. Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem* **70**, 437-473 (2001).
20. Hu, S.*, et al.* Capillary sieving electrophoresis/micellar electrokinetic capillary chromatography for two-dimensional protein fingerprinting of single mammalian cells. *Analytical chemistry* **76**, 4044-4049 (2004).
21. Link, A.J.*, et al.* Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* **17**, 676-682 (1999).

22. Wolters, D.A., Washburn, M.P. & Yates, J.R., 3rd. An automated multidimensional protein identification technology for shotgun proteomics. *Analytical chemistry* **73**, 5683-5690 (2001).
23. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **312**, 212-217 (2006).
24. Caprioli, R.M., Farmer, T.B. & Gile, J. Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. *Analytical chemistry* **69**, 4751-4760 (1997).
25. Mann, M., Hojrup, P. & Roepstorff, P. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol Mass Spectrom* **22**, 338-345 (1993).
26. Roepstorff, P. MALDI-TOF mass spectrometry in protein chemistry. *Exs* **88**, 81-97 (2000).
27. Wilm, M. Principles of electrospray ionization. *Molecular & cellular proteomics : MCP* (2011).
28. Siuzdak, G.*, et al.* Mass spectrometry and viral analysis. *Chem Biol* **3**, 45-48 (1996).
29. Wilm, M. & Mann, M. Analytical properties of the nanoelectrospray ion source. *Analytical chemistry* **68**, 1-8 (1996).
30. Taylor, G. Disintegration of water drops in an electric field. *Proc. R. Soc.* **A280**, 383-397 (1964).
31. Dole, M., Mack, L.L. & Hines, R.L. Molecular Beams of Macroions. *J Chem Phys* **49**, 2240-& (1968).
32. Kebarle, P. A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J Mass Spectrom* **35**, 804-817 (2000).
33. Kebarle, P.a.T., L. From ions in solution to ions in the gas phase. *Anal Chem*, 972A-986A (1993).
34. Mamyrin, B.A.K., V. I.; Shmikk, D. V.; Zagulin, V. A. The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution. *Sov. Phys. JETP* **37**, 45-48 (1973).
35. Paul, W.a.S., H. . A new mass spectrometer without magnetic field. *Z. Naturforsch.* **8A**, 448–450 (1953).
36. Panisko, E.A., Conrads, T.P., Goshe, M.B. & Veenstra, T.D. The postgenomic age: characterization of proteomes. *Exp Hematol* **30**, 97-107 (2002).
37. Nagaraj, N.*, et al.* Systems-wide perturbation analysis with near complete coverage of the yeast proteome by single-shot UHPLC runs on a bench-top Orbitrap. *Molecular & cellular proteomics : MCP* (2011).
38. de Godoy, L.M.*, et al.* Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**, 1251-1254 (2008).
39. Anderson, N.L. & Anderson, N.G. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* **1**, 845-867 (2002).
40. Bae, S.H.*, et al.* Strategies for the enrichment and identification of basic proteins in proteome projects. *Proteomics* **3**, 569-579 (2003).
41. Molloy, M.P.*, et al.* Extraction of membrane proteins by differential solubilization for separation using two-dimensional gel electrophoresis. *Electrophoresis* **19**, 837-844 (1998).
42. Righetti, P.G., Castagna, A., Herbert, B., Reymond, F. & Rossier, J.S. Prefractionation techniques in proteome analysis. *Proteomics* **3**, 1397-1407 (2003).
43. Baggerman, G., Vierstraete, E., De Loof, A. & Schoofs, L. Gel-based versus gel-free proteomics: a review. *Comb Chem High Throughput Screen* **8**, 669-677 (2005).

44. Gygi, S.P., Rist, B., Griffin, T.J., Eng, J. & Aebersold, R. Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. *J Proteome Res* **1**, 47-54 (2002).
45. 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* **21**, 532-538 (2003).
46. Thongboonkerd, V. Proteomics. *Forum Nutr* **60**, 80-90 (2007).
47. Shapiro, A.L., Vinuela, E. & Maizel, J.V., Jr. Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. *Biochemical and biophysical research communications* **28**, 815-820 (1967).
48. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685 (1970).
49. Awdeh, Z.L., Williamson, A.R. & Askonas, B.A. Isoelectric focusing in polyacrylamide gel and its application to immunoglobulins. *Nature* **219**, 66-67 (1968).
50. Kenrick, K.G. & Margolis, J. Isoelectric focusing and gradient gel electrophoresis: a two-dimensional technique. *Analytical biochemistry* **33**, 204-207 (1970).
51. O'Farrell, P.H. High resolution two-dimensional electrophoresis of proteins. *The Journal of biological chemistry* **250**, 4007-4021 (1975).
52. Neuhoff, V., Arold, N., Taube, D. & Ehrhardt, W. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **9**, 255-262 (1988).
53. Merril, C.R., Goldman, D., Sedman, S.A. & Ebert, M.H. Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. *Science* **211**, 1437-1438 (1981).
54. Chevallet, M., Luche, S. & Rabilloud, T. Silver staining of proteins in polyacrylamide gels. *Nat Protoc* **1**, 1852-1858 (2006).
55. Gorg, A., Weiss, W. & Dunn, M.J. Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **4**, 3665-3685 (2004).
56. Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y. & Aebersold, R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 9390-9395 (2000).
57. Corthals, G.L., Wasinger, V.C., Hochstrasser, D.F. & Sanchez, J.C. The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis* **21**, 1104-1115 (2000).
58. Fountoulakis, M., Takacs, M.F. & Takacs, B. Enrichment of low-copy-number gene products by hydrophobic interaction chromatography. *Journal of chromatography. A* **833**, 157-168 (1999).
59. Righetti, P.G., Castagna, A. & Herbert, B. Prefractionation techniques in proteome analysis. *Anal Chem* **73**, 320A-326A (2001).
60. Badock, V., Steinhusen, U., Bommert, K. & Otto, A. Prefractionation of protein samples for proteome analysis using reversed-phase high-performance liquid chromatography. *Electrophoresis* **22**, 2856-2864 (2001).
61. Luche, S., Santoni, V. & Rabilloud, T. Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. *Proteomics* **3**, 249-253 (2003).
62. Oh-Ishi, M., Satoh, M. & Maeda, T. Preparative two-dimensional gel electrophoresis with agarose gels in the first dimension for high molecular mass proteins. *Electrophoresis* **21**, 1653-1669 (2000).
63. Santoni, V., Molloy, M. & Rabilloud, T. Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **21**, 1054-1070 (2000).

64. Rabilloud, T. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics* **2**, 3-10 (2002).
65. Fountoulakis, M., Juranville, J.F., Berndt, P., Langen, H. & Suter, L. Two-dimensional database of mouse liver proteins. An update. *Electrophoresis* **22**, 1747-1763 (2001).
66. Xu, X.E., Jiang, Y.H. & Shi, Q. [Tumor proteomics: the past, present and future]. *Ai Zheng* **27**, 1009-1017 (2008).
67. Yates, J.R., Ruse, C.I. & Nakorchevsky, A. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng* **11**, 49-79 (2009).
68. Shen, Y. & Smith, R.D. Proteomics based on high-efficiency capillary separations. *Electrophoresis* **23**, 3106-3124 (2002).
69. Anspach, J.A., Maloney, T.D. & Colon, L.A. Ultrahigh-pressure liquid chromatography using a 1-mm id column packed with 1.5-microm porous particles. *J Sep Sci* **30**, 1207-1213 (2007).
70. Motoyama, A., Venable, J.D., Ruse, C.I. & Yates, J.R., 3rd. Automated ultra-high-pressure multidimensional protein identification technology (UHP-MudPIT) for improved peptide identification of proteomic samples. *Analytical chemistry* **78**, 5109-5118 (2006).
71. Everley, R.A. & Croley, T.R. Ultra-performance liquid chromatography/mass spectrometry of intact proteins. *Journal of chromatography. A* **1192**, 239-247 (2008).
72. Whitesides, G.M. The origins and the future of microfluidics. *Nature* **442**, 368-373 (2006).
73. Janasek, D., Franzke, J. & Manz, A. Scaling and the design of miniaturized chemical-analysis systems. *Nature* **442**, 374-380 (2006).
74. Lion, N.*, et al.* Microfluidic systems in proteomics. *Electrophoresis* **24**, 3533-3562 (2003).
75. Astorga-Wells, J., Jornvall, H. & Bergman, T. A microfluidic electrocapture device in sample preparation for protein analysis by MALDI mass spectrometry. *Analytical chemistry* **75**, 5213-5219 (2003).
76. Park, S.R. & Swerdlow, H. Concentration of DNA in a flowing stream for high-sensitivity capillary electrophoresis. *Analytical chemistry* **75**, 4467-4474 (2003).
77. Astorga-Wells, J., Vollmer, S., Bergman, T. & Jornvall, H. Microfluidic systems and proteomics: applications of the electrocapture technology to protein and peptide analysis. *Anal Biochem* **345**, 10-17 (2005).
78. Astorga-Wells, J., Vollmer, S., Tryggvason, S., Bergman, T. & Jornvall, H. Microfluidic electrocapture for separation of peptides. *Analytical chemistry* **77**, 7131-7136 (2005).
79. Gustafsson, M., Hirschberg, D., Palmberg, C., Jornvall, H. & Bergman, T. Integrated sample preparation and MALDI mass spectrometry on a microfluidic compact disk. *Analytical chemistry* **76**, 345-350 (2004).
80. Kånge R, S.T., Werner E, et al. Detection and identification of phosphopeptides captured by IMAC and prepared for MALDI MS analysis within a CD microlaboratory. *Proceedings of the 51th ASMS Conference on Mass Spectrometry and Allied Topics, 2003.* (2003).
81. Kange, R.*, et al.* Comparison of different IMAC techniques used for enrichment of phosphorylated peptides. *J Biomol Tech* **16**, 91-103 (2005).
82. Ideker, T., Galitski, T. & Hood, L. A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet* **2**, 343-372 (2001).
83. Chuang, H.Y., Hofree, M. & Ideker, T. A decade of systems biology. *Annu Rev Cell Dev Biol* **26**, 721-744 (2010).
84. Jain, E.*, et al.* Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics* **10**, 136 (2009).

85. Geer, L.Y., *et al.* The NCBI BioSystems database. *Nucleic Acids Res* **38**, D492-496 (2010).
86. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* **28**, 45-48 (2000).
87. Clauser, K.R., Baker, P. & Burlingame, A.L. Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Analytical chemistry* **71**, 2871-2882 (1999).
88. 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* **20**, 3551-3567 (1999).
89. Yates, J.R., 3rd. Mass spectrometry and the age of the proteome. *J Mass Spectrom* **33**, 1-19 (1998).
90. Henzel, W.J., *et al.* Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci U S A* **90**, 5011-5015 (1993).
91. Colinge, J., Masselot, A., Giron, M., Dessingy, T. & Magnin, J. OLAV: towards high-throughput tandem mass spectrometry data identification. *Proteomics* **3**, 1454-1463 (2003).
92. Eng J.K., M.A.L.a.Y.J.R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry* **5**, 14 (1994).
93. Duncan, M.W., Aebersold, R. & Caprioli, R.M. The pros and cons of peptide-centric proteomics. *Nature biotechnology* **28**, 659-664 (2010).
94. Elias, J.E. & Gygi, S.P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4**, 207-214 (2007).
95. The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res* **38**, D142-148 (2010).
96. Hubbard, T.J., *et al.* Ensembl 2009. *Nucleic Acids Res* **37**, D690-697 (2009).
97. Kersey, P.J., *et al.* The International Protein Index: an integrated database for proteomics experiments. *Proteomics* **4**, 1985-1988 (2004).
98. Cote, R.G., *et al.* The Protein Identifier Cross-Referencing (PICR) service: reconciling protein identifiers across multiple source databases. *BMC Bioinformatics* **8**, 401 (2007).
99. Waegele, B., *et al.* CRONOS: the cross-reference navigation server. *Bioinformatics* **25**, 141-143 (2009).
100. Jones, P., *et al.* PRIDE: new developments and new datasets. *Nucleic Acids Res* **36**, D878-883 (2008).
101. Siepen, J.A., *et al.* ISPIDER Central: an integrated database web-server for proteomics. *Nucleic Acids Res* **36**, W485-490 (2008).
102. Ashburner, M., *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* **25**, 25-29 (2000).
103. Dennis, G., Jr., *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**, P3 (2003).
104. Kanehisa, M., *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Res* **36**, D480-484 (2008).
105. Keshava Prasad, T.S., *et al.* Human Protein Reference Database--2009 update. *Nucleic Acids Res* **37**, D767-772 (2009).
106. Shannon, P., *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).
107. Tryggvason, K., Patrakka, J. & Wartiovaara, J. Hereditary proteinuria syndromes and mechanisms of proteinuria. *N Engl J Med* **354**, 1387-1401 (2006).

108. Ballermann, B.J. Glomerular endothelial cell differentiation. *Kidney Int* **67**, 1668-1671 (2005).
109. Wartiovaara, J.*, et al.* Nephrin strands contribute to a porous slit diaphragm scaffold as revealed by electron tomography. *J Clin Invest* **114**, 1475-1483 (2004).
110. Reiser, J., Kriz, W., Kretzler, M. & Mundel, P. The glomerular slit diaphragm is a modified adherens junction. *J Am Soc Nephrol* **11**, 1-8 (2000).
111. Tryggvason, K. & Wartiovaara, J. How does the kidney filter plasma? *Physiology (Bethesda)* **20**, 96-101 (2005).
112. Avasthi, P.S. & Koshy, V. Glomerular endothelial glycocalyx. *Contrib Nephrol* **68**, 104-113 (1988).
113. Kestila, M.*, et al.* Positionally cloned gene for a novel glomerular protein--nephrin--is mutated in congenital nephrotic syndrome. *Mol Cell* **1**, 575-582 (1998).
114. Donoviel, D.B.*, et al.* Proteinuria and perinatal lethality in mice lacking NEPH1, a novel protein with homology to NEPHRIN. *Mol Cell Biol* **21**, 4829-4836 (2001).
115. Shih, N.Y.*, et al.* Congenital nephrotic syndrome in mice lacking CD2-associated protein. *Science* **286**, 312-315 (1999).
116. Boute, N.*, et al.* NPHS2, encoding the glomerular protein podocin, is mutated in autosomal recessive steroid-resistant nephrotic syndrome. *Nat Genet* **24**, 349-354 (2000).
117. Kos, C.H.*, et al.* Mice deficient in alpha-actinin-4 have severe glomerular disease. *J Clin Invest* **111**, 1683-1690 (2003).
118. Asanuma, K.*, et al.* Synaptopodin regulates the actin-bundling activity of alpha-actinin in an isoform-specific manner. *J Clin Invest* **115**, 1188-1198 (2005).
119. Michaud, J.L. & Kennedy, C.R. The podocyte in health and disease: insights from the mouse. *Clin Sci (Lond)* **112**, 325-335 (2007).
120. Alport, A.C. Hereditary Familial Congenital Haemorrhagic Nephritis. *Br Med J* **1**, 504-506 (1927).
121. Longo, I.*, et al.* COL4A3/COL4A4 mutations: from familial hematuria to autosomal-dominant or recessive Alport syndrome. *Kidney Int* **61**, 1947-1956 (2002).
122. Jais, J.P.*, et al.* X-linked Alport syndrome: natural history in 195 families and genotype-phenotype correlations in males. *J Am Soc Nephrol* **11**, 649-657 (2000).
123. Barker, D.F.*, et al.* Identification of mutations in the COL4A5 collagen gene in Alport syndrome. *Science* **248**, 1224-1227 (1990).
124. Kestila, M.*, et al.* Positionally cloned gene for a novel glomerular protein--nephrin--is mutated in congenital nephrotic syndrome. *Mol Cell* **1**, 575-582 (1998).
125. Pierson, M., Cordier, J., Hervouuet, F. & Rauber, G. [an Unusual Congenital and Familial Congenital Malformative Combination Involving the Eye and Kidney]. *J Genet Hum* **12**, 184-213 (1963).
126. Zenker, M.*, et al.* Congenital nephrosis, mesangial sclerosis, and distinct eye abnormalities with microcoria: an autosomal recessive syndrome. *Am J Med Genet A* **130A**, 138-145 (2004).
127. Boute, N.*, et al.* NPHS2, encoding the glomerular protein podocin, is mutated in autosomal recessive steroid-resistant nephrotic syndrome. *Nat Genet* **24**, 349-354 (2000).
128. Ruf, R.G.*, et al.* Patients with mutations in NPHS2 (podocin) do not respond to standard steroid treatment of nephrotic syndrome. *J Am Soc Nephrol* **15**, 722-732 (2004).
129. Weber, S.*, et al.* NPHS2 mutation analysis shows genetic heterogeneity of steroid-resistant nephrotic syndrome and low post-transplant recurrence. *Kidney Int* **66**, 571-579 (2004).

130. Caridi, G.*, et al.* Podocin mutations in sporadic focal-segmental glomerulosclerosis occurring in adulthood. *Kidney Int* **64**, 365 (2003).
131. Kaplan, J.M.*, et al.* Mutations in ACTN4, encoding alpha-actinin-4, cause familial focal segmental glomerulosclerosis. *Nature genetics* **24**, 251-256 (2000).
132. Reiser, J.*, et al.* TRPC6 is a glomerular slit diaphragm-associated channel required for normal renal function. *Nature genetics* **37**, 739-744 (2005).
133. Kim, J.M.*, et al.* CD2-associated protein haploinsufficiency is linked to glomerular disease susceptibility. *Science* **300**, 1298-1300 (2003).
134. Hinkes, B.*, et al.* Positional cloning uncovers mutations in PLCE1 responsible for a nephrotic syndrome variant that may be reversible. *Nature genetics* **38**, 1397-1405 (2006).
135. Galloway, W.H. & Mowat, A.P. Congenital microcephaly with hiatus hernia and nephrotic syndrome in two sibs. *J Med Genet* **5**, 319-321 (1968).
136. Jeanpierre, C.*, et al.* Identification of constitutional WT1 mutations, in patients with isolated diffuse mesangial sclerosis, and analysis of genotype/phenotype correlations by use of a computerized mutation database. *Am J Hum Genet* **62**, 824-833 (1998).
137. Takemoto, M.*, et al.* A new method for large scale isolation of kidney glomeruli from mice. *The American journal of pathology* **161**, 799-805 (2002).
138. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* **72**, 248-254 (1976).
139. He, L.*, et al.* The glomerular transcriptome and a predicted protein-protein interaction network. *Journal of the American Society of Nephrology : JASN* **19**, 260-268 (2008).
140. Ficenec, D.*, et al.* Computational knowledge integration in biopharmaceutical research. *Brief Bioinform* **4**, 260-278 (2003).
141. Kang, H.A. & Hershey, J.W. Effect of initiation factor eIF-5A depletion on protein synthesis and proliferation of Saccharomyces cerevisiae. *The Journal of biological chemistry* **269**, 3934-3940 (1994).
142. Lee, Y., Kim, H.K., Park, H.E., Park, M.H. & Joe, Y.A. Effect of N1-guanyl-1,7-diaminoheptane, an inhibitor of deoxyhypusine synthase, on endothelial cell growth, differentiation and apoptosis. *Mol Cell Biochem* **237**, 69-76 (2002).
143. Clement, P.M.*, et al.* Identification and characterization of eukaryotic initiation factor 5A-2. *Eur J Biochem* **270**, 4254-4263 (2003).
144. Clement, P.M., Johansson, H.E., Wolff, E.C. & Park, M.H. Differential expression of eIF5A-1 and eIF5A-2 in human cancer cells. *Febs J* **273**, 1102-1114 (2006).
145. Chen, G.*, et al.* Protein profiles associated with survival in lung adenocarcinoma. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 13537-13542 (2003).
146. Yoshida, Y.*, et al.* Two-dimensional electrophoretic profiling of normal human kidney glomerulus proteome and construction of an extensible markup language (XML)-based database. *Proteomics* **5**, 1083-1096 (2005).
147. Lenkkeri, U.*, et al.* Structure of the gene for congenital nephrotic syndrome of the finnish type (NPHS1) and characterization of mutations. *Am J Hum Genet* **64**, 51-61 (1999).
148. Sitek, B.*, et al.* Novel approaches to analyse glomerular proteins from smallest scale murine and human samples using DIGE saturation labelling. *Proteomics* **6**, 4337-4345 (2006).
149. Hirschberg, D.*, et al.* Identification of endothelial proteins by MALDI-MS using a compact disc microfluidic system. *Protein J* **23**, 263-271 (2004).

150. Jiang, X.S.*, et al.* Two-dimensional gel electrophoresis maps of the proteome and phosphoproteome of primitively cultured rat mesangial cells. *Electrophoresis* **26**, 4540-4562 (2005).
151. Takemoto, M.*, et al.* Large-scale identification of genes implicated in kidney glomerulus development and function. *Embo J* **25**, 1160-1174 (2006).
152. Miyamoto, M.*, et al.* In-depth proteomic profiling of the normal human kidney glomerulus using two-dimensional protein prefractionation in combination with liquid chromatography-tandem mass spectrometry. *Journal of proteome research* **6**, 3680-3690 (2007).
153. Delahunty, C. & Yates, J.R., 3rd. Protein identification using 2D-LC-MS/MS. *Methods* **35**, 248-255 (2005).
154. Rickard, E.C., Strohl, M.M. & Nielsen, R.G. Correlation of electrophoretic mobilities from capillary electrophoresis with physicochemical properties of proteins and peptides. *Analytical biochemistry* **197**, 197-207 (1991).
155. Kawakami, T.*, et al.* Protein identification from product ion spectra of peptides validated by correlation between measured and predicted elution times in liquid chromatography/mass spectrometry. *Proteomics* **5**, 856-864 (2005).
156. Astorga-Wells, J.*, et al.* Membrane protein identifications by mass spectrometry using electrocapture-based separation as part of a two-dimensional fractionation system. *Anal Biochem* **381**, 33-42 (2008).
157. Tryggvason, S., Nukui, M., Oddsson, A., Tryggvason, K. & Jornvall, H. Glomerulus proteome analysis with two-dimensional gel electrophoresis and mass spectrometry. *Cell Mol Life Sci* **64**, 3317-3335 (2007).
158. Waanders, L.F.*, et al.* Quantitative proteomic analysis of single pancreatic islets. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 18902-18907 (2009).
159. Sun, Y.*, et al.* Glomerular transcriptome changes associated with lipopolysaccharide-induced proteinuria. *American journal of nephrology* **29**, 558-570 (2009).
160. Teramoto, K.*, et al.* Microarray analysis of glomerular gene expression in murine lupus nephritis. *J Pharmacol Sci* **106**, 56-67 (2008).
161. Nukui M., H.L., Patrakka J., Takemoto M., Betsholtz C., Tegnér J. and Tryggvason K. Stage specific glomerular transcriptome profiling in ADR-induced nephropathy mice. *Manuscript*.
162. Norlin J., H.L., Tryggvason K. and Betsholtz C. . Glomerular transcriptome analysis and the presence of alternatively activated macrophages in diabetic nephropathy. *Manuscript*.
163. Woroniecka, K.I.*, et al.* Transcriptome analysis of human diabetic kidney disease. *Diabetes* **60**, 2354-2369 (2011).
164. Hodgin, J.B.*, et al.* A molecular profile of focal segmental glomerulosclerosis from formalin-fixed, paraffin-embedded tissue. *Am J Pathol* **177**, 1674-1686 (2010).
165. Kwoh, C., Shannon, M.B., Miner, J.H. & Shaw, A. Pathogenesis of nonimmune glomerulopathies. *Annu Rev Pathol* **1**, 349-374 (2006).
166. Hudkins, K.L.*, et al.* BTBR Ob/Ob mutant mice model progressive diabetic nephropathy. *J Am Soc Nephrol* **21**, 1533-1542 (2010).